IMPORT REQUIRED LIBRARIES

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

IMPORT THE FILES

In [25]:
```
tx = pd.read_csv("QVI_transaction_data (2).csv")      # transactions
cust = pd.read_csv("QVI_purchase_behaviour.csv")
print(tx.shape, cust.shape)
print(tx.columns)
print(cust.columns)
tx.columns   = tx.columns.str.strip()
cust.columns = cust.columns.str.strip()
key = "LYLTY_CARD_NBR"
tx[key]   = tx[key].astype(str)
cust[key] = cust[key].astype(str)
cust = cust.drop_duplicates(subset=key)
combined = tx.merge(cust, on=key, how="left")
print(combined.shape)
# % of transactions without a matching customer record:
missing = combined.filter(regex="^(?!.*" + key + ").*$")  # just to avoid key confusion
# Better: check a known cust column, e.g. 'LIFESTAGE' if it exists
if "LIFESTAGE" in combined.columns:
    print("Unmatched rows:", combined["LIFESTAGE"].isna().mean()*100, "%")
total_sales = combined["TOT_SALES"].sum()
print("Total sales:", total_sales)
combined
```

```
(264836, 8) (72637, 3)
Index(['DATE', 'STORE_NBR', 'LYLTY_CARD_NBR', 'TXN_ID', 'PROD_NBR',
       'PROD_NAME', 'PROD_QTY', 'TOT_SALES'],
      dtype='object')
Index(['LYLTY_CARD_NBR', 'LIFESTAGE', 'PREMIUM_CUSTOMER'], dtype='object')
(264836, 10)
Unmatched rows: 0.0 %
Total sales: 1934415.0000000002
```

Out[25]:

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	TOT_SALES	LIFESTAGE	PREMIUM
0	43390	1	1000	1	5	Natural Chip Compny SeaSalt175g	2	6.0	YOUNG SINGLES/COUPLES	
1	43599	1	1307	348	66	CCs Nacho Cheese 175g	3	6.3	MIDAGE SINGLES/COUPLES	
2	43605	1	1343	383	61	Smiths Crinkle Cut Chips Chicken 170g	2	2.9	MIDAGE SINGLES/COUPLES	
3	43329	2	2373	974	69	Smiths Chip Thinly S/Cream&Onion 175g	5	15.0	MIDAGE SINGLES/COUPLES	
4	43330	2	2426	1038	108	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.8	MIDAGE SINGLES/COUPLES	
...	...	...	...	...	...	...	...	...	...	
264831	43533	272	272319	270088	89	Kettle Sweet Chilli And Sour Cream 175g	2	10.8	YOUNG SINGLES/COUPLES	
264832	43325	272	272358	270154	74	Tostitos Splash Of Lime 175g	1	4.4	YOUNG SINGLES/COUPLES	
264833	43410	272	272379	270187	51	Doritos Mexicana 170g	2	8.8	YOUNG SINGLES/COUPLES	
264834	43461	272	272379	270188	42	Doritos Corn Chip Mexican Jalapeno 150g	2	7.8	YOUNG SINGLES/COUPLES	
264835	43365	272	272380	270189	74	Tostitos Splash Of Lime 175g	2	8.8	YOUNG SINGLES/COUPLES	

264836 rows × 10 columns

TOTAL NUMBER OF CUSTOMER

In [27]:
```
dataset.describe()
```

Out[27]:

	LYLTY_CARD_NBR
count	7.263700e+04
mean	1.361859e+05
std	8.989293e+04
min	1.000000e+03
25%	6.620200e+04
50%	1.340400e+05
75%	2.033750e+05
max	2.373711e+06

In [ ]:
```
total_customer = 241584
```

AVERAGE NUMBER OF TRANSACTION PER CUSTOMER

In [28]:
```
dataset.shape
```

Out[28]: (72637, 3)

In [29]:
```
total_customer = 241584
transaction = 264834
avg_transaction = total_customer/transaction
print(avg_transaction)
```

```
0.9122091574344684
```

In [ ]:
```

```