Name= Pawar Vaishnavi Subhash

Roll no=649

PRN 202201090095

Batch=F3

**ASSIGNMENT (4B)**

import numpy as np

import pandas as pd

all\_data=pd.read\_csv("/content/1686715083343\_all\_data.csv")

all\_data.head()

Order ID Product Quantity Ordered Price Each Order Date Purchase Address 0176559.0 Bose SoundSport Headphones 1.099.9904-07-2019 22:30682 Chestnut St, Boston, MA 022151176560.0 Google Phone1.0600.0004-12-2019 14:38669 Spruce St, Los Angeles, CA 900012176560.0Wired Headphones1.011.9904-12-2019 14:38669 Spruce St, Los Angeles, CA 900013176561.0Wired Headphones1.011.9905/30/19 9:27333 8th St, Los Angeles,

CA 900014176562.0USB-C Charging Cable1.011.9504/29/19 13:03381 Wilson St, Sat Francisco, CA 94016

all\_data.shape

(69, 6)

drop rows of nan

#find NAN

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()

all\_data.shape

| **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 |

(67, 6)

get rid of text order date column

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

print(all\_data)

Order ID Product Quantity Ordered Price Each \

0 176559.0 Bose SoundSport Headphones 1.0 99.99

1 176560.0 Google Phone 1.0 600.00

2 176560.0 Wired Headphones 1.0 11.99

3 176561.0 Wired Headphones 1.0 11.99

4 176562.0 USB-C Charging Cable 1.0 11.95

.. ... ... ... ...

64 259329.0 Lightning Charging Cable 1.0 14.95

65 259330.0 AA Batteries (4-pack) 2.0 3.84

66 259331.0 Apple Airpods Headphones 1.0 150.00

67 259332.0 Apple Airpods Headphones 1.0 150.00

68 259333.0 Bose SoundSport Headphones 1.0 99.99

Order Date Purchase Address

0 04-07-2019 22:30 682 Chestnut St, Boston, MA 02215

1 04-12-2019 14:38 669 Spruce St, Los Angeles, CA 90001

2 04-12-2019 14:38 669 Spruce St, Los Angeles, CA 90001

3 05/30/19 9:27 333 8th St, Los Angeles, CA 90001

4 04/29/19 13:03 381 Wilson St, San Francisco, CA 94016

.. ... ...

64 09-05-2019 19:00 480 Lincoln St, Atlanta, GA 30301

65 09/25/19 22:01 763 Washington St, Seattle, WA 98101

66 09/29/19 7:00 770 4th St, New York City, NY 10001

67 09/16/19 19:21 782 Lake St, Atlanta, GA 30301

68 09/19/19 18:03 347 Ridge St, San Francisco, CA 94016

[67 rows x 6 columns]

make column correct column

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

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

augument data with additional columns

add month column

[ ]

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

Order IDProductQuantity OrderedPrice EachOrder DatePurchase AddressMonth0176559.0Bose SoundSport Headphones1.099.9904-07-2019 22:30682 Chestnut St, Boston, MA 0221541176560.0Google Phone1.0600.0004-12-2019 14:38669 Spruce St, Los Angeles, CA 9000142176560.0Wired Headphones1.011.9904-12-2019 14:38669 Spruce St, Los Angeles, CA 9000143176561.0Wired Headphones1.011.9905/30/19 9:27333 8th St, Los Angeles, CA 9000154176562.0USB-C Charging Cable1.011.9504/29/19 13:03381 Wilson St, San Francisco, CA 940164

add city column

[ ]

def get\_city(address):  
  return address.split(",")[1].strip(" ")  
def get\_state(address):  
  return address.split(",")[2].strip(" ")[1]  
  
all\_data['City']=all\_data['Purchase Address'].apply(lambda x:f"{get\_city(x)} ({get\_state(x)})")  
all\_data.head()  
   #print(all\_data)

Order IDProductQuantity OrderedPrice EachOrder DatePurchase AddressMonthCity0176559.0Bose SoundSport Headphones1.099.9904-07-2019 22:30682 Chestnut St, Boston, MA 022154Boston (MA)1176560.0Google Phone1.0600.0004-12-2019 14:38669 Spruce St, Los Angeles, CA 900014Los Angeles (CA)2176560.0Wired Headphones1.011.9904-12-2019 14:38669 Spruce St, Los Angeles, CA 900014Los Angeles (CA)3176561.0Wired Headphones1.011.9905/30/19 9:27333 8th St, Los Angeles, CA 900015Los Angeles (CA)4176562.0USB-C Charging Cable1.011.9504/29/19 13:03381 Wilson St, San Francisco, CA 940164San Francisco (CA)

waht was the best month for sales?how much was earned that month?

all\_data['Sales']=all\_data('Quantity Ordered').astype('int')all\_data('Price Each').astype('Float')

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

#2)WHICH CITY SOLD THE MOST PRODUCT?

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

print(Dummycity)

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

#print(max(city\_max))

#waht products are most often sold together

df=all\_data[all\_data['Order ID'].duplicated(keep=False)]

df['Grouped']=df.groupby('Order ID')['Product'].transform(lambda x:','.join(x))

df2=df[['Order ID','Grouped']].drop\_duplicates()

print(df['Grouped'])

Google Phone, Wired Headphones

2 Google Phone ,Wired Headphones

Name: Grouped, dtype: object

from itertools import combinations

from collections import Counter

count=Counter()

for row in df2['Grouped']:

row\_list=row.split(',')

count.update(Counter(combinations(row\_list,2)))

for key,value in count.most\_common(10):

print(key,value)

o/p

('Google Phone', 'Wired Headphones') 1

1 product\_group=all\_data.groupby('Product')

2 quantity\_ordered = product\_group.sum()['Quantity Ordered']

1 print(quantity\_ordered)

Product AA Batteries (4-pack) 64.0

AAA Batteries (4-pack) 109.0

Apple Airpods Headphones 3.0

Bose SoundSport Headphones 3.0

Google Phone 1.0

Lightning Charging Cable 4.0

USB-C Charging Cable 8.0

Wired Headphones 7.0

Name: Quantity Ordered, dtype: float64

1 prices=all\_data.groupby('Product').mean()['Price Each']

1 print(prices)

Product

AA Batteries (4-pack) 3.84

AAA Batteries (4-pack) 2.99

Apple Airpods Headphones 150.00

Bose SoundSport Headphones 99.99

Google Phone 600.00

Lightning Charging Cable 14.95

USB-C Charging Cable 11.95

Wired Headphones 11.99

Name: Price Each, dtype: float64