

“Año del Fortalecimiento de la Soberanía Nacional”



Trabajo final de Análisis de datos con Python (Nivel 1)

Autores:

Chunga Castilla, José Augusto

Coello Mejía, Gustavo Adolfo

Huamán Paredes, María Sofía

Meléndez Rodríguez, Daniel Josué

Rivera Correa, César Paul

Asesor:

Ing. Pedro Rotta Saavedra

Piura 2022

CONTENIDO

INTRODUCCIÓN	3
ANÁLISIS DEL SISTEMA	4
EJEMPLOS AL CORRER EL SISTEMA	5
CONCLUSIONES	10

INTRODUCCIÓN

El manejo y análisis de bases de datos son considerados fundamentales para una buena organización industrial. A medida que las empresas se van haciendo más grandes en el mercado, la data con la que trabajan se va convirtiendo cada vez más tediosa. Lo cual, conlleva a problemas de ineficiencia e improductividad en los empleados. Para ello, hoy en día, los lenguajes de programación han facilitado el manejo de big data. Gracias a que su adaptabilidad y su capacidad de creación de funciones han otorgado a los trabajadores una herramienta capaz de mejorar la eficiencia en su producción, además de un buen manejo de la información a la hora de tomar decisiones.

Uno de los lenguajes de programación más conocido es Python. Este nos permite crear funciones y bucles que ayudan a optimizar el manejo de las bases de datos. En el presente informe, se desarrollará un sistema de programación el cual generará una mejora en el análisis de datos. La finalidad de dicho programa es el de facilitar al empleado la actualización de los datos, la filtración de dicha base y la de poder fabricar gráficas que sean de fácil interpretación para la toma de decisiones.

Las oportunidades de mejora con la programación son sustanciales. Para lo cual este ejemplo nos ayuda a observar las ventajas competitivas que puede tener una empresa con personas que manejen el lenguaje de programación de Python. A medida que se va requiriendo mayores funciones para optimizar, el código puede ser mejorado. Lo cual se puede inferir que el uso de un lenguaje de programación brinda una mejora constante para la empresa. Además, que facilita el trabajo de los trabajadores y ayuda a mejorar la productividad y la eficiencia. Ocasionando además una nueva cultura, trayendo consigo una constante optimización de todos los procesos de la empresa.

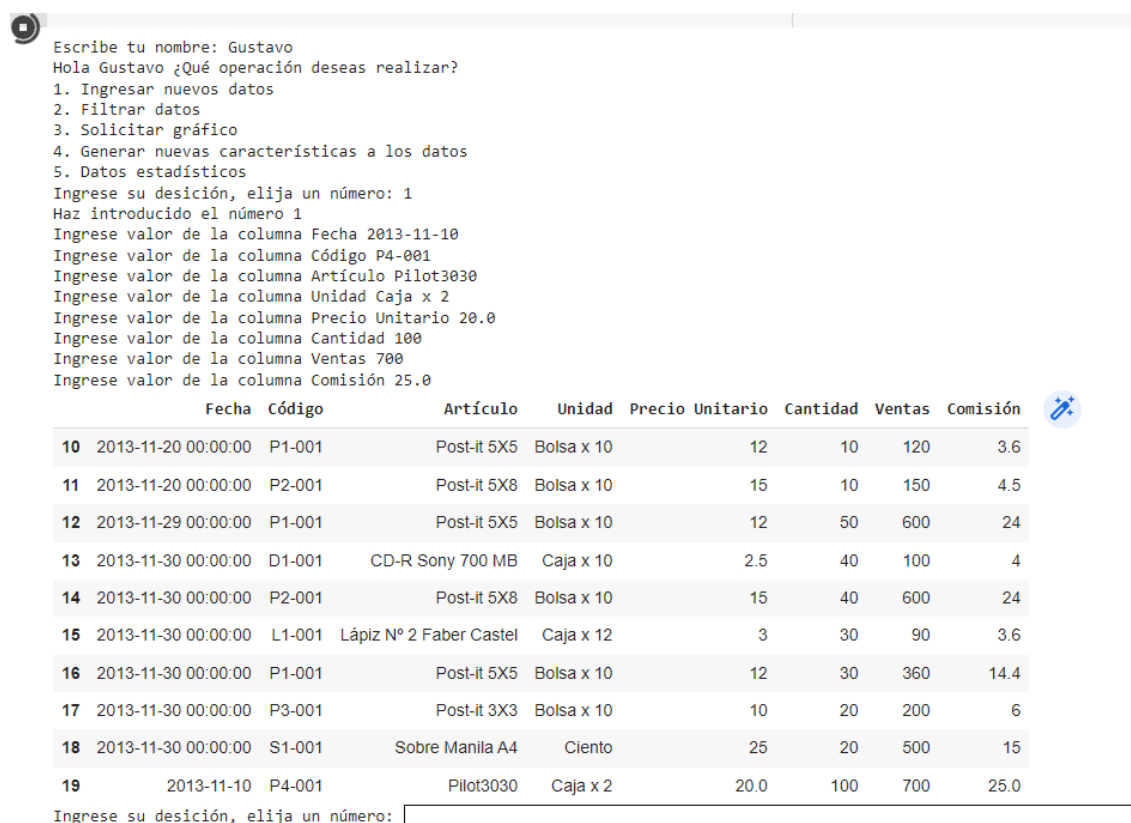
ANÁLISIS DEL SISTEMA

La creación de este código está basado en lo enseñado en clase, ya que el trabajo 1 está basado en algunos criterios, nosotros tomamos dichos criterios para poder darle al usuario lo que el desea elegir, y así sea más fácil el análisis de datos. Para poder definir el nombre del usuario y sea más fácil la interacción, utilizamos la función 'input' luego con un 'print' podemos imprimir en la pantalla un saludo, y se le da al usuario una serie de opciones para que pueda decidir. Colocamos una variable llamada "decisión" la cual llamamos a una función llamada 'int' la cual el programa interpreta que debe ir un número entero y así poder colocar cualquiera de las opciones mostradas, mediante bucles 'while' nosotros podemos poner una comparación, así cuando sea equivalente la decisión con el valor comparado se ejecuta una opción, en el caso del primer 'while' nos permite agregar una fila a nuestra base de datos existentes. Esta base de datos la hemos exportado de Excel y la podemos visualizar en el programa mediante '.head()' sin antes haber llamado a la librería de pandas como 'pd' definimos la función 'columnas(df)' para que luego mediante la función 'añadir_fila' podremos utilizarla, utilizamos en ella un 'for' como condicional y luego llamarlo. Mediante el comando 'df.append' agregamos la nueva fila que deseamos para poder visualizar los resultados utilizamos 'df.tail()' como no podíamos visualizar los datos aun así entonces colocamos 'display' lo cual nos muestra la fila que hemos agregado. Para la segunda opción de filtrar los datos igualmente utilizamos 'while', al llegar a este bucle nos damos cuenta que se debe hacer un filtrado, así que para que el cliente desee filtrar cualquier valor le damos la opción para que pueda filtrar cualquier característica de la base de datos esto lo hacemos mediante el comando '.describe()' . Luego para que el programa pueda comparar el valor colocado por el usuario utilizamos la herramienta '.loc' y para visualizar las coincidencias con el programa, lo visualizamos mediante 'print', y así con todas las opciones del filtrado. Para el tercer 'while' necesitamos importar 'matplotlib' para poder realizar gráficos como es un gráfico "n" dimensional definimos cuales irán en el gráfico y poder añadirlos mediante "list()" para luego definir las características del grafico mediante la opción 'plt.scatter' . Para el cuarto 'while' necesitamos añadir una nueva característica a la base de datos así que significa nuevas columnas para ello se nos ocurrió que el usuario pueda elegir entre 2 características que le gustaría agregar mediante la opción 'base_de_datos_df' la cual ya ha sido definido con anterioridad y mediante corchetes podremos colocar cada característica a su fila correspondiente. Finalmente para el ultimo 'while' queremos ver gráficos más enfocado con el tema estadístico, así que le decimos al usuario que características desea comparar definiendo una última decisión y mediante el bucle 'while' colocamos las opciones que escoja llamamos a 'matplotlib' ya que se verán nuevamente gráficos, y mediante la opción 'sns.relplot'

podemos definir los ejes del gráficos y de que base de datos se está tomando y para que esto se pueda visualizar utilizamos ‘plt.show’ para que esto no se haga un bucle infinito colocamos ‘break’ y finalmente colocamos un ‘else’ para que si no se escoge ninguna de las opciones mostradas se cierre el programa inmediatamente.

EJEMPLOS AL CORRER EL SISTEMA

Para la aplicación del programa, vamos a verificar que todas las opciones señaladas funcionan correctamente. Para ello, empezamos con la opción 1. La cual especifica que podemos ingresar nuevos datos a la base de datos dada. Empezamos señalando la opción que queremos realizar, en este caso es la 1. Y empieza la serie de preguntas que nos piden para añadir todos los detalles del nuevo dato. Esto lo podemos observar en la figura 1.



Escribe tu nombre: Gustavo
 Hola Gustavo ¿Qué operación deseas realizar?
 1. Ingresar nuevos datos
 2. Filtrar datos
 3. Solicitar gráfico
 4. Generar nuevas características a los datos
 5. Datos estadísticos
 Ingrese su desición, elija un número: 1
 Haz introducido el número 1
 Ingrese valor de la columna Fecha 2013-11-10
 Ingrese valor de la columna Código P4-001
 Ingrese valor de la columna Artículo Pilot3030
 Ingrese valor de la columna Unidad Caja x 2
 Ingrese valor de la columna Precio Unitario 20.0
 Ingrese valor de la columna Cantidad 100
 Ingrese valor de la columna Ventas 700
 Ingrese valor de la columna Comisión 25.0

	Fecha	Código	Artículo	Unidad	Precio Unitario	Cantidad	Ventas	Comisión
10	2013-11-20 00:00:00	P1-001	Post-it 5X5	Bolsa x 10	12	10	120	3.6
11	2013-11-20 00:00:00	P2-001	Post-it 5X8	Bolsa x 10	15	10	150	4.5
12	2013-11-29 00:00:00	P1-001	Post-it 5X5	Bolsa x 10	12	50	600	24
13	2013-11-30 00:00:00	D1-001	CD-R Sony 700 MB	Caja x 10	2.5	40	100	4
14	2013-11-30 00:00:00	P2-001	Post-it 5X8	Bolsa x 10	15	40	600	24
15	2013-11-30 00:00:00	L1-001	Lápiz N° 2 Faber Castel	Caja x 12	3	30	90	3.6
16	2013-11-30 00:00:00	P1-001	Post-it 5X5	Bolsa x 10	12	30	360	14.4
17	2013-11-30 00:00:00	P3-001	Post-it 3X3	Bolsa x 10	10	20	200	6
18	2013-11-30 00:00:00	S1-001	Sobre Manila A4	Ciento	25	20	500	15
19	2013-11-10	P4-001	Pilot3030	Caja x 2	20.0	100	700	25.0

Ingrese su desición, elija un número:

Figura 1

Después de observar que podemos añadir correctamente nuevos datos y además de actualizar constantemente nuestra base de datos. Entonces pasamos a verificar la segunda opción. La opción 2 nos da la capacidad de poder filtrar nuestra base de datos con respecto a sus características. Para lo cual, señalamos la opción 2 al programa y elegimos la característica que deseamos filtrar. Para este ejemplo, utilizamos la característica “Código”, la cual le corresponde la opción “b” en el programa. Después de señalar la opción anteriormente mencionada,

introducimos el código que queremos filtrar. Y como resultado observamos la base de datos filtrada con la característica dispuesta. Esto lo podemos observar en la figura 2.

```
Ingrese su desición, elija un número: 2
Haz introducido el número 2
¿Qué deseas filtrar?, Elige entre éstas opciones:
a. Fecha
b. Código
c. Artículo
d. Unidad
e. Precio Unitario
f. Cantidad
g. Ventas
h. Comisión
Ingrese su decisión, elija una letra: b
Elija un código: P1-001
```

	Fecha	Código	Artículo	...	Cantidad	Ventas	Comisión
1	2013-11-10	P1-001	Post-it 5X5	...	40	480	19.2
10	2013-11-20	P1-001	Post-it 5X5	...	10	120	3.6
12	2013-11-29	P1-001	Post-it 5X5	...	50	600	24.0
16	2013-11-30	P1-001	Post-it 5X5	...	30	360	14.4

[4 rows x 8 columns]

Figura 2

A continuación, vamos a verificar la funcionalidad de la opción 3. Dicha función consiste en realizar gráficos con la data de la base de datos definida. Para ello, llamamos al programa que realiza la opción 3. Para esta opción al tener data numérica y data descriptiva, se utiliza un gráfico n dimensional. El gráfico se muestra al terminar el programa ya que se termina cuando solicitamos la acción requerida. Esto lo podemos observar en la figura 3.

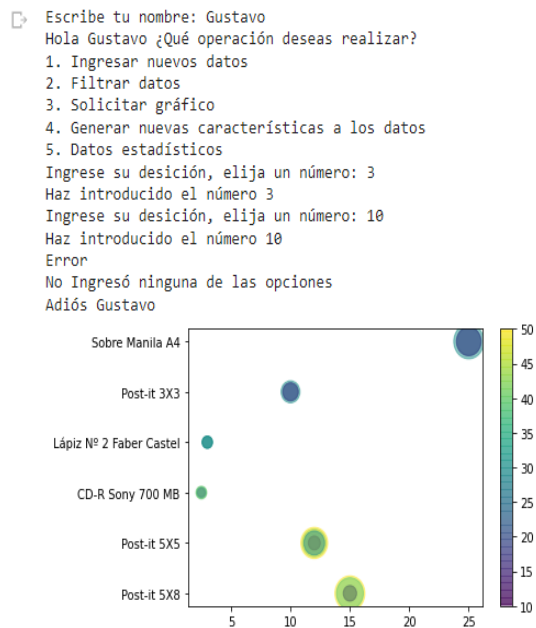


Figura 3

Para la opción 4, llamamos al programa de nuevo para elegir la opción 4 que queremos realizar. Esta opción se encarga de agregar una característica más a la base de datos definida. Esto ayuda mucho a mejorar la especificación de la data para próximas tomas de decisiones. Para este ejemplo, tenemos dos opciones para agregar. Una es el lugar de distribución y otra es el nombre del distribuidor. Nosotros elegimos que el lugar de distribución es una característica importante para tomar en cuenta, por lo cual la agregamos. Esto lo podemos observar en la figura 4.

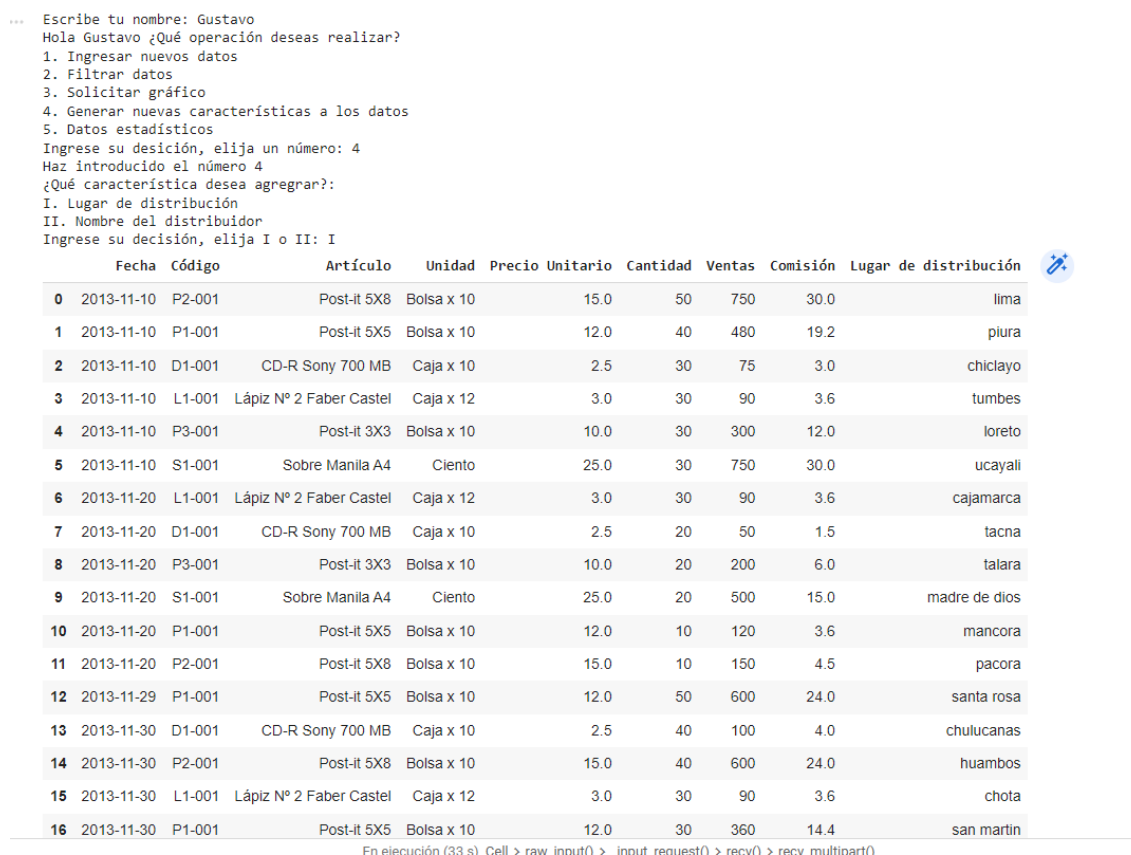


Figura 4

Y, por último, para comprobar a en toda su conformidad el programa construido. Se procederá a realizar la opción 5. Dicha opción consiste otorgar datos estadísticos de la base datos. Cabe resaltar que la base de datos solo tiene 4 características numéricas, por lo cual, los datos estadísticos serán en base a esas características. Para este ejemplo, observaremos la relación del precio unitario y las ventas. El programa al recibir las indicaciones nos da el gráfico estadístico de las dos variables elegidas, a través de la cual, podemos observar su comportamiento y ver cuál es mejor precio que tiene un mayor impacto en las ventas. Esto se puede observar en la figura 5.


```

Escribe tu nombre: Gustavo
Hola Gustavo ¿Qué operación deseas realizar?
1. Ingresar nuevos datos
2. Filtrar datos
3. Solicitar gráfico
4. Generar nuevas características a los datos
5. Datos estadísticos
Ingrese su desición, elija un número: 5
Haz introducido el número 5
¿Qué gráfico estadístico deseas visualizar?:
A. Precio Unitario vs Ventas
B. Precio Unitario vs Cantidad
C. Ventas vs Comisión
Elija una Opción: A

```

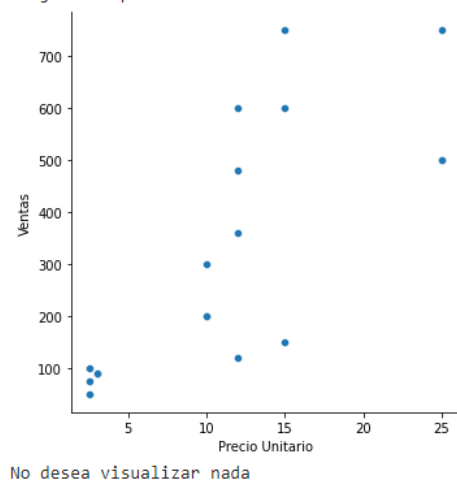


Figura 5

Con esto se completa el ejemplo. En este ejemplo hemos podido observar la funcionalidad del programa para cada una de las opciones. Y además la facilidad y la eficiencia que genera este código brinda una mejor productividad para los empleados de la empresa.

CONCLUSIONES

- i. El lenguaje con el que cuenta Python nos permite crear programas como el desarrollado en este trabajo, analizar datos a profundidad, ordenarlos, análisis estadísticos con gráficos y demás. Representamos el flujo del programa, usando una tabla digital, además con apoyo del bucle 'while' para que el usuario pueda elegir, apoyados en el uso de otros comandos que se dirigen por filtraciones mediante las decisiones del usuario, de esta forma en el programa cumple con el objetivo que el usuario pueda usar un programa que funcione hasta que él decida cerrarlo. Lo representado nos proporciona un programa sofisticado y útil para amplios ámbitos
- ii. El programa Google Colab es una excelente herramienta para programar y ejecutar códigos Python. También, nos proporciona paquetes preinstalados y listos para usar tales como las librerías numpy, pandas, entre otros. Además, al ejecutarse y almacenarse en la nube se pueden hacer trabajos colaborativos y también importar archivos desde GitHub.
- iii. Otra función muy importante de Google Colab, que nos ha ayudado a lo largo este trabajo ha sido la ayuda propia de la interfaz a la cual se accede escribiendo un comando seguido de "??" y ejecutándolo para que nos brinde información y el uso correcto ese comando.
- iv. Siempre se debe tener en cuenta la indexación porque nos puede llegar a generar muchos errores a lo largo del código. Si no indexamos de manera correcta o si no nos percatamos hasta que parte del código queremos que cumpla un while, por ejemplo, puede que nos quedemos atrapados en un loop al ejecutar el programa.
- v. Se debe tener en cuenta la instalación e importación correcta de librerías antes de ejecutar el código Python pese a que, si no se hace con anterioridad, no podrá ejecutarse el código en ningún momento y nos arrojaría un error.
- vi. El trabajo nos ha ayudado a involucrarnos en la solución de un problema real, aplicando todos los conceptos vistos en clase, esto es muy importante para fortalecer nuestras habilidades como futuros ingenieros, además nos ha incentivado en seguir estudiando la ciencia de los datos.