

Universidades de Burgos, León y
Valladolid

Máster universitario

Inteligencia de Negocio y Big Data en Entornos Seguros



**TFM del Máster Inteligencia de Negocio
y Big Data en Entornos Seguros**

**Toma de decisiones basada en
procesos automatizados**

Presentado por Ignacio González Moreno
en Universidad de Valladolid 4 de julio de 2023

Tutor: Francisco J. Andújar

Resumen

Durante los últimos años la sociedad está siendo participe de cambios sumamente trascendentales en el ámbito de las tecnologías. Uno de los más destacados es, sin duda, la aparición de herramientas RPA (*Robot Process Automation*), las cuales permiten simular acciones de la misma manera que lo haría un ser humano. Esta interacción directa desde una aplicación a un sistema tiene tal alcance que está produciendo un impacto directo en nuestro desarrollo personal y profesional. Mediante estos procesos automatizados es posible acceder a datos que hasta ahora únicamente podían ser gestionados por aquellos actores que generaban esta información. El día a día está lleno de decisiones que se deben tomar, y por tanto, cualquier herramienta que aporte información y aumente las posibilidades de tomar una decisión acertada, es un motivo claro de estudio.

En este proyecto se ha estudiado como automatizar y extraer información para un caso de uso concreto (una web de venta de entradas de eventos deportivos), generando diferentes procesos automatizados que trabajan todas las fases del dato, transformándolo en información que aporte y sea de provecho para la toma de decisiones. Mediante la aplicación UiPath se lanzan una serie de robots desatendidos, los cuales recogen diariamente los datos sobre la venta de entradas del portal, y además calculan los indicadores más interesantes para el tratamiento de la información. Estos datos e indicadores, una vez almacenados, son utilizados por la aplicación Tableau para crear un panel o dashboard que facilitará la lectura, comprensión e interpretación de los datos.

Si bien este proyecto se ha centrado en un caso de uso concreto, todo el proceso llevado a cabo sería aplicable a cualquier otro ámbito online, tanto profesional como personal, sirviendo este trabajo como una guía para aquellos que pretendan profundizar en el desarrollo de procesos automatizados RPA.

Descriptores

Big Data, Inteligencia de Negocio, Automatización Robótica de Procesos, Eventos deportivos, Web Scraping, Toma de decisiones automática, UiPath, Tableau

Abstract

During the last years, we have been involved in significant changes in the field of technology. One of the most remarkable is the appearance of RPA (*Robot Process Automation*) tools that allow us to simulate actions in the same way we would do it by ourselves. This direct interaction from an application to a system under test is so powerful, having a direct impact on our personal and professional development. Through these automated processes we can access to data that would otherwise only be managed by actors that generated this information. Everyday our life is full of decisions that we need to make, and therefore, any tool that provides information and increases the chances of making a right decision is a clear reason for study

In this project, we have studied how to automate and extract information for a specific use case (a ticketing website for sports events), generating different automated processes that work all data phases, transforming it into information that contributes and allow us decision making. From UiPath application, an unattended robots are launched, which collect daily data on the portal's ticket sales, and also calculate the most interesting indicators for data processing. These data and indicators are used by Tableau application to create a panel or dashboard that will allows to read and understand the data.

Although this project has focused on a specific use case, the entire process would be applicable to any other online environment, both professional and personal. This work gives guidelines for those who intend to deep into the development of automated processes.

Keywords

Big Data, Business Intelligence, Robot Process Automation, Sports events, Web Scraping, Decission Process Automation, UiPath, Tableau

Índice general

Índice general	iii
Índice de figuras	v
Índice de tablas	vii
1. Introducción	3
Memoria	3
2. Objetivos del proyecto	7
2.1. Objetivos del software	7
2.2. Objetivos personales	8
3. Conceptos teóricos	9
3.1. RPA: Robot Process Automation	9
3.2. Web Scraping	11
3.3. DPA: Decision Process Automation	13
4. Técnicas metodológicas y herramientas	15
4.1. Metodología	15
4.2. Elección de herramientas	18
5. Aspectos relevantes del desarrollo del proyecto	29
5.1. Contexto del dato	29
5.2. Técnicas de recogida. Extracción	37
5.3. Proceso de Transformación del dato	53

5.4. Almacenamiento de los datos	58
5.5. Cloud UiPath: Orchestrator	61
5.6. Visualización de los datos	65
5.7. Análisis de los resultados	70
6. Conclusiones y Líneas de trabajo futuras	75
Apéndices	80
Apéndice A Plan de Proyecto Software	83
A.1. Planificación temporal	83
A.2. Estudio de viabilidad	84
Apéndice B Especificación de Requisitos	89
B.1. Catalogo y Especificación de requisitos	89
Apéndice C Documentación técnica de programación	93
C.1. Estructura de directorios	93
C.2. Instalación y ejecución del proyecto	96
C.3. Incidencias encontradas	96
Bibliografía	99

Índice de figuras

3.1. Componentes clave de un RPA [6].	11
3.2. Flujo típico de Web Scraping [2].	12
4.3. Fases típicas de un ciclo de vida sobre un flujo RPA [48].	16
4.4. Comunicación entre componentes para ejecución de un flujo desatendido de automatización con UiPath [50].	21
4.5. Componentes Tableau [50].	24
5.6. Visión general del estadio [39].	30
5.7. Listado de próximos partidos en <i>entradas.com</i> durante el mes de marzo de 2023	32
5.8. Selección de entradas en <i>entradas.com</i>	32
5.9. Selección de zona en <i>entradas.com</i>	33
5.10. Selección de sector en <i>entradas.com</i>	33
5.11. Información de sector en <i>entradas.com</i>	34
5.12. Nueva selección de zona en <i>entradas.com</i>	36
5.13. Nueva selección de sector en <i>entradas.com</i>	37
5.14. Detalle de extracción de tabla de partidos con UiPath	39
5.15. Flujo: Flowchart_Init.xaml	40
5.16. Detalle de recogida de atributo <i>aaname</i> en zona Fondo Goya	42
5.17. Flujo: CollectData.xaml	43
5.18. Etiquetas SVG: Path y Polygon	46
5.19. Atributos de recogida	46
5.20. Primer escenario. Extracción de atributos	48
5.21. Etiquetas disponibles segundo escenario	49
5.22. Trabajando con la funcionalidad Computer vision	50
5.23. Flujo principal de llamadas en segundo escenario	51
5.24. Detalle secuencia: Goya collect sector	52
5.25. Detalle: Goya_FONDO PLANTA 0.xaml	54

5.26. Flujo principal de cálculo de KPI	57
5.27. Flujo de calculo de indicadores para la zona Goya	59
5.28. Carpeta contenedora de procesos automatizados	63
5.29. SUT	64
5.30. Monitorización del proceso automatizado	64
5.31. Boceto inicial de visualización de datos	66
5.32. Almacenes de datos en Tableau	67
5.33. Diagrama conexiones entre almacenes de dato	68
5.34. Creación de dashboard en Tableau Desktop	70
5.35. Real Madrid - Fenerbahçe: Tendencia de asientos disponibles para un partido.	71
5.36. Real Madrid - Barcelona: Partido sin entradas a la venta.	71
5.37. Real Madrid - LDLC Asvel Villeurbanne: Tendencia incremental.	72
5.38. Tendencia incremental de asientos disponibles en un partido.	72
A.1. Planificación TFM	85

Índice de tablas

5.1. Relación KPI con variables calculadas	58
5.2. Relación de ficheros sobre Listado inicial de eventos	60
5.3. Estructura de almacén TablaEvento.xlsx	60
5.4. Relación de ficheros sobre Escaneo de sectores	61
5.5. Estructura de almacén de datos por zonas	61
5.6. Relación de ficheros sobre el almacén de KPI	62
5.7. Estructura de almacén de datos por zonas	62
A.1. Presupuesto del proyecto	86
A.2. Licencias utilizadas para el desarrollo del proyecto	87
B.1. Requisito 1: Valoración inicial del TFM. Alcance	89
B.2. Requisito 2.1: Puesta en marcha de RPA: Solución UiPath.	90
B.3. Requisito 2.2: Puesta en marcha de Tableau Desktop	90
B.4. Funcionalidad 1: Extracción del dato (Primer Escenario: SVG).	90
B.5. Funcionalidad 2: Almacenamiento. Carga y tratamiento del dato	90
B.6. Funcionalidad 3.1: Integración del proyecto en Cloud	90
B.7. Funcionalidad 3.2: Habilitar conector Tableau. Ejecución directa desde cuadro de mando (Finalmente sacado del alcance)	91
B.8. Funcionalidad 4: Creación de la visualización y Extracción de indicadores	91
B.9. Funcionalidad 5: Extracción del dato (Segundo Escenario: Computer Vision).	91
B.10. Requisito 3: Pruebas del proyecto.	91
B.11. Requisito 4: Go live.	92
B.12. Requisito 5.1: Documentación (proyecto software en Confluence)	92
B.13. Requisito 5.2: Documentación (Memoria).	92
C.1. Incidencias relacionadas directamente con el código del proyecto.	97

C.2. Incidencias relacionadas con la recogida del dato. 97

Memoria

Introducción

Es ampliamente aceptado que el Big Data se ha convertido en un elemento clave en el ámbito empresarial para la gestión de grandes volúmenes de datos. Según un reciente estudio de la empresa Smartsheet, más del 40 % de los empleados dedican aproximadamente una cuarta parte de la semana a realizar tareas manuales repetitivas que pueden automatizarse [42]. La mayoría de las organizaciones siguen dependiendo de la extracción manual de datos para recuperar la información pertinente de fuentes no estructuradas para su uso empresarial.

Esta necesidad de manejar grandes cantidades de datos ha dejado de ser algo exclusivo de grandes empresas y se ha trasladado también a un ámbito personal, donde es posible aplicarlo en nuestro día a día. Un dato disponible en la web es perfectamente transformable en información que nos permita tomar decisiones a nivel personal, como por ejemplo, afrontar la compra de un nuevo vehículo familiar, una vivienda o incluso para plantear unas vacaciones y estudiar el mercado previamente.

La recogida y gestión masiva de datos ofrecida desde Internet es posible sin grandes inversiones. Puede ser muy provechosa antes de acometer alguna acción comercial y orientar hacia cuál es el momento oportuno para hacer inversiones, indicarnos si históricamente bajan los precios de una acción en una determinada temporada del año, conocer que zonas del teatro son las más solicitadas tras recoger lo ocurrido en eventos anteriores, etc.

El contenido de este trabajo desarrolla esta idea, adoptando el proceso ETL (*Extract Transform Load*) que permite a las organizaciones mover datos desde múltiples fuentes, formatearlos, limpiarlos y cargarlos en otro almacén de datos para analizar y operar un proceso de negocio. Este proceso se ha visto en profundidad a lo largo de varias asignaturas del Máster, pero aquí

se ha incluido un punto de vista generalista y enfocado a la automatización. Se trabajan todas las fases de tratamiento del dato y se identifican las herramientas adecuadas para obtener la información que deseamos, para así transformar todos estos datos en información valiosa para una posible toma de decisiones.

Es importante destacar que el flujo de extracción del dato tratado en este proyecto es aplicable a todo recurso. Cada uno de nosotros navega por Internet desde sitios implementados en múltiples tecnologías; también utilizamos aplicaciones nativas en móviles o programas clientes de escritorio. Todas estas fuentes de datos muestran la información de muchas maneras y formas; y de todas ellas, es posible extraer su dato para un propósito específico y obtener un beneficio.

La automatización, por tanto, es fundamental, ya que si no se aplica, la extracción de datos sería imposible. Mecanizar el proceso y automatizarlo es un punto clave en todo el proyecto, con el propósito de aportar formas innovadoras de desarrollo de procesos para incrementar el conocimiento y la toma de decisiones. La base para llevarlo a cabo será un desarrollo de código, un proyecto software donde hacer posible este escenario. Con este objetivo, se han utilizado herramientas RPA (*Robot Process Automation*) que permiten realizar el proceso de *Data Scrapping* requerido, que es sin duda, la parte más destacada de todo el proyecto.

Semanalmente se producen eventos deportivos, partidos que realizan los clubes en un mismo estadio y que generan datos sobre ocupaciones de esos eventos y sus precios. Estos datos pueden transformarse en una verdadera fuente de valor para nuestros intereses. Es decir, partiendo de una información publicada en un recurso web, es posible analizar/descubrir cuáles son los metadatos (atributos web) necesarios a extraer y generar información a partir de ellos con el fin de encaminar nuestras decisiones/acciones de la manera más acertada. Por ejemplo, con esta información podríamos determinar cuáles son las zonas más solicitadas por los aficionados, qué asientos quedan vacíos semana a semana, cuánto tiempo tardan en venderse todas las entradas de un determinado sector, etc. Para desarrollar el caso de estudio donde aplicar esta idea, se ha seleccionado una web de venta de entradas sobre eventos deportivos de un reconocido equipo de baloncesto, en concreto Real Madrid de baloncesto.

Para llevar a cabo esta tarea, nos encontramos con varios retos a superar: la obtención de estos datos debe ser directamente desde los portales web de venta de entradas, desde donde intentaremos encontrar una vía para automatizar esta extracción. El dato extraído directamente de la web es un

dato no estructurado [46], es decir, un dato que no sigue ningún esquema o modelo de datos predefinido, y debemos ser capaces de tratarlo con las herramientas adecuadas. Finalmente, debemos transformar todos estos datos en información sensible para nuestro beneficio.

Entendiendo que todo dato puede convertirse en información y que la repercusión planteada aquí es aplicable a cualquier ámbito, las opciones que nos aporta el uso de aplicaciones de automatización de procesos [20] para llevarlo a cabo son incalculables. Por último, destacar la importancia del proceso desarrollado en este proyecto. Aunque el caso de uso elegido concreta un flujo de extracción y tratamiento de datos, todo el proceso realizado sería perfectamente trasladable a otros ámbitos, incluido el empresarial. Los objetivos comerciales de una empresa pueden pasar por analizar el contenido web ofrecido de cualquier sector o competencia y utilizarlo en su ciclo de toma de decisiones. El proceso desarrollado en este proyecto ofrece una guía a seguir para llevar a cabo este tipo de tareas.

El contenido del resto de la memoria se estructura de la siguiente manera:

Objetivos del proyecto, donde se explicarán los objetivos que se pretenden conseguir con la realización de este trabajo.

Conceptos teóricos, apartado donde se profundizará en los conceptos teóricos utilizados para el desarrollo del proyecto.

Técnicas y herramientas, donde se expondrán las herramientas y técnicas usadas para el desarrollo del proyecto.

Aspectos relevantes del desarrollo, donde se incidirá en los aspectos más importantes durante el proceso de desarrollo.

Conclusiones y líneas de trabajo futuras, donde se expondrán las conclusiones a las que se ha llegado después del desarrollo del proyecto, así como futuras líneas de desarrollo y posibles ampliaciones del sistema desarrollado.

Además, se incluyen los siguientes apéndices:

Plan de Proyecto Software, donde se expone el plan y la viabilidad del proyecto.

Requisitos, donde se incidirán los aspectos relacionados con las necesidades de desarrollo del proceso de automatización implementado.

Documentación técnica de programación, donde se indica la estructura usada en el proyecto y las ubicaciones a nivel técnico del código

desarrollado para los diferentes flujos automatizados que se engloban dentro de este Trabajo de Fin de máster.

Objetivos del proyecto

Este apartado explica de forma precisa y concreta cuales son los objetivos que se persiguen con la realización del proyecto, tanto desde el punto de vista del producto a construir como desde una perspectiva personal. Los objetivos concretos se especifican a continuación.

2.1. Objetivos del software

El objetivo de este trabajo de fin de máster es conseguir un flujo de automatización que dé soporte en la toma de decisiones en un ámbito de negocio específico. En este caso, se ha escogido como caso de uso la venta de entradas en eventos deportivos, siendo nuestro objetivo afrontar con la mayor información posible el proceso de compra de entradas para un determinado evento.

Para ello, se han definido una serie de preguntas bajo la cual se cimienta el trabajo:

- ¿Cualquier dato disponible en internet es útil y aporta información en la toma de decisiones?
- ¿Es posible automatizar la fase de extracción del dato, teniendo como origen un recurso web desestructurado?
- ¿A día de hoy, existen herramientas que nos permiten realizar la automatización de todo el proceso, en este tipo de escenarios?

Se pretende responder a estas preguntas a lo largo de la memoria, aportando detalle e indicando en cada proceso, lo aplicado para llevarlo a cabo. Por tanto, los objetivos que se persiguen con la realización del proyecto son:

- OBJ1 Descubrir y ser capaz de utilizar algunas de las plataformas tecnológicas disponibles en el mercado que permiten desarrollar proyectos sobre procesos automatizados en el ámbito del Big Data.
- OBJ2 Aplicar, validar y evaluar métodos de datos sobre conjuntos y flujos de datos masivos y desestructurados.
- OBJ3 Capacitación para el diseño de soluciones que permitan extraer información relevante y valiosa para el soporte a la toma de decisiones.
- OBJ4 Diseño y desarrollo de visualizaciones a partir de información extraída de datos masivos y complejos.

2.2. Objetivos personales

Desde un punto de vista personal, los principales objetivos a conseguir con la realización de este proyecto han sido: i) el desarrollo de un proyecto Big Data en todas sus fases, desde la obtención del dato hasta su visualización, y ii) la posibilidad de profundizar en el uso de algunas tecnologías que son de verdadero interés en el ámbito de la automatización, como es el caso de UiPath.

Aplicar estas técnicas en un contexto real, utilizando una importante plataforma de venta de entradas, ha supuesto todo un reto. El poder acceder a esos datos generados diariamente, tratarlos y sacar provecho de ellos, era algo impensable antes de la realización de este proyecto. Finalmente, otro de los objetivos ha sido trabajar y evaluar la aplicación Tableau Desktop para la visualización de los datos.

Sin duda, el mayor objetivo aquí no es otro que desarrollar aptitudes que puedan ser trasladables al ámbito personal/empresarial que estén en la línea de las materias estudiadas en el Máster.

Conceptos teóricos

La toma de decisiones es un rasgo fundamentalmente humano. Las empresas de hoy deben procesar una cantidad asombrosa de datos para tomar decisiones. Evaluar volúmenes extraordinarios de datos es tarea casi imposible incluso para los analistas más experimentados. Por esto, hoy en día es necesario entender por qué es importante contar con información para tomar decisiones.

Ante cualquier situación, contar con datos útiles y oportunos facilita la decisión ante una circunstancia que se presenta, incrementa la obtención de posibilidades de éxito y nos encamina para acertar en nuestras decisiones. Múltiples sistemas pueden ofrecernos ayuda en este ámbito [32], como la automatización de procesos o RPA, o el *Web Scraping*. Estas técnicas ponen a nuestra disposición datos que posteriormente podrán transformarse en información y construir un DPA (*Decision Process Automation*) donde de un vistazo se puedan evaluar los volúmenes de datos tratados, simplificando el proceso de toma de decisiones.

En las siguientes secciones se explicarán estos conceptos: en la Sección 3.1 se hablará de RPA, la Sección 3.2 tratará sobre *Web Scraping*, y finalmente, la Sección 3.3 presentará el concepto de DPA.

3.1. RPA: Robot Process Automation

La automatización robótica de procesos, o por su acrónimo en inglés, RPA [26] [55], utiliza tecnologías de automatización para imitar el comportamiento humano sobre una secuencia de acciones concretas que se repiten a lo largo de un espacio de tiempo. Estos flujos definidos de antemano pueden ser tareas administrativas dentro del ámbito laboral o personal. Acciones

como extraer datos, completar formularios y mover archivos, entre otros, son claramente procesos potencialmente automatizables. La combinación de llamadas API (*Application Programming Interface*) e interacciones de interfaz de usuario (UI) integran y realizan estas tareas repetitivas entre las aplicaciones involucradas .

Al implementar scripts que emulan procesos humanos, las herramientas de RPA completan la ejecución autónoma de diversas actividades y transacciones en sistemas de software no relacionados. Esta forma de automatización utiliza software basado en reglas para realizar actividades de procesos comerciales en un gran volumen, liberando recursos humanos para priorizar tareas más complejas. Las técnicas RPA permiten a los equipos de trabajo encargados de la toma de decisiones acelerar sus esfuerzos de transformación digital y generar un mayor retorno de la inversión (*Return On Investment* o ROI [45]) de su personal. Además, la tecnología RPA ayuda a reducir los errores y mejorar la calidad de los datos. Al automatizar esas tareas que carecen de valor estratégico, pero sí de valor táctico, se reduce el riesgo de errores humanos y se mejora la precisión de los datos.

Por tanto, las herramientas RPA son una forma de automatizar fácilmente tareas concretas y relativamente sencillas que de lo contrario deberían efectuarse de forma manual.

Una tarea RPA posee ciertas características clave:

- Debe ser sencilla, coherente y repetible.
- Es una tarea repetitiva que no precisa conocimiento técnico.
- Es un proceso existente o previsto que ofrecen una productividad, eficiencia y/o coste-beneficio significativo.
- Es una tarea que ofrece oportunidades para mejorar la experiencia del cliente.

Para ser eficaz, la tecnología RPA debe ayudar de principio a fin, desde el descubrimiento de oportunidades de automatización hasta la creación rápida de robots de alto rendimiento, y desde la gestión de miles de flujos de trabajo automatizados hasta la medición del impacto de sus automatizaciones.

Brevemente, un RPA consta de tres componentes: una primera aplicación que permite desarrollar los flujos de automatización y genera los scripts del proceso, un segundo componente que hace de parte gestora de estos flujos donde se administran y programan sus lanzamientos y una tercera parte que

son los robots que se encargan de ejecutar en último término esos procesos en la máquina destino. La Figura 3.1 muestra estos componentes claves y sitúa el flujo de una típica herramienta RPA.

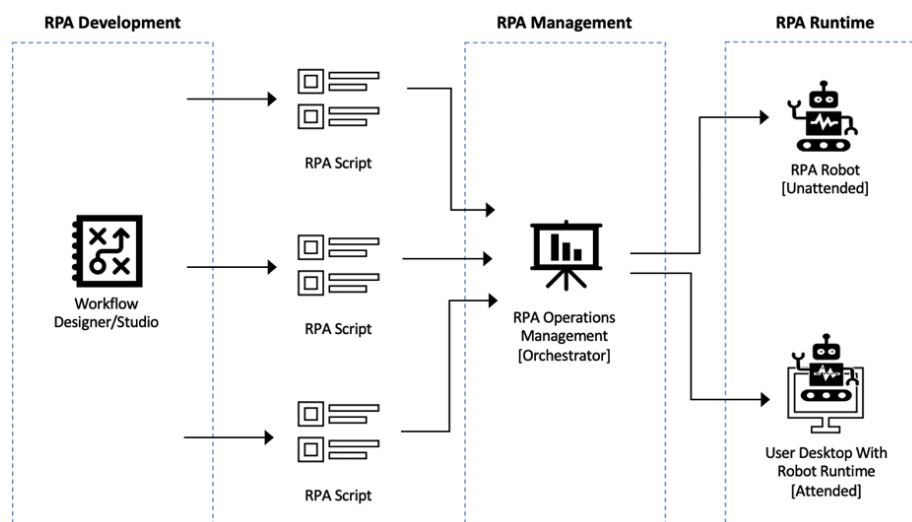


Figura 3.1: Componentes clave de un RPA [6].

3.2. Web Scraping

Se denomina *Web Scraping* [15] al proceso que permite a los usuarios extraer datos de diferentes sitios web mediante técnicas automatizadas. El servicio rastrea cuidadosamente las páginas necesarias que contengan los datos requeridos, toma la información y exporta los datos a un archivo para su posterior análisis. Uno de los formatos más populares para recibir los datos del web scraping son los archivos .csv o .xlsx, pero se pueden obtener en otros formatos.

El propósito de aplicación de esta técnica puede dirigirse a ámbitos como:

- Datos para investigadores.
- Datos sobre mercado laboral.
- Contactos de ventas y marketing.
- Seguimiento de sitios web en medios sociales.
- Seguimiento y control de precios.

Las páginas web se construyen con lenguajes de marcado basados en texto (HTML y XHTML) y suelen contener gran cantidad de datos útiles en forma de texto. Sin embargo, la mayoría de las páginas web están diseñadas para un usuario final humano y no para facilitar su uso automatizado. Por ello, se han desarrollado herramientas y programas informáticos especializados para facilitar la extracción de esos datos de las páginas web.

La Figura 3.2 muestra el flujo típico del *Web Scraping*, el cual está impulsado por robots o rastreadores web que funcionan mediante la dinámica de buscar y extraer (copiar). Su actividad se centra en extraer solo los datos específicos de ciertos sitios webs independientemente de su tecnología, para finalmente transformarlos en datos estructurados de los que poder obtener información.



Figura 3.2: Flujo típico de Web Scraping [2].

En este punto es donde enlaza el RPA con el Web Scraping y ambos conceptos se combinan para extraer datos de manera automatizada de casi cualquier sitio web. Para hacernos una idea de la utilidad y lo extendido de esta técnica, destacar que existen empresas que directamente ofrecen herramientas de *Web Scraping* como un servicio para sus clientes, como FindDataLab.com [17] o ProWebScraper.com [36].

Aspecto Legal

El *Web Scraping* es una técnica valiosa y barata para todo tipo de actores involucrados en el competitivo mercado mundial. Sin embargo, debe realizarse con respeto y responsabilidad hacia los propietarios de los datos y los administradores del sitio. Hay algunas indicaciones [35] que se deben

tener en cuenta para realizar un *Web Scraping* legal y evitar problemas en este sentido.

Antes de empezar a extraer datos, es necesario decidir qué información se quiere recibir, de qué sitios web y en qué formato. El siguiente paso es pensar cómo utilizar los datos recogidos. Aquí hay que responder a dos preguntas importantes:

1. ¿Los datos se van a publicar o serán usados únicamente para las necesidades personales o internas de la empresa?
2. ¿Puede la extracción de datos causar algún perjuicio a los propietarios de los mismos (por ejemplo, pérdidas de reputación o financieras)?

En cuanto a la primera pregunta, si los datos se extraen para uso y análisis personal, entonces el *Web Scraping* es legal y ético. Sobre la segunda, es importante asegurarse de que la información de los sitios no contenga datos personales y que la realización del proceso no infrinja las condiciones de servicio del sitio web. Si es así, el *Web Scraping* puede hacerse sin pedir permiso al propietario de los datos.

En cuanto a limitaciones técnicas, los procesos automatizados de extracción de datos pueden sobrecargar los servidores de un sitio web al solicitar datos muchas más veces que un ser humano. Es necesario cuidar esta periodicidad y no afectar de ninguna manera al rendimiento y al ancho de banda del servidor web. Si esto se sobrepasa, la mayoría de los servidores web bloquearán automáticamente la IP desde donde se realiza el *Web Scraping*, impidiendo que se siga accediendo a sus páginas web.

Así pues, un *Web Scraping* legal puede acceder a los datos del sitio web como un visitante, escanear y copiar cualquier información pública que esté disponible para el usuario normal. Pero no puede, por ejemplo, dañar la codificación del sitio, destruir obstáculos digitales seguros ni interferir de ningún modo en el funcionamiento normal del sitio web.

3.3. DPA: Decision Process Automation

El software de toma de decisiones es una aplicación informática que ayuda a personas y organizaciones a elegir y dar criterio, normalmente clasificando, priorizando o eligiendo entre una serie de opciones. Este tipo de aplicaciones, que llevan existiendo desde 1973 [16], están centradas en

elegir entre alternativas con un análisis multicriterio. La DPA ofrece una forma de optimizar la toma de decisiones automatizando el modo en que las organizaciones analizan lo que ha ocurrido y lo que está ocurriendo para predecir con mayor precisión lo que podría ocurrir [41].

Del mismo modo que se crea un “gemelo digital” para modelar los parámetros de rendimiento de un motor de avión, la DPA crea un modelo digital de una organización o de un proceso. Examina datos históricos, tendencias y decisiones; evalúa las condiciones actuales analizando los flujos de datos entrantes; y modela escenarios y resultados potenciales. Con la DPA, las organizaciones pueden examinar mejor las oportunidades en evolución y formular soluciones nunca vistas.

El siguiente paso lógico desde la automatización robótica de procesos (RPA) es el convertirse en un proceso de decisión automatizada. DPA no elimina la RPA, sino que utiliza como entrada los datos generados por ella. Fundamentalmente, es un motor de decisión de orden superior que se ejecuta en un modelo virtual de las operaciones de una empresa.

Técnicas metodológicas y herramientas

En este punto se detallan los aspectos más importantes sobre los que se cimienta el proyecto, tanto a nivel de desarrollo del proyecto con las metodologías usadas, como lo referente a herramientas para poder llevarlo a cabo. Los detalles esenciales relativos a la elección y valoración de lo técnico encuentran su justificación en este apartado.

4.1. Metodología

Diseño e implementación de un Proceso de automatización de procesos.

La tarea de diseñar una Arquitectura de Automatización [28] está estrechamente relacionada con el desarrollo de un producto software. Es necesario definir el espacio conceptual, qué capas se van a tratar, qué servicios se van a usar, cómo se va a interaccionar con el sistema destino, etc., todo ello para permitir la realización de una solución de automatización end to end (E2E).

Hay un número de actividades fundamentales requeridas para diseñar un proceso de automatización, las cuales pueden ser ordenadas de acuerdo a las necesidades del proyecto. Estas actividades se abordan en las siguientes secciones, y se necesitarán más o menos actividades dependiendo de la complejidad de la arquitectura.

Por tanto, para lograr el objetivo de finalizar con éxito el proyecto de automatización, se han seguido los siguientes puntos clave [28]:

1. Capturar los requisitos necesarios para definir una arquitectura adecuada.
2. Definir claramente las acciones a llevar a cabo dentro del proceso automatizado.
3. Soportar el desarrollo de componentes que permitan interactuar con los elementos web implicados en la automatización.
4. Evaluar complejidad para una implementación determinada de una arquitectura.
5. Hacer lo posible por habilitar la reutilización de los componentes de automatización que se desarrollen, ya que es fundamental para superar posibles cambios en tecnologías o herramientas software. (5.1)
6. Facilitar el mantenimiento y la evolución de los procesos automatizados.

De este modo podrá determinarse si es conveniente considerar una solución de automatización u otra y alcanzar la operatividad completa en nuestros flujos.

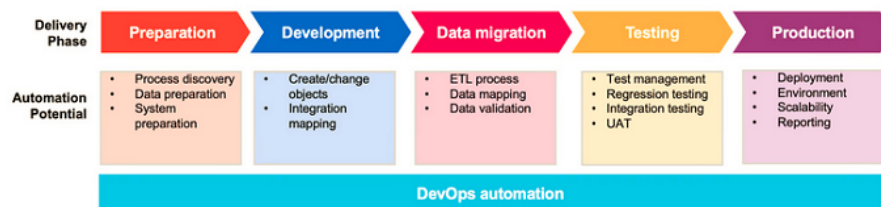


Figura 4.3: Fases típicas de un ciclo de vida sobre un flujo RPA [48].

Para el desarrollo de los flujos RPA, podemos seguir las fases típicas de desarrollo de un proyecto software. Estas fases se pueden observar en la Figura 4.3 y se detallan a continuación:

1. Fase de preparación: El primer paso de cualquier proyecto software es comprender bien cómo se realiza el trabajo.
2. Fase de desarrollo: Todo proyecto de desarrollo requiere implementar código. Creación/actualización de objetos, la corrección de código, normas de codificación, etc.

3. Fase de migración: Los desarrolladores realizan varias actividades de migración durante la fase de desarrollo. Actividades estándar como preparación de máquinas virtuales (VM), gestión de usuarios, configuraciones de políticas, actividades de elevación y desplazamiento, y configuraciones de red y seguridad son algunos ejemplos.
4. Fase de pruebas: Las pruebas consumen una gran cantidad de tiempo en cualquier proyecto de desarrollo y son una buena oportunidad para la automatización.
5. Fase de puesta en producción: tareas relativas a las acciones finales de despliegue del producto.

El proceso de Web Scraping planteado en este proyecto es muy adecuado para adoptar este escenario, ya que se trata de una tarea sencilla que se puede documentar con precisión y tiene una secuencia definida de pasos conocida de antemano. Desde aquí, concretamente en nuestro escenario, la estrategia de automatización de procesos debe cubrir:

- Qué partes de la aplicación deben tratarse.
- En qué medida y con qué periodicidad.
- Qué atributos se deben extraer.
- Donde almacenar persistentemente el dato extraído.

Dentro de los aspectos metodológicos, se incluyen también procedimientos de desarrollo. Se debe tener en cuenta que el código de automatización a crear es sensible a la interfaz, es decir, el código se verá afectado por cambios en la interfaz de usuario, siendo esto un punto clave que condiciona todo el proyecto.

El diseño de las pruebas también se ha tenido en cuenta y se han implementado a modo de guión donde pueden utilizarse diversos enfoques para crear la secuencia de acciones requerida.

Herramientas para la Gestión del Proyecto

- **Confluence:** Repositorio de documentación usado para recoger toda la información relevante del proyecto. Confluence [3] ofrece un espacio de trabajo en equipo donde alojar el conocimiento generado y registrar

la colaboración entre usuarios. Se han utilizado las páginas dinámicas que ofrece para planificar y capturar la información sobre el proyecto. Dentro de sus espacios están estructurados y organizados las distintas fases del proyecto, de modo que todos los miembros autorizados del equipo tengan visibilidad al conocimiento y puedan acceder a la información que necesiten. Toda la documentación de este proyecto es accesible desde esta referencia [24].

- **JIRA:** Aplicación web para gestión de proyectos, gestión de tareas y gestión de errores. Ofrece ayuda a equipos de todo tipo para la seguimiento de sus proyectos. Jira [4] es especialmente útil en proyectos de desarrollo de software como este, ya que proporciona tableros de Scrum [27] y Kanban [5] listos para usar como centros de gestión de tareas y requisitos, donde se asignan a flujos de trabajo personalizables. Todo el seguimiento de tareas de este proyecto es accesible desde esta referencia [25].

4.2. Elección de herramientas

En las fases preliminares del proyecto se tiene en cuenta las herramientas que se usarán para llevarlo a cabo, tanto en la parte de desarrollo del RPA como en la de visualización del dato. Existen muchas opciones en el mercado y es necesario identificar la herramienta más adecuada para el objetivo de este proyecto, sobre todo en lo que respecta a las licencias.

Para su elección, se han tenido en cuenta informes de las consultoras de mayor relevancia y ampliamente reconocidas por las empresas del sector [19], [18]. De todas ellas se ha valorado su complejidad al aplicar el escenario planteado en el TFM, sus particularidades de licenciamiento y en el caso de RPA, las tecnologías a automatizar que cubre el producto.

UiPath

La automatización de procesos robotizados puede considerarse como una pasarela, una forma de introducir la automatización y obtener ventajas de negocio con un bajo coste y prácticamente sin riesgo. El concepto es sencillo: un “robot” de software replica interacciones entre persona y sistema, automatizando lo que de otro modo serían tareas repetitivas y tediosas. RPA [34] cubre así el vacío entre la interacción manual y la plena automatización.

RPA es particularmente atractivo en nuestro caso de uso, donde una vez implementado, es posible ejecutarlo permanentemente incluso de manera

paralela haciendo uso de varios robots a la vez. En este punto, también se han valorado diferentes arquitecturas de los productos de automatización, es decir, como se van a trabajar con ellos: Cloud u On-premise, si es complejo el desarrollo de automatizaciones, si requiere de conocimientos previos en algún lenguaje de programación, si la curva de aprendizaje es asumible, etc.

La herramienta seleccionada para el desarrollo de los flujos de automatización ha sido UiPath en su versión Community. UiPath es una herramienta líder en el sector[18], y con este licenciamiento gratuito nos permite la posibilidad de crear un robot automatizado para realizar las tareas planteadas y ejecutarlas de manera desatendida.

UiPath es una plataforma que cubre todo el desarrollo de procesos automatizados y no requiere de conocimientos de código ni lenguajes de programación para su creación [52]. Permite crear flujos personalizados de nivel empresarial para ofrecer automatizaciones multiplataforma sobre tareas manuales sin crear ni una sola línea de código..

Componentes

La solución UiPath ofrece varios componentes para dar respuesta a las diferentes necesidades sobre automatización de procesos. Entre estos componentes destacan:

UiPath Studio: es la aplicación de UiPath para desarrollo de flujos de automatización. Ofrece una interfaz con una gran variedad de herramientas para diseñar los procesos de automatizaciones atendidas, desatendidas y las pruebas.

Este IDE permite la automatización de tareas a los usuarios con poca o nula experiencia de programación ofreciendo múltiples paneles para acceder más fácilmente a funcionalidades específicas y hacer más fácil el proceso de automatización. Es decir, se abstrae del código utilizando actividades que se traducen en acciones y son utilizables únicamente con un movimiento 'drag & drop'.

UiPath Automation Cloud: solución SaaS para usuarios UiPath basada en la nube, lista para configurar y utilizar inmediatamente. Automation Cloud ayuda a gestionar todo el trabajo y los recursos de automatización en un solo lugar y permite desplegar, operar y escalar procesos en solo unos segundos. A través de diferentes servicios es posible gestionar usuarios, licencias, tenants y acceder a funcionalidades del UiPath Orchestrator, una aplicación web del Cloud que permite orquestar

robots en la ejecución de procesos. UiPath Orchestrator gestiona los recursos que se utilizarán en los proyectos de automatización y que consumirán los robots, así como el acceso a los mismos. También actúa como punto de integración con soluciones y aplicaciones de terceros. En resumidas cuentas, UiPath Orchestrator nos permite crear robots, entornos, máquinas y procesos, ejecutar trabajos, crear disparadores, etc., todo desde una ubicación centralizada y segura en la nube.

UiPath Automation Cloud está disponible tanto a los usuarios con licencia Community como a los usuarios con licencia Enterprise. Además, sirve de unión entre el desarrollo (UiPath Studio) y el sistema destino de las automatizaciones. Es decir, los desarrollos pueden subirse desde Studio al Automation Cloud y desde allí, lanzarlos vía Robot (tanto atendido como desatendido) a la máquina destino donde las acciones serán ejecutadas.

Robots: En UiPath existen dos tipos de robot disponibles: robots atendidos y robots desatendidos. Los usuarios de UiPath pueden activar robots utilizando UiPath Assistant para ejecutar los procesos en sus máquinas locales. Estos se denominan robots asistidos (attended). El segundo tipo de robots se denominan desatendidos y son robots que se despliegan en máquinas independientes y no necesitan intervención humana. La ejecución de procesos para este tipo de robots se inicia y gestiona desde el Orquestador. La comunicación entre los robots y el orquestador se realiza a través de un servicio de robot que se instala en la máquina destino. Cada 30 segundos, el servicio envía datos al Orquestador y espera a que éste responda. Cuando hay un trabajo disponible para ser ejecutado, el Orquestador responde pasando la información del trabajo al robot.

Por tanto, un flujo completo sería el siguiente:

1. Desde Studio se construye un proceso de automatización.
2. Una vez completado, se publica como paquete NuGet, que no es más que un archivo ZIP con la extensión .nupkg que contiene código compilado (archivos DLL) de otros archivos relacionados con ese código.
3. Se publica en el Orchestrator (Cloud). Cada paquete representa una versión y Orchestrator los almacena todos dentro de una carpeta.

4. Para ejecutar el proyecto de automatización, los paquetes publicados deben estar asociados a carpetas. La razón de esto está relacionado con la forma en que se ejecuta la automatización.
5. Finalmente el Robot debe tener acceso/permiso a la carpeta para poder ejecutar la automatización.

En la Figura 4.4 se puede observar este flujo de comunicaciones comentado entre los componentes de UiPath

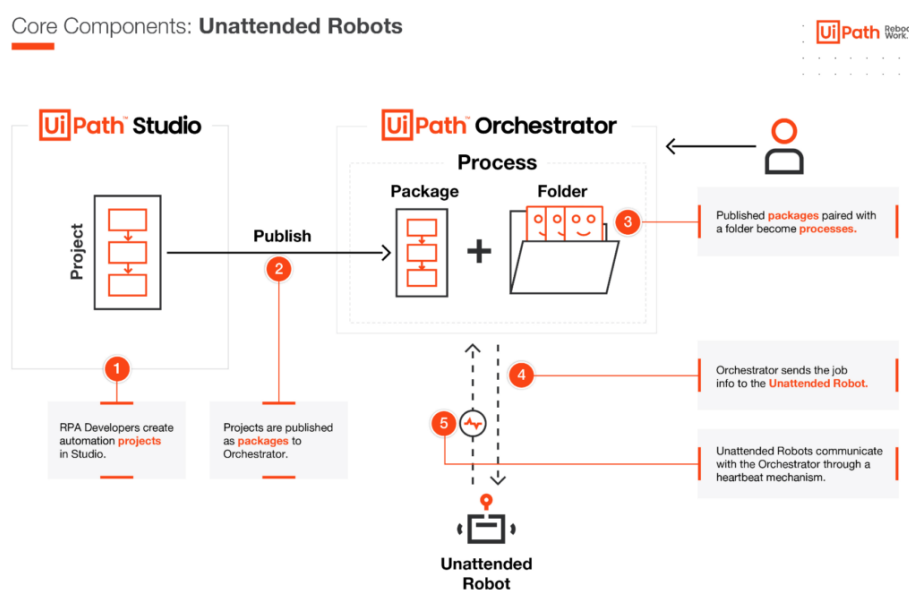


Figura 4.4: Comunicación entre componentes para ejecución de un flujo desatendido de automatización con UiPath [50].

Licenciamiento

Es destacable también que UiPath permite el uso de su aplicación sobre un licenciamiento gratuito. Para ambientes no empresariales, la versión Community es perfecta, ya que ofrece acceso completo a toda la solución y permite gestionar licencias de robot con algunas restricciones [49].

Tiene una duración ilimitada en el tiempo y se puede activar en línea. El UiPath Robot disponible asociado a la Community Edition sólo puede conectarse al UiPath Orchestrator Community Edition a través de tu propio tenant.

Ventajas y desventajas del uso de UiPath

En este apartado se detallan los pros y contras del uso de UiPath como herramienta de automatización de procesos RPA, para una mejor comprensión por parte del lector.

Entre las principales ventajas de UiPath, podemos destacar:

1. Libera a las personas de las tareas rutinarias y repetitivas. Un RPA libera a los recursos para que puedan dedicarse a tareas con mayor valor.
2. Puede automatizar tareas repetitivas que no requieran de criterios complejos de decisión humanos.
3. Fácil de implementar en los casos de uso adecuados, presenta un riesgo bajo, puesto que replica tareas manuales que ya existen.
4. No es necesario volver a formar o alterar procesos existentes.
5. Generación de procesos fiables y siempre disponibles sin interrupción.
6. Gran comunidad detras del producto.

Y entre sus desventajas:

1. No es tan adecuado para tareas que dependan de decisiones complejas o cuenten con múltiples caminos.
2. No puede solucionar procesos que no estén bien diseñados o que sean ineficientes.
3. Limitado en alcance. No es flexible a cambios y la adaptación es lenta.
4. El coste de inversión puede ser alto al principio.
5. Con la versión Community no es posible la ejecución de un robot de manera desatendida.

En cualquier caso, las ventajas superan a las contras y UiPath es la mejor de las opciones del mercado, principalmente porque la Comunidad UiPath es el colectivo de automatización más grande del mundo, conectando a clientes, partners, freelancers, entusiastas y principiantes. El último dato oficial de 2021, indicaba que la comunidad tenia más de 1,5 millones de miembros

[11]. Esto es sin duda un punto fundamental, ya que la comunidad puede resolver gran parte de las dudas surgidas durante el desarrollo, si es que no están ya resueltas en sus incontables posts existentes.

Tableau

La resolución de un problema puede realizarse desde una capa de visualización donde el dato queda representado de manera útil al usuario. La plataforma Tableau nos permite visualizar los datos de forma interactiva, ayudándonos a interpretar los mismos y haciendo que sea más fácil explorar y administrar los datos. Asimismo, permite descubrir y compartir información más rápidamente a fin de generar grandes oportunidades en los negocios. Además, Tableau permite obtener el máximo partido de los datos desde la conexión con la fuente hasta la explotación, siendo una solución eficaz, segura y flexible de análisis integral del dato.

Por ello, Tableau será la solución para desarrollar la capa de visualización final a usar durante el proyecto, desde donde ver y comprender los datos extraídos previamente. Partiendo de una implementación que no necesita conocimientos técnicos ni de programación para funcionar, se irán incorporando secciones en la visualización final hasta representar gráficamente la información deseada.

Sin ser la única empresa que ofrece este servicio [13], la herramienta ha despertado mucho interés entre los analistas de datos y estudiantes de diversos sectores formando una comunidad de más de un millón de miembros activos, que desde un foro[43] resuelven dudas y se apoyan entre sí, además de realizar eventos únicos como conferencias anuales. Al igual que ocurría con UiPath, uno de los puntos mas destacados de Tableau es la utilidad de su comunidad.

Componentes

El ecosistema Tableau ofrece varios componentes para dar respuesta a diferentes necesidades del sector. Dentro de ellos, podemos destacar los siguientes componentes.

Tableau Desktop: este componente ofrece todo lo necesario para acceder a los datos, visualizarlos y analizarlos. Con una interfaz intuitiva y bajo una aplicación de escritorio nos permite desarrollar nuestro cuadro de mando con los componentes e indicadores seleccionados mas relevantes.

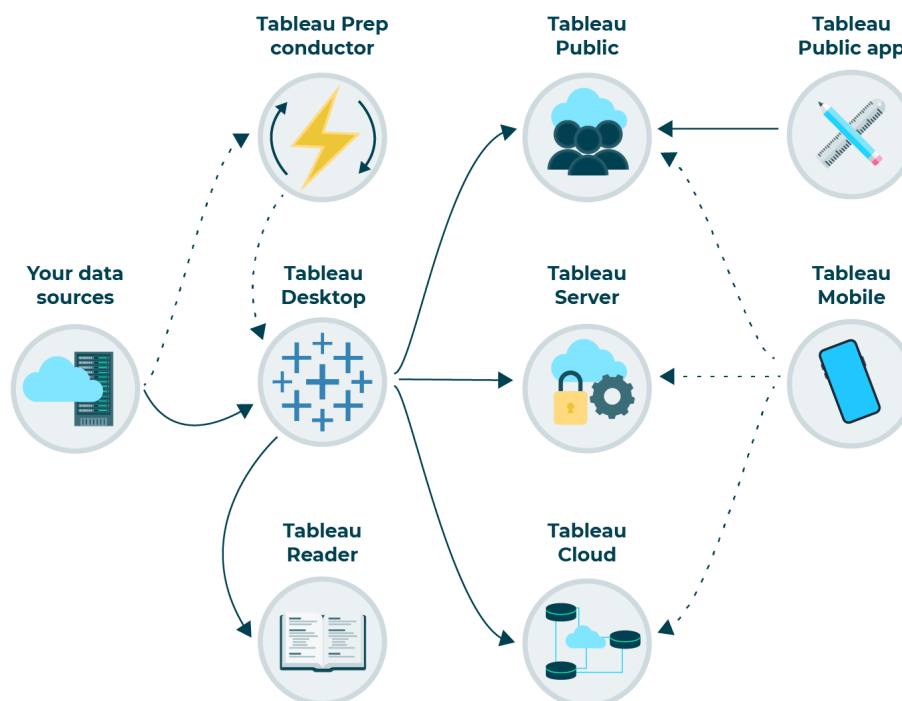


Figura 4.5: Componentes Tableau [50].

Tableau Desktop ayuda a realizar tareas de preprocesamiento muy sencillas, permitiendo importar los datos desde el almacén para realizar operaciones de unión y mezcla de datos. También Es posible modificar las formas de los datos y crear visualizaciones interactivas en cuestión de minutos a través de múltiples variaciones, pero filtrando e impidiendo crear gráficos que vayan en contra de las mejores prácticas de visualización de datos.

Tableau Public: Componente Tableau que ofrece una plataforma gratuita en línea diseñada para explorar y compartir las visualizaciones de datos creadas con el resto de la comunidad, contando con el repositorio de visualizaciones de datos más grande del mundo [1]. Desde este cloud es posible crear un porfolio propio de visualizaciones en línea para darse a conocer y ofrecer servicios a la comunidad. El poder acceder a esta enorme cantidad de visualizaciones de datos impulsa y facilita el desarrollo de las habilidades del usuario.

El ecosistema Tableau, cuyos principales servicios podemos ver en la Figura 4.5, es una opción excelente para que usuarios de cualquier nivel jueguen con los datos una vez creado el conjunto de datos.

Licenciamiento

Tableau ofrece un licenciamiento para estudiantes, haciendo posible el aprendizaje y uso de Tableau Desktop para el desarrollo de proyectos académicos. La duración de este tipo de licencia es de 12 meses. Todos los recursos de la plataforma de aprendizaje virtual también son accesibles una vez que se ha verificado la cuenta.

Destacar que también es posible usar la licencia para estudiantes [44] en un ámbito de trabajo o práctica profesional sin documentación adicional.

Ventajas y desventajas del uso de Tableau

Con Tableau, los datos se transforman en información visual de una manera muy sencilla, pero esta plataforma no está exenta de virtudes y también de carencias. Entre sus principales ventajas, podemos destacar:

1. **Alto rendimiento:** Tableau puede manejar grandes cantidades de datos, permitiendo crear diferentes tipos de visualizaciones sin que ello afecte al rendimiento de los cuadros de mando. Además, también permite al usuario establecer conexiones “en directo” con distintas fuentes de datos, como SQL, etc.
2. **Compatible con móviles:** Tableau dispone de una aplicación móvil para IOS y Android que añade movilidad a los usuarios de Tableau y les permite mantener las estadísticas a su alcance en todo momento. Gracias a ello, Tableau tiene un gran porcentaje de usuarios que utilizan activamente el móvil. La aplicación móvil soporta prácticamente todas las funcionalidades de la versión de escritorio y online.
3. **Comunidad de Tableau:** la comunidad de Tableau cuenta con usuarios y desarrolladores que atienden las consultas rápidamente, es atractiva y resolutiva. Dispone de diversos y completos recursos online, guías, formación, foros online, etc.
4. **Fácil de actualizar:** actualizaciones soportadas y fáciles de implementar.

5. **Bajo coste:** Tableau es relativamente una solución de bajo coste en comparación con otros homólogos de Big Data como Qlik y PowerBI [7].
6. **Facilidad de uso:** dispone de una interfaz de usuario sencilla, permitiendo trabajar utilizando acciones 'drag and drop', entre otras, que son intuitivas y fáciles de aprender.

En cuanto a sus inconvenientes, destacan:

1. **No hay actualización automática de los informes.** Por ello, requiere esfuerzo extra, ya que el usuario no tiene una opción automática para actualizar sus informes con la ayuda de programación. Por lo tanto, se requiere algún esfuerzo manual para actualizar los datos en back-end.
2. **No es una solución completa:** aunque el software Tableau es fácil de usar para aplicaciones de Inteligencia de Negocio, no proporciona ninguna plataforma para desarrollar aplicaciones analíticas que se puedan compartir ampliamente basados en un estándar.
3. **No hay control de versiones:** una vez que los cuadros de mando y los informes se publican en el servidor no se puede volver a los niveles anteriores de datos en Tableau. No es posible volver atrás y recuperar versiones antiguas de una determinada visualización.
4. **Conocimientos de SQL:** es necesario tener conocimientos de SQL para crear conjuntos de datos ricos y complejos a partir de múltiples fuentes de datos.
5. **Visualizaciones complejas requiere de mucha formación y recursos:** el uso básico de la aplicación no exige de muchos conocimientos. El potencial de visualización de la herramienta es casi ilimitado, pero la curva de aprendizaje es increíblemente pronunciada para los usuarios que no son analistas. Llegar a conocer las capacidades de la herramienta y aprovechar al máximo la funcionalidad de la herramienta es tarea difícil sin una formación exhaustiva.
6. **Resoluciones de pantalla en Tableau Dashboards:** puede haber ocasiones en las que la resolución de pantalla del creador de Tableau sea diferente a la del usuario final, siendo posible que el diseño de los dashboards se vea alterado y se distorsione ligeramente.

A pesar de estos inconvenientes, con la herramienta Tableau se pueden realizar todas las operaciones necesarias para desarrollar el proyecto de manera ágil. Tiene perfecta integración con UiPath, poniendo a nuestra disposición conectores que permiten obtener directamente los datos a visualizar en Tableau desde los procesos automatizados de UiPath. Además, permite creación de indicadores, preparar tendencias sobre los datos o insertar links a urls. Por todos estos detalles y por su facilidad de uso, la convierten en la plataforma adecuada para generar las visualizaciones de este proyecto.

Aspectos relevantes del desarrollo del proyecto

En este apartado se exponen los aspectos más importantes del desarrollo del proyecto y los detalles de mayor relevancia dentro de las fases de análisis, diseño e implementación.

5.1. Contexto del dato

Contexto Pabellón

El primer paso para definir un alcance dentro del proyecto es concretar donde se va a hacer uso de él. Es decir, se ha definido un origen del dato, seleccionado la web que se va estudiar. Concretamente, se ha tomado como caso de uso la venta de entradas del equipo de baloncesto del Real Madrid [30], cuyos partidos como equipo local se celebran en el Palacio de los Deportes de la Comunidad de Madrid — WiZink Center. La Figura 5.6 muestra una visión general del estadio WiZink Center.

Sin embargo, antes de continuar debemos valorar la envergadura de lo que implica este proyecto, siendo necesario un estudio sobre la estructura a tratar: el pabellón. Para poder acometer con éxito nuestro proyecto, será necesario conocer datos como número total de aforo, zonas existentes y su distribución en el campo, entradas totales por sector, precio de cada una de esas zonas, etc.



Figura 5.6: Visión general del estadio [39].

Realizando una primera búsqueda sobre el aforo de este recinto, los resultados no son claros:

- Según la página oficial del Real Madrid de baloncesto el aforo es de 15.000 espectadores [40].
- Según Wikipedia, el aforo de este pabellón llega hasta los 13.109 en los encuentros de baloncesto [53].
- Según revistas especializadas, la cancha del Real Madrid dispone de 12.900 espectadores [21].
- Según el promotor del recinto (WiZink), la cifra es de 13.478 personas sentadas en grada [54].

Por tanto, dada la disparidad de cifras, para la elaboración de este proyecto se ha tomado como referencia a la hora de realizar cálculos de ocupación y recaudación la cifra dada por el promotor del recinto, la entidad bancaria en línea WiZink. Es decir, consideraremos que **el aforo es de 13.478 espectadores**.

Los indicadores clave extraídos del contexto pabellón y que serán utilizados para realizar las métricas mostradas en el proyecto son:

- A la hora de realizar cálculos de ocupación y recaudación, se tomará como **cifra de referencia un aforo de 13.478 espectadores**.
- El **número de abonados** [12] al Real Madrid de baloncesto se sitúa entre **6.000 y 8.000** y el precio de esos abonos puede consultarse en la web de abonados [39].
- Para cada partido se recogen **datos de venta de entradas/localidades disponibles**, pero se parte de una cifra de referencia inicial que calculamos así:
 $7.000 \text{ abonos aprox.} / 4 \text{ zonas} = \mathbf{1750 \text{ abonos por zona}}$.
- Por tanto, el **total aforo por partido**: $13.478 - 7.000 \text{ abonos} = \mathbf{6.478 \text{ localidades}}$ a la venta cada partido.

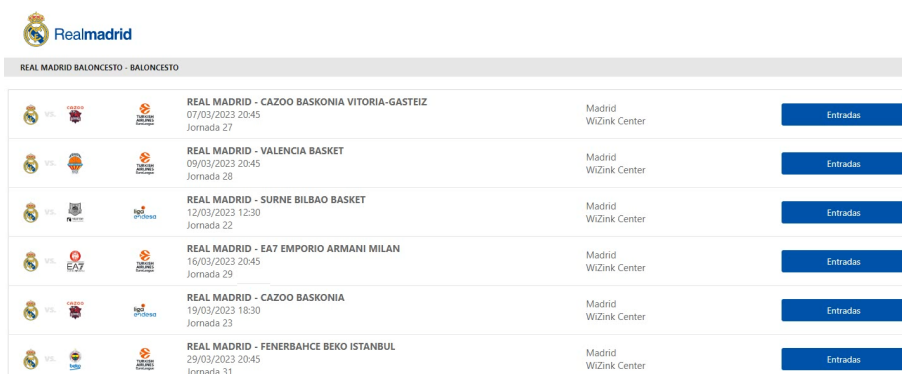
Un desglose completo de toda este análisis e información recopilada entorno a en este contexto puede leerse con mayor detalle en la web del proyecto [23].

Portal de Venta de entradas

El acceso a cómo funcionan hoy en día la industria de los eventos y el negocio de las entradas, en sus aspectos fundamentales es algo complejo [47]. Aunque la historia de la venta y distribución de entradas para espectáculos es tan antigua como el tiempo (piense en el Coliseo de Roma). La diferencia entre entonces y ahora es que los propios recintos ya no se ocupan de la parte operativa de la venta de entradas. TicketMaster o entradas.com son plataformas que ofrecen servicios de venta de tickets que empresas como el Real Madrid usan para vender las entradas en su nombre.

Hasta la llegada de estos actores la distribución, las operaciones y la gestión del inventario eran tareas extremadamente difíciles para el local (o empresa que usa el local) que no poseía tecnología ni software. Por eso, con su propio software de venta, este tipo de portales se han convertido en la vía de entrada para la adquisición de cualquier asiento en un evento, sea deportivo o no.

Con este contexto situamos la fuente de la que extraeremos el dato a tratar en este proyecto. El origen del dato, que es la base del proyecto, se















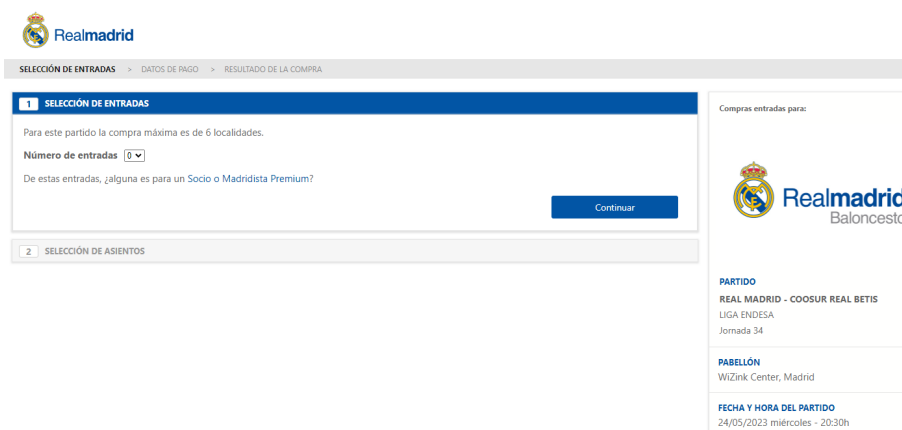
REAL MADRID BALONCESTO - BALONCESTO			
		REAL MADRID - CAZOO BASKONIA VITORIA-GASTEIZ 07/03/2023 20:45 Jornada 27	Madrid WiZink Center Entradas
		REAL MADRID - VALENCIA BASKET 09/03/2023 20:45 Jornada 28	Madrid WiZink Center Entradas
		REAL MADRID - SURNE BILBAO BASKET 12/03/2023 12:30 Jornada 22	Madrid WiZink Center Entradas
		REAL MADRID - EA7 EMPORIO ARMANI MILAN 16/03/2023 20:45 Jornada 29	Madrid WiZink Center Entradas
		REAL MADRID - CAZOO BASKONIA 19/03/2023 18:30 Jornada 23	Madrid WiZink Center Entradas
		REAL MADRID - FENERBAHCE BEKO ISTANBUL 29/03/2023 20:45 Jornada 31	Madrid WiZink Center Entradas

Figura 5.7: Listado de próximos partidos en *entradas.com* durante el mes de marzo de 2023



SELECCIÓN DE ENTRADAS > DATOS DE PAGO > RESULTADO DE LA COMPRA

1 SELECCIÓN DE ENTRADAS

Para este partido la compra máxima es de 6 localidades.


Número de entradas

De estas entradas, alguna es para un Socio o Madridista Premium?

[Continuar](#)

2 SELECCIÓN DE ASIENTOS

Compras entradas para:

 **Real Madrid**
Baloncesto

PARTIDO
REAL MADRID - COOSUR REAL BETIS
LIGA ENDESA
Jornada 34

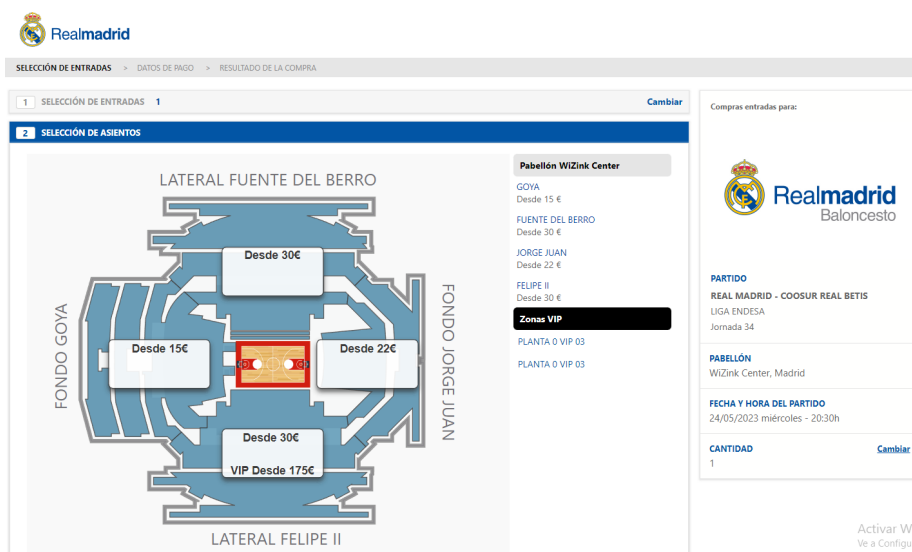
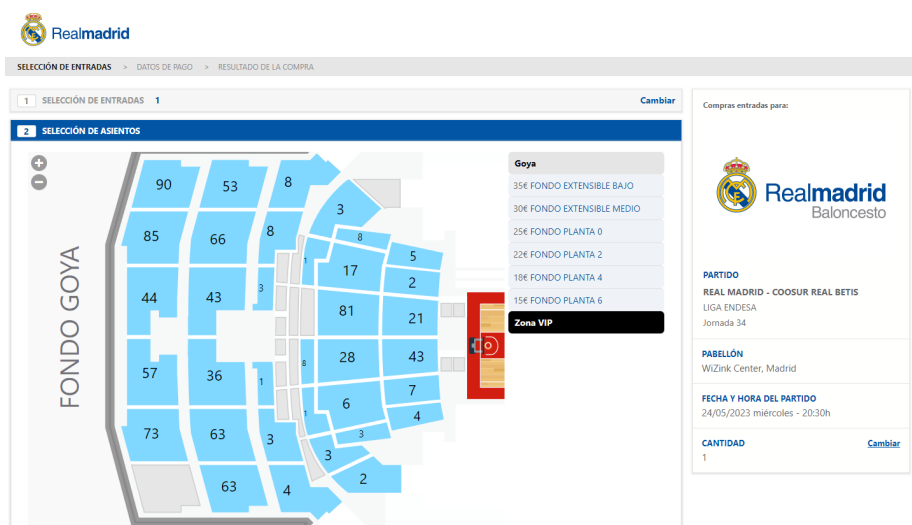
PABELLÓN
WiZink Center, Madrid

FECHA Y HORA DEL PARTIDO
24/03/2023 miércoles - 20:30h

Figura 5.8: Selección de entradas en *entradas.com*

recoge del sitio web de venta de entradas del Real Madrid de Baloncesto [30]. Este sitio ofrece como página de entrada una lista de todos los partidos planificados y que como equipo local, el Real Madrid va a jugar en el WiZink Center[40], tal y como se muestra en la Figura 5.7. Todos los eventos mostrados en la lista corresponden al mes en curso y en cada uno de ellos se permite iniciar un proceso de compra independiente de localidades haciendo click en el correspondiente botón “Entrada”.

La siguiente pantalla, mostrada en la Figura 5.8, solicita al comprador información para poder continuar. Debemos introducir un número de entradas a adquirir y pulsar “Continuar” si queremos pasar a la siguiente fase del proceso de compra. En este punto se ofrece al comprador una vista completa del alzado del pabellón (Figura 5.9). Aquí todavía no es posible la selección

Figura 5.9: Selección de zona en *entradas.com*Figura 5.10: Selección de sector en *entradas.com*

del asiento, antes se debe acceder a una de las cuatro zonas disponibles del campo.

Una vez dentro de la zona seleccionada, aparece en detalle cada uno de los sectores disponibles, tal y como se muestra en la Figura 5.10. Aunque el proceso de compra no se ha finalizado aún, y faltaría una última acción para seleccionar el asiento concreto donde completar la compra de la entrada,

es aquí donde finalizamos nuestra exploración, ya que hemos encontrado el dato que queremos conseguir, quedando sólo extraerlo. Es en este punto del proceso donde la página web de venta de entradas nos permite capturar datos relevantes. El método para hacerlo consiste en posicionar el ratón por encima de cada uno de los sectores, de la manera mostrada en la Figura 5.11.

Los principales datos que podemos capturar son:

1. Nombre del sector.
2. Número de asientos disponibles del sector.
3. Precio de las localidades dentro del sector.

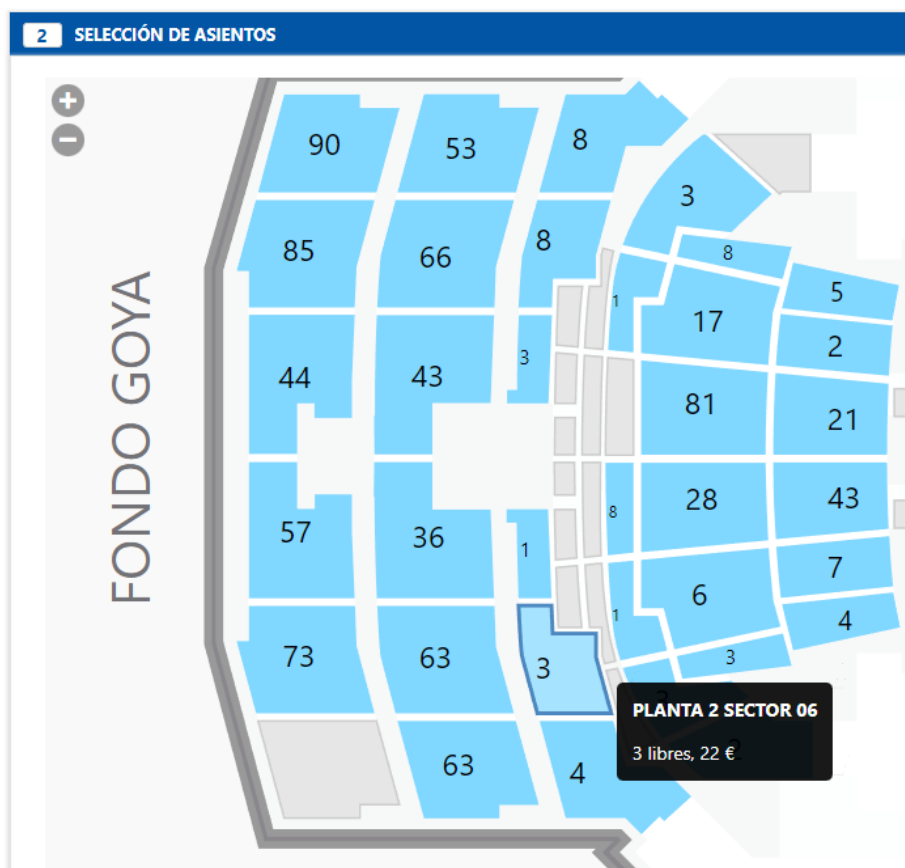


Figura 5.11: Información de sector en *entradas.com*

Estos datos extraídos directamente de un sitio web se denominan “no estructurados”, lo cual significa, de forma resumida, que se trata de conjuntos de datos cuyo origen no es un formato de base de datos estructurado y por tanto, no provienen de un almacén previo. Los datos no estructurados tienen estructura interna, pero no están predefinidos por modelos de datos. En comparación con el resto de datos estructurados, el volumen de datos con esta denominación supone el 80 % del total y el porcentaje no deja de crecer, tendiendo a duplicarse cada año [14].

Algunas particularidades encontradas en la exploración de esta web, que también se han tenido en cuenta a lo largo de este proyecto, son las siguientes:

- No requiere de credenciales de acceso para iniciar un proceso de compra.
- No contiene CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*), lo cual es ideal para lanzar procesos automatizados.
- Se detecta que la propia plataforma *entradas.com* no ofrece una única vía para acceder al canal de compra de entradas, o mejor dicho, es posible encontrar otros accesos [31], que permiten al aficionado iniciar un proceso de compra.

Cambios inesperados

Cuando pensamos en un sitio web entendemos que la apariencia, las tecnologías que corren por debajo y los pasos que realizamos hasta llegar a un punto, “siempre” van a ser los mismos. Se ha comprobado durante el desarrollo de este proyecto que esta premisa no siempre se cumple.

Durante la recogida diaria del dato, concretamente el día 26 de enero de 2023, nos encontramos que el proceso automatizado ha fallado. Revisando logs y evidencias se observa que el problema no está asociado al proceso desarrollado, sino a la propia web de venta de entradas. Por tanto, se realiza de nuevo el proceso de compra manualmente y se comprueba que, efectivamente, el servicio online de venta de entradas ha cambiado de un día para otro y la manera de presentar la información en pantalla al usuario también. La web, siendo la misma, se ha actualizado. Los pasos iniciales como el listado de eventos (Figura 5.7) no han cambiado, pero al acceder a la página desde donde son visibles las zonas disponibles del pabellón, la manera de presentar el recinto de baloncesto al cliente es diferente ya que han utilizado otra tecnología para ofrecer al usuario la misma información.

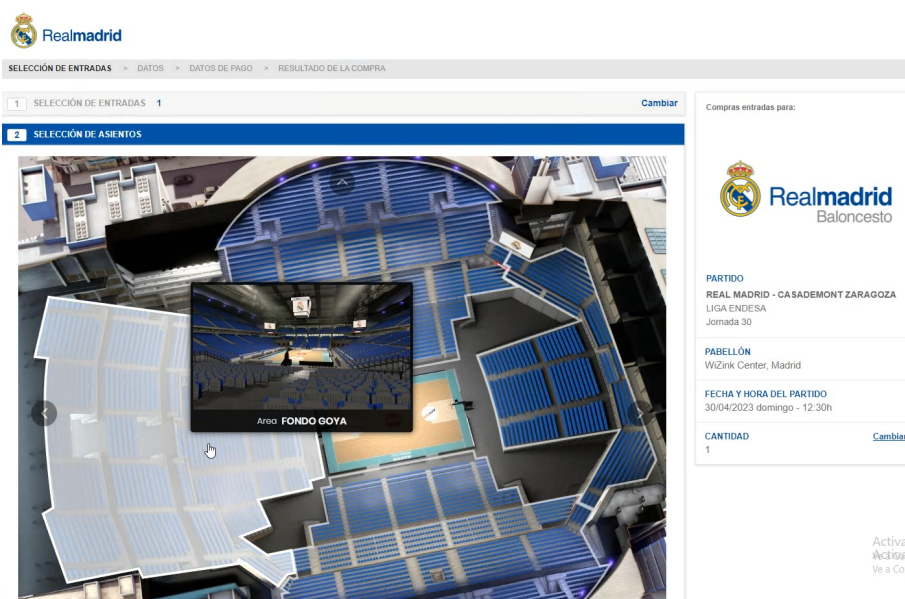


Figura 5.12: Nueva selección de zona en *entradas.com*

Como se puede observar en la Figura 5.12, la página web ha modificado la manera de mostrar la vista global del pabellón al cliente. Además, si como usuarios accedemos a un sector concreto del pabellón, también podemos observar claramente más cambios en la web (Figura 5.13).

Sin duda, este cambio tecnológico llevado a cabo por parte del portal de venta ha supuesto un impacto en el desarrollo del proyecto que es necesario valorar. El cambio no es trivial y es necesario un nuevo análisis técnico sobre como abordar y plantear posibles soluciones.

Esto supone un riesgo para el proceso de automatización que no estaba previsto y han sido necesarias adaptaciones de los procesos inicialmente desarrollados para la recogida de datos. La página de venta de tickets *entradas.com* ha sido modificada de un día para otro y a partir de este momento, es necesario monitorizar con más detalle la web, para tener un mayor control sobre el impacto que nos pueda ocasionar en el proyecto.

Surgen en este punto algunas preguntas: ¿Cómo es posible que un portal de entradas cambie de un día para otro la manera de ofrecer al cliente un flujo de venta de entradas? ¿Cuál es el propósito por el que realizan este tipo de acciones?

Una posible respuesta es que el cambio de tecnología responda a una estrategia para protegerse de ataques antirobots. Los ticket bots [37] son

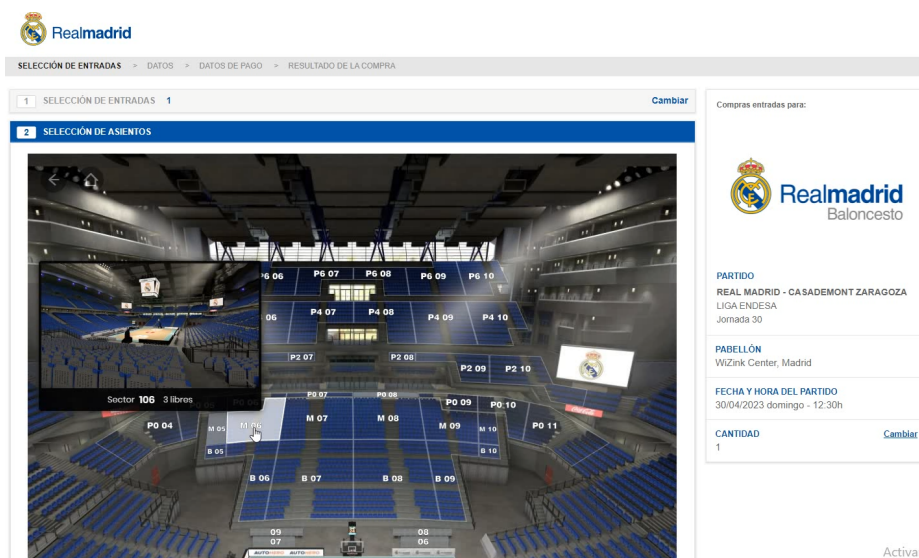


Figura 5.13: Nueva selección de sector en *entradas.com*

programas informáticos diseñados para ayudar a comprar entradas mediante tareas automatizadas, como la búsqueda de detalles sobre precios, la comprobación del inventario en busca de asientos nuevos o la compra y reventa de entradas. Los bots de reventa suelen imitar el comportamiento de los usuarios humanos, sólo que más rápido y en mayores volúmenes, lo que significa compra de entradas de manera desleal.

Los “scalper bot”[38] también son un problema enorme en el mundo de la venta de entradas, ya que representan casi el 40 % de todo el tráfico de los sitios web de venta de entradas. Son una de las principales razones por las que muchas veces es imposible conseguir entradas para todo tipo de eventos en directo [33].

Las medidas aplicadas para solventar este contratiempo y cómo se ha conseguido llegar a capturar los mismos datos en este nuevo escenario se detallan en la siguiente sección, donde se habla sobre las técnicas utilizadas en ambos casos.

5.2. Técnicas de recogida. Extracción

Una vez realizado el estudio sobre todo lo que afecta al proceso de adquisición del dato, junto con todas las entradas y salidas de páginas necesarias

para llevarlo a cabo, se han identificado los aspectos más interesantes sobre como técnicamente se ha accedido al dato.

Cuando se genera un flujo de automatización con la herramienta UiPath Studio, se crean una serie de ficheros de extensión *.xaml*, un formato de lenguaje extensible basado en XML usado en el desarrollo de aplicaciones y optimizado para describir gráficamente interfaces de usuario visuales desde un punto de vista gráfico. Estos ficheros se utilizarán para almacenar la información de las actividades que se añaden a los flujos de automatización creados.

A continuación, se detallan las técnicas de recogida utilizadas en los primeros pasos del proceso de compra de entradas, que ya se ha identificado que en ningún caso cambian de tecnología.

Flowchart_Init.xaml

Partiendo de la exploración anterior, el proceso de automatización se inicia con la apertura de una página web (html) de compra de entradas [30] para generar un listado inicial de eventos (ver Figura 5.7).

Estas acciones se desarrollan en el diagrama de flujo Flowchart_Init.xaml, cuyo punto de entrada es la apertura del navegador y finaliza guardando la lista de los partidos en un fichero Excel. Las actividades, que son los bloques de construcción utilizados en UiPath Studio para diseñar proyectos de automatización, extraen información sobre los partidos de baloncesto disponibles para comprar entradas para después almacenarla en un archivo.

A continuación, se enumeran secuencialmente las tareas según se han desarrollado en el proceso automatizado:

1. **RecoverCurrentEvents:** recupera todos los eventos actuales (partidos) disponibles en la web. Todos los datos útiles deben almacenarse en una tabla de datos. Contiene llamadas a las siguientes actividades:
 - 1.1 **Open browser:** acción de abrir el navegador con una url dada.
 - 1.2 **Extract Table Data:** extrae en un objeto *dataTable*, un tipo de variable que puede almacenar datos en formato filas y columnas, todos los datos relevantes para cada evento/partido: datos del partido, partido, fecha, hora, jornada, competición, rival, url a página de venta y lugar del evento.

En la imagen 5.14 se observa como UiPath reconoce los diferentes campos de la pantalla para facilitar su extracción.

- 1.3 **For Each Row in Data Table:** transforma con operaciones las cadenas de datos extraídas, para unificar y ofrecer un dato limpio para su almacenamiento en diferentes filas dentro del dataTable.

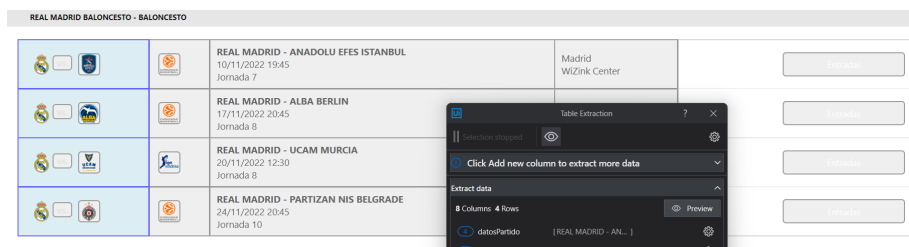


Figura 5.14: Detalle de extracción de tabla de partidos con UiPath

2. **Upload TablaEvento File:** flujo que considera que si existe un fichero previo de eventos, y por tanto, hay que actualizarlo con los nuevos valores (partidos) recogidos de la web.
3. **Assign CurrentEvents File:** flujo que considera que no existe fichero previo de eventos. Almacena todos los currentEvents extraídos a un objeto dataTable.
4. **SaveOnExcel:** conjunto de actividades que trasladan el objeto dataTable a un fichero Excel almacenado en el equipo desde donde se ejecuta la automatización. El fichero Excel se llama TablaEvento.xlsx y la ruta donde se almacenan los archivos creados es:


```
"%AppData%/TFM_ScrapingData/")+"FileName.xlsx"
```
5. **CollectData:** La última actividad de este primer flujo invoca a otro flujo encargado de la recopilación de datos de cada uno de los partidos recuperados.

La Figura 5.15 muestra las actividades de este flujo. Este flujo es el punto de entrada del proceso automatizado, es decir, el primer flujo que siempre se va a ejecutar cuando se ejecute el proyecto completo.

CollectData.xaml

El siguiente paso del proceso de compra consiste en ir a la página principal de venta de entradas de cada uno de los partidos de la lista recuperada

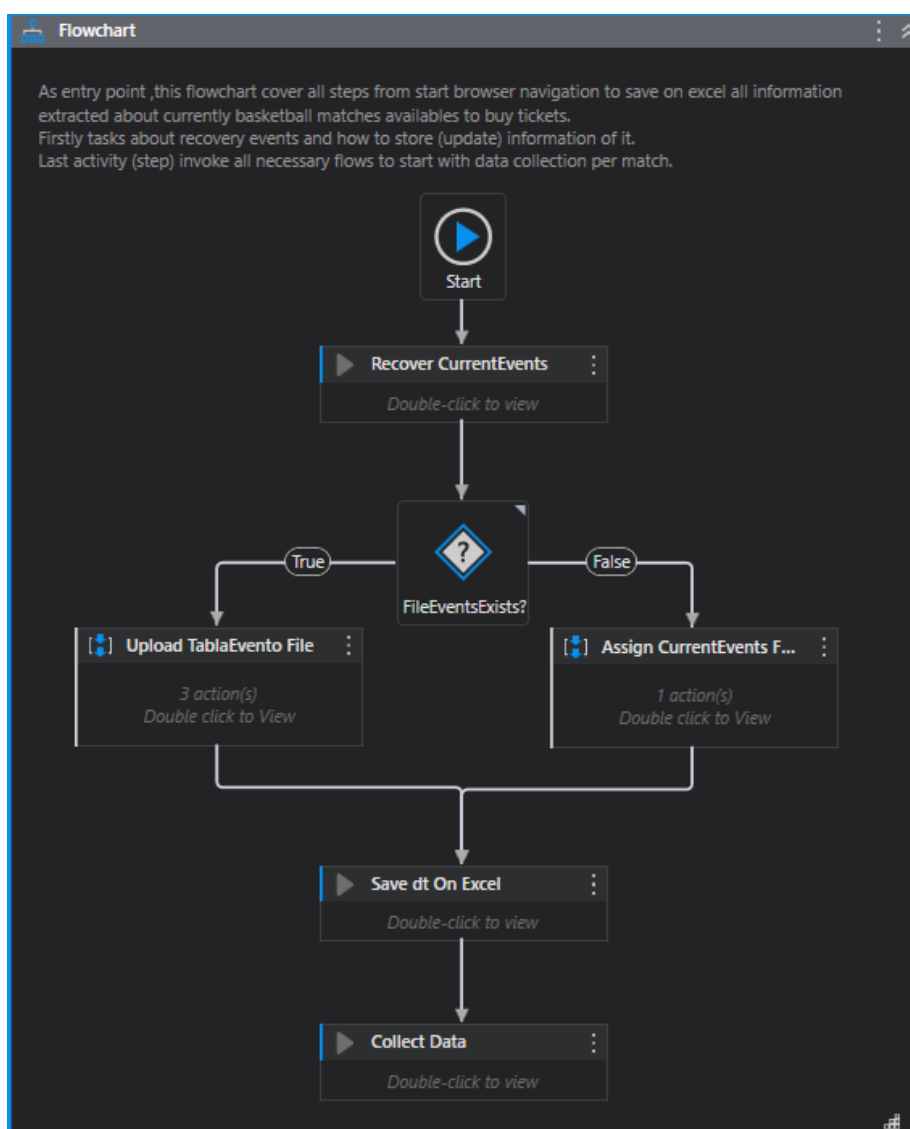


Figura 5.15: Flujo: Flowchart_Init.xaml

anteriormente. Para esto, el proceso utiliza la url de cada partido que ofrece la web y que se ha extraído anteriormente, abre un nuevo navegador desde donde se visitará una a una las 4 zonas disponibles del pabellón e irá recuperando sector a sector todos los datos que se han identificado en la anterior sección.

De nuevo, se enumeran secuencialmente las tareas realizadas por el proceso automatizado. El flujo comienza lee el fichero creado por el proceso *Flowchart_Init* y ejecuta las siguientes actividades para cada una de las filas del mismo (partidos de la lista).

- 1.- **Get url Partido:** desde el fichero *TablaEvento.xlsx* donde anteriormente se han almacenado todos los datos relativos a cada partido, se extrae la url del partido a tratar.
- 2.- **Navigate to partido url:** actividad que realiza la navegación a la vista de las gradas de baloncesto del partido a tratar. Esta página contiene un menú desplegable donde se selecciona el “Número de entradas” a comprar. En la automatización desarrollada, siempre se selecciona el valor 1.
- 3.- **Go zone FondoGoya:** aquí se recupera el valor del atributo “aaname” o “Attribute name” (Figura 5.16), el cual indica el nombre del elemento sobre el que se desea realizar la acción, en este caso la zona FondoGoya. El propósito es detectar si existen localidades disponibles en esta zona y continuar con el proceso en esta zona o saltar a la siguiente.
- 4.- **Go zone FondoJorgeJuan:** flujo que trata la zona FondoJorgeJuan de la misma manera que se ha comentado para la Zona Fondo Goya en el punto 3.
- 5.- **Go zone LateralFuenteBerro:** flujo que trata la zona LateralFuenteBerro de la misma manera que se ha comentado para la Zona Fondo Goya en el punto 3.
- 6.- **Go zone LateralFelipeII:** flujo que trata la zona LateralFelipeII de la misma manera que se ha comentado para la Zona Fondo Goya en el punto 3.
- 7.- **Condicionales:**
 - 7.1.- **seats availables in Zone ?** condicional incluido que comprueba si en la zona a tratar hay asientos libres o por el contrario ya



Figura 5.16: Detalle de recogida de atributo aaaname en zona Fondo Goya

se han vendido todas las localidades de la zona. El flujo continúa con el proceso de recogida del dato en la zona si existen localidades disponibles para esa zona o salta a la siguiente zona en caso contrario.

7.2.- ExcelZoneFileExists ? condicional incluido que redirecciona el flujo comprobando si el almacén de datos (el Excel) para la zona tratada está creado o no. Si la respuesta es negativa, crea un dataTable vacío y dirige al método "SectorScanning.xam". En caso positivo nos lleva a secuencia "Filter By partido", donde se recuperan los datos contenidos en la Excel correspondiente a la zona, se filtran por el partido concreto que esté tratando y se continua con la siguiente condición.

7.3.- PartidoRowCount <> 0 ? condicional incluido que divide el flujo comprobando si el partido tratado tiene ya filas insertadas en la Excel o no. La última fecha contenida para el partido será el dato que desencadena la redirección del flujo en uno u otro sentido. Si no existen filas para el partido a tratar, este condicional dirige el flujo al método "SectorScanning.xaml".

7.4.- lastDateinExcel.Equals(Today) ? condicional para comprobar que la fecha recuperada anteriormente es igual al día en curso. Esta condición es la que limita la extracción de los datos a una vez al día, como se explica en la Sección 5.5. Si quisiéramos permitir

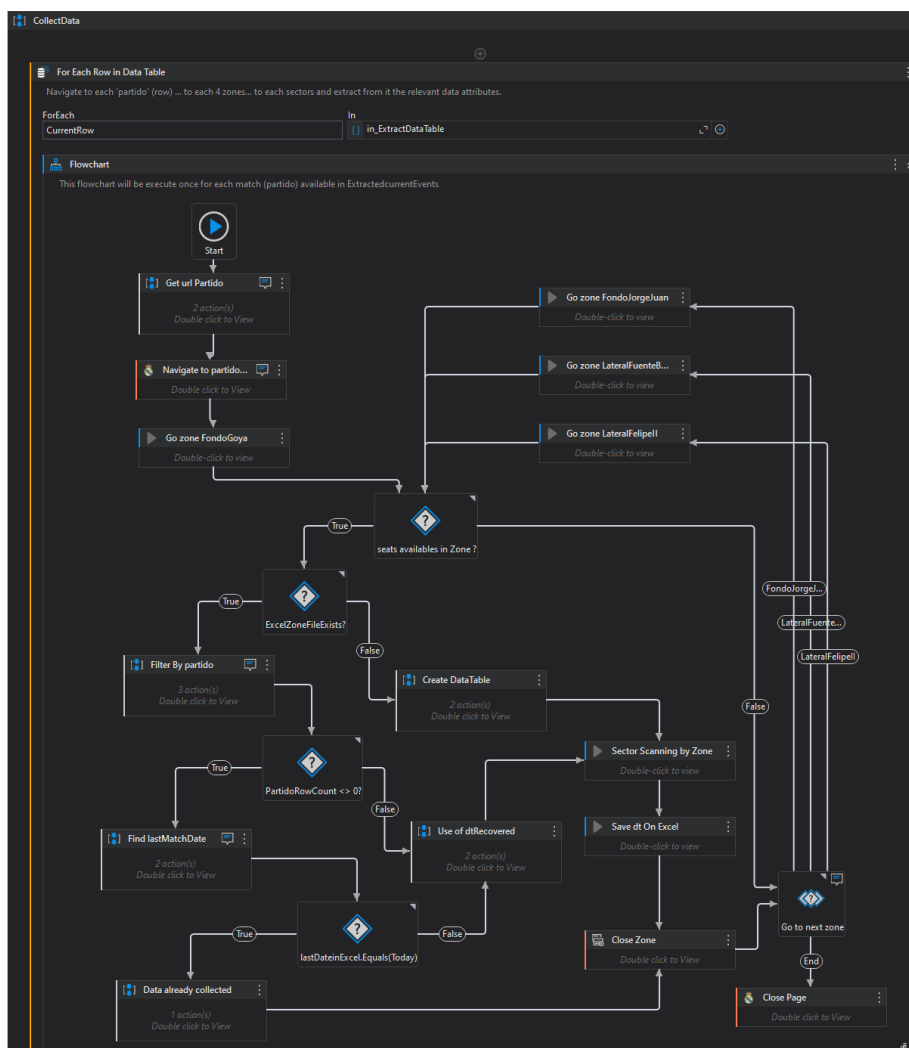


Figura 5.17: Flujo: CollectData.xaml

una periodicidad mayor, debemos modificar este punto. Después de esta actividad el flujo va directo a “Close Zone”.

8.- Go to next zone: actividad multidecisión que divide el flujo de control en cuatro ramas, de las cuales se ejecuta una sola en función de la siguiente zona a tratar. El orden de ejecución es el siguiente:

1. Fondo Goya
2. Fondo JorgeJuan
3. Lateral Fuente del Berro
4. Lateral Felipe II

9.- Sector Scanning by Zone: actividad que contiene una invocación a otro flujo donde se controla la recopilación concreta de los datos de cada uno de los sectores de la zona tratada. El flujo concreto es: “SectorScanning.xam”

10.- Save dt On Excel: tarea idéntica a la actividad 4 (*SaveOnExcel*) del flujo *Flowchart_Init*.

11.- Close Zone: acciones para cerrar la web de la zona que acabamos de tratar.

12.- Close Page: una vez finalizada la inspección/extracción de las cuatro zonas para el partido que estamos tratando, se lleva acabo el cierre de la página web para continuar con el siguiente partido (si lo hubiera).

El flujo indicado en Figura 5.17 se repetirá para cada uno de los partidos (filas) recuperadas en la primera sección. A continuación, se deben definir las acciones para extraer los atributos concretos de cada uno de los sectores

Como se explicó en la subsección 5.1, ante los cambios sufridos en la web de venta de entradas, es necesario también cambiar las acciones de recogida de datos, ya que tecnológicamente no podemos acceder al dato de la misma manera. Por tanto, a partir de aquí se diferenciarán dos escenarios para recogida de atributos. La denotación de primer y segundo escenario refiere únicamente a flujos creados antes y después del cambio en la página web. Las acciones de recogida, es decir, el Web Scrapping aplicado, es diferente y por tanto, los flujos creados también.

Primer escenario. Tecnología sobre sectores: Gráficos vectoriales SVG

El contexto tecnológico del primer escenario es una web html clásica que utiliza gráficos vectoriales escalables (Scalable Vector Graphics) para mostrar cada uno de los sectores del campo. SVG es un tipo único de formato de imágenes para gráficos basados en vectores y está escrito en lenguaje XML. El elemento SVG es un contenedor que define un sistema de coordenadas, el cual es usado para agregar un fragmento de un SVG dentro de un documento HTML. En nuestro caso, la web usa estos fragmentos SVG para crear cada uno de los sectores del campo que vemos en pantalla.

Los dibujos e imágenes SVG se crean utilizando una amplia gama de elementos dedicados a la construcción, el dibujo y el diseño de imágenes. Uno de los elementos SVG es el elemento `<polygon>`, que define una forma cerrada formada por un conjunto de segmentos de líneas rectas conectados, donde el último punto conecta con el primer punto. El elemento `<polygon>` se utiliza para crear un gráfico que contiene al menos tres lados. Otro elemento SVG usado en esta web es la etiqueta `<path>`, que es un elemento genérico para definir una forma. Todas las formas básicas pueden crearse con un metadato path. La Figura 5.18 muestra como la web usa estas dos etiquetas para formar el dibujo (la forma) de los sectores del pabellón sobre la pantalla.

Llegados a este punto del proceso, la automatización debe capturar los atributos que nos interesa recoger de cada uno de estos sectores del campo (polígonos SVG). En esta primera implementación, la tecnología utilizada por la web permite que estos atributos sean accesibles directamente desde la web, por lo que podrán ser obtenidos inspeccionando el código fuente de los diferentes elementos de la misma. Los atributos capturados, también detallados en la Figura 5.19, son los siguientes:

- `data-price-desc`: Denominación que recibe un grupo de sectores agrupados por precio. Corresponde con cada una de las líneas en las que se divide un fondo/lateral.
- `data-zone-name`: Nombre concreto del sector.
- `data-default-price`: Precio del sector.
- `data-available-seats`: Número de localidades disponible del sector.

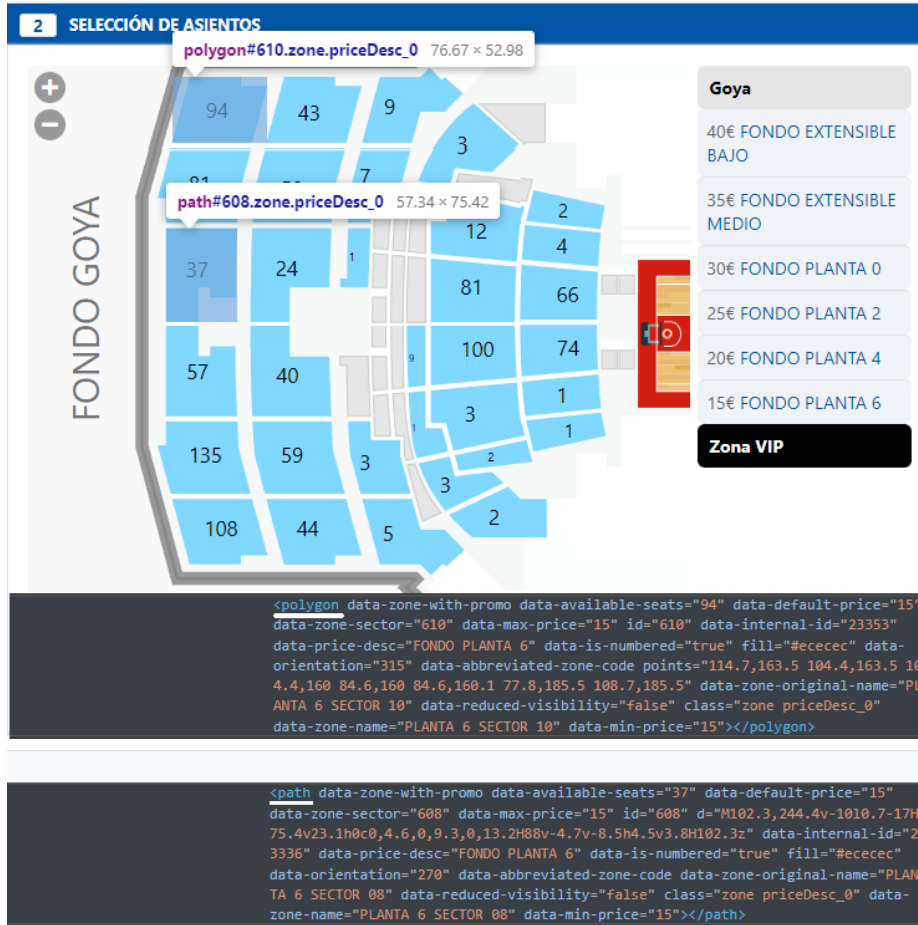


Figura 5.18: Etiquetas SVG: Path y Polygon



Figura 5.19: Atributos de recogida

SectorScanning.xaml

El flujo desarrollado en UiPath para recoger los datos anteriormente descritos se llama “SectorScanning”. Cada vez que se lanza el proceso global de automatización del proyecto, este flujo se llama cuatro veces, una vez por cada zona del pabellón. Aunque para todos los sectores que aparecen en pantalla se recogen los mismo atributos, obviamente las zonas del campo son diferentes y sus sectores también. Por tanto, es necesario pasar uno por uno por todos ellos e identificar (entre otros atributos) los asientos disponibles.

Es en este método donde se realiza el “data scrapping” sobre la pantalla para todos los atributos identificados. Es decir, es aquí donde se extrae el dato no estructurado directamente de la web para poder tratarlo posteriormente. En la Figura 5.20 se pueden observar el flujo creado en UiPath para la recogida de estos atributos.

Segundo escenario: Computer Vision.

Como hemos visto en el apartado anterior, la propia web nos ofrecía el dato y permitía recoger los atributos más representativos. Sin embargo, tras el cambio de la web, esto deja de ser posible, haciendo necesario modificar el método de recogida de datos.

Partiendo del mismo punto donde se visualiza por completo el pabellón (Figura 5.12), es evidente que no es posible trabajar igual que antes. Cuando la web se presenta en este modo de visualización, la información en la pantalla aparece después de situar el ratón por encima de cada uno de los sectores mostrados en la página. Internamente, esta acción desencadena la llamada a una función que devuelve los atributos que nos interesan.

Si bien es posible inspeccionar la web como hacíamos antes, ahora no aparecen los atributos de manera directa como si ocurría en el primer escenario. En la Figura 5.21 se observa claramente el cambio, donde ninguno de los atributos tratados en el primer escenario son expuestos directamente por la web. Técnicamente esta situación supone un reto, puesto que desde nuestra interfaz de desarrollo necesitamos usar otro tipo de actividades que nos permitan capturar el dato.

Para solventar este contratiempo, se tuvo que estudiar que otros mecanismos ofrece la herramienta UiPath para este tipo de situaciones, y finalmente se encontró una familia de acciones denominada Computer Vision (CV), que eran adecuadas para este tipo de problemas. Las actividades de CV contienen acciones de automatización de Interfaz gráfica (UI) refactorizadas,

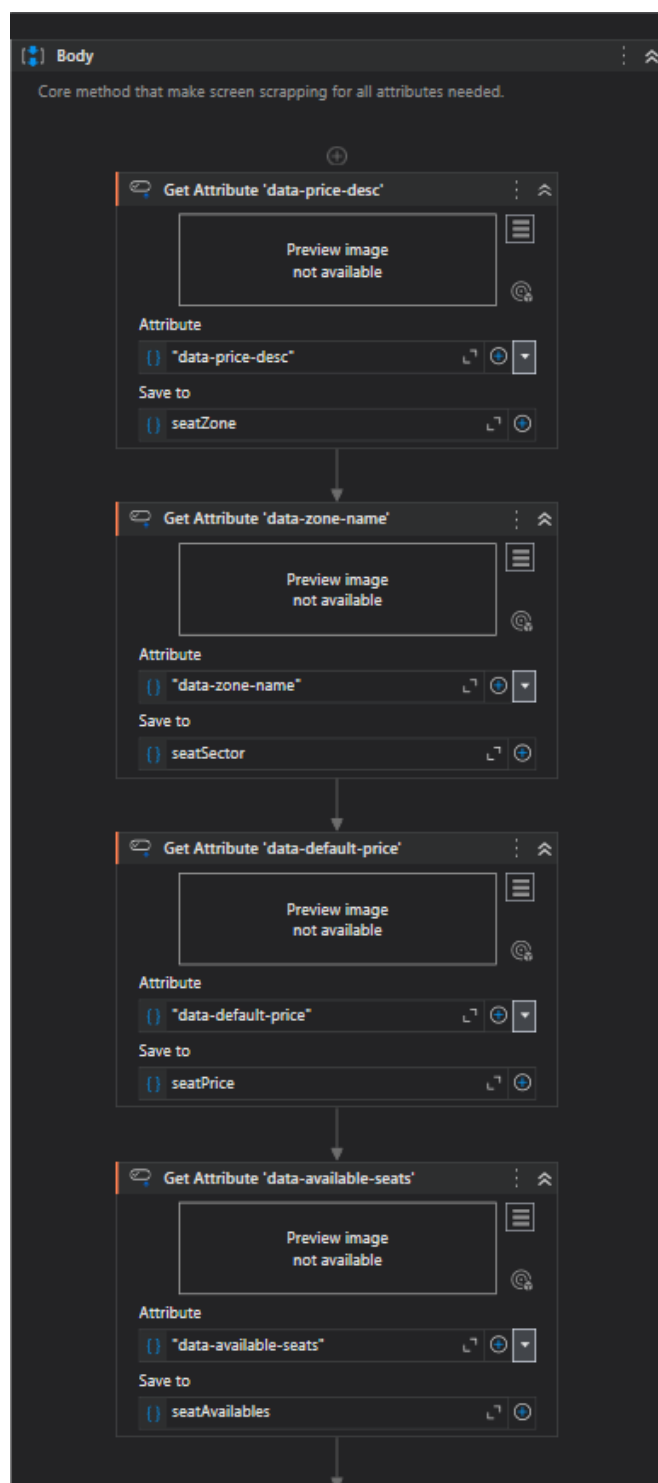


Figura 5.20: Primer escenario. Extracción de atributos

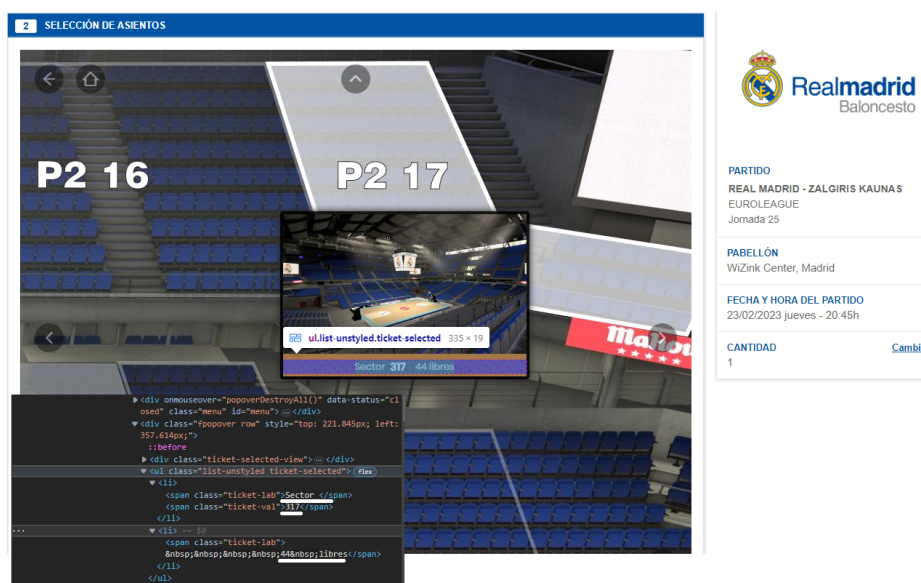


Figura 5.21: Etiquetas disponibles segundo escenario

como por ejemplo Click, Type Into (pulsar Intro), Get Text (Recoger Texto), etc. La principal diferencia entre las actividades de CV y las aplicables directamente a los atributos de la web es el uso de una red neuronal de CV, la cual ha sido desarrollada internamente por UiPath. La red neuronal es capaz de identificar elementos de UI como botones, campos de entrada de texto o casillas de verificación sin el uso de atributos, actuando directamente mediante reconocimiento visual en pantalla.

Estas actividades envían imágenes de la página o ventana que se está automatizando a la red neuronal, donde automáticamente se analiza y se identifican todos los elementos de la UI. A estas imágenes es posible asociar uno o varios anclajes o “anchors”, los cuales se utilizan para señalar la ubicación exacta de otros elementos de UI con los que se está interactuando, asegurando que la acción que se pretende realizar sea correcta. La Figura 5.22 presenta como UiPath identifica las diferentes partes interactivables de la página y permite relacionar un elemento principal (foco) con el resto de anclajes o anchors.

Esto puede confundir y llevar a pensar que CV es una herramienta OCR (Optical Character Recognition). Las herramientas OCR se basan completamente en la imagen y si esas imágenes cambian ligeramente de estructura, entonces los resultados obtenidos no serán fiables en esas situaciones. Sin embargo, las actividades de CV ayudan a identificar los elementos de la



Figura 5.22: Trabajando con la funcionalidad Computer vision

imagen como elementos individuales. Son más fiables que las herramientas OCR y soportan aplicaciones basadas en Flash, Silverlight, PDFs o directamente por imágenes. Otra diferencia es que la funcionalidad de CV requiere conexión con el servidor para procesarse.

Teniendo en cuentas las particularidades de este segundo escenario y las nuevas actividades que debían utilizarse para recoger los atributos, se desarrolló un nuevo flujo en UiPath llamado SectorScanningv2.

SectorScanningv2.xaml

Al igual que en el primer escenario, cada vez que se lanza el proceso global de automatización del proyecto, el flujo “SectorScanningv2” se llama cuatro veces, una vez por cada zona del pabellón. Aunque para todos los sectores que aparecen en pantalla se recogen los mismos atributos, obviamente las zonas del campo son diferentes y sus sectores también. Por tanto, es necesario recorrer los sectores una a uno y capturar los asientos disponibles. En la Figura 5.23 se detalla el flujo de llamadas para cada zona desarrollado en UiPath para tratar este escenario.

Dentro de cada una de estas llamadas a los sectores de cada zona, se han desarrollado otros flujo donde, para cada fondo/lateral, se establece su nombre y su precio. Esto es así porque en este escenario el valor de estos dos atributos es fijo, ya que no es posible recuperar estos datos directamente desde la web, y por tanto, estos atributos se establecen de antemano.

En la Figura 5.24 se observa como se han desarrollado estas actividades para la zona Fondo Goya. Para cada uno de los sectores dentro de esta zona, se ha desarrollado un flujo que individualiza el desarrollo y lo hace más manejable ante posibles cambios. Los flujos de los sectores tratados, ordenados según su cercanía a la pista de baloncesto, son los siguientes:

- Goya_FONDO_EXTENSIBLE_BAJO.xaml

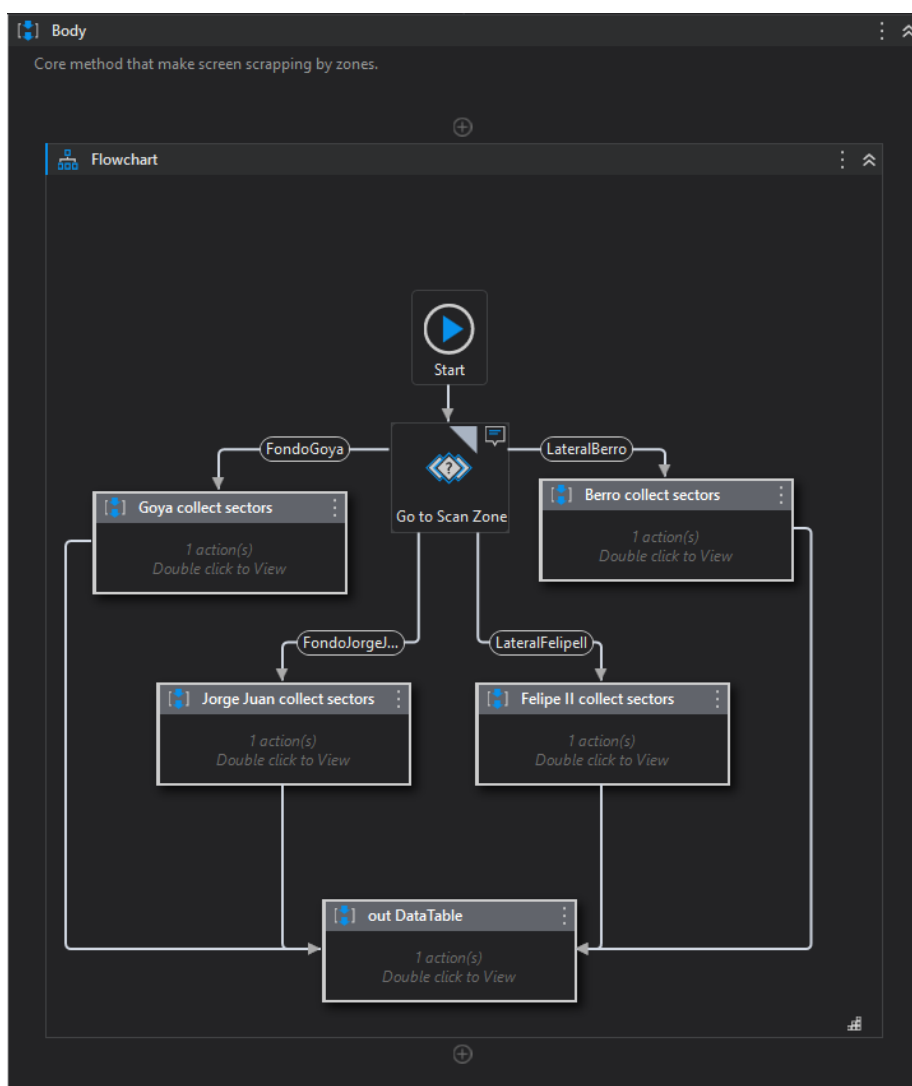


Figura 5.23: Flujo principal de llamadas en segundo escenario

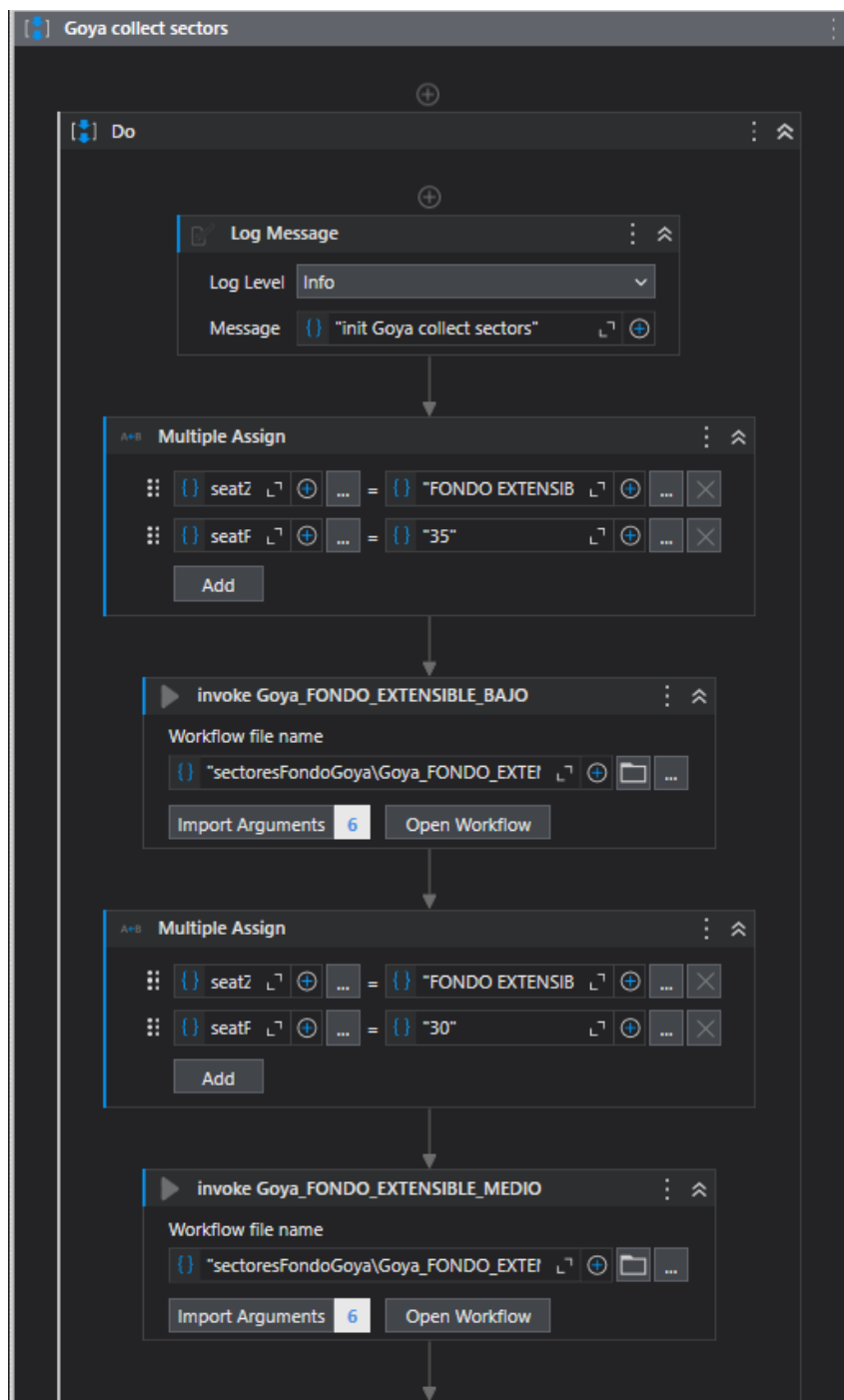


Figura 5.24: Detalle secuencia: Goya collect sector

- Goya_FONDO_EXTENSIBLE_MEDIO.xaml
- Goya_FONDO PLANTA 0.xaml
- Goya_FONDO PLANTA 2.xaml
- Goya_FONDO PLANTA 4.xaml
- Goya_FONDO PLANTA 6.xaml

En el proyecto de desarrollo, todos estos flujos están contenidos en una misma carpeta. Nótese que el proceso aplicado en el resto de zonas es el mismo. El detalle completo de todos los flujos desarrollados para cada uno de los sectores de cada zona se ha incluido en el Apéndice C.

Dentro de cada uno de estos flujos de sectores se trabaja principalmente con actividades de la familia Computer Vision. En primer lugar, se identifican las partes con las que queremos interactuar en la página web a tratar. Las acciones utilizadas son situar el ratón encima del elemento (*Computer Vision: CV Hover - Text*) que desencadena que la web muestre un diálogo o popUp en la pantalla con el número de localidades disponibles del sector. Y la segunda acción utilizada es la recogida de texto (*CV Get Text - Any Text*), que recoge este dato mostrado en pantalla usando la funcionalidad de reconocimiento por OCR. La última actividad de este flujo es en realidad una invocación a otro flujo llamado AddSectorData.xaml, donde se añade al dataTable una nueva fila con la información extraída. En la Figura 5.25 podemos observar estos elementos. El primer elemento pertenece a ese “lienzo” capturado sobre el que ejecutar las acciones CV, que en nuestro ejemplo se refiere al Fondo Goya. El resto de actividades hacen referencia a las otras acciones comentadas arriba.

Finalmente, destacar que esta segunda alternativa para la recuperación del dato es menos fiable que la ofrecida en el primer escenario. Esta implementación es más propensa a fallos debido a que el dato es directamente extraído de una imagen generada mediante visión artificial. Esto supone una necesidad de mantenimiento mayor y por tanto, un proyecto más delicado, puesto pequeños cambios en la web supondrán también adaptaciones en el desarrollo.

5.3. Proceso de Transformación del dato

Dentro del proceso de ETL (Extraction, Transformation, Load) [56], la parte de transformación se ha realizado con la misma herramienta que la

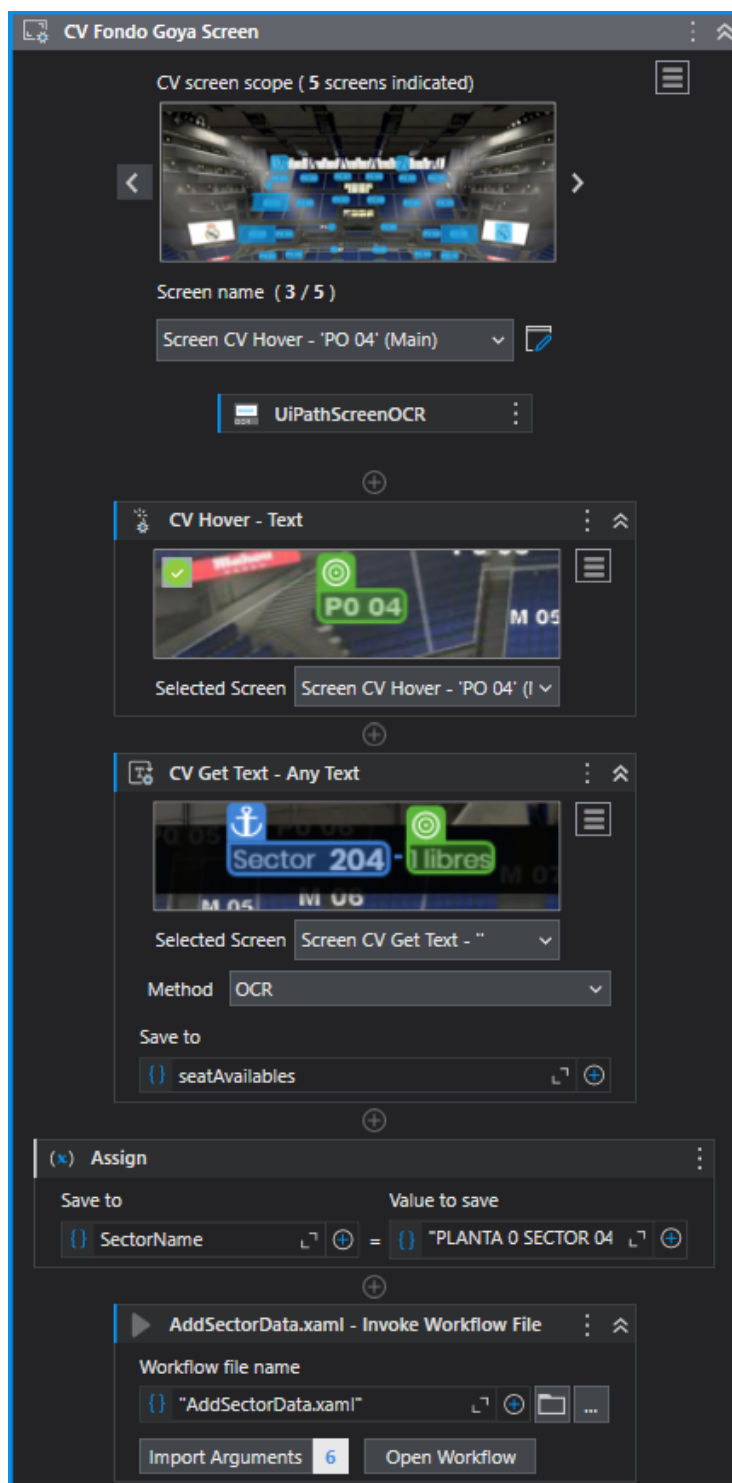


Figura 5.25: Detalle: Goya_FONDO PLANTA 0.xaml

extracción. Partiendo de los datos extraídos en el punto anterior, se han desarrollado unos flujos que realizan varias operaciones sobre dichos datos para transformarlos en información útil.

KPI

Una vez obtenido el dato hay que valorar que información queremos ofrecer al usuario final y extraer de este análisis los indicadores a tratar. Un indicador es un valor medible que muestra la eficacia con la que se está cumpliendo un objetivo. Habitualmente se denomina Key Performance Indicator (KPI) a esta forma de medir y representar la información. Permite también hacer comparaciones con el resto de datos para determinar si está cumpliendo o no sus objetivos. Elegir los KPI que son relevantes dependerá totalmente de estos objetivos y de la capacidad que tengamos de extraer información de la fuente de datos para medir su rendimiento.

A continuación, se han identificado algunos de los indicadores que se pretenden extraer de los datos. También se ha añadido en pseudocódigo los cálculos aplicados para conseguir la información. Los KPI identificados son los siguientes:

KPI 1: Total Entradas Vendidas: total de entradas vendidas en el partido seleccionado.

$$\text{Total Entradas Vendidas} = \text{Total aforo} - (\text{sum}(\text{seatsAvailables}))$$

KPI 2: Total Entradas Disponibles: entradas aún disponibles para el partido seleccionado. Si es partido ya ha finalizado, su valor es 0.

$$\text{Total Entradas Disponibles} = \text{sum}(\text{seatsAvailables})$$

KPI 3: Total Ventas Hoy: indicador doble sobre entradas vendidas en el día seleccionado [nº entradas].

$$\text{Total Ventas Hoy} = \text{Ayer}(\text{sum}(\text{seatsAvailables})) - \text{Hoy}(\text{sum}(\text{seatsAvailables}))$$

KPI 4: Total ventas partido (€): cantidad de dinero recaudado (en euros) por la venta de localidades hasta el último día de recogida.

$$\text{Total ventas partido} = \sum \text{Partido}(\text{precioZonas} * (\text{seatsAvailables_inicio}) - \text{sum}(\text{seatsAvailables_fin}))$$

KPI 5: Total Ventas Anuales: Total acumulado de ventas anuales, medido en €.

$$\text{Total Ventas Anuales} = \text{año}(\text{KPI 4})$$

KPI 6: Media de Ocupación: indicador histórico sobre media total de ocupación de asientos.

$$\text{Media de Ocupacion} = \text{avg}(\text{date}(\text{KPI 1}))$$

KPI 7: % Ocupación partido.

$$\% \text{ Ocupación partido} = \%(\text{KPI 1})$$

KPI 8: Tendencia: Tendencia de ocupación global del recinto con respecto al partido anterior.

$$\text{Tendencia} = \text{KPI 8}_{\text{partidoSeleccionado}} - \text{KPI 8}_{\text{partidoAnterior}}$$

Para los cálculos que han requerido de datos sobre precios y por tanto, hacer uso de cifras relativas al campo, se ha tenido en cuenta lo indicado en la Sección 5.1.

Flujos de cálculo de indicadores

Una vez determinada la información que se pretende extraer y mostrar, se procede a desarrollar los flujos para generar los diferentes indicadores. Este apartado entra dentro del análisis exploratorio de datos, ya que se valora el conjunto de datos para comprobar si pueden ser empleados para detectar tanto las características más sobresalientes como las características inesperadas.

La dificultad técnica de esta parte no es elevada, puesto que únicamente se tratan operaciones matemáticas aplicadas sobre los datos extraídos y filtros de los datos. La Figura 5.26 muestran las actividades del flujo principal de desarrollo.

De nuevo, se enumeran secuencialmente las tareas según se van realizando por el proceso automatizado. El proceso comienza con el flujo mostrado en la Figura 5.26, que a su vez, es otro invocador de los siguientes métodos:

1. OpenExcelFiles: Abre el archivo con los datos de la zona a tratar.

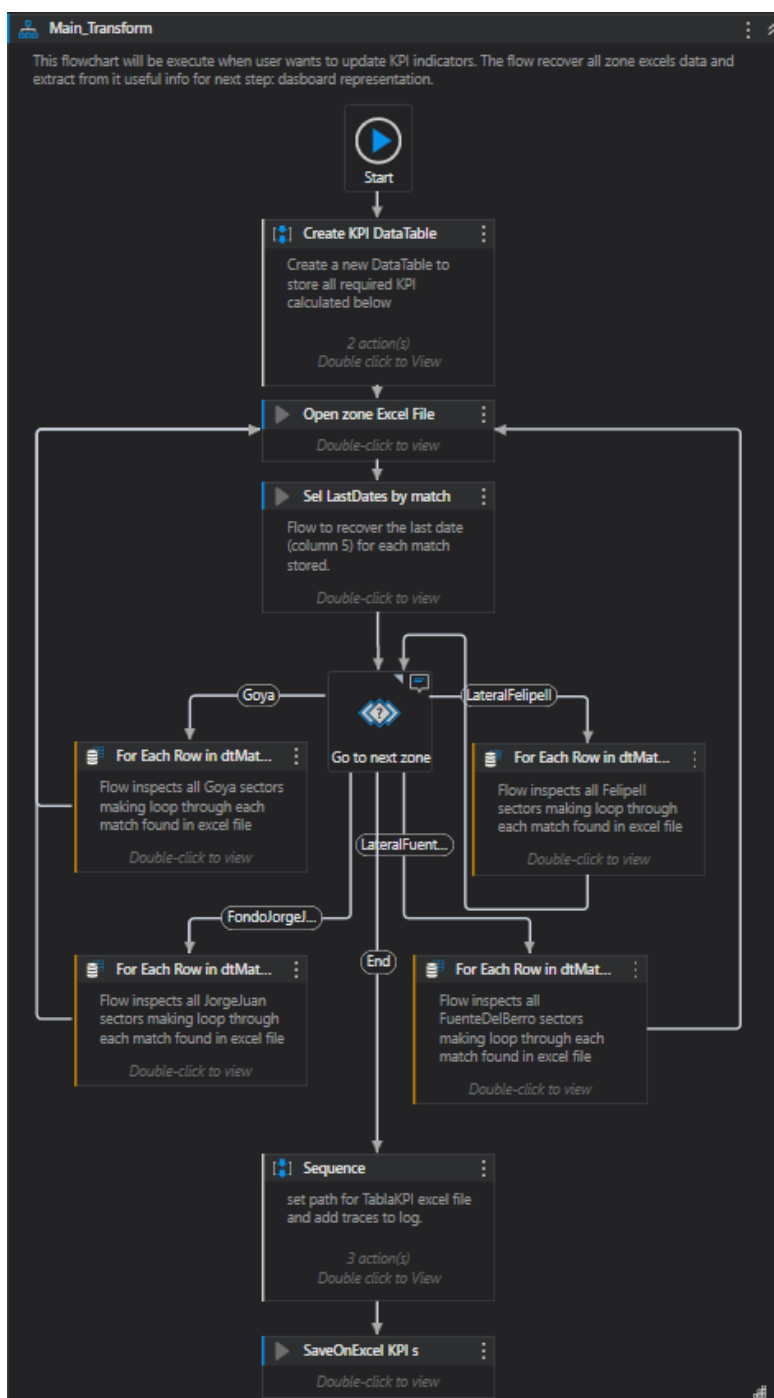


Figura 5.26: Flujo principal de cálculo de KPI

Indicador	Variable
KPI 1	seatsSold
KPI 2	seatAvailables
KPI 4	totalSales
KPI 7	perSeatsBooked

Tabla 5.1: Relación KPI con variables calculadas

2. ManipulateData: Flujo que recupera la última fecha almacenada de la zona a tratar para usarla en cálculos futuros.
3. SaveOnExcel: Conjunto de actividades que trasladan el objeto dataTable a un fichero Excel almacenado en el equipo desde donde se ejecuta la automatización.

El tratamiento para la extracción de los indicadores también debe ser diferenciado por zonas, puesto que así se ha organizado el almacenamiento de los datos. Por tanto, es necesario abrir cada uno de estos archivos, recuperar los datos y transformar ese dato en información. Al igual que en anteriores apartados, se muestra la zona Fondo Goya como ejemplo en la Figura 5.27, siendo aplicable al resto de zonas el mismo proceso. Una vez extraída la información tenemos ya listos los indicadores, los cuales se pueden observar en la Tabla 5.1.

5.4. Almacenamiento de los datos

En este apartado se detallan todas las acciones que se han requerido para habilitar el guardado de los datos recogidos por el proceso de automatización. La persistencia de los datos en ficheros internos con un formato fijo es necesaria ya que nos permite una posterior consulta y es la base de la visualización final del proyecto.

La organización y estructura de los campos también es importante, ya que debemos tener asignado a cada dato un campo en el fichero destino y a cada sector/zona desde donde se extrae ese campo, un identificador. Es decir, debemos ser capaces a posteriori de situar el dato en su zona física del campo que le corresponde, añadiendo también una identificación temporal para cada uno de ellos. Diariamente se extrae una “foto” de los datos disponibles en la web y el guardado de estos datos no solo abarca al propio dato, sino

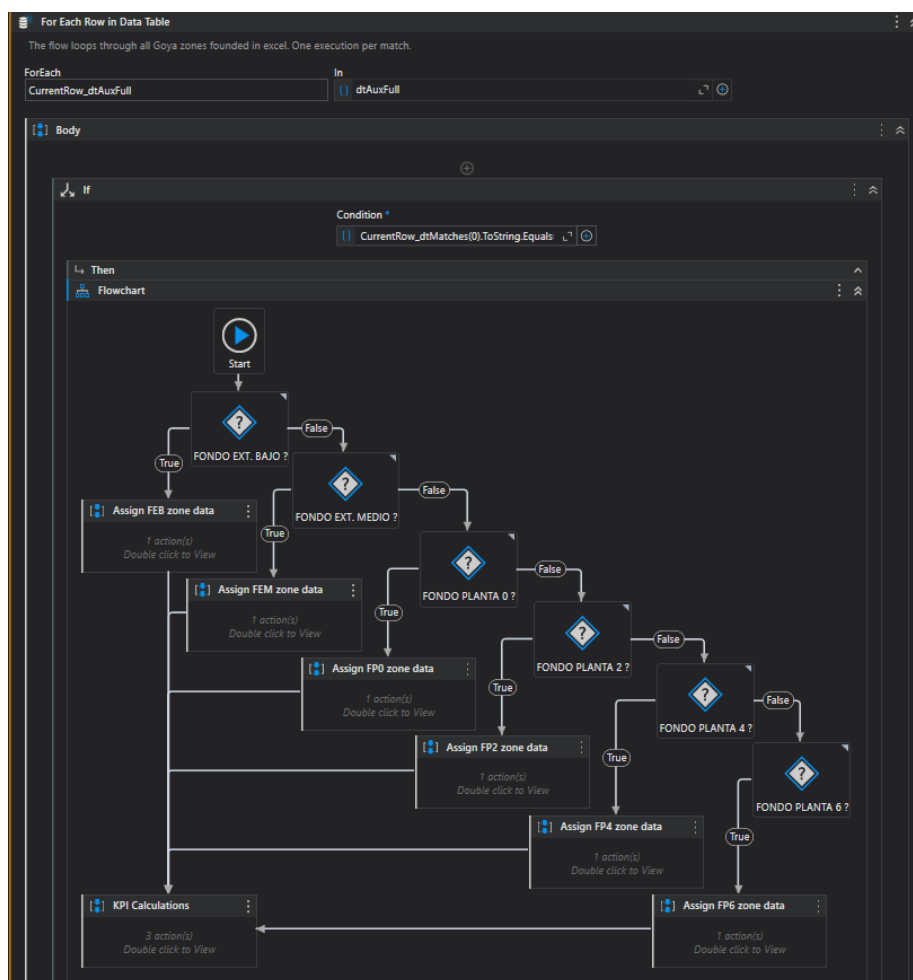


Figura 5.27: Flujo de cálculo de indicadores para la zona Goya

también como queremos recuperarlos o hacer uso de ellos. Esto también se ha tenido en cuenta a la hora de estructurar el almacenamiento del dato.

El sistema elegido para la grabación de los datos han sido ficheros .xlsx (Microsoft Excel). Por su facilidad de uso desde la herramienta y por ser casi un estándar en el mercado, es lo más conveniente y cercano al proyecto aquí presentado. Con el volumen estimado de datos que se manejan en el proyecto, es lo más conveniente para ofrecer al lector una visualización rápida y estructurada de la información.

A continuación, se indica la estructura de los ficheros de almacenamiento para cada uno de los flujos de extracción de datos y que ejecutan acciones de almacenamiento en disco:

Fichero Desarrollo	Flujo	Archivo Destino
RecoverCurrentEvents.xaml	Sequence	TablaEvento.xlsx

Tabla 5.2: Relación de ficheros sobre Listado inicial de eventos

Campo	Tipo	Descripción
Partido	String	Nombre identificador del partido
FechaHora	DateTime	Timestamp del evento
Jornada	String	Texto con jornada a la que aplica el evento
Competición	String	url con imagen de la competición a la que aplica el evento
Rival	String	url con imagen del escudo del club contrincante
EntradasUrl	String	url de acceso directo a la compra del evento

Tabla 5.3: Estructura de almacén TablaEvento.xlsx

Almacén de Listado inicial de eventos

En la Tabla 5.2 se muestra la relación entre el fichero desarrollado para la funcionalidad con su correspondiente almacén de datos, mientras que la relación de campos utilizados en la estructura del almacén de datos se muestra en Tabla 5.3:

Almacén de Escaneo de sectores

En la Tabla 5.4 se muestra la relación entre ficheros desarrollados para la funcionalidad de escaneo de sectores con sus correspondientes almacenes de datos. En este caso tenemos cuatro almacenes de datos (uno para cada zona).

La relación de campos utilizados en la estructura del almacén de datos se muestra en la Tabla 5.5. La estructura interna de los campos es la misma para cada uno de los ficheros de zonas. Como muestra esta tabla, la colección de datos contiene un documento por cada zona del campo, el cuál tiene seis atributos. Para esta colección se ha optado por establecer como índice el campo Partido. La razón de hacerlo así es que con esta elección podemos obtener una consulta cubierta completa, es decir, que la consulta

Fichero Desarrollo	Flujo	Archivo Destino
SectorScanning.xaml	Goya collect sectors	TablafondoGoya.xlsx
SectorScanning.xaml	Jorge Juan collect sectors	TablaJorgeJuan.xlsx
SectorScanning.xaml	Berro collect sectors	TablaLateralFelipe.xlsx
SectorScanning.xaml	Felipe II collect sectors	TablaLateralFuenteBerro.xlsx

Tabla 5.4: Relación de ficheros sobre Escaneo de sectores

Campo	Tipo	Descripción
match	String	Nombre identificador del partido.
seatZone	String	Nombre de zona (fila) correspondiente.
seatSector	String	Nombre de sector (box) correspondiente.
seatPrice	String	Precio de la zona.
seatAvailables	String	Localidades disponibles.
date	DateTime	Timestamp de recogida del dato.

Tabla 5.5: Estructura de almacén de datos por zonas

satisfaga íntegramente la recuperación de todos los valores relativos al evento seleccionado.

Almacén de Indicadores

Por último, faltan almacenar los diferentes indicadores. En la Tabla 5.6 se muestra la correlación entre el fichero desarrollado para la funcionalidad de generar los indicadores con su correspondiente almacén de datos en disco, mientras que la relación de campos utilizados en la estructura del almacén de datos de indicadores se muestra en la tabla 5.7.

5.5. Cloud UiPath: Orchestrator

El Orchestrator es componente cloud de la solución UiPath que no permite gestionar todos los desarrollos realizados directamente desde la

Fichero Desarrollo	Flujo	Archivo Destino
Main.xaml	TFMaster_Transform	TablaKPI.xlsx

Tabla 5.6: Relación de ficheros sobre el almacén de KPI

Campo	Tipo	Descripción
match	String	Nombre identificador del partido.
Zone	String	Nombre de zona (fila) correspondiente.
seatZone	String	Nombre de sector (box) correspondiente
seatPrice	String	Precio de la zona
LastDate	DateTime	Timestamp de último dato recogido.
totalCapacity	String	Capacidad total de la zona. [link a Contexto+Baloncesto]
seatAvailables	String	Localidades disponibles del sector.
seatSold	String	Localidades vendidas del sector.
perSeatsBooked	String	Porcentaje de localidades vendidas.
totalSales	String	Importe de ventas totales del sector.

Tabla 5.7: Estructura de almacén de datos por zonas

web, siendo de vital importancia en el desarrollo de este proyecto. Este componente cloud ofrece una serie de funcionalidades muy útiles a la hora de manejar las diferentes automatizaciones que se han desarrollado.

Todos los desarrollos y las automatizaciones se han llevado a cabo utilizando el IDE propio de Uipath (Studio). Una vez que hemos finalizado los procesos y están listos para lanzarse de manera desatendida, es posible subir los paquetes (proyectos) desarrollados y gestionarlos desde el Orchestrator para su integración en la plataforma. Esto nos permite desde administrar las versiones de los propios paquetes subidos, crear disparadores con ejecuciones programadas de estos procesos, crear plantillas de máquinas destino donde lanzar las automatizaciones de manera desatendida, administrar diferentes carpetas donde tener organizados todos nuestros desarrollos, etc.

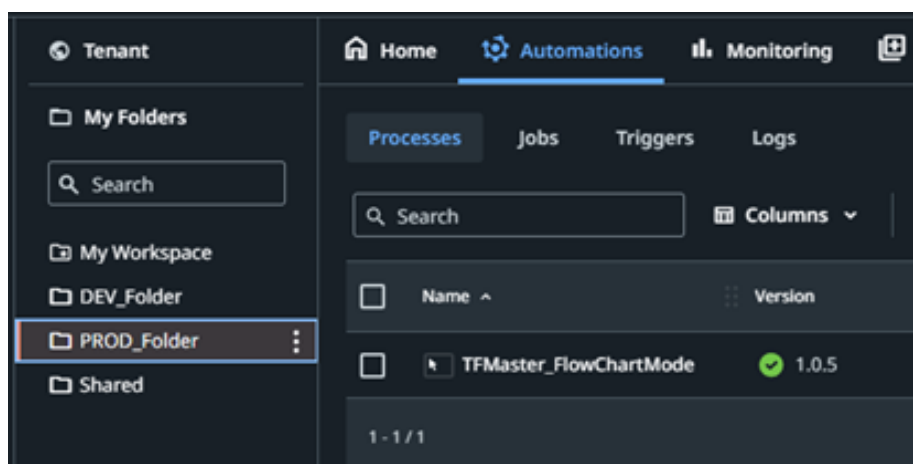


Figura 5.28: Carpeta contenedora de procesos automatizados

Ejecución

Una vez subido el desarrollo desde UiPath Studio al Orchestrator, tenemos ya disponible el trabajo para poder ejecutarlo. Desde la carpeta `PROD_Folder`, mostrada en la Figura 5.28, es posible lanzar el proceso. Para ello, también es necesario definir una máquina destino. El System Under Test (SUT) es la máquina donde se dirigirá nuestra automatización, la cual se recomienda que tenga las mismas características técnicas que la utilizada en fase de desarrollo y pruebas para evitar posibles problemas.

Por tanto, el último paso será instalar en la máquina destino el paquete UiPath Assistant, que permitirá que nuestra máquina sea accesible directamente desde el cloud. De forma resumida, este paquete instala un servicio en la máquina destino que permite esta comunicación. A través de una cuenta propia del sistema, se habilita la conexión entre máquina y cloud. La Figura 5.29 muestra la relación de máquinas disponibles que son accesibles desde el cloud. Desde aquí también pueden darse de alta nuevas máquinas destino y será donde se aloje la única licencia disponible que ofrece la versión Community para la ejecución de robots.

Con estos pasos ya es posible lanzar nuestro proceso hacia la máquina destino donde se ejecutará. Para ello hay que hacer uso de un robot que se asocia al proceso justo antes de lanzarlo.

Una vez ejecutado ya es posible monitorizar los logs que se van generando y recibir la salida de ejecución desde la plataforma. Desde el propio cloud es posible extraer métricas de ejecución y visualizar diferentes diagramas

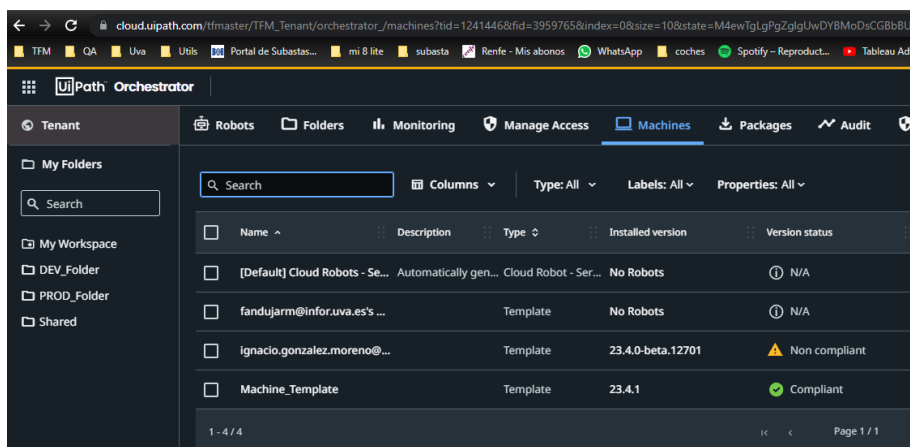


Figura 5.29: SUT



Figura 5.30: Monitorización del proceso automatizado

sobre el estado de ejecución de los procesos de automatización lanzados, tal y como se puede ver en la Figura 5.30.

Desde aquí es posible evaluar los resultados esperados y medir valores para, por ejemplo, reconducir flujos e ir afinando los posibles errores encontrados. Se extraen métricas sobre eficiencia, tiempo y estado de las ejecuciones.

La periodicidad de captura de información (*Web Scraping*) de los procesos automatizados planteados en este TFM es de **1 ejecución automatizada al día**, para así no vulnerar lo estipulado en la Subsección 3.2.

5.6. Visualización de los datos

Se aborda en esta última sección la comunicación de los resultados. En muchos casos, la visualización puede ser la realización de un informe escrito, presentación o similar sobre las conclusiones extraídas del análisis de los datos. En este caso, el desafío consiste en ser capaz de resumir y filtrar de forma satisfactoria la información generada, de forma que destaquen de forma eficiente e intuitiva los aspectos a los que se desea dar mayor relevancia, ya que se pretende obtener una reacción positiva del usuario final, incluso por personas no expertas. La presentación visual debe ser útil, pero simple, de forma que intuitivamente la información se comprenda de forma agrupada en un dashboard o cuadro de mando. Entre los objetivos a satisfacer, podemos destacar:

1. Facilitar la comprensión de los datos.
2. Determinar qué datos son importantes y cuáles no lo son.
3. Simplificar la comunicación.
4. Mostrar el contexto y avance de un proyecto a usuarios no expertos.

Partiendo de los KPI definidos previamente, se creó un primer boceto gráfico sobre cómo sería su representación en un dashboard final y su disposición de elementos sobre la pantalla. Este boceto (Figura 5.31), realizado a mano alzada, servirá de guía para el desarrollo de la visualización final.

A continuación, se detallan los pasos realizados para conseguir esta representación de los datos haciendo uso de la herramienta seleccionada para esta tarea: Tableau Desktop.

Conectarse al conjunto de datos

El primer paso consiste en importar cada uno de los almacenes de datos que se han creado y alimentado a lo largo de cada una de las ejecuciones de nuestro robot. Diariamente se han poblado con datos los diferentes ficheros de zonas junto con sus partidos. En la Figura 5.32 se pueden observar las conexiones habilitadas en Tableau.

Partiendo de cada uno de estos ficheros se crean relaciones de varios a varios, que habilitarán la conexión entre almacenes y nos permitirá realizar consultas a todos los datos contenidos en las diferentes tablas. En la Figura 5.33 se muestra el diagrama con las diferentes conexiones creadas.

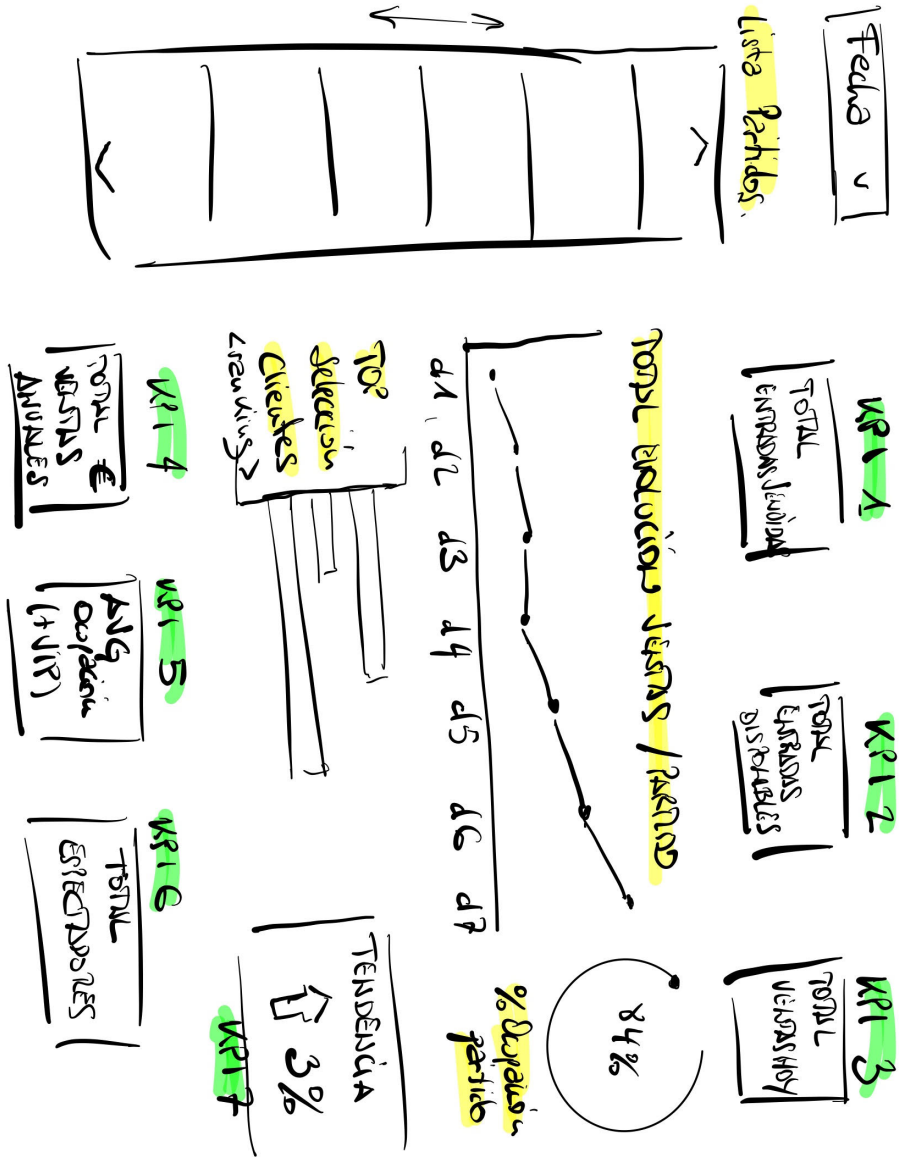


Figura 5.31: Boceto inicial de visualización de datos

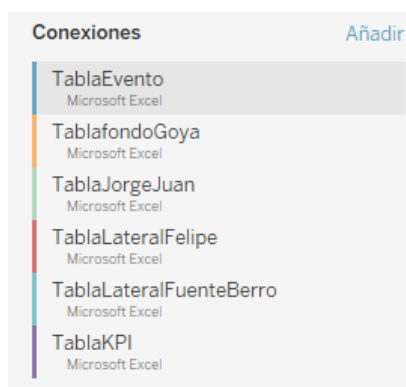


Figura 5.32: Almacenes de datos en Tableau

Una vez cargadas las tablas y teniendo los datos disponibles desde la plataforma Tableau, podemos comenzar a desarrollar la visualización de datos.

Desarrollo cuadro de mando

Finalmente, comenzamos la última fase del proyecto, consistente en dar forma a los datos extraídos. Para ello, se creará un cuadro de mando que será accesible desde un navegador web y estará vinculado a las fuentes de datos anteriormente descritas. Los cuadros de mando o dashboards [29] permiten combinar varias gráficas en una sola pantalla. Un cuadro de mando es un tipo de interfaz gráfica de usuario que suele ofrecer vistas rápidas de los indicadores clave que son relevantes para un objetivo o proceso concreto.

El punto de partida para crear cualquier gráfico o mapa en Tableau es colocar campos en Columnas y Filas, que para la mayoría de los gráficos corresponden a los ejes X e Y, respectivamente. Por cada gráfico que se quiera generar, se debe generar una nueva Hoja y trabajar en el gráfico de forma individual. El listado de hojas creadas desde los datos disponibles es la siguiente:

- Eventos: Fecha del evento mostrado.
- ListaPartidos: Desplegable con la lista de los partidos disponibles.
- imgTorneo: Imagen representativa de la competición que se juega.
- scoreboard: Hoja con escudo de equipo local y visitante junto con el resultado final del encuentro.

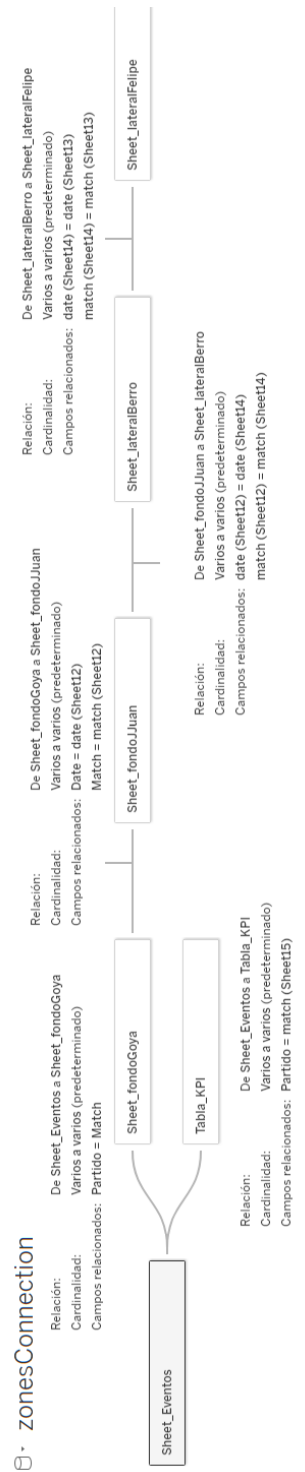


Figura 5.33: Diagrama conexiones entre almacenes de dato

- slotEvents: Días con datos recogidos del partido seleccionado.
- TotalbyZonesAGR: Acumulado de asientos disponibles totales por día.
- TodaybyZonesAGR: Número de asientos vendidos cada día.
- TotalbyZones: Acumulado de asientos disponibles por zona y día. Incluye etiqueta con el número concreto.
- LastDateValues: Valores de asientos disponibles en el último día de venta de entradas, para cada una de las zonas y el total.
- KPI: Panel inferior que contiene los indicadores: capacidad de pabellón, asientos disponibles, número de asientos vendidos y porcentaje ocupación partido.
- KPI_Partido_Ventas: Recaudación por venta de entradas en partido medido en euros.
- KPI_Temporada_per: Porcentaje de tendencia de ocupación por temporada.
- KPI_Temporada_ventas: Recaudación por venta de entradas en temporada (acumulativo) medido en euros.

Los cuadros de mando permiten organizar los gráficos que hayamos construido en las diferentes hojas en una única visualización, con aspecto de un panel de control, en el que podemos ver la información relevante de un vistazo. Los gráficos deben haber sido creados antes en Hojas individuales, puesto que esta parte del interfaz simplemente agrupa datos ya creados.

La Figura 5.34 muestra como se han dispuesto las diferentes hojas (paneles) en la pantalla durante el proceso de creación del cuadro de mando. La distribución de elementos en pantalla se ha procurado que sea de la manera más simplificada, destacando los indicadores principales para resaltar su importancia.

Despliegue en web

Ahora que tenemos un gráfico pulido como para que se pueda inspeccionar por el público, podemos pasar al proceso de publicarlo. Tableau permite publicar los resultados como páginas web de forma muy sencilla. Para poder acceder a esta funcionalidad, debemos crear un perfil de usuario en el servidor público Tableau Public como se comentó en la sección 4.2. La

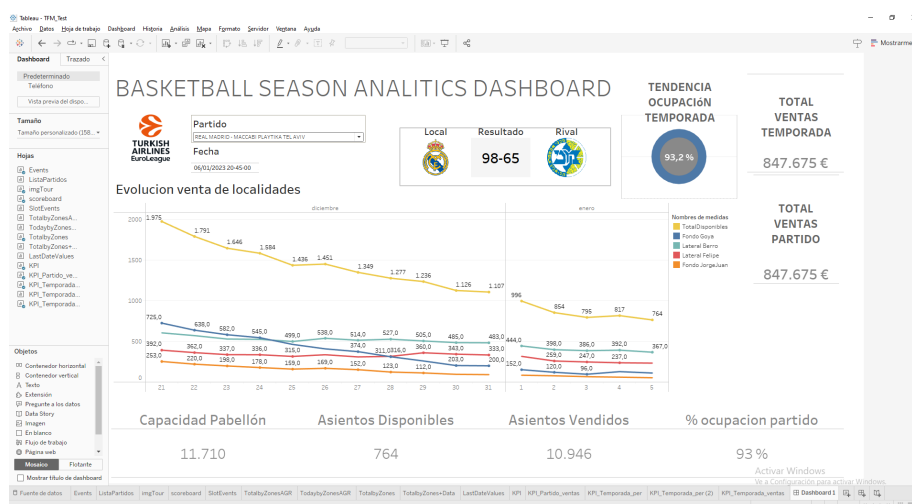


Figura 5.34: Creación de dashboard en Tableau Desktop

aplicación asigna al usuario una URL pública de perfil, la cual se emplea para dar acceso a otras personas a la visualización creada. Para almacenar la visualización en el servidor, hay que compartir el cuadro de mando creado desde Tableau Desktop. Si las operaciones se realizan de forma correcta, al acabar este proceso la visualización estará disponible en el perfil del servidor de Tableau Public. En el caso de este proyecto, la url de acceso al cuadro de mando es la indicada aquí: [22]

5.7. Análisis de los resultados

Por último, en esta sección se valorarán los resultados obtenidos para algunos partidos concretos y analizamos tendencias y puntos destacados del proceso.

Escenario 1: Tendencia típica de ocupación de partido.

Un caso típico de ocupación de localidades en un partido, es el mostrado en la Figura 5.35, donde a medida que avanzan los días de venta de localidades, los asientos disponibles de cada una de las zonas van disminuyendo. Tanto en la línea superior que representa el acumulado de las otras cuatro, como en las que representan a cada zona, se observa un tendencia descendente. El resto de indicadores se muestran en la página ofreciendo la información indicada.

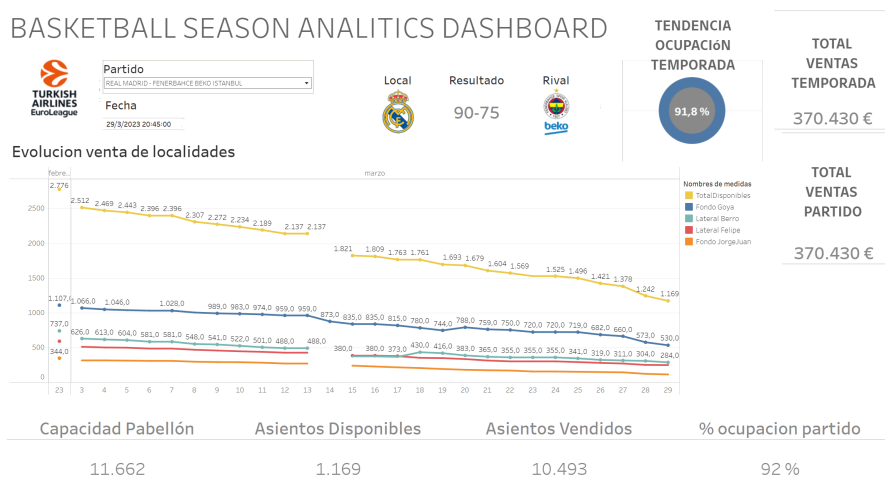


Figura 5.35: Real Madrid - Fenerbahçe: Tendencia de asientos disponibles para un partido.

Escenario 2: Partido sin entradas disponibles.

Otro escenario encontrado son aquellos partidos donde se han agotado las entradas a la venta. Por el interés despertado por estos encuentros a lo largo del periodo de venta, los asientos disponibles han disminuido hasta alcanzar el valor cero. La Figura 5.36, que pertenece al partido Real Madrid - Barcelona nos muestra esta situación

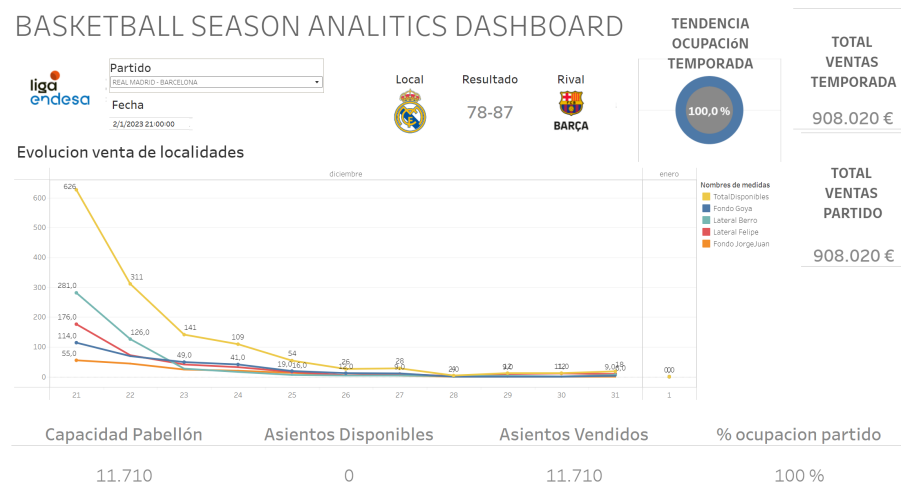


Figura 5.36: Real Madrid - Barcelona: Partido sin entradas a la venta.

Escenario 3: Según avanza la venta de localidades el número de asientos se incrementa.

Por último, comentar otro escenario que a primera vista puede parecer extraño, pero que ocurre en muchos partidos que hemos tratado. Si observamos las tendencias de venta de entradas por partido, vemos en algunos puntos concretos (días) el número de asientos como se muestra en Figura 5.37, en lugar de disminuir, se incrementa. Este comportamiento que nos 'confunde' el dato, hace que si observamos las gráficas de evolución de ventas sobre una determinada zona, a veces se observen tendencias como la indicada en Figura 5.38.

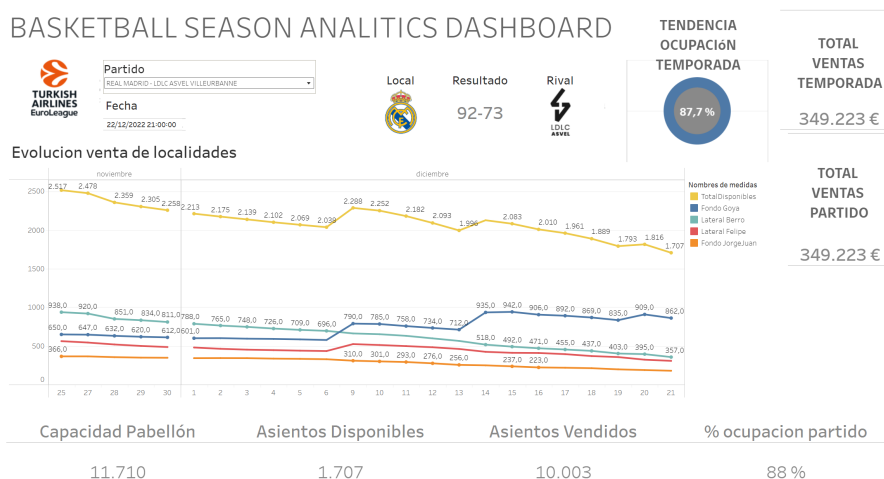


Figura 5.37: Real Madrid - LDLC Asvel Villeurbanne: Tendencia incremental.



Figura 5.38: Tendencia incremental de asientos disponibles en un partido.

Este tipo de situaciones, se ha consultado a la propia empresa de gestión de entradas que ha respondido de la siguiente manera:

Le informamos que las localidades disponibles son las que se pueden ver en la web en todo momento. Al acceder al proceso de compra muchas personas, puede suceder que bloqueen las localidades durante el periodo de compra y, si finalmente no completan el proceso, éstas queden liberadas. No se guarda ningún cupo de entradas, todas las localidades disponibles se ponen a la venta en el momento en que se abre la venta.

Con este análisis se finaliza el repaso a los aspectos destacados sobre el desarrollo práctico del Proyecto Fin de Máster.

Conclusiones y Líneas de trabajo futuras

El desarrollo de este TFM se ha realizado sobre los dos pilares fundamentales sobre los que se basa el Máster cursado, como son la Inteligencia de Negocio y el Big Data. Por un lado, se ha desarrollado una herramienta para dar soporte en la toma de decisiones en un ámbito de negocio como es el mundo del baloncesto, a la hora de afrontar la mejor opción de compra de una entrada o abono de temporada. Por otro lado, en el desarrollo de este proyecto se han pasado por todas las etapas de tratamiento de datos dentro de un proyecto Big Data, como son los procesos ETL, el almacenamiento, así como la visualización final de los mismos, de manera que la información sea de utilidad y aporte valor.

Las conclusiones mas destacadas son:

1. **Dinamismo diario del ticket.** El “tablero” de las entradas disponibles para un evento concreto cambia diariamente y no siempre sigue una línea descendente. Si a medida que se acerca la fecha del evento se van realizando compras de entradas, lo esperable es que el número de entradas disponibles el día $i + 1$ siempre será menor que el el día i . Sin embargo, el análisis del dato recogido nos dice que esto no siempre es cierto. Esta información aporta algo de luz sobre como manejan las entradas los equipos profesionales. Por ejemplo, un abonado que tiene una localidad asignada libera su entrada para un evento concreto y rechaza su asistencia. El promotor del evento decide hacer disponible esa entrada y ponerla a la venta en el portal. Esta situación incrementa el número de asientos libres en el sector afectado, aumentando (en lugar de disminuir) la disponibilidad de entradas en el sector. Este

tipo de situaciones son muy habituales y se observa en casi todos los partidos monitorizados.

2. **Dinamismo en el precio del ticket.** Todo lo que pueda indicar un cambio en la demanda de los aficionados se utilizará para un mejor precio. Los precios entre partidos cambian dependiendo de su interés y la liberación de asientos al final del periodo de compra también es intencionado.
3. **Dinamismo en el diseño de la plataforma.** Gracias a las situaciones encontradas a lo largo del desarrollo del proyecto, se ha podido estudiar como la capacidad de automatización de una plataforma real es algo cambiante y pueden alterar los planes iniciales del proyecto. Estas situaciones, aún siendo desconocidas en un principio, han sido muy enriquecedoras y me han permitido obtener un mejor conocimiento sobre el contexto RPA y su ámbito de aplicación.
4. **Reproducibilidad del proceso.** Si bien este proyecto no podría utilizarse directamente sobre otro caso de uso, todo el proceso de automatización, captura del dato, visualización del dato, etc., realizado en este proyecto es trasladable a otros procesos que utilicen este mismo concepto de compra online: teatros, cine, conciertos, etc. La adaptación a diferentes ámbitos es lo que hace que una herramienta RPA tenga un ROI muy alto.
5. **Posibles malos usos de la tecnología.** En el contexto de venta de entradas, otra lectura a extraer es que un mal uso de esta tecnología permite crear “scalper bots”, generando robots de compra de entradas para una posible reventa y haciendo injusta la venta de entradas por Internet. En este proyecto solo se ha pretendido recoger información, pero añadiendo nuevos flujos de trabajo habría sido posible crear un robot con fines maliciosos.
6. **Simplifica la comprensión del dato.** El proceso de automatización aquí presentado ofrece información basada en cómo ha funcionado la venta de entradas en el pasado, al tiempo que va incorporando los flujos de datos del presente. De este modo, es el usuario el que entra en juego para aportar el punto de perspicacia y previsión, así como su habilidad para comprender el contexto tratado. Esto encaja directamente con la clasificación de aplicaciones DPA tratada en los conceptos teóricos 3.3.

Además, al inicio de la memoria formulábamos una serie de preguntas. Una vez finalizado el proyecto, es el momento de dar una respuesta a cada una de ellas:

- *¿Cualquier dato disponible en internet es útil y aporta información en la toma de decisiones?*

Toda la información disponible en internet NO es útil. Para extraer valor de un dato, se requiere aportarle estructura y transformarlo para poder trabajar con él. Pero para hacer todo esto es necesario recogerlo, poder disponer de él y almacenarlo. Por tanto, la más importante conclusión o mayor valor adquirido es mostrar, a través de este caso de uso, que es posible extraer de “cualquier” ámbito online información para nuestro beneficio. Las posibilidades que se abren aquí son tan amplias como proyectos queramos acometer.

- *¿Es posible automatizar la fase de extracción del dato, teniendo como origen un recurso web desestructurado?*

A lo largo de este proyecto hemos profundizado en el concepto del web scraping, Se ha aplicado y detallado en dos escenarios tecnológicos como proceder a implementarlo, mostrando que la automatización es posible, incluso en dos escenarios muy diferentes.

- *¿A día de hoy, existen herramientas que nos permiten realizar la automatización de todo el proceso, en este tipo de escenarios?*

En este proyecto se ha trabajado con la herramienta UiPath, líder en el sector del RPA, y que además permite integrarse con innumerables herramientas de terceros para potenciar su uso. En cualquier caso, existen muchas otras herramientas que también ofrecen similares funcionalidades, siendo de vital importancia el valorar lo amplio que es el espectro tecnológico que abarcan.

Además, se indica cómo se puede mejorar el proyecto, identificando varias líneas para continuar con el desarrollo realizado aquí:

1. Implementar el conector Tableau-UiPath para poder ejecutar el robot automatizado directamente desde el cuadro de mando. (B.7).

Debido a los imprevistos surgidos durante la realización del proyecto, este requisito inicial no se llegó a realizar. Se pretendía integrar la visualización desarrollada en Tableau con el proceso de automatización

desarrollado desde UiPath. UiPath ofrece un conector [51] desde el cual es posible lanzar directamente desde Tableau los trabajos automatizados y que están disponibles en el cloud, de tal manera que se podría incluir en el cuadro de mando un elemento (boton, link, etc.) que desencadene una llamada a la ejecución del proceso automatizado en UiPath. Con esto se conectan los dos entornos y tendríamos una solución central que contiene las dos funcionalidades: por un lado, la visualización del dato y por otro, el poder la inicializar bajo demanda los procesos automatizados que generan nuevos datos en los almacenes, que serían actualizados directamente en la visualización.

2. Desarrollo de más indicadores.

El dato puede tener muchos lados del que extraer información. En este proyecto hemos destacado algunos indicadores, pero sin duda hay muchos más que pueden trabajarse como son: un historial entre temporadas, comparativa de precios entre sectores/zonas/partidos, valorar las competiciones que tienen mas demanda, etc. Es decir, en este proyecto se ha desarrollado una visualización de datos de manera sencilla, pero es posible usar una amplia variedad de funciones para manipular datos que podrían potenciar aun más el proyecto.

3. Unir los dos procesos automatizados de recogida de datos.

Durante el desarrollo del proyecto se han creado dos proyectos de automatización: *TFMaster_FlowChartMode* y *TFM_Flow_v2*, correspondientes a Escenario 1 y Escenario 2 tratados en la memoria. La idea de esta mejora es unir estos dos procesos en uno, para que sea el proceso mismo el que diferencie entre uno u otro escenario. Hasta ahora, es el operador que lanza manualmente los procesos el que ejecuta uno u otro, dependiendo de si el escenario presente en la web de entradas es el identificado como 1 o el 2.

4. Valoración de otros canales de venta.

Ya que hay otros canales para realizar la compra de entradas, podría iniciarse un estudio sobre la conveniencia o no de continuar recopilando el dato desde la web elegida o, por el contrario, valorar otras alternativas.

5. Adaptación del proceso a otros eventos.

Siguiendo esta línea, se podría valorar otro tipo de eventos que utilicen este mismo canal de compra y este mismo pabellón como localización del evento, e intentar adaptar el proceso automatizado a otro tipo de

escenarios de venta de entradas. Es posible que desde este caso de uso sobre la venta de entradas de baloncesto no se alcance a valorar la repercusión del proyecto, pero si se planteara este proceso para un evento de mucha mayor trascendencia como puede ser una Super Bowl o las series finales de la NBA, podemos hacernos una idea de cuan valiosa puede ser la información extraída por este método.

Por último, destacar que personalmente, espero que este trabajo sea una fuente de referencia para aquellos que pretendan profundizar en el desarrollo de procesos automatizados RPA sobre plataformas de compra de entradas en diferentes escenarios posibles, como conciertos, teatro, eventos deportivos, cine, etc.

Apéndices

Apéndice A

Plan de Proyecto Software

En este apéndice se refleja la planificación temporal de las tareas realizadas durante el desarrollo del proyecto, así como un estudio de la viabilidad económica y legal del mismo.

A.1. Planificación temporal

Este trabajo se ha desarrollado dentro del marco del TFM del Máster en Inteligencia de Negocio y Big Data en Entornos Seguros. Acorde a la guía del mismo, el alcance y la duración del TFM se corresponde de 9 créditos ECTS, equivalentes a 225 horas. El proyecto se inició a primeros de octubre de 2022 y finalizó el 1 de junio de 2023, llevándose a cabo las diferentes tareas dentro de este intervalo de tiempo.

Cuando se conocen las actividades requeridas de nuestro proyecto, se definen las fases para iniciar el proyecto de automatización e identificar los medios para llevarlo a cabo. La planificación del proyecto comienza incluyendo una identificación de requisitos que se detallan en el Apéndice B, donde se definen tareas y acciones a llevar a cabo y la determinación de la estrategia de automatización de pruebas detallada en Sección 4.1.

Se plantea aquí por tanto una estructura de planificación sobre la construcción de diferentes fases, con el objetivo de agrupar cada uno de los requisitos detectados para el desarrollo de este proyecto de automatización. El diagrama creado con la representación temporal de cada una de estas fases puede comprobarse en Figura A.1. Concretamente, las fases del proyecto han sido las siguientes:

- Fase 1:** Documentación. Esta fase comprende las acciones de investigación de las diferentes posibilidades a la hora de implementar el proyecto en cuanto a arquitectura, modelos y herramientas a utilizar para llevar a cabo con éxito y además, las tareas realizadas periódicamente para documentar todo lo sucedido a lo largo del proyecto.
- Fase 2:** Preparación de la infraestructura. Una vez elegidas las herramientas y la metodología a seguir, en esta fase se realizó un estudio de las técnicas y herramientas concretas seleccionadas. Esto es, se realizó la puesta en marcha y configuración del entorno de desarrollo y del cloud para herramientas UiPath y Tableau. También se ha incluido en esta fase todo lo necesario para preparar la máquina destino donde se ejecutaran los procesos automatizados.
- Fase 3:** Desarrollo y almacenamiento del dato. En esta fase se realizó todo lo relativo a la codificación de los requisitos del proyecto, los cuales pueden consultarse en el Apéndice B.
- Fase 4:** Pruebas. En esta fase se realizaron pruebas sobre de los diferentes procesos automatizados desarrollados en este proyecto. Se incluye aquí también las acciones derivadas por los diferentes problemas encontrados durante la fase de desarrollo. Se puede consultar mas detalle en la sección C.3.
- Fase 5:** Puesta en producción y Soporte. Esta fase incluye lo necesario para la puesta en producción del proyecto. Desde el despliegue del primer escenario hasta los últimos retoques técnicos que han sido necesarios tanto en la parte del desarrollo del proceso como el la parte de visualización de los datos generados en Tableau.

Dentro de la planificación también se han tenido en cuenta los puntos clave de cualquier proyecto de desarrollo software, sea automatizado o no. La tarea de diseñar una arquitectura de automatización [28] está muy alineada con la arquitectura de un producto software. Por tanto, es necesario definir el espacio conceptual, que capas vamos a tratar, que servicios vamos a usar, etc., para permitir la realización de una solución de automatización completa, tanto internamente como externamente.

A.2. Estudio de viabilidad

En esta sección se va a detallar la viabilidad económica y legal del desarrollo de este proyecto.

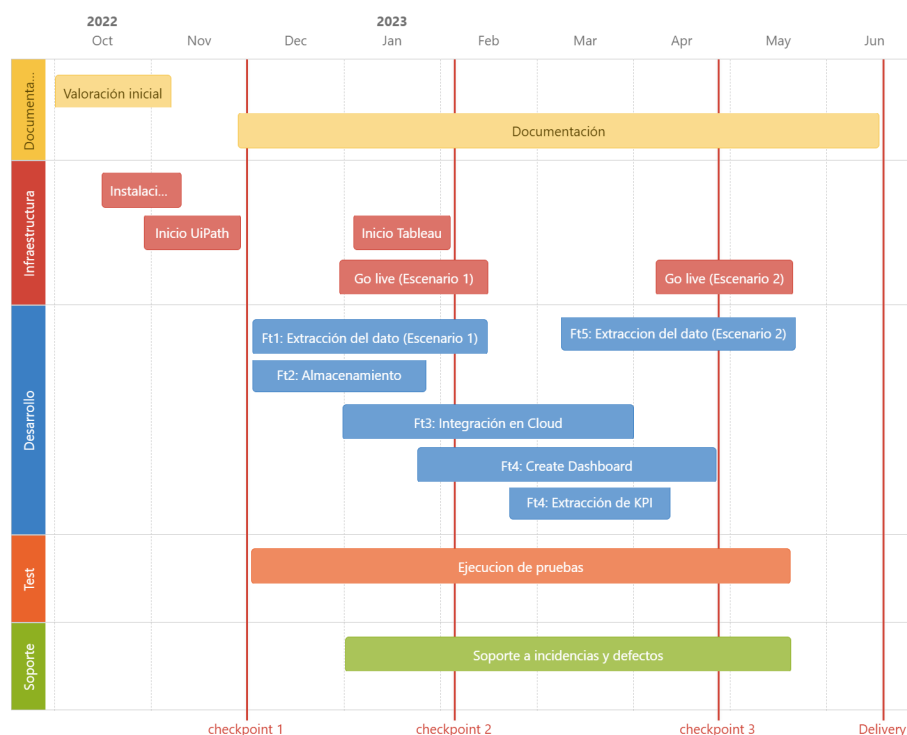


Figura A.1: Planificación TFM

Viabilidad económica

A partir de la planificación anterior y los recursos técnicos utilizados, podemos elaborar una previsión de la viabilidad económica del proyecto desarrollado. Se van a diferenciar para ellos dos tipos de recursos, como son los recursos técnicos y los recursos humanos.

Para el desarrollo del proyecto, tanto a nivel documental como de codificación, se ha llevado a cabo utilizando un ordenador sobremesa con las siguientes características:

- Procesador Intek i7 3770 3.40 Ghz
- Memoria RAM de 12GB DDR3 1600 MHz
- Disco de 512GB HDD SATA
- Disco de 1Tb HDD (Extraíble)
- Internet Fibra Smart 1Gb

Descripción	Meses	Coste mensual	Total
Recursos técnicos			
Ordenador personal	8	8,33€	66,59€
Internet	8	20€	160€
Recursos humanos			
Sueldo AP (media jornada)	8	958,08€	7664,64€
Cotización seguridad social	8	226,11€	1808,86€
Coste Total			9700,09€

Tabla A.1: Presupuesto del proyecto

Según la tabla aplicable a las sociedades [10], los equipos informáticos se amortizan o bien aplicando un porcentaje del 25 % sobre el precio de adquisición o bien en un máximo de 8 años. Siendo el precio total del mismo de 799€, la amortización anual del ordenador sería de 99,88€, lo que nos deja un coste mensual de 8,33€. Por tanto el coste asociado al tiempo de uso correspondería a 66,59€. En cuanto a gastos técnicos también incluimos el gasto de internet, que tiene un coste mensual de 20€. Por tanto, el coste total asociado a la duración del proyecto es de 226,64€.

Además, hay que tener en cuenta los recursos humanos utilizados. En este caso se tendrá en cuenta el rol de Analista Programador, encargado de analizar y desarrollar el código de las distintas implementaciones realizadas en el proyecto. Acorde al “XVII Convenio colectivo estatal de empresas de consultoría y estudios de mercado y de la opinión pública” [9] el sueldo bruto anual sería de 22.993,74€. En este caso se ha compatibilizado el desarrollo del proyecto con el trabajo diario, así que por simplicidad, se va a considerar la mitad de la jornada. El coste mensual a jornada completa sería de 1916,15€ y, por tanto, a media jornada, el coste se corresponde con 958,08€. El coste total de los ocho meses sería de 7664,64€. Finalmente, habría que añadirle el coste de la cotización a la seguridad social, que en este caso, acorde a las bases y tipos de cotización de 2022 [8], sería del 23,60 %. Todos estos costes pueden consultarse en la Tabla A.1.

Viabilidad legal

En esta sección se va a tratar la viabilidad legal del proyecto desarrollado como Trabajo Fin de Máster. Teniendo en cuenta los desarrollos realizados, en la Tabla A.2 se detallan las licencias utilizadas.

Las imágenes de los iconos de los distintos equipos de baloncesto utilizadas en la aplicación web provienen directamente de la web de venta y son

Aplicacion	Descripción	Licencia
Atlassian Confluence	SaaS	Licencia gratuita
Atlassian JIRA	SaaS	Licencia gratuita
UiPath Studio	Desktop IDE	Community Edition
UiPath Automation Cloud	SaaS	Community Edition
Tableau Desktop	Desktop IDE	Licencia para estudiantes (duración 12 meses)
Tableau Public	SaaS	Plataforma gratuita en línea
Office 365	SaaS	Programa para estudiantes y docentes (Uva)
GitHub	SaaS	Licencia con derechos de autor pre-determinadas
Overleaf	SaaS	Licencia libre

Tabla A.2: Licencias utilizadas para el desarrollo del proyecto

utilizados sin ningún fin comercial y sin perjudicar cualquier función de marca. Su uso tiene únicamente finalidad informativa sin pretender un efecto llamada o publicitario, únicamente como ámbito académico de este proyecto. En cuanto a la legalidad de la técnica Web Scraping, este tema fue tratado en detalle en la Sección 3.2.

Apéndice *B*

Especificación de Requisitos

Partiendo de la metodología y plan inicial, se indica el marco de trabajo para tratar el TFM con una visión de necesidades sobre especificaciones a alto nivel. Se desglosan los requisitos (Rq) y funcionalidades (Ft) que se han detallado a lo largo de la memoria, especificando en cada caso como abordar cada necesidad y a que proceso aplica. El detalle de los requisitos extraídos para el proyecto se muestra en el siguiente catalogo de requisitos.

B.1. Catalogo y Especificación de requisitos

A continuación, las siguientes tablas muestran los requisitos del proyecto para una mejor comprensión por parte del lector.

Acciones	Estudio del caso de uso planteado. Alcance y punto de partida
	Análisis de mercado
Objetivos	Definir contexto a tratar en el TFM. Escenario concreto del caso de uso a aplicar
	Selección de herramienta que permita el desarrollo de nuestros objetivos

Tabla B.1: Requisito 1: Valoración inicial del TFM. Alcance

Acciones	Instalación. Habilitar herramienta
Objetivos	Configuración + herramienta disponible

Tabla B.2: Requisito 2.1: Puesta en marcha de RPA: Solución UiPath.

Acciones	Instalación. Habilitar herramienta
Objetivos	Configuración + herramienta disponible

Tabla B.3: Requisito 2.2: Puesta en marcha de Tableau Desktop

Acciones	Creación de automatización para extraer la lista de eventos (partidos)
	Creación de automatización para inspeccionar las diferentes zonas/sectores del pabellón
Objetivos	Recopilar datos sobre eventos y zonas

Tabla B.4: Funcionalidad 1: Extracción del dato (Primer Escenario: SVG).

Acción	Almacenamiento interno de los datos recogidos.
Objetivos	Ofrecer un soporte a los datos para persistir su información y hacerla accesible a posteriori

Tabla B.5: Funcionalidad 2: Almacenamiento. Carga y tratamiento del dato

Acciones	Preparar y configurar UiPath cloud
	Subir código generado sobre automatizaciones al cloud
	Crear disparadores y prepara ejecuciones de los diferentes procesos creados
Objetivos	Preparar UiPath Orchestrator para hacer las automatizaciones disponibles desde la web

Tabla B.6: Funcionalidad 3.1: Integración del proyecto en Cloud

Acciones	Análisis del conector. Puesta en marcha
	Integrar Tableau con el conector UiPath
Objetivos	Configuración de conector y pruebas hasta su puesta en marcha

Tabla B.7: Funcionalidad 3.2: Habilitar conector Tableau. Ejecución directa desde cuadro de mando (Finalmente sacado del alcance)

Acciones	Crear esquema inicial. Bocetos
	Valoración de KPI a incluir.
	Crear flujos UiPath para extracción de estos KPIs
	Construcción de Dashboard en herramienta Tableau. Create + Test + Public
Objetivos	Habilitar panel de consulta sobre datos extraídos.

Tabla B.8: Funcionalidad 4: Creación de la visualización y Extracción de indicadores

Acciones	Análisis del cambio de tecnología sufrido en la web de compra de entradas.
	Estudio de cómo llevar a cabo las automatizaciones en el nuevo escenario
	Desarrollo de un refactorización del código o creación de nuevos procesos
Objetivos	Adaptar los procesos automatizados de recogida de datos a la nueva web de entradas

Tabla B.9: Funcionalidad 5: Extracción del dato (Segundo Escenario: Computer Vision).

Acciones	Verificación de procesos de automatización en los diferentes escenarios encontrados
Objetivos	Alcanzar la calidad esperada en los procesos. Reducción de falsos positivos/negativos

Tabla B.10: Requisito 3: Pruebas del proyecto.

Acciones	Lanzamiento/Ejecución de procesos automatizados
Objetivos	Recogida y almacenamiento de datos recogidos por la automatización

Tabla B.11: Requisito 4: Go live.

Acciones	Creación de contenido en plataforma de seguimiento del proyecto
Objetivos	Centralizar la información generada en una sola herramienta

Tabla B.12: Requisito 5.1: Documentación (proyecto software en Confluence)

Acciones	Creación de contenido en memoria del TFM
Objetivos	Generar documento entregable para defensa de proyecto

Tabla B.13: Requisito 5.2: Documentación (Memoria).

Apéndice C

Documentación técnica de programación

En este apéndice se va a explicar como poner en marcha los diferentes desarrollos realizados para la construcción de los procesos automatizados implementados en este proyecto.

C.1. Estructura de directorios

A continuación, se detalla la estructura de directorios de cada uno de los desarrollos realizados en este proyecto.

En primer lugar, el proyecto de automatización sobre el escenario 1 (datos SVG) se encuentra en el directorio *TFMaster/TFMaster_FlowChartMode/*. El propósito de esta automatización es disponer de un proceso automatizado que permita extraer datos sobre un evento deportivo desde una conocida página de venta de entradas. El método de extracción del dato es a través de los atributos que contienen los elementos web. En un ultimo pase se trasladan a archivos externos para persistir esta información. El árbol correspondiente a este proyecto es el siguiente:

```
.
├── FlowchartInit.xaml
├── RecoverCurrentEvents.xaml
├── FondoGoya.xaml
├── FondoJorgeJuan.xaml
├── LateralFelipe.xaml
└── LateralFuenteBerro.xaml
```

```

├── RecoverCurrentEvents.xaml
├── SectorScanning.xaml
├── SaveOnExcel.xaml
└── project.json

```

El segundo proyecto, la automatización en el segundo escenario (Computer Vision) se encuentra en el directorio *TFMaster/TFM_Flow_v2/*. El propósito de esta automatización es disponer de un proceso automatizado que permita extraer datos sobre un evento deportivos desde una conocida página de venta de entradas. El método de extracción del dato es vía OCR directamente según se muestra en pantalla web, con funcionalidades Computer Vision que ofrece UiPath. En un ultimo pase se trasladan a archivos externos para persistir esta información. El árbol correspondiente a este proyecto es el siguiente:

```

├── sectoresFondoGoya
│   ├── Goya_FONDO_EXTENSIBLE_BAJO.xaml
│   ├── Goya_FONDO_EXTENSIBLE_MEDIO.xaml
│   ├── Goya_FONDO_EXTENSIBLE_0.xaml
│   ├── Goya_FONDO_EXTENSIBLE_2.xaml
│   ├── Goya_FONDO_EXTENSIBLE_4.xaml
│   └── Goya_FONDO_EXTENSIBLE_6.xaml
├── sectoresFondoJorgeJuan
│   ├── JorgeJuan_FONDO_EXTENSIBLE_BAJO.xaml
│   ├── JorgeJuan_FONDO_EXTENSIBLE_MEDIO.xaml
│   ├── JorgeJuan_FONDO_EXTENSIBLE_0.xaml
│   └── JorgeJuan_FONDO_EXTENSIBLE_2.xaml
├── sectoresLateralBerro
│   ├── Berro_FONDO_EXTENSIBLE_BAJO.xaml
│   ├── Berro_FONDO_EXTENSIBLE_MEDIO.xaml
│   ├── Berro_FONDO_EXTENSIBLE_0.xaml
│   ├── Berro_FONDO_EXTENSIBLE_2.xaml
│   └── Berro_FONDO_EXTENSIBLE_2 SUPERIOR.xaml
├── sectoresLateralFelipe
│   ├── Felipe_FONDO_EXTENSIBLE_BAJO.xaml
│   ├── Felipe_FONDO_EXTENSIBLE_MEDIO.xaml
│   ├── Felipe_FONDO_EXTENSIBLE_0.xaml
│   ├── Felipe_FONDO_EXTENSIBLE_2.xaml
│   └── Felipe_FONDO_EXTENSIBLE_2 SUPERIOR.xaml
├── Flowchart_Init.xaml
└── RecoverCurrentEvents.xaml

```

```
├── CollectData.xaml
├── SectorScanning_v2.xaml
├── FondoGoya.xaml
├── FondoJorgeJuan.xaml
├── LateralFuenteBerro.xaml
├── LateralFelipe.xaml
├── AddSectorData.xaml
├── SaveOnExcel.xaml
└── project.json
```

Finalmente, el proyecto de automatización de la generación de los KPI se encuentra en el directorio *TFMaster/TFMaster_Transform_*/. El propósito de esta automatización es disponer de un proceso automatizado que permita generar los indicadores definidos desde los datos existentes en almacenes locales. La transformación del dato realizada en este flujo, se etiqueta en diferentes KPI y se vuelca en un nuevo archivo externo para su posterior tratamiento. El árbol correspondiente a este proyecto es el siguiente:

```
├── Main.xaml
├── OpenExcelFiles.xaml
├── ManipulateData.xaml
├── ExtractInfobysector.xaml
├── SaveOnExcel.xaml
└── project.json
```

Repositorios utilizados

La solución de automatización planteada en este proyecto tiene su propio ciclo de vida de software. Hay que recordar que los diferentes proyectos creados, como cualquier otro software, necesita estar bajo control de versiones y sus prestaciones documentadas. De lo contrario, se hace muy difícil desplegar diferentes partes de la misma y hacerlas trabajar juntas, o trabajar en ciertos entornos. Se ha utilizado para la gestión de versiones el sistema GitHub, una plataforma de alojamiento, propiedad de Microsoft, que ofrece a los desarrolladores la posibilidad de crear repositorios de código y guardarlos en la nube de forma segura, usando un sistema de control de versiones llamado Git. GitHub permite centralizar el contenido del repositorio para poder colaborar con los otros miembros de nuestra organización. El código fuente de los proyectos desarrollados se aloja en el siguiente repositorio:

<https://github.com/iGitMoreno/TFMaster>

C.2. Instalación y ejecución del proyecto

A continuación, se detalla el procedimiento de despliegue, destacando su claridad y facilidad de puesta en marcha. Este procedimiento depende de la versión; por lo tanto, también está incluido en el control de versiones.

1. **Instalación local de procesos** Importar proyectos desde repositorio Git y abrir proyecto desde la aplicación UiPath Studio.
2. **Dependencias de software**
 - 2.1 UiPath Assistant debe estar instalado y configurado en el sistema destino, donde se ejecutarán las automatizaciones.
 - 2.2 Es necesaria la instalación de la extensión correspondiente al navegador usado¹ para habilitar el proceso de lanzar la automatización en el mismo.
 - 2.3 Desde Uipath Studio los proyectos deben subirse al Orquestador que se ha configurado y automáticamente estarán disponibles para su uso.
 - 2.4 El directorio `%APPDATA %_TFMExcelData`, que es usado para guardar los ficheros de almacenes de datos generados por la automatización, debe estar creado en la máquina destino. La automatización creará un fichero por cada una de las zonas del pabellón.
 - 2.5 La ejecución de los flujos desde el Orquestador (Cloud) realiza operaciones de E/S en ficheros Excel. Por tanto Excel debe estar disponible en el sistema donde se ejecutará el proceso de automatización.

C.3. Incidencias encontradas

Durante el desarrollo del proyecto se han detectado varias incidencias que se han ido corrigiendo, tanto incidencias externas, como la ya comentada sobre el cambio de web, e incidencias internas propias de la automatización del desarrollo software.

A continuación, las Tablas [C.1](#) y [C.2](#) detallan los diferentes problemas encontrados durante la fase de desarrollo en UiPath. Esto pretende ofrecer una guía y ayuda al futuro programador.

#1	Object invoked has disconnected from its clients
#2	UiPath swaps day with Month
#3	Uso de EnvironmentVariables
#4	Ajuste sobre apertura de compra de asientos a abonados/socios y zonas VIP
#5	Zonas sin localidades disponibles

Tabla C.1: Incidencias relacionadas directamente con el código del proyecto.

#1	Según avanza la venta de localidades el número de asientos se incrementa.
#2	Cambio en tecnología web para proceso de adquisición de entradas.

Tabla C.2: Incidencias relacionadas con la recogida del dato.

Para obtener información detallada sobre la gestión del proyecto y las incidencias encontradas, pueden consultar el siguiente enlace:

<https://tfmaster.atlassian.net/wiki/spaces/TFM/pages/5144577/Relacion+de+Problemas+encontrados>

¹<https://docs.uipath.com/studio/standalone/2022.10/user-guide/about-extensions>

Bibliografía

- [1] R. Aguilar. El poder de Tableau Public (report). <https://blog.comunidados.com/el-poder-de-tableau-public-ced88e59e73d>, 2022. [Internet; Accedido 12 de mayo de 2023].
- [2] Arbusta. Web scrapping: qué es y qué desafíos de negocio resuelve (web). <https://arbusta.net/web-scrapping-desafios-soluciones/>, 2021. [Internet; Accedido el 12 de mayo de 2023].
- [3] Atlassian. Confluence (Web). <https://www.atlassian.com/software/confluence>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [4] Atlassian. Jira Software (Web). <https://www.atlassian.com/software/jira>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [5] Atlassian. What is kanban? (Web). <https://www.atlassian.com/agile/kanban>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [6] AWS. Increase Operational Efficiency and Drive Faster Business Outcomes with UiPath Robots on AWS (Web). <https://aws.amazon.com/es/blogs/apn/increase-operational-efficiency-and-drive-faster-business-outcomes-with-uipath-robots-on-aws/>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [7] J. Beresford. State of the Self Service BI market Jan 2022 (report). https://www.linkedin.com/pulse/state-self-service-bi-market-jan-2022-james-beresford/?trk=articles_directory, 2022. [Internet; Accedido 12 de mayo de 2023].

- [8] Boletín oficial de Estado. Bases y Tipos de Cotización (report). <https://www.seg-social.es/Trabajadores/CotizacionRecaudacionTrabajadores/3653>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [9] Boletín oficial de Estado. Convenio colectivo estatal de empresas (report). <https://www.boe.es/boe/dias/2018/03/06/pdfs/BOE-A-2018-3156.pdf>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [10] Boletín oficial de Estado. Tablas amortización (report). https://se.de.agenciatributaria.gob.es/Sede/ayuda/manuales-videos-folleto/manuales-practicos/folleto-actividades-economicas/3-impuesto-sobre-renta-personas-fisicas/3_5-estimacion-directa-simplificada/3_5_4-tabla-amortizacion-simplificada.html, 2023. [Internet; Accedido 12 de mayo de 2023].
- [11] T. Clancy. UiPath Community Grows to More Than 1.5 Million Members (report). <https://www.uipath.com/newsroom/uipath-community-grows-to-more-than-1.5-million-members-and-uipath-announces-three-new-features-to-increase-career-opportunities>, 2021. [Internet; Accedido 12 de mayo de 2023].
- [12] V. Colmenarejo. Merchandising y abonados, los puntos negros de la gestión del Real Madrid de basket (Web). <https://karusito.com/2022/09/12/merchandising-y-abonados-los-puntos-negros-de-la-gestion-del-madrid-de-basket/>, 2022. [Internet; Accedido el 12 de mayo de 2023].
- [13] Davoy.tech. Power BI vs Tableau vs Google Data Studio (in 2021) (Web). <https://davoy.tech/power-bi-vs-tableau-vs-google-data-studio-in-2021/>, 2021. [Internet; Accedido el 12 de mayo de 2023].
- [14] Dell. Datos no estructurados (web). <https://www.dell.com/es-es/dt/learn/data-storage/unstructured-data.htm>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [15] R. Diouf, E. N. Sarr, O. Sall, B. Birregah, M. Bousso, and S. N. Mbaye. Web Scraping: State-of-the-Art and Areas of Application. *International Conference on Big Data (Big Data)*, pages 6040–6042, 2019.
- [16] J. S. Dyer. A Time-Sharing Computer Program for the Solution of the Multiple Criteria Problem. *Management Science*, 19(12):1379–1383, 1973.

- [17] FinddataLab. Web scraping service: Turn any website into data and have it delivered directly to you in any format (Web). <https://finddataLab.com/>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [18] Forrester. The Forrester Wave (TM): Robotic Process Automation, Q1 2023 (Web). https://www.uiopath.com/resources/automation-analyst-reports/forrester-wave-rpa?utm_source=LinkedIn&utm_medium=CPC&utm_term=SponsoredContent&utm_content=forresterwave2021&utm_campaign=AR21FWR, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [19] Gartner. 2022 Gartner Magic Quadrant (TM) for Robotic Process Automation (report). <https://www.uiopath.com/blog/rpa/gartner-magic-quadrant-rpa-report>, 2022. [Internet; Accedido 12 de mayo de 2023].
- [20] Gartner. Automatización de procesos robotizados (web). <https://www.gartner.es/es/finanzas/insights/robotica-en-las-finanzas>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [21] Gigantes. Las pistas de baloncesto con más capacidad del mundo (web). <https://www.gigantes.com/mas-basket/pistas-baloncesto-con-mas-capacidad/>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [22] I. González-Moreno. Basketball Season Analytics Dashboard (Web). https://public.tableau.com/app/profile/ignacio.vdata/viz/TFM_Test/Dashboard1, 2023. [Internet; Accedido 12 de mayo de 2023].
- [23] I. González-Moreno. Contexto: Baloncesto + pabellón (wizink center) tfmaster. <https://tfmaster.atlassian.net/wiki/spaces/TFM/pages/6848513/Contexto+Baloncesto+Pabellon+Wizink+Center>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [24] I. González-Moreno. TFM Confluence (tfmaster). <https://tfmaster.atlassian.net/wiki/home>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [25] I. González-Moreno. TFM JIRA (tfmaster). <https://tfmaster.atlassian.net/wiki/home>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [26] P. Hofmann, C. Samp, and U. N. Robotic process automation. *Electron Markets* 30, 6:99–106, 2020.

- [27] M. Hron and N. Obwegeser. Why and how is scrum being adapted in practice: A systematic review. *Journal of Systems and Software*, 183(6):0164–1212, 2022.
- [28] ISTQB. Ingeniero de Automatización de Pruebas (Report). https://es.sstqb.com/_files/ugd/acfdb9_ced9879fa2e64e0b8b71eb1f011f2094.pdf, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [29] H. Losbichler, L. Perkhofer, P. Hofer, and J. Winzer. Reporting 2.0 – Interactive Visualizations and Dashboards for Big Data (Report). <https://accid.org/wp-content/uploads/2021/12/REPORT1.pdf>, 2021. [Internet; Accedido el 12 de mayo de 2023].
- [30] R. Madrid. Web oficial de venta de entradas del Real Madrid. acceso 1. <https://realmadrid.entradas.com/sports-web/event/RMB>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [31] R. Madrid. Web oficial de venta de entradas del Real Madrid. acceso 2. <https://www.realmadrid.com/entradas/baloncesto>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [32] M. Marabelli, S. Newell, and V. Handunge. The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *Journal of Strategic Information Systems*, 30(1):1–15, 2022.
- [33] T. Marks. Why ticket prices are going through the roof (web). <https://www.consumerreports.org/money/why-ticket-prices-are-going-through-the-roof/>, 2016. [Internet; Accedido 12 de mayo de 2023].
- [34] N. Ostdick. Everything you need to know about screen scraping software (web). <https://www.uipath.com/blog/rpa/screen-scraping-software-everything-you-need-to-know>, 2017. [Internet; Accedido el 12 de mayo de 2023].
- [35] ProWebScraper. Top 5 myths of price scraping legality (web). <https://prowebscraper.com/blog/is-price-scraping-legal/>, 2022. [Internet; Accedido el 12 de mayo de 2023].
- [36] ProWebScraper. No Code Web Scraping Tool (Web). <https://prowebscraper.com/>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [37] Queue-it. Everything you need to know about ticket bots (web). <https://queue-it.com/blog/ticket-bots/>, 2023. [Internet; Accedido 12 de mayo de 2023].

- [38] Queue-it. How do tickets sell out in seconds? Debunking the instant sellout.ºnsale myth (Web). <https://academy.uipath.com/learning-plans/uipath-rpa-associate-certification-training>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [39] Real Madrid. Abonos Baloncesto 2022/2023 (Web). <https://www.realmadrid.com/entradas/baloncesto/abonos-temporada-22-23>, 2022. [Internet; Accedido el 12 de mayo de 2023].
- [40] Real Madrid. WiZink Center (Web). <https://www.realmadrid.com/baloncesto/pabellon>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [41] P. Satyavolu. Decision-Making: The New Frontier for Automation (Report). https://www.cognizant.com/en_us/insights/documents/decision-making-the-new-frontier-for-automation-codex-5467.pdf, 2020. [Internet; Accedido el 12 de mayo de 2023].
- [42] Smartsheet. Automation in the workplace 2017 (web). <https://www.smartsheet.com/sites/default/files/smartsheet-automation-workplace.pdf>, 2017. [Internet; Accedido el 12 de mayo de 2023].
- [43] Tableau. Get Answers from the Tableau Community (Web). <https://community.tableau.com/s/explore-forums>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [44] Tableau. Tableau para estudiantes (web). <https://www.tableau.com/es-es/academic/students>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [45] Teamim. Measuring the ROI of Robotic Process Automation (RPA) Implementation (Web). <https://www.teamim.com/blog/measuring-roi-for-rpa/>, 2021. [Internet; Accedido el 12 de mayo de 2023].
- [46] Telefonicatech. ¿qué son los datos no estructurados? (web). <https://aiofthings.telefonicatech.com/recursos/datapedia/datos-no-estructurados>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [47] TickPick. How The Ticket Industry Works | From the Artist to Ticket-Master to Brokers to You (Web). <https://www.tickpick.com/blog/ticket-industry-ticket-resale-ticketmaster/>, 2019. [Internet; Accedido 12 de mayo de 2023].

- [48] UiPath. Accelerating digital transformation (web). <https://www.uipath.com/blog/product-and-updates/how-cios-can-achieve-end-to-end-automation>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [49] UiPath. Automation Cloud, About Licensing (Web). https://www.uipath.com/resources/automation-analyst-reports/forrester-wave-rpa?utm_source=LinkedIn&utm_medium=CPC&utm_term=SponsoredContent&utm_content=forresterwave2021&utm_campaign=AR21FWR, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [50] UiPath. Communication between components (web). <https://www.softoneconsultancy.com/uipath-associate-certification-training/>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [51] UiPath. UiPath Add-In For Tableau (Web). <https://docs.uipath.com/add-ins/other/latest/user-guide/tableau-about>, 2023. [Internet; Accedido 12 de mayo de 2023].
- [52] UiPath. UiPath Plattform Report. https://html.cdn.contenttrave.com/crcloud/crscorm/uploads/uipath_lms_11218/encryptedfile/507141/v3.0/scormcontent/assets/x30CuNT8nnPjqt7_PM9SE0DDLsvHIwYz-The-UiPath-Platform.pdf, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [53] Wikipedia. Palacio de los Deportes (Web). https://es.wikipedia.org/wiki/Palacio_de_Deportes_de_la_Comunidad_de_Madrid, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [54] WizinkCenter. Aforos y configuraciones Report. <https://www.wizinkcenter.es/descargas/comunicacion/AforosWiZinkCenter.pdf>, 2023. [Internet; Accedido el 12 de mayo de 2023].
- [55] R. T. Yarlagadda. The RPA and AI Automation. *International Journal of Creative Research Thoughts (IJCRT)*, 6, 2018.
- [56] S. Zhao. What is ETL? (Web). <https://www.edq.com/blog/what-is-etl-extract-transform-load/>, 2018. [Internet; Accedido 12 de mayo de 2023].