

Name - Tejas Sonawane

PRN - 202201090073

Roll No. 559

Batch E 3

```
import numpy as np
import pandas as pd
```

```
all_data = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/1686715083343_all_data.csv")
all_data.head()
```

	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
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
4	176562.0	USB-C Charging Cable	1.0	11.95	04/29/19 13:03	381 Wilson St, San Francisco, CA 94016

```
#FIND MAN
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()
```

<bound method NDFrame.head of Address							
36	NaN	NaN	NaN	NaN	NaN	NaN	NaN
51	NaN	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	
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	

```
all_data = all_data[all_data['Order Date'].str[0:2]!='0']
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
	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		
	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		

[57 rows x 6 columns]

```
all_data['çşrant:zt y OrXžered'] = pd.to_nura•rt c (aJl_dat a['quantity Ordered'])
aJl_data['PrLce Gain'] = pa.co_nurærfc(a11_data$'nr1Ee GacA' )
```

```
all_data['Month') - pd.to_dztctimr(z1ł_dsfo['Opder Da-;']).dt.month
all_data.head()
```

	Order ID	Product	Quantity Onóereü	Price Each	Order Date	Puzelhmoc Eddrcæ*	Month
0	178559.0	Bosa SounäSpor Headpl-ones	1.0	99.99	04-07-2018 22:30	692 Ł:e*tnu• St. Eostor, L/A ?221fi	4
	178560.0	Google Phone	1.0	600.00	04-12-2018 14:38	669 Spruce St, Los Angeles, CA 90001	4
2	UB D.0	filed Haadphone6	1.0	11.9%	V-12-2018 11:38	669 Spruce St, Los Angeles, CA 90001	4

Aod City Column

```
def get_city(ass):
    return address.split(",")[1].strip(" ")

def get_state(address):
    return address.split(",")[3].strip(" ")[1]

all_data['city'] = all_data['address'].apply(lambda x: get_city(x) + "(" + get_state(x) + ")")
all_data.head()
```

	Order ID	Product	Quantity Qrdøred	Price Each	qrdşr Data	Purchas-a Addr•sx	month	C1ty
0	178559.0	Bose SounoSpcl  9eaMpnnne°	1.0	99.99	T1-07-Ju18 22:00	682 Mæstnut St, Bosçen, MA 02210	4	Dnstrn jA1
1	17a D.Łi	Goęqlę Phong	1.0	600.00	04-12-10 1ç.0R	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (A)
2	178560.0	Wired Headphones	1.0	11.99	04-12-10 14.5B	669 Spruce St, Los Angeles, CA 90001	4	Lœ Adalah IA]

- Data Exploration

Question 1 - What wss the best month for sales and how much was earnea in that month?

```
all_data['Sales'] = all_data['quantity ordered'].astype('int')*all_data['Price Eacs']*1.05
all_data.groupby(['Month']).sum()

*tpyt hgn• ingurr• 3:2•4cs4a 7 3Ss65d*-7 rutur*Ł4rnz ne. 7hr gefault value or nur• - ž c_only žn Data&nam•GroupBj
° 11_aat°.8nó•pü r(E'«anth' ] ]-&u• { j
```

	Order ID	quantity Ordered	Price Each	Sales
Month				
4	7335546.0	123.ç	685.80	121D.76
5	353124.0	2.0	111.9P	111.96
6	184076.0	1.0	14.95	14.95
8	728962.0	9.0	23.92	50.83
9	2378802.0	17.Łi	C:91d4	CiC62
10	550924.0	11.0	10.87	39.69
11	740314.0	19.Łi	13.66	65.31
12	550825.0	12.0	8.97	50.83

- Question 2 - Which city sold the most product?

PM 2:52 6/22/23

```

all_data.groupby('City').sum()
print all_data.groupby('City').sum()
all_data.groupby('City').sum()
# print(max(city_max))

<panaaz - ar 6.groupby - genc - c - DataFra>eGroutay ob;0ct - t g::re2C0bt szf0>

```

## П 4 which products are most often sold together?

```

cl* - at 1_dat a[at 1_data[ "Pn?er E0' j. ditpl1cated (keep- £aLse) j

```

```

%Referenced : https:// stackoverflow.com/question s / 2729 8J 78/ccasE atencte- strings- frgra- severa
d+{ 'Grouped ']= d-f.groupby( " Ehrder IB' )[' Product' ]. transforo(Laebda x' ", '. $oln {x})
df2=df[ ' Ordee IO' . " Grouped " ]].drop_dup11catPs{ j

```

```

1 Google Phone, Nired Headphone-8
2 Google Phone, H1red r'adphonPs
be: 6x>uped, dtype: object
<{python*input*S7*7905ebdbe5d9 }>: : Sett Ingu1thCopyHarnIng :
A value 1s Crying to be set on a copy of a slice £noe a aatarraae.
my uxlng .me[ixxInsezen, cnl_ lnaexer] = value tnsteas

```

```

Sae the cayeats in the dociaentation {££g
dE{ 'GrOuped' )* Bf.groupbye{ 'ordtr IO' }[' Product' ]. trans ora{ lmbda x: ' .'.koro(x)}

```

```

-Fnon ro1leEt1ons Import Courier

```

```

count = Counter()

for rop lo Qf2[ 'trouped' ]:
    xn_lls1 = rop.solit ',')
    count.update(Counter(combinations (row_list, 2)))

for key, value in count.lTost_£onna (J6): prJnt t££y, Yalue)

('Google Phone ', '41red Hendphc nes' ) 1

```

## g 3 which products sold the most? Why do u think it sold the most?

```

pn£xtuEt roup = all_data.groupby( 'Product ' )
quant 1ty_adercd - pro4uct-gr'ciup - sw*( j[" O+x«t1tx Ordered ' ]
pr*nt ( quant 1ty_or'be red)

```

```

As eaterLes fa-pack ed.e
AAA BatterGes (4- pack } 10S.6
Apple Airpods ftee-dphDries 3.e
Base Sourussport headphones 3.a Google
Pnone 1.a
Liyhtn Ing Charginy Cab1s 4.6
U5B-C Charging Cable 8.6
k/1red Headphone s 7.6
file* @ia«t1ty ordered, dtype: rLoat64
< {python-Input -26-ddc2efsf2ab> : 2* FutureAlarnJng' The de-FauJC value of- nuoea!c_on1y la oatafiraa1eGro<gz'By.sun l s depmc aCed. l n a fut
quant 1ty rdered = pr'ouaE top .sun( )[' *quant 1ty Ordered " ]

```

```

p-1at( quant 1ty_o rdered )

```

```

rproduct
Aa Batteries (4-pack ) 6s.d
AAbi Batterles (4- pack) 1B9.0
AppJe xi i"pods vauPhone s 3.0
bosc so**asPb*£ Hezapho let 3.0

Li/h/£/ Ch&rgtrg cabJe 4.0
"JsB-C Charktng CabLe 8.0
'Jlred f£eadphone s 7.e'
\luna. @.ant i ty Or4'lrrd, c1type - 71oa1£.4

```

```

prlces = all_data.groupby('Product').mean()['nricc Each']

```

```
AA Batteries (ü-pack)          1. B4
u eattcrics f4-pack            1. C'
Apple A1rpods rh•aopnones      L5o.ee
Bose sou£dsport l•leadphoaes   99. 99
GoogJe PTione                  óaB.6a
Li ght n }n/, / hary,} n;• Cah]ç 14. 95
Urr8 -C Ctisirging Ciihië      i1. 95
Nired rMadphone s              IL. JJ
de : Pat E e Eac h, dtype i float64
<1python-input-22-ff09c559I5e5>il: FutureWarning: the default value of nuneric_onlp in DataFrameGrguABY.*ean 1s deprecated. In a fu
```

