



Instituto Tecnológico y de Estudios Superiores de Monterrey

Herramientas Computacionales: El arte de la analítica (Grupo 121)

TC1002S.121

Actividad Evaluable: Obtención de estadísticas descriptivas

Manual de código

Profesor:
Luciano Garcia Bañuelos

Fecha: 28 de Octubre del 2021

Equipo 6 | Integrantes:

Fredy Yahir Canseco Santos	A01735589
Iván Guzmán Sosa	A01734212
José Ángel Ramírez Ramírez	A01735529

El programa es interactivo y se divide por celdas con “# %%”

Se importan las librerías de pandas, seaborn y matplotlib.pyplot como pd, sns y plt respectivamente. Estas librerías sirven para análisis de datos además de creación y diseño de gráficos.

Usando seaborn se asigna un tema a usar en cada gráfico que se cree, en el caso de este código se usa “white”.

Se crea una variable “df” que contiene los datos del archivo dividido por comas, “ulabox_orders_with_categories_partials_2017.csv”.

Usando pandas se crea una tabla solo con los datos relevantes de archivo de extensión csv, una que muestre la información de los datos en cada variable, una para cada uno de sus valores máximos y mínimos, también para el promedio la mediana y la desviación estándar.

Usando matplotlib se asigna un tamaño para una figura “f”, posteriormente con seaborn y pandas se crea una gráfica de mapa de calor entre el porcentaje por departamento de venta, el total de ítems y el porcentaje de descuento.

```
# %%
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

sns.set_theme (style= "white")

# %%
df= pd.read_csv
("ulabox_orders_with_categories_partials_2017.csv")

df[["total_items", "discount%", "Food%", "Fresh%", "Drinks%", "Home%", "
Beauty%", "Health%", "Baby%", "Pets%"]]

# %%
df[["total_items",
"discount%", "Food%", "Fresh%", "Drinks%", "Home%", "Beauty%", "Health%",
"Baby%", "Pets%"]].info()

# %%
df[["total_items",
"discount%", "Food%", "Fresh%", "Drinks%", "Home%", "Beauty%", "Health%",
"Baby%", "Pets%"]].max()

# %%
df[["total_items",
"discount%", "Food%", "Fresh%", "Drinks%", "Home%", "Beauty%", "Health%",
"Baby%", "Pets%"]].min()
```

```
# %%
df[["total_items",
    "discount%", "Food%", "Fresh%", "Drinks%", "Home%", "Beauty%", "Health%",
    "Baby%", "Pets%"]].mean()
# %%
df[["total_items",
    "discount%", "Food%", "Fresh%", "Drinks%", "Home%", "Beauty%", "Health%",
    "Baby%", "Pets%"]].median()

# %%
df[["total_items",
    "discount%", "Food%", "Fresh%", "Drinks%", "Home%", "Beauty%", "Health%",
    "Baby%", "Pets%"]].std()

# %%
f = plt.figure(figsize=(15, 8))
sns.heatmap(df.corr(), annot=True, linewidths=0.5)
```