MARKET BASKET INSIGHTS

Phase 2: Innovation

Introduction:

Market Basket Analysis is a Data Mining technique used by retailers to increase sales by better understanding customer purchasing patterns. It involves analysing large data sets, such as purchase history, to reveal product groupings and products that are likely to be purchased together. and Apriori algorithm is a widely-used and well-known association rule algorithm and is a popular algorithm used in market basket analysis. it is also considered accurate and overtop AIS and SETM algorithms. Implementation of market analysis requires a background in statistics and data science and some algorithmic computer programming skills. One example is the Shopping Basket Analysis tool in Microsoft Excel, which analysis transaction data contained in a spreadsheet and performs market basket analysis. A transaction ID must relate to the items to be analysed.

Advanced Association Analysis Techniques:

1. Apriori Algorithm: Apriori is a classic algorithm for association rule mining. It identifies patterns in data, such as frequent item-sets and association rules, which can reveal hidden relationships between variables.

2. FP-Growth: FP-Growth is another frequent itemset mining algorithm that is more efficient than Apriori for large datasets. It constructs a compact data structure called an FP-tree to find frequent item-sets.

3. Sequential Pattern Mining: If your data involves sequences (e.g., time series data or clickstream data), techniques like Sequential Pattern Mining can help discover patterns of events or behaviours over time.

Visualization Tools:

1. Heatmaps: Heatmaps are excellent for displaying associations and correlations between variables. You can use them to visualize the strength of relationships between items or attributes.

2. Sankey Diagrams: Sankey diagrams are useful for showing flow and connections between categories or stages in a process. They are effective in visualizing sequential patterns.

3. Network Graphs: Network graphs can represent complex relationships between entities, making them valuable for visualizing associations in data with multiple interconnected variables.

4. Interactive Dashboards: Tools like Tableau, Power BI, or custom-built dashboards using libraries like D3.js allow users to explore associations interactively. Users can filter, drill down, and see associations in real-time.

5. 3D Visualization: For multidimensional data, 3D visualization techniques can be used to provide a more immersive view of associations and patterns.

Enhanced Insights Presentation:

1. Interactive Reports: Create interactive reports that allow users to explore associations themselves.

2. Anomaly Detection: Overlay anomaly detection on your visualizations to highlight unusual patterns or further investigation

3. Predictive Analytics: If relevant, use predictive models in conjunction with association analysis to forecast future trends based on historical associations.

Machine Learning and Deep Learning:

For more complex and non-linear relationships, consider using machine learning or deep learning techniques such as neural networks to uncover associations and patterns in the data

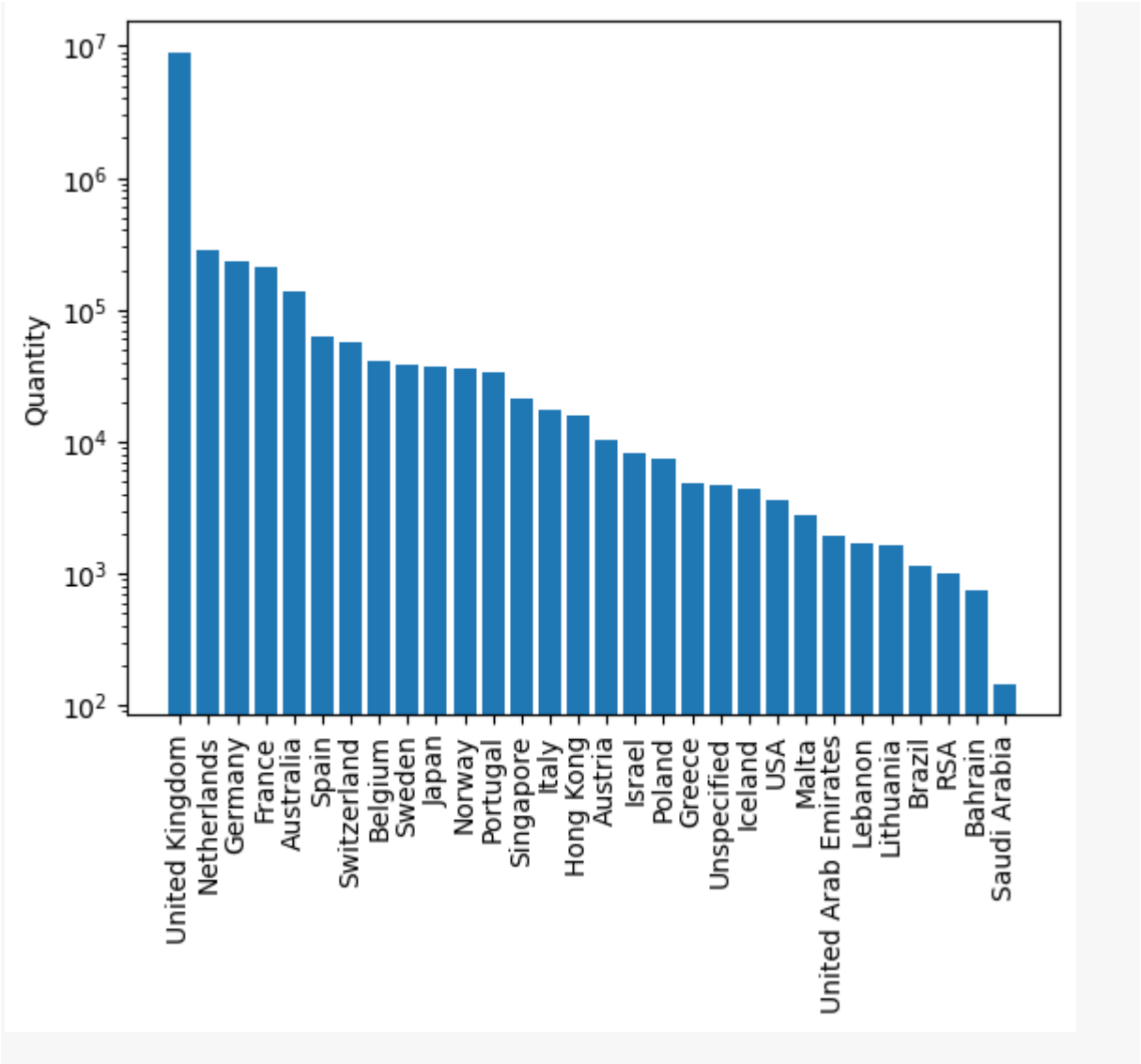
Code:

```
import numpy as np
import pandas as pd
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
from matplotlib import pyplot as plt
df=pd.read_excel("/kaggle/input/market-basket-analysis/Assignment-1_Data.xlsx")
df.info()
df.isnull().sum()
df.dropna(subset=["Itemname"],inplace=True)
```

```
df = df[df["Quantity"]>0]

df.isnull().sum()

df['CustomerID'].fillna(99999, inplace=True)

df["SumPrice"]=df["Quantity"]*df["Price"]

best_selling_items = df.groupby(['Country', 'Itemname']).agg({'Quantity':
'sum'}).reset_index()

best_selling_items = best_selling_items.groupby('Country').apply(lambda x:
x[x['Quantity'] == x['Quantity'].max()]).reset_index(drop=True)

best_selling_items.sort_values("Quantity",ascending=False)

total_sales_country = df.groupby(['Country']).agg({'SumPrice': 'sum'}).reset_index()

total_sales_country = total_sales_country.sort_values('SumPrice',
ascending=False).reset_index(drop=True)

total_sales_country

plt.bar(total_sales_country["Country"],total_sales_country["SumPrice"])

plt.yscale('log')

plt.ylabel('Quantity')

plt.xticks(rotation=90)

plt.show()

only_uk = df[df["Country"]=="United Kingdom"]

only_uk.groupby("Itemname")["Quantity"].sum().sort_values(ascending=False)

total_sales_item = df.groupby(['Itemname']).agg({'Price': 'mean', 'Quantity': 'sum',
'SumPrice': 'sum'}).reset_index()

total_sales_item['Count'] = df.groupby(['Itemname']).size().value

total_sales_item = total_sales_item.sort_values("SumPrice", ascending=False)

total_sales_item

transactions = df.groupby(['BillNo'])['Itemname'].apply(list)

transactions

one_hot = pd.get_dummies(df['Itemname'])

one_hot

one_hot['BillNo']=df['BillNo']

one_hot

transaction_matrix = pd.merge(transactions, one_hot, on='BillNo')
```

```
transaction_matrix
transaction_matrix[one_hot.columns[:-1]] = (transaction_matrix[one_hot.columns[:-1]] >= 1).astype(int)
transaction_matrix
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
print(transaction_matrix.dtypes)
transaction_matrix.iloc[:, 1:] = transaction_matrix.iloc[:, 1:].astype(bool)
frequent_itemsets = apriori(transaction_matrix.iloc[:, 1:], min_support=0.01, use_colnames=True)
frequent_itemsets
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
rules
rules.sort_values('lift', ascending=False).head(10)
import mpld3
fig, ax = plt.subplots()
scatter = ax.scatter(rules['lift'], rules['confidence'], alpha=0.5
tooltips = []
for i in range(len(rules)):
    rule = rules.iloc[i]
    tooltip = f"Rule: {rule['antecedents']} -> {rule['consequents']}\nSupport: {rule['support']:.3f}\nConfidence: {rule['confidence']:.3f}\nLift: {rule['lift']:.3f}"
    tooltips.append(tooltip)
mpld3.plugins.connect(fig, mpld3.plugins.PointHTMLTooltip(scatter, tooltips))
ax.set_xlabel("Lift")
ax.set_ylabel("Confidence")
ax.set_title("Association Rules Scatter Plot")
mpld3.display()
rules[(rules['lift'] > 40) & (rules['lift'] < 50)]
```

Output:



Itemname
PAPER CRAFT, LITTLE BIRDIE 80995
MEDIUM CERAMIC TOP STORAGE JAR 77036
WORLD WAR 2 GLIDERS ASSTD DESIGNS 49526
JUMBO BAG RED RETROSPOT 44268
WHITE HANGING HEART T-LIGHT HOLDER 35744
...
HEN HOUSE W CHICK IN NEST 1
BLACKCHRISTMAS TREE 30CM 1
GOLD COSMETICS BAG WITH BUTTERFLY 1
WATERING CAN SINGLE HOOK PISTACHIO 1
*Boombox Ipod Classic 1

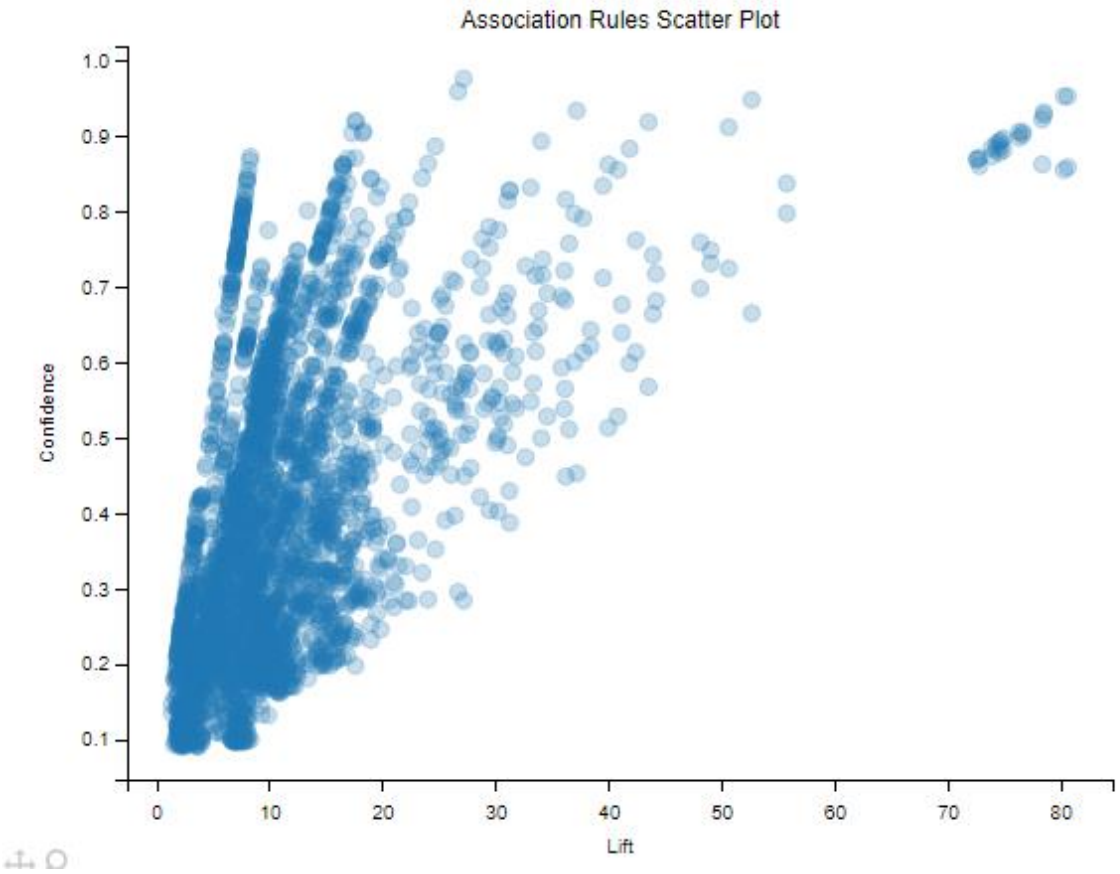
Name: Quantity, Length: 4046, dtype: int64

	support	itemsets
0	0.015809	(10 COLOUR SPACEBOY PEN)
1	0.012567	(12 MESSAGE CARDS WITH ENVELOPES)
2	0.017887	(12 PENCIL SMALL TUBE WOODLAND)
3	0.018242	(12 PENCILS SMALL TUBE RED RETROSPOT)
4	0.017887	(12 PENCILS SMALL TUBE SKULL)
...	...	...
1891	0.011249	(JUMBO BAG RED RETROSPOT, JUMBO SHOPPER VINTAG...
1892	0.011249	(LUNCH BAG CARS BLUE, LUNCH BAG BLACK SKULL.,...
1893	0.010388	(LUNCH BAG CARS BLUE, LUNCH BAG BLACK SKULL.,...
1894	0.010286	(LUNCH BAG SUKI DESIGN, LUNCH BAG BLACK SKULL...
1895	0.010286	(CHARLOTTE BAG PINK POLKADOT, CHARLOTTE BAG SU...

	Itemname	*Boombox Ipod Classic	*USB Office Mirror Ball	10 COLOUR SPACEBOY PEN	12 COLOURED PARTY BALLOONS	12 DAISY PEGS IN WOOD BOX	12 EGG HOUSE PAINTED WOOD	12 HANGING EGGS HAND PAINTED	12 IVORY ROSE PEG PLACE SETTINGS	12 MESSAGE CARDS WITH ENVELOPES	...	returned
BillNo												
536365	[WHITE HANGING HEART T-LIGHT HOLDER, WHITE MET...	0	0	0	0	0	0	0	0	0	...	0
536366	[HAND WARMER UNION JACK, HAND WARMER RED POLKA...	0	0	0	0	0	0	0	0	0	...	0
536367	[ASSORTED COLOUR BIRD ORNAMENT, POPPY'S PLAYHO...	0	0	0	0	0	0	0	0	0	...	0
	[JAM											

Out[10]:

United Kingdom	486167
Germany	9042
France	8408
Spain	2485
Netherlands	2363
Belgium	2031
Switzerland	1967
Portugal	1501
Australia	1185
Norway	1072
Italy	758
Sweden	451
Unspecified	446
Austria	398
Poland	330
Japan	321
Israel	295
Hong Kong	284
Singapore	222
Iceland	182



Conclusion:

Organizations are using this technique wisely and making billions by playing with the mind of the customer. It is an effective way of improving your sales without having to put extra effort into marketing that won't give you results as incredible as with this technique. So go ahead and try it on all the data you have in your repository to recognize patterns that may surprise you to the roots. Market basket analysis (MBA) finds great application for the marketing perspective to the retailers. It helps retailers in optimizing their marketing campaigns and strategize future sales by understanding their customers better Apriori is the commonly cited algorithm by the data scientist that identifies frequent items in the database. It is useful for unsupervised learning and requires no training and thus no predictions. This algorithm is used especially for large data sets where useful relationships among the items are to be determined. This shortcut states that all items in a frequent itemset must also be frequent. It helps in saving a lot of computational time.