

Informe del proyecto de Introducción a la Ciencia de Datos

Fuentes de datos utilizadas:

- 30 Mipymes → carpeta mipymes
- Mercatoria → carpeta mercatoria
- El toque → carpeta el toque

Los datos de Mercatoria fueron recopilados mediante scrapping. Se normalizaron los precios, que originalmente se encontraban en USD, mediante una función que recibe el precio del dólar según los cálculos de ElToque y realiza la aritmética correspondiente para convertirlos a moneda nacional.

Todos los datos se encuentran en sus respectivas carpetas dentro de una carpeta llamada json.

De Mercatoria se recopilaron datos sobre el arroz, frijoles, y yogurt, incluyendo el precio, país de origen, peso/gramaje, y marca. Del yogurt y los frijoles también se recopila el tipo de frijol y sabor del yogurt.

En otra carpeta llamada notebook, se encuentra el jupyter notebook y una biblioteca llamada Tools, desde donde se importan las funciones necesarias para la realización del análisis.

Se utilizó la biblioteca plotly para la graficación.

En una primera instancia se analiza el promedio de los siguientes productos (arroz, frijoles, pollo, aceite, sal y el azúcar) productos básicos para cocinar, lo que nos permite notar la diferencia entre los precios y promediar su costo. La función utilizada para el análisis fue promedio_canasta_basica.

Una segunda gráfica muestra cuatro gráficos de donut que permite notar la diferencia entre mercatoria y las mipymes, subdividiendo entre arroz y frijoles, los países que más importan estos. Las funciones utilizadas fueron pais_moda_frijoles, pais_moda_arroz y mercatoria_moda. La lógica del programa es buscar primero por la marca del producto, se guarda en una lista y luego se comprueba si está o no y. Esto nos permite contar el país y asegurarnos de no contarlos más de una vez.

La siguiente gráfica también es una gráfica de barras pero es parte de este análisis, pero permite ilustrar mejor los países que tienen en conjunto, dividiendo una parte frijoles y la otra arroz. En el .py la biblioteca llamada graph, en la función llamada barr_com la primera parte del código realiza este proceso de escoger solo los países en común.

Se realizó una gráfica heatmap que muestra el precio del kilogramo de frijol por tipo de frijol y país exportador, la función correspondiente es heatmap_tabla en Tools. Se utilizan únicamente los datos de frijoles de Mercatoria.

Las dos últimas gráficas muestran datos referentes al yogurt en Mercatoria. La primera es una gráfica de frecuencias absolutas de los sabores de yogurt encontrados, la segunda un gráfico de pastel de los yogurts según el tipo (regular, probiótico, etc.). Los datos utilizados para crear esta gráfica se agregaron con la función yogurt_data de la biblioteca Tools, la cual itera sobre toda la lista de yogurts recopilados de Mercatoria y cuenta en un diccionario la cantidad de apariciones de cada sabor y tipo de yogurt.