

Python for Data Science

Emtech Institute

Nombre: García Ruiz Andrea

Grupo: 2

Fecha de entrega: 06/09/2020

Índice

1. Introducción	2
2. Definición del código	2
3. Solución del problema	3
4. Conclusiones	4

Introducción

El presente proyecto forma parte del curso "Python for Data Science" impartido por Emtech.

El desarrollo del proyecto consta de la resolución de un caso ficticio, relacionado con una tienda de productos llamada Lifestore. La descripción del caso es la siguiente:

LifeStore es una tienda virtual que maneja una amplia gama de artículos, recientemente, la Gerencia de ventas, se percató que la empresa tiene una importante acumulación de inventario. Asimismo, se ha identificado una reducción en las búsquedas de un grupo importante de productos, lo que ha redundado en una disminución sustancial de sus ventas del último trimestre.

A lo largo del proyecto se trabajará en cumplir las especificaciones requeridas por la gerencia, lo cual ayudará a mejorar las estrategias de la tienda y mejorar sus ventas.

En los puntos subsecuentes, se comentará más a detalle el enfoque utilizado para la resolución del problema y se hablará un poco del código implementado.

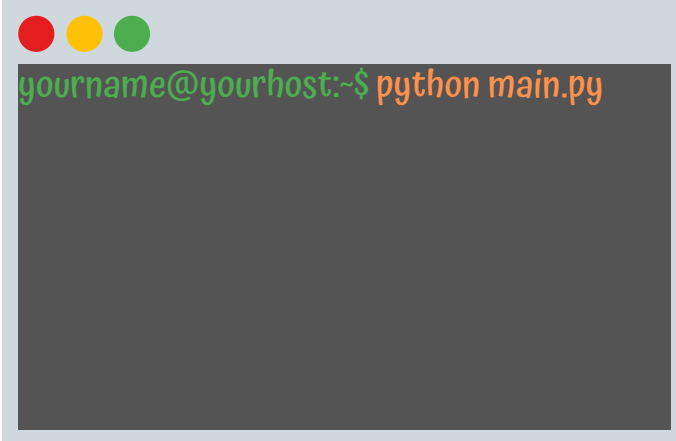
Definición del código

El código se encuentra alojado en un repositorio de Github, al que es posible acceder con la siguiente liga:

<https://github.com/andreaagr/CursoSantanderEmtech>

Todos los archivos referentes al proyecto se encuentran contenidos en una carpeta llamado "Proyecto 1". Dentro del proyecto tenemos diversos archivos que ayudan a generar los datos requeridos, sin embargo, el que reúne los datos de todos y permite desplegar la información de una forma amigable es `main.py`.

De manera que, para la ejecución del proyecto bastará usar la instrucción:

A terminal window with a light blue title bar and three colored window control buttons (red, yellow, green) on the left. The terminal has a dark gray background. The prompt 'yourname@yourhost:~\$' is in green, and the command 'python main.py' is in orange. The rest of the terminal area is empty.

```
yourname@yourhost:~$ python main.py
```

La función de los demás archivos se enuncia a continuación:

- **lifestore_file.py:** Almacena la información referente a productos, ventas y búsquedas de la tienda.
- **top50.py:** Agrupa los productos según la cantidad de ventas que han tenido y los ordena. De este archivo se puede llamar a la lista `quantities_top50` e imprimir únicamente los 50 productos con mayores ventas.
- **most_searched.py:** Realiza el mismo proceso que `top50.py`, sólo que los productos se agrupan a partir de su cantidad de búsquedas. La lista que se obtiene de este archivo es `quantities_searched`, a partir de la cual es posible imprimir los 100 productos con más búsquedas.
- **categories_sales.py:** Se genera una lista llamada `sales_by_category`, en la que las ventas realizadas están divididas por categoría, ya que la información se obtiene de `quantities_top50`, no es necesario ordenar el producto final. Usando la información de esta lista, se pueden obtener los 50 productos con menos ventas.
- **categories_searches.py:** El proceso que realiza este archivo es el mismo que el del punto anterior, sólo esta vez se dividen las búsquedas por categorías. El resultado obtenido será la lista `least_searched`.
- **rating.py:** En `rating.py` se genera el promedio de calificaciones obtenidas en reseñas para cada producto, así como, el porcentaje de devoluciones. Posteriormente en `best_products` los productos se agrupan dependiendo de su porcentaje de devolución y luego se ordenan a partir de su promedio de calificación.

De esta forma en main.py se pueden obtener los 20 productos con mejores reseñas y los 20 productos con peores reseñas.

- **statistics.py:** En este archivo las ventas se agrupan en sales_per_year, que almacena tanto el mes como el año en el que se realizó cada compra. También se calculan las ventas mensuales por año y se realiza el cálculo de ventas mensuales promedio, tomando en cuenta las ventas de los 3 años.

Todos los archivos se encuentran comentados para facilitar su entendimiento.

Solución del problema

A continuación se explica el enfoque para la resolución de los puntos requeridos por la gerencia:

- **Productos más vendidos/buscados:** Para ambos casos, primero debía realizarse un conteo de las ventas o búsquedas de cada producto y después ordenar esta información de mayor a menor.
- **Productos menos vendidos/buscados por categoría:** En mi caso contemple los productos con cero ventas y cero búsquedas, así que de existir estos deberán imprimirse.

Los datos se obtienen de las listas generadas en el punto 1, de ahí se sacan los registros requeridos (50 o 100) y se van acomodando en la categoría correspondiente. Así obtenemos la lista que se puede desplegar para conocimiento del usuario.

- **Productos con mejores/peores reseñas:** Ya que se debían tomar en cuenta tanto reseñas como devoluciones, decidí agrupar los registros dependiendo de la cantidad de devoluciones. Sin embargo, esto no supone el mejor enfoque, ya que puede que 2 objetos tengan la misma cantidad de devoluciones pero diferente cantidad de ventas.

Así que en lugar de sólo considerar la cantidad, saque la proporción de devoluciones (devoluciones/ventas). Posteriormente se agruparon los productos dependiendo su proporción de devoluciones y dentro de cada proporción se ordenaron de mayor a menor o viceversa.

- **Total de ingresos:** Para el total de ingresos se tomaron en cuenta las devoluciones. En el despliegue de la información se muestra: el total de ventas, el total en reembolsos y por último el total luego de restar estas 2 cantidades.
- **Ventas promedio mensuales:** Se obtuvieron sumando las ventas de cada mes en los 3 años y luego dividiéndolas entre esta misma cantidad. Si tuviéramos más datos, nos daría un mejor panorama de en que meses reportamos más ventas independientemente del año.
- **Total anual:** Ya que teníamos 3 años diferentes en los que se habían registrado ventas, se sacó la suma total de ventas de cada uno de ellos.
- **Meses con más ventas al año:** De igual forma al tener información de 3 años diferentes, la información se dividió por año, así que se muestran los meses con más ventas según el año.

Conclusión

Las acciones que debería llevar a cabo Lifestore para mejorar las ventas de sus productos, basándonos en la información recabada son:

- Colocar en un lugar más visible a todos aquellos productos que no hayan tenido búsquedas
- Analizar los costos de los productos que hayan tenido búsquedas pero no ventas, y de esta forma adaptarlos para incentivar su compra
- Hay meses en los que Lifestore no presentó ventas, una recomendación sería revisar factores externos que eviten que la gente realice compras en esos meses.
- Revisar aquellos productos con bajas reseñas y con alto índice de devolución.

Finalmente en lo que concierne al proyecto, revisé diversos conceptos y logré familiarizarme con el lenguaje de programación Python.

Además de analizar información que no está muy alejada de la realidad, se lograron sacar conclusiones para la futura toma de decisiones.