My Homeworks 4.6.1 Combining & Exporting Data

Merging the Instacart Orders Data with the orders_products_prior DataFrame

In [5]:
```
# Import libraries
import pandas as pd
import numpy as np
import os

# Tell Python to remember a main folder path
path = r'C:\Users\Asus\Music\Instacart Basket Analysis'

# Import dataset orders_products_prior.csv
df_ords_prior = pd.read_csv(os.path.join(path, 'Data', 'Original Data', 'orders_products_prior.csv'), index_col = False)
# Import dataset orders_checked.csv
df_ords = pd.read_csv(os.path.join(path, 'Data', 'Prepared Data', 'orders_cleaned.csv'), index_col = False)
df_prods = pd.read_csv(os.path.join(path, 'Data', 'Prepared Data', 'products_cleaned.csv'), index_col = False)

# Merging orders and order_products_prior
orders_products_combined = pd.merge(df_ords, df_ords_prior, on='order_id', how='inner')
```

2. Export the Merged File in Pickle Format

In [7]:
```
orders_products_combined.to_pickle('orders_products_combined.pkl')
```

3. Import the orders_products_combined DataFrame from the Pickle File

In [9]:
```
orders_products_combined = pd.read_pickle('orders_products_combined.pkl')
```

4. Check the Shape of the Imported DataFrame

In [11]:
```
print(orders_products_combined.shape)
```

```
(32434489, 11)
```

5. Combine the orders_products_combined DataFrame with Your Products Data Set

Using our cleaned df_prods dataframe for this merge.

In [13]:
```
ords_prods_merge = pd.merge(orders_products_combined, df_prods, on='product_id', how='inner')
```

6. Confirm the Results of the Merge Using the Merge Flag

create a merge flag to see the source of each row.

In [ ]:
```
ords_prods_merge['_merge'] = ords_prods_merge['_merge'].astype('category')
print(ords_prods_merge['_merge'].value_counts())
```

7. Export the Newly Created DataFrame in a Suitable Format

In [21]:
```
ords_prods_merge.to_csv('ords_prods_merge.pkl', compression='gzip')
```

8. Organize Your Notebooks and Project Folder

Ensure my project folder has a Data subfolder. all your exported data files into the Data folder.

In [ ]:
```
Instacart Basket Analysis/
│
├── Script/
│   ├── orders_products_combined.pkl
│   ├── ords_prods_merge.csv.gz
│   └── products_cleaned.csv
│
├── Notebooks/
│   ├── my Homeworks 4.6.1 Combining & Exporting Data.ipynb
│
└── README.md
```

9. Save the Two Notebooks and Send Them to Your Tutor

i saved couples of notebooks in one notebook

In [65]:
```
# Data Cleaning Notebook

## Importing Libraries
import pandas as pd
```

Merging DataFrames

In [76]:
```
# Merging orders and order_products_prior
orders_products_combined = pd.merge(df_ords, df_ords_prior, on='order_id', how='inner')
```

Exporting Data

In [81]:
```
orders_products_combined.to_pickle('orders_products_combined.pkl')
```

Checking the shape of the imported dataframe

In [84]:
```
print(orders_products_combined.shape)
```

```
(32434489, 11)
```

Merging with cleaned products data

In [91]:
```
df_prods = pd.read_csv(os.path.join(path, 'Data', 'Prepared Data', 'products_cleaned.csv'), index_col = False)
ords_prods_merge = pd.merge(orders_products_combined, df_prods, on='product_id', how='inner')
```

Exporting merged dataframe

In [104…
```
ords_prods_merge.to_csv('ords_prods_merge.csv.gz', compression='gzip')
```

In [ ]:
```
```

In [23]:
```
import pandas as pd

# Assume ords_prods_merge is already loaded from the previous steps
# Creating 'price_label' column
ords_prods_merge['price_label'] = ords_prods_merge['prices'].apply(lambda x: 'High' if x > 10 else ('Low' if x < 5 else 'Medium'))

# Creating 'busiest_day' column
# Count the number of orders per day
day_orders = ords_prods_merge['order_dow'].value_counts()

# Determine the busiest day
```

```
busiest_day = day_orders.idxmax()
ords_prods_merge['busiest_day'] = ords_prods_merge['order_dow'].apply(lambda x: 'Busiest day' if x == busiest_day else 'Other day')
```

In [ ]: