**"Año del Fortalecimiento de la Soberanía Nacional”**

UNIVERSIDAD DE PIURA

Logotipo

Descripción generada automáticamente

**FACULTAD DE INGENIERÍA**

ANÁLISIS DE DATOS CON PYTHON

GRUPO 08

**“TRABAJO 1”**

**DOCENTE:** Ing. Pedro Rotta Saavedra

**INTEGRANTES:**

* + - * García Chinguel, Emersson
      * More Sernaque, Jhon Darwin
      * Roque de la Cruz, Pedro Ignacio
      * Vilca Aguilar, Luis Ángel

**INFORME FINAL**

Piura, 23 de enero de 2021

Tabla de contenido

[1 INTRODUCCIÓN 3](#_Toc93860297)

[2 PLANTEAMIENTO DEL TRABAJO 3](#_Toc93860298)

[3 ANÁLISIS DEL SISTEMA 4](#_Toc93860299)

[4 DESARROLLO DEL SISTEMA 5](#_Toc93860300)

[4.1 Ingresar datos a la base de datos y que se actualicen. 5](#_Toc93860301)

[4.2 Filtrar datos por medio de características. 6](#_Toc93860302)

[4.3 Solicitar variables estadísticas de los datos. 8](#_Toc93860303)

[4.4 Solicitar gráficos de análisis entre algunas variables. 8](#_Toc93860304)

[4.5 Generar nuevas características para los datos. (Nuevas columnas) 10](#_Toc93860305)

[5 CONCLUSIONES 12](#_Toc93860306)

[6 ANEXOS 13](#_Toc93860307)

# INTRODUCCIÓN

Este informe está destinado a abordar el desarrollo de nuestro trabajo realizado durante el curso de “Análisis de datos con Python Nivel 1”, por lo cual escogimos realizar el trabajo del tipo 01, que incluye planteamiento de trabajo, análisis del sistema, desarrollo de este y conclusiones.

En el planteamiento del problema se describe el alcance que abarcará el programa junto con las herramientas a utilizar; en el análisis del sistema se explican lo que debe realizar el programa con la ayuda de un diagrama. En el desarrollo, se explicarán las funcionalidades de cada parte del código. Finalmente, las conclusiones, que mostrarán lo aprendido en el curso y en este trabajo.

# PLANTEAMIENTO DEL TRABAJO

Se pide realizar un programa de sistematización para análisis de datos. El programa debe permitir, mediante la terminal, ingresar datos, guardarlos en una base de datos en Excel, y permitir mediante opciones que el usuario pueda hacer estas acciones:

* Ingresar datos a la base de datos y que se actualicen.
* Filtrar datos por medio de características.
* Solicitar gráficos de análisis entre algunas variables.
* Solicitar variables estadísticas de los datos.
* Generar nuevas características para los datos. (Nuevas columnas)

Se pueden usar las siguientes librerías para su ejecución: Numpy, Pandas, Matplotlib, Seaborn y Plotly



Figura 1: Nombres de las librerías a utilizar en el sistema

Fuente: Elaboración propia

# ANÁLISIS DEL SISTEMA

De acuerdo al criterio descrito anteriormente, se requiere que el sistema/programa, se ejecute por consola y no mediante una interfaz gráfica, por lo cual se deberá establecer un loop principal para que el usuario pueda navegar por todas las funciones del programa, y escoger aquella función que necesite, desde el ingreso de datos hasta la solicitud de gráficas de análisis, siempre y cuando el usuario decida continuar, se le seguirán mostrando los paneles para elegir la función requerida, *se explicará de una mejor forma en el apartado siguiente.*

Imagen que contiene reloj, dibujo, señal

Descripción generada automáticamente

Figura 2: Diagrama del sistema

Fuente: Elaboración propia

# DESARROLLO DEL SISTEMA

## Ingresar datos a la base de datos y que se actualicen.

Texto

Descripción generada automáticamente

Figura 3: Función para ingresar datos

Fuente: Elaboración propia

La función para ingresar datos inicia pidiendo que se agregue el dato que se encuentra en la primera columna del dataframe, así como el formato de este, seguido, con un input se procede a agregar la información en la lista “info[ ]”; el proceso se repite hasta la última columna. Luego se procede a crear un diccionario con la lista de las columnas y la información ingresada con el fin de agregarlo al dataframe original y actualizarlo, finalmente se pregunta al usuario si desea agregar otro elemento y dependiendo de su elección se continúa o sale del programa, asimismo se reinicia la lista “info[ ]”. A continuación, se presenta un ejemplo.

Imagen que contiene Texto

Descripción generada automáticamente

Figura 4: Ejemplo de agregar datos en una tabla

Fuente: Elaboración propia

En esta imagen se aprecia como se agregaron dos datos al dataframe original. Se ha agregado una fila adicional. Cabe recalcar que se han incluido números flotantes en las columnas de cuartos y baños con el objetivo de poder hacer cálculos con sus respectivos datos.

## Filtrar datos por medio de características.

En cuanto al filtrado de datos, la función solicitara el nombre de una columna al usuario, la cual se validará si existe dentro del dataframe, si es que no existe, simplemente la función terminara su ejecución, por otro lado, si es que existe, se le pedirá el parámetro de entrada y su tipo de dato, para que la función pueda filtrar dentro de todos los registros del dataframe, y devolver los valores encontrados al usuario, y si es que no existen, se le devolverá un mensaje de todas formas.

Texto

Descripción generada automáticamente

Figura 5: Función para filtrar datos

Fuente: Elaboración propia

Texto

Descripción generada automáticamente

Figura 6: Ejemplo de filtrar los datos

Fuente: Elaboración propia

## Solicitar variables estadísticas de los datos.

Texto

Descripción generada automáticamente

Figura 7: Función para solicitar datos estadísticos

Fuente: Elaboración propia

En el caso de la función de datos estadísticos, al iniciar se solicita el nombre de la columna que se desea analizar, la cual se agrega a una lista “variable[ ]”. Después, el programa solicita al usuario si desea agregar más nombres de columnas, si ese es el caso se procede a seguir llenando la matriz hasta que el usuario ya no desee, finalmente se presenta la información con los datos estadísticos más importantes. Para ello, se presenta un ejemplo.

Texto

Descripción generada automáticamente

Figura 8: Ejemplo para hallar los datos estadísticos

Fuente: Elaboración propia

En la imagen se presenta dos columnas seleccionadas del dataframe con sus respectivas variables estadísticas: cantidad de datos, media, desviación estándar, valor mínimo, valor máximo y los percentiles.

## Solicitar gráficos de análisis entre algunas variables.

En cuanto a la función de para solicitar gráficas de análisis, el programa solo pedirá al usuario que ingrese un tipo de gráfico disponible, ya sea del tipo **a**, **b** o **c**, si el usuario ingresa una opción fuera de rango, le saldrá un mensaje gracias a la función *chooseCollection (dato, colección**[ ])***,** que se encarga de revisar que cualquier parámetro exista dentro de una colección de opciones.

Una vez, se escoja un tipo de gráfico, se le solicitará al usuario los parámetros que requiere dicho gráfico, por ejemplo; ejeX, ejeY, hue (opcional), alpha (opcional), bins (opcional).

Cuanto el programa reciba todos los parámetros, usando la librería de matplotlib y seaborn, se muestran los gráficos, como se muestra en la imagen de arriba.

Texto

Descripción generada automáticamente

Figura 9:Funión para elegir el tipo de gráfico

Fuente: Elaboración propia

**Gráfico

Descripción generada automáticamente con confianza media**

Figura 10: Gráfico de densidad

Fuente: Elaboración Propia



Figura 11: Función para la configuración de los gráficos

Fuente: Elaboración propia

## Generar nuevas características para los datos. (Nuevas columnas)

Para evaluar este requisito del programa se creó una función denominada *Agregar\_columna()* el cual cumple la función de agregar las columnas que el usuario desee a la tabla digital principal. Ver figura 12

A continuación, se muestra un ejemplo en la figura 13 donde se aprecia que se agregó la columna *Ubicación*, estos datos se agregaron aleatoriamente de algunos lugares de Piura. Se aprecia que dicha columna se ha agregado satisfactoriamente.

Texto

Descripción generada automáticamente

Figura 12: Función para agregar columnas en la tabla digital

Fuente: Elaboración propia

Pantalla de computadora con letras

Descripción generada automáticamente con confianza media

Figura 13: Tabla digital y una columna adicional ingresada por el usuario

Fuente: Elaboración propia

# CONCLUSIONES

* El Lenguaje de Programación Python, permite trabajar de una manera cada vez más sistematizada por disponer de una sintaxis clara y sencilla, en esto se puede combinar los diferentes paradigmas de programación de los que se dispone y por ende se consigue la reducción en el tiempo de trabajo.
* Sistematizar estas operaciones para el análisis de datos, conlleva cumplir con la sintaxis correcta, saber identificar el tipo de dato, sus funciones integradas, usar de manera adecuada sus librerías, etc. Esto para poner especial énfasis en saber qué librería, método, función, a utilizar con el fin de tener un programa muy eficiente.
* En los programas que se han escrito y ejecutado, se ha puesto hincapié en la manera distinta de organizar la información y en su representación jerárquico-relacional entre los datos que permiten realizar las tareas para los que se ha escrito, y las instrucciones necesarias para seleccionar registros y campos requeridos de una base de datos.

# ANEXOS

Texto

Descripción generada automáticamente

Figura 14: Ingreso de datos con la función ingresar\_datos()

Fuente: Elaboración propia

Texto

Descripción generada automáticamente

Figura 15: Filtrar datos con filtrarDataFrame(database\_df)

Fuente: Elaboración propia

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Figura 16: Respuesta del filtrado

Fuente: Elaboración propia

Interfaz de usuario gráfica

Descripción generada automáticamente

Figura 17: Gráfico generado con la función chooseChart()

Fuente: Elaboración propia

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Figura 18: Datos estadísticos generados por la función datos\_estadisticos()

Fuente: Elaboración propia

Texto

Descripción generada automáticamente

Figura 19: Agregar columnas con la función Agregar\_columna()

Fuente: Elaboración propia

Texto

Descripción generada automáticamente

Figura 20: Comprobación de que se agregó la columna

Fuente: Elaboración propia