

GENERACIÓN DE VISUALES Y ANALÍTICAS A PARTIR DE DATOS OBTENIDOS POR UN RPA EN EMPRESA DEL SECTOR EDUCATIVO

Gaviria Moncayo Jeison Esteban
Giraldo Mena Luisa Fernanda

Nota de Aceptación

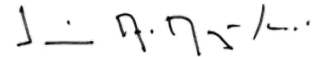
Certificamos que el presente Trabajo de Grado Satisface, en alcances y calidad, todos los requisitos que demanda un Trabajo de Grado de Maestría.



Director

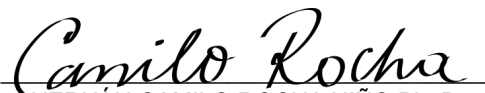


Jurado




Jurado

Aprobado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana Cali, para optar el título de Magister en Ciencia de Datos.



HERNÁN CAMILO ROCHA NIÑO Ph. D.
Decano Facultad de Ingeniería y Ciencias



JUAN CARLOS MARTÍNEZ ARIAS
Director Posgrados de Ingeniería y Ciencias



Acta de Correcciones al Documento de Trabajo de Grado

Santiago de Cali, 5 de julio del 2023

Autor: Gaviria Moncayo Jeison Esteban, Giraldo Mena Luisa Fernanda

Título del Trabajo de Grado: “GENERACIÓN DE VISUALES Y ANALÍTICAS A PARTIR DE DATOS OBTENIDOS POR UN RPA EN EMPRESA DEL SECTOR EDUCATIVO”

Director: Juan Carlos Martínez Arias

Como indica el artículo 2.13 de las Directrices para Trabajo de Grado de Maestría, he verificado que el estudiante indicado arriba ha implementado todas las correcciones que los Jurados del Proyecto de Trabajo de Grado definieron que se efectuaran, como consta en el Acta de Evaluación correspondiente.



Firma del Director del Trabajo de Grado

Santiago de Cali, 29 de Mayo del 2023

Doctora

Gloría Inés Alvarez V.

Directora Maestría en Ciencia de Datos
Facultad de Ingeniería y Ciencias
Pontificia Universidad Javeriana de Cali

Asunto: Presentación para evaluación del proyecto aplicado

Cordial Saludo,

Con el fin de cumplir con los requisitos exigidos por la Universidad para optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto denominado “GENERACIÓN DE VISUALES Y ANALÍTICAS A PARTIR DE DATOS OBTENIDOS POR UN RPA EN EMPRESA DEL SECTOR EDUCATIVO”, el cual fue realizado por los estudiantes Jeison Esteban Gaviria Moncayo y Luisa Fernanda Giraldo Mena con códigos 5478378, 02232532 respectivamente, pertenecientes a la Maestría en Ciencia de Datos, bajo la dirección de Juan Carlos Martínez Arias.

El suscrito director del Proyecto Aplicado autoriza para que se proceda a hacer la evaluación de este proyecto, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado y sustentado oficialmente.

Atentamente,



Jeison Esteban Gaviria Moncayo
C.C. 1.085.339.491 de Pasto



Luisa Fernanda Giraldo Mena
C.C. 1.151.960.065 de Cali



Juan Carlos Martínez Arias
C.C. 7549319 de Armenia



Santiago de Cali, 9 de Julio de 2022

Alianza Superior SAS

NIT: 900.556.828-2

CERTIFICA QUE:

Que ESTEBAN GAVIRIA, identificado con cédula de ciudadanía 1.085.339.491 está vinculado con Alianza Superior/ Griky como DATA ANALYST, Se encuentra realizando un proyecto para nuestra compañía donde está realizando el uso de modelos de Ciencia de datos y Machine Learning. Conocemos y respaldamos el proceso educativo de nuestros empleados por esta razón autorizamos que este proyecto pueda usado externamente exclusivamente con fines académicos al igual que todos los datos que de este se requieran.

Atentamente,



Michael Stiven Zapata Cordoba
Coordinador De Inteligencia De Negocios Y Analítica
Michael.zapata@griky.co
318 444 4396

FICHA RESUMEN

PROYECTO APLICADO – MAESTRÍA EN CIENCIA DE DATOS

TÍTULO: GENERACIÓN DE VISUALES Y ANALÍTICAS A PARTIR DE DATOS OBTENIDOS POR UN RPA EN EMPRESA DEL SECTOR EDUCATIVO

1. **ÁREA DE TRABAJO:** Analítica, visualización y automatización.
2. **TIPO DE PROYECTO (Aplicado, Innovación, Investigación):** Aplicado.
3. **ESTUDIANTE(S):** Gaviria Moncayo Jeison Esteban, Giraldo Mena Luisa Fernanda
4. **CORREO ELECTRÓNICO:**
estebangaviria13@javerianacali.edu.co
luisagiraldo96@javerianacali.edu.co
5. **DIRECCIÓN Y TELEFONO:**
6. **DIRECTOR:** Juan Carlos Martínez Arias
7. **VINCULACIÓN DEL DIRECTOR:** Profesor Pontificia Universidad Javeriana.
8. **CORREO ELECTRÓNICO DEL DIRECTOR:** juancmartinez@javerianacali.edu.co
9. **OTROS GRUPOS O EMPRESAS:** Griky.
10. **PALABRAS CLAVE (al menos 5):** analítica, visualización, web scraping, Python, robot, automatización, PowerBi.
11. **FECHA DE INICIO:** julio 2022.
12. **FECHA DE FINALIZACIÓN:** 8 Meses.
13. **RESUMEN:**

Para una compañía cuyo foco es ofrecer educación virtual a bajo costo y de alta calidad, uno de sus mayores retos es lograr conocer el comportamiento de sus usuarios mediante visuales y analíticas que permitan tomar decisiones en pro de su crecimiento. Actualmente, el principal problema se basa en la obtención de los datos, ya que se realiza de manera manual y puede generar errores en la descarga. Además, requiere de recursos humanos dedicados a esta tarea, incrementando los gastos de la compañía a medida que el volumen de datos aumenta. Por este motivo se desarrolló un robot que permite automatizar la descarga de datos para su posterior análisis y visualización. Para ello, inicialmente se realizó una búsqueda de literatura acerca de trabajos relacionados que sirvieron de guía para el desarrollo del proyecto. Posteriormente, se definieron las herramientas a utilizar y su debida documentación. Después, se dio inicio al desarrollo del robot y se realizaron pruebas de funcionamiento con data real. Para finalizar, se desarrollaron visualizaciones y analíticas a partir de los datos obtenidos que permitieron cumplir con el objetivo del proyecto.



Pontificia Universidad
JAVERIANA
Cali

**GENERACIÓN DE VISUALES Y ANALÍTICAS A PARTIR DE DATOS OBTENIDOS
POR UN RPA EN EMPRESA DEL SECTOR EDUCATIVO**

*Gaviria Moncayo Jeison Esteban¹, Giraldo Mena Luisa Fernanda²
54783781, 02232532*

*Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos*

Director
Juan Carlos Martínez Arias

FACULTAD DE INGENIERÍA Y CIENCIAS
MAestrÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, MAYO 29 DE 2023

TABLA DE CONTENIDO

	Página
FIGURAS.....	6
TABLAS.....	7
ANEXOS.....	8
INTRODUCCIÓN	9
1. DEFINICIÓN DEL PROBLEMA.....	10
1.1. PLANTEAMIENTO DEL PROBLEMA	10
1.2. FORMULACIÓN DEL PROBLEMA	11
2. OBJETIVOS DEL PROYECTO.....	12
2.1. OBJETIVO GENERAL	12
2.2. OBJETIVOS ESPECÍFICOS	12
2.3. RESULTADOS ESPERADOS.....	12
3. MARCO TEÓRICO DE REFERENCIA Y ANTECEDENTES.....	13
3.1. AUTOMATIZACIÓN ROBÓTICA DE PROCESOS- RPA.....	13
3.2. WEB SCRAPING.....	14
3.2.1. Lenguajes y Librerías más utilizadas	15
3.2.2. Usos del Web Scraping.....	15
3.3. ANÁLISIS DE DATOS.....	16
3.3.1. Usos del análisis de datos.....	16
3.3.2. Técnicas de análisis de datos.....	17
3.3.3. Tipos de análisis de datos	18
3.3.4. Métodos para el análisis de datos	18
3.3.5. Aprendizaje automático y análisis de datos	19
3.4. VISUALIZACIÓN DE DATOS.....	19
3.4.1. Ventajas de la visualización de datos	20
3.4.2. Gráficos comunes.....	20
3.4.3. Aprendizaje automático y análisis de datos	21
3.5. TRABAJOS RELACIONADOS.....	21
4. SELECCIÓN DE HERRAMIENTAS.....	24
4.1. HERRAMIENTAS PARA DESARROLLO DE RPA.....	25
4.1.1. Lenguajes de programación	25
4.1.2. Herramientas de automatización de pruebas.....	26

4.1.3.	Navegadores web	27
4.1.4.	Almacenamiento en nube	28
4.1.5.	Entornos de desarrollo integrado.....	29
4.2.	HERRAMIENTAS PARA ANÁLISIS DE DATOS Y VISUALIZACIÓN	30
5.	DESARROLLO DEL RPA	32
5.1.	DEFINICIÓN DE PARÁMETROS Y VARIABLES	32
5.2.	CONSTRUCCIÓN DEL ROBOT	33
5.2.1.	Importación de librerías	33
5.2.2.	Definición de función principal.....	35
5.2.3.	Acceso a la plataforma	35
5.2.4.	Conteo de cursos	35
5.2.5.	Selección de filtros	36
5.2.6.	Ciclo repetitivo de descarga	37
5.2.7.	Lectura, consolidación y carga de archivos.....	37
5.3.	Inconvenientes presentados en el desarrollo del robot	39
5.4.	Representación gráfica del proceso	40
5.5.	Prueba de Funcionamiento del RPA	43
6.	PROCESAMIENTO DE DATOS	46
7.	RESULTADOS VISUALES Y ANALÍTICAS	49
7.1.	Gráficos seleccionados	49
7.2.	Dashboard Progreso de Usuarios en Griky.....	57
8.	CONCLUSIONES Y TRABAJOS FUTUROS	60
8.1.	CONCLUSIONES.....	60
8.1.1.	Conclusiones generales	60
8.1.2.	Conclusiones sobre el análisis de datos.....	60
8.1.3.	Conclusiones sobre ciencia de datos	60
8.2.	TRABAJOS FUTUROS.....	62
9.	REFERENCIAS BIBLIOGRÁFICAS	63
	ANEXOS.....	66

FIGURAS

	Página
Figura 1. Código Importación de librerías	34
Figura 2. Definición de función principal	35
Figura 3. Acceso a la plataforma	35
Figura 4. Conteo de cursos	36
Figura 5. Selección de filtros	36
Figura 6. Ejemplo resultado	37
Figura 7. Ciclo repetitivo de descarga	37
Figura 8. Lectura, consolidación y carga de archivos	38
Figura 9. Solución problemas con iframe	39
Figura 10. Flujo RPA	41
Figura 11. Flujo de proceso de RPA para todos clientes.	42
Figura 12. Diagrama de flujo función get_Progreso_Interno	46
Figura 13. Diagrama de Flujo transformación de datos	47
Figura 14. Diagrama de Flujo Modelo de datos	48
Figura 15. Resumen General	49
Figura 16. Métrica General: Monthly Active Users	50
Figura 17. Cantidad de Matrículas por Estado	51
Figura 18. Tasa de Finalización por Cliente	52
Figura 19. Total Cursos Inscritos VS Cursos Finalizados por Cliente	53
Figura 20. Cantidad de Matrículas por Estado- Cliente: Classrun	54
Figura 21. Total matrículas VS Total matrículas Finalizadas por Curso y Cliente: Classrun	54
Figura 22. Métrica General: Monthly Active Users- Cliente: Classrun	55
Figura 23. Dashboard Progreso Usuarios en Griky 1	58
Figura 24. Dashboard Progreso Usuarios en Griky 2	59

TABLAS

Página

Tabla 1. Comparación de tecnologías de automatización: RPA, Inteligencia Artificial y Machine Learning	14
Tabla 2. Trabajos relacionados con herramientas de automatización.	23
Tabla 3. Lenguaje de Programación Python Vs R.....	25
Tabla 4. Selenium Vs Puppeteer.....	26
Tabla 5. Google Chrome Vs Mozilla Firefox	28
Tabla 6. One Drive Vs Google Drive	29
Tabla 7. Visual Studio Code Vs PyCharm.....	29
Tabla 8. Power BI Vs Tableau	30
Tabla 9. Variables Definidas.....	32
Tabla 10. Parametros Definidos	33
Tabla 11. Tiempo Descarga Manual VS Tiempo Descarga por RPA en horas.....	43

ANEXOS

	Página
Anexo 1 Dashboard- Proyecto-griky.....	67
Anexo 2 get_Progreso_Interno.....	69
Anexo 3 get_Usuarios_Interno.....	69
Anexo 4 Progreso_Interno	70
Anexo 5 Usuarios_Interno	71

INTRODUCCIÓN

Actualmente, en el mundo empresarial, la innovación es clave para el éxito de las compañías. Como afirmó el autor Jay Samit, "Para tener éxito, la innovación no se trata solo de la creación de valor, sino de la captura de valor". Esta idea plantea un desafío fundamental para las empresas: no solo deben crear ideas innovadoras que generen valor, sino también asegurarse de mantenerlo y evitar caer en una zona de confort.

En este contexto, Griky, una plataforma de educación en línea ha aceptado este desafío y se ha convertido en una empresa comprometida en brindar educación de alta calidad a bajo costo. Griky se ha posicionado como una red que facilita la transformación universitaria al conectar diferentes universidades con proveedores de contenido a través de una plataforma de aprendizaje permanente. Su objetivo es ofrecer a los usuarios la oportunidad de acceder a un aprendizaje ilimitado, flexible y pertinente para toda la vida.

Ante un futuro cada vez más incierto, Griky ha desarrollado rápidamente plataformas de aprendizaje permanente con marca propia para universidades, que incluyen cursos ilimitados y garantizan una implementación ágil y sencilla. Para la compañía, la capacidad de tomar decisiones rápidas y efectivas es de vital importancia, dado que la cantidad de datos y usuarios se encuentra en constante crecimiento.

En el marco de este proyecto, se ha creado una herramienta RPA (Automatización Robótica de Procesos) que desempeña tareas repetitivas de descarga de datos de manera automática y diaria. Esto ha permitido que Griky cuente con datos relacionados al progreso de los usuarios en todas sus plataformas, generando análisis y visualizaciones que brindan una comprensión profunda del comportamiento del consumidor. Estas herramientas analíticas son fundamentales para respaldar la toma de decisiones estratégicas de la compañía.

En resumen, el proyecto de implementación de la herramienta RPA en Griky ha sido una iniciativa exitosa que ha fortalecido la capacidad de la compañía para comprender y responder a las necesidades de sus usuarios. A través de la automatización de procesos y el análisis de datos, Griky ha logrado capturar y mantener el valor generado por sus ideas innovadoras. Este enfoque ha permitido que la empresa logre mantenerse a la vanguardia en el mercado de la educación en línea y seguir brindando una experiencia educativa de alta calidad a un costo accesible.

1. DEFINICIÓN DEL PROBLEMA

Los datos son el insumo principal para generar certificados de finalización de cursos; su principal función es ayudar a la toma de decisiones por medio de analíticas y visuales que permiten comprender el comportamiento de los usuarios en la plataforma. Actualmente, el proceso de descarga de información es demorado, tedioso y al ser manual puede incurrir en errores en los reportes generados: omitiendo documentos, generando duplicados, procedimientos repetitivos y no automatizados. Según Michael Zapata, Coordinador de Business Intelligence en Griky, una persona en promedio puede demorarse entre 1.5 y 2 horas en descargar reportes de información para uno de los clientes de la compañía. Además, el procedimiento actual no es escalable: necesita incrementar la cantidad de recursos humanos a medida que la plataforma crece para realizar tareas repetitivas que no generan valor en el proceso de negocio de la empresa.

1.1. PLANTEAMIENTO DEL PROBLEMA

Griky es una plataforma de educación en línea con sede compartida entre Bogotá y Miami que busca conectar a las universidades con los mejores proveedores de contenido mediante una plataforma de aprendizaje permanente ofreciendo cursos a bajo costo con alta calidad [1].

Una de las actividades de mayor importancia para la plataforma es conocer y analizar los datos que se generan para crear analíticas que permiten tomar decisiones en pro del crecimiento de la empresa. Inicialmente, el proceso de descarga de datos y generación de reportes relacionados con el estado de un usuario en un curso dado se ha venido realizando de manera manual debido a que el volumen de información no era tan significativo. Sin embargo, en los últimos años Griky ha tenido un crecimiento exponencial, por lo que se está dificultando ejecutar los procesos de descarga de la forma en que se han venido realizando; generando la necesidad de incrementar los recursos humanos para efectuar dicho proceso debido al creciente aumento en el volumen de datos; incurriendo en gastos, pérdida de tiempo e información y crecimiento lento, llevando a la compañía a tener procesos no escalables que no generan valor.

La creación de una herramienta que realiza procedimientos repetitivos de descarga de datos de forma automática minimiza errores humanos y utilización de recursos. La automatización de procesos repetitivos mediante *web scraping* permite que la disponibilidad de datos sea acorde al creciente volumen generado. Los datos obtenidos son el insumo principal para la generación de visuales y analíticas que permiten comprender el comportamiento de los usuarios, ayudando a incrementar la competitividad de la empresa y la efectiva toma de decisiones para así mantenerse y crecer en el mercado. Por parte de Griky, el señor Michael Zapata quien se desempeña como

coordinador de Inteligencia Empresarial brindó apoyo en temas relacionados con el proyecto, ya que la herramienta es de utilidad para esta área.

1.2. FORMULACIÓN DEL PROBLEMA

Pregunta de investigación:

- ¿Cómo desarrollar una herramienta personalizada que realice procedimientos repetitivos de descarga de datos de forma automática logrando minimizar errores, reducir gastos y utilización de recursos humanos; con la finalidad de disponer de datos para realizar analíticas y visuales que permitan tomar decisiones en pro del crecimiento la compañía?

Preguntas de sistematización:

- ¿Qué proyectos están relacionados con la automatización de procesos y la generación de analíticas y visuales a partir de datos sin procesar?
- ¿Qué herramientas se deben utilizar para la automatización, visualización y análisis de datos?
- ¿Cómo desarrollar un robot que realice tareas repetitivas de descarga de datos de forma automática?
- ¿Cómo realizar pruebas de funcionamiento del robot con data real y corregir posibles errores?
- ¿Cómo crear visualizaciones y analíticas que permitan comprender el comportamiento de los datos generados?

2. OBJETIVOS DEL PROYECTO

2.1. OBJETIVO GENERAL

Desarrollar una herramienta personalizada que realice los procedimientos repetitivos de descarga de datos de forma automática minimizando errores, reduciendo gastos y utilización de recursos humanos, con la finalidad de disponer de datos para realizar analíticas y predicciones que permitan tomar decisiones en pro del crecimiento de la compañía.

2.2. OBJETIVOS ESPECÍFICOS

- Identificar y revisar proyectos similares relacionados con la automatización de procesos y la generación de analíticas y visuales a partir de datos sin procesar.
- Definir las herramientas a utilizar y su debida documentación: lenguaje de programación, librerías, web framework, software de análisis y visualización de datos que permitan dar solución al problema planteado.
- Desarrollar un robot que realice las tareas repetitivas de descarga de datos de forma automática.
- Realizar pruebas de funcionamiento del robot con data real obtenida en las diferentes plataformas de la empresa.
- Crear visualizaciones y analíticas que permitan comprender el comportamiento de los datos obtenidos.

2.3. RESULTADOS ESPERADOS

- Robot de automatización que replique las acciones de un usuario en el proceso de descarga de datos, convirtiendo el proceso manual en uno escalable.
- Generación de visuales y analíticas con los datos obtenidos por el robot.
- Disminución de gastos operativos que pueden ser utilizados en crecimiento e innovación de la compañía.
- Disminución de errores generados al momento de realizar el proceso de extracción de data.

3. MARCO TEÓRICO DE REFERENCIA Y ANTECEDENTES

La ciencia de datos se ha convertido en una herramienta fundamental para la toma de decisiones en diferentes ámbitos, incluyendo la educación. En este proyecto, se utilizan técnicas de automatización de procesos robóticos (RPA, por sus siglas en inglés), *web scraping*, análisis y visualización de datos para obtener información valiosa sobre el progreso de los usuarios en la plataforma de aprendizaje en línea de Griky.

Para el desarrollo del proyecto se requirió de conocimientos en diferentes áreas, a continuación, se mencionan algunas de ellas:

3.1. AUTOMATIZACIÓN ROBÓTICA DE PROCESOS- RPA

La automatización de procesos robóticos (RPA) es una tecnología emergente que permite automatizar tareas repetitivas y estructuradas en aplicaciones informáticas, como la descarga de archivos en la plataforma de Griky. Esto se logra mediante la creación de "robots" que simulan las acciones humanas en el uso de estas aplicaciones.

Existen diversos campos donde se puede utilizar RPA para automatizar tareas repetitivas y mejorar la eficiencia de los procesos. Algunos de los campos son:

- **Finanzas:** para automatizar tareas como la gestión de facturas, la reconciliación de cuentas y la gestión de presupuestos.
- **Recursos humanos:** para automatizar tareas como el reclutamiento, la contratación, la gestión de nóminas y la gestión de beneficios para los empleados.
- **Servicio al cliente:** para automatizar tareas como el registro de tickets, la gestión de correos electrónicos y la atención al cliente.
- **Logística:** para automatizar tareas como la gestión de inventarios, la planificación de rutas y la gestión de entregas.
- **Salud:** para automatizar tareas en el sector de la salud, como la gestión de citas médicas, la facturación y la gestión de reclamaciones de seguros.
- **Manufactura:** para automatizar tareas en la línea de producción, como la gestión de pedidos y el seguimiento de la producción.
- **Administración pública:** para automatizar tareas en la administración pública, como la gestión de documentos, la gestión de procesos y la atención al ciudadano.

Antes de la llegada de la tecnología RPA, estas tareas se realizaban manualmente, lo que requería mucho tiempo y esfuerzo. Con la introducción de la automatización de procesos robóticos, estas

tareas se han vuelto más eficientes y menos propensas a errores, lo que ha llevado a un aumento en la productividad y la eficiencia en muchos sectores [2] [3].

Diferencias entre RPA y otras tecnologías de automatización:

A continuación, se presenta en la tabla 1 las principales diferencias entre RPA, IA y Machine Learning en términos de su enfoque, aplicaciones y capacidades.

Tabla 1. Comparación de tecnologías de automatización: RPA, Inteligencia Artificial y Machine Learning

TECNOLOGÍA DE AUTOMATIZACIÓN	DESCRIPCIÓN
RPA	Automatiza tareas repetitivas basadas en reglas predefinidas, utilizando software para emular la interacción humana con los sistemas. Es ideal para procesos estructurados y de alto volumen.
INTELIGENCIA ARTIFICIAL	Permite a las máquinas realizar tareas que normalmente requerirían inteligencia humana, como la capacidad de aprender, razonar y resolver problemas. Es ideal para procesos no estructurados y de bajo volumen.
MACHINE LEARNING	Un subconjunto de la inteligencia artificial, que permite a las máquinas aprender de los datos y mejorar su rendimiento sin ser programadas explícitamente. Es ideal para procesos con grandes conjuntos de datos y en constante cambio.

Fuente: Elaboración propia con base en websites CFB Bots [4] y Google Cloud [5]

Cada tecnología tiene sus propias fortalezas y limitaciones, por lo que es importante considerar cuidadosamente el proceso que se desea automatizar para así mismo elegir la tecnología adecuada en consecuencia.

3.2. WEB SCRAPING

Web scraping: Es una técnica utilizada para extraer datos de páginas web de manera automatizada y estructurada. La automatización permite llevar a cabo tareas repetitivas sin necesidad de las personas; se relaciona con robots, los cuales se entienden como una máquina programable capaz de realizar varias funciones o tareas complejas, manipular objetos y realizar automáticamente operaciones, incluyendo diferentes tipos de movimientos, en respuesta a su entorno. Esta técnica

se ha vuelto fundamental en la era de la información, en donde la cantidad de datos disponibles en la web es inmensa y el procesamiento manual de dichos datos es prácticamente imposible [6]. En el caso de este proyecto, se utilizó para recopilar información de la plataforma de aprendizaje en línea de Griky, lo cual permitirá tener una visión general del progreso de los usuarios en diferentes áreas [7] [8] [9].

El Web Scraping es una práctica muy utilizada en diversas áreas, como en el sector empresarial para la obtención de información de la competencia, en el sector financiero para el análisis de tendencias en el mercado, en el sector de la salud para el monitoreo de enfermedades, y en el sector educativo para la obtención de información acerca de los estudiantes, entre otros.

3.2.1. Lenguajes y Librerías más utilizadas

Existen diversos lenguajes de programación, librerías y herramientas que se utilizan para la ejecución del *Web Scraping*. Algunos de los lenguajes más utilizados son Python, Java, Ruby y Perl. En cuanto a las librerías, algunas de las más utilizadas son *BeautifulSoup*, *Scrapy*, y *Requests* [10]. Por otro lado, herramientas como Selenium son muy populares en el *Web Scraping*, ya que permiten simular la interacción de un usuario en una página web y extraer datos de forma automatizada [11].

Python es uno de los lenguajes de programación más utilizados en el *Web Scraping* debido a su facilidad de uso, su velocidad y su gran cantidad de librerías disponibles. *BeautifulSoup* es una de las librerías más utilizadas en Python para el *Web Scraping* debido a su capacidad para analizar y extraer información de documentos HTML y XML. *Scrapy* es otra librería de Python que se utiliza para el *Web Scraping* de forma más estructurada y ordenada [10].

Selenium es una herramienta de Python que se utiliza para la automatización de pruebas en páginas web, pero también se utiliza para el *Web Scraping* debido a su capacidad para interactuar con páginas web dinámicas y complejas. *Requests* es una librería de Python que se utiliza para hacer solicitudes HTTP y obtener información de páginas web de forma sencilla [11].

3.2.2. Usos del Web Scraping

El *Web Scraping* se utiliza en diversos ámbitos y fines, algunos de los más comunes son:

- **Obtención de información de la competencia:** las empresas utilizan el *Web Scraping* para obtener información acerca de sus competidores, como precios, ofertas, promociones, entre otros.
- **Análisis de tendencias en el mercado:** el *Web Scraping* se utiliza en el sector financiero para el análisis de tendencias en el mercado de valores y la obtención de información acerca de las empresas cotizantes.
- **Monitoreo de enfermedades:** el *Web Scraping* se utiliza en la salud para el monitoreo de enfermedades, la obtención de información acerca de su propagación y la identificación de patrones.
- **Obtención de información acerca de los estudiantes:** en el sector educativo, el *Web Scraping* se utiliza para la obtención de información acerca de los estudiantes, como su rendimiento académico, asistencia, entre otros.

El *Web Scraping* se ha vuelto fundamental en la era de la información, permitiendo la obtención de grandes cantidades de datos de forma automatizada. La elección del lenguaje y la librería adecuados, así como la definición de los objetivos y los fines del *Web Scraping* son clave para su éxito. En el sector educativo, el *Web Scraping* puede ser utilizado para la obtención de información valiosa acerca de los estudiantes, lo que permitiría una mejora en la toma de decisiones y en la gestión de las instituciones educativas.

3.3. ANÁLISIS DE DATOS

El análisis de datos es una técnica que se utiliza para extraer información útil a partir de grandes conjuntos de datos. Esta técnica se aplica en diferentes campos como la medicina, la ingeniería, la economía, la estadística, la psicología, entre otros, para identificar patrones, tendencias y relaciones en los datos recopilados. El análisis de datos puede ser realizado manualmente o con el uso de herramientas de software especializadas. En este proyecto, se utilizó para examinar la información recopilada de la plataforma de Griky y obtener *insights* valiosos sobre el rendimiento de los usuarios en diferentes áreas de aprendizaje [12].

3.3.1. Usos del análisis de datos

A continuación, se describen algunos casos de uso del análisis de datos y una breve descripción de cómo puede ser útil en determinado campo:

- **Identificación de oportunidades de negocio:** analizar datos puede ayudar a las empresas a identificar nuevas oportunidades de mercado y a desarrollar estrategias para aprovecharlas. Por ejemplo, se pueden analizar los patrones de compra de los clientes para identificar productos o servicios que tienen una alta demanda y que aún no están siendo ofrecidos por la competencia.
- **Mejora del rendimiento de la empresa:** el análisis de datos también puede utilizarse para mejorar la eficiencia y la productividad de la empresa. Por ejemplo, se pueden analizar los procesos de producción para identificar cuellos de botella y mejorar el flujo de trabajo.
- **Toma de decisiones informadas:** puede proporcionar información valiosa para la toma de decisiones informadas en la empresa. Por ejemplo, se pueden analizar los datos financieros para identificar tendencias y tomar decisiones sobre la inversión de capital.
- **Identificación de riesgos:** el análisis de datos puede ayudar a las empresas a identificar posibles riesgos y amenazas. Por ejemplo, se pueden analizar los datos de seguridad para identificar patrones de actividad sospechosa y tomar medidas preventivas.
- **Personalización de la experiencia del cliente:** contribuye en las empresas a personalizar la experiencia del cliente. Por ejemplo, se pueden analizar los datos de compra y de navegación en la página web para ofrecer productos o servicios relevantes a cada cliente.

3.3.2. Técnicas de análisis de datos

Existen diversas técnicas de análisis de datos que pueden utilizarse en función de los objetivos y el tipo de datos que se estén analizando. A continuación, se describen algunas de las técnicas más comunes:

- **Análisis estadístico:** se utiliza para analizar los datos numéricos y cuantitativos. Se pueden utilizar diversas herramientas estadísticas para medir la tendencia central, la dispersión, la correlación y la probabilidad de los datos. El análisis estadístico puede ayudar a identificar patrones en los datos y a realizar predicciones sobre futuros resultados.
- **Análisis de regresión:** consiste en analizar la relación entre dos o más variables. La regresión puede ser lineal o no lineal y se utiliza para predecir el valor de una variable en función de otra variable.
- **Análisis de cluster:** esta técnica se utiliza para agrupar los datos en función de sus características y similitudes. El análisis de cluster puede ayudar a identificar patrones y segmentos dentro de los datos.
- **Análisis de texto:** en esta técnica se analiza datos no estructurados, como los comentarios de los clientes o los informes de los empleados. El análisis de texto puede ayudar a identificar temas y tendencias en los datos.

- **Análisis de series temporales:** consiste en analizar los datos que cambian con el tiempo. El análisis de series temporales puede ayudar a identificar patrones y tendencias en los datos y a realizar predicciones para el futuro.

3.3.3. Tipos de análisis de datos

Existen diferentes tipos de análisis de datos que se pueden aplicar según los objetivos que se quieran alcanzar. A continuación, se detallan los tipos de análisis más comunes:

- **Análisis descriptivo:** se utiliza para resumir y describir los datos de forma clara y concisa. En este análisis se utilizan técnicas estadísticas para calcular medidas como la media, la mediana, la moda, la varianza, etc. El análisis descriptivo es útil para resumir grandes cantidades de datos y hacerlos más comprensibles.
- **Análisis exploratorio:** este tipo de análisis se utiliza para explorar los datos y detectar patrones, relaciones y tendencias. El objetivo del análisis exploratorio es descubrir nuevas relaciones no previstas y llegar a una comprensión más profunda de los datos. Este análisis se realiza mediante la visualización de datos y técnicas estadísticas.
- **Análisis inferencial:** utilizado para hacer inferencias y generalizaciones a partir de los datos. El objetivo del análisis inferencial es llegar a conclusiones sobre una población a partir de una muestra de datos. Este análisis se realiza mediante la aplicación de pruebas estadísticas y técnicas de estimación.
- **Análisis predictivo:** predice el comportamiento futuro de los datos en función de patrones y tendencias pasadas. El análisis predictivo se utiliza en la toma de decisiones en los negocios, la ciencia y la ingeniería. Este análisis se realiza mediante la aplicación de técnicas de modelado estadístico y aprendizaje automático.
- **Análisis prescriptivo:** recomienda acciones a partir de los datos. El análisis prescriptivo se utiliza en la toma de decisiones estratégicas en los negocios y en la gestión de operaciones. Este análisis se realiza mediante la aplicación de técnicas de optimización y simulación.

3.3.4. Métodos para el análisis de datos

Existen diferentes métodos para el análisis de datos que se pueden aplicar según los objetivos que se quieran alcanzar. A continuación, se detallan los métodos más comunes:

- **Análisis univariado:** este método se utiliza para analizar una sola variable a la vez. El objetivo del análisis univariado es resumir y describir los datos de forma clara y concisa.

En este análisis se utilizan técnicas estadísticas para calcular medidas como la media, la mediana, la moda, la varianza, etc.

- **Análisis bivariado:** este método se utiliza para analizar dos variables simultáneamente. El objetivo del análisis bivariado es descubrir las relaciones entre dos variables y determinar si hay alguna correlación entre ellas. En este análisis se utilizan técnicas estadísticas como la correlación y la regresión.
- **Análisis multivariado:** este método se utiliza para analizar tres o más variables simultáneamente. El objetivo del análisis multivariado es descubrir las relaciones complejas entre varias variables y determinar si hay alguna correlación entre ellas. En este análisis se utilizan técnicas estadísticas como el análisis de componentes principales, el análisis factorial y el análisis de clúster.

3.3.5. Aprendizaje automático y análisis de datos

El aprendizaje automático es una técnica de inteligencia artificial que utiliza algoritmos para analizar grandes cantidades de datos y aprender patrones y relaciones entre ellos. Se utiliza en muchos campos, como la medicina, la banca, la publicidad y la seguridad informática.

En el análisis de datos, el aprendizaje automático se utiliza para predecir resultados futuros a partir de datos históricos. Por ejemplo, se puede utilizar el aprendizaje automático para predecir el precio de una acción en la bolsa o el número de ventas de un producto en función de los datos de ventas históricos. Las técnicas de aprendizaje automático más comunes son la regresión, la clasificación y el clustering [12].

El aprendizaje automático también se utiliza en el análisis de datos para identificar patrones y tendencias en los datos. Por ejemplo, se puede utilizar el aprendizaje automático para identificar patrones de fraude en las transacciones de una empresa o para identificar patrones de comportamiento en los clientes. Las técnicas de aprendizaje automático más comunes para la identificación de patrones son el análisis de componentes principales y el análisis de redes neuronales.

En resumen, el aprendizaje automático es una técnica de inteligencia artificial que se utiliza en el análisis de datos para predecir resultados futuros y para identificar patrones y tendencias en los datos. Es una herramienta valiosa para las empresas que buscan mejorar su rendimiento y tomar decisiones informadas a partir de sus datos [13].

3.4. VISUALIZACIÓN DE DATOS

La visualización de datos es una técnica que consiste en representar información de manera gráfica,

con el objetivo de facilitar su análisis y comprensión. La visualización de datos permite identificar patrones, tendencias y relaciones entre los datos de una manera más clara y efectiva que con el análisis de datos tradicional.

3.4.1. Ventajas de la visualización de datos

Algunas de las ventajas de la visualización de datos son:

- **Facilita la identificación de patrones y tendencias:** la visualización de datos permite identificar patrones y tendencias que pueden no ser evidentes en los datos sin procesar.
- **Permite comparar datos de manera efectiva:** la visualización de datos permite comparar datos de manera efectiva, ya que se pueden representar varios conjuntos de datos en un solo gráfico.
- **Ayuda a identificar relaciones entre datos:** la visualización de datos permite identificar relaciones entre datos que pueden no ser evidentes en una tabla de datos.
- **Facilita la toma de decisiones:** la visualización de datos permite tomar decisiones informadas de manera más efectiva, ya que se puede entender y analizar la información de manera más clara.

3.4.2. Gráficos comunes

Existen diversos tipos de gráficos que se utilizan en la visualización de datos, cada uno con su propia función y características. A continuación, se presentan algunos de los gráficos más comunes en la visualización de datos:

- **Gráfico de barras:** es un tipo de gráfico que se utiliza para comparar valores entre diferentes categorías. En un gráfico de barras, las barras verticales o horizontales representan los valores de cada categoría.
- **Gráfico de líneas:** se utiliza para mostrar tendencias a lo largo del tiempo. En un gráfico de líneas, las líneas conectan los valores de cada punto en el tiempo.
- **Gráfico circular:** indica la proporción de diferentes categorías. En un gráfico circular, las proporciones se representan como porcentajes de un círculo completo.
- **Gráfico de dispersión:** el gráfico de dispersión es un tipo de gráfico que se utiliza para mostrar la relación entre dos variables. En un gráfico de dispersión, cada punto representa un par de valores de las dos variables.
- **Mapa de calor:** muestra la distribución de valores en un conjunto de datos. En un mapa de calor, los valores se muestran como colores en una matriz.

3.4.3. Aprendizaje automático y análisis de datos

Existen diversas herramientas para la visualización de datos, cada una con sus propias características y funcionalidades. A continuación, se presentan algunas de las herramientas más utilizadas para la visualización de datos:

- **Tableau:** Es una herramienta de visualización de datos que permite crear gráficos, mapas, tablas y paneles interactivos. Es una herramienta muy intuitiva y fácil de usar, que permite conectar con diversos tipos de fuentes de datos.
- **Power BI:** herramienta creada por Microsoft que permite crear gráficos, tablas y paneles interactivos. Es una herramienta muy potente, que permite conectar con diversas fuentes de datos y trabajar con grandes conjuntos de datos.
- **Excel:** Es una herramienta de hoja de cálculo que también puede utilizarse para la visualización de datos. Excel permite crear gráficos y tablas simples, y es una herramienta muy ampliamente utilizada.
- **D3.js:** D3.js es una biblioteca de JavaScript para la visualización de datos en la web. Es una herramienta muy potente y flexible, que permite crear gráficos y visualizaciones personalizadas.
- **Python:** Python es un lenguaje de programación que cuenta con diversas bibliotecas para la visualización de datos, como Matplotlib y Seaborn. Estas bibliotecas permiten crear gráficos y visualizaciones personalizadas.

3.5. TRABAJOS RELACIONADOS

Se realizó una investigación de trabajos relacionados al desarrollo de herramientas de automatización de procesos que permitan realizar extracción de datos, así como generación de visuales y analíticas en el sector educativo.

Nicole Vicente Stenhouse es creador de HabScraper, una herramienta automatizada para la extracción de datos con web scraping que nace de la necesidad de obtener datos de páginas web de una forma automática realizando, por ejemplo, consultas en un buscador, accediendo a una determinada página web y realizando cierto número de pasos repetitivos para obtener la data necesaria. HabScraper replica las acciones de navegación que puede realizar un usuario y extrae de forma automática datos de contacto de manera masiva en determinadas páginas web; el proyecto describe el diseño, funcionamiento e implementación de la herramienta. Este proyecto tiene gran similitud en cuanto a lo que se desarrolló en el presente documento, ya que, se requiere replicar tareas repetitivas que hoy en día realiza un recurso humano para descargar información, además, el proyecto de HabScraper brinda una documentación interesante utilizando una herramienta de

ciencia de datos como es *web scraping* para automatización de procesos, por lo que se usó como guía para el diseño y desarrollo del robot [14].

Juan Sebastián Casas Casallas es autor de un interesante proyecto que consiste en el diseño de tableros de control de indicadores para el análisis de la información acerca de la situación de los graduados de la Institución Universitaria Politécnico Grancolombiano mediante el uso de Microsoft PowerBI. Este proyecto brinda información importante contemplando múltiples analíticas y visuales como: temáticas de interés del usuario, nivel de satisfacción, estado del usuario, predicciones en relación con futuras inscripciones, entre otros. Estas analíticas se aplicaron en el presente proyecto y fueron de gran utilidad al momento de realizar el análisis y visualización de datos utilizando herramientas como PowerQuery y PowerBI [15].

El proyecto "Web Scraping and Data Visualization for Online Education Platforms" [4] utiliza técnicas de *web scraping* para recopilar datos de diferentes plataformas de educación en línea y crear visualizaciones de datos para analizar el rendimiento de los estudiantes. El proyecto incluye la creación de un robot de *web scraping* para recopilar datos y la utilización de herramientas de visualización de datos como D3.js para crear gráficos y visualizaciones interactivas. Este proyecto proporciona información útil sobre cómo utilizar técnicas de *web scraping* y visualización de datos en el ámbito educativo [16].

El trabajo de investigación titulado "Big data analytics in education: A review of literature and a case study" [17] analiza el uso del análisis de big data en la educación y su impacto en la toma de decisiones educativas. En particular, el estudio examina el potencial de los datos generados en el entorno educativo para mejorar la eficacia de los programas de aprendizaje y el rendimiento estudiantil. Este trabajo es útil para comprender la importancia de los datos generados en el contexto educativo y cómo se utilizan para mejorar el aprendizaje.

El sitio web Scrapfly.io realizó un proyecto-tutorial en donde detalla cómo utilizar Selenium y Python para realizar web scraping de datos dinámicos generados por JavaScript o AJAX en páginas web. El proyecto explica cómo navegar por las páginas, esperar a que se carguen los elementos y hacer clic o desplazarse por la pantalla para obtener los datos deseados. El código Python se ejecuta en un navegador sin cabeza (*headless browser*), que es un navegador que no tiene interfaz gráfica y que se puede controlar mediante comandos. Esta técnica se aplica al proyecto para extraer datos de la plataforma de Griky, que utiliza elementos dinámicos para mostrar el progreso de los usuarios [18].

A continuación, en la Tabla 2 se muestran varios proyectos relacionados con el análisis de datos y la automatización de procesos utilizando diversas técnicas y herramientas. Cada proyecto se presenta con una breve descripción y una lista de las técnicas y herramientas utilizadas en el

proyecto. Esta tabla ofrece una visión general de diferentes enfoques para el análisis de datos y la automatización de procesos.

Tabla 2. Trabajos relacionados con herramientas de automatización.

Proyecto	Descripción	Técnicas y herramientas utilizadas
HabScraprer	Herramienta automatizada para la extracción de datos con <i>web scraping</i> que replica acciones de navegación y extrae de forma automática datos de contacto.	Web scraping, Python
Diseño de tableros de control de indicadores	Diseño de tableros de control de indicadores para el análisis de la información acerca de la situación de los graduados.	Microsoft PowerBI, Power Query, R.
Web Scraping and Data Visualization for Online Education	Utiliza técnicas de <i>web scraping</i> para recopilar datos de diferentes plataformas de educación en línea y crear visualizaciones de datos.	Web scraping, D3.js,
Big data analytics in education: A review of literature	Analiza el uso del análisis de big data en la educación y su impacto en la toma de decisiones educativas.	Big data, Tableau. UiPath.
Web Scraping with Selenium and Python Tutorial + Example Project	Proyecto-tutorial que detalla cómo utilizar Selenium y Python para realizar <i>web scraping</i> de datos dinámicos generados por JavaScript o AJAX.	Web scraping, Selenium, Python, Headless browser.

Fuente: Elaboración propia con base en trabajos relacionados.

En resumen, los trabajos relacionados con el desarrollo de herramientas de automatización en el sector educativo utilizan técnicas como *Web Scraping*, Automatización de procesos, Big data y herramientas como PowerBI, Power Query, Tableau, UiPath, Selenium, R y Python. Estas herramientas pueden utilizarse para recopilar datos de diferentes fuentes, realizar análisis y visualizaciones de datos para mejorar la eficacia de los programas de aprendizaje y el rendimiento estudiantil. En la siguiente sección se muestra una comparación de las herramientas que se utilizaron en el proyecto y el por qué se eligieron sobre otras.

4. SELECCIÓN DE HERRAMIENTAS

En este capítulo, se llevará a cabo la selección de las herramientas que serán utilizadas en el proyecto de automatización de procesos en la empresa Griky. Antes de iniciar con la selección de herramientas se presenta un breve contexto de la empresa.

Griky es una plataforma educativa en línea que tiene como misión principal facilitar el acceso a la educación de calidad a través de soluciones tecnológicas innovadoras. Su plataforma ofrece una amplia gama de cursos en línea que se dividen en diferentes catálogos, provenientes tanto de creación propia de la empresa como de otros proveedores de contenido, además, recursos de aprendizaje adaptativos y herramientas de evaluación para estudiantes de diferentes niveles educativos. La empresa se ha asociado con instituciones educativas y expertos en diversos campos para proporcionar contenido educativo actualizado y de alta calidad.

El proyecto en cuestión tiene como objetivo principal automatizar ciertos procesos dentro de Griky con el fin de optimizar su eficiencia y mejorar la experiencia de los usuarios. Para lograr esto, se utilizaron técnicas de automatización de software y RPA (Robotic Process Automation) con el objetivo de lograr la descarga automatizada de datos relacionados al progreso de los usuarios en la plataforma de la empresa, estos datos son esenciales para conocer el comportamiento de los usuarios de diferentes clientes de la compañía y permitir la toma de decisiones estratégicas basadas en datos. Al realizar tareas repetitivas y basadas en reglas de manera automática, se libera tiempo y recursos para que el equipo de Griky pueda centrarse en actividades que generen valor en la compañía.

En resumen, el objetivo principal de este capítulo es seleccionar las herramientas adecuadas para la construcción de la herramienta RPA para Griky que finalmente permitió obtener datos necesarios para realizar el análisis del progreso de los usuarios en la plataforma de la empresa. Estas herramientas se utilizarán para descargar, analizar, procesar y visualizar datos, optimizando así la eficiencia de los procesos internos y mejorando la experiencia general de los usuarios.

En el contexto del desarrollo de soluciones tecnológicas, es fundamental contar con las herramientas adecuadas que permitan enfrentar los retos de manera efectiva. En este sentido, fue importante definir con claridad las herramientas a utilizar y documentar su uso para garantizar su correcta implementación y eficiencia en la solución del problema planteado.

En el caso particular del proyecto, se definieron las herramientas de programación, librerías, web framework y software de análisis y visualización de datos que permitieron enfrentar el desafío que se presentó. En este sentido, es importante destacar que estas herramientas fueron seleccionadas cuidadosamente en función de su capacidad para resolver las necesidades específicas del proyecto, así como por su eficiencia y escalabilidad.

A continuación, se describen detalladamente las herramientas seleccionadas y se comparan con herramientas similares. Se proporciona la información necesaria del porque su uso fue efectivo en el proyecto. Esto permitió contar con una base sólida para el desarrollo de la solución y garantizar el éxito.

4.1. HERRAMIENTAS PARA DESARROLLO DE RPA

La Robotic Process Automation (RPA) se ha convertido en una herramienta valiosa para la automatización de tareas y procesos empresariales. Para desarrollar soluciones de RPA, se requiere del uso de herramientas específicas que permitan crear, implementar y mantener robots de manera efectiva.

En esta sección, se presentan las principales herramientas para el desarrollo de RPA, considerando aquellas que se han consolidado en el mercado y que ofrecen un amplio conjunto de funcionalidades y características para el desarrollo de soluciones de automatización empresarial.

4.1.1. Lenguajes de programación

Python y R son dos lenguajes de programación ampliamente utilizados en el campo de la ciencia de datos y el análisis estadístico. Si bien ambos lenguajes son adecuados para realizar tareas de análisis de datos y modelado estadístico, difieren en varios aspectos importantes, incluyendo su sintaxis, facilidad de uso, velocidad y escalabilidad.

En la Tabla 3 se realizó una revisión de la literatura especializada y se recolectaron datos de diversas fuentes para llevar a cabo la comparación entre el lenguaje de Programación Python y R.

Tabla 3. Lenguaje de Programación Python Vs R

Características	Python	R
Lenguaje	Interpretado, de alto nivel, multipropósito	Interpretado, de alto nivel, orientado a estadísticas
Comunidad	Gran comunidad de desarrolladores	Gran comunidad de estadísticos y analistas
Sintaxis	Sintaxis clara y legible	Sintaxis complicada
Velocidad de ejecución	Rápido en tareas de computación numérica y científica	Menos rápido en tareas de computación numérica y científica
Gráficos	Biblioteca Matplotlib de Python para visualización de datos	Biblioteca ggplot2 de R para visualización de datos
Aprendizaje	Fácil de aprender y usar	Curva de aprendizaje empinada
Herramientas de análisis	Numerosas bibliotecas para	Numerosas bibliotecas para

	análisis de datos, aprendizaje automático, procesamiento de imágenes, entre otros	análisis de datos, estadísticas y gráficos
--	---	--

Fuente: IBM Cloud Team, IBM Cloud, 2021 [19]

Basado en esta tabla comparativa, se concluyó que ambas herramientas son excelentes para el análisis de datos y estadísticas, y ambas tienen una gran comunidad de usuarios. Sin embargo, Python es una herramienta más versátil, con una sintaxis clara y legible, lo que la hace fácil de aprender y usar. También tiene una amplia variedad de bibliotecas para diferentes tareas, lo que la hace una opción sólida para el análisis de datos, aprendizaje automático, procesamiento de imágenes y otras tareas. En cuanto a la velocidad de ejecución, Python es más rápido en tareas de computación numérica y científica en comparación con R. Además, su biblioteca Matplotlib es una herramienta potente y fácil de usar para la visualización de datos.

Estas fueron las razones por las cuales se seleccionó Python como herramienta para el proyecto:

- Es una herramienta más versátil.
- Fácil de aprender y usar.
- Gran variedad de bibliotecas para diferentes tareas y mayor velocidad de ejecución en tareas de computación numérica y científica.
- La biblioteca Matplotlib es excelente para la visualización de datos, lo que hace que Python sea una herramienta ideal para el análisis de datos y estadísticas en este proyecto.

4.1.2. Herramientas de automatización de pruebas

Selenium y Puppeteer son dos herramientas de automatización de pruebas muy populares en el desarrollo de software moderno. Aunque ambas herramientas tienen el mismo objetivo de automatizar la interacción de un navegador web con una página web, difieren en varios aspectos.

En la Tabla 4 se analizó las principales diferencias entre Selenium y Puppeteer, destacando sus fortalezas y debilidades en cuanto a la facilidad de uso, la documentación, la capacidad de depuración, la velocidad y la escalabilidad.

Tabla 4. Selenium Vs Puppeteer

Características	Selenium	Puppeteer
Lenguaje de Programación	Soporta múltiples lenguajes, incluyendo Python, Java, C#, Ruby, JavaScript, entre otros.	Es una biblioteca de Node.js y solo admite JavaScript.
Interacción con	Selenium es compatible con múltiples	Puppeteer solo es compatible con

navegadores	navegadores web, incluyendo Chrome, Firefox, Safari, IE, entre otros.	el motor de renderizado Chromium.
Acciones en el navegador	Selenium ofrece una amplia variedad de acciones y operaciones que se pueden realizar en un navegador, incluyendo hacer clic, escribir, desplazarse, capturar pantallas, entre otras.	Puppeteer también ofrece muchas funciones similares a Selenium, pero su API es más intuitiva y fácil de usar.
Documentación	La documentación de Selenium es amplia y detallada. Además, hay una gran cantidad de recursos en línea disponibles para ayudar a los usuarios.	Puppeteer también tiene una documentación completa y útil, pero como es una biblioteca más nueva, es posible que no haya tantos recursos en línea disponibles.
Comunidad de Usuarios	Selenium es una herramienta de automatización de pruebas muy popular y ampliamente utilizada, lo que significa que hay una gran comunidad de usuarios en línea que pueden brindar soporte y compartir conocimientos.	Puppeteer es una biblioteca relativamente nueva, por lo que su comunidad de usuarios aún está en desarrollo. Aunque su popularidad está creciendo rápidamente.

Fuente: Elaboración propia con base en websites Kulyk, 2022 [20] y Vasilis, 2023 [21]

Después de comparar ambas herramientas, Selenium es una mejor opción que Puppeteer para la automatización de pruebas web. Aunque ambas herramientas ofrecen funciones similares, Selenium es compatible con una variedad más amplia de navegadores y lenguajes de programación, lo que lo hace más flexible y accesible para los usuarios. Además, tiene una comunidad de usuarios más grande y establecida que puede brindar soporte y compartir conocimientos. Por último, dado que ya se ha elegido Python como lenguaje de programación para el proyecto, es importante tener en cuenta que Selenium es compatible con Python, lo que lo hace más adecuado para la integración con el resto del proyecto. Por lo mismo, se eligió Selenium usando Python para la automatización de pruebas web.

4.1.3. Navegadores web

Google Chrome y Mozilla Firefox son dos de los navegadores web más utilizados en todo el mundo. Ambos navegadores compiten por ofrecer una experiencia de navegación web rápida, segura y confiable. Si bien existen muchas similitudes entre ambos navegadores, también existen diferencias significativas que pueden afectar la elección de los usuarios.

En la Tabla 5 se presentan algunas de las características más relevantes de Google Chrome y Mozilla Firefox, como la velocidad de carga de páginas, la compatibilidad con extensiones y complementos, la privacidad y seguridad, entre otros aspectos importantes. Esta comparación permitió seleccionar el navegador que mejor se adaptó a las necesidades y preferencias para el proyecto.

Tabla 5. Google Chrome Vs Mozilla Firefox

Característica	Google Chrome	Mozilla Firefox
Compatibilidad con Python	Excelente	Buena
Compatibilidad con Selenium	Excelente	Buena
Velocidad de carga de páginas	Excelente	Buena
Interfaz de usuario	Moderna y sencilla	Clásica y personalizable
Seguridad	Buena	Excelente
Estabilidad	Excelente	Buena
Popularidad	Muy popular	Popular

Fuente: Elaboración propia con base en websites MERKLE, 2020 [22] y LAMBDATEST [23]

Ambos navegadores son compatibles con Python y Selenium, pero Google Chrome se destaca por su excelente compatibilidad y rendimiento en la automatización de pruebas de software y RPA. Además, Google Chrome es conocido por su velocidad de carga de páginas, lo que lo hace ideal para RPA que implica una gran cantidad de navegación web. En cuanto a seguridad y estabilidad, ambos navegadores son sólidos, aunque Firefox se destaca por su alto nivel de seguridad.

En resumen, aunque Mozilla Firefox es un excelente navegador, para la automatización de pruebas de software y RPA, Google Chrome es la mejor opción debido a su excelente compatibilidad con Python y Selenium, su velocidad de carga de páginas y su popularidad en la comunidad de desarrollo de software.

4.1.4. Almacenamiento en nube

El almacenamiento en la nube se ha convertido en una herramienta cada vez más popular para almacenar y compartir archivos en línea. En lugar de tener que depender del almacenamiento físico en discos duros externos o dispositivos de almacenamiento USB, la nube ofrece una solución más flexible y accesible para almacenar y acceder a los datos.

La Tabla 6 compara dos de los servicios de almacenamiento en la nube más populares, OneDrive de Microsoft y Google Drive de Google. Ambos servicios ofrecen a los usuarios la posibilidad de almacenar y compartir archivos en línea, acceder a ellos desde cualquier dispositivo y colaborar en documentos utilizando aplicaciones de productividad.

Tabla 6. One Drive Vs Google Drive

Característica	OneDrive	Google Drive
Almacenamiento	5GB	15GB
Integración con Microsoft Office	Integrado completamente con Microsoft Office, lo que facilita la edición y colaboración en documentos de Office	Puede trabajar con archivos de Office, pero no tiene la misma integración y funcionalidad que OneDrive
Acceso y sincronización de archivos	Excelente integración con sistemas operativos de Microsoft, permitiendo la sincronización de archivos de manera más fluida	Integración en otros sistemas operativos como Mac, pero no tiene la misma fluidez y funcionalidad que OneDrive
Compartir archivos	Permite compartir archivos con otras personas y establecer permisos de acceso y edición	Ofrece la misma funcionalidad que OneDrive, pero no tiene la misma integración con Microsoft Office
Seguridad	OneDrive cuenta con medidas de seguridad avanzadas y cuenta con una política de privacidad sólida, ofreciendo un alto nivel de protección de datos	Google Drive también cuenta con medidas de seguridad avanzadas y una buena política de privacidad, pero algunas personas pueden tener preocupaciones sobre la privacidad de sus datos con Google

Fuentes: Elaboración propia con base en Websites Microsoft [24] y Google One [25]

Teniendo en cuenta que la empresa Griky utiliza la suit de Microsoft, es recomendable utilizar OneDrive como herramienta de almacenamiento en la nube debido a su excelente integración con los programas de la suite. Además, ofrece una seguridad y privacidad muy buenas, lo que la hace una opción confiable para el almacenamiento de información importante de la empresa.

4.1.5. Entornos de desarrollo integrado

Visual Studio Code y PyCharm son dos de los entornos de desarrollo integrado (IDE) más populares utilizados por los programadores para desarrollar aplicaciones en diferentes lenguajes de programación, como Python, JavaScript, TypeScript, entre otros.

En la Tabla 7 se presentan algunas de las características y diferencias entre Visual Studio Code y PyCharm.

Tabla 7. Visual Studio Code Vs PyCharm

Características	Visual Studio Code	PyCharm
Precio	Gratis y de código abierto	Freemium, la versión básica es gratuita, pero la versión completa es de pago

Interfaz de usuario	Interfaz de usuario simple e intuitiva	Interfaz de usuario compleja y abrumadora al principio
Compatibilidad de lenguaje	Compatible con múltiples lenguajes, incluyendo Python	Enfocado principalmente en Python, pero también admite otros lenguajes
Integración con herramientas externas	Admite una amplia variedad de herramientas externas y extensiones	Proporciona muchas herramientas integradas para el desarrollo, pero no admite muchas herramientas externas
Rendimiento	Menos exigente en recursos de sistema y más rápido	Más exigente en recursos de sistema y más lento
Comunidad y soporte	Gran comunidad y amplia documentación en línea	Buena comunidad y soporte técnico, pero menos documentación disponible en línea

Fuente: Elaboración propia con base en websites Visual Studio Code [26] y PyCharm [27]

En conclusión, aunque PyCharm es una herramienta poderosa y popular para el desarrollo en Python, Visual Studio Code es una excelente opción para escribir código en Python debido a su interfaz de usuario simple e intuitiva, su compatibilidad con múltiples lenguajes y su gran cantidad de herramientas y extensiones externas. Además, su rendimiento es mejor que el de PyCharm y tiene una gran comunidad de desarrolladores con una amplia documentación en línea disponible.

4.2. HERRAMIENTAS PARA ANÁLISIS DE DATOS Y VISUALIZACIÓN

Las herramientas para análisis de datos y visualización son un conjunto de programas y aplicaciones diseñados para ayudar a los interesados a analizar y visualizar datos de manera efectiva. Estas herramientas son esenciales para el procesamiento y análisis de grandes cantidades de datos, lo que ayuda a las empresas y organizaciones a tomar decisiones y a mejorar sus operaciones.

Power BI y Tableau son dos de las herramientas más populares para el análisis y la visualización de datos utilizadas por empresas y organizaciones de todo el mundo. Ambas herramientas ofrecen una amplia gama de características y funcionalidades para ayudar a los usuarios a descubrir ideas y patrones ocultos en grandes cantidades de datos, y para presentar estos datos de manera efectiva a través de gráficos, tablas y otros elementos visuales.

En la Tabla 8 se presentan algunas de las características más destacadas de cada herramienta.

Tabla 8. Power BI Vs Tableau

Característica	Power BI	Tableau
Precio	Ofrece una versión gratuita y planes de pago, con precios más bajos que	Precios más altos que Power BI y no cuenta con una versión gratuita.

	Tableau	
Integración	Integra fácilmente con otras herramientas de Microsoft, como Excel y Azure	Ofrece integraciones con diversas herramientas de terceros, pero no cuenta con una integración tan directa con herramientas de Microsoft
Facilidad de uso	Interface intuitiva y fácil de usar, especialmente para usuarios que ya están familiarizados con la suite de Microsoft	Interface también intuitiva, pero requiere un poco más de tiempo para aprender a usar todas las características y funciones
Visualizaciones	Gran cantidad de visualizaciones disponibles, incluyendo visualizaciones personalizadas. Las actualizaciones y mejoras son frecuentes	Cuenta con una gran cantidad de visualizaciones disponibles, aunque no tanto como Power BI, y la frecuencia de actualizaciones no es tan alta
Capacidad de procesamiento de datos	Capacidad para procesar grandes cantidades de datos y manejar conjuntos de datos complejos	Similar a Power BI en términos de capacidad de procesamiento, aunque algunos usuarios han reportado que Tableau se vuelve más lento cuando se trabaja con grandes conjuntos de datos
Comunidad y soporte	Cuenta con una gran comunidad de usuarios y un amplio soporte de Microsoft	También cuenta con una comunidad activa de usuarios, pero el soporte no es tan amplio como el de Power BI y no cuenta con el respaldo de una empresa tan grande como Microsoft

Fuente: Elaboración propia con base en websites Biswal,2023 [28] y Kaur, [29]

Con base en la tabla comparativa anterior, se tomó la decisión de usar Power BI como herramienta de visualización de datos para este proyecto. Aunque ambas herramientas son muy buenas, Power BI tiene la ventaja de integrarse mejor con otras herramientas de Microsoft, lo que es beneficioso para una empresa que ya utiliza esta suite. Además, cuenta con una versión gratuita y planes de pago más asequibles que Tableau, lo que puede ser una ventaja si se tiene en cuenta el presupuesto. También tiene una comunidad de usuarios más grande y un soporte más amplio de Microsoft. Por último, tiene una gran cantidad de visualizaciones disponibles, lo que es útil para generar los informes requeridos a partir de los datos obtenidos por el RPA y almacenados en One Drive.

5. DESARROLLO DEL RPA

La automatización robótica de procesos (RPA, por sus siglas en inglés) se ha convertido en una herramienta esencial para mejorar la eficiencia y la productividad de las empresas. A través de la programación de robots software, es posible automatizar tareas repetitivas y manuales, liberando así a los empleados de trabajos monótonos y permitiéndoles concentrarse en actividades de mayor valor agregado. En este contexto, el desarrollo del RPA se ha convertido en un tema crucial para aquellas organizaciones que desean mantenerse competitivas en el mercado. En esta sección, se explora en detalle el proceso de desarrollo del RPA, incluyendo los pasos necesarios para la implementación exitosa de esta tecnología.

Se desarrolló un robot de automatización haciendo uso de diferentes herramientas, el cual tiene como funcionalidad seguir una serie de pasos repetitivos con el objetivo de generar un reporte que resume el progreso de usuarios en un curso dado.

5.1. DEFINICIÓN DE PARÁMETROS Y VARIABLES

La automatización robótica de procesos (RPA) involucra la programación de robots software para llevar a cabo tareas específicas. Para que estos robots puedan realizar las tareas correctamente, es necesario definir los parámetros y variables adecuados. Los parámetros son las entradas que se proporcionan al robot, mientras que las variables son los valores que el robot utiliza para realizar los cálculos necesarios para completar la tarea. En esta sección, se muestra en detalle la definición de parámetros y variables en el contexto del RPA. Se definen estos elementos y cómo se asegura su precisión para garantizar la eficacia y la eficiencia de la automatización robótica de procesos.

Antes de iniciar la construcción del robot se listaron algunas variables que contienen información de acceso y de botones necesarios para navegar en la plataforma, a continuación, en la tabla 9 se muestran algunas de las variables más relevantes:

Tabla 9. Variables Definidas

Nombre	Descripción
user	Nombre de usuario para ingresar a la plataforma
psw	Contraseña de acceso a la plataforma
aplicar	Botón de acceso
nube	Botón en donde se alojan los archivos a descargar

driver	Variable en la que se aloja el web driver para Google Chrome
--------	--

De la misma forma fue necesario definir unos parámetros de entrada que se muestran en la tabla 10 para el funcionamiento del robot tales como:

Tabla 10. Parametros Definidos

Nombre	Descripción
url	Dirección de la página a extraer información
incremento	Número de cursos a descargar por ciclo
lim_inf	Tiempo mínimo de espera en descargar un archivo en segundos
lim_sup	Tiempo máximo de espera en descargar un archivo en segundos

5.2. CONSTRUCCIÓN DEL ROBOT

Una vez se definieron las variables y parámetros necesarios para la tarea que se va a automatizar, llega el momento de construir el robot que realiza las mismas. La construcción del robot es un proceso crucial en la automatización robótica de procesos (RPA), ya que determina la capacidad del robot para ejecutar la tarea de manera eficiente y efectiva. En esta sección, se muestran los pasos realizados para construir el robot RPA, incluyendo la selección de herramientas y tecnologías, la programación del robot, y la integración de este en el sistema de la empresa. Además, se habla de las consideraciones importantes que deben tenerse en cuenta durante el proceso de construcción para garantizar que el robot cumpla con los requisitos de la tarea y se integre de manera efectiva en el sistema de la empresa.

5.2.1. Importación de librerías

Se importaron las librerías necesarias para el procesamiento de datos y para llevar a cabo el procedimiento de automatización.

Figura 1. Código Importación de librerías

```
from lib2to3.pgen2 import driver
from tabnanny import check
from typing import final
from unicodedata import name
from urllib import response
import pandas
import time
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as ec
import requests
from selenium.webdriver.support.events import EventFiringWebDriver, AbstractEventListener
from selenium.common.exceptions import TimeoutException
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.common.by import By
from selenium.webdriver.common.action_chains import ActionChains
```

Cada una de estas librerías tiene un propósito específico y son utilizadas para mejorar el proceso de desarrollo de software, desde el análisis sintáctico y la corrección de errores en el código, hasta el análisis y la manipulación de datos, el manejo de solicitudes HTTP, la automatización de pruebas web y mucho más. El uso de estas librerías ahorra tiempo y esfuerzo en la creación de nuevas funcionalidades y mejora la eficiencia en el proceso de desarrollo de software. Es importante conocer estas librerías y saber cómo utilizarlas para mejorar la calidad del código y la eficacia en la programación.

Todas las librerías mencionadas anteriormente son relevantes y útiles en diferentes situaciones y contextos. Sin embargo, algunas de las librerías más populares y ampliamente utilizadas son:

- Pandas: es una de las librerías más populares en Python para el análisis y la manipulación de datos en formato de tabla. Es especialmente útil para la limpieza y transformación de datos, así como para el análisis y la visualización de datos.
- Requests: es una de las librerías más utilizadas para realizar solicitudes HTTP y trabajar con APIs y servicios web. Es fácil de usar y proporciona una gran cantidad de opciones y funciones para trabajar con datos de la web.
- Selenium: es una de las herramientas más utilizadas para la automatización de pruebas web y la interacción con navegadores web. Es muy útil para simular la interacción de un usuario con una aplicación web y realizar pruebas automatizadas.

Estas librerías son solo algunas de las más populares y utilizadas, cada una de las librerías mencionadas tiene su propio propósito y utilidad en diferentes situaciones y contextos.

5.2.2. Definición de función principal

Se define la función principal que tiene como entrada los parámetros antes mencionados

Figura 2. Definición de función principal

```
def robot(url, incremento, lim_inf, lim_sup):  
  
    #Variables usuario y contraseña:  
    credenciales_excel = r"D:\PROYECTOS\ROBOT GRIKY\credenciales.xlsx"  
    df = pandas.read_excel(credenciales_excel)  
    user = df["username"][0]  
    psw = df["password"][0]  
  
    #Aplicar filtros y descargar  
    aplicar = '//*[@id="submit-button"]'  
    nube = '//*[@id="main"]/div/nav/ul/li/a/i'  
  
    #Web driver para google Chrome  
    driver = webdriver.Chrome()
```

La función *robot()* retorna n cantidad de archivos descargados en formato .xls de una página (url) dada.

5.2.3. Acceso a la plataforma

En esta sentencia de código se hace el acceso a la plataforma con las variables de acceso ya definidas. Además, se coloca un tiempo de espera de dos segundos entre cada paso para que la pagina tenga tiempo de reaccionar.

Figura 3. Acceso a la plataforma

```
# Login:  
driver.find_element(By.CSS_SELECTOR, "#login-email").send_keys(user)  
# time.sleep(2)  
driver.find_element(By.CSS_SELECTOR, "#login-password").send_keys(psw)  
time.sleep(2)  
driver.find_element(By.CSS_SELECTOR, "#login > button").click()  
time.sleep(2)
```

5.2.4. Conteo de cursos

Una vez el robot accede a la plataforma, debe contar la cantidad de cursos disponibles en el sitio web antes de iniciar el proceso de descarga:

Figura 4. Conteo de cursos

```
# información inicial:  
total_cursos = len(driver.find_elements(By.XPATH,  
'//*[@id="course_section_contents"]/div/div/div[2]/ul/li'))
```

La función `len()` permite contar la cantidad de elementos de la lista especificada por medio de un `XPATH`.

5.2.5. Selección de filtros

A continuación, se seleccionan todos los filtros necesarios para hacer la descarga.

Figura 5. Selección de filtros

```
# Seleccionar filtros:  
time.sleep(2)  
driver.find_element(By.CSS_SELECTOR, "#user_properties_section").click()  
time.sleep(1)  
cuenta_filtros = len(driver.find_elements(By.NAME, "selected_properties"))  
  
# Ciclo para elegir todos los atributos disponibles:  
for i in range(1, cuenta_filtros + 1):  
    driver.find_element(  
        By.XPATH, '//*[@id="id_selected_properties"]/li[' + str(i) + "]/label"  
    ).click()  
  
time.sleep(2)  
  
# Filtros para elegir formato de descarga (xls):  
driver.find_element(By.CSS_SELECTOR, "#format_section").click()  
driver.find_element(  
    By.CSS_SELECTOR, "#table-export-selection > li:nth-child(2)"  
).click()
```

La sentencia `for` recorre todos los filtros disponibles y da como resultado la selección de todas las propiedades del usuario y la selección del formato de exportación como se muestra en la siguiente imagen.

Figura 6. Ejemplo resultado

Seleccione las propiedades del usuario para mostrar ^

Nota: Si no se seleccionan propiedades de usuario, se utilizarán las propiedades predeterminadas.

<input checked="" type="checkbox"/> Apellidos	<input checked="" type="checkbox"/> Año de nacimiento	<input checked="" type="checkbox"/> Ciudad	<input checked="" type="checkbox"/> Correo	<input checked="" type="checkbox"/> Descripción
<input checked="" type="checkbox"/> Género	<input checked="" type="checkbox"/> Nombre	<input checked="" type="checkbox"/> Nombre de usuario	<input checked="" type="checkbox"/> País	<input checked="" type="checkbox"/> Tema

Elija el formato de exportación ^

XLS report

5.2.6. Ciclo repetitivo de descarga

En esta sección se genera el ciclo de descarga de datos que repite el proceso hasta descargar todos los cursos disponibles en la plataforma.

Figura 7. Ciclo repetitivo de descarga

```
# INICIO CICLO
while contador <= cant_descargas * incremento:
    while contador <= ciclos_descarga:
        digitar_cursos = ("#course_section_contents > div > div > div.panel > ul > li:nth-child("
            + str(contador)+ ") > label" )
        driver.find_element(By.CSS_SELECTOR, digitar_cursos).click()
        contador += 1
        ciclos_descarga += incremento

    driver.find_element(By.XPATH, aplicar).click()
    time.sleep(5)

driver.find_element(By.XPATH, '//*[@id="dialog-container"]')
driver.find_element(By.XPATH, '//*[@id="dialog-container"]/div[1]')
```

La anterior sentencia de código repite el proceso de descarga según la cantidad de cursos que tiene un sitio web dado, por lo que es funcional para descarga de archivos de todos los sitios web de la compañía.

5.2.7. Lectura, consolidación y carga de archivos

Una vez descargados todos los archivos en una carpeta local, se debe leerlos y consolidarlos en un dataframe, el código siguiente realiza este proceso:

Figura 8. Lectura, consolidación y carga de archivos

```
def borrar():
    py_files = glob.glob('C:/Users/Administrator/Downloads/*.xls')
    for xls_file in py_files:
        try:
            os.remove(xls_file)
        except OSError as e:
            print(f"Error:{ e.strerror}")
            pass

def generar_archivo(path_destino):

    path = 'C:/Users/Administrator/Downloads'
    xls_files = glob.glob(os.path.join(path, "*.xls"))
    today = time.strftime('%m-%d-%Y')
    dataframes= []

    if len(xls_files) > 0:
        # Loop over the list of xls files
        for i in xls_files:

            # read the xls file
            df = pd.read_excel(i)
            dataframes.append(df)

            print('Cantidad de archivos:'+ str(len(dataframes)))
            df_consolidado = pd.concat(dataframes)
            df_consolidado.to_excel(path_destino+"Profesionales_"+today+".xlsx", index=False)

    py_files = glob.glob('C:/Users/Administrator/Downloads/*.xls')

    for xls_file in py_files:
        try:
            os.remove(xls_file)
        except OSError as e:
            print(f"Error:{ e.strerror}")
```

La función *borrar()* recorre con ayuda de un ciclo *for* la carpeta de descargas y borra todos los elementos que sean de tipo *.xls*.

La función *generar_archivo()* busca si existen archivos de formato *.xls* en la carpeta de descargas, si existen, entonces añade cada uno de ellos a un dataframe y lo almacena con la fecha del día de la descarga en una carpeta local que se sincroniza automáticamente con un repositorio en OneDrive.

5.3. Inconvenientes presentados en el desarrollo del robot

El robot no podía acceder a todas las páginas de la compañía dado que algunas de ellas tenían un iframe que impedía al robot acceder con normalidad, para solucionar este inconveniente se realizó una profunda investigación en la documentación de Selenium y en foros en internet. Para dar solución a este error se utilizó la siguiente sentencia de código que permite cambiar entre un iframe y otro:

Figura 9. Solución problemas con iframe

```
#Quitar widget eanx
try:
    driver.switch_to.frame(1)
    driver.find_element(By.CLASS_NAME, "widget")
    driver.find_element(By.XPATH, "/html/body/div/div[1]/div/div/button").click()
    driver.switch_to.default_content()
except:
    driver.switch_to.default_content()
    pass

#Quitar widget griky
try:
    driver.switch_to.frame(2)
    driver.find_element(By.CLASS_NAME, "widget")
    driver.find_element(By.XPATH, "/html/body/div/div[1]/div/div/button").click()
    driver.switch_to.default_content()
except:
    driver.switch_to.default_content()
    pass
```

La sentencia de código anterior permitió que el robot pueda avanzar en el proceso de descarga sin importar la incorporación de un widget en algunos sitios web específicos de la compañía.

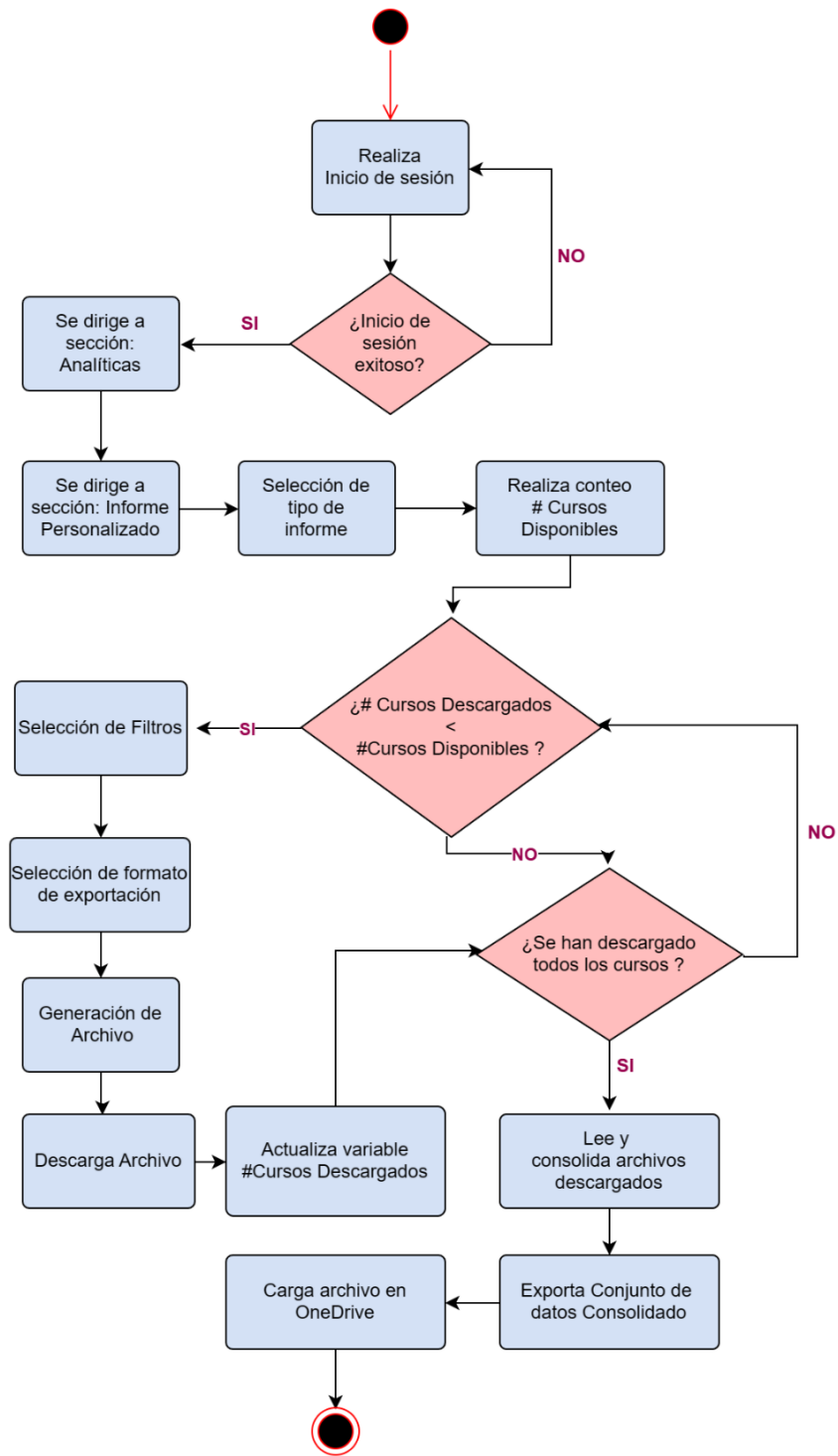
Finalmente, el proceso automatizado realiza una serie de pasos repetitivos y genera un archivo con toda la información relacionada al progreso de usuarios en cursos de una de las plataformas de Griky.

5.4. Representación gráfica del proceso

En esta sección se presentan dos diagramas de flujo que muestran de forma clara y concisa el proceso de descarga de información desarrollado. Estos diagramas, también conocidos como diagramas de actividad, son una herramienta visual útil para entender y explicar los distintos pasos del proceso de manera clara y sencilla. Cada diagrama muestra los pasos detallados a seguir en cada etapa, desde el inicio del proceso hasta su finalización.

En la figura 10, se representa el proceso de descarga de información realizado con el RPA. Desde el inicio de sesión en la página web de un cliente dado hasta la carga del archivo generado en OneDrive, se muestran todas las etapas necesarias para obtener y consolidar la información requerida.

Figura 10. Flujo RPA

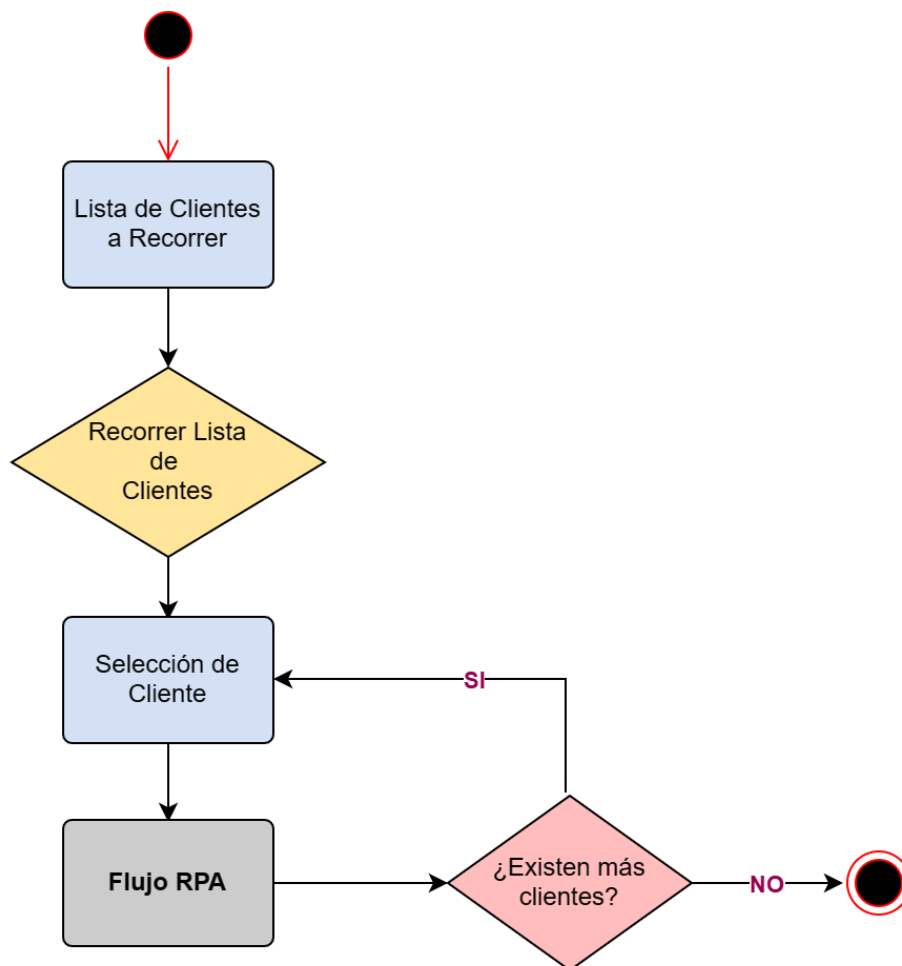


Fuente: Elaboración propia

Una vez generado y cargado el archivo en OneDrive, se puede acceder y utilizar la información de manera eficiente y sencilla. Este proceso simplifica la obtención de la información necesaria y evita la necesidad de realizar el proceso de manera manual, ahorrando tiempo y recursos valiosos.

La siguiente parte del proceso es recorrer una lista de clientes para obtener y consolidar información de manera automatizada utilizando el proceso "Flujo RPA" previamente descrito. Como se observa en la figura 11, cada uno de los pasos del proceso se repite para cada uno de los clientes en la lista, asegurando que se obtenga la información requerida de manera eficiente y sin errores.

Figura 11. Flujo de proceso de RPA para todos clientes.



Fuente: Elaboración propia

Una vez finalizado el proceso para todos los clientes en la lista, el archivo generado para cada uno de ellos se carga en el repositorio correspondiente en OneDrive. De esta forma, se asegura un almacenamiento seguro y accesible de la información generada. Este proceso reduce significativamente el tiempo y los recursos necesarios para obtener la información requerida y garantiza una mayor precisión y consistencia en los resultados obtenidos.

5.5. Prueba de Funcionamiento del RPA

Probar el correcto funcionamiento del RPA (Automatización de Procesos Robóticos) es una etapa crucial en el desarrollo de cualquier proyecto de automatización. En esta etapa se tiene como objetivo evaluar el rendimiento y la eficacia del robot en diferentes escenarios y situaciones. Se realizaron pruebas para verificar si el robot funcionaba correctamente bajo diferentes escenarios, si cumplía con los requisitos y objetivos del proyecto, y si era capaz de manejar excepciones y errores de manera adecuada. En este sentido, la prueba de funcionamiento del RPA es un paso fundamental para asegurar la calidad y fiabilidad del proyecto de automatización y garantizar su éxito. En esta sección se presentan las diferentes fases y metodologías de prueba que se pueden emplear para validar el correcto funcionamiento del robot. Además, de una pequeña explicación de uno de los inconvenientes presentados.

Para el momento de pruebas de descarga se hizo uso de un servidor Windows con 16gb de ram y 100gb de disco duro en donde se alojó el robot y se inició la ejecución para uno de los sitios web de la compañía con resultado exitoso. Posteriormente se hizo la ejecución con diferentes sitios web de la compañía, en donde se encontró que el RPA no funcionaba correctamente dado que algunos de estos sitios web presentaban un widget que impedía avanzar con el flujo normal; el problema y la solución referente al widget se detalla en la sección 5.3.

Se eligieron los sitios web de Classrun, Corona, Eanx y Uma para las pruebas porque son los clientes más grandes de Griky y representan la mayor cantidad de tiempo en descarga manual. De esta manera, al probar y validar el funcionamiento del aplicativo con estos clientes, se garantiza que sea capaz de manejar grandes volúmenes de cursos y usuarios. A continuación, en la tabla 11 se muestran los resultados del tiempo total gastado en la descarga con y sin uso del RPA.

Tabla 11. Tiempo Descarga Manual VS Tiempo Descarga por RPA en horas.

Nombre Sitio web	Url	Cantidad de cursos	Cantidad de usuarios	Tiempo de descarga manual en horas	Tiempo de descarga por RPA en horas
Classrun	https://campus.class-run.com/	1126	18099	10	6

Corona	https://go.coronaaprende.com	555	7550	8	4.9
Eanx	https://campus.eanx.io/courses	878	4191	5	3.1
Uma	https://umaplus.uma.edu.pe/	358	3986	4.5	3

En la tabla 11 se muestran los resultados del tiempo total gastado en la descarga de cursos en cuatro sitios web de la empresa, comparando la descarga manual con la descarga utilizando RPA (Automatización de Procesos Robóticos). En la tabla se observa una reducción del tiempo de descarga de alrededor del 40% en promedio cuando se utiliza el RPA. Los sitios web evaluados son: Classrun, Corona, Eanx y Uma. Se presentan datos como la cantidad de cursos disponibles en cada sitio, la cantidad de usuarios, el tiempo total gastado en la descarga manual en horas, y el tiempo total gastado en la descarga utilizando RPA en horas.

En cuanto a los ahorros en tiempo, es importante destacar que, aunque se observa una reducción del tiempo de descarga de alrededor del 40% en promedio al utilizar el RPA, estos ahorros no son tan grandes como se esperaría. Esto se debe a que el proceso automatizado no depende únicamente del robot, sino que también está condicionado por la generación de archivos por parte de la página web que contienen la información del progreso de cada curso. El robot debe esperar a que estos archivos sean generados y realizar el proceso de descarga repetitivamente hasta completar todos los cursos para un cliente dado. Por lo tanto, las limitaciones impuestas por la página web pueden afectar el rendimiento y el tiempo total requerido en la descarga, limitando así el ahorro de tiempo obtenido con el aplicativo.

En las pruebas realizadas, se buscó probar el buen funcionamiento del aplicativo de descarga automática, evaluando su rendimiento y eficacia en diferentes escenarios y situaciones. El objetivo de estas pruebas fue garantizar que el RPA cumpliera con los requisitos y objetivos del proyecto. Específicamente, se buscó probar que el robot funcionara correctamente bajo diferentes escenarios (sitios web de diferentes clientes), verificando si era capaz de manejar excepciones y errores de manera adecuada. Además, se buscó evaluar el cumplimiento de los requisitos y objetivos del proyecto por parte del robot durante la descarga de cursos desde los sitios web de los clientes.

Después de realizar estas pruebas, se tiene la garantía de que el aplicativo de descarga automática funciona de manera eficiente y confiable, ya que se pusieron a prueba diferentes casos y situaciones que podrían afectar su rendimiento. Esto asegura que el robot es capaz de manejar las tareas de descarga de cursos de manera exitosa, cumpliendo con los requerimientos del proyecto y garantizando una alternativa efectiva a la descarga manual.

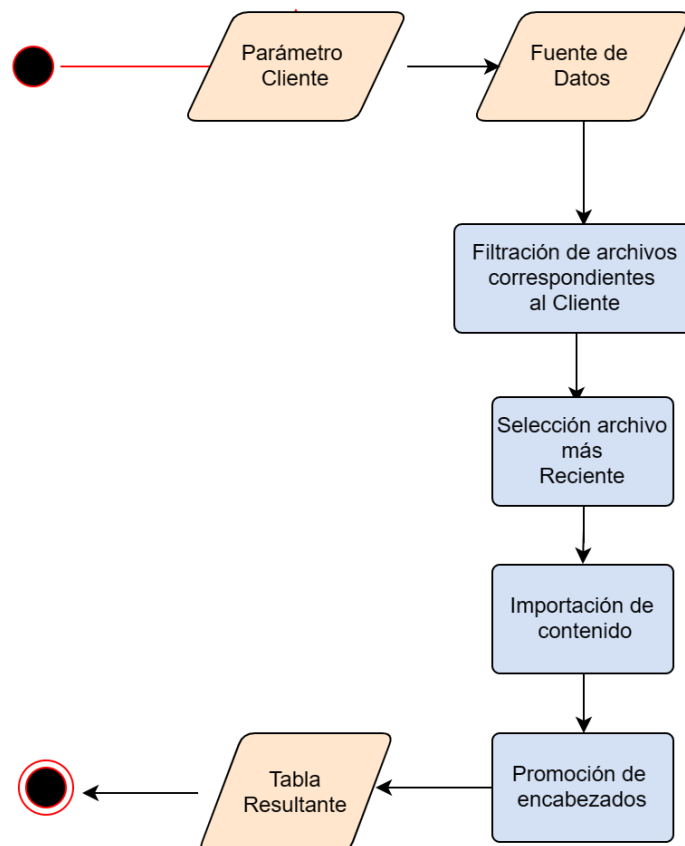
Los resultados evidencian que la implementación de RPA es una solución efectiva para reducir significativamente el tiempo necesario para descargar cursos en sitios web de la empresa. Se evidenció que la implementación de RPA logró reducir significativamente el tiempo necesario para

la descarga, mejorando la eficiencia del proceso. Sin embargo, los ahorros en tiempo no son tan grandes como se esperaría debido a las limitaciones impuestas por la generación de archivos por parte de las páginas web, lo cual condiciona el rendimiento y el tiempo total requerido en la descarga. Estas pruebas estratégicas y menos operativas permiten garantizar el buen funcionamiento del aplicativo de descarga automática y ofrecen una visión más clara sobre la eficacia del RPA en el proceso de automatización.

6. PROCESAMIENTO DE DATOS

En esta sección se detalla el procedimiento utilizado para obtener la información del progreso de usuarios en la plataforma de educación para un cliente en específico y de forma consolidada considerando todos los clientes. Para ello, se utilizó Power Query dentro de Power BI para crear la función "*get_Progreso_Interno*". Esta función toma como origen de datos archivos en excel generados por el RPA que se encuentran almacenados en OneDrive, los filtra para seleccionar el último archivo generado, se importa el contenido y se genera una tabla con los datos de progreso para el cliente deseado. En la figura 12, se muestra el diagrama de flujo de la función "*get_Progreso_Interno*", el código de la función de encuentra en el *Anexo 2*.

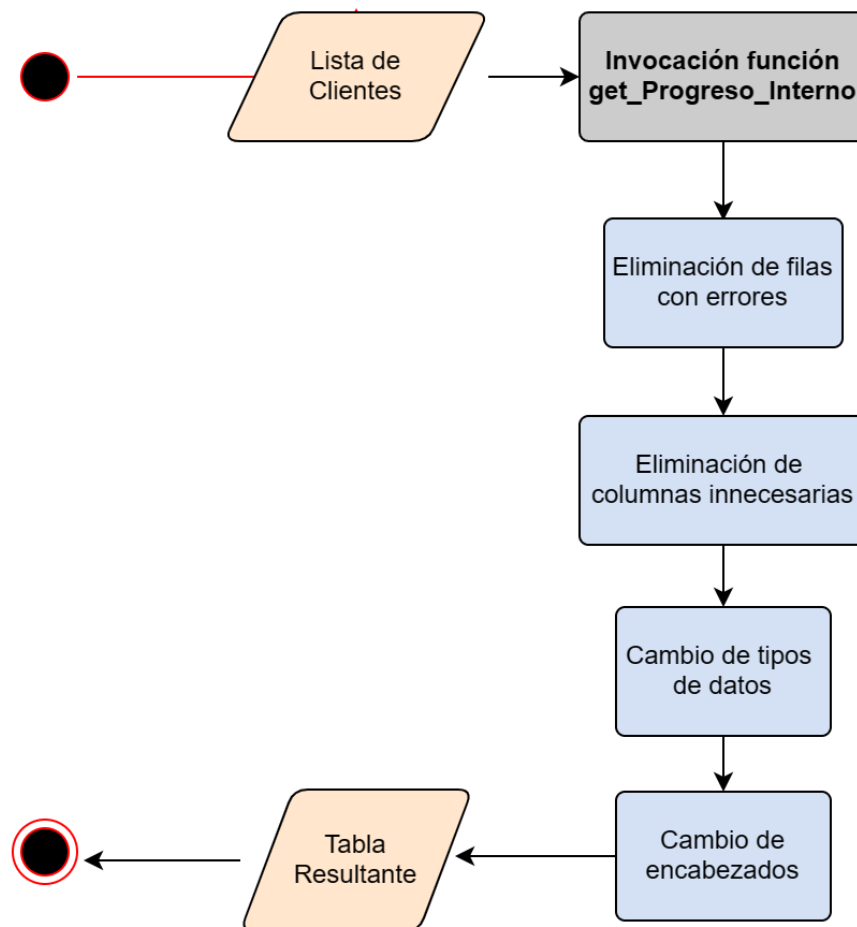
Figura 12. Diagrama de flujo función *get_Progreso_Interno*



Fuente: Elaboración propia

Posteriormente, se creó una tabla llamada "progreso_interno" que utiliza una lista de clientes para invocar la función "get_Progreso_Interno" para cada uno de ellos. Una vez obtenidos los datos, se procedió a realizar una limpieza y procesamiento de estos para dejarlos listos para su utilización en Power BI. Esto incluyó la eliminación de filas con errores y de columnas innecesarias, cambio de tipos de datos de cada columna, cambio de encabezados, entre otros. En la figura 13, se muestra el diagrama de flujo de la tabla "progreso_interno".

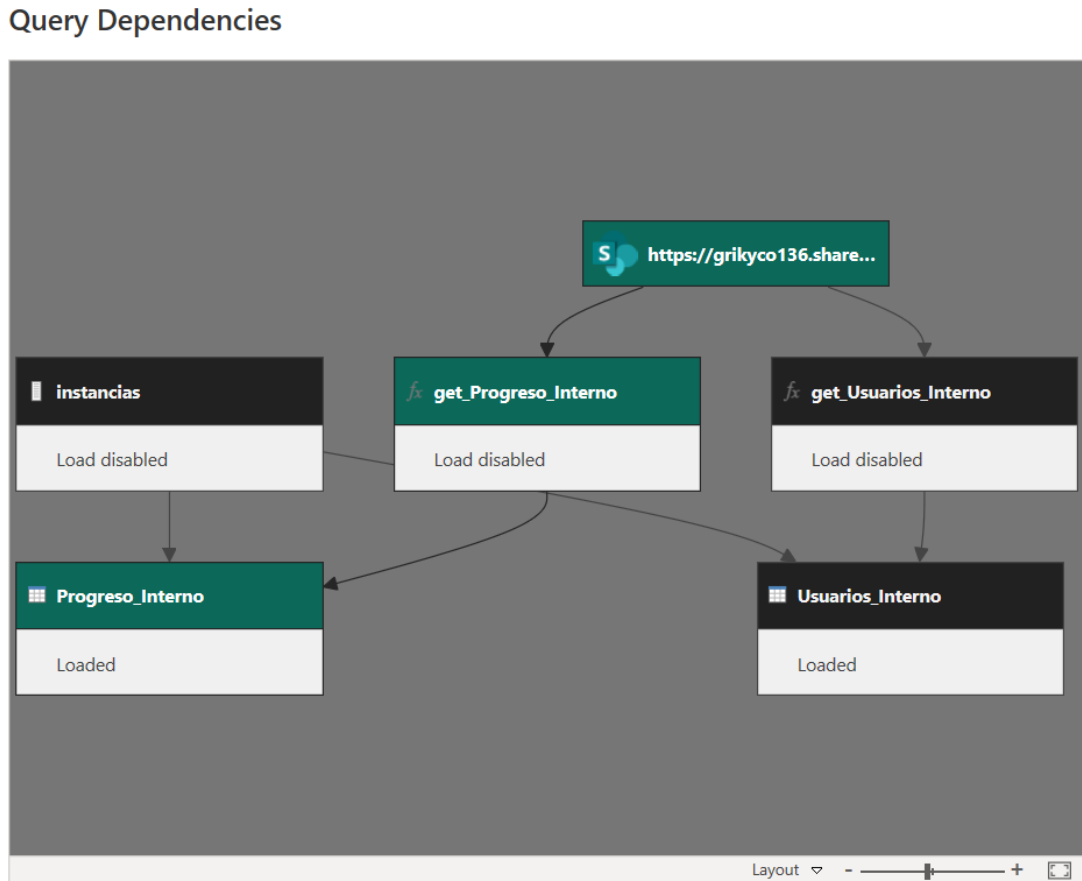
Figura 13. Diagrama de Flujo transformación de datos



Fuente: Elaboración propia

En la Figura 14, se presenta el diagrama de Power Query que resume el modelo de datos generado. Este incluye la función y tabla previamente mencionadas, junto con otra función y tabla que se enfocan en obtener información específica del usuario. La función se llama "get_Usuarios_Interno" y la tabla se llama "Usuarios_Interno", ambas realizan un proceso similar a la función "get_Progreso_Interno" y a la tabla "Progreso_Interno" mencionadas recientemente.

Figura 14. Diagrama de Flujo Modelo de datos



Fuente: Elaboración propia- Flujo elaborado en Power Query

La utilización de Power Query dentro de Power BI permitió procesar de manera eficiente los datos del progreso de usuarios en la plataforma de educación, permitiendo obtener información relevante y lista para utilizar. El proceso descrito en esta sección permitió importar, filtrar, y generar tablas para cada cliente, así como también realizar una limpieza y procesamiento de estos para su posterior utilización en Power BI. El resultado final permite continuar con la etapa de visualización de datos obtenidos.

7. RESULTADOS VISUALES Y ANALÍTICAS

La visualización de los resultados y las analíticas son un aspecto crucial en cualquier proyecto de automatización. Una vez que se ha implementado la solución de RPA, es importante analizar los datos y resultados generados por el robot para obtener información valiosa que pueda ser utilizada para tomar decisiones en la empresa. En primer lugar, se muestra un resumen general de algunos datos relevantes para conocer el estado actual de los usuarios en la plataforma como se indica en la Figura 15.

Figura 15. Resumen General



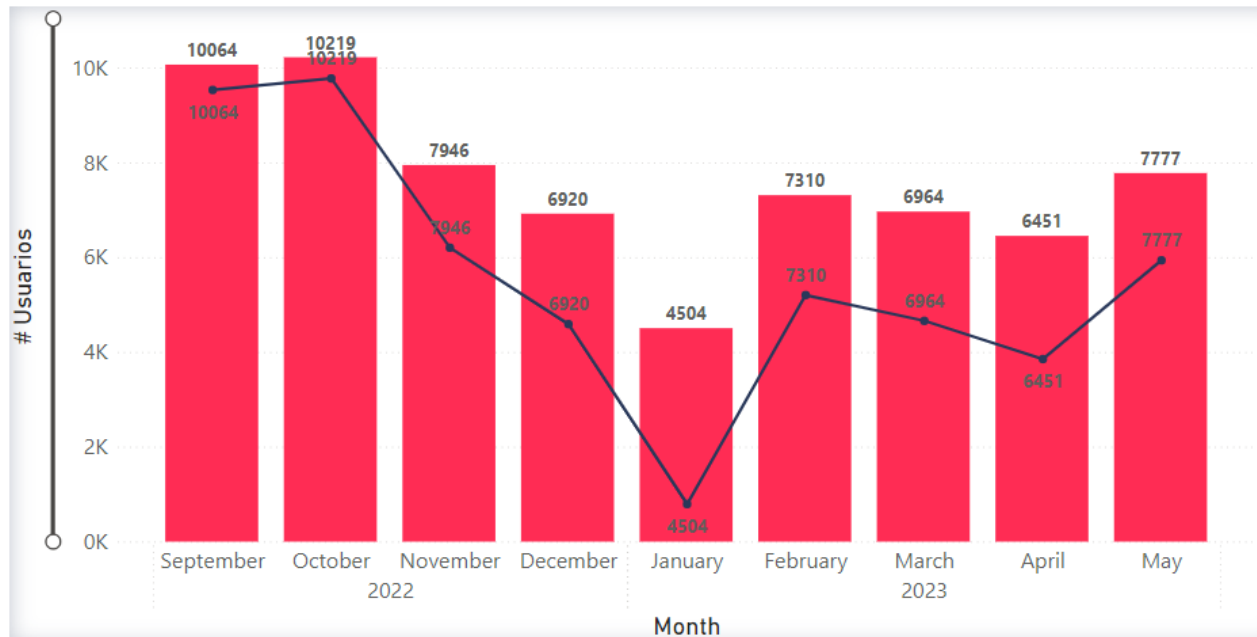
Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

En la Figura 15 se resumen algunos datos que permiten conocer de manera general el comportamiento de los estudiantes en la plataforma considerando todos los clientes. Por ejemplo, la Tasa de finalización indica que solo el 18% de los cursos matriculados se han finalizado. Este dato puede interpretarse de varias formas, por un lado, podría indicar que los estudiantes abandonan los cursos antes de finalizarlos debido a un problema en la estructura o diseño de estos, o tal vez porque no han logrado mantener la motivación y el compromiso necesarios para completarlos. También podría indicar que la plataforma cuenta con un gran número de cursos disponibles y los estudiantes prefieren probar diferentes opciones antes de decidir comprometerse con uno en particular. En cualquier caso, este dato es importante para entender el comportamiento de los estudiantes y puede ser utilizado para mejorar la calidad de los cursos y aumentar su tasa de finalización.

7.1. Gráficos seleccionados

A continuación, en la Figura 16, se indica una métrica interesante para comprender el comportamiento de los usuarios en la plataforma de Griky; los MAU's, este indicador es utilizado por empresas de redes sociales y otras compañías para contar el número de usuarios únicos que han visitado un sitio web en el último mes [30]. Es una métrica importante para medir el crecimiento y la rentabilidad de una plataforma, ya que representa la cantidad de usuarios activos en función del tiempo.

Figura 16. Métrica General: Monthly Active Users

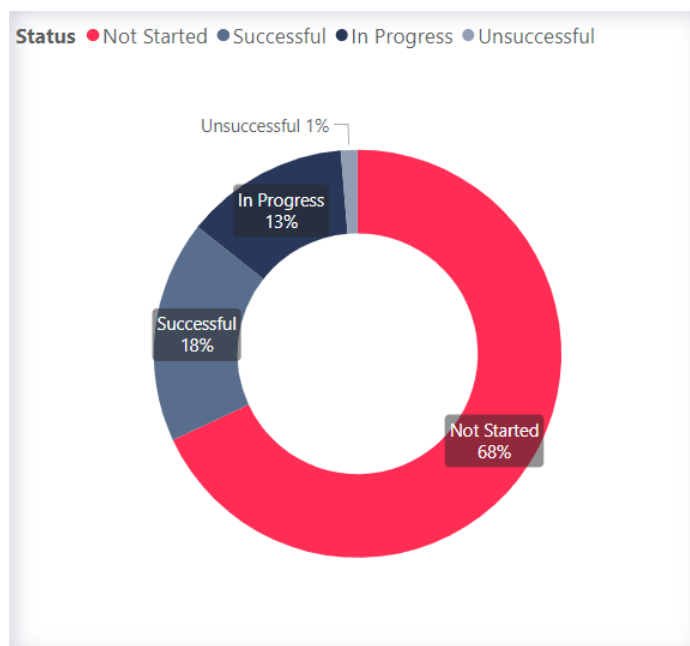


Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

La Figura 16 muestra una tendencia negativa de los MAU's en la plataforma desde octubre de 2022 hasta mayo de 2023. Se observa una disminución en la cantidad de usuarios activos a partir de octubre de 2022, que podría ser explicada por el hecho de que gran parte de los clientes de Griky son universidades y en ese período los estudiantes salen a vacaciones. Sin embargo, a partir de enero de 2023 se observa un aumento en la cantidad de MAU's, lo que sugiere que una vez que los estudiantes regresan a clases, la plataforma se convierte en una herramienta valiosa para el aprendizaje y el desarrollo. En general, es importante considerar diferentes factores que puedan afectar la tendencia de los MAU's para entender mejor el comportamiento de los usuarios en la plataforma y mejorar la calidad de los servicios ofrecidos.

La Figura 17 muestra los datos de las matrículas clasificadas por su estado. Se puede observar que la mayoría de las matrículas corresponden a cursos que no se han iniciado (Not Started) correspondiente al 68% del total. Sin embargo, también se observa un número significativo de cursos en progreso (In Progress) con un 13% y cursos que fueron completados exitosamente (Successful) con un 18%. En general, estos datos pueden ser utilizados para entender el comportamiento de los estudiantes en relación a la finalización de los cursos, y pueden ayudar a la plataforma a enfocar sus estrategias en mejorar la experiencia y motivación para completar los cursos.

Figura 17. Cantidad de Matrículas por Estado



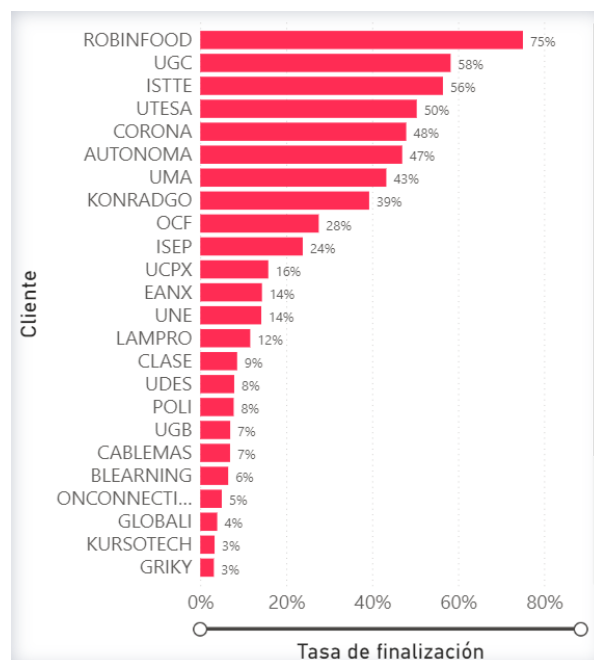
Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

En particular, es necesario poner especial atención en los cursos que no se han iniciado y en los que están en progreso, ya que estos representan la mayor cantidad de matrículas en la plataforma. Se podría explorar la razón detrás de la falta de iniciación en los cursos y realizar mejoras para incentivar a los estudiantes a completarlos. Por otro lado, también hay que considerar aquellos cursos que no fueron completados exitosamente, ya que pueden requerir mejoras en términos de diseño y contenido. En resumen, la información que proporciona este gráfico puede ser utilizada para tomar medidas con el objetivo de mejorar la experiencia del usuario y aumentar la tasa de finalización de los cursos.

Para Griky es importante enfocar el análisis de los datos de los usuarios en la plataforma a nivel de clientes, por eso, es importante poder identificar aquellos clientes con un mejor rendimiento y eficacia en cuanto a la finalización de cursos se refiere. Para ello, se presentan dos figuras clave que ayudan a entender y comparar el comportamiento de los clientes en la plataforma.

La Figura 18 muestra un diagrama de barras que ordena los clientes con la mayor tasa de finalización de cursos. En este ordenamiento se observa que el cliente RobinFood tiene la tasa más alta con el 75%, seguido por UGC, ISTTE con 58% y 56% respectivamente. Esta información es importante para entender el rendimiento de los clientes en relación a la finalización de los cursos, y puede ayudar a la plataforma a enfocar sus estrategias para mejorar la tasa de finalización principalmente en los clientes que tienen la tasa más baja.

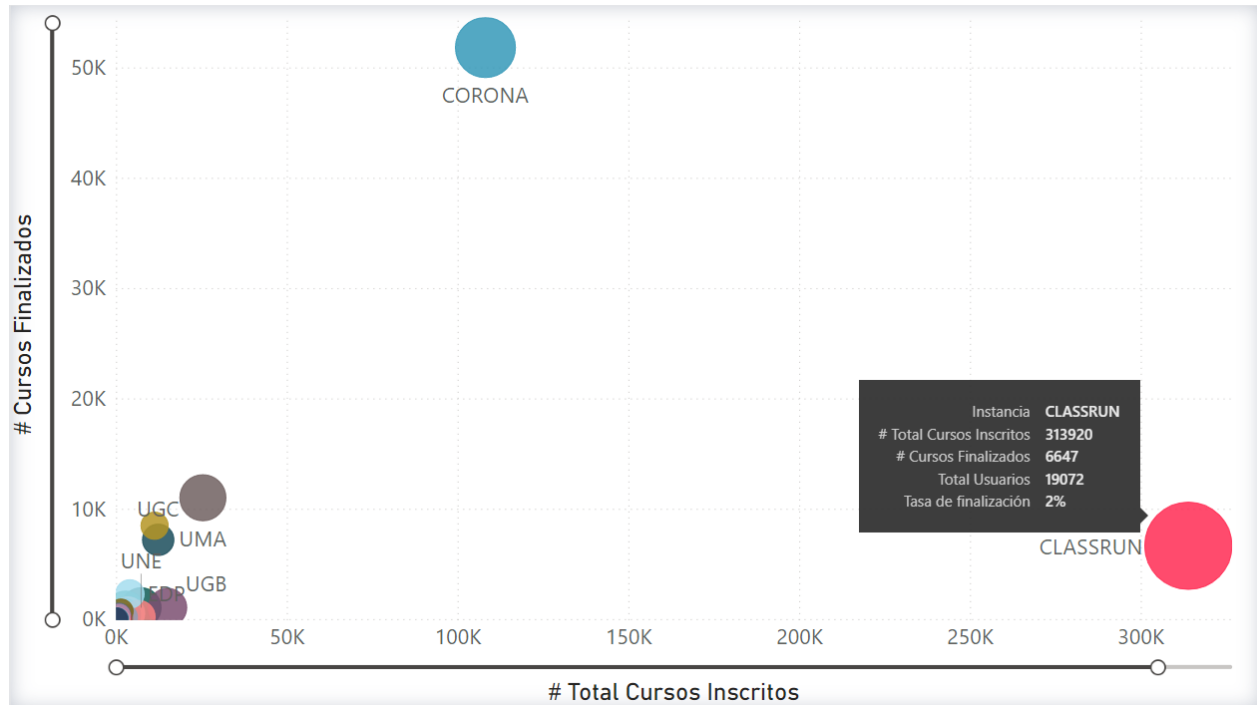
Figura 18. Tasa de Finalización por Cliente



Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

Para complementar la información anterior, la Figura 19 muestra un diagrama de dispersión que compara el número de cursos finalizados versus el número de cursos inscritos por cliente. Este gráfico es importante ya que permite identificar clientes que, aunque tienen una tasa de finalización baja, pueden tener una gran cantidad de cursos inscritos y, por lo tanto, representan un segmento importante en la plataforma. Por ejemplo, se puede observar que el cliente *Classrun* tiene una tasa de finalización baja (2%), pero ha inscrito más del tripe de cursos que el siguiente con mayor inscripción (314k). Esta información puede ser útil para entrar en contacto con el cliente y buscar soluciones para mejorar su experiencia en la plataforma y, en consecuencia, mejorar su tasa de finalización de cursos.

Figura 19. Total Cursos Inscritos VS Cursos Finalizados por Cliente



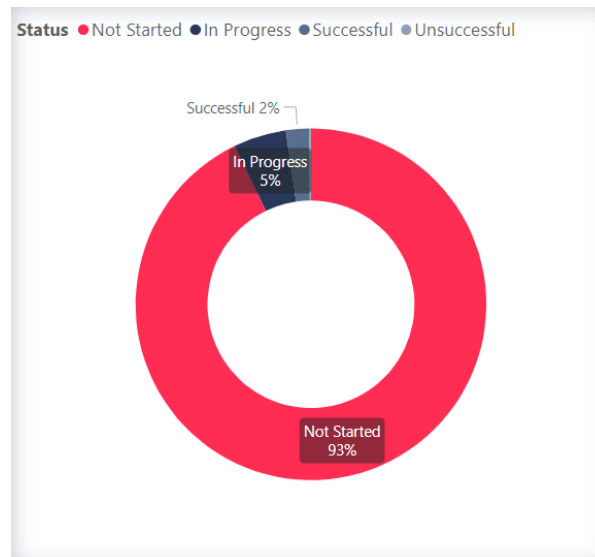
Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

Teniendo en cuenta lo anterior, ahora se va a analizar un cliente en específico: "Classrun". Este cliente representa un caso especial debido a su comportamiento en la plataforma, ya que a pesar de haber inscrito más del triple de cursos que el cliente con siguiente mayor inscripción, su tasa de finalización es muy baja, solo del 2%.

La baja tasa de finalización de "Classrun" podría ser atribuida a varios factores, tales como la complejidad de los cursos, la falta de motivación de los estudiantes o la falta de compromiso. También puede ser resultado de problemas en la estructura y diseño de los cursos realizados por "Classrun", o de un déficit en la calidad de los materiales educativos que ofrece a sus estudiantes.

Un dato interesante es que el 93% de matrículas en las que se han inscrito usuarios de *Classrun* no han iniciado, mientras que solo el 2% reflejan matrículas o cursos finalizados como se observa en la Figura 20.

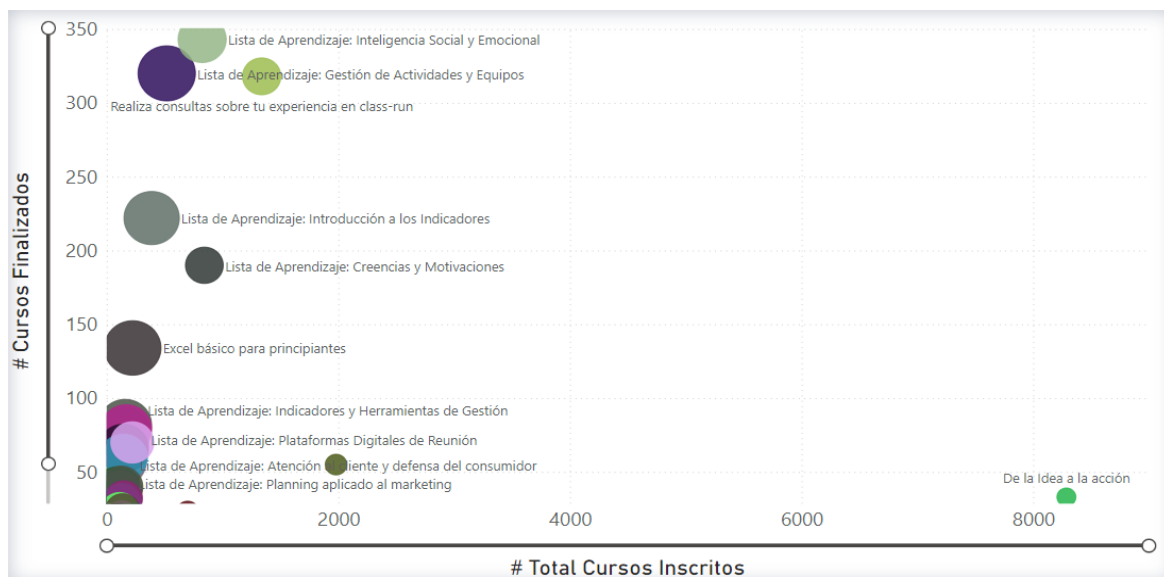
Figura 20. Cantidad de Matrículas por Estado- Cliente: Classrun



Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

Para profundizar en el análisis de por qué los usuarios de "Classrun" se inscriben en cursos pero no los consumen, se presenta el siguiente diagrama de dispersión que compara la cantidad de matrículas versus la cantidad de veces que se ha finalizado un curso para este cliente (Figura 21). Esta figura resulta importante para entender cuál es el comportamiento de los usuarios en relación con la finalización de los cursos ofrecidos por "Classrun" y para poder determinar posibles factores que estén influyendo en su baja tasa de finalización.

Figura 21. Total matrículas VS Total matrículas Finalizadas por Curso y Cliente: Classrun



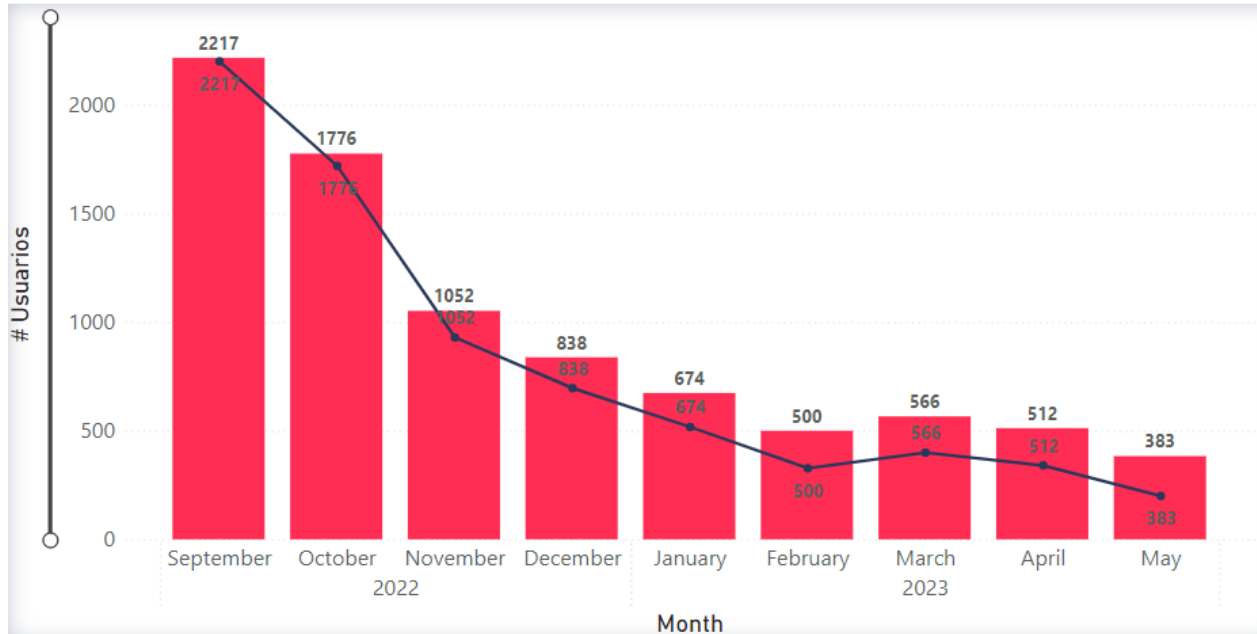
Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

Pensando en mejorar la tasa de finalización en "Classrun", es importante analizar los cursos que presentan una mayor tasa de finalización (parte superior del gráfico) y determinar las posibles razones detrás de su éxito. Una estrategia útil puede ser analizar la categoría de los cursos que presentan una tasa de finalización alta, como también el nivel de dificultad, la duración de estos, etc.

De esta forma se puede obtener información más detallada sobre las preferencias de los estudiantes y los factores que influyen en lograr una tasa de finalización elevada. Por ejemplo, si se observa que los videos cortos y prácticos tienen mejor aceptación, se pueden implementar estas características en cursos que tienen una baja tasa de finalización.

Además del análisis anterior, es importante destacar la información proporcionada por la gráfica de MAUs (Usuarios Activos Mensuales) para el cliente "Classrun" (Figura 22). Se observa que hay un comportamiento negativo en la pendiente de la gráfica desde septiembre de 2022 a mayo de 2023. Esta tendencia puede estar relacionada con la baja tasa de finalización de los cursos ofrecidos por "Classrun", ya que un bajo compromiso de los estudiantes en la plataforma podría estar influyendo en este comportamiento.

Figura 22. Métrica General: Monthly Active Users- Cliente: Classrun



Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

Es importante que la plataforma realice una revisión a fondo de la experiencia de usuario que se ofrece, incluyendo revisar el contenido y el diseño de los cursos, las estrategias utilizadas para motivar a los estudiantes a completarlos, y la facilidad con la que se accede a los mismos dentro de la plataforma. Es crucial mejorar la calidad de la educación ofrecida y revisar el proceso de inscripción a cursos para que los usuarios no abandonen la plataforma y sigan participando activamente en ella.

En general, la información proporcionada por estas gráficas es valiosa para entender el comportamiento de los usuarios en la plataforma, lo que permitirá enfocar las estrategias y desarrollar herramientas para mejorar la participación y completitud de los cursos, lo que a su vez repercutirá en la retención y lealtad de los estudiantes, así como en la satisfacción de los clientes.

Por tanto, analizar los datos del cliente "Classrun" con detenimiento para identificar las causas de su baja tasa de finalización, y diseñar estrategias enfocadas en mejorar la experiencia del usuario y su compromiso con la plataforma. Por ejemplo, se puede ofrecer a los estudiantes incentivos para finalizar los cursos, como descuentos en próximas matrículas o reconocimientos por logros alcanzados. En cualquier caso, la situación de "Classrun" pone de manifiesto la importancia de una permanente supervisión y mejora en la calidad de los cursos ofrecidos por la plataforma para garantizar el éxito de los estudiantes y el negocio.

En conclusión, los datos de progreso de los usuarios en la plataforma Griky proporcionados a través de la herramienta RPA permiten identificar patrones de comportamiento de los estudiantes y clientes, lo que es crucial para mejorar la experiencia de usuario y para la fidelización de estos. A partir del análisis de datos, se pueden enfocar estrategias y soluciones específicas para mejorar la tasa de finalización de cursos, la retención y lealtad de los estudiantes, y la satisfacción de los clientes. En definitiva, la implementación del robot RPA para recolectar y analizar estos datos es una herramienta valiosa para optimizar la plataforma y mejorar el rendimiento de la empresa.

7.2. Dashboard Progreso de Usuarios en Griky

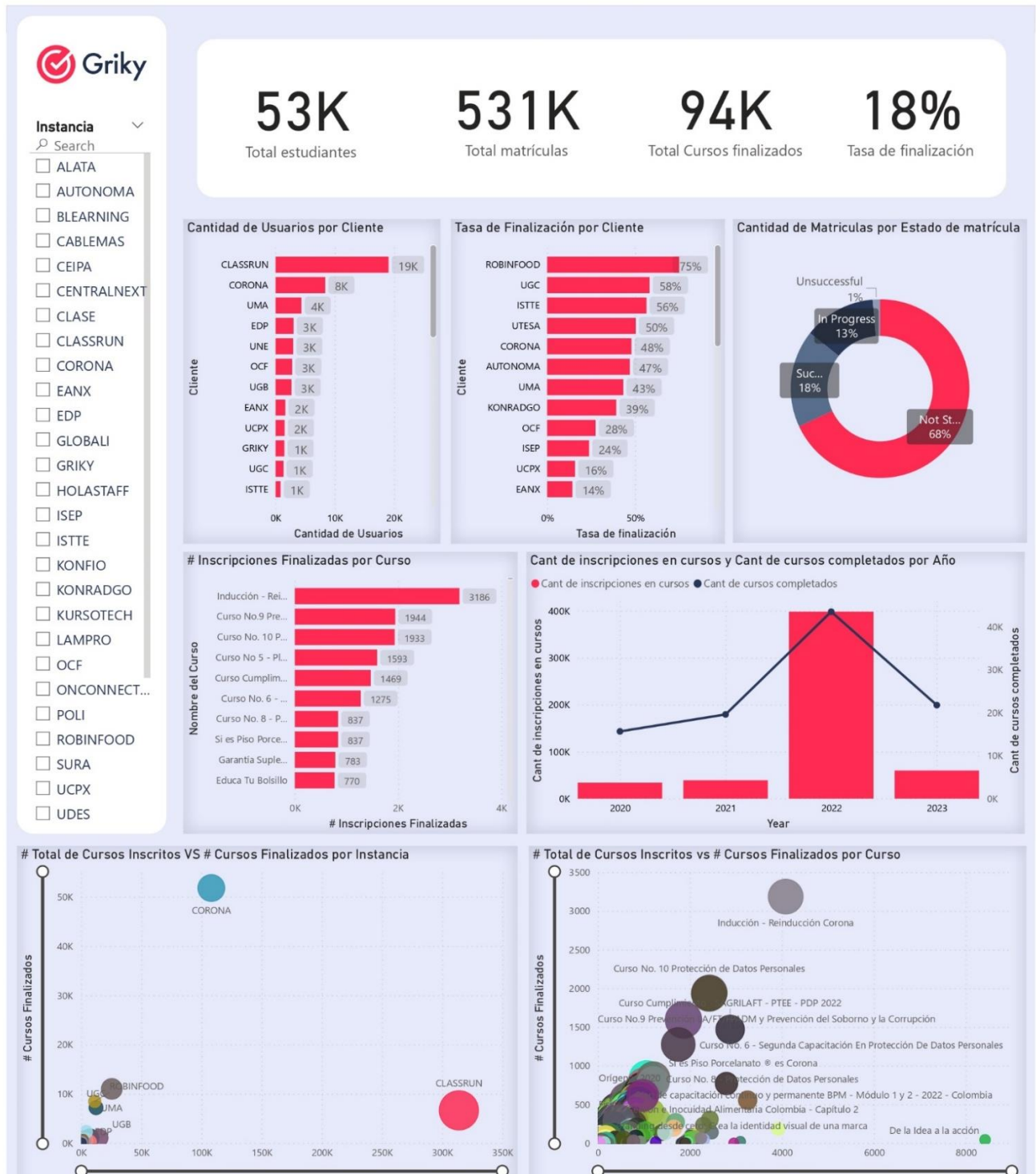
Con el objetivo de agrupar las visualizaciones descritas en el inciso anterior, se creó un dashboard que ofrece varias ventajas a la hora de conocer el progreso de los usuarios de un cliente dado. Este dashboard cuenta con filtros que permiten segmentar y analizar los datos de manera más efectiva, lo que facilita la identificación de patrones y tendencias. Además, este dashboard ofrece la posibilidad de mostrar la información a los interesados de manera clara y concisa, lo que facilita la toma de decisiones basadas en datos. Gracias a estas funciones, el dashboard se convierte en una herramienta fundamental para medir y optimizar el rendimiento de los usuarios, permitiendo a los equipos tomar acciones concretas para mejorar la experiencia del cliente y maximizar los resultados.

El dashboard fue construido utilizando la plataforma Power BI, que ofrece diversas ventajas técnicas para su desarrollo y uso. Power BI ofrece integración de diversas fuentes de datos, visualizaciones interactivas, generación de informes personalizados y compartición segura. También permite la creación de paneles personalizados y establecimiento de alertas para detectar problemas y tomar medidas inmediatas [29].

Gracias a la implementación del RPA, ahora se cuenta con datos que permiten actualizar el dashboard con una frecuencia diaria. El dashboard es dinámico y se puede interactuar con el haciendo uso del siguiente enlace: [Griky-Dashboard](#).

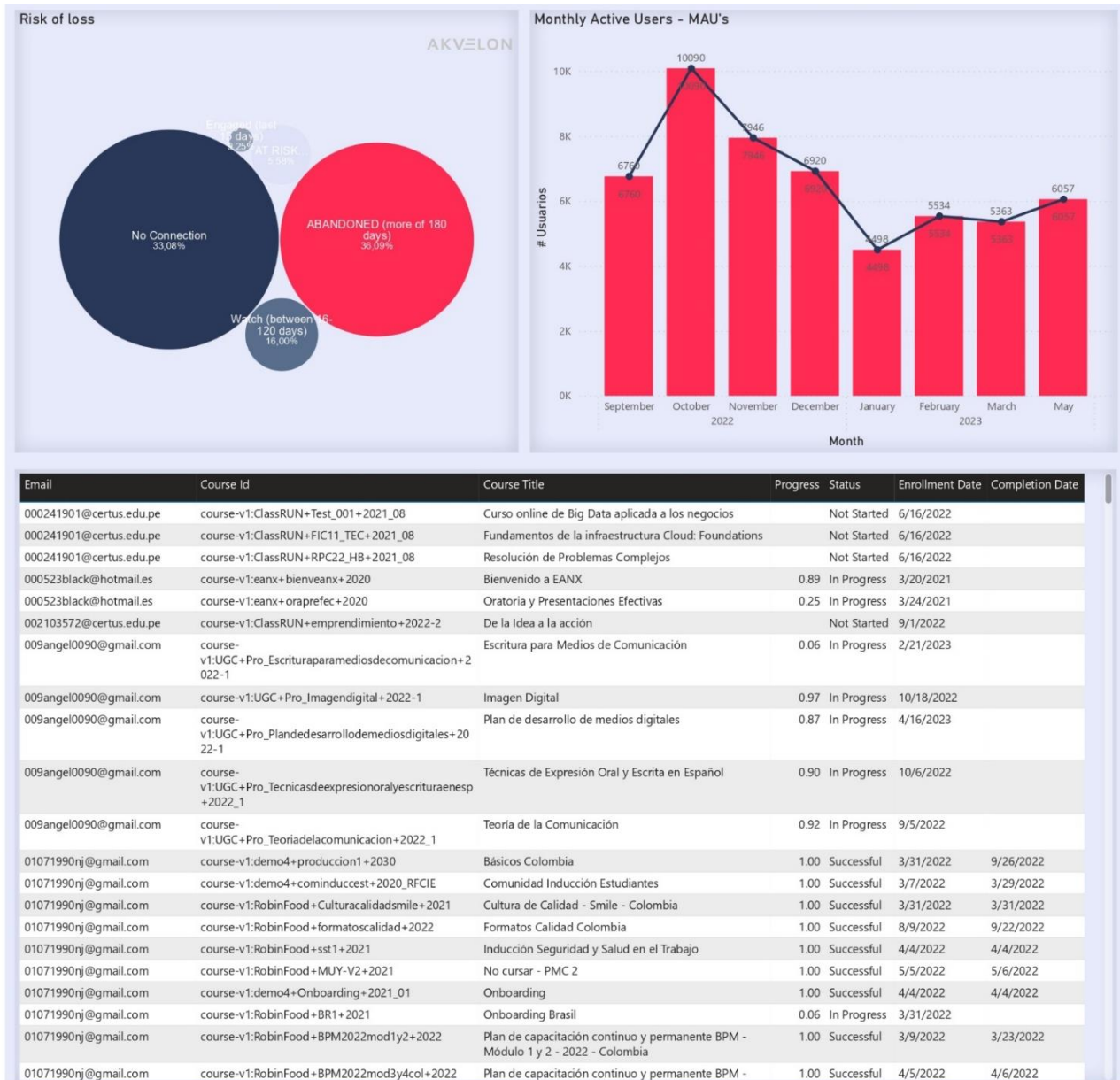
A continuación, en las figuras 23 y 24 se presenta el dashboard que resume las gráficas detalladas en la sección 7.1. Para obtener el mayor provecho a la información presentada en el dashboard, se recomienda hacer uso de los filtros en la parte izquierda del mismo para hacer un análisis por cliente. En la parte final del dashboard se presenta una tabla que resume la información del progreso de usuarios en un curso dado, esta información puede ser exportada en un archivo Excel haciendo uso del enlace mencionado anteriormente y con unas credenciales de acceso que son generadas y administradas por el propietario del dashboard.

Figura 23. Dashboard Progreso Usuarios en Griky 1



Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

Figura 24. Dashboard Progreso Usuarios en Griky 2



Fuente: Elaboración Propia. Disponible en: [Griky-Dashboard](#)

8. CONCLUSIONES Y TRABAJOS FUTUROS

8.1. CONCLUSIONES

8.1.1. Conclusiones generales

- La implementación de un robot RPA en la plataforma Griky ha permitido recopilar y analizar datos relevantes sobre el comportamiento de los estudiantes en las plataformas de los clientes de la empresa.
- El análisis de estos datos ha proporcionado información valiosa para comprender los patrones de comportamiento, las tendencias y los desafíos que enfrenta la plataforma.
- Las visualizaciones y métricas presentadas han permitido identificar áreas de mejora, como la tasa de finalización de cursos y la retención de estudiantes.

8.1.2. Conclusiones sobre el análisis de datos

- El análisis de la tasa de finalización de cursos ha revelado patrones interesantes, como diferentes tendencias dependiendo del cliente en estudio, lo que ha permitido identificar los clientes con un alto rendimiento y los que presentan tendencias negativas en el tiempo.
- Se ha identificado que los cursos no iniciados representan la mayoría de las matrículas totales (alrededor del 68%), lo que sugiere posibles áreas de mejora en la motivación y compromiso de los estudiantes; los estudiantes se matriculan en cursos pero no los finalizan.
- El análisis de los clientes con mejores tasas de finalización ha permitido identificar estrategias exitosas que se pueden implementar para mejorar la tasa de finalización en otros clientes. Por ejemplo, basándose en los tipos de cursos que consumen los clientes exitosos.
- La identificación de clientes con altas tasas de matrículas, pero baja tasa de finalización ha resaltado la importancia de abordar las posibles barreras que impiden la finalización de los cursos.

8.1.3. Conclusiones sobre ciencia de datos

- La aplicación de técnicas de ciencia de datos y análisis estadístico ha brindado una visión profunda de los datos generados por la plataforma Griky.
- La capacidad de recopilar, procesar y visualizar grandes volúmenes de datos de manera automatizada a través del robot RPA ha mejorado significativamente la eficiencia y precisión del análisis.

- El enfoque en la identificación de patrones, tendencias y factores influyentes ha permitido tomar decisiones más informadas y desarrollar estrategias personalizadas para mejorar la experiencia del usuario.
- La implementación de un dashboard para la presentación visual de los resultados ha facilitado la interpretación y la comunicación de los hallazgos a los stakeholders y equipos de trabajo.

En cuanto al cumplimiento de los objetivos planteados en el proyecto, se puede afirmar que se han logrado satisfactoriamente. El objetivo principal de obtener datos relacionados con el progreso de los usuarios y utilizarlos para tomar decisiones estratégicas ha sido alcanzado con éxito. La implementación de la herramienta RPA ha brindado a Griky la capacidad de recopilar y analizar datos de manera efectiva, lo que ha permitido mejorar la experiencia del usuario y fortalecer la posición de la compañía en el mercado de la educación en línea.

En resumen, el proyecto de implementación de la herramienta RPA permitió obtener datos de gran utilidad para Griky y ha sido un paso importante hacia la mejora continua y la innovación en el campo de la educación en línea. Los logros obtenidos, la capacidad de mantenerlos y el cumplimiento de los objetivos establecidos demuestran el compromiso de Griky con la excelencia educativa y su capacidad para adaptarse y prosperar en un entorno empresarial altamente competitivo.

8.2. TRABAJOS FUTUROS

Los datos recopilados por la herramienta RPA creada en este proyecto proporcionan una gran cantidad de información que puede ser utilizada para mejorar la experiencia del usuario y el rendimiento de la empresa. En este sentido, hay varios trabajos futuros que podrían abordarse a partir de estos datos, tales como:

- Desarrollo de modelos de machine learning para recomendaciones de cursos: una herramienta que puede ser muy útil para mejorar el compromiso y la motivación de los usuarios en la plataforma es la recomendación personalizada de cursos. Para ello, se puede emplear la técnica de machine learning para generar recomendaciones basadas en los datos recopilados por el robot RPA, como los cursos previamente tomados y la tasa de completitud de estos.
- Generación de engagement en los usuarios de Griky: otro trabajo futuro que se puede abordar a partir de los datos de progreso de los usuarios en Griky es la generación de engagement. Esto se puede realizar mediante la creación de herramientas que fomenten la participación y el intercambio de conocimiento entre estudiantes, la gamificación de los cursos, o la personalización de la plataforma, en base a los gustos y preferencias de cada usuario.
- Análisis más exhaustivo de los datos: una vez recopilados los datos de progreso de los usuarios, se abre la posibilidad de realizar un análisis más exhaustivo de los mismos. Esto permitiría comprender mejor el comportamiento de los usuarios en la plataforma, y enfocar estrategias más específicas para mejorar su experiencia.

En conclusión, gracias al robot RPA generado se cuenta con datos valiosos para mejorar la plataforma y el rendimiento de la empresa. Los trabajos futuros que se pueden abordar a partir de estos datos son variados, y permitirán a Griky ofrecer una experiencia más personalizada y atractiva para los usuarios.

9. REFERENCIAS BIBLIOGRÁFICAS

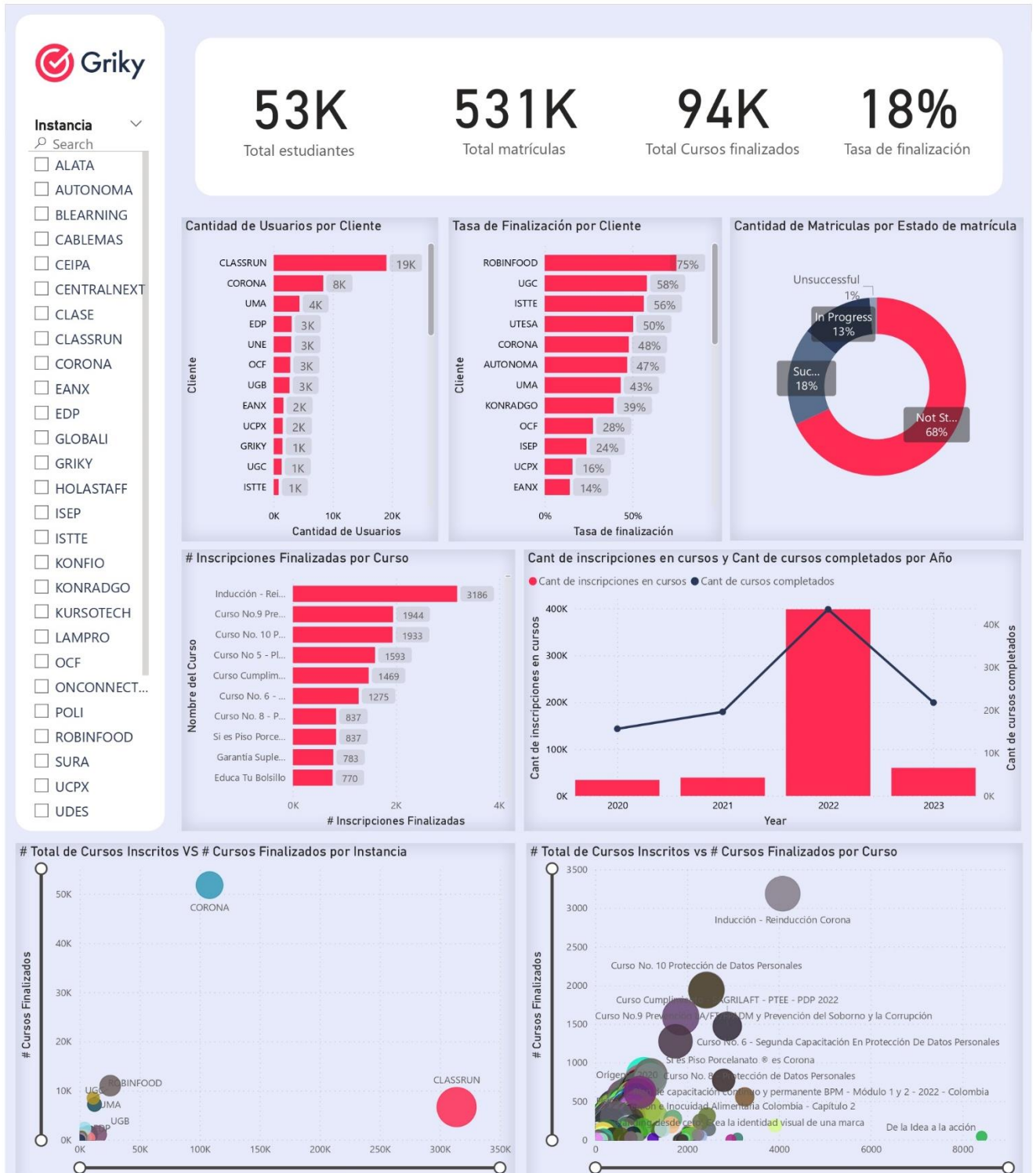
- [1] J. Caparros, «Griky, plataforma de educación en línea, obtiene US\$5 millones al sobrepasar los 200.000 estudiantes,» 29 Noviembre 2021. [En línea]. Available: <https://forbes.co/2021/11/29/emprendedores/griky-plataforma-de-educacion-en-linea-obtiene-us5-millones-al-sobrepasar-los-200-000-estudiantes/>. [Último acceso: 02 Julio 2022].
- [2] O. Castillo, «10 casos de uso de un RPA en el entorno empresarial,» telcel, [En línea]. Available: <https://www.telcel.com/empresas/tendencias/notas/casos-de-uso-de-rpa-en-entorno-empresarial>. [Último acceso: 22 Abr. 2023].
- [3] itpeers, «5 áreas donde la automatización será esencial,» itpeers, 01 Oct. 2019. [En línea]. Available: <https://itpeers.com/es/2019/10/01/5-areas-onde-a-automatizacao-sera-imprescindivel/>. [Último acceso: 22 Abr. 2023].
- [4] CFB Bots, «The Difference between Robotic Process Automation and Artificial Intelligence,» CFB Bots, 20 Abr. 2018. [En línea]. Available: <https://cfb-bots.medium.com/the-difference-between-robotic-process-automation-and-artificial-intelligence-4a71b4834788#:~:text=Save-,The%20Difference%20between%20Robotic%20Process%20Automation%20and%20Artificial%20Intelligence,of%20human%20intelligence%20.> [Último acceso: 22 Abr. 2023].
- [5] Google Cloud, «Artificial intelligence (AI) vs. machine learning (ML),» Google Cloud, [En línea]. Available: <https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning#:~:text=While%20artificial%20intelligence%20encompasses%20the,accurate%20results%20by%20identifying%20patterns..> [Último acceso: 22 Abr. 2023].
- [6] M. Martí, «Web Scraping Introducción y Herramientas,» [En línea]. Available: <https://sitelabs.es/web-scraping-introduccion-y-herramientas/>. [Último acceso: 22 Junio 2022].
- [7] Red Hat, «El concepto de automatización,» Red Hat, 21 Marzo 2018. [En línea]. Available: <https://www.redhat.com/es/topics/automation>. [Último acceso: Junio 22 2022].
- [8] Gobierno de Canarias, «Ejemplificación. La robótica,» Gobierno de Canarias, [En línea]. Available: http://www3.gobiernodecanarias.org/aciisi/cienciasmc/web/u9/ejemplificacion_u9.html. [Último acceso: 22 Junio 2022].
- [9] Netec, «¿Qué es programación?,» Netec, [En línea]. Available: <https://www.netec.com/que-es-programacion>. [Último acceso: 22 Junio 2022].
- [10] S. Chaves, Formadores IT, 21 Feb. 2022. [En línea]. Available: <http://www.formadoresit.es/el-lenguaje-python-y-sus-principales-librerias/>. [Último acceso: 13 May. 2023].
- [11] R. Zambrano, «Cómo hacer Web Scraping con Selenium,» OpenWebinars, 21 Jun. 2019. [En línea]. Available: <https://openwebinars.net/blog/como-hacer-web-scraping-con-selenium/#:~:text=Qu%C3%A9%20es%20Selenium,como%20si%20fuese%20un%20humano..> [Último acceso: 13 May. 2023].
- [12] InnovaciónDigital360, «Análisis de datos: Concepto, metodología y técnicas,» InnovaciónDigital360, 07 Sep. 2022. [En línea]. Available: <https://www.innovaciondigital360.com/big-data/analisis-de-datos-tecnicas-y-metodologias-para-la-aplicacion-de-analytics/>. [Último acceso: 13 May. 2023].
- [13] Google Cloud, «¿Qué es el aprendizaje automático?,» Google, [En línea]. Available: <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419>. [Último acceso: 13 May. 2023].
- [14] N. Stenhouse, «HabScrap: herramienta automatizada para la extracción de datos con web scraping,» 2017-18. [En línea]. Available: <https://dspace.uib.es/xmlui/handle/11201/151095>. [Último acceso: 22 Junio 2022].
- [15] J. Casas, «Diseño de tableros de control de indicadores para el análisis de la información acerca de la situación de los graduados de la Institución Universitaria Politécnico,» 2019. [En línea]. Available: <https://alejandria.poligran.edu.co/handle/10823/1626>. [Último acceso: 22 Junio 2022].
- [16] C. Pandita, «Web Scraping and Data Visualization for Online Education Platforms,» de *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Chennai, India, 2021.
- [17] E. M. J. Murumba, «Big Data Analytics in Higher Education: A Review,» 2017. [En línea]. Available: https://www.academia.edu/33600670/Big_Data_Analytics_in_Higher_Education_A_Review. [Último acceso: 04 Mar. 2023].
- [18] ScrapFly, «Web Scraping with Selenium and Python Tutorial + Example Project,» ScrapFly Blog, 10 Ene. 2022. [En línea]. Available: <https://scrapfly.io/blog/web-scraping-with-selenium-and-python/>. [Último acceso: 01 Abr. 2023].
- [19] IBM Cloud Team, IBM Cloud, «Python vs. R: What's the Difference?,» 23 Marzo 2021. [En línea]. Available: <https://www.ibm.com/cloud/blog/python-vs-r>. [Último acceso: 04 Marzo 2023].
- [20] O. Kulyk, «Puppeteer vs. Selenium - Which Is Better? + Bonus,» Scraping Ant, 04 Dic. 2022. [En línea]. Available:

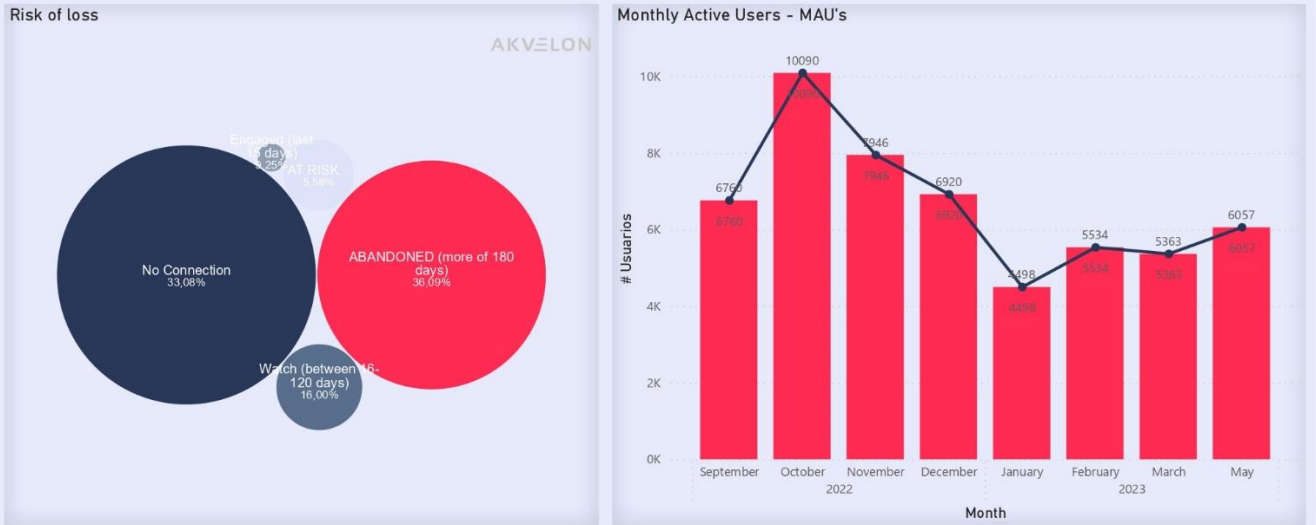
- <https://scrapingant.com/blog/puppeteer-vs-selenium>. [Último acceso: 04 Mar. 2023].
- [21] T. Vasilis, «Puppeteer vs. Selenium for automation,» 23 Feb. 2023. [En línea]. Available: <https://blog.apify.com/puppeteer-vs-selenium-for-automation/>. [Último acceso: 04 Mar. 2023].
- [22] MERKLE, «BrowserStack: testing multi-navegador al alcance de la web,» 01 Sep. 2020. [En línea]. Available: <https://www.merkle.com/es/es/blog/browserstack-testing-multi-navegador>. [Último acceso: 04 Mar. 2023].
- [23] Lambdatest, «Plataforma de Pruebas de automatización en línea,» Lambdatest, (s.f). [En línea]. Available: <https://www.lambdatest.com/intl/es/automation-testing>. [Último acceso: 04 Mar. 2023].
- [24] Microsoft, «Compare cloud storage pricing and plans,» Microsoft, (s.f). [En línea]. Available: <https://www.microsoft.com/en-us/microsoft-365/onedrive/compare-onedrive-plans?activetab=tab%3aprimary1>. [Último acceso: 04 Mar. 2023].
- [25] Google One, «One membership to get more out of Google,» Google, (s.f). [En línea]. Available: <https://one.google.com/about>. [Último acceso: 04 Mar. 2023].
- [26] Visual Studio Code, «Python in Visual Studio Code,» (s.f). [En línea]. Available: <https://code.visualstudio.com/docs/languages/python>. [Último acceso: 04 Mar. 2023].
- [27] PyCharm, «The Python IDE,» Ene. 2023. [En línea]. Available: <https://www.jetbrains.com/pycharm/>. [Último acceso: 04 Mar. 2023].
- [28] A. Biswal, «Power BI Vs Tableau: Difference and Comparison,» 30 Ene. 2023. [En línea]. Available: <https://www.simplilearn.com/tutorials/power-bi-tutorial/power-bi-vs-tableau#:~:text=Power%20BI%20is%20considered%20to,popular%20with%20seasoned%20data%20analysts>. [Último acceso: 04 Mar. 2023].
- [29] R. Kaur, «Power BI vs. Tableau vs. Qlik Sense: Which BI Tool is the Winner?,» (s.f). [En línea]. Available: <https://www.selecthub.com/business-intelligence/tableau-vs-qlikview-vs-microsoft-power-bi/>. [Último acceso: 04 Mar. 2023].
- [30] C. Tardi, «Monthly Active Users (MAU): Definition and How the Indicator Is Used,» Investopedia, 20 Jul. 2022. [En línea]. Available: [https://www.investopedia.com/terms/m/monthly-active-user-mau.asp#:~:text=Monthly%20active%20users%20\(MAU\)%20is,%2C%20email%20address%2C%20or%20username..](https://www.investopedia.com/terms/m/monthly-active-user-mau.asp#:~:text=Monthly%20active%20users%20(MAU)%20is,%2C%20email%20address%2C%20or%20username..) [Último acceso: 025 May. 2023].
- [31] A. Martinez, «Análisis,» Concepto Definición, 02 Agosto 2021. [En línea]. Available: <https://conceptodefinition.de/analisis/>. [Último acceso: 22 Junio 2022].
- [32] A. G. J. Pérez, «Definición de Analítico,» Definición De, 2015. [En línea]. Available: <https://definicion.de/analitico/>. [Último acceso: 22 Junio 2022].
- [33] A. G. J. Pérez, «Definición de visual,» Definición De, 2015. [En línea]. Available: <https://definicion.de/visual/>. [Último acceso: 22 Junio 2022].
- [34] M. M. J. Pérez, «Definición de datos,» Definición De, 2009. [En línea]. Available: <https://definicion.de/datos/>. [Último acceso: 22 Junio 2022].
- [35] S. Cao, «Research on Python Data Visualization,» 2021. [En línea]. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/1757/1/012122>. [Último acceso: 22 Junio 2022].
- [36] V. D. a. G. A. Dronova, «Python as an automation tool in IS. Protecting,» *Python as an automation tool in IS. Protecting*, 2022.
- [37] X. M. D. Roperro, «Proyecto BI para el análisis del proceso de captación de alumnos en centros educativos,» Enero 2021. [En línea]. Available: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/126486/6/droperroaTFG0121memoria.pdf>. [Último acceso: 22 Junio 2022].
- [38] Griky, 2022. [En línea]. Available: <https://www.griky.co/>. [Último acceso: 02 Julio 2022].
- [39] A. Purdila, «Visual Studio Code: My New Favorite Code Editor,» 26 Ene. 2017. [En línea]. Available: <https://webdesign.tutsplus.com/tutorials/visual-studio-code-my-new-favorite-code-editor--cms-28012>. [Último acceso: 04 Mar. 2023].
- [40] M. López, «Creación de una plataforma tecnológica para el análisis y generación de conocimiento como apoyo en la toma de decisiones sobre indicadores académicos de la facultad de Ciencias en la Epoch,» Septiembre 2018. [En línea]. Available: <https://repositorio.pucesa.edu.ec/handle/123456789/2480>. [Último acceso: 22 Junio 2022].
- [41] C. Albon, «Automated web scraping with Python,» Packt Publishing Ltd, (s.f). [En línea]. [Último acceso: 04 Mar. 2023].
- [42] K. Bryson, «RPA for beginners: a practical guide to automating repetitive tasks,» Apress, 2019. [En línea].
- [43] J. D. T. B. & T. B. Kelleher, «Data science an introduction,» CRC Press, 2018. [En línea].

- [44] P. Kulkarni, «Data visualization with Power BI cookbook: creating compelling interactive dashboards,» Packt Publishing Ltd, 2019. [En línea].

ANEXOS

Anexo 1 Dashboard- Proyecto-griky





Email	Course Id	Course Title	Progress	Status	Enrollment Date	Completion Date
000241901@certus.edu.pe	course-v1:ClassRUN+Test_001+2021_08	Curso online de Big Data aplicada a los negocios		Not Started	6/16/2022	
000241901@certus.edu.pe	course-v1:ClassRUN+FIC11_TEC+2021_08	Fundamentos de la infraestructura Cloud: Foundations		Not Started	6/16/2022	
000241901@certus.edu.pe	course-v1:ClassRUN+RPC22_HB+2021_08	Resolución de Problemas Complejos		Not Started	6/16/2022	
000523black@hotmail.es	course-v1:eanx+bienveanx+2020	Bienvenido a EANX	0.89	In Progress	3/20/2021	
000523black@hotmail.es	course-v1:eanx+oraprefec+2020	Oratoria y Presentaciones Efectivas	0.25	In Progress	3/24/2021	
002103572@certus.edu.pe	course-v1:ClassRUN+emprendimiento+2022-2	De la Idea a la acción		Not Started	9/1/2022	
009angel0090@gmail.com	course-v1:UGC+Pro_Escrituraparamediosdecomunicacion+2022-1	Escritura para Medios de Comunicación	0.06	In Progress	2/21/2023	
009angel0090@gmail.com	course-v1:UGC+Pro_Imagedigital+2022-1	Imagen Digital	0.97	In Progress	10/18/2022	
009angel0090@gmail.com	course-v1:UGC+Pro_Plannedesarrollodemediosdigitales+2022-1	Plan de desarrollo de medios digitales	0.87	In Progress	4/16/2023	
009angel0090@gmail.com	course-v1:UGC+Pro_Tecnicasdeexpresionalyescrituraenesp+2022_1	Técnicas de Expresión Oral y Escrita en Español	0.90	In Progress	10/6/2022	
009angel0090@gmail.com	course-v1:UGC+Pro_Teoriadela comunicacion+2022_1	Teoría de la Comunicación	0.92	In Progress	9/5/2022	
01071990nj@gmail.com	course-v1:demo4+produccion1+2030	Básicos Colombia	1.00	Successful	3/31/2022	9/26/2022
01071990nj@gmail.com	course-v1:demo4+cominducces+2020_RFCIE	Comunidad Inducción Estudiantes	1.00	Successful	3/7/2022	3/29/2022
01071990nj@gmail.com	course-v1:RobinFood+Culturacalidadsmile+2021	Cultura de Calidad - Smile - Colombia	1.00	Successful	3/31/2022	3/31/2022
01071990nj@gmail.com	course-v1:RobinFood+formatoscalidad+2022	Formatos Calidad Colombia	1.00	Successful	8/9/2022	9/22/2022
01071990nj@gmail.com	course-v1:RobinFood+sst1+2021	Inducción Seguridad y Salud en el Trabajo	1.00	Successful	4/4/2022	4/4/2022
01071990nj@gmail.com	course-v1:RobinFood+MUJ-V2+2021	No cursar - PMC 2	1.00	Successful	5/5/2022	5/6/2022
01071990nj@gmail.com	course-v1:demo4+Onboarding+2021_01	Onboarding	1.00	Successful	4/4/2022	4/4/2022
01071990nj@gmail.com	course-v1:RobinFood+BR1+2021	Onboarding Brasil	0.06	In Progress	3/31/2022	
01071990nj@gmail.com	course-v1:RobinFood+BPM2022mod1y2+2022	Plan de capacitación continuo y permanente BPM - Módulo 1 y 2 - 2022 - Colombia	1.00	Successful	3/9/2022	3/23/2022
01071990nj@gmail.com	course-v1:RobinFood+BPM2022mod3y4col+2022	Plan de capacitación continuo y permanente BPM -	1.00	Successful	4/5/2022	4/6/2022

Anexo 2 get_Progreso_Interno

```

let
    Source = (instancia as text) =>
let
    //instancia = "Autonoma",
    Source = SharePoint.Files("https://grikyco136.sharepoint.com/sites/GrikyDashboard", [ApiVersion = 15]),
    #"Filtered Rows" = Table.SelectRows(Source, each ([Folder Path] =
"https://grikyco136.sharepoint.com/sites/GrikyDashboard/Documentos compartidos/Reporting/"&instancia&"/Profesionales/")),
    #"Removed Other Columns" = Table.SelectColumns(#"Filtered Rows",{"Name", "Content"}),
    #"Split Column by Delimiter" = Table.SplitColumn(#"Removed Other Columns", "Name", Splitter.SplitTextByDelimiter("_",
QuoteStyle.None), {"Name.1", "Name.2", "Name.3"}),
    #"Changed Type" = Table.TransformColumnTypes(#"Split Column by Delimiter",{{"Name.1", type text}, {"Name.2", type
text}, {"Name.3", type text}}),
    #"Filtered Rows1" = Table.SelectRows(#"Changed Type", each ([Name.3] <> null)),
    #"Split Column by Delimiter1" = Table.SplitColumn(#"Filtered Rows1", "Name.3", Splitter.SplitTextByDelimiter(".",
QuoteStyle.None), {"Name.3.1", "Name.3.2"}),
    #"Changed Type1" = Table.TransformColumnTypes(#"Split Column by Delimiter1",{{"Name.3.1", type date}, {"Name.3.2",
type text}}),
    #"Removed Errors" = Table.RemoveRowsWithErrors(#"Changed Type1", {"Name.3.1"}),
    #"Filtered Rows2" = Table.SelectRows(#"Removed Errors", let latest = List.Max(#"Removed Errors"[Name.3.1]) in each
[Name.3.1] = latest),
    Content = #"Filtered Rows2"{0}[Content],
    #"Imported Excel Workbook" = Excel.Workbook(Content),
    #"Removed Other Columns1" = Table.SelectColumns(#"Imported Excel Workbook",{ "Data" }),
    Data = #"Removed Other Columns1"{0}[Data],
    #"Promoted Headers" = Table.PromoteHeaders(#"Data", [PromoteAllScalars=true]),
    #"Removed Duplicates" = Table.Distinct(#"Promoted Headers")
in
    #"Removed Duplicates"
in
    Source

```

Anexo 3 get_Usuarios_Interno

```

let
    Source = (instancia as text) =>
let

```

```
//instancia = "Classrun",
Source = SharePoint.Files("https://grikyco136.sharepoint.com/sites/GrikyDashboard", [ApiVersion = 15]),
#"Filtered Rows" = Table.SelectRows(Source, each ([Folder Path] =
"https://grikyco136.sharepoint.com/sites/GrikyDashboard/Documentos compartidos/Reporting/"&instancia&"/Usuarios/")),
#"Removed Other Columns" = Table.SelectColumns("#Filtered Rows",{"Name", "Content"}),
#"Split Column by Delimiter" = Table.SplitColumn("#Removed Other Columns", "Name", Splitter.SplitTextByDelimiter("_",
QuoteStyle.None), {"Name.1", "Name.2", "Name.3"}),
#"Changed Type" = Table.TransformColumnTypes("#Split Column by Delimiter",{{"Name.1", type text}, {"Name.2", type
text}, {"Name.3", type text}}),
#"Filtered Rows1" = Table.SelectRows("#Changed Type", each ([Name.3] <> null)),
#"Split Column by Delimiter1" = Table.SplitColumn("#Filtered Rows1", "Name.3", Splitter.SplitTextByDelimiter(".",
QuoteStyle.None), {"Name.3.1", "Name.3.2"}),
#"Changed Type1" = Table.TransformColumnTypes("#Split Column by Delimiter1",{{"Name.3.1", type date}, {"Name.3.2",
type text}}),
#"Removed Errors" = Table.RemoveRowsWithErrors("#Changed Type1", {"Name.3.1"}),
#"Filtered Rows2" = Table.SelectRows("#Removed Errors", let latest = List.Max("#Removed Errors"[Name.3.1]) in each
[Name.3.1] = latest),
Content = #"Filtered Rows2"{0}[Content],
#"Imported Excel Workbook" = Excel.Workbook(Content),
#"Removed Other Columns1" = Table.SelectColumns("#Imported Excel Workbook",{"Data"}),
Data = #"Removed Other Columns1"{0}[Data],
#"Promoted Headers" = Table.PromoteHeaders(Data, [PromoteAllScalars=true])
in
#"Promoted Headers"
in
Source
```

Anexo 4 Progreso_Interno

```
let
Source = instancias,
#"Converted to Table" = Table.FromList(Source, Splitter.SplitByNothing(), null, null, ExtraValues.Error),
#"Invoked Custom Function" = Table.AddColumn("#Converted to Table", "Interno", each get_Progreso_Interno([Column1])),
#"Removed Errors1" = Table.RemoveRowsWithErrors("#Invoked Custom Function", {"Interno"}),
#"Expanded Interno" = Table.ExpandTableColumn("#Removed Errors1", "Interno", {"Course Title", "Course Id", "Name",
"Email", "Username", "Gender", "Phone Number", "Identification Number", "Position Held", "Job Description", "Status",
"Progress", "Current Score", "Badges", "Posts", "Total Time Spent", "Enrollment Date", "Completion Date", "Country", "City",
```

```
"Company", "Job Code", "What topics would you like to learn or deepen?"), {"Course Title", "Course Id", "Name", "Email",
"Username", "Gender", "Phone Number", "Identification Number", "Position Held", "Job Description", "Status", "Progress",
"Current Score", "Badges", "Posts", "Total Time Spent", "Enrollment Date", "Completion Date", "Country", "City", "Company",
"Job Code", "What topics would you like to learn or deepen?")),
#"Removed Columns" = Table.RemoveColumns("#Expanded Interno", {"Username", "Gender", "Phone Number",
"Identification Number", "Position Held", "Job Description", "Country", "City", "Company", "Job Code", "What topics would
you like to learn or deepen?")),
#"Renamed Columns" = Table.RenameColumns("#Removed Columns", {"Column1", "Instancia"}),
#"Changed Type with Locale" = Table.TransformColumnTypes("#Renamed Columns", {"Completion Date", type date}),
"es-CO"),
#"Changed Type with Locale1" = Table.TransformColumnTypes("#Changed Type with Locale", {"Enrollment Date", type
date}), "es-CO"),
#"Split Column by Delimiter" = Table.SplitColumn("#Changed Type with Locale1", "Total Time Spent",
Splitter.SplitTextByDelimiter(":", QuoteStyle.Csv), {"Total Time Spent.1", "Total Time Spent.2", "Total Time Spent.3"}),
#"Changed Type" = Table.TransformColumnTypes("#Split Column by Delimiter", {"Total Time Spent.1", Int64.Type},
{"Total Time Spent.2", Int64.Type}, {"Total Time Spent.3", Int64.Type}),
#"Added Custom" = Table.AddColumn("#Changed Type", "Total Time Spent", each [Total Time Spent.1]+[Total Time
Spent.2]/60 + [Total Time Spent.3]/3600),
#"Removed Columns1" = Table.RemoveColumns("#Added Custom", {"Total Time Spent.1", "Total Time Spent.2", "Total
Time Spent.3"}),
#"Changed Type1" = Table.TransformColumnTypes("#Removed Columns1", {"Total Time Spent", type number}, {"Current
Score", type number}, {"Progress", type number}),
#"Added Custom1" = Table.AddColumn("#Changed Type1", "Catalogo", each "Interno"),
#"Removed Errors" = Table.RemoveRowsWithErrors("#Added Custom1", {"Instancia", "Course Title", "Course Id", "Name",
"Email", "Status", "Progress", "Current Score", "Badges", "Posts", "Enrollment Date", "Completion Date", "Total Time Spent",
"Catalogo"}),
#"Uppercased Text" = Table.TransformColumns("#Removed Errors", {"Instancia", Text.Upper, type text})
```

in

```
#"Uppercased Text"
```

Anexo 5 Usuarios_Interno

let

```
Source = instancias,
```

```
#"Converted to Table" = Table.FromList(Source, Splitter.SplitByNothing(), null, null, ExtraValues.Error),
#"Invoked Custom Function" = Table.AddColumn("#Converted to Table", "Interno", each get_Usuarios_Interno([Column1])),
#"Removed Errors" = Table.RemoveRowsWithErrors("#Invoked Custom Function", {"Interno"}),
#"Expanded Interno" = Table.ExpandTableColumn("#Removed Errors", "Interno", {"Name", "Email", "Username", "Gender",
"Phone Number", "Identification Number", "Position Held", "Job Description", "Enrollments", "Finished", "Failed", "In
Progress", "Not Started", "Average Final Score", "Badges", "Posts", "Total Time Spent", "Last Login", "Country", "City",
"Company", "Job Code", "What topics would you like to learn or deepen?"}, {"Name", "Email", "Username", "Gender", "Phone
Number", "Identification Number", "Position Held", "Job Description", "Enrollments", "Finished", "Failed", "In Progress", "Not
Started", "Average Final Score", "Badges", "Posts", "Total Time Spent", "Last Login", "Country", "City", "Company", "Job
Code", "What topics would you like to learn or deepen?")),
#"Renamed Columns" = Table.RenameColumns("#Expanded Interno", {"Column1", "Instancia"}),
#"Removed Other Columns" = Table.SelectColumns("#Renamed Columns", {"Instancia", "Name", "Email", "Enrollments",
"Finished", "Failed", "In Progress", "Not Started", "Average Final Score", "Badges", "Posts", "Total Time Spent", "Last Login"}),
#"Uppercased Text" = Table.TransformColumns("#Removed Other Columns", {"Instancia", Text.Upper, type text}),
#"Renamed Columns1" = Table.RenameColumns("#Uppercased Text", {"Instancia", "Cliente"}),
#"Changed Type" = Table.TransformColumnTypes("#Renamed Columns1", {"Total Time Spent", type text}),
#"Split Column by Delimiter" = Table.SplitColumn("#Changed Type", "Total Time Spent", Splitter.SplitTextByDelimiter(":",
QuoteStyle.Csv), {"Total Time Spent.1", "Total Time Spent.2", "Total Time Spent.3"}),
```

```
#"Changed Type1" = Table.TransformColumnTypes("#Split Column by Delimiter",{{"Total Time Spent.1", Int64.Type}, {"Total Time Spent.2", Int64.Type}, {"Total Time Spent.3", Int64.Type}}),
```

```
#"Added Custom1" = Table.AddColumn("#Changed Type1", "Total Time Spent", each [Total Time Spent.1]+ [Total Time Spent.2]/60 + [Total Time Spent.3]/3600),
```

```
#"Removed Columns" = Table.RemoveColumns("#Added Custom1",{"Total Time Spent.1", "Total Time Spent.2", "Total Time Spent.3"}),
```

```
#"Changed Type with Locale" = Table.TransformColumnTypes("#Removed Columns", {{"Last Login", type date}}, "es-CO"),
```

```
#"Changed Type2" = Table.TransformColumnTypes("#Changed Type with Locale",{{"Total Time Spent", type number}}),
```

```
#"Lowercased Text" = Table.TransformColumns("#Changed Type2",{{"Email", Text.Lower, type text}}),
```

```
#"Removed Duplicates" = Table.Distinct("#Lowercased Text", {"Email"})
```

in

```
#"Removed Duplicates"
```