task-1

June 14, 2023

0.1 Data Preparation and Customer Analysis

[1]:
```
#import libraries
import pandas as pd
import datetime as date
import numpy as np
import seaborn as sns
%matplotlib inline
import matplotlib.pyplot as plt
from scipy import stats
```

[2]:
```
purchase_df= pd.read_csv(r"C:\Users\Preeti\Desktop\Certifications\Quantium Data␣
↪Analytics\QVI_purchase_behaviour.csv")
```

[3]:
```
purchase_df.head()
```

[3]:

	LYLTY_CARD_NBR	LIFESTAGE	PREMIUM_CUSTOMER
0	1000	YOUNG SINGLES/COUPLES	Premium
1	1002	YOUNG SINGLES/COUPLES	Mainstream
2	1003	YOUNG FAMILIES	Budget
3	1004	OLDER SINGLES/COUPLES	Mainstream
4	1005	MIDAGE SINGLES/COUPLES	Mainstream

[4]:
```
transaction_df= pd.read_excel(r"C:\Users\Preeti\Desktop\Certifications\Quantium␣
↪Data Analytics\QVI_transaction_data.xlsx")
```

[5]:
```
transaction_df.head()
```

[5]:

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	\
0	43390	1	1000	1	5	
1	43599	1	1307	348	66	
2	43605	1	1343	383	61	
3	43329	2	2373	974	69	
4	43330	2	2426	1038	108	

	PROD_NAME	PROD_QTY	TOT_SALES
0	Natural Chip Compny SeaSalt175g	2	6.0
1	CCs Nacho Cheese 175g	3	6.3

```
2    Smiths Crinkle Cut  Chips Chicken 170g         2        2.9
3    Smiths Chip Thinly  S/Cream&Onion 175g         5       15.0
4  Kettle Tortilla ChpsHny&Jlpno Chili 150g         3       13.8
```

0.1.1 Transaction Data

[6]:
```
#transforming date
transaction_df['DATE']= pd.to_datetime(transaction_df['DATE'], origin =␣
↪'1899-12-30', unit='D')
```

[7]:
```
transaction_df['PROD_NAME'].describe()
```

[7]:
```
count                                   264836
unique                                     114
top       Kettle Mozzarella   Basil & Pesto 175g
freq                                      3304
Name: PROD_NAME, dtype: object
```

[8]:
```
#finding the most frequent number
import collections
freq=collections.Counter([j  for s in transaction_df["PROD_NAME"] for j in s.
↪split()])
```

[9]:
```
# sorting in descending orders
fre = pd.DataFrame([freq.keys(), freq.values()], index=['Word','Frequency']).
↪transpose().sort_values(by='Frequency',ascending=False)
```

[10]:
```
fre=fre[[s[0] not in ['0','1','2','3','4','5','6','7','8','9','&'] for s in␣
↪fre['Word'] ]]
```

[11]:
```
fre.head()
```

[11]:
```
      Word Frequency
11   Chips     49770
16  Kettle     41288
8   Smiths     28860
29    Salt     27976
6   Cheese     27890
```

[12]:
```
#dropping salsa items
transaction_df.drop(transaction_df[[("Salsa" in s)  for s in␣
↪transaction_df['PROD_NAME']]].index,inplace=True)
```

[13]:
```
transaction_df[[('Salsa' in s) for s in transaction_df['PROD_NAME']]]
```

[13]:
```
Empty DataFrame
Columns: [DATE, STORE_NBR, LYLTY_CARD_NBR, TXN_ID, PROD_NBR, PROD_NAME,
```

```
PROD_QTY, TOT_SALES]
Index: []
```

[14]:
```
transaction_df.describe()
```

[14]:
```
            STORE_NBR  LYLTY_CARD_NBR        TXN_ID       PROD_NBR  \
count  246742.000000    2.467420e+05  2.467420e+05  246742.000000
mean      135.051098    1.355310e+05  1.351311e+05      56.351789
std        76.787096    8.071528e+04  7.814772e+04      33.695428
min         1.000000    1.000000e+03  1.000000e+00       1.000000
25%        70.000000    7.001500e+04  6.756925e+04      26.000000
50%       130.000000    1.303670e+05  1.351830e+05      53.000000
75%       203.000000    2.030840e+05  2.026538e+05      87.000000
max       272.000000    2.373711e+06  2.415841e+06     114.000000

            PROD_QTY      TOT_SALES
count  246742.000000  246742.000000
mean        1.908062       7.321322
std         0.659831       3.077828
min         1.000000       1.700000
25%         2.000000       5.800000
50%         2.000000       7.400000
75%         2.000000       8.800000
max       200.000000     650.000000
```

[15]:
```
transaction_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 246742 entries, 0 to 264835
Data columns (total 8 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   DATE            246742 non-null  datetime64[ns]
 1   STORE_NBR       246742 non-null  int64
 2   LYLTY_CARD_NBR  246742 non-null  int64
 3   TXN_ID          246742 non-null  int64
 4   PROD_NBR        246742 non-null  int64
 5   PROD_NAME       246742 non-null  object
 6   PROD_QTY        246742 non-null  int64
 7   TOT_SALES       246742 non-null  float64
dtypes: datetime64[ns](1), float64(1), int64(5), object(1)
memory usage: 16.9+ MB
```

[16]:
```
transaction_df.isna().sum()
```

[16]:
```
DATE              0
STORE_NBR         0
```

```
LYLTY_CARD_NBR    0
TXN_ID            0
PROD_NBR          0
PROD_NAME         0
PROD_QTY          0
TOT_SALES         0
dtype: int64
```

Removing Anomalies

[17]:
```
transaction_df['PROD_QTY'].describe()
```

[17]:
```
count    246742.000000
mean          1.908062
std           0.659831
min           1.000000
25%           2.000000
50%           2.000000
75%           2.000000
max         200.000000
Name: PROD_QTY, dtype: float64
```

[18]:
```
transaction_df[transaction_df['PROD_QTY']>5]
```

[18]:

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	TOT_SALES
69762	2018-08-19	226	226000	226201	4	Dorito Corn Chp	Supreme 380g	200	650.0
69763	2019-05-20	226	226000	226210	4	Dorito Corn Chp	Supreme 380g	200	650.0

[19]:
```
transaction_df.drop(labels=transaction_df[transaction_df['PROD_QTY']==200].
↪index,inplace=True)
```

[20]:
```
ts = transaction_df.groupby('DATE').count()
ts.head()
```

[20]:

DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	TOT_SALES
2018-07-01	663	663	663	663	663	663	663
2018-07-02	650	650	650	650	650	650	650
2018-07-03	674	674	674	674	674	674	674
2018-07-04	669	669	669	669	669	669	669
2018-07-05	660	660	660	660	660	660	660

```
DATE
2018-07-01        663
2018-07-02        650
2018-07-03        674
2018-07-04        669
2018-07-05        660
```

[21]:
```
#missing date
set(pd.date_range('2018-07-01', end='2019-06-30',freq='D'))-set((ts.index))
```

[21]:
```
{Timestamp('2018-12-25 00:00:00', freq='D')}
```

[22]:
```
ts.loc['2018-12-25']=np.nan
```

[23]:
```
ts[ts.index=='2018-12-25']
```

[23]:
```
            STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  PROD_NAME  PROD_QTY  \
DATE
2018-12-25        NaN             NaN     NaN       NaN        NaN       NaN

            TOT_SALES
DATE
2018-12-25        NaN
```

[24]:
```
from plotly.offline import init_notebook_mode, iplot
init_notebook_mode(connected=True)
import plotly.offline as offline
offline.init_notebook_mode()
!pip install cufflinks
import cufflinks as cf
cf.go_offline()
```

```
Requirement already satisfied: cufflinks in c:\users\preeti\anaconda\lib\site-
packages (0.17.3)
Requirement already satisfied: six>=1.9.0 in c:\users\preeti\anaconda\lib\site-
packages (from cufflinks) (1.16.0)
Requirement already satisfied: ipywidgets>=7.0.0 in
c:\users\preeti\anaconda\lib\site-packages (from cufflinks) (7.6.5)
Requirement already satisfied: ipython>=5.3.0 in
c:\users\preeti\anaconda\lib\site-packages (from cufflinks) (8.7.0)
Requirement already satisfied: colorlover>=0.2.1 in
c:\users\preeti\anaconda\lib\site-packages (from cufflinks) (0.3.0)
Requirement already satisfied: pandas>=0.19.2 in
c:\users\preeti\anaconda\lib\site-packages (from cufflinks) (1.5.3)
Requirement already satisfied: setuptools>=34.4.1 in
c:\users\preeti\anaconda\lib\site-packages (from cufflinks) (66.0.0)
Requirement already satisfied: plotly>=4.1.1 in
```

```
c:\users\preeti\anaconda\lib\site-packages (from cufflinks) (5.9.0)
Requirement already satisfied: numpy>=1.9.2 in
c:\users\preeti\anaconda\lib\site-packages (from cufflinks) (1.24.1)
Requirement already satisfied: decorator in c:\users\preeti\anaconda\lib\site-
packages (from ipython>=5.3.0->cufflinks) (5.1.1)
Requirement already satisfied: pickleshare in c:\users\preeti\anaconda\lib\site-
packages (from ipython>=5.3.0->cufflinks) (0.7.5)
Requirement already satisfied: matplotlib-inline in
c:\users\preeti\anaconda\lib\site-packages (from ipython>=5.3.0->cufflinks)
(0.1.6)
Requirement already satisfied: backcall in c:\users\preeti\anaconda\lib\site-
packages (from ipython>=5.3.0->cufflinks) (0.2.0)
Requirement already satisfied: traitlets>=5 in
c:\users\preeti\anaconda\lib\site-packages (from ipython>=5.3.0->cufflinks)
(5.7.1)
Requirement already satisfied: colorama in c:\users\preeti\anaconda\lib\site-
packages (from ipython>=5.3.0->cufflinks) (0.4.5)
Requirement already satisfied: pygments>=2.4.0 in
c:\users\preeti\anaconda\lib\site-packages (from ipython>=5.3.0->cufflinks)
(2.11.2)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.11 in
c:\users\preeti\anaconda\lib\site-packages (from ipython>=5.3.0->cufflinks)
(3.0.20)
Requirement already satisfied: stack-data in c:\users\preeti\anaconda\lib\site-
packages (from ipython>=5.3.0->cufflinks) (0.2.0)
Requirement already satisfied: jedi>=0.16 in c:\users\preeti\anaconda\lib\site-
packages (from ipython>=5.3.0->cufflinks) (0.18.1)
Requirement already satisfied: nbformat>=4.2.0 in
c:\users\preeti\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflinks)
(5.7.0)
Requirement already satisfied: ipykernel>=4.5.1 in
c:\users\preeti\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflinks)
(6.15.2)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in
c:\users\preeti\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflinks)
(1.0.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in
c:\users\preeti\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflinks)
(3.5.2)
Requirement already satisfied: ipython-genutils~=0.2.0 in
c:\users\preeti\anaconda\lib\site-packages (from ipywidgets>=7.0.0->cufflinks)
(0.2.0)
Requirement already satisfied: pytz>=2020.1 in
c:\users\preeti\anaconda\lib\site-packages (from pandas>=0.19.2->cufflinks)
(2022.7)
Requirement already satisfied: python-dateutil>=2.8.1 in
c:\users\preeti\anaconda\lib\site-packages (from pandas>=0.19.2->cufflinks)
(2.8.2)
```

```
Requirement already satisfied: tenacity>=6.2.0 in
c:\users\preeti\anaconda\lib\site-packages (from plotly>=4.1.1->cufflinks)
(8.0.1)
Requirement already satisfied: jupyter-client>=6.1.12 in
c:\users\preeti\anaconda\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (6.1.12)
Requirement already satisfied: pyzmq>=17 in c:\users\preeti\anaconda\lib\site-
packages (from ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (23.2.0)
Requirement already satisfied: debugpy>=1.0 in
c:\users\preeti\anaconda\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (1.5.1)
Requirement already satisfied: nest-asyncio in
c:\users\preeti\anaconda\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (1.5.5)
Requirement already satisfied: tornado>=6.1 in
c:\users\preeti\anaconda\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (6.1)
Requirement already satisfied: psutil in c:\users\preeti\anaconda\lib\site-
packages (from ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (5.9.0)
Requirement already satisfied: packaging in c:\users\preeti\anaconda\lib\site-
packages (from ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (23.0)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in
c:\users\preeti\anaconda\lib\site-packages (from
jedi>=0.16->ipython>=5.3.0->cufflinks) (0.8.3)
Requirement already satisfied: fastjsonschema in
c:\users\preeti\anaconda\lib\site-packages (from
nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (2.16.2)
Requirement already satisfied: jsonschema>=2.6 in
c:\users\preeti\anaconda\lib\site-packages (from
nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (4.16.0)
Requirement already satisfied: jupyter-core in
c:\users\preeti\anaconda\lib\site-packages (from
nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (4.11.2)
Requirement already satisfied: wcwidth in c:\users\preeti\anaconda\lib\site-
packages (from prompt-toolkit<3.1.0,>=3.0.11->ipython>=5.3.0->cufflinks) (0.2.5)
Requirement already satisfied: notebook>=4.4.1 in
c:\users\preeti\anaconda\lib\site-packages (from
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (6.5.2)
Requirement already satisfied: pure-eval in c:\users\preeti\anaconda\lib\site-
packages (from stack-data->ipython>=5.3.0->cufflinks) (0.2.2)
Requirement already satisfied: asttokens in c:\users\preeti\anaconda\lib\site-
packages (from stack-data->ipython>=5.3.0->cufflinks) (2.0.5)
Requirement already satisfied: executing in c:\users\preeti\anaconda\lib\site-
packages (from stack-data->ipython>=5.3.0->cufflinks) (0.8.3)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
c:\users\preeti\anaconda\lib\site-packages (from
jsonschema>=2.6->nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (0.18.0)
Requirement already satisfied: attrs>=17.4.0 in
```

```
c:\users\preeti\anaconda\lib\site-packages (from
jsonschema>=2.6->nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (22.1.0)
Requirement already satisfied: pywin32>=1.0 in
c:\users\preeti\anaconda\lib\site-packages (from jupyter-
core->nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (305.1)
Requirement already satisfied: prometheus-client in
c:\users\preeti\anaconda\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(0.14.1)
Requirement already satisfied: argon2-cffi in c:\users\preeti\anaconda\lib\site-
packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(21.3.0)
Requirement already satisfied: Send2Trash>=1.8.0 in
c:\users\preeti\anaconda\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(1.8.0)
Requirement already satisfied: nbconvert>=5 in
c:\users\preeti\anaconda\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(6.4.4)
Requirement already satisfied: jinja2 in c:\users\preeti\anaconda\lib\site-
packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(2.11.3)
Requirement already satisfied: terminado>=0.8.3 in
c:\users\preeti\anaconda\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(0.17.1)
Requirement already satisfied: nbclassic>=0.4.7 in
c:\users\preeti\anaconda\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(0.4.8)
Requirement already satisfied: notebook-shim>=0.1.0 in
c:\users\preeti\anaconda\lib\site-packages (from nbclassic>=0.4.7->notebook>=4.4
.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.2.2)
Requirement already satisfied: jupyter-server>=1.8 in
c:\users\preeti\anaconda\lib\site-packages (from nbclassic>=0.4.7->notebook>=4.4
.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.23.4)
Requirement already satisfied: mistune<2,>=0.8.1 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.8.4)
Requirement already satisfied: testpath in c:\users\preeti\anaconda\lib\site-
packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidg
ets>=7.0.0->cufflinks) (0.6.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.5.0)
```

```
Requirement already satisfied: jupyterlab-pygments in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.1.2)
Requirement already satisfied: beautifulsoup4 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (4.11.1)
Requirement already satisfied: entrypoints>=0.2.2 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.4)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.13)
Requirement already satisfied: defusedxml in c:\users\preeti\anaconda\lib\site-
packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidg
ets>=7.0.0->cufflinks) (0.7.1)
Requirement already satisfied: bleach in c:\users\preeti\anaconda\lib\site-
packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidg
ets>=7.0.0->cufflinks) (4.1.0)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\preeti\anaconda\lib\site-packages (from jinja2->notebook>=4.4.1->widget
snbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (2.0.1)
Requirement already satisfied: pywinpty>=1.1.0 in
c:\users\preeti\anaconda\lib\site-packages (from terminado>=0.8.3->notebook>=4.4
.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (2.0.2)
Requirement already satisfied: argon2-cffi-bindings in
c:\users\preeti\anaconda\lib\site-packages (from argon2-cffi->notebook>=4.4.1->w
idgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (21.2.0)
Requirement already satisfied: anyio<4,>=3.1.0 in
c:\users\preeti\anaconda\lib\site-packages (from jupyter-server>=1.8->nbclassic>
=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks
) (3.5.0)
Requirement already satisfied: websocket-client in
c:\users\preeti\anaconda\lib\site-packages (from jupyter-server>=1.8->nbclassic>
=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks
) (0.58.0)
Requirement already satisfied: cffi>=1.0.1 in c:\users\preeti\anaconda\lib\site-
packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbexte
nsion~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.15.1)
Requirement already satisfied: soupsieve>1.2 in
c:\users\preeti\anaconda\lib\site-packages (from beautifulsoup4->nbconvert>=5->n
otebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks)
(2.3.2.post1)
Requirement already satisfied: webencodings in
c:\users\preeti\anaconda\lib\site-packages (from bleach->nbconvert>=5->notebook>
=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.1)
Requirement already satisfied: sniffio>=1.1 in
c:\users\preeti\anaconda\lib\site-packages (from anyio<4,>=3.1.0->jupyter-server
>=1.8->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>
```

```
=7.0.0->cufflinks) (1.2.0)
Requirement already satisfied: idna>=2.8 in c:\users\preeti\anaconda\lib\site-
packages (from anyio<4,>=3.1.0->jupyter-server>=1.8->nbclassic>=0.4.7->notebook>
=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (3.4)
Requirement already satisfied: pycparser in c:\users\preeti\anaconda\lib\site-
packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->
widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (2.21)
```

[25]:
```
#plot showing missing date
ts['TXN_ID'].iplot(kind='bar',xTitle='Day',yTitle= "Number of transactions",␣
↪title = "Transactions over time")
```

[26]:
```
#adding features
def fun(s):
    a=[]
    for i in s:
        if i in ['0','1','2','3','4','5','6','7','8','9']:
            a.append(i)
    return int("".join(a))
```

[27]:
```
transaction_df['PACKAGE_SIZE']=transaction_df['PROD_NAME'].apply(fun)
transaction_df.head()
```

[27]:

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	\
0	2018-10-17	1	1000	1	5	
1	2019-05-14	1	1307	348	66	
2	2019-05-20	1	1343	383	61	
3	2018-08-17	2	2373	974	69	
4	2018-08-18	2	2426	1038	108	

	PROD_NAME	PROD_QTY	TOT_SALES	PACKAGE_SIZE
0	Natural Chip Compny SeaSalt175g	2	6.0	175
1	CCs Nacho Cheese 175g	3	6.3	175
2	Smiths Crinkle Cut Chips Chicken 170g	2	2.9	170
3	Smiths Chip Thinly S/Cream&Onion 175g	5	15.0	175
4	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.8	150

[28]:
```
#Histogram for pack_size
transaction_df['PACKAGE_SIZE'].iplot(kind='hist',xTitle='Number of␣
↪transactions',yTitle='Packets size',title='Package Size')
```

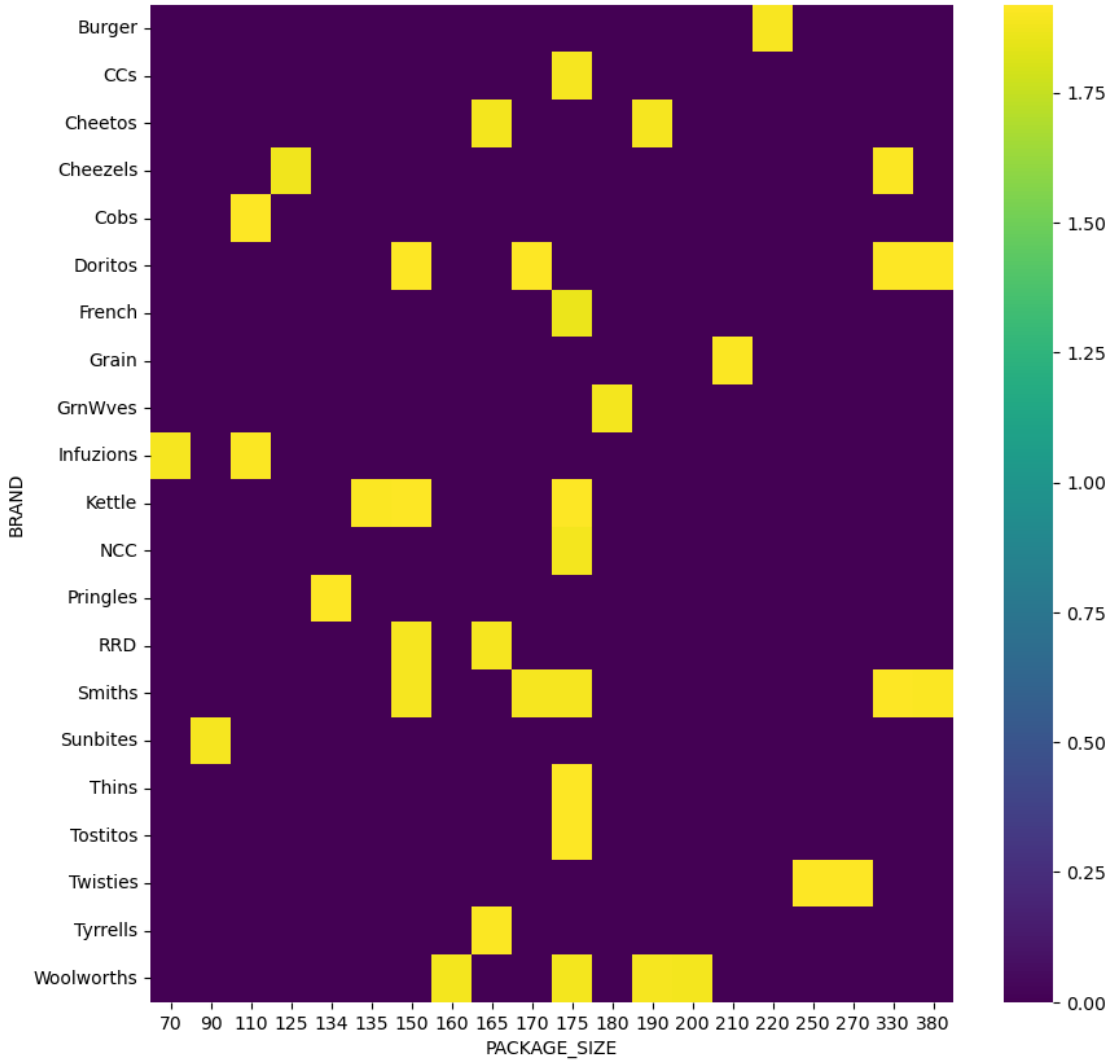
[29]:
```
transaction_df['BRAND']=[s.split()[0] for s in transaction_df['PROD_NAME']]
transaction_df['BRAND'].replace('Dorito','Doritos',inplace=True)
transaction_df['BRAND'].replace('Infzns','Infuzions',inplace=True)
transaction_df['BRAND'].replace('Smith','Smiths',inplace=True)
transaction_df['BRAND'].replace('Snbts','Sunbites',inplace=True)
transaction_df['BRAND'].replace('Red','RRD',inplace=True)
```

```
transaction_df['BRAND'].replace('Old','Old El Paso',inplace=True)
transaction_df['BRAND'].replace('WW','Woolworths',inplace=True)
transaction_df['BRAND'].replace('Natural','NCC',inplace=True)
```

[30]:
```
#Histogram for brands
transaction_df['BRAND'].iplot(kind='hist',xTitle='Brand',yTitle='Packets␣
↪sold',title='Popular brands')
```

[31]:
```
#heatmap showing packet quantity mostly bought according to brand and packet␣
↪size
plt.figure(figsize=(10,10))
sns.heatmap(pd.
↪pivot_table(data=transaction_df,index='BRAND',columns='PACKAGE_SIZE',values='PROD_QTY').
↪fillna(0),cmap='viridis')
```

[31]: <AxesSubplot: xlabel='PACKAGE_SIZE', ylabel='BRAND'>



0.1.2 Purchase Data

[32]:
```
purchase_df.head()
```

[32]:
```
   LYLTY_CARD_NBR               LIFESTAGE PREMIUM_CUSTOMER
0            1000   YOUNG SINGLES/COUPLES          Premium
1            1002   YOUNG SINGLES/COUPLES       Mainstream
2            1003          YOUNG FAMILIES           Budget
3            1004   OLDER SINGLES/COUPLES       Mainstream
4            1005  MIDAGE SINGLES/COUPLES       Mainstream
```

[33]:
```
purchase_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72637 entries, 0 to 72636
Data columns (total 3 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   LYLTY_CARD_NBR    72637 non-null  int64 
 1   LIFESTAGE         72637 non-null  object
 2   PREMIUM_CUSTOMER  72637 non-null  object
dtypes: int64(1), object(2)
memory usage: 1.7+ MB
```

[34]:
```
purchase_df.describe(include='all')
```

[34]:
```
        LYLTY_CARD_NBR LIFESTAGE PREMIUM_CUSTOMER
count     7.263700e+04     72637            72637
unique             NaN         7                3
top                NaN  RETIREES       Mainstream
freq               NaN     14805            29245
mean      1.361859e+05       NaN              NaN
std       8.989293e+04       NaN              NaN
min       1.000000e+03       NaN              NaN
25%       6.620200e+04       NaN              NaN
50%       1.340400e+05       NaN              NaN
75%       2.033750e+05       NaN              NaN
max       2.373711e+06       NaN              NaN
```

[35]:
```
purchase_df.isna().sum()
```

[35]:
```
LYLTY_CARD_NBR      0
LIFESTAGE           0
PREMIUM_CUSTOMER    0
dtype: int64
```

Final Dataset(Joining both datasets)

[36]:
```
final_df= pd.merge(transaction_df,purchase_df, on = 'LYLTY_CARD_NBR')
```

[37]:
```
final_df.head()
```

[37]:

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	TOT_SALES	PACKAGE_SIZE	BRAND	LIFESTAGE	PREMIUM_CUSTOMER
0	2018-10-17	1	1000	1	5	Natural Chip Compny SeaSalt175g	2	6.0	175	NCC	YOUNG SINGLES/COUPLES	Premium
1	2019-05-14	1	1307	348	66	CCs Nacho Cheese 175g	3	6.3	175	CCs	MIDAGE SINGLES/COUPLES	Budget
2	2018-11-10	1	1307	346	96	WW Original Stacked Chips 160g	2	3.8	160	Woolworths	MIDAGE SINGLES/COUPLES	Budget
3	2019-03-09	1	1307	347	54	CCs Original 175g	1	2.1	175	CCs	MIDAGE SINGLES/COUPLES	Budget
4	2019-05-20	1	1343	383	61	Smiths Crinkle Cut Chips Chicken 170g	2	2.9	170	Smiths	MIDAGE SINGLES/COUPLES	Budget

[38]:
```
final_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 246740 entries, 0 to 246739
Data columns (total 12 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   DATE              246740 non-null  datetime64[ns]
 1   STORE_NBR         246740 non-null  int64
 2   LYLTY_CARD_NBR    246740 non-null  int64
 3   TXN_ID            246740 non-null  int64
 4   PROD_NBR          246740 non-null  int64
 5   PROD_NAME         246740 non-null  object
 6   PROD_QTY          246740 non-null  int64
 7   TOT_SALES         246740 non-null  float64
 8   PACKAGE_SIZE      246740 non-null  int64
 9   BRAND             246740 non-null  object
 10  LIFESTAGE         246740 non-null  object
 11  PREMIUM_CUSTOMER  246740 non-null  object
dtypes: datetime64[ns](1), float64(1), int64(6), object(4)
memory usage: 24.5+ MB
```

[39]:
```
final_df.isna().sum()
```

[39]:
```
DATE                0
STORE_NBR           0
LYLTY_CARD_NBR      0
TXN_ID              0
PROD_NBR            0
PROD_NAME           0
PROD_QTY            0
TOT_SALES           0
PACKAGE_SIZE        0
BRAND               0
LIFESTAGE           0
PREMIUM_CUSTOMER    0
dtype: int64
```

[40]:
```
final_df.to_csv('QVI_data.csv')
```

[41]:
```
final_df[['TOT_SALES','PREMIUM_CUSTOMER']].groupby('PREMIUM_CUSTOMER').sum().
 ↪sort_values('TOT_SALES',ascending=False)
```

[41]:

PREMIUM_CUSTOMER	TOT_SALES
Mainstream	700865.40
Budget	631406.85
Premium	472905.45

Who spends the most on chips (total sales), describing customers by lifestage ?

how premium their general purchasing behaviour is?

[42]:
```
#Who spends the most on chips (total sales), describing customers by lifestage␣
 ↪and
#how premium their general purchasing behaviour is
a=final_df[['LIFESTAGE','PREMIUM_CUSTOMER','TOT_SALES']].
 ↪groupby(['PREMIUM_CUSTOMER','LIFESTAGE']).sum()
a.sort_values('TOT_SALES',ascending=False)
```

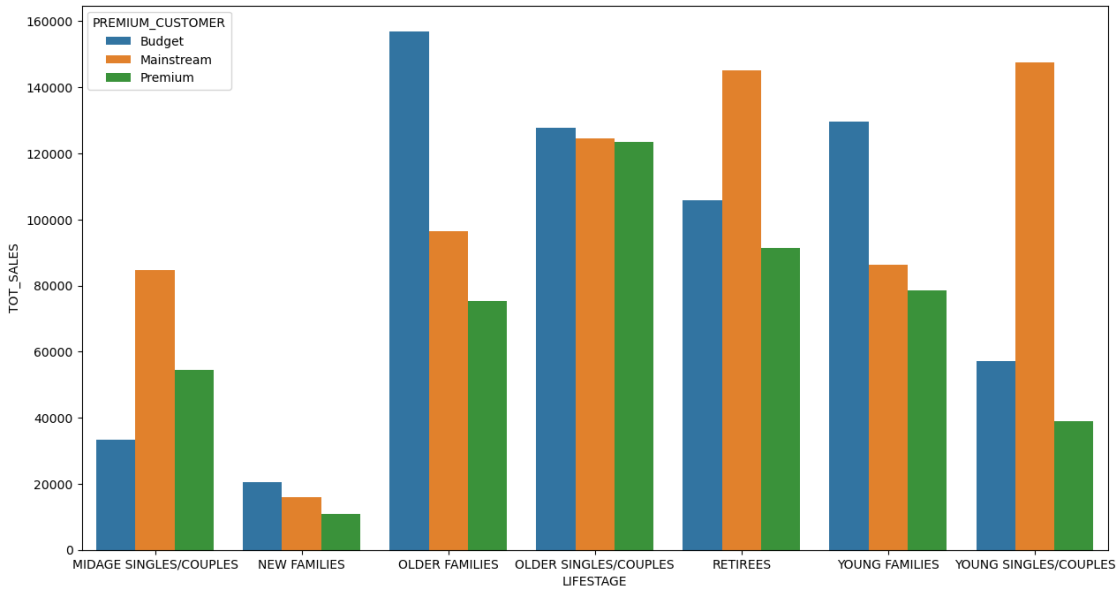
[42]:

PREMIUM_CUSTOMER	LIFESTAGE	TOT_SALES
Budget	OLDER FAMILIES	156863.75
Mainstream	YOUNG SINGLES/COUPLES	147582.20
	RETIREES	145168.95
Budget	YOUNG FAMILIES	129717.95
	OLDER SINGLES/COUPLES	127833.60
Mainstream	OLDER SINGLES/COUPLES	124648.50
Premium	OLDER SINGLES/COUPLES	123537.55

```
Budget          RETIREES                105916.30
Mainstream      OLDER FAMILIES           96413.55
Premium         RETIREES                 91296.65
Mainstream      YOUNG FAMILIES           86338.25
                MIDAGE SINGLES/COUPLES   84734.25
Premium         YOUNG FAMILIES           78571.70
                OLDER FAMILIES           75242.60
Budget          YOUNG SINGLES/COUPLES    57122.10
Premium         MIDAGE SINGLES/COUPLES   54443.85
                YOUNG SINGLES/COUPLES    39052.30
Budget          MIDAGE SINGLES/COUPLES   33345.70
                NEW FAMILIES             20607.45
Mainstream      NEW FAMILIES             15979.70
Premium         NEW FAMILIES             10760.80
```

[43]:
```
plt.figure(figsize=(15,8))
sns.barplot(y=a.reset_index()['TOT_SALES'],x=a.reset_index()['LIFESTAGE'],hue=a.
 ↪reset_index()['PREMIUM_CUSTOMER'])
```

[43]: <AxesSubplot: xlabel='LIFESTAGE', ylabel='TOT_SALES'>



[44]:
```
a.iplot(title="Sales per segment",yTitle='Total sales',xTitle='Segment')
```

[45]:
```
# How many customers are in each segment
b=purchase_df.groupby(['PREMIUM_CUSTOMER','LIFESTAGE']).count()
b.columns=['CUSTOMER_COUNT']
b.sort_values('CUSTOMER_COUNT',ascending=False)
```

[45]:

PREMIUM_CUSTOMER	LIFESTAGE	CUSTOMER_COUNT
Mainstream	YOUNG SINGLES/COUPLES	8088
	RETIREES	6479
	OLDER SINGLES/COUPLES	4930
Budget	OLDER SINGLES/COUPLES	4929
Premium	OLDER SINGLES/COUPLES	4750
Budget	OLDER FAMILIES	4675
	RETIREES	4454
	YOUNG FAMILIES	4017
Premium	RETIREES	3872
Budget	YOUNG SINGLES/COUPLES	3779
Mainstream	MIDAGE SINGLES/COUPLES	3340
	OLDER FAMILIES	2831
	YOUNG FAMILIES	2728
Premium	YOUNG SINGLES/COUPLES	2574
	YOUNG FAMILIES	2433
	MIDAGE SINGLES/COUPLES	2431
	OLDER FAMILIES	2274
Budget	MIDAGE SINGLES/COUPLES	1504
	NEW FAMILIES	1112
Mainstream	NEW FAMILIES	849
Premium	NEW FAMILIES	588

[46]:
```
b.iplot(title="Number of customers per segment",yTitle='No of Customers',xTitle='Segment')
```

This contributes to there being more sales to these customer segments but this is not a major driver for the Budget - Older families segment. Higher sales may also be driven by more units of chips being bought per customer.

0.1.3 How many chips are bought per customer by segment

[47]:
```
#How many chips are bought per customer by segment
c=final_df[['LIFESTAGE','PREMIUM_CUSTOMER','TOT_SALES']].groupby(['LIFESTAGE','PREMIUM_CUSTOMER']).count()
c.sort_values('TOT_SALES',ascending=False).head(5)
```

[47]:

LIFESTAGE	PREMIUM_CUSTOMER	TOT_SALES
OLDER FAMILIES	Budget	21514
RETIREES	Mainstream	19970
YOUNG SINGLES/COUPLES	Mainstream	19544
YOUNG FAMILIES	Budget	17763
OLDER SINGLES/COUPLES	Budget	17172

[48]:
```
c.iplot(title="Number of packets sold per segment",yTitle='No of␣
↪Packets',xTitle='Segment')
```

0.1.4 What's the average chip price by customer segment

[49]:
```
# What's the average chip price by customer segment
final_df['CHIP_PRICE']=final_df['TOT_SALES']/final_df['PROD_QTY']
d=final_df[['LIFESTAGE','PREMIUM_CUSTOMER','CHIP_PRICE']].
↪groupby(['PREMIUM_CUSTOMER','LIFESTAGE']).mean()
d.sort_values("CHIP_PRICE",ascending=False)
```

[49]:

PREMIUM_CUSTOMER	LIFESTAGE	CHIP_PRICE
Mainstream	YOUNG SINGLES/COUPLES	4.065642
	MIDAGE SINGLES/COUPLES	3.994241
Budget	RETIREES	3.924404
Premium	RETIREES	3.920942
Budget	NEW FAMILIES	3.917688
Mainstream	NEW FAMILIES	3.916133
Premium	OLDER SINGLES/COUPLES	3.893182
Budget	OLDER SINGLES/COUPLES	3.882096
Premium	NEW FAMILIES	3.872110
Mainstream	RETIREES	3.844294
	OLDER SINGLES/COUPLES	3.814665
Premium	MIDAGE SINGLES/COUPLES	3.770698
	YOUNG FAMILIES	3.762150
Budget	YOUNG FAMILIES	3.760737
	OLDER FAMILIES	3.745340
	MIDAGE SINGLES/COUPLES	3.743328
Mainstream	OLDER FAMILIES	3.737077
	YOUNG FAMILIES	3.724533
Premium	OLDER FAMILIES	3.717000
	YOUNG SINGLES/COUPLES	3.665414
Budget	YOUNG SINGLES/COUPLES	3.657366

[50]:
```
d.iplot(title="Avg pay per packet per segment",yTitle='Avg␣
↪Payment',xTitle='Segment')
```

[51]:
```
#  The customer's total spend over the period
# to understand what proportion of their grocery spend is on chips
transaction1=pd.read_excel("QVI_transaction_data.xlsx")
totsalespercust=transaction1[['LYLTY_CARD_NBR','TOT_SALES']].
↪groupby(['LYLTY_CARD_NBR']).sum().reset_index()
ratio=final_df[['LYLTY_CARD_NBR','TOT_SALES']].
↪merge(totsalespercust,on='LYLTY_CARD_NBR').rename(columns={'TOT_SALES_x':
↪'TRAN_SALE','TOT_SALES_y':'CUST_TOT_SALE'})
```

```
ratio['RATIO']=ratio['TRAN_SALE']/ratio['CUST_TOT_SALE']
ratio.sort_values('RATIO')
```

[51]:

	LYLTY_CARD_NBR	TRAN_SALE	CUST_TOT_SALE	RATIO
174208	152094	1.9	112.1	0.016949
75460	48155	1.9	100.7	0.018868
174557	168140	1.7	86.5	0.019653
16284	104061	1.7	85.9	0.019790
30772	55244	1.7	85.7	0.019837
…	…	…	…	…
163956	49312	11.4	11.4	1.000000
163855	47486	7.4	7.4	1.000000
163852	47465	10.8	10.8	1.000000
162683	12139	8.6	8.6	1.000000
246739	272380	8.8	8.8	1.000000

[246740 rows x 4 columns]

[52]:
```
#  Proportion of customers in each customer segment overall to compare against␣
 ↪the
# mix of customers who purchase chips
e=final_df[['LIFESTAGE','PREMIUM_CUSTOMER','TOT_SALES']].
 ↪groupby(['PREMIUM_CUSTOMER','LIFESTAGE']).count()
e["TOT_SALES"]/(e['TOT_SALES'].sum())
```

[52]:

PREMIUM_CUSTOMER	LIFESTAGE	
Budget	MIDAGE SINGLES/COUPLES	0.019012
	NEW FAMILIES	0.011445
	OLDER FAMILIES	0.087193
	OLDER SINGLES/COUPLES	0.069596
	RETIREES	0.057652
	YOUNG FAMILIES	0.071991
	YOUNG SINGLES/COUPLES	0.034745
Mainstream	MIDAGE SINGLES/COUPLES	0.044966
	NEW FAMILIES	0.008855
	OLDER FAMILIES	0.053664
	OLDER SINGLES/COUPLES	0.069146
	RETIREES	0.080935
	YOUNG FAMILIES	0.048419
	YOUNG SINGLES/COUPLES	0.079209
Premium	MIDAGE SINGLES/COUPLES	0.030850
	NEW FAMILIES	0.006031
	OLDER FAMILIES	0.042162
	OLDER SINGLES/COUPLES	0.067115
	RETIREES	0.049591
	YOUNG FAMILIES	0.043706
	YOUNG SINGLES/COUPLES	0.023717

```
Name: TOT_SALES, dtype: float64
```

0.2 t-test

[53]:
```
#Mainstream vs premium
stats.ttest_ind([4.065642,3.994241],[3.770698,3.665414])
```

[53]:
```
Ttest_indResult(statistic=4.903408005498769, pvalue=0.039164352682153285)
```

[54]:
```
#Mainstream vs budget
stats.ttest_ind([4.065642,3.994241],[3.657366,3.743328])
```

[54]:
```
Ttest_indResult(statistic=5.898899732826305, pvalue=0.027555775534860754)
```

0.2.1 The t-test results in a p-value of 0.03 and 0.02 , i.e. the unit price for mainstream, young and mid-age singles and couples ARE significantly higher than that of budget or premium, young and midage singles and couples. Now we are focussing on the mainstream, young and mid-age singles and couples brands that these two customer segments prefer more than others

[55]:
```
midage=final_df[(final_df['PREMIUM_CUSTOMER']=='Mainstream') &␣
 ↪(final_df['LIFESTAGE']=='MIDAGE SINGLES/COUPLES')]
young=final_df[(final_df['PREMIUM_CUSTOMER']=='Mainstream') &␣
 ↪(final_df['LIFESTAGE']=='YOUNG SINGLES/COUPLES')]
print(f"MIDAGE SINGLES/COUPLES\n{midage['BRAND'].value_counts().head(5)}")
print(f"YOUNG SINGLES/COUPLES\n{young['BRAND'].value_counts().head(5)}")
```

```
MIDAGE SINGLES/COUPLES
Kettle       2136
Smiths       1276
Doritos      1210
Pringles     1159
Infuzions     679
Name: BRAND, dtype: int64
YOUNG SINGLES/COUPLES
Kettle       3844
Doritos      2379
Pringles     2315
Smiths       1921
Infuzions    1250
Name: BRAND, dtype: int64
```

0.2.2 Kettle, Smiths and Doritos are popular among MIDAGE and Kettle, Pringles and Doritos are popular among YOUNG

[56]:
```
print(f"MIDAGE SINGLES/COUPLES\n{midage['PACKAGE_SIZE'].value_counts().
  ↪head(5)}")
print(f"YOUNG SINGLES/COUPLES\n{young['PACKAGE_SIZE'].value_counts().head(5)}")
```

```
MIDAGE SINGLES/COUPLES
175    2975
150    1777
134    1159
110    1124
170     882
Name: PACKAGE_SIZE, dtype: int64
YOUNG SINGLES/COUPLES
175    4997
150    3080
134    2315
110    2051
170    1575
Name: PACKAGE_SIZE, dtype: int64
```

0.2.3 Both the segments buy 175g,150g and 134 packets mostly

0.2.4 Therefore if someone buys Doritos Kettle can be recommended and vice-versa. Same for Pringles and Kettle.

[57]:
```
!pip install nbconvert[webpdf]
```

```
Requirement already satisfied: nbconvert[webpdf] in
c:\users\preeti\anaconda\lib\site-packages (6.4.4)
Requirement already satisfied: nbformat>=4.4 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (5.7.0)
Requirement already satisfied: traitlets>=5.0 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (5.7.1)
Requirement already satisfied: entrypoints>=0.2.2 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (0.4)
Requirement already satisfied: bleach in c:\users\preeti\anaconda\lib\site-
packages (from nbconvert[webpdf]) (4.1.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (1.5.0)
Requirement already satisfied: beautifulsoup4 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (4.11.1)
Requirement already satisfied: mistune<2,>=0.8.1 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (0.8.4)
Requirement already satisfied: testpath in c:\users\preeti\anaconda\lib\site-
packages (from nbconvert[webpdf]) (0.6.0)
Requirement already satisfied: defusedxml in c:\users\preeti\anaconda\lib\site-
```

```
packages (from nbconvert[webpdf]) (0.7.1)
Requirement already satisfied: pygments>=2.4.1 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (2.11.2)
Requirement already satisfied: jupyterlab-pygments in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (0.1.2)
Requirement already satisfied: jupyter-core in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (4.11.2)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in
c:\users\preeti\anaconda\lib\site-packages (from nbconvert[webpdf]) (0.5.13)
Requirement already satisfied: jinja2>=2.4 in c:\users\preeti\anaconda\lib\site-
packages (from nbconvert[webpdf]) (2.11.3)
Collecting pyppeteer<1.1,>=1
  Downloading pyppeteer-1.0.2-py3-none-any.whl (83 kB)
     ---------------------------------------- 83.4/83.4 kB 2.4 MB/s eta 0:00:00
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\preeti\anaconda\lib\site-packages (from jinja2>=2.4->nbconvert[webpdf])
(2.0.1)
Requirement already satisfied: nest-asyncio in
c:\users\preeti\anaconda\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert[webpdf]) (1.5.5)
Requirement already satisfied: jupyter-client>=6.1.5 in
c:\users\preeti\anaconda\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert[webpdf]) (6.1.12)
Requirement already satisfied: jsonschema>=2.6 in
c:\users\preeti\anaconda\lib\site-packages (from
nbformat>=4.4->nbconvert[webpdf]) (4.16.0)
Requirement already satisfied: fastjsonschema in
c:\users\preeti\anaconda\lib\site-packages (from
nbformat>=4.4->nbconvert[webpdf]) (2.16.2)
Requirement already satisfied: importlib-metadata>=1.4 in
c:\users\preeti\anaconda\lib\site-packages (from
pyppeteer<1.1,>=1->nbconvert[webpdf]) (6.0.0)
Collecting websockets<11.0,>=10.0
  Downloading websockets-10.4-cp39-cp39-win_amd64.whl (101 kB)
     ---------------------------------------- 101.4/101.4 kB ? eta 0:00:00
Requirement already satisfied: tqdm<5.0.0,>=4.42.1 in
c:\users\preeti\anaconda\lib\site-packages (from
pyppeteer<1.1,>=1->nbconvert[webpdf]) (4.64.1)
Requirement already satisfied: urllib3<2.0.0,>=1.25.8 in
c:\users\preeti\anaconda\lib\site-packages (from
pyppeteer<1.1,>=1->nbconvert[webpdf]) (1.26.14)
Requirement already satisfied: certifi>=2021 in
c:\users\preeti\anaconda\lib\site-packages (from
pyppeteer<1.1,>=1->nbconvert[webpdf]) (2022.12.7)
Collecting pyee<9.0.0,>=8.1.0
  Downloading pyee-8.2.2-py2.py3-none-any.whl (12 kB)
Requirement already satisfied: appdirs<2.0.0,>=1.4.3 in
c:\users\preeti\anaconda\lib\site-packages (from
```

```
pyppeteer<1.1,>=1->nbconvert[webpdf]) (1.4.4)
Requirement already satisfied: soupsieve>1.2 in
c:\users\preeti\anaconda\lib\site-packages (from
beautifulsoup4->nbconvert[webpdf]) (2.3.2.post1)
Requirement already satisfied: webencodings in
c:\users\preeti\anaconda\lib\site-packages (from bleach->nbconvert[webpdf])
(0.5.1)
Requirement already satisfied: packaging in c:\users\preeti\anaconda\lib\site-
packages (from bleach->nbconvert[webpdf]) (23.0)
Requirement already satisfied: six>=1.9.0 in c:\users\preeti\anaconda\lib\site-
packages (from bleach->nbconvert[webpdf]) (1.16.0)
Requirement already satisfied: pywin32>=1.0 in
c:\users\preeti\anaconda\lib\site-packages (from jupyter-
core->nbconvert[webpdf]) (305.1)
Requirement already satisfied: zipp>=0.5 in c:\users\preeti\anaconda\lib\site-
packages (from importlib-metadata>=1.4->pyppeteer<1.1,>=1->nbconvert[webpdf])
(3.11.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
c:\users\preeti\anaconda\lib\site-packages (from
jsonschema>=2.6->nbformat>=4.4->nbconvert[webpdf]) (0.18.0)
Requirement already satisfied: attrs>=17.4.0 in
c:\users\preeti\anaconda\lib\site-packages (from
jsonschema>=2.6->nbformat>=4.4->nbconvert[webpdf]) (22.1.0)
Requirement already satisfied: python-dateutil>=2.1 in
c:\users\preeti\anaconda\lib\site-packages (from jupyter-
client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert[webpdf]) (2.8.2)
Requirement already satisfied: tornado>=4.1 in
c:\users\preeti\anaconda\lib\site-packages (from jupyter-
client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert[webpdf]) (6.1)
Requirement already satisfied: pyzmq>=13 in c:\users\preeti\anaconda\lib\site-
packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert[webpdf])
(23.2.0)
Requirement already satisfied: colorama in c:\users\preeti\anaconda\lib\site-
packages (from tqdm<5.0.0,>=4.42.1->pyppeteer<1.1,>=1->nbconvert[webpdf])
(0.4.5)
Installing collected packages: pyee, websockets, pyppeteer
Successfully installed pyee-8.2.2 pyppeteer-1.0.2 websockets-10.4
```

[ ]: