

Análisis de datos con Python

Alejandro Cimental Chávez
Universidad de Guanajuato
Herramientas Informáticas Y Gestión de la Información
(Dated: June 17, 2021)

Con la finalidad de aprender a utilizar python y algunas de sus librerías, se desarrollara un análisis de dos bases de datos. La primera será una base de datos proporcionada por el INE para predecir la cantidad de casillas necesarias en las próximas votaciones del país. Además el alumno podrá escoger una base de datos de su elección para realizar el análisis correspondiente y obtener resultados de interés.

I. INTRODUCCION

A. Instituto Nacional Electoral

El INE es uno de los órganos constitucionales autónomos de México, siendo este el encargado de regular los procesos electorales de todo el país. De igual forma el INE parte de otros procesos de participación ciudadana como consultas populares, además de ser la entidad que emite las credenciales para votar a través del registro federal de electores.[1]



FIG. 1. Instituto Nacional Electoral logo

B. Obtencion de datos electorales y análisis

A través del portal electrónico del INE se hizo uso de los datos abiertos del padrón electoral desde el mes de junio del año 2019 hasta diciembre del 2020 para realizar una predicción del padrón electoral para el mes de febrero del 2021 y así poder asignar la cantidad de casillas electorales que se tendría que emplear en las elecciones que se llevaran a cabo en el mes de Junio del 2021. Esta vez se utilizó Python y específicamente su librería pandas como herramienta de análisis de datos. Desde 2008 pandas ha sido una librería que ayuda al análisis y manipulación de datos ofreciendo al usuario estructuras de

datos y operaciones para manipular tablas numéricas y series de datos.



FIG. 2. Librería de software pandas

A diferencia de excel, Python ofrece la ventaja de poder trabajar con grandes bases de datos sin sufrir una demora considerable en sus operaciones. Con una simple línea de código es posible obtener precisamente los datos que se están buscando, acomodarlos en cierto orden, realizar operaciones matemáticas y mucho más. Si bien es cierto que se requiere un poco más de conocimiento para utilizarlo, para el usuario adecuado que busca analizar una gran base de datos es la herramienta ideal para el trabajo.

C. Base de datos de elección personal

Para el segundo proyecto de análisis de datos utilice una base de datos de tiendas departamentales en Estados Unidos donde se muestran los envíos, productos, clientes, ganancias, ventas, categoría de productos, etc. Se hizo un análisis de ventas en varias categorías para poder obtener información que pudiera beneficiar a la cadena de tiendas de alguna manera. Se pudo obtener en que estados las ventas eran mayores, que categoría era la que generaba más ganancias a la empresa al igual que los productos más vendidos. Con estos datos la empresa podría tomar decisiones que mejoren las ganancias y talvez dirigirse a un mercado más específico y tratar de evitar pérdidas con las categorías que su demanda es menor.

II. RESULTADOS

En la siguiente grafica podemos apreciar que de los años de venta en la lista, el 2014 fue el mejor superando los \$700000 dolares.

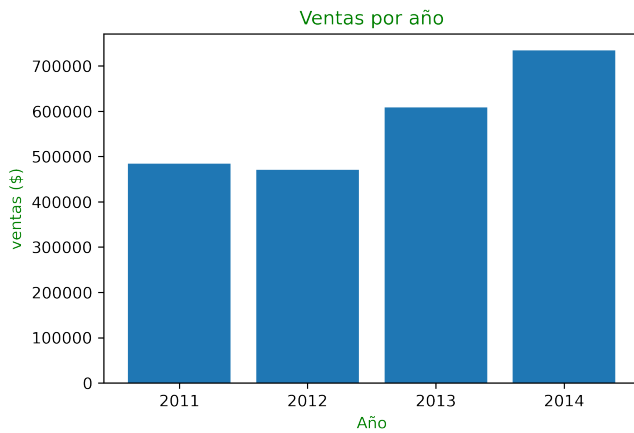


FIG. 3. Ventas por año

En el gráfico circular se puede ver el porcentaje de ventas por cada categoría. En general los artículos de oficina son mucho más baratos que un mueble o un equipo de computo lo que podría explicar porque los pedidos en esa categoría son mucho mayores.

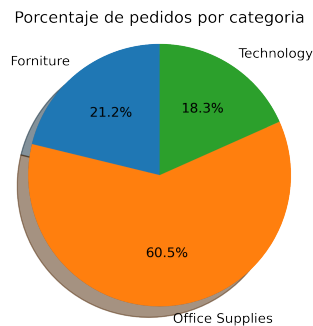


FIG. 4. Porcentaje de compras por categoría

Aún cuando la gran mayoría de pedidos son en el área de oficina, la mayor ganancia la genera la categoría de tecnología. Como ya se comentaba esto tiene que ver con los precios promedio de los artículos en cada una de las categorías.

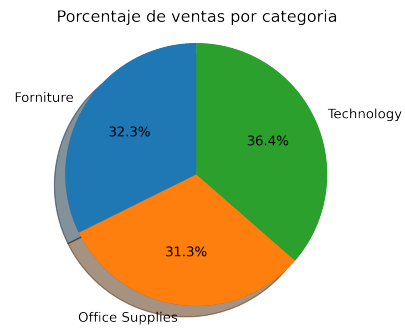


FIG. 5. Porcentaje de ganancias por categoría

El análisis también nos permite conocer cuál estado fue en la que se vendió más. Como era de esperarse, las ventas fueron superiores en los estados con una mayor población o una ciudad mayor como lo son Los Angeles, Manhattan o Dallas.

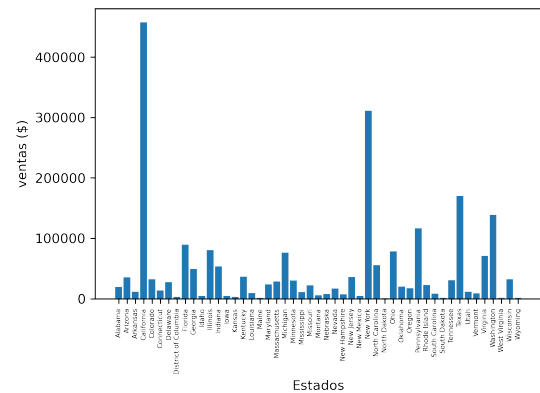


FIG. 6. Porcentaje de ganancias por categoría

III. CONCLUSIONES

Las graficas presentadas con anterioridad son solo algunas del total de graficas obtenidas con el análisis de datos realizado. Estos datos podrían permitir a la empresa tomar decisiones como en que mes deberían tener mas inventario, que clase de productos deberían invertir en comprar y almacenar o tal vez que clase de productos deberían dejar de vender ya que la demanda es poca y representan capital estancado. Por otro lado el análisis de datos es una practica necesaria en cualquier ámbito ya que arroja señales o guías de hacia donde se debería dirigir la operación para su mejoría.

[1] WIKIPEDIA *Instituto Nacional Electoral*. Fecha de consulta: Marzo, 2021. Obtenido desde: <https://en.wikipedia.org/wiki/InstitutoNacionalElectoral>

[2] GeeksforGeeks *Plot a pie chart in Python using Matplotlib*. Fecha de consulta: Junio, 2021. Obtenido desde: <https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/>

[3] YOUTUBE *Complete Python Pandas Data Science Tutorial*. Fecha de consulta: Junio, 2021. Obtenido desde:

<https://www.youtube.com/watch?v=vmEHCJofslg>