Task 1-Data Preparation and Customer Analytics

Conduct analysis on your client's transaction dataset and identify customer purchasing behaviours to generate insights and provide commercial recommendations.

Background information for the task

We need to present a strategic recommendation to Julia that is supported by data which she can then use for the upcoming category review however to do so we need to analyse the data to understand the current purchasing trends and behaviours. The client is particularly interested in customer segments and their chip purchasing behaviour. Consider what metrics would help describe the customers' purchasing behaviour.

Main goals of this task are :

1. Examine transaction data - check for missing data, anomalies, outliers and clean them
2. Examine customer data - similar to above transaction data
3. Data analysis and customer segments - create charts and graphs, note trends and insights
4. Deep dive into customer segments - determine which segments should be targetted

Importing Libraries

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df = pd.read_excel('/content/drive/MyDrive/Colab Notebooks/QVI_transaction_data.xlsx')
```

```
df.head()
```

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	TOT_SALES
0	43390	1	1000	1	5	Natural Chip Compny SeaSalt175g	2	6.0
1	43599	1	1307	348	66	CCs Nacho Cheese 175g	3	6.3
2	43605	1	1343	383	61	Smiths Crinkle Cut Chips Chicken 170g	2	2.9
3	43329	2	2373	974	69	Smiths Chip Thinly S/Cream&Onion 175g	5	15.0

```
df.describe()
```

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_QTY	TOT_S
count	264836.000000	264836.00000	2.648360e+05	2.648360e+05	264836.000000	264836.000000	264836.00
mean	43464.036260	135.08011	1.355495e+05	1.351583e+05	56.583157	1.907309	7.30
std	105.389282	76.78418	8.057998e+04	7.813303e+04	32.826638	0.643654	3.08
min	43282.000000	1.00000	1.000000e+03	1.000000e+00	1.000000	1.000000	1.50
25%	43373.000000	70.00000	7.002100e+04	6.760150e+04	28.000000	2.000000	5.40
50%	43464.000000	130.00000	1.303575e+05	1.351375e+05	56.000000	2.000000	7.40
75%	43555.000000	203.00000	2.030942e+05	2.027012e+05	85.000000	2.000000	9.20
max	43646.000000	272.00000	2.373711e+06	2.415841e+06	114.000000	200.000000	650.00

```
df1 = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/QVI_purchase_behaviour.csv')
```

```
df1.head()
```

```
df1.describe()
```

	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

```
df.isnull().sum()
```

```
DATE                0
STORE_NBR           0
LYLTY_CARD_NBR      0
TXN_ID              0
PROD_NBR            0
PROD_NAME           0
PROD_QTY            0
TOT_SALES           0
dtype: int64
```

```
df1.isnull().sum()
```

```
LYLTY_CARD_NBR      0
LIFESTAGE           0
PREMIUM_CUSTOMER    0
dtype: int64
```

▶ **Checking & Removing Outliers**

[ ] ↳ *4 cells hidden*

▶ **Date Column to Date-Time Format**

[ ] ↳ *2 cells hidden*

▶ **Product Name Columns to check all items are Chips**

[ ] ↳ *5 cells hidden*

▶ **PROD_QTY values above 50.75**

[ ] ↳ *1 cell hidden*

▶ Two outliers of value 200 in PROD_QTY will be removed. Both entries are by the same customer and will be examined by this customer's transactions

[ ] ↳ *51 cells hidden*

Trends and Insights :

Top 3 total sales contributor segment are

- Older families (Budget) $156,864
- Young Singles/Couples (Mainstream) $147,582
- Retirees (Mainstream) $145,169

1. Young Singles/Couples (Mainstream) has the highest population, followed by Retirees (Mainstream). Which explains their high total sales.

2. Despite Older Families not having the highest population, they have the highest frequency of purchase, which contributes to their high total sales.

3. Older Families followed by Young Families has the highest average quantity of chips bought per purchase.

4. The Mainstream category of the "Young and Midage Singles/Couples" have the highest spending of chips per purchase. And the difference to the non-Mainstream "Young and Midage Singles/Couples" are statistically significant.

5. Chips brand Kettle is dominating every segment as the most purchased brand.

6. Observing the 2nd most purchased brand, "Young and Midage Singles/Couples" is the only segment with a different preference (Doritos) as compared to others' (Smiths).

7. Most frequent chip size purchased is 175gr followed by the 150gr chip size for all segments.

Views and Recommendations:

1. Older Families:

Focus on the Budget segment. Strength: Frequent purchase. We can give promotions that encourages more frequency of purchase. Strength: High quantity of chips purchased per visit. We can give promotions that encourage them to buy more quantity of chips per purchase.

2. Young Singles/Couples:

Focus on the Mainstream segment. This segment is the only segment that had Doritos as their 2nd most purchased brand (after Kettle). To specifically target this segment it might be a good idea to collaborate with Doritos merchant to do some branding promotion catered to "Young Singles/Couples - Mainstream" segment. Strength: Population quantity. We can spend more effort on making sure our promotions reach them, and it reaches them frequently.

3. Retirees:

Focus on the Mainstream segment. Strength: Population quantity. Again, since their population quantity is the contributor to the high total sales, we should spend more effort on making sure our promotions reaches as many of them as possible and frequent.

4. General:

All segments has Kettle as the most frequently purchased brand, and 175gr (regardless of brand) followed by 150gr as the preferred chip size. When promoting chips in general to all segments it is good to take advantage of these two points.

Colab paid products - Cancel contracts here

✓ 0s completed at 3:05 PM ● ✕