

# EDA Using Python

July 19, 2024

```
[2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import getpass
import seaborn as sns
import snowflake.connector
```

## Connecting to Snowflake

```
[4]: conn = snowflake.connector.connect(
    user = 'DEB',
    password = getpass.getpass('Your Snowflake Password: '),
    ##password = 'Qwerty@12345',
    ## account = svfpjax-gg07076
    account = 'svfpjax-gg07076',
    database='RETAILS',
    schema='PUBLIC',
    warehouse='COMPUTE_WH',
)
```

Your Snowflake Password: .....

```
[5]: cur = conn.cursor()
```

```
[6]: select_demographic_RAW = 'SELECT * FROM demographic_RAW'
select_campaign_DESC_RAW = 'SELECT * FROM CAMPAIGN_DESC_RAW'
select_campaign_RAW = 'SELECT * FROM CAMPAIGN_RAW'
select_product_RAW = 'SELECT * FROM PRODUCT_RAW'
select_coupon_RAW = 'SELECT * FROM COUPON_RAW'
select_coupon_REDEMPT_RAW = 'SELECT * FROM COUPON_REDEMPT_RAW'
select_TRANSACTION_RAW = 'SELECT * FROM TRANSACTION_RAW'
```

```
[7]: cur.execute(select_demographic_RAW)
demographic_RAW = cur.fetch_pandas_all()
```

```
[8]: demographic_RAW.head(5)
```

```
[8]:  AGE_DESC  MARITAL_STATUS_CODE  INCOME_DESC  HOMEOWNER_DESC  HH_COMP_DESC  \
0      65+                      A      35-49K      Homeowner    2 Adults No Kids
```

1	45-54	A	50-74K	Homeowner	2 Adults No Kids
2	25-34	U	25-34K	Unknown	2 Adults Kids
3	25-34	U	75-99K	Homeowner	2 Adults Kids
4	45-54	B	50-74K	Homeowner	Single Female

	HOUSEHOLD_SIZE_DESC	KID_CATEGORY_DESC	HOUSEHOLD_KEY
0	2	None/Unknown	1
1	2	None/Unknown	7
2	3	1	8
3	4	2	13
4	1	None/Unknown	16

```
[9]: cur.execute(select_CAMPAIGN_DESC_RAW)
CAMPAIGN_DESC_RAW = cur.fetch_pandas_all()
```

```
[10]: CAMPAIGN_DESC_RAW.head(5)
```

```
[10]:
```

	DESCRIPTION	CAMPAIGN	START_DAY	END_DAY
0	TypeB	24	659	719
1	TypeC	15	547	708
2	TypeB	25	659	691
3	TypeC	20	615	685
4	TypeB	23	646	684

```
[11]: cur.execute(select_CAMPAIGN_RAW)
CAMPAIGN_RAW = cur.fetch_pandas_all()
```

```
[12]: CAMPAIGN_RAW.head(5)
```

```
[12]:
```

	DESCRIPTION	HOUSEHOLD_KEY	CAMPAIGN
0	TypeA	17	26
1	TypeA	27	26
2	TypeA	212	26
3	TypeA	208	26
4	TypeA	192	26

```
[13]: cur.execute(select_PRODUCT_RAW)
PRODUCT_RAW = cur.fetch_pandas_all()
```

```
[14]: PRODUCT_RAW.head(5)
```

```
[14]:
```

	PRODUCT_ID	MANUFACTURER	DEPARTMENT	BRAND	COMMODITY_DESC \
0	25671	2	GROCERY	National	FRZN ICE
1	26081	2	MISC. TRANS.	National	NO COMMODITY DESCRIPTION
2	26093	69	PASTRY	Private	BREAD
3	26190	69	GROCERY	Private	FRUIT - SHELF STABLE
4	26355	69	GROCERY	Private	COOKIES/CONES

	SUB_COMMODITY_DESC	CURR_SIZE_OF_PRODUCT
0	ICE - CRUSHED/CUBED	22 LB
1	NO SUBCOMMODITY DESCRIPTION	
2	BREAD:ITALIAN/FRENCH	
3	APPLE SAUCE	50 OZ
4	SPECIALTY COOKIES	14 OZ

```
[15]: cur.execute( select_COUPON_RAW)
COUPON_RAW = cur.fetch_pandas_all()
```

```
[16]: COUPON_RAW.head(5)
```

```
[16]:
```

	COUPON_UPC	PRODUCT_ID	CAMPAIGN
0	10000089061	27160	4
1	10000089064	27754	9
2	10000089073	28897	12
3	51800009050	28919	28
4	52100000076	28929	25

```
[17]: cur.execute(select_COUPON_REDEMPT_RAW)
COUPON_REDEMPT_RAW = cur.fetch_pandas_all()
```

```
[18]: COUPON_REDEMPT_RAW.head(5)
```

```
[18]:
```

	HOUSEHOLD_KEY	DAY	COUPON_UPC	CAMPAIGN
0	1	421	10000085364	8
1	1	421	51700010076	8
2	1	427	54200000033	8
3	1	597	10000085476	18
4	1	597	54200029176	18

```
[19]: cur.execute(select_TRANSACTION_RAW)
TRANSACTION_RAW = cur.fetch_pandas_all()
```

```
[20]: TRANSACTION_RAW.head(5)
```

```
[20]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE	\
0	2375	26984851472	1	1004906	1	1.39	
1	2375	26984851472	1	1033142	1	0.82	
2	2375	26984851472	1	1036325	1	0.99	
3	2375	26984851472	1	1082185	1	1.21	
4	2375	26984851472	1	8160430	1	1.50	

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC	COUPON_MATCH_DISC
0	364	-0.60	1631	1	0	0
1	364	0.00	1631	1	0	0
2	364	-0.30	1631	1	0	0
3	364	0.00	1631	1	0	0

4	364	-0.39	1631	1	0	0
---	-----	-------	------	---	---	---

```
[21]: conn.close()
      cur.close()
```

[21]: False

```
[22]: demographic_RAW.isnull().sum()
```

```
[22]: AGE_DESC          0
      MARITAL_STATUS_CODE  0
      INCOME_DESC       0
      HOMEOWNER_DESC    0
      HH_COMP_DESC      0
      HOUSEHOLD_SIZE_DESC 0
      KID_CATEGORY_DESC  0
      HOUSEHOLD_KEY      0
      dtype: int64
```

```
[23]: demographic_RAW.describe()
```

```
[23]: HOUSEHOLD_KEY
      count    2500.00000
      mean     1250.50000
      std      721.83216
      min       1.00000
      25%      625.75000
      50%     1250.50000
      75%     1875.25000
      max     2500.00000
```

```
[24]: CAMPAIGN_DESC_RAW.isnull().sum()
```

```
[24]: DESCRIPTION    0
      CAMPAIGN      0
      START_DAY     0
      END_DAY       0
      dtype: int64
```

```
[25]: CAMPAIGN_RAW.isnull().sum()
```

```
[25]: DESCRIPTION    0
      HOUSEHOLD_KEY  0
      CAMPAIGN      0
      dtype: int64
```

```
[26]: PRODUCT_RAW.isnull().sum()
```

```
[26]: PRODUCT_ID      0
      MANUFACTURER    0
      DEPARTMENT      0
      BRAND           0
      COMMODITY_DESC   0
      SUB_COMMODITY_DESC 0
      CURR_SIZE_OF_PRODUCT 0
      dtype: int64
```

```
[27]: COUPON_RAW.isnull().sum()
```

```
[27]: COUPON_UPC      0
      PRODUCT_ID     0
      CAMPAIGN       0
      dtype: int64
```

```
[28]: COUPON_REDEMPT_RAW.isnull().sum()
```

```
[28]: HOUSEHOLD_KEY    0
      DAY            0
      COUPON_UPC      0
      CAMPAIGN       0
      dtype: int64
```

```
[29]: TRANSACTION_RAW.isnull().sum()
```

```
[29]: HOUSEHOLD_KEY    0
      BASKET_ID       0
      DAY            0
      PRODUCT_ID     0
      QUANTITY        0
      SALES_VALUE     0
      STORE_ID        0
      RETAIL_DISC     0
      TRANS_TIME      0
      WEEK_NO         0
      COUPON_DISC     0
      COUPON_MATCH_DISC 0
      dtype: int64
```

### Data Modifications

```
[30]: from datetime import datetime, timedelta
```

```
[31]: start_date = pd.to_datetime('2019-01-01')
```

```
[32]: start_date
```

```
[32]: Timestamp('2019-01-01 00:00:00')
```

```
[33]: TRANSACTION_RAW.head(5)
```

```
[33]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE	\
0	2375	26984851472	1	1004906	1	1.39	
1	2375	26984851472	1	1033142	1	0.82	
2	2375	26984851472	1	1036325	1	0.99	
3	2375	26984851472	1	1082185	1	1.21	
4	2375	26984851472	1	8160430	1	1.50	

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC	COUPON_MATCH_DISC	\
0	364	-0.60	1631	1	0	0	
1	364	0.00	1631	1	0	0	
2	364	-0.30	1631	1	0	0	
3	364	0.00	1631	1	0	0	
4	364	-0.39	1631	1	0	0	

```
[34]: start_date = pd.to_datetime('2019-01-01')
```

```
[35]: TRANSACTION_RAW['DATE'] = start_date + pd.to_timedelta(TRANSACTION_RAW['DAY'],  
    ↪unit = 'D')
```

```
[36]: TRANSACTION_RAW.head(10)
```

```
[36]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE	\
0	2375	26984851472	1	1004906	1	1.39	
1	2375	26984851472	1	1033142	1	0.82	
2	2375	26984851472	1	1036325	1	0.99	
3	2375	26984851472	1	1082185	1	1.21	
4	2375	26984851472	1	8160430	1	1.50	
5	2375	26984851516	1	826249	2	1.98	
6	2375	26984851516	1	1043142	1	1.57	
7	2375	26984851516	1	1085983	1	2.99	
8	2375	26984851516	1	1102651	1	1.89	
9	2375	26984851516	1	6423775	1	2.00	

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC	COUPON_MATCH_DISC	\
0	364	-0.60	1631	1	0	0	
1	364	0.00	1631	1	0	0	
2	364	-0.30	1631	1	0	0	
3	364	0.00	1631	1	0	0	
4	364	-0.39	1631	1	0	0	
5	364	-0.60	1642	1	0	0	
6	364	-0.68	1642	1	0	0	
7	364	-0.40	1642	1	0	0	
8	364	0.00	1642	1	0	0	

9	364	-0.79	1642	1	0	0
---	-----	-------	------	---	---	---

```

DATE
0 2019-01-02
1 2019-01-02
2 2019-01-02
3 2019-01-02
4 2019-01-02
5 2019-01-02
6 2019-01-02
7 2019-01-02
8 2019-01-02
9 2019-01-02

```

```
[37]: CAMPAIGN_DESC_RAW.head(5)
```

```
[37]:
```

	DESCRIPTION	CAMPAIGN	START_DAY	END_DAY
0	TypeB	24	659	719
1	TypeC	15	547	708
2	TypeB	25	659	691
3	TypeC	20	615	685
4	TypeB	23	646	684

```
[38]: CAMPAIGN_DESC_RAW['Start_date'] = start_date + pd.
      ↪to_timedelta(CAMPAIGN_DESC_RAW['START_DAY'],unit='D')
```

```
[39]: CAMPAIGN_DESC_RAW['End_date']=start_date + pd.
      ↪to_timedelta(CAMPAIGN_DESC_RAW['END_DAY'],unit='D')
```

```
[40]: CAMPAIGN_DESC_RAW.head(10)
```

```
[40]:
```

	DESCRIPTION	CAMPAIGN	START_DAY	END_DAY	Start_date	End_date
0	TypeB	24	659	719	2020-10-21	2020-12-20
1	TypeC	15	547	708	2020-07-01	2020-12-09
2	TypeB	25	659	691	2020-10-21	2020-11-22
3	TypeC	20	615	685	2020-09-07	2020-11-16
4	TypeB	23	646	684	2020-10-08	2020-11-15
5	TypeB	21	624	656	2020-09-16	2020-10-18
6	TypeB	22	624	656	2020-09-16	2020-10-18
7	TypeA	18	587	642	2020-08-10	2020-10-04
8	TypeB	19	603	635	2020-08-26	2020-09-27
9	TypeB	17	575	607	2020-07-29	2020-08-30

```
[41]: CAMPAIGN_DESC_RAW['Campaign_Duration'] = CAMPAIGN_DESC_RAW['END_DAY'] -
      ↪CAMPAIGN_DESC_RAW['START_DAY']
```

```
[42]: CAMPAIGN_DESC_RAW.head(10)
```

```
[42]:
```

	DESCRIPTION	CAMPAIGN	START_DAY	END_DAY	Start_date	End_date	\
0	TypeB	24	659	719	2020-10-21	2020-12-20	
1	TypeC	15	547	708	2020-07-01	2020-12-09	
2	TypeB	25	659	691	2020-10-21	2020-11-22	
3	TypeC	20	615	685	2020-09-07	2020-11-16	
4	TypeB	23	646	684	2020-10-08	2020-11-15	
5	TypeB	21	624	656	2020-09-16	2020-10-18	
6	TypeB	22	624	656	2020-09-16	2020-10-18	
7	TypeA	18	587	642	2020-08-10	2020-10-04	
8	TypeB	19	603	635	2020-08-26	2020-09-27	
9	TypeB	17	575	607	2020-07-29	2020-08-30	

```

Campaign_Duration
0          60
1         161
2          32
3          70
4          38
5          32
6          32
7          55
8          32
9          32

```

```
[43]: COUPON_REDEMPT_RAW['Date'] = start_date + pd.
      ↪to_timedelta(COUPON_REDEMPT_RAW['DAY'],unit='D')
```

```
[44]: COUPON_REDEMPT_RAW.head(10)
```

```
[44]:
```

	HOUSEHOLD_KEY	DAY	COUPON_UPC	CAMPAIGN	Date
0	1	421	10000085364	8	2020-02-26
1	1	421	51700010076	8	2020-02-26
2	1	427	54200000033	8	2020-03-03
3	1	597	10000085476	18	2020-08-20
4	1	597	54200029176	18	2020-08-20
5	8	422	53600000078	8	2020-02-27
6	13	396	53700048182	5	2020-02-01
7	13	424	10000085364	8	2020-02-29
8	13	434	53600000078	8	2020-03-10
9	13	447	52370020076	8	2020-03-23

```
[45]: CAMPAIGN_DESC_RAW['Campaign_Duration'].mean()
```

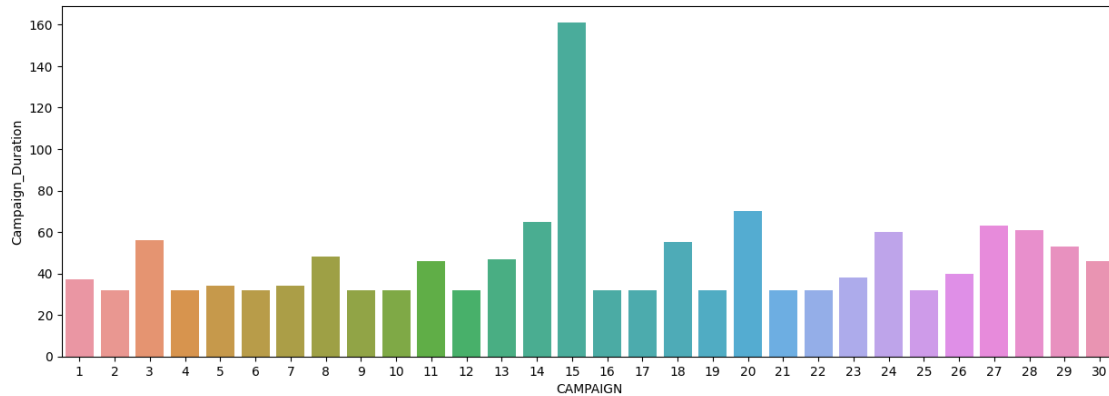
```
[45]: 46.6
```

The Average Campaign Duration is 46.6 days



```
[46]: plt.figure(figsize=(15,5))
sns.barplot(x='CAMPAIGN',y='Campaign_Duration',data = CAMPAIGN_DESC_RAW)
```

```
[46]: <Axes: xlabel='CAMPAIGN', ylabel='Campaign_Duration'>
```



### Campaign 15 Lasted more than 160 days

```
[47]: CAMPAIGN_DESC_RAW.groupby('DESCRIPTION').aggregate({'CAMPAIGN':
↳ 'count', 'Campaign_Duration': 'mean'})
```

```
[47]:
```

	CAMPAIGN	Campaign_Duration
DESCRIPTION		
TypeA	5	47.200000
TypeB	19	37.631579
TypeC	6	74.500000

There have been 19 type B campaigns, whose average duration was 38 days. In comparison, there has been 6 type C campaigns of 75 days on average.

```
[48]: CAMPAIGN_DESC_RAW['Start_month'] = CAMPAIGN_DESC_RAW['Start_date'].dt.
↳ strftime('%m')
```

```
[49]: CAMPAIGN_DESC_RAW['End_month'] = CAMPAIGN_DESC_RAW['End_date'].dt.strftime('%m')
```

```
[50]: CAMPAIGN_DESC_RAW['Start_Year'] = CAMPAIGN_DESC_RAW['Start_date'].dt.
↳ strftime('%Y')
```

```
[51]: CAMPAIGN_DESC_RAW['End_Year'] = CAMPAIGN_DESC_RAW['End_date'].dt.strftime('%Y')
```

```
[52]: CAMPAIGN_DESC_RAW.head(10)
```

```
[52]:
```

	DESCRIPTION	CAMPAIGN	START_DAY	END_DAY	Start_date	End_date	\
0	TypeB	24	659	719	2020-10-21	2020-12-20	
1	TypeC	15	547	708	2020-07-01	2020-12-09	

2	TypeB	25	659	691	2020-10-21	2020-11-22
3	TypeC	20	615	685	2020-09-07	2020-11-16
4	TypeB	23	646	684	2020-10-08	2020-11-15
5	TypeB	21	624	656	2020-09-16	2020-10-18
6	TypeB	22	624	656	2020-09-16	2020-10-18
7	TypeA	18	587	642	2020-08-10	2020-10-04
8	TypeB	19	603	635	2020-08-26	2020-09-27
9	TypeB	17	575	607	2020-07-29	2020-08-30

	Campaign_Duration	Start_month	End_month	Start_Year	End_Year
0	60	10	12	2020	2020
1	161	07	12	2020	2020
2	32	10	11	2020	2020
3	70	09	11	2020	2020
4	38	10	11	2020	2020
5	32	09	10	2020	2020
6	32	09	10	2020	2020
7	55	08	10	2020	2020
8	32	08	09	2020	2020
9	32	07	08	2020	2020

```
[53]: CAMPAIGN_RAW['HOUSEHOLD_KEY'].nunique()
```

```
[53]: 1584
```

```
[54]: TRANSACTION_RAW['HOUSEHOLD_KEY'].nunique()
```

```
[54]: 2494
```

There are 1584 households have participated to the campaign, leaving 910 households who never participated to any campaign.

```
[55]: CAMPAIGN_RAW.head(5)
```

```
[55]:  DESCRIPTION  HOUSEHOLD_KEY  CAMPAIGN
0      TypeA           17           26
1      TypeA           27           26
2      TypeA          212           26
3      TypeA          208           26
4      TypeA          192           26
```

```
[56]: CAMPAIGN_RAW.groupby('HOUSEHOLD_KEY').aggregate({'CAMPAIGN': 'count'})
```

```
[56]:           CAMPAIGN
HOUSEHOLD_KEY
1              8
2              1
3              3
```

```

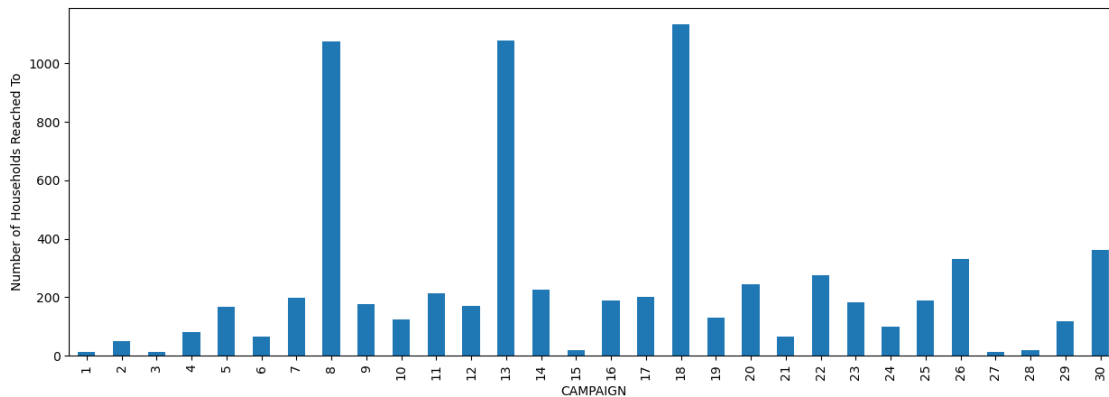
4          1
6          4
...
2496       5
2497       3
2498       6
2499       2
2500       8

```

[1584 rows x 1 columns]

```
[57]: plt.figure(figsize=(15,5))
      CAMPAIGN_RAW.groupby('CAMPAIGN')['HOUSEHOLD_KEY'].count().plot.bar()
      plt.ylabel('Number of Households Reached To')
```

```
[57]: Text(0, 0.5, 'Number of Households Reached To')
```



**In Campaign 18 maximum number of households are participated.**

```
[58]: COUPON_RAW.head(5)
```

```
[58]:
```

	COUPON_UPC	PRODUCT_ID	CAMPAIGN
0	10000089061	27160	4
1	10000089064	27754	9
2	10000089073	28897	12
3	51800009050	28919	28
4	52100000076	28929	25

```
[59]: Coupon_Given = COUPON_RAW.groupby('CAMPAIGN').
      ↪ aggregate(Total_product=('PRODUCT_ID', 'nunique'),
      ↪ Total_coupon=('COUPON_UPC', 'nunique'))

      Coupon_Given.head(10)
```

```
[59]:
```

	Total_product	Total_coupon
CAMPAIGN		
1	383	11
2	301	16
3	506	34
4	200	12
5	443	11
6	18	1
7	259	8
8	17178	209
9	825	18
10	393	14

```
[60]: Coupon_Given.sort_values(by='Total_product',ascending=False).head(10)
```

```
[60]:
```

	Total_product	Total_coupon
CAMPAIGN		
13	35616	207
18	35513	209
8	17178	209
26	3396	181
30	3205	181
22	1917	21
24	1887	2
16	1025	13
25	826	17
9	825	18

```
[61]: CAMPAIGN_DESC_RAW.head(5)
```

```
[61]:
```

	DESCRIPTION	CAMPAIGN	START_DAY	END_DAY	Start_date	End_date	\
0	TypeB	24	659	719	2020-10-21	2020-12-20	
1	TypeC	15	547	708	2020-07-01	2020-12-09	
2	TypeB	25	659	691	2020-10-21	2020-11-22	
3	TypeC	20	615	685	2020-09-07	2020-11-16	
4	TypeB	23	646	684	2020-10-08	2020-11-15	

	Campaign_Duration	Start_month	End_month	Start_Year	End_Year
0	60	10	12	2020	2020
1	161	07	12	2020	2020
2	32	10	11	2020	2020
3	70	09	11	2020	2020
4	38	10	11	2020	2020

```
[62]: Coupon_Given = Coupon_Given.merge(right =CAMPAIGN_DESC_RAW,on='CAMPAIGN',how='left')
```

```
[63]: Coupon_Given.head(5)
```

```
[63]:
```

	CAMPAIGN	Total_product	Total_coupon	DESCRIPTION	START_DAY	END_DAY	\
0	1	383	11	TypeB	346	383	
1	2	301	16	TypeB	351	383	
2	3	506	34	TypeC	356	412	
3	4	200	12	TypeB	372	404	
4	5	443	11	TypeB	377	411	

	Start_date	End_date	Campaign_Duration	Start_month	End_month	Start_Year	\
0	2019-12-13	2020-01-19	37	12	01	2019	
1	2019-12-18	2020-01-19	32	12	01	2019	
2	2019-12-23	2020-02-17	56	12	02	2019	
3	2020-01-08	2020-02-09	32	01	02	2020	
4	2020-01-13	2020-02-16	34	01	02	2020	

	End_Year
0	2020
1	2020
2	2020
3	2020
4	2020

```
[64]: Coupon_Given.columns
```

```
[64]: Index(['CAMPAIGN', 'Total_product', 'Total_coupon', 'DESCRIPTION', 'START_DAY',
            'END_DAY', 'Start_date', 'End_date', 'Campaign_Duration', 'Start_month',
            'End_month', 'Start_Year', 'End_Year'],
            dtype='object')
```

```
[65]: Coupon_Given.loc[:
↪, ('CAMPAIGN', 'Total_product', 'Total_coupon', 'Start_date', 'End_date', 'DESCRIPTION', 'Campaign
↪sort_values(by='Total_product',
              ascending=False).head(10)
```

```
[65]:
```

	CAMPAIGN	Total_product	Total_coupon	Start_date	End_date	DESCRIPTION	\
12	13	35616	207	2020-05-19	2020-07-05	TypeA	
17	18	35513	209	2020-08-10	2020-10-04	TypeA	
7	8	17178	209	2020-02-17	2020-04-05	TypeA	
25	26	3396	181	2019-08-13	2019-09-22	TypeA	
29	30	3205	181	2019-11-20	2020-01-05	TypeA	
21	22	1917	21	2020-09-16	2020-10-18	TypeB	
23	24	1887	2	2020-10-21	2020-12-20	TypeB	
15	16	1025	13	2020-07-15	2020-08-16	TypeB	
24	25	826	17	2020-10-21	2020-11-22	TypeB	
8	9	825	18	2020-03-11	2020-04-12	TypeB	

	Campaign_Duration
12	47

17	55
7	48
25	40
29	46
21	32
23	60
15	32
24	32
8	32

campaign 13,18,8 are the one with most product in them.

```
[66]: COUPON_RAW.head(5)
```

```
[66]:   COUPON_UPC  PRODUCT_ID  CAMPAIGN
0  10000089061      27160         4
1  10000089064      27754         9
2  10000089073      28897        12
3  51800009050      28919        28
4  52100000076      28929        25
```

```
[67]: PRODUCT_RAW.head(5)
```

```
[67]:   PRODUCT_ID  MANUFACTURER  DEPARTMENT  BRAND  COMMODITY_DESC \
0      25671           2      GROCERY  National      FRZN ICE
1      26081           2  MISC. TRANS.  National  NO COMMODITY DESCRIPTION
2      26093          69      PASTRY  Private      BREAD
3      26190          69      GROCERY  Private  FRUIT - SHELF STABLE
4      26355          69      GROCERY  Private  COOKIES/CONES

      SUB_COMMODITY_DESC  CURR_SIZE_OF_PRODUCT
0      ICE - CRUSHED/CUBED      22 LB
1  NO SUBCOMMODITY DESCRIPTION
2      BREAD:ITALIAN/FRENCH
3      APPLE SAUCE      50 OZ
4      SPECIALTY COOKIES      14 OZ
```

```
[68]: coupon_product = COUPON_RAW.merge(right=PRODUCT_RAW,on='PRODUCT_ID',how='left')
```

```
[69]: coupon_product.head(5)
```

```
[69]:   COUPON_UPC  PRODUCT_ID  CAMPAIGN  MANUFACTURER  DEPARTMENT  BRAND \
0  10000089061      27160         4           69      GROCERY  Private
1  10000089064      27754         9           69      GROCERY  Private
2  10000089073      28897        12           69      GROCERY  Private
3  51800009050      28919        28          236      GROCERY  National
4  52100000076      28929        25          317      GROCERY  National
```

	COMMODITY_DESC	SUB_COMMODITY_DESC \
0	SHORTENING/OIL	VEGETABLE/SALAD OIL
1	FRZN VEGETABLE/VEG DSH	FRZN BAGGED VEGETABLES - PLAIN
2	EGGS	EGGS - X-LARGE
3	REFRGRATD DOUGH PRODUCTS	REFRIGERATED COOKIES-CHUB
4	CHEESE	STRING CHEESE

	CURR_SIZE_OF_PRODUCT
0	48 OZ
1	16 OZ
2	A D 1 DZ
3	16.5 OZ
4	AB 8 OZ

**Top 10 product on which the Coupon has been applied**

```
[70]: coupon_product['COMMODITY_DESC'].value_counts().head(10)
```

```
[70]: COMMODITY_DESC
HAIR CARE PRODUCTS      6261
MAKEUP AND TREATMENT    5585
FRZN MEAT/MEAT DINNERS  4893
FROZEN PIZZA            3658
BEEF                    3314
ICE CREAM/MILK/SHERBTS  3257
CHEESE                  3146
YOGURT                  2460
ORAL HYGIENE PRODUCTS   2437
LUNCHMEAT               2425
Name: count, dtype: int64
```

**Most prominent products among coupons created are bathroom products such as hair care and makeup.**

```
[71]: COUPON_REDEMPT_RAW.head(10)
```

```
[71]:  HOUSEHOLD_KEY  DAY  COUPON_UPC  CAMPAIGN      Date
0             1  421  10000085364         8  2020-02-26
1             1  421  51700010076         8  2020-02-26
2             1  427  54200000033         8  2020-03-03
3             1  597  10000085476        18  2020-08-20
4             1  597  54200029176        18  2020-08-20
5             8  422  53600000078         8  2020-02-27
6            13  396  53700048182         5  2020-02-01
7            13  424  10000085364         8  2020-02-29
8            13  434  53600000078         8  2020-03-10
9            13  447  52370020076         8  2020-03-23
```

```
[72]: COUPON_RAW.head(5)
```

```
[72]:
```

	COUPON_UPC	PRODUCT_ID	CAMPAIGN
0	10000089061	27160	4
1	10000089064	27754	9
2	10000089073	28897	12
3	51800009050	28919	28
4	52100000076	28929	25

```
[73]: COUPON_RAW['COUPON_UPC'].nunique()
```

```
[73]: 1135
```

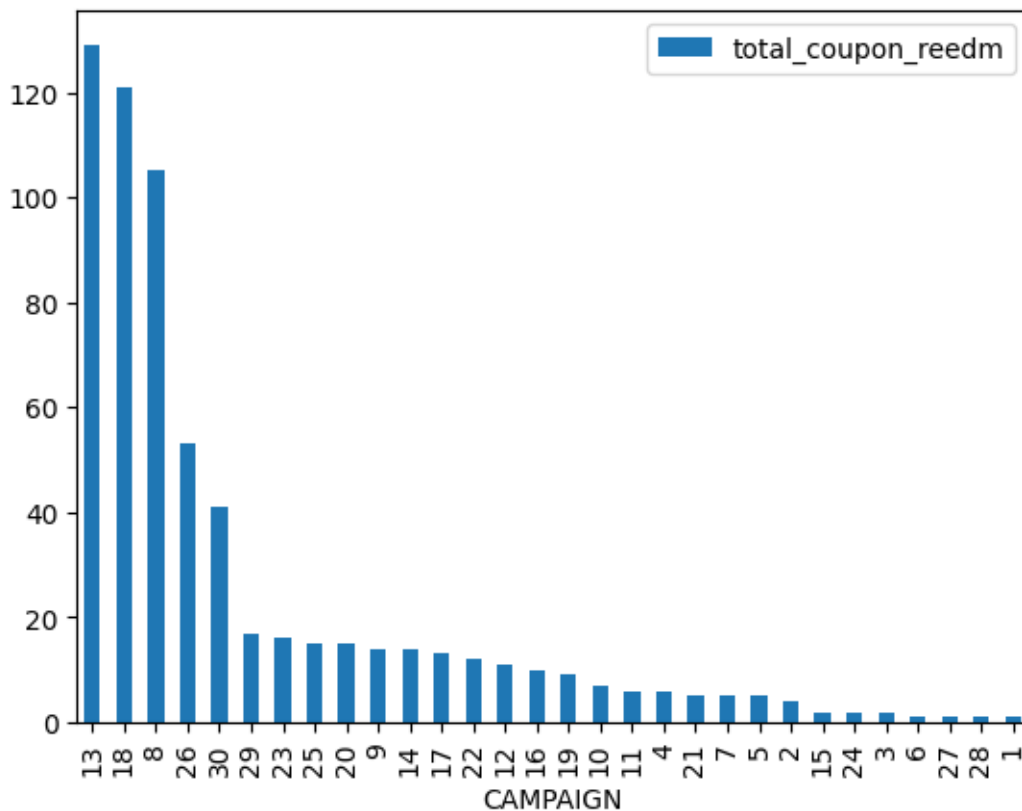
```
[74]: COUPON_REDEMPT_RAW['COUPON_UPC'].nunique()
```

```
[74]: 556
```

**Only 556 Coupon has been used out of 1135**

```
[75]: COUPON_REDEMPT_RAW.groupby('CAMPAIGN').  
      ↪agg(total_coupon_reedm=('COUPON_UPC', 'nunique')).  
      ↪sort_values(by='total_coupon_reedm', ascending=False).plot.bar()
```

```
[75]: <Axes: xlabel='CAMPAIGN'>
```





```
[76]: Coupon_redeem = COUPON_REDEMPT_RAW.groupby('CAMPAIGN').
      ↪agg(total_coupon_reedm=('COUPON_UPC','nunique'))
```

```
Coupon_redeem
```

```
[76]:
```

	total_coupon_reedm
CAMPAIGN	
1	1
2	4
3	2
4	6
5	5
6	1
7	5
8	105
9	14
10	7
11	6
12	11
13	129
14	14
15	2
16	10
17	13
18	121
19	9
20	15
21	5
22	12
23	16
24	2
25	15
26	53
27	1
28	1
29	17
30	41

```
[77]: Coupon_redeem.sort_values(by = 'total_coupon_reedm',ascending=False).head(10)
```

```
[77]:
```

	total_coupon_reedm
CAMPAIGN	
13	129
18	121
8	105

26	53
30	41
29	17
23	16
25	15
20	15
9	14

The most coupon has been used in campaign 13, 18, 8

```
[78]: Coupon_redeem = Coupon_redeem.merge(right=Coupon_Given, on='CAMPAIGN',
      ↪how='left')
```

```
[79]: Coupon_redeem.columns
```

```
[79]: Index(['CAMPAIGN', 'total_coupon_reedm', 'Total_product', 'Total_coupon',
      'DESCRIPTION', 'START_DAY', 'END_DAY', 'Start_date', 'End_date',
      'Campaign_Duration', 'Start_month', 'End_month', 'Start_Year',
      'End_Year'],
      dtype='object')
```

```
[80]: Coupon_redeem.head(10)
```

```
[80]:
```

	CAMPAIGN	total_coupon_reedm	Total_product	Total_coupon	DESCRIPTION	\
0	1	1	383	11	TypeB	
1	2	4	301	16	TypeB	
2	3	2	506	34	TypeC	
3	4	6	200	12	TypeB	
4	5	5	443	11	TypeB	
5	6	1	18	1	TypeC	
6	7	5	259	8	TypeB	
7	8	105	17178	209	TypeA	
8	9	14	825	18	TypeB	
9	10	7	393	14	TypeB	

	START_DAY	END_DAY	Start_date	End_date	Campaign_Duration	Start_month	\
0	346	383	2019-12-13	2020-01-19	37	12	
1	351	383	2019-12-18	2020-01-19	32	12	
2	356	412	2019-12-23	2020-02-17	56	12	
3	372	404	2020-01-08	2020-02-09	32	01	
4	377	411	2020-01-13	2020-02-16	34	01	
5	393	425	2020-01-29	2020-03-01	32	01	
6	398	432	2020-02-03	2020-03-08	34	02	
7	412	460	2020-02-17	2020-04-05	48	02	
8	435	467	2020-03-11	2020-04-12	32	03	
9	463	495	2020-04-08	2020-05-10	32	04	

End\_month Start\_Year End\_Year

0	01	2019	2020
1	01	2019	2020
2	02	2019	2020
3	02	2020	2020
4	02	2020	2020
5	03	2020	2020
6	03	2020	2020
7	04	2020	2020
8	04	2020	2020
9	05	2020	2020

```
[81]: Coupon_redeem['Coupon_redeem_rate']=(Coupon_redeem['total_coupon_reedm']/
↳Coupon_redeem['Total_coupon'])*100
```

```
[82]: Coupon_redeem.head().sort_values(by = 'Coupon_redeem_rate',ascending = False)
```

```
[82]:
```

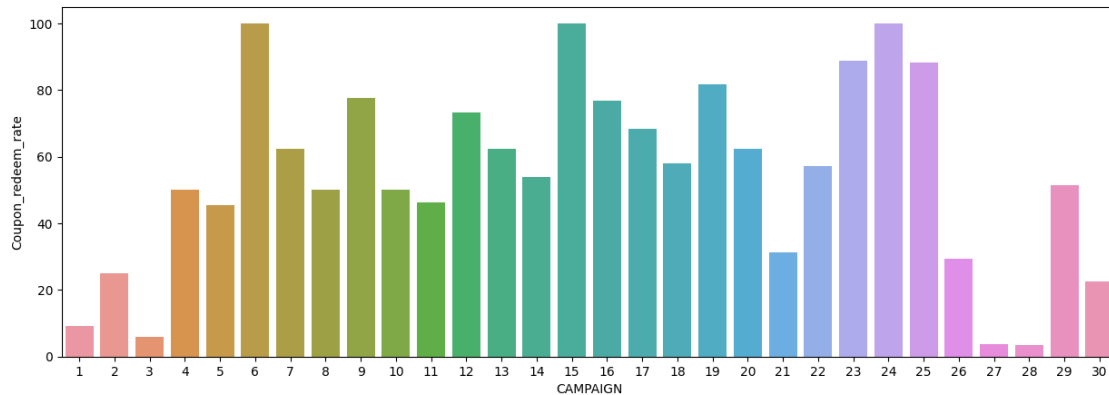
	CAMPAIGN	total_coupon_reedm	Total_product	Total_coupon	DESCRIPTION	\
3	4	6	200	12	TypeB	
4	5	5	443	11	TypeB	
1	2	4	301	16	TypeB	
0	1	1	383	11	TypeB	
2	3	2	506	34	TypeC	

	START_DAY	END_DAY	Start_date	End_date	Campaign_Duration	Start_month	\
3	372	404	2020-01-08	2020-02-09	32	01	
4	377	411	2020-01-13	2020-02-16	34	01	
1	351	383	2019-12-18	2020-01-19	32	12	
0	346	383	2019-12-13	2020-01-19	37	12	
2	356	412	2019-12-23	2020-02-17	56	12	

	End_month	Start_Year	End_Year	Coupon_redeem_rate
3	02	2020	2020	50.000000
4	02	2020	2020	45.454545
1	01	2019	2020	25.000000
0	01	2019	2020	9.090909
2	02	2019	2020	5.882353

```
[83]: plt.figure(figsize=(15,5))
sns.barplot(x='CAMPAIGN',y='Coupon_redeem_rate',data=Coupon_redeem)
```

```
[83]: <Axes: xlabel='CAMPAIGN', ylabel='Coupon_redeem_rate'>
```



```
[84]: trnx_bucket = TRANSACTION_RAW.groupby('BASKET_ID').aggregate({'SALES_VALUE':
    ↪ 'sum', 'COUPON_DISC': 'sum', 'COUPON_MATCH_DISC': 'sum'})

trnx_bucket.head(10)
```

```
[84]:
```

BASKET_ID	SALES_VALUE	COUPON_DISC	COUPON_MATCH_DISC
26984851472	5.91	0	0
26984851516	12.43	0	0
26984896261	11.37	0	0
26984905972	2.55	0	0
26984945254	3.53	0	0
26984951769	5.79	0	0
26985025264	44.78	0	0
26985040735	139.22	0	0
26985052379	0.99	0	0
26985165432	0.00	0	0

```
[85]: trnx_bucket['Use_coupon'] = trnx_bucket['COUPON_DISC']!=0
```

```
[86]: trnx_bucket.head(10)
```

```
[86]:
```

BASKET_ID	SALES_VALUE	COUPON_DISC	COUPON_MATCH_DISC	Use_coupon
26984851472	5.91	0	0	False
26984851516	12.43	0	0	False
26984896261	11.37	0	0	False
26984905972	2.55	0	0	False
26984945254	3.53	0	0	False
26984951769	5.79	0	0	False
26985025264	44.78	0	0	False
26985040735	139.22	0	0	False
26985052379	0.99	0	0	False

26985165432	0.00	0	0	False
-------------	------	---	---	-------

```
[87]: trnx_bucket.sort_values('SALES_VALUE',ascending=False).head(10)
```

```
[87]:
```

	SALES_VALUE	COUPON_DISC	COUPON_MATCH_DISC	Use_coupon
BASKET_ID				
27534431747	465.70	-3	-1	True
27798435271	463.17	-2	-2	True
27093315740	400.79	0	0	False
28941937397	400.56	-1	0	True
28210676830	391.05	0	0	False
27950470466	376.38	-2	0	True
29034966796	369.48	0	0	False
28919755135	368.95	0	0	False
28675556859	365.86	-3	0	True
29221810662	358.98	0	0	False

```
[88]: trnx_bucket['Use_coupon'].value_counts()
```

```
[88]: Use_coupon
False    47892
True     2587
Name: count, dtype: int64
```

**Total coupon used 2587**

```
[89]: round(trnx_bucket['SALES_VALUE'].mean(),2)
```

```
[89]: 26.93
```

The average basket value is \$27

```
[90]: trnx_bucket.groupby('Use_coupon').aggregate( sales_mean=('SALES_VALUE','mean'),
                                                    COUPON_DISC_mean=
                                                    ↳
↳('COUPON_DISC','mean'),
                                                    ↳
↳COUPON_MATCH_DISC=('COUPON_MATCH_DISC','mean'),
                                                    No_coupon ↳
↳('SALES_VALUE','count'))
```

```
[90]:
```

	sales_mean	COUPON_DISC_mean	COUPON_MATCH_DISC	No_coupon
Use_coupon				
False	25.009980	0.000000	0.000000	47892
True	62.533348	-2.587167	-0.334364	2587

**The average basket value without a coupon is \$25.**

**The average basket value with a coupon is \$63.**

**The average discount generated by coupons is \$3.**

*It shows that customeer purchase more product when coupon is given to them*

```
[91]: trnx_desc = TRANSACTION_RAW.merge(right= trnx_bucket,on='BASKET_ID',how='left')
```

```
[92]: trnx_desc.head(10)
```

```
[92]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE_x	\
0	2375	26984851472	1	1004906	1	1.39	
1	2375	26984851472	1	1033142	1	0.82	
2	2375	26984851472	1	1036325	1	0.99	
3	2375	26984851472	1	1082185	1	1.21	
4	2375	26984851472	1	8160430	1	1.50	
5	2375	26984851516	1	826249	2	1.98	
6	2375	26984851516	1	1043142	1	1.57	
7	2375	26984851516	1	1085983	1	2.99	
8	2375	26984851516	1	1102651	1	1.89	
9	2375	26984851516	1	6423775	1	2.00	

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC_x	\
0	364	-0.60	1631	1	0	
1	364	0.00	1631	1	0	
2	364	-0.30	1631	1	0	
3	364	0.00	1631	1	0	
4	364	-0.39	1631	1	0	
5	364	-0.60	1642	1	0	
6	364	-0.68	1642	1	0	
7	364	-0.40	1642	1	0	
8	364	0.00	1642	1	0	
9	364	-0.79	1642	1	0	

	COUPON_MATCH_DISC_x	DATE	SALES_VALUE_y	COUPON_DISC_y	\
0	0	2019-01-02	5.91	0	
1	0	2019-01-02	5.91	0	
2	0	2019-01-02	5.91	0	
3	0	2019-01-02	5.91	0	
4	0	2019-01-02	5.91	0	
5	0	2019-01-02	12.43	0	
6	0	2019-01-02	12.43	0	
7	0	2019-01-02	12.43	0	
8	0	2019-01-02	12.43	0	
9	0	2019-01-02	12.43	0	

	COUPON_MATCH_DISC_y	Use_coupon
0	0	False
1	0	False
2	0	False
3	0	False

4	0	False
5	0	False
6	0	False
7	0	False
8	0	False
9	0	False

```
[93]: trnx_desc.columns
```

```
[93]: Index(['HOUSEHOLD_KEY', 'BASKET_ID', 'DAY', 'PRODUCT_ID', 'QUANTITY',
        'SALES_VALUE_x', 'STORE_ID', 'RETAIL_DISC', 'TRANS_TIME', 'WEEK_NO',
        'COUPON_DISC_x', 'COUPON_MATCH_DISC_x', 'DATE', 'SALES_VALUE_y',
        'COUPON_DISC_y', 'COUPON_MATCH_DISC_y', 'Use_coupon'],
        dtype='object')
```

```
[94]: trnx_desc.
      ↪drop(['SALES_VALUE_y', 'COUPON_DISC_y', 'COUPON_MATCH_DISC_y'],axis=1,inplace=True)
```

```
[95]: trnx_desc.head(10)
```

```
[95]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE_x	\
0	2375	26984851472	1	1004906	1	1.39	
1	2375	26984851472	1	1033142	1	0.82	
2	2375	26984851472	1	1036325	1	0.99	
3	2375	26984851472	1	1082185	1	1.21	
4	2375	26984851472	1	8160430	1	1.50	
5	2375	26984851516	1	826249	2	1.98	
6	2375	26984851516	1	1043142	1	1.57	
7	2375	26984851516	1	1085983	1	2.99	
8	2375	26984851516	1	1102651	1	1.89	
9	2375	26984851516	1	6423775	1	2.00	

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC_x	\
0	364	-0.60	1631	1	0	
1	364	0.00	1631	1	0	
2	364	-0.30	1631	1	0	
3	364	0.00	1631	1	0	
4	364	-0.39	1631	1	0	
5	364	-0.60	1642	1	0	
6	364	-0.68	1642	1	0	
7	364	-0.40	1642	1	0	
8	364	0.00	1642	1	0	
9	364	-0.79	1642	1	0	

	COUPON_MATCH_DISC_x	DATE	Use_coupon
0	0	2019-01-02	False
1	0	2019-01-02	False

2	0	2019-01-02	False
3	0	2019-01-02	False
4	0	2019-01-02	False
5	0	2019-01-02	False
6	0	2019-01-02	False
7	0	2019-01-02	False
8	0	2019-01-02	False
9	0	2019-01-02	False

```
[96]: trnx_desc= trnx_desc.merge(right=PRODUCT_RAW,on='PRODUCT_ID',how='left')

trnx_desc.head(10)
```

```
[96]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE_x \
0	2375	26984851472	1	1004906	1	1.39
1	2375	26984851472	1	1033142	1	0.82
2	2375	26984851472	1	1036325	1	0.99
3	2375	26984851472	1	1082185	1	1.21
4	2375	26984851472	1	8160430	1	1.50
5	2375	26984851516	1	826249	2	1.98
6	2375	26984851516	1	1043142	1	1.57
7	2375	26984851516	1	1085983	1	2.99
8	2375	26984851516	1	1102651	1	1.89
9	2375	26984851516	1	6423775	1	2.00

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC_x \
0	364	-0.60	1631	1	0
1	364	0.00	1631	1	0
2	364	-0.30	1631	1	0
3	364	0.00	1631	1	0
4	364	-0.39	1631	1	0
5	364	-0.60	1642	1	0
6	364	-0.68	1642	1	0
7	364	-0.40	1642	1	0
8	364	0.00	1642	1	0
9	364	-0.79	1642	1	0

	COUPON_MATCH_DISC_x	DATE	Use_coupon	MANUFACTURER	DEPARTMENT \
0	0	2019-01-02	False	69	PRODUCE
1	0	2019-01-02	False	2	PRODUCE
2	0	2019-01-02	False	69	PRODUCE
3	0	2019-01-02	False	2	PRODUCE
4	0	2019-01-02	False	69	PRODUCE
5	0	2019-01-02	False	69	GROCERY
6	0	2019-01-02	False	321	DRUG GM
7	0	2019-01-02	False	586	GROCERY
8	0	2019-01-02	False	1266	GROCERY



9 0 2019-01-02 False 586 GROCERY

	BRAND	COMMODITY_DESC	SUB_COMMODITY_DESC \
0	Private	POTATOES	POTATOES RUSSET (BULK&BAG)
1	National	ONIONS	ONIONS SWEET (BULK&BAG)
2	Private	VEGETABLES - ALL OTHERS	CELERY
3	National	TROPICAL FRUIT	BANANAS
4	Private	ORGANICS FRUIT & VEGETABLES	ORGANIC CARROTS
5	Private	BAKED BREAD/BUNS/ROLLS	HAMBURGER BUNS
6	National	BROOMS AND MOPS	SPONGES: BATH HOUSEHOLD
7	National	COOKIES/CONES	TRAY PACK/CHOC CHIP COOKIES
8	National	PNT BTR/JELLY/JAMS	PEANUT BUTTER
9	National	COOKIES/CONES	GRAHAM CRACKERS

	CURR_SIZE_OF_PRODUCT
0	5 LB
1	40 LB
2	
3	40 LB
4	1 LB
5	12 OZ
6	
7	18 OZ
8	12 OZ
9	11 OZ

```
[97]: COMMODITY_Coupon = trnx_desc.groupby('COMMODITY_DESC').
      ↪aggregate(total_quantity=('QUANTITY', 'count'),
               Use_coupon=('Use_coupon', 'sum'),
               Coupon=('COUPON_DISC_x', 'sum'))
```

```
[98]: COMMODITY_Coupon.head(10)
```

```
[98]:
```

COMMODITY_DESC	total_quantity	Use_coupon	Coupon
	1303	872	-1396
(CORP USE ONLY)	12	0	0
ADULT INCONTINENCE	74	3	-1
AIR CARE	928	193	-124
ANALGESICS	681	89	-3
ANTACIDS	307	36	-7
APPAREL	138	11	0
APPLES	2001	287	0
AUDIO/VIDEO PRODUCTS	149	16	-40
AUTOMOTIVE PRODUCTS	216	25	-6

```
[99]: COMMODITY_Coupon['Coupon%']=round((COMMODITY_Coupon['Use_coupon']/
↪COMMODITY_Coupon['total_quantity'])*100,2)
```

```
[100]: COMMODITY_Coupon.sort_values('Coupon%',ascending=False).head(30)
```

```
[100]:
```

	total_quantity	Use_coupon	Coupon	Coupon%
COMMODITY_DESC				
NATURAL VITAMINS	1	1	-5	100.00
	1303	872	-1396	66.92
VALENTINE	5	2	0	40.00
BABY FOODS	3664	1114	-38	30.40
RW FRESH PROCESSED MEAT	7	2	0	28.57
DIAPERS & DISPOSABLES	864	231	-215	26.74
BABY HBC	817	181	-82	22.15
AIR CARE	928	193	-124	20.80
FRZN BREAKFAST FOODS	1743	347	-61	19.91
CAT LITTER	531	104	-11	19.59
FACIAL TISS/DNR NAPKIN	1237	242	-29	19.56
GLASSWARE & DINNERWARE	36	7	0	19.44
LAUNDRY ADDITIVES	921	176	-25	19.11
FRUIT - SHELF STABLE	3267	611	-19	18.70
VITAMINS	334	61	-18	18.26
HOUSEHOLD CLEANG NEEDS	1962	355	-96	18.09
FD WRAPS/BAGS/TRSH BG	2358	416	-48	17.64
ORAL HYGIENE PRODUCTS	1715	302	-104	17.61
FROZEN - BOXED(GROCERY)	127	22	-14	17.32
FROZEN BREAD/DOUGH	1241	214	-62	17.24
DISHWASH DETERGENTS	1198	206	-43	17.20
PICKLE/RELISH/PKLD VEG	1350	232	-69	17.19
HAIR CARE PRODUCTS	2119	363	-79	17.13
DOG FOODS	1923	326	-64	16.95
TEAS	730	123	-19	16.85
RICE CAKES	250	42	0	16.80
CEREAL/BREAKFAST	304	51	-10	16.78
PAPER TOWELS	1775	296	-18	16.68
BATH TISSUES	2410	397	-125	16.47
CONVENIENT BRKFST/WHLSM SNACKS	3878	638	-90	16.45

```
[101]: TRANSACTION_RAW.groupby(['HOUSEHOLD_KEY','WEEK_NO','DAY']).
↪aggregate({'SALES_VALUE':'sum','RETAIL_DISC':'sum',
                                                    'COUPON_DISC'↪
↪:'sum','COUPON_MATCH_DISC':'sum'})
```

```
[101]:
```

	SALES_VALUE	RETAIL_DISC	COUPON_DISC	\
HOUSEHOLD_KEY WEEK_NO DAY				
1 8 51	78.66	-16.54	-1	
10 67	41.10	-8.59	0	

	13	88	26.90	-6.72	0
	14	94	63.43	-11.08	-2
	15	101	53.45	-16.42	0
...			...	...	...
2500	21	145	50.23	-2.02	-3
	22	150	55.03	-2.17	0
	26	174	110.30	-49.05	-1
		175	10.59	0.00	-2
		176	20.26	-2.51	0

COUPON_MATCH_DISC			
HOUSEHOLD_KEY	WEEK_NO	DAY	
1	8	51	0
	10	67	0
	13	88	0
	14	94	-1
	15	101	0
...			...
2500	21	145	0
	22	150	0
	26	174	0
		175	0
		176	0

[40506 rows x 4 columns]

```
[102]: TRANSACTION_RAW.head(5)
```

```
[102]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE	\
0	2375	26984851472	1	1004906	1	1.39	
1	2375	26984851472	1	1033142	1	0.82	
2	2375	26984851472	1	1036325	1	0.99	
3	2375	26984851472	1	1082185	1	1.21	
4	2375	26984851472	1	8160430	1	1.50	

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC	COUPON_MATCH_DISC	\
0	364	-0.60	1631	1	0	0	
1	364	0.00	1631	1	0	0	
2	364	-0.30	1631	1	0	0	
3	364	0.00	1631	1	0	0	
4	364	-0.39	1631	1	0	0	

	DATE
0	2019-01-02
1	2019-01-02
2	2019-01-02
3	2019-01-02

4 2019-01-02

```
[103]: TRANSACTION_RAW.groupby(TRANSACTION_RAW['DATE'].dt.year).  
        ↪aggregate({'SALES_VALUE':'sum','RETAIL_DISC':'sum',  
        ↪'COUPON_DISC':'sum'})
```

```
[103]:      SALES_VALUE  RETAIL_DISC  COUPON_DISC  
DATE  
2019    1359551.73    -248978.11         -6693
```

Sales Value increases as Retail Discount and Coupon Discount increases

Dropping the columns

```
[104]: from datetime import datetime
```

```
[105]: CAMPAIGN_DESC_RAW.head(5)
```

```
[105]:  DESCRIPTION  CAMPAIGN  START_DAY  END_DAY  Start_date  End_date  \  
0      TypeB      24      659      719 2020-10-21 2020-12-20  
1      TypeC      15      547      708 2020-07-01 2020-12-09  
2      TypeB      25      659      691 2020-10-21 2020-11-22  
3      TypeC      20      615      685 2020-09-07 2020-11-16  
4      TypeB      23      646      684 2020-10-08 2020-11-15  
  
      Campaign_Duration  Start_month  End_month  Start_Year  End_Year  
0              60          10          12          2020      2020  
1             161          07          12          2020      2020  
2              32          10          11          2020      2020  
3              70          09          11          2020      2020  
4              38          10          11          2020      2020
```

```
[106]: CAMPAIGN_DESC_RAW.drop(['START_DAY','END_DAY'],axis=1,inplace=True)
```

```
[107]: CAMPAIGN_DESC_RAW.head(5)
```

```
[107]:  DESCRIPTION  CAMPAIGN  Start_date  End_date  Campaign_Duration  Start_month  \  
0      TypeB      24 2020-10-21 2020-12-20              60          10  
1      TypeC      15 2020-07-01 2020-12-09             161          07  
2      TypeB      25 2020-10-21 2020-11-22              32          10  
3      TypeC      20 2020-09-07 2020-11-16              70          09  
4      TypeB      23 2020-10-08 2020-11-15              38          10  
  
      End_month  Start_Year  End_Year  
0          12      2020      2020  
1          12      2020      2020  
2          11      2020      2020
```

3	11	2020	2020
4	11	2020	2020

```
[108]: CAMPAIGN_DESC_RAW.head(5)
```

```
[108]: DESCRIPTION  CAMPAIGN  Start_date  End_date  Campaign_Duration  Start_month  \
0      TypeB      24  2020-10-21  2020-12-20      60      10
1      TypeC      15  2020-07-01  2020-12-09     161      07
2      TypeB      25  2020-10-21  2020-11-22      32      10
3      TypeC      20  2020-09-07  2020-11-16      70      09
4      TypeB      23  2020-10-08  2020-11-15      38      10

      End_month  Start_Year  End_Year
0      12      2020      2020
1      12      2020      2020
2      11      2020      2020
3      11      2020      2020
4      11      2020      2020
```

```
[109]: CAMPAIGN_DESC_RAW['Start_date'] = pd.
      ↪to_datetime(CAMPAIGN_DESC_RAW['Start_date']).apply(lambda x: x.date())
```

```
[110]: CAMPAIGN_DESC_RAW['End_date'] = pd.to_datetime(CAMPAIGN_DESC_RAW['End_date']).
      ↪apply(lambda x: x.date())
```

```
[111]: COUPON_REDEMPT_RAW.head(5)
```

	HOUSEHOLD_KEY	DAY	COUPON_UPC	CAMPAIGN	Date
0	1	421	10000085364	8	2020-02-26
1	1	421	51700010076	8	2020-02-26
2	1	427	54200000033	8	2020-03-03
3	1	597	10000085476	18	2020-08-20
4	1	597	54200029176	18	2020-08-20

```
[112]: COUPON_REDEMPT_RAW.drop(['DAY'],axis=1,inplace=True)
```

```
[113]: COUPON_REDEMPT_RAW.head(5)
```

	HOUSEHOLD_KEY	COUPON_UPC	CAMPAIGN	Date
0	1	10000085364	8	2020-02-26
1	1	51700010076	8	2020-02-26
2	1	54200000033	8	2020-03-03
3	1	10000085476	18	2020-08-20
4	1	54200029176	18	2020-08-20

```
[114]: COUPON_REDEMPT_RAW['Date']=pd.to_datetime(COUPON_REDEMPT_RAW['Date']).
      ↪apply(lambda x: x.date())
```

```
[115]: TRANSACTION_RAW.head(5)
```

```
[115]:
```

	HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	QUANTITY	SALES_VALUE	\
0	2375	26984851472	1	1004906	1	1.39	
1	2375	26984851472	1	1033142	1	0.82	
2	2375	26984851472	1	1036325	1	0.99	
3	2375	26984851472	1	1082185	1	1.21	
4	2375	26984851472	1	8160430	1	1.50	

	STORE_ID	RETAIL_DISC	TRANS_TIME	WEEK_NO	COUPON_DISC	COUPON_MATCH_DISC	\
0	364	-0.60	1631	1	0	0	
1	364	0.00	1631	1	0	0	
2	364	-0.30	1631	1	0	0	
3	364	0.00	1631	1	0	0	
4	364	-0.39	1631	1	0	0	

	DATE
0	2019-01-02
1	2019-01-02
2	2019-01-02
3	2019-01-02
4	2019-01-02

```
[116]: TRANSACTION_RAW.drop(['DAY', 'WEEK_NO'],axis=1,inplace=True)
```

```
[117]: TRANSACTION_RAW.head(5)
```

```
[117]:
```

	HOUSEHOLD_KEY	BASKET_ID	PRODUCT_ID	QUANTITY	SALES_VALUE	STORE_ID	\
0	2375	26984851472	1004906	1	1.39	364	
1	2375	26984851472	1033142	1	0.82	364	
2	2375	26984851472	1036325	1	0.99	364	
3	2375	26984851472	1082185	1	1.21	364	
4	2375	26984851472	8160430	1	1.50	364	

	RETAIL_DISC	TRANS_TIME	COUPON_DISC	COUPON_MATCH_DISC	DATE
0	-0.60	1631	0	0	2019-01-02
1	0.00	1631	0	0	2019-01-02
2	-0.30	1631	0	0	2019-01-02
3	0.00	1631	0	0	2019-01-02
4	-0.39	1631	0	0	2019-01-02

```
[118]: TRANSACTION_RAW['DATE']=pd.to_datetime(TRANSACTION_RAW['DATE']).apply(lambda x:
↳x.date())
```

Now loading the table to Db

```
[119]: from sqlalchemy import create_engine
from sqlalchemy.engine import URL
```

```
import snowflake.connector as snowCtx
from snowflake.connector.pandas_tools import write_pandas
import pandas as pd
import getpass
```

```
[120]: conn = snowflake.connector.connect(
        user = 'DEB',
        password = getpass.getpass('Your Snowflake Password: '),
        ##password = 'Qwerty@12345',
        ## account = svfpjax-gg07076
        account = 'svfpjax-gg07076',
        database='RETAILS',
        schema='PUBLIC',
        warehouse='COMPUTE_WH',
    )
```

Your Snowflake Password: .....

```
[121]: cur = conn.cursor()
```

```
[122]: cur.execute(''' CREATE OR REPLACE TABLE COUPON_REDEMPT_NEW
(HOUSEHOLD_KEY NUMBER(38,0),
COUPON_UPC NUMBER(38,0),
CAMPAIGN NUMBER(38,0),
Date Date
)''')
```

[122]: <snowflake.connector.cursor.SnowflakeCursor at 0x1f79800fcd0>

```
[123]: cur.execute('''CREATE OR REPLACE TABLE TRANSACTION_NEW
(HOUSEHOLD_KEY NUMBER(38,0),
BASKET_ID NUMBER(38,0),
PRODUCT_ID NUMBER(38,0),
QUANTITY NUMBER(38,0),
SALES_VALUE FLOAT,
STORE_ID NUMBER(38,0),
RETAIL_DISC FLOAT,
TRANS_TIME NUMBER(38,0),
COUPON_DISC FLOAT,
COUPON_MATCH_DISC FLOAT,
Date Date
)
''')
```

[123]: <snowflake.connector.cursor.SnowflakeCursor at 0x1f79800fcd0>

```
[124]: cur.execute('''
CREATE OR REPLACE TABLE CAMPAIGN_DESC_NEW
```

```
(DESCRIPTION VARCHAR(10),
CAMPAIGN NUMBER(38,0),
Start_date date,
End_date date,
Campaign_Duration NUMBER(38,0),
Start_month VARCHAR(10),
End_month VARCHAR(10),
Start_Year INT,
End_Year INT)'''
```

[124]: <snowflake.connector.cursor.SnowflakeCursor at 0x1f79800fcd0>

```
[125]: success, nchunks, nrows, _ = write_pandas(conn,
    ↳ CAMPAIGN_DESC_RAW, 'CAMPAIGN_DESC_NEW', quote_identifiers=False)
print(str(success)+' '+str(nchunks)+' '+str(nrows))
```

True,1,30

```
[126]: success, nchunks, nrows, _ = write_pandas(conn,
    ↳ COUPON_REDEMPT_RAW, 'COUPON_REDEMPT_NEW', quote_identifiers=False)
print(str(success)+' '+str(nchunks)+' '+str(nrows))
```

True,1,2318

```
[127]: success, nchunks, nrows, _ = write_pandas(conn, TRANSACTION_RAW,
    ↳ 'TRANSACTION_NEW', quote_identifiers=False)
print(str(success)+' '+str(nchunks)+' '+str(nrows))
```

True,1,450416

```
[128]: cur.close()
conn.close()
```

[ ]:

[ ]:

[ ]:

[ ]: