

```
In [1]: import pandas as pd  
import numpy as np
```

```
In [2]: dataset = pd.read_csv('QVI_data.csv')
```

```
In [3]: dataset.head()
```

Out[3]:

	LYLTY_CARD_NBR	DATE	STORE_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY
0	1000	2018-10-17	1	1	5	Natural Chip Compny SeaSalt175g	2
1	1002	2018-09-16	1	2	58	Red Rock Deli Chikn&Garlic Aioli 150g	1
2	1003	2019-03-07	1	3	52	Grain Waves Sour Cream&Chives 210G	1
3	1003	2019-03-08	1	4	106	Natural ChipCo Hony Soy Chckn175g	1
4	1004	2018-11-02	1	5	96	WW Original Stacked Chips 160g	1



```
In [4]: total_sale = sum(dataset['TOT_SALES'])
```

```
In [6]: total_sale
```

Out[6]: 1933114.9999996515

There is not any customer column in a dataset but we can get customers by TX_ID, because it is unique for every individual.

In [7]: `dataset.describe()`

Out[7]:

	LYLTY_CARD_NBR	STORE_NBR	TXN_ID	PROD_NBR	PROD_QTY	TO
count	2.648340e+05	264834.000000	2.648340e+05	264834.000000	264834.000000	264834.000000
mean	1.355488e+05	135.079423	1.351576e+05	56.583554	1.905813	1.905813
std	8.057990e+04	76.784063	7.813292e+04	32.826444	0.343436	0.343436
min	1.000000e+03	1.000000	1.000000e+00	1.000000	1.000000	1.000000
25%	7.002100e+04	70.000000	6.760050e+04	28.000000	2.000000	2.000000
50%	1.303570e+05	130.000000	1.351365e+05	56.000000	2.000000	2.000000
75%	2.030940e+05	203.000000	2.026998e+05	85.000000	2.000000	2.000000
max	2.373711e+06	272.000000	2.415841e+06	114.000000	5.000000	5.000000



In [14]: `total_customers = 241584`

TOTAL NO OF TRANSACTION PER CUSTOMERS

In [10]: `dataset.shape`

Out[10]: (264834, 12)

In [15]: `avg_transaction = total_customers/264834`

In [16]: `avg_transaction`

Out[16]: 0.9122091574344684

In []: