



Tecnológico de Monterrey

Herramientas computacionales: el arte de la analítica

Entregable

Actividad Evaluable: Mapas de calor y boxplots


Presenta

Adrián Aguilar Sánchez	A01651592
Juan Francisco García Rodríguez	A01660981
César Andrés Ceballos Castillo	A01661893
Klaus Manuel Cedillo Arredondo	A01653257
Javier Valente Rodríguez	A01662693

Docente

Sergio Ruiz Loza

Para la realización de este documento y posterior solución a cuestionario, se empleó la base de datos llama “*avocado*”.

 Responde las siguientes preguntas:

1. ¿Hay alguna variable que no aporte información?

En un principio no teníamos conocimientos sobre las variables numéricas tales como:

- 4046
- 4225
- 4770

Sin embargo, luego de indagar un poco en internet, nos dimos cuenta que, se referían a tipos de aguacate, por lo que en realidad no carecen de relevancia, aunque en una primera impresión, posterior a la investigación, así lo consideramos.

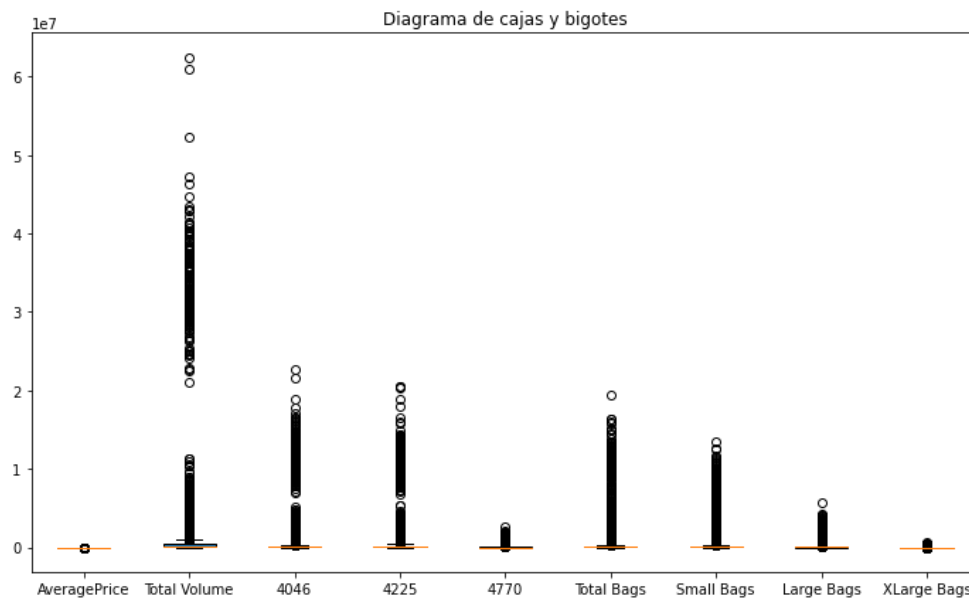
2. ¿Si tuvieras que eliminar variables, ¿cuáles quitarías y por qué?

Al momento de cargar la base de datos a *Python* con ayuda de *pandas*, se añade una columna llamada “*Unnamed: 0*”, la cual no consideramos relevante y por ello sería motivo de exclusión.

3. ¿Existen variables que tengan datos extraños?

No sabemos con exactitud a lo que se refieran los datos que se encuentran en la columna “*Unnamed: 0*”

4. Si comparas las variables, ¿todas están en rangos similares?



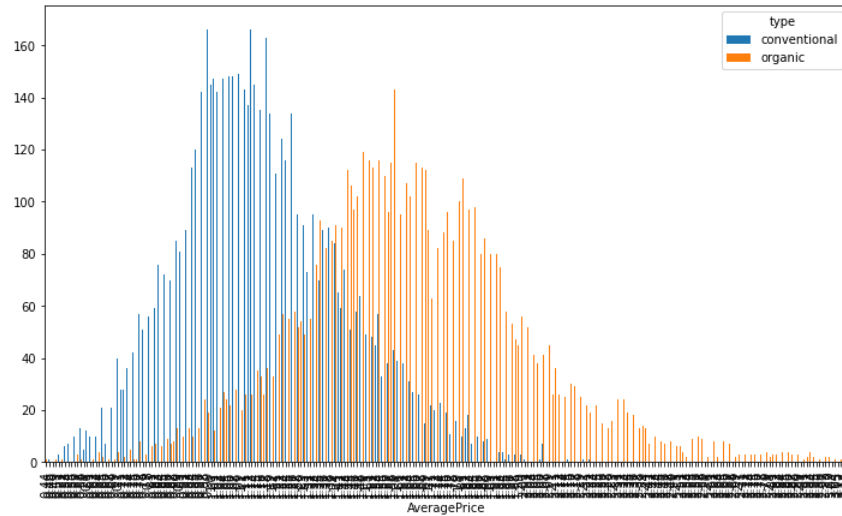
El *Average Price* y el *Total Volume* están fuera del rango en comparación con las demás variables

5. ¿Crees que esto afecte?

No realmente, pues a pesar de tener muchos datos sigue siendo analizable.

6. ¿Puedes encontrar grupos que se parezcan? ¿Qué grupos son estos?

Se realizó una gráfica comparando el tipo de aguacate con el precio, como la que muestra a continuación:



Inferimos que los precios aumentan para los aguacates orgánicos.