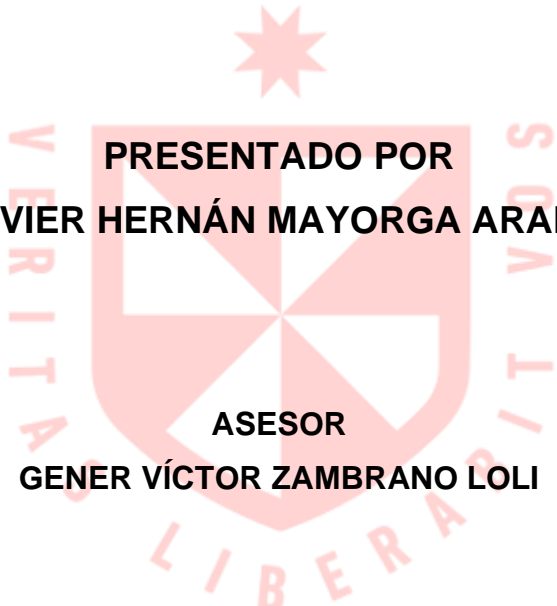


FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE COMPUTACIÓN Y SISTEMAS

**SISTEMA DE AUTOMATIZACIÓN DE DESCARGA DE  
DATOS (RPA), PROCESAMIENTO (ETL) Y ANÁLISIS DE  
DATOS (DATAWAREHOUSE) PARA EL TRASPASO DE  
CLIENTES A CLARO PERÚ**



PRESENTADO POR  
JAVIER HERNÁN MAYORGA ARANA

ASESOR  
GENER VÍCTOR ZAMBRANO LOLI

TRABAJO DE SUFICIENCIA PROFESIONAL  
PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE  
COMPUTACIÓN Y SISTEMAS

LIMA – PERÚ  
2024



**CC BY-NC-ND**

**Reconocimiento – No comercial – Sin obra derivada**

El autor sólo permite que se pueda descargar esta obra y compartirla con otras personas, siempre que se reconozca su autoría, pero no se puede cambiar de ninguna manera ni se puede utilizar comercialmente.

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

## Reporte de Turnitin

### Reporte de similitud

NOMBRE DEL TRABAJO

**Sistema de Automatización de Descarga de Datos (RPA), Procesamiento (ETL) y Análisis de Datos (Dataw**

AUTOR

**JAVIER HERNÁN MAYORGA ARANA**

RECuento de PALABRAS

**13828 Words**

RECuento DE CARACTERES

**81692 Characters**

RECuento DE PÁGINAS

**100 Pages**

TAMAÑO DEL ARCHIVO

**3.8MB**

FECHA DE ENTREGA

**Jun 27, 2024 8:24 AM GMT-5**

FECHA DEL INFORME

**Jun 27, 2024 8:25 AM GMT-5**

#### ● 5% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 3% Base de datos de Internet
- Base de datos de Crossref
- 4% Base de datos de trabajos entregados
- 0% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

#### ● Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 8 palabras)

## RESUMEN

La primera etapa del informe corresponde a la recolección de información del sitio web [ChecaTusLineas](#) de OSIPTEL, mediante técnicas automatizadas de Scraping, y se recopilan todos los datos de forma eficiente y coherente, eliminando la necesidad de obtener la data manualmente.

Posteriormente, se emplea un Automatización Robótica (RPA) para automatizar tareas repetitivas involucradas en el procesamiento y manipulación de datos. Los robots de RPA están configurados para realizar tareas como limpieza, normalización y transformación de datos, lo que garantiza la precisión y reduce el tiempo de procesamiento.

Python servirá como lenguaje de programación principal para extraer y procesar datos. Utilizaremos bibliotecas como Pandas para una manipulación de datos eficiente. Estos conjuntos de datos procesados se integrarán mediante herramientas ETL (Extract, Transform, Load), lo que garantizará una transición sin problemas desde diversas fuentes a un almacén o repositorio de datos centralizado. Los procesos ETL abarcan pasos esenciales manipulación de la información, lo que simplifica la armonización y estandarización de la información para su análisis futuro.

Finalmente, al aprovechar las herramientas y técnicas de inteligencia empresarial (BI), los datos integrados se visualizan y analizan para obtener información útil y rápida que servirá para ofrecer paquetes de minutos y promociones para que los clientes de otros operadores telefónicos se traspasen a Claro. Los paneles, informes y análisis interactivos permiten explorar tendencias, identificar patrones y descubrir oportunidades ocultas en beneficio de Net Axxes.

Palabras Claves: Python, Selenium, Pandas, Amazon Glue, Snowflake, ETL y BI.

## **ABSTRACT**

The first stage of the report corresponds to the extraction of data from the OSIPTEL - [ChecaTusLineas](#) website, using automated Scraping techniques, and all data is collected in an efficient and coherent manner, eliminating the need to obtain the data manually.

Subsequently, RPA is employed to automate repetitive tasks involved in data processing and manipulation. RPA bots are configured to perform tasks such as data cleansing, normalization, and transformation, ensuring accuracy and reducing processing time.

Python will serve as the primary programming language for extracting and processing data. We will utilize libraries such as Pandas for efficient data manipulation. These processed data sets will be integrated using ETL (Extract, Transform, Load) tools, ensuring a smooth transition from diverse sources to a centralized data warehouse or repository. ETL processes encompass essential information handling steps, which facilitate the harmonization and standardization of data for further analysis.

Finally, by leveraging business intelligence (BI) tools and techniques, the integrated data is visualized and analyzed to obtain useful and quick information that will serve to offer bundled minutes and promotions for customers of other telephone operators to transfer to Claro. Dashboards, reports, and interactive analytics allow you to explore trends, identify patterns, and uncover hidden opportunities for the benefit of Net Axxes.

Keywords: Python, Selenium, Pandas, Amazon Glue, Snowflake, ETL, and BI.