Market Basket Analysis

In this assignment, we will leverage the Apriori algorithm, a powerful technique rooted in associative learning, to conduct a comprehensive market basket analysis.

In [11]:

```
#loading the packages
import pandas as pd
import numpy as np
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

In [2]:

```
# Load the datasets
online_sales = pd.read_csv('C:\\Users\\sujoydutta\\Desktop\\Data analysis\\Projects\\Marketing insights\\Online
```

In [3]:

```
#getting information of the dataset
online_sales.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52924 entries, 0 to 52923
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   CustomerID        52924 non-null  int64  
 1   Transaction_ID    52924 non-null  int64  
 2   Transaction_Date  52924 non-null  int64  
 3   Product_SKU       52924 non-null  object 
 4   Product_name      52924 non-null  object 
 5   Product_Category  52924 non-null  object 
 6   Quantity          52924 non-null  int64  
 7   Avg_Price         52924 non-null  float64
 8   Delivery_Charges  52924 non-null  float64
 9   Coupon_Status     52924 non-null  object 
dtypes: float64(2), int64(4), object(4)
memory usage: 4.0+ MB
```

In [5]:

```
#taking subset
columns_of_interest = ['CustomerID', 'Transaction_Date', 'Product_name']
subset=online_sales[columns_of_interest]
subset
```

Out[5]:

	CustomerID	Transaction_Date	Product_name
0	17850	20190101	Nest Learning Thermostat 3rd Gen-USA - Stainle...
1	17850	20190101	Nest Learning Thermostat 3rd Gen-USA - Stainle...
2	17850	20190101	Google Laptop and Cell Phone Stickers
3	17850	20190101	Google Men's 100% Cotton Short Sleeve Hero Tee...
4	17850	20190101	Google Canvas Tote Natural/Navy
...	...	...	...
52919	14410	20191231	Nest Cam Indoor Security Camera - USA
52920	14410	20191231	Google Zip Hoodie Black
52921	14410	20191231	Nest Learning Thermostat 3rd Gen-USA - White
52922	14600	20191231	Nest Protect Smoke + CO White Wired Alarm-USA
52923	14600	20191231	Nest Protect Smoke + CO White Battery Alarm-USA

52924 rows × 3 columns

In [7]:

```
# Grouping products by CustomerID and Transaction_Date, join them with commas
transaction_data = subset.groupby(['CustomerID', 'Transaction_Date'])['Product_name'].apply(','.join).reset_ind
transaction_data
```

Out[7]:

	CustomerID	Transaction_Date	Product_name
0	12346	20190915	Android Men's Engineer Short Sleeve Tee Charco...
1	12347	20190324	Four Color Retractable Pen,Red Spiral Google N...
2	12347	20191101	Google Doodle Decal,Google Twill Cap,Windup An...
3	12347	20191102	Nest Learning Thermostat 3rd Gen-USA - White,N...
4	12348	20190622	26 oz Double Wall Insulated Bottle,Google Styl...
...	...	...	...
3203	18269	20190405	Android BTTF Cosmos Graphic Tee,Android Men's ...
3204	18269	20190620	Google Men's Vintage Tank
3205	18277	20191023	Nest Learning Thermostat 3rd Gen-USA - Stainle...
3206	18283	20190729	Google Leather Perforated Journal,Recycled Pap...
3207	18283	20191010	Keyboard DOT Sticker,Google Laptop and Cell Ph...

3208 rows × 3 columns

In [12]:

```
# Split the Product_name column by comma and create dummy variables (one-hot encoding)
onehot_data = transaction_data['Product_name'].str.get_dummies(',')
```

In [13]:

```
# Concatenate the one-hot encoded data with the original dataset
transaction_data = pd.concat([transaction_data, onehot_data], axis=1)
transaction_data
```

Out[13]:

	CustomerID	Transaction_Date	Product_name	1 oz Hand Sanitizer	20 oz Stainless Steel Insulated Tumbler	22 oz Android Bottle	22 oz YouTube Bottle Infuser	23 oz Wide Mouth Sport Bottle	24 oz YouTube Sergeant Stripe Bottle	25L Classic Rucksack	...	YouTube Twill Cap	YouTu Wome Favo Wh
0	12346	20190915	Android Men's Engineer Short Sleeve Tee Charco...	0	0	0	0	0	0	0	...	0	
1	12347	20190324	Four Color Retractable Pen,Red Spiral Google N...	0	0	0	0	0	1	0	...	0	
2	12347	20191101	Google Doodle Decal,Google Twill Cap,Windup An...	0	0	0	0	0	0	0	...	0	
3	12347	20191102	Nest Learning Thermostat 3rd Gen-USA - White,N...	0	0	0	0	0	0	0	...	0	
4	12348	20190622	26 oz Double Wall Insulated Bottle,Google Styl...	0	0	0	0	0	0	0	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
3203	18269	20190405	Android BTTF Cosmos Graphic Tee,Android Men's ...	0	0	0	0	0	0	0	...	0	
3204	18269	20190620	Google Men's Vintage Tank	0	0	0	0	0	0	0	...	0	
3205	18277	20191023	Nest Learning Thermostat 3rd Gen-USA - Stainle...	0	0	0	0	0	0	0	...	0	
3206	18283	20190729	Google Leather Perforated Journal,Recycled Pap...	0	0	0	0	0	0	0	...	1	
3207	18283	20191010	Keyboard DOT Sticker,Google Laptop and Cell Ph...	0	0	0	1	0	0	0	...	0	

3208 rows × 407 columns

In [14]:

```
# Drop the original Product_name column if you no longer need it
transaction_data.drop(columns=['Product_name'], inplace=True)
```

In [17]:

```
#dropping useless columns
transaction_data.drop(columns=['CustomerID','Transaction_Date'], inplace=True)
```

In [18]:

```
# Set a minimum support threshold (e.g., 1% of transactions)
min_support = 0.01

# Apply Apriori algorithm to find frequent itemsets
frequent_itemsets = apriori(transaction_data, min_support=min_support, use_colnames=True)
```

```
C:\Users\sujoydutta\anaconda3\lib\site-packages\mlxtend\frequent_patterns\fpcommon.py:110: DeprecationWarning: 
DataFrames with non-bool types result in worse computationalperformance and their support might be discontinued 
in the future.Please use a DataFrame with bool type
  warnings.warn(
```

In [19]:

```
#  Generating association rules
rules = association_rules(frequent_itemsets, metric='lift', min_threshold=1.0)
rules
```

Out[19]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Google 22 oz Water Bottle)	(1 oz Hand Sanitizer)	0.176746	0.036471	0.018392	0.104056	2.853103	0.011945	1.075435	0.788947
1	(1 oz Hand Sanitizer)	(Google 22 oz Water Bottle)	0.036471	0.176746	0.018392	0.504274	2.853103	0.011945	1.660703	0.674089
2	(1 oz Hand Sanitizer)	(Google Kick Ball)	0.036471	0.081359	0.010599	0.290598	3.571798	0.007631	1.294952	0.747283
3	(Google Kick Ball)	(1 oz Hand Sanitizer)	0.081359	0.036471	0.010599	0.130268	3.571798	0.007631	1.107846	0.783798
4	(1 oz Hand Sanitizer)	(Google Laptop and Cell Phone Stickers)	0.036471	0.200748	0.015274	0.418803	2.086213	0.007953	1.375183	0.540371
...	...	...	...	...	...	...	...	...	...	...
278421	(Nest Cam Indoor Security Camera - USA)	(Nest Secure Alarm System Starter Pack - USA, ...	0.471633	0.010599	0.010287	0.021811	2.057929	0.005288	1.011462	0.972951
278422	(Nest Protect Smoke + CO White Wired Alarm-USA)	(Nest Secure Alarm System Starter Pack - USA, ...	0.236908	0.014963	0.010287	0.043421	2.901974	0.006742	1.029750	0.858883
278423	(Nest Thermostat E - USA)	(Nest Secure Alarm System Starter Pack - USA, ...	0.125935	0.011534	0.010287	0.081683	7.082151	0.008834	1.076389	0.982536
278424	(Nest Learning Thermostat 3rd Gen-USA - White)	(Nest Secure Alarm System Starter Pack - USA, ...	0.221945	0.011845	0.010287	0.046348	3.912774	0.007658	1.036180	0.956779
278425	(Nest Learning Thermostat 3rd Gen-USA - Stainl...	(Nest Secure Alarm System Starter Pack - USA, ...	0.475374	0.010910	0.010287	0.021639	1.983400	0.005100	1.010966	0.945084

278426 rows × 10 columns

In [20]:

```
# Filtering rules based on confidence and lift thresholds
filtered_rules = rules[(rules['confidence'] >= 0.5) & (rules['lift'] >= 1.0)]
filtered_rules
```

Out[20]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
1	(1 oz Hand Sanitizer)	(Google 22 oz Water Bottle)	0.036471	0.176746	0.018392	0.504274	2.853103	0.011945	1.660703	0.674089
10	(1 oz Hand Sanitizer)	(Nest Cam Indoor Security Camera - USA)	0.036471	0.471633	0.024002	0.658120	1.395405	0.006801	1.545472	0.294088
12	(1 oz Hand Sanitizer)	(Nest Cam Outdoor Security Camera - USA)	0.036471	0.479115	0.023691	0.649573	1.355777	0.006217	1.486429	0.272348
14	(1 oz Hand Sanitizer)	(Nest Learning Thermostat 3rd Gen-USA - Stainl...	0.036471	0.475374	0.023379	0.641026	1.348466	0.006042	1.461458	0.268198
28	(20 oz Stainless Steel Insulated Tumbler)	(Nest Cam Indoor Security Camera - USA)	0.033042	0.471633	0.022756	0.688679	1.460200	0.007172	1.697178	0.325932
...	...	...	...	...	...	...	...	...	...	...
278164	(Nest Protect Smoke + CO White Battery Alarm-U...	(Nest Secure Alarm System Starter Pack - USA, ...	0.020262	0.032419	0.010287	0.507692	15.660355	0.009630	1.965399	0.955505
278173	(Nest Protect Smoke + CO White Battery Alarm-U...	(Nest Cam Outdoor Security Camera - USA, Nest ...	0.018703	0.038653	0.010287	0.550000	14.229032	0.009564	2.136326	0.947441
278187	(Nest Secure Alarm System Starter Pack - USA, ...	(Nest Protect Smoke + CO White Battery Alarm-U...	0.019638	0.035536	0.010287	0.523810	14.740184	0.009589	2.025374	0.950831
278202	(Nest Protect Smoke + CO White Battery Alarm-U...	(Nest Cam IQ - USA, Nest Cam Outdoor Security ...	0.020262	0.036471	0.010287	0.507692	13.920316	0.009548	1.957168	0.947358
278237	(Nest Protect Smoke + CO White Battery Alarm-U...	(Nest Secure Alarm System Starter Pack - USA, ...	0.020574	0.034289	0.010287	0.500000	14.581818	0.009581	1.931421	0.950987

64545 rows × 10 columns

Comment:This table is self explanatory,let us take an example of the first rule: this rule suggests that there is a moderate association between purchasing (1 oz Hand Sanitizer) and (Google 22 oz Water Bottle). Customers who buy (1 oz Hand Sanitizer) are 50.43% more likely to also buy (Google 22 oz Water Bottle) than if they were purchased independently. The lift value of 2.85 indicates that this association is stronger than random chance, and the conviction value of 166.07 suggests a strong belief in the association.