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

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
```

```
df=pd.read_csv('/content/drive/MyDrive/Market_Basket_Optimisation (1).csv')
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
```

DATA PREPROCESSING

```
df.head()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```

	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	h
0	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	mineral water	milk	energy bar	whole wheat rice	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	low fat yogurt	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

```
df.shape
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
(7500, 20)
```

MISSING VALUES

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

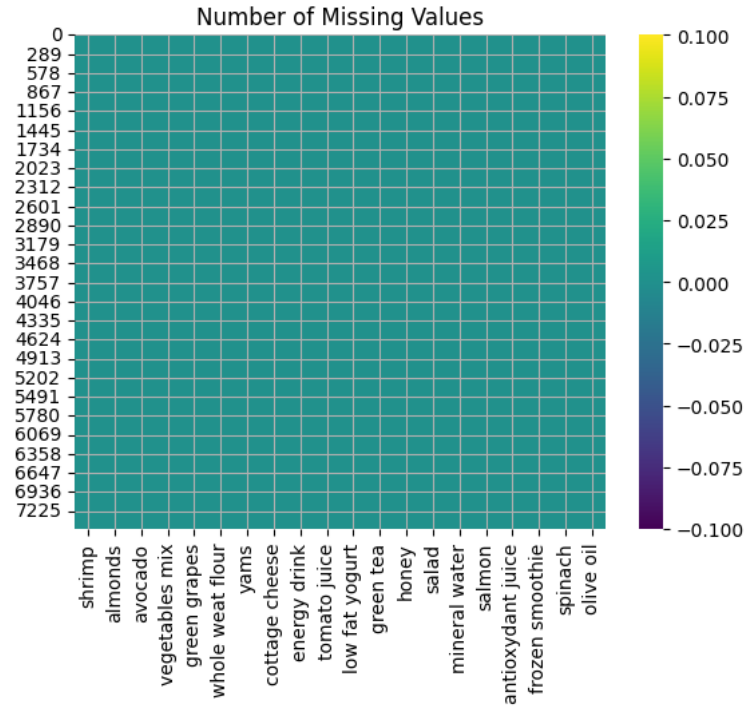
```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
shrimp                  0
almonds              1754
avocado              3112
vegetables mix       4156
green grapes         4972
whole weat flour     5637
yams                 6132
cottage cheese       6520
energy drink         6847
tomato juice         7106
low fat yogurt       7245
green tea            7347
honey                7414
salad                7454
mineral water        7476
salmon               7493
antioxydant juice    7497
frozen smoothie      7497
spinach              7498
olive oil            7500
dtype: int64
```

```
df=df.fillna(value=0)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
```

```
df.isnull().sum()
sns.heatmap(df.isnull(),cmap='viridis')
plt.grid()
plt.title("Number of Missing Values")
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
Text(0.5, 1.0, 'Number of Missing Values')
```



BINARY MATRIX FORMAT

```
df.head()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```

	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	h
0	burgers	meatballs	eggs	0	0	0	0	0	0	0	0	0	
1	chutney	0	0	0	0	0	0	0	0	0	0	0	
2	turkey	avocado	0	0	0	0	0	0	0	0	0	0	
3	mineral water	milk	energy bar	whole wheat rice	green tea	0	0	0	0	0	0	0	
4	low fat yogurt	0	0	0	0	0	0	0	0	0	0	0	

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7500 entries, 0 to 7499
Data columns (total 20 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   shrimp             7500 non-null   object
 1   almonds            7500 non-null   object
 2   avocado            7500 non-null   object
 3   vegetables mix     7500 non-null   object
 4   green grapes       7500 non-null   object
 5   whole weat flour   7500 non-null   object
```

```
 6   yams              7500 non-null   object
 7   cottage cheese    7500 non-null   object
 8   energy drink      7500 non-null   object
 9   tomato juice      7500 non-null   object
 10  low fat yogurt    7500 non-null   object
 11  green tea         7500 non-null   object
 12  honey             7500 non-null   object
 13  salad             7500 non-null   object
 14  mineral water     7500 non-null   object
 15  salmon            7500 non-null   object
 16  antioxydant juice 7500 non-null   object
 17  frozen smoothie   7500 non-null   object
 18  spinach           7500 non-null   object
 19  olive oil         7500 non-null   float64
dtypes: float64(1), object(19)
memory usage: 1.1+ MB
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
```

```
df = pd.get_dummies(df, columns = ['shrimp','almonds','avocado','vegetables mix','green grapes','whole weat flour',
                                   'yams','cottage cheese','energy drink','tomato juice','low fat yogurt','green tea','honey','salad','mi
                                   'antioxydant juice','frozen smoothie','spinach','olive oil'])
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
```

```
df.head()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```

	shrimp_almonds	shrimp_antioxydant juice	shrimp_asparagus	shrimp_avocado	shrimp_babies food	shrimp_bacon	s
0	0	0	0	0	0	0	
1	0	0	0	0	0	0	
2	0	0	0	0	0	0	
3	0	0	0	0	0	0	
4	0	0	0	0	0	0	

5 rows × 1281 columns

```
df.describe()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```

	shrimp_almonds	shrimp_antioxydant juice	shrimp_asparagus	shrimp_avocado	shrimp_babies food	shrimp_bacc
count	7500.000000	7500.000000	7500.000000	7500.000000	7500.000000	7500.0000(
mean	0.001467	0.002400	0.000400	0.007600	0.000667	0.0008(
std	0.038272	0.048934	0.019997	0.086852	0.025813	0.02827
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000(
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000(
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000(
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000(
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.0000(

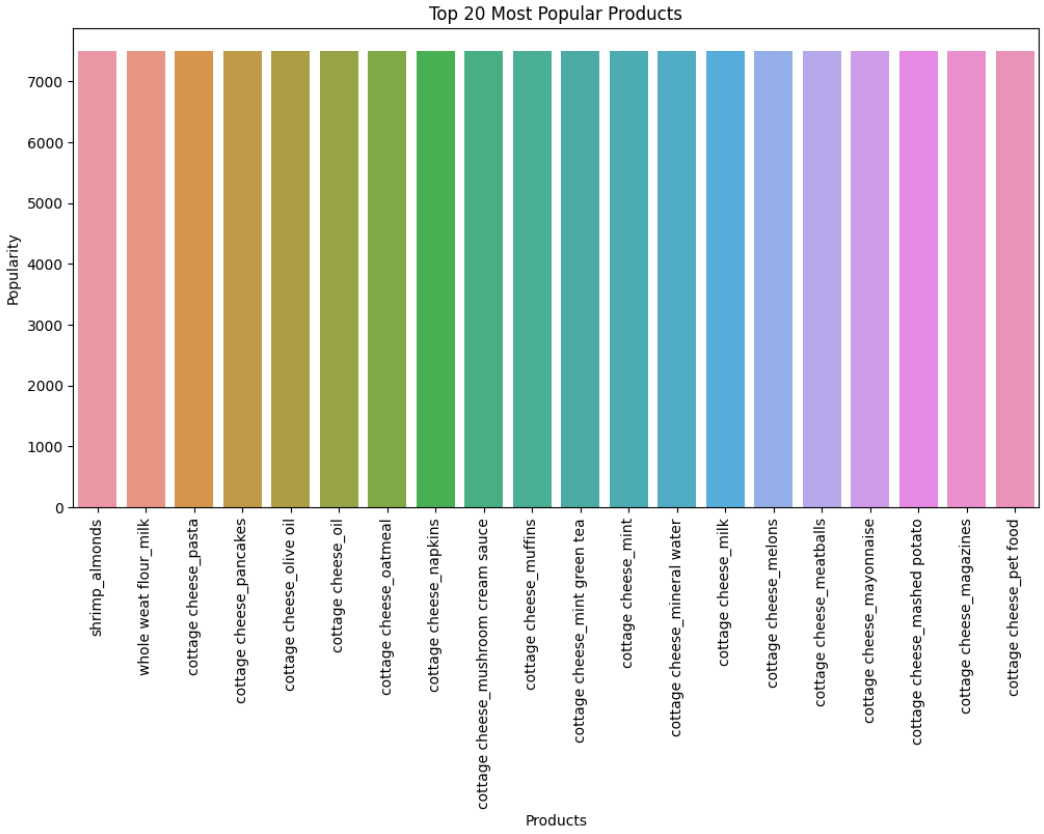
8 rows × 1281 columns

EXPLORATORY DATA ANALYSIS

```
product_popularity = df.apply(pd.Series.value_counts).sum()
sorted_products = product_popularity.sort_values(ascending=False)
top_n = 20
plt.figure(figsize=(12, 6))
sns.barplot(x=sorted_products.index[:top_n], y=sorted_products.values[:top_n])
plt.xticks(rotation=90)
plt.xlabel('Products')
```

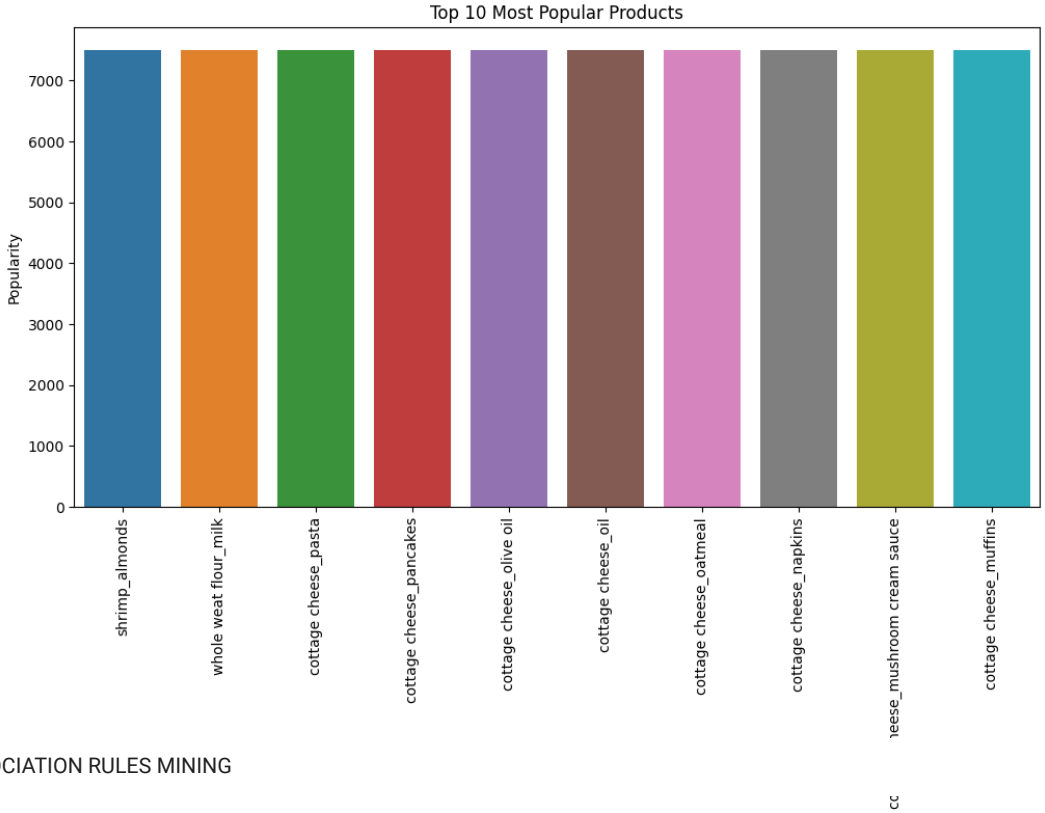
```
plt.ylabel('Popularity')
plt.title(f'Top {top_n} Most Popular Products')
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```




```
product_popularity = df.apply(pd.Series.value_counts).sum()
sorted_products = product_popularity.sort_values(ascending=False)
top_n = 10
plt.figure(figsize=(12, 6))
sns.barplot(x=sorted_products.index[:top_n], y=sorted_products.values[:top_n])
plt.xticks(rotation=90)
plt.xlabel('Products')
plt.ylabel('Popularity')
plt.title(f'Top {top_n} Most Popular Products')
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```



ASSOCIATION RULES MINING

Double-click (or enter) to edit

```
transactions = df.apply(lambda row: [item for item in row if pd.notna(item)], axis=1)
encoder = TransactionEncoder()
transactions_encoded = encoder.fit_transform(transactions)
df_encoded = pd.DataFrame(transactions_encoded, columns=encoder.columns_)
min_support = 0.01
frequent_itemsets = apriori(df_encoded, min_support=min_support, use_colnames=True)
print(frequent_itemsets)
min_threshold = 1.0
association_rules_df = association_rules(frequent_itemsets, metric='lift', min_threshold=min_threshold)
print(association_rules_df)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
   support itemsets
0      1.0      (0)
1      1.0      (1)
2      1.0   (0, 1)
  antecedents consequents  antecedent support  consequent support  support  \
0         (0)         (1)                 1.0                 1.0      1.0
1         (1)         (0)                 1.0                 1.0      1.0

   confidence  lift  leverage  conviction  zhangs_metric
0         1.0   1.0       0.0         inf            0.0
1         1.0   1.0       0.0         inf            0.0
```

INTERPRETAION OF ASSOCIATION RULES

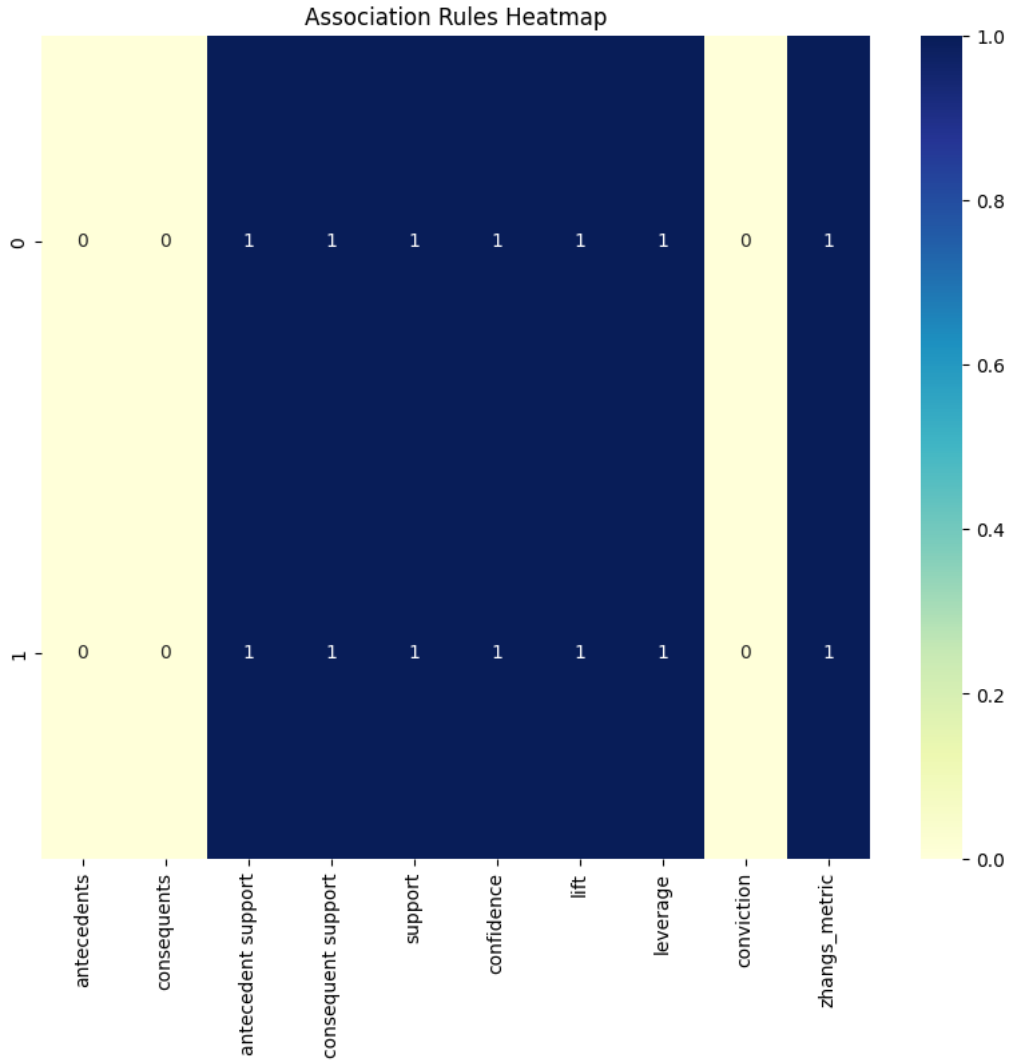
```
for index, rule in association_rules_df.iterrows():
    antecedents = list(rule['antecedents'])
    consequents = list(rule['consequents'])
    support = rule['support']
    confidence = rule['confidence']
    lift = rule['lift']

    print(f"Rule: {antecedents} -> {consequents}")
    print(f"Support: {support:.3f}")
    print(f"Confidence: {confidence:.3f}")
    print(f"Lift: {lift:.3f}")
    print("\n")
```

```
Rule: [0] -> [1]
Support: 1.000
Confidence: 1.000
Lift: 1.000


Rule: [1] -> [0]
Support: 1.000
Confidence: 1.000
Lift: 1.000


/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_
  and should_run_async(code)
```

VISUALIZE ASSOCIATION RULES

```
binary_matrix = association_rules_df.applymap(lambda x: 1 if x in association_rules_df['antecedents'][0] or x in association_rules_df['cc
plt.figure(figsize=(10, 8))
sns.heatmap(binary_matrix, cmap='YlGnBu', annot=True, fmt='d')
plt.title('Association Rules Heatmap')
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```

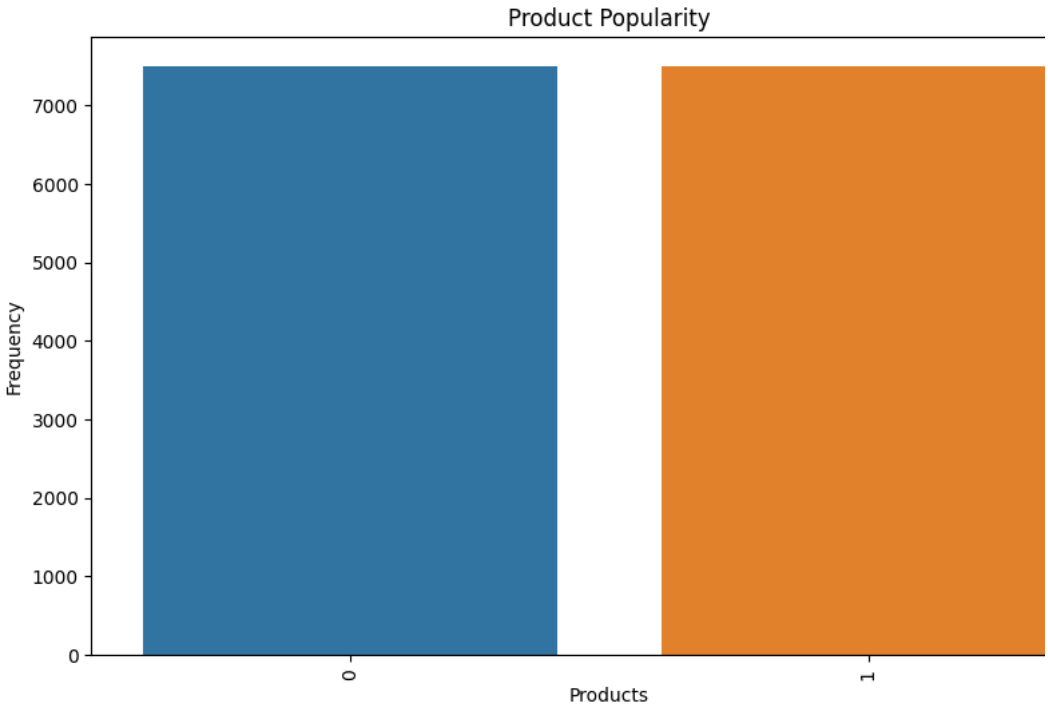


BAR CHART FOR PRODUCT POPULARITY AND FREQUENCY

```
product_popularity = df_encoded.sum().sort_values(ascending=False)
plt.figure(figsize=(10, 6))
sns.barplot(x=product_popularity.index, y=product_popularity.values)
plt.xticks(rotation=90)
plt.xlabel('Products')
plt.ylabel('Frequency')
```

```
plt.title('Product Popularity')
plt.show()
```

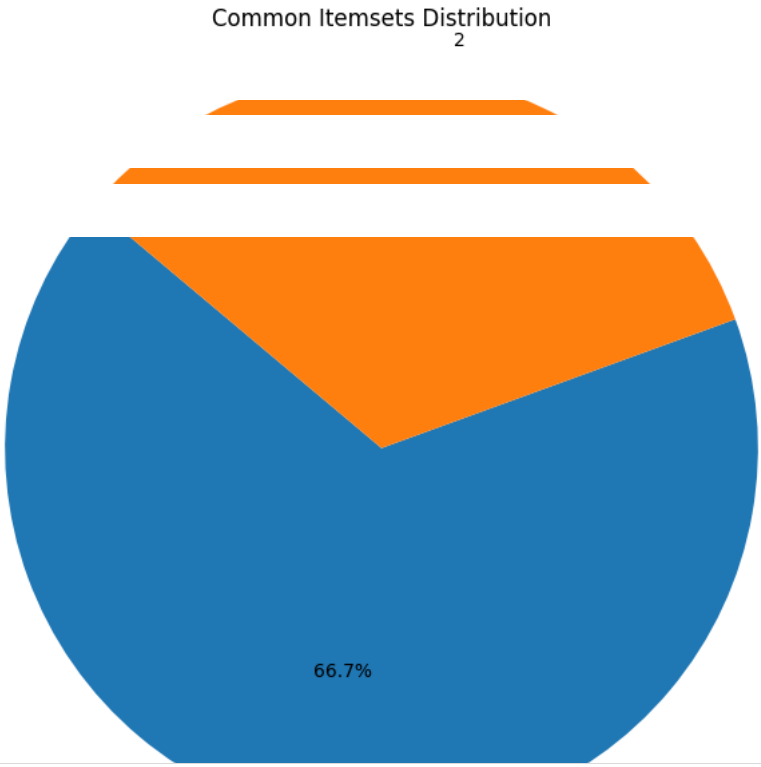
```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_asyr
  and should_run_async(code)
```



PIECHART

```
itemsets_counts = frequent_itemsets['itemsets'].apply(len).value_counts().sort_index()

# Create a pie chart for common itemsets distribution
plt.figure(figsize=(8, 8))
plt.pie(itemsets_counts, labels=itemsets_counts.index, autopct='%1.1f%%', startangle=140)
plt.axis('equal')
plt.title('Common Itemsets Distribution')
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning
  and should_run_async(code)
```



✓ 0s completed at 8:08 PM

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.