

untitled4

August 3, 2024

```
[4]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.style as style
import seaborn as sns
import datetime
import xlrd
%matplotlib inline

# Ignoring any warnings.

import warnings
warnings.simplefilter(action="ignore", category=FutureWarning)
qvi_data=pd.read_csv(r"C:\Users\Divya\Downloads\QVI_data.csv")
qvi_data
```

```
[4]:
```

	LYLTY_CARD_NBR	DATE	STORE_NBR	TXN_ID	PROD_NBR	\
0	1000	2018-10-17	1	1	5	
1	1002	2018-09-16	1	2	58	
2	1003	2019-03-07	1	3	52	
3	1003	2019-03-08	1	4	106	
4	1004	2018-11-02	1	5	96	
...	
264829	2370701	2018-12-08	88	240378	24	
264830	2370751	2018-10-01	88	240394	60	
264831	2370961	2018-10-24	88	240480	70	
264832	2370961	2018-10-27	88	240481	65	
264833	2373711	2018-12-14	88	241815	16	

	PROD_NAME	PROD_QTY	TOT_SALES	\
0	Natural Chip Compny SeaSalt175g	2	6.0	
1	Red Rock Deli Chikn&Garlic Aioli 150g	1	2.7	
2	Grain Waves Sour Cream&Chives 210G	1	3.6	
3	Natural ChipCo Hony Soy Chckn175g	1	3.0	
4	WW Original Stacked Chips 160g	1	1.9	
...	
264829	Grain Waves Sweet Chillli 210g	2	7.2	

264830	Kettle Tortilla ChpsFeta&Garlic	150g	2	9.2
264831	Tyrrells Crisps	Lightly Salted 165g	2	8.4
264832	Old El Paso Salsa	Dip Chnky Tom Ht300g	2	10.2
264833	Smiths Crinkle Chips	Salt & Vinegar 330g	2	11.4

	PACK_SIZE	BRAND	LIFESTAGE	PREMIUM_CUSTOMER
0	175	NATURAL	YOUNG SINGLES/COUPLES	Premium
1	150	RRD	YOUNG SINGLES/COUPLES	Mainstream
2	210	GRNWVES	YOUNG FAMILIES	Budget
3	175	NATURAL	YOUNG FAMILIES	Budget
4	160	WOOLWORTHS	OLDER SINGLES/COUPLES	Mainstream
...
264829	210	GRNWVES	YOUNG FAMILIES	Mainstream
264830	150	KETTLE	YOUNG FAMILIES	Premium
264831	165	TYRRELLS	OLDER FAMILIES	Budget
264832	300	OLD	OLDER FAMILIES	Budget
264833	330	SMITHS	YOUNG SINGLES/COUPLES	Mainstream

[264834 rows x 12 columns]

```
[2]: pip install xlrd
```

Collecting xlrd

Downloading xlrd-2.0.1-py2.py3-none-any.whl.metadata (3.4 kB)

Downloading xlrd-2.0.1-py2.py3-none-any.whl (96 kB)

```
----- 0.0/96.5 kB ? eta -:--:--
----- 0.0/96.5 kB ? eta -:--:--
---- 10.2/96.5 kB ? eta -:--:--
----- 41.0/96.5 kB 388.9 kB/s eta 0:00:01
----- 81.9/96.5 kB 573.4 kB/s eta 0:00:01
----- 96.5/96.5 kB 612.0 kB/s eta 0:00:00
```

Installing collected packages: xlrd

Successfully installed xlrd-2.0.1

Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 24.0 -> 24.2

[notice] To update, run: python.exe -m pip install --upgrade pip

```
[3]: pip install datetime
```

Collecting datetimeNote: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 24.0 -> 24.2

[notice] To update, run: python.exe -m pip install --upgrade pip

```

Downloading DateTime-5.5-py3-none-any.whl.metadata (33 kB)
Collecting zope.interface (from datetime)
  Downloading zope.interface-6.4.post2-cp312-cp312-win_amd64.whl.metadata (44
kB)
----- 0.0/44.1 kB ? eta -:--:--
----- 30.7/44.1 kB 1.3 MB/s eta 0:00:01
----- 44.1/44.1 kB 720.4 kB/s eta 0:00:00
Requirement already satisfied: pytz in
c:\users\divya\appdata\local\programs\python\python312\lib\site-packages (from
datetime) (2024.1)
Collecting setuptools (from zope.interface->datetime)
  Using cached setuptools-72.1.0-py3-none-any.whl.metadata (6.6 kB)
Downloading DateTime-5.5-py3-none-any.whl (52 kB)
----- 0.0/52.6 kB ? eta -:--:--
----- 51.2/52.6 kB 2.7 MB/s eta 0:00:01
----- 52.6/52.6 kB 921.9 kB/s eta 0:00:00
Downloading zope.interface-6.4.post2-cp312-cp312-win_amd64.whl (206 kB)
----- 0.0/206.5 kB ? eta -:--:--
----- 61.4/206.5 kB 1.1 MB/s eta 0:00:01
----- 112.6/206.5 kB 1.3 MB/s eta 0:00:01
----- 163.8/206.5 kB 1.1 MB/s eta 0:00:01
----- 206.5/206.5 kB 1.0 MB/s eta 0:00:00
Using cached setuptools-72.1.0-py3-none-any.whl (2.3 MB)
Installing collected packages: setuptools, zope.interface, datetime
Successfully installed datetime-5.5 setuptools-72.1.0 zope.interface-6.4.post2

```

```
[5]: qvi_data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 264834 entries, 0 to 264833
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   LYLTY_CARD_NBR        264834 non-null int64
 1   DATE                  264834 non-null object
 2   STORE_NBR             264834 non-null int64
 3   TXN_ID                264834 non-null int64
 4   PROD_NBR              264834 non-null int64
 5   PROD_NAME             264834 non-null object
 6   PROD_QTY              264834 non-null int64
 7   TOT_SALES             264834 non-null float64
 8   PACK_SIZE             264834 non-null int64
 9   BRAND                 264834 non-null object
10   LIFESTAGE             264834 non-null object
11   PREMIUM_CUSTOMER      264834 non-null object
dtypes: float64(1), int64(6), object(5)
memory usage: 24.2+ MB

```

```
[6]: qvi_data.shape
```

```
[6]: (264834, 12)
```

```
[7]: qvi_data.isnull().sum()
```

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

```
[9]: qvi_data["YEAR_MONTH"]=pd.to_datetime(qvi_data["DATE"]).dt.to_period("M")
     qvi_data
```

```
[9]:
```

	LYLTY_CARD_NBR	DATE	STORE_NBR	TXN_ID	PROD_NBR	\
0	1000	2018-10-17	1	1	5	
1	1002	2018-09-16	1	2	58	
2	1003	2019-03-07	1	3	52	
3	1003	2019-03-08	1	4	106	
4	1004	2018-11-02	1	5	96	
...	
264829	2370701	2018-12-08	88	240378	24	
264830	2370751	2018-10-01	88	240394	60	
264831	2370961	2018-10-24	88	240480	70	
264832	2370961	2018-10-27	88	240481	65	
264833	2373711	2018-12-14	88	241815	16	

	PROD_NAME	PROD_QTY	TOT_SALES	\
0	Natural Chip Compny SeaSalt175g	2	6.0	
1	Red Rock Deli Chikn&Garlic Aioli 150g	1	2.7	
2	Grain Waves Sour Cream&Chives 210G	1	3.6	
3	Natural ChipCo Hony Soy Chckn175g	1	3.0	
4	WW Original Stacked Chips 160g	1	1.9	
...	
264829	Grain Waves Sweet Chilli 210g	2	7.2	
264830	Kettle Tortilla ChpsFeta&Garlic 150g	2	9.2	
264831	Tyrrells Crisps Lightly Salted 165g	2	8.4	
264832	Old El Paso Salsa Dip Chnky Tom Ht300g	2	10.2	

264833 Smiths Crinkle Chips Salt & Vinegar 330g 2 11.4

	PACK_SIZE	BRAND	LIFESTAGE	PREMIUM_CUSTOMER	\
0	175	NATURAL	YOUNG SINGLES/COUPLES	Premium	
1	150	RRD	YOUNG SINGLES/COUPLES	Mainstream	
2	210	GRNWVES	YOUNG FAMILIES	Budget	
3	175	NATURAL	YOUNG FAMILIES	Budget	
4	160	WOOLWORTHS	OLDER SINGLES/COUPLES	Mainstream	
...	
264829	210	GRNWVES	YOUNG FAMILIES	Mainstream	
264830	150	KETTLE	YOUNG FAMILIES	Premium	
264831	165	TYRRELLS	OLDER FAMILIES	Budget	
264832	300	OLD	OLDER FAMILIES	Budget	
264833	330	SMITHS	YOUNG SINGLES/COUPLES	Mainstream	

	YEAR_MONTH
0	2018-10
1	2018-09
2	2019-03
3	2019-03
4	2018-11
...	...
264829	2018-12
264830	2018-10
264831	2018-10
264832	2018-10
264833	2018-12

[264834 rows x 13 columns]

```
[10]: total_sales=qvi_data.groupby(["STORE_NBR", "YEAR_MONTH"])["TOT_SALES"].sum()
total_sales=total_sales.to_frame()
total_sales
```

```
[10]:
```

	STORE_NBR	YEAR_MONTH	TOT_SALES
	1	2018-07	206.9
		2018-08	176.1
		2018-09	278.8
		2018-10	188.1
		2018-11	192.6

	272	2019-02	395.5
		2019-03	442.3
		2019-04	445.1
		2019-05	314.6
		2019-06	312.1

[3169 rows x 1 columns]

```
[11]: total_customers=qvi_data.groupby(["STORE_NBR", "YEAR_MONTH"])["LYLTY_CARD_NBR"].  
      ↪nunique()  
      total_customers=total_customers.to_frame()  
      total_customers
```

```
[11]:
```

		LYLTY_CARD_NBR
STORE_NBR	YEAR_MONTH	
1	2018-07	49
	2018-08	42
	2018-09	59
	2018-10	44
	2018-11	46
...		...
272	2019-02	45
	2019-03	50
	2019-04	54
	2019-05	34
	2019-06	34

[3169 rows x 1 columns]

```
[13]: transactions_per_customer=qvi_data.groupby(["STORE_NBR",  
      ↪"YEAR_MONTH"])["TXN_ID"].nunique()/qvi_data.groupby(["STORE_NBR",  
      ↪"YEAR_MONTH"])["LYLTY_CARD_NBR"].nunique()  
      transactions_per_customer=transactions_per_customer.to_frame()  
      transactions_per_customer
```

```
[13]:
```

		0
STORE_NBR	YEAR_MONTH	
1	2018-07	1.061224
	2018-08	1.023810
	2018-09	1.050847
	2018-10	1.022727
	2018-11	1.021739
...		...
272	2019-02	1.066667
	2019-03	1.060000
	2019-04	1.018519
	2019-05	1.176471
	2019-06	1.088235

[3169 rows x 1 columns]

```
[14]: dataframe_list=[total_sales, total_customers, transactions_per_customer]
dataframe=pd.concat(dataframe_list, axis=1)
dataframe.columns=["TOT_SALES", "TOT_CUST", "TXN_PER_CUST"]
dataframe
```

```
[14]:
```

		TOT_SALES	TOT_CUST	TXN_PER_CUST
STORE_NBR	YEAR_MONTH			
1	2018-07	206.9	49	1.061224
	2018-08	176.1	42	1.023810
	2018-09	278.8	59	1.050847
	2018-10	188.1	44	1.022727
	2018-11	192.6	46	1.021739
...	
272	2019-02	395.5	45	1.066667
	2019-03	442.3	50	1.060000
	2019-04	445.1	54	1.018519
	2019-05	314.6	34	1.176471
	2019-06	312.1	34	1.088235

[3169 rows x 3 columns]

```
[15]: recorded_stores=pd.pivot_table(qvi_data, index="STORE_NBR",
↪columns="YEAR_MONTH", values="TXN_ID", aggfunc="count")
recorded_stores
```

```
[15]:
```

YEAR_MONTH	2018-07	2018-08	2018-09	2018-10	2018-11	2018-12	2019-01	\
STORE_NBR								
1	52.0	43.0	62.0	45.0	47.0	47.0	36.0	
2	41.0	43.0	37.0	43.0	40.0	38.0	45.0	
3	138.0	134.0	119.0	119.0	118.0	129.0	121.0	
4	160.0	151.0	138.0	155.0	139.0	133.0	168.0	
5	120.0	112.0	125.0	107.0	111.0	125.0	118.0	
...	
268	52.0	54.0	34.0	48.0	51.0	43.0	38.0	
269	139.0	132.0	124.0	148.0	136.0	133.0	144.0	
270	139.0	154.0	126.0	119.0	133.0	149.0	155.0	
271	129.0	101.0	114.0	114.0	122.0	117.0	120.0	
272	52.0	48.0	36.0	51.0	45.0	47.0	50.0	
YEAR_MONTH	2019-02	2019-03	2019-04	2019-05	2019-06			
STORE_NBR								
1	55.0	49.0	43.0	51.0	43.0			
2	32.0	46.0	49.0	50.0	42.0			
3	139.0	130.0	110.0	123.0	122.0			
4	102.0	135.0	137.0	126.0	134.0			
5	106.0	97.0	109.0	104.0	127.0			
...			

268	37.0	47.0	50.0	52.0	40.0
269	133.0	122.0	139.0	130.0	127.0
270	125.0	143.0	132.0	128.0	127.0
271	102.0	101.0	109.0	127.0	129.0
272	48.0	53.0	56.0	40.0	37.0

[272 rows x 12 columns]

```
[16]: recorded_stores.isnull().sum()
```

```
[16]: YEAR_MONTH
2018-07    6
2018-08    9
2018-09    8
2018-10    7
2018-11    8
2018-12    9
2019-01    9
2019-02    8
2019-03    7
2019-04    7
2019-05    9
2019-06    8
Freq: M, dtype: int64
```

```
[17]: unrecorded_stores=[]
for i in recorded_stores.index:
    if recorded_stores.loc[i].isnull().any():
        unrecorded_stores.append(i)
unrecorded_stores
```

```
[17]: [11, 31, 44, 76, 85, 92, 117, 193, 206, 211, 218, 252]
```

```
[18]: dataframe=dataframe.drop(unrecorded_stores, axis=0)
dataframe
```

```
[18]:
```

	STORE_NBR	YEAR_MONTH	TOT_SALES	TOT_CUST	TXN_PER_CUST
	1	2018-07	206.9	49	1.061224
		2018-08	176.1	42	1.023810
		2018-09	278.8	59	1.050847
		2018-10	188.1	44	1.022727
		2018-11	192.6	46	1.021739
...		
272		2019-02	395.5	45	1.066667
		2019-03	442.3	50	1.060000
		2019-04	445.1	54	1.018519

2019-05	314.6	34	1.176471
2019-06	312.1	34	1.088235

[3120 rows x 3 columns]

```
[19]: pre_trial_data=dataframe.loc[dataframe.index.
    ↳get_level_values("YEAR_MONTH")<"2019-02"]
pre_trial_data=pre_trial_data.reset_index()
pre_trial_data
```

```
[19]:
```

	STORE_NBR	YEAR_MONTH	TOT_SALES	TOT_CUST	TXN_PER_CUST
0	1	2018-07	206.9	49	1.061224
1	1	2018-08	176.1	42	1.023810
2	1	2018-09	278.8	59	1.050847
3	1	2018-10	188.1	44	1.022727
4	1	2018-11	192.6	46	1.021739
...
1815	272	2018-09	304.7	32	1.125000
1816	272	2018-10	430.6	44	1.136364
1817	272	2018-11	376.2	41	1.097561
1818	272	2018-12	403.9	47	1.000000
1819	272	2019-01	423.0	46	1.086957

[1820 rows x 5 columns]

```
[20]: control_stores=pre_trial_data[(pre_trial_data.STORE_NBR!=77 ) & (pre_trial_data.
    ↳STORE_NBR!=86) & (pre_trial_data.STORE_NBR!=88)][["TOT_SALES", "TOT_CUST",
    ↳"TXN_PER_CUST"]].groupby(pre_trial_data.STORE_NBR).sum()
control_stores
```

```
[20]:
```

	TOT_SALES	TOT_CUST	TXN_PER_CUST
STORE_NBR			
1	1386.90	317	7.327967
2	1128.50	272	7.359700
3	7526.15	744	8.209829
4	9127.00	849	8.535253
5	5739.70	651	8.791906
...
268	1549.05	304	7.373037
269	6664.50	746	8.921035
270	6697.95	734	9.147187
271	5765.10	652	8.671966
272	2744.35	302	7.620124

[257 rows x 3 columns]

```
[21]: trial_stores=pre_trial_data[(pre_trial_data.STORE_NBR==77 ) | (pre_trial_data.
    ↪STORE_NBR==86) | (pre_trial_data.STORE_NBR==88)][["TOT_SALES", "TOT_CUST",
    ↪"TXN_PER_CUST"]].groupby(pre_trial_data.STORE_NBR).sum()
trial_stores
```

```
[21]:
```

	TOT_SALES	TOT_CUST	TXN_PER_CUST
STORE_NBR			
77	1699.00	299	7.405289
86	6119.85	697	8.798544
88	9383.60	880	8.523817

```
[22]: difference=control_stores.loc[control_stores.corrwith(trial_stores.loc[77],
    ↪method="pearson", axis=1).nlargest(5).index]

difference=(trial_stores.loc[77]-difference).sort_values(by="TOT_SALES",
    ↪ascending=False)
difference["DIFFERENCE"]=difference["TOT_SALES"]-difference["TOT_SALES"].mean()
difference.sort_values(by="DIFFERENCE", ascending=False)
```

```
[22]:
```

	TOT_SALES	TOT_CUST	TXN_PER_CUST	DIFFERENCE
STORE_NBR				
139	1493.2	257.0	0.405289	609.34
135	1486.9	256.0	0.012432	603.04
161	1459.0	252.0	0.405289	575.14
233	39.2	1.0	0.115969	-844.66
46	-59.0	-3.0	0.094215	-942.86

```
[23]: difference=control_stores.loc[control_stores.corrwith(trial_stores.loc[86],
    ↪axis=1).nlargest(5).index]
difference=(trial_stores.loc[86]-difference).sort_values(by="TOT_SALES",
    ↪ascending=False)
difference["DIFFERENCE"]=difference["TOT_SALES"]-difference["TOT_SALES"].mean()
difference.sort_values(by="DIFFERENCE", ascending=False)
```

```
[23]:
```

	TOT_SALES	TOT_CUST	TXN_PER_CUST	DIFFERENCE
STORE_NBR				
258	5934.85	670.0	1.798544	4066.46
215	3411.85	386.0	1.486773	1543.46
225	29.25	3.0	0.023669	-1839.14
196	-6.45	1.0	0.040716	-1874.84
57	-27.55	-2.0	0.031815	-1895.94

```
[24]: difference=control_stores.loc[control_stores.corrwith(trial_stores.loc[88],
    ↪axis=1).nlargest(5).index]
difference=(trial_stores.loc[88]-difference).sort_values(by="TOT_SALES",
    ↪ascending=False)
```

```
difference["DIFFERENCE"]=difference["TOT_SALES"]-difference["TOT_SALES"].mean()
difference.sort_values(by="DIFFERENCE", ascending=False)
```

```
[24]:
```

	TOT_SALES	TOT_CUST	TXN_PER_CUST	DIFFERENCE
STORE_NBR				
60	1697.1	154.0	0.052504	783.5
75	1420.1	129.0	0.078986	506.5
72	865.1	79.0	0.085871	-48.5
203	439.9	38.0	0.135139	-473.7
165	145.8	18.0	0.093389	-767.8

```
[25]: trial_stores_one=pre_trial_data.loc[pre_trial_data.STORE_NBR.isin([77])].
      ↪reset_index()
trial_stores_two=pre_trial_data.loc[pre_trial_data.STORE_NBR.isin([86])].
      ↪reset_index()
trial_stores_three=pre_trial_data.loc[pre_trial_data.STORE_NBR.isin([88])].
      ↪reset_index()

control_stores_one=pre_trial_data.loc[pre_trial_data.STORE_NBR.isin([46])].
      ↪reset_index()
control_stores_two=pre_trial_data.loc[pre_trial_data.STORE_NBR.isin([57])].
      ↪reset_index()
control_stores_three=pre_trial_data.loc[pre_trial_data.STORE_NBR.isin([165])].
      ↪reset_index()

stores=pd.concat([trial_stores_one, trial_stores_two, trial_stores_three,
      ↪control_stores_one, control_stores_two, control_stores_three], axis=0)
stores
```

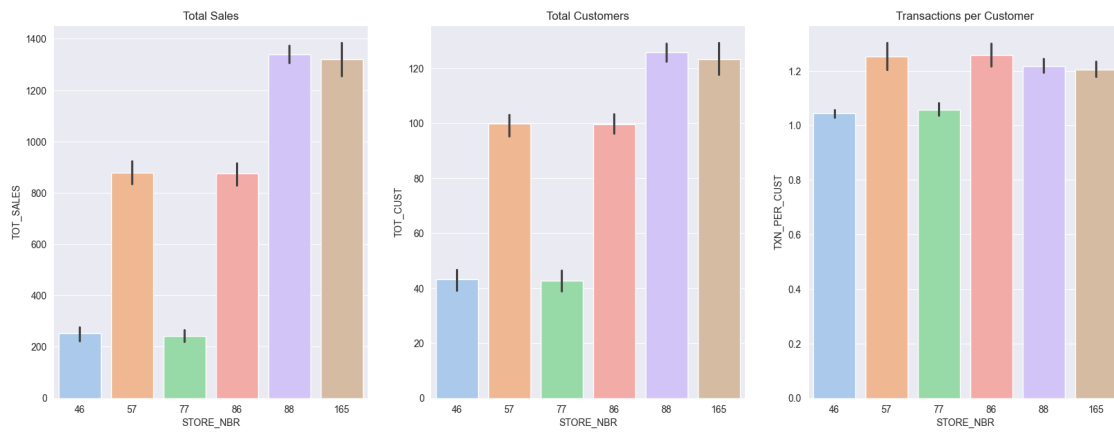
```
[25]:
```

	index	STORE_NBR	YEAR_MONTH	TOT_SALES	TOT_CUST	TXN_PER_CUST
0	504	77	2018-07	296.80	51	1.078431
1	505	77	2018-08	255.50	47	1.021277
2	506	77	2018-09	225.20	42	1.047619
3	507	77	2018-10	204.50	37	1.027027
4	508	77	2018-11	245.30	41	1.073171
5	509	77	2018-12	267.30	46	1.043478
6	510	77	2019-01	204.40	35	1.114286
0	560	86	2018-07	892.20	99	1.272727
1	561	86	2018-08	764.05	94	1.170213
2	562	86	2018-09	914.60	103	1.242718
3	563	86	2018-10	948.40	109	1.266055
4	564	86	2018-11	918.00	100	1.250000
5	565	86	2018-12	841.20	98	1.224490
6	566	86	2019-01	841.40	94	1.372340
0	574	88	2018-07	1310.00	129	1.186047
1	575	88	2018-08	1323.80	131	1.206107

2	576	88	2018-09	1423.00	124	1.266129
3	577	88	2018-10	1352.40	123	1.260163
4	578	88	2018-11	1382.80	130	1.200000
5	579	88	2018-12	1325.20	126	1.174603
6	580	88	2019-01	1266.40	117	1.230769
0	294	46	2018-07	253.00	45	1.066667
1	295	46	2018-08	240.70	44	1.045455
2	296	46	2018-09	233.00	41	1.048780
3	297	46	2018-10	275.10	47	1.042553
4	298	46	2018-11	273.10	42	1.047619
5	299	46	2018-12	306.90	50	1.060000
6	300	46	2019-01	176.20	33	1.000000
0	371	57	2018-07	839.60	103	1.203883
1	372	57	2018-08	915.40	102	1.274510
2	373	57	2018-09	792.80	99	1.171717
3	374	57	2018-10	965.80	104	1.307692
4	375	57	2018-11	830.00	100	1.170000
5	376	57	2018-12	951.00	104	1.259615
6	377	57	2019-01	852.80	87	1.379310
0	1099	165	2018-07	1457.00	133	1.255639
1	1100	165	2018-08	1206.60	109	1.256881
2	1101	165	2018-09	1281.20	122	1.172131
3	1102	165	2018-10	1234.40	118	1.169492
4	1103	165	2018-11	1291.20	126	1.166667
5	1104	165	2018-12	1345.40	121	1.206612
6	1105	165	2019-01	1422.00	133	1.203008

```
[26]: sns.set_style("darkgrid")
figure, axis=plt.subplots(1, 3, figsize=(20, 7))
sns.barplot(x="STORE_NBR", y="TOT_SALES", data=stores, ax=axis[0],
           palette="pastel")
axis[0].set_title("Total Sales")
sns.barplot(x="STORE_NBR", y="TOT_CUST", data=stores, ax=axis[1],
           palette="pastel")
axis[1].set_title("Total Customers")
sns.barplot(x="STORE_NBR", y="TXN_PER_CUST", data=stores, ax=axis[2],
           palette="pastel")
axis[2].set_title("Transactions per Customer")
figure.suptitle("Comparison of the Total Sales, Total Customers, and
           Transactions per Customer for Each the Trial Stores and the Control Stores
           During the Pre-Trial Duration")
plt.show()
```

Comparison of the Total Sales, Total Customers, and Transactions per Customer for Each the Trial Stores and the Control Stores During the Pre-Trial Duration



[]: