

```
In [36]: import pandas as pd

# Load the CSV and Excel files
# Read the CSV and the Excel files
purchase_behaviour_df = pd.read_csv('QVI_purchase_behaviour.csv', header=0)
transaction_data_df = pd.read_excel('QVI_transaction_data.xlsx')

# Display the first few rows of both dataframes for initial inspection
purchase_behaviour_df.head(), transaction_data_df.head()
```

```
Out[36]: (  LYLTY_CARD_NBR          LIFESTAGE PREMIUM_CUSTOMER
0           1000    YOUNG SINGLES/COUPLES          Premium
1           1002    YOUNG SINGLES/COUPLES          Mainstream
2           1003          YOUNG FAMILIES          Budget
3           1004    OLDER SINGLES/COUPLES          Mainstream
4           1005    MIDAGE SINGLES/COUPLES          Mainstream,

   DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
0  43390         1           1000        1         5
1  43599         1           1307       348         66
2  43605         1           1343       383         61
3  43329         2           2373       974         69
4  43330         2           2426      1038        108

          PROD_NAME  PROD_QTY  TOT_SALES
0   Natural Chip      Compny SeaSalt175g         2         6.0
1              CCs Nacho Cheese      175g         3         6.3
2  Smiths Crinkle Cut  Chips Chicken 170g         2         2.9
3  Smiths Chip Thinly  S/Cream&Onion 175g         5        15.0
4  Kettle Tortilla ChpsHny&Jlpno Chili 150g         3        13.8 )
```

```
In [39]: # Check for duplicates and drop them
purchase_behaviour_df = purchase_behaviour_df.drop_duplicates()
transaction_data_df = transaction_data_df.drop_duplicates()

# Standardize column names to lowercase for consistency
purchase_behaviour_df.columns = purchase_behaviour_df.columns.str.lower()
transaction_data_df.columns = transaction_data_df.columns.str.lower()

# Check for and remove null values
purchase_behaviour_df = purchase_behaviour_df.dropna()
transaction_data_df = transaction_data_df.dropna()

# Merge the datasets on 'lylty_card_nbr'
merged_df = pd.merge(transaction_data_df, purchase_behaviour_df, on='lylty_card_nbr')

# Display the shape of the merged dataframe and the first few rows
merged_df.shape, merged_df.head()
```

```
Out[39]: ((264835, 10),
          date store_nbr  lylty_card_nbr  txn_id  prod_nbr  \
0  43390      1          1000      1      5
1  43599      1          1307     348     66
2  43605      1          1343     383     61
3  43329      2          2373     974     69
4  43330      2          2426    1038    108

          prod_name  prod_qty  tot_sales  \
0  Natural Chip      Compny SeaSalt175g      2      6.0
1          CCs Nacho Cheese      175g      3      6.3
2  Smiths Crinkle Cut  Chips Chicken 170g      2      2.9
3  Smiths Chip Thinly  S/Cream&Onion 175g      5     15.0
4  Kettle Tortilla ChpsHny&Jlpno Chili 150g      3     13.8

          lifestage  premium_customer
0  YOUNG SINGLES/COUPLES      Premium
1  MIDAGE SINGLES/COUPLES      Budget
2  MIDAGE SINGLES/COUPLES      Budget
3  MIDAGE SINGLES/COUPLES      Budget
4  MIDAGE SINGLES/COUPLES      Budget )
```

In []: