

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

```
dataset = pd.read_csv("QVI_data.csv")
```

```
dataset.head()
```

	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

	PACK_SIZE \	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

	BRAND	LIFESTAGE	PREMIUM_CUSTOMER
0	NATURAL	YOUNG SINGLES/COUPLES	Premium
1	RRD	YOUNG SINGLES/COUPLES	Mainstream
2	GRNWVES	YOUNG FAMILIES	Budget
3	NATURAL	YOUNG FAMILIES	Budget
4	WOOLWORTHS	OLDER SINGLES/COUPLES	Mainstream

```
total_sales=sum(dataset['TOT_SALES'])
print(total_sales)
```

```
1933115.0
```

```
dataset.describe()
```

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

	PROD_QTY	TOT_SALES	PACK_SIZE
count	264834.000000	264834.000000	264834.000000
mean	1.905813	7.299346	182.425512
std	0.343436	2.527241	64.325148
min	1.000000	1.500000	70.000000
25%	2.000000	5.400000	150.000000
50%	2.000000	7.400000	170.000000
75%	2.000000	9.200000	175.000000
max	5.000000	29.500000	380.000000

```
dataset.shape
```

```
(264834, 12)
```

```
total_customers = 241584
```

```
transaction = 264834
```

```
avg_trans = total_customers/transaction
```

```
print(avg_trans)
```

```
0.9122091574344684
```