Yash Sasane

Roll no : 681 PRN : 202201050052

Batch : F4

from google.colab import drive drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

import numpy as np import pandas as pd all\_data=pd.read\_csv('/content/drive/MyDrive/1686715083343\_all\_data (6).csv') all\_data.head()

**Order Quantity Price Order**

**Product Purchase Address**

**ID Ordered Each Date**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 176559.0 | Bose SoundSport Headphones | 1.0 | 99.99 | 04-07-2019 22:30 | 682 Chestnut St,  Boston, MA 02215 |
| **1** | 176560.0 | Google Phone | 1.0 | 600.00 | 04-12-2019 14:38 | 669 Spruce St, Los  Angeles, CA 90001 |
| **2** | 176560.0 | Wired  Headphones | 1.0 | 11.99 | 04-12-2019 14:38 | 669 Spruce St, Los  Angeles, CA 90001 |
| **3** | 176561.0 | Wired Headphones | 1.0 | 11.99 | 05/30/19  9:27 | 333 8th St, Los  Angeles, CA 90001 |

#FindNAN nan\_df=all\_data[all\_data.isna().any(axis=1)] display(nan\_df.head()) all\_data.shape

all\_data=all\_data.dropna(how='all') all\_data.head()

**Order ID Product Quantity Ordered Price Each Order Date Purchase Address**

**36** NaN NaN NaN NaN NaN NaN

**51**

NaN

NaN

NaN

NaN

NaN

NaN

**Order**

**ID**

**Product**

**Quantity**

**Ordered**

**Price**

**Each**

**Order Date**

**Purchase Address**

**0**

176559.0

Bose SoundSport

Headphones

1.0

99.99

04-07-2019

22:30

682

Chestnut St,

Boston, MA 02215

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 176560.0 | Google Phone | 1.0 | 600.00 | 04-12-2019 14:38 | 669 Spruce St, Los  Angeles, CA 90001 |

all\_data = all\_data[all\_data['Order Date'].str[0:2]!='Or'] print(all\_data)

all\_data['Quantity Ordered'] = pd.to\_numeric(all\_data['Quantity Ordered']) all\_data['Price Each'] = pd.to\_numeric(all\_data['Price Each'])

all\_data['Month'] = all\_data['Order Date'].str[0:2] all\_data['Month'] = all\_data['Month'].astype('int32') all\_data.head()

**Order Quantity Price Order Purchase**

**Product Month**

**ID Ordered Each Date Address**

|  |  |  |
| --- | --- | --- |
| Bose 04-07-   1. 176559.0 SoundSport 1.0 99.99 2019 from pandas.core.ops.methodsHeadphones import add\_flex\_arithmetic\_methods22:30 def get\_city(address):   return address.split(",")[1].strip(" ") 04-12-   1. 176560.0 Google Phone 1.0 600.00 2019 def get\_state(address): 14:38  return address.split(",")[2].split(" ")[1] 04-12-   Wired   1. 176560.0 1.0 11.99 2019 | 682 Chestnut St,  Boston, MA  02215  669 Spruce St,  Los Angeles, CA  90001  669 Spruce St, Los Angeles, CA | 4  4  4 |

all\_data['city'] = all\_data[Headphones"Purchase Address"].apply(lambda14:38 x:f"{get\_city(x90001)} {get\_state(x)}") all\_data.head()

333 8th St, Los

Wired 05/30/19

**3** 176561.0 1.0 11.99 Angeles, CA 5

**Order** Headphones**Quantity Price Order Purchase**9:27

**Product Month** 90001**city Sal**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  | 381 Wilson St | |  |
| **0** | 176559.0 | Bose  SoundSport Headphones | 1.0 | 99.99 | 04-072019  22:30 | 682  Chestnut  St,  Boston, MA 02215 | 4 | Boston  MA | 99. |
| **1** | 176560.0 | Google Phone | 1.0 | 600.00 | 04-122019  14:38 | 669  Spruce  St, Los  Angeles, CA 90001 | 4 | Los  Angeles  CA | 600. |
| **2** | 176560.0 | Wired  Headphones | 1.0 | 11.99 | 04-122019  14:38 | 669  Spruce  St, Los Angeles, | 4 | Los  Angeles  CA | 11. |

**ID Ordered Each Date Address**

all\_data['Sales'] = all\_data['Quantity Ordered'].astype('int') \* all\_data['Price Each'].astype('float') all\_data.groupby(['Month']).sum()

<ipython-input-11-788baa00bdec>:2: FutureWarning: The default value of numeric\_onl all\_data.groupby(['Month']).sum()

**Order ID Quantity Ordered Price Each**

**Sales**

**Month**

|  |  |  |  |
| --- | --- | --- | --- |
| **4** 7335546.0 | 123.0 | 885.80 | 1210.76 |
| **5** 353124.0 | 2.0 | 111.98 | 111.98 |
| **6** 184076.0 | 1.0 | 14.95 | 14.95 |
| **8** 726962.0 | 9.0 | 23.92 | 50.83 |
| **9** 2378802.0 | 17.0 | 591.44 | 616.62 |
| **10** 550924.0 | 11.0 | 10.67 | 39.69 |
| **11** 740314.0 | 19.0 | 13.66 | 65.31 |
| **12** 550635.0 | 17.0 | 8.97 | 50.83 |

Dummycity=all\_data.groupby(['city']) print(Dummycity)

#city\_max=all\_data.groupby(['City']).sum()

#print(max(city\_max))

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7fd864fee8f0>

product\_group = all\_data.groupby('Product') quantity\_ordered = product\_group.sum()['Quantity Ordered']

<ipython-input-17-4815a60ac30b>:2: FutureWarning: The default value of numeric\_only in DataFrameGroupBy.sum is deprecated. In a future quantity\_ordered = product\_group.sum()['Quantity Ordered']

[Colab paid products](https://colab.research.google.com/signup?utm_source=footer&utm_medium=link&utm_campaign=footer_links) - [Cancel contracts here](https://colab.research.google.com/cancel-subscription) completed at 10:05 PM

**

0

s