Data Mining Lab-3(Apriori algo)

DataSet Used:

Grocery Store dataset used

Link: https://drive.google.com/file/d/1Wd9Q2xTd6AYN3v5Ao4fOje6VrRHC_7oN/view?usp=drive_link

Source Code:

```
import pandas as pd
df1=pd.read_csv('/content/Groceries_dataset.csv')

grocery = df1.groupby(['Member_number',
'Date'])['itemDescription'].apply(list).tolist()
print(grocery)
!pip install mlxtend --quiet
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori,association_rules
te=TransactionEncoder()
te_array=te.fit(grocery).fit_transform(grocery)
print(te_array)
freq_items=apriori(df2,use_colnames=True,min_support=0.01)
display(freq_items)
rules = association_rules(freq_items, metric='confidence',
min_threshold=0.01)
display(rules)
import matplotlib.pyplot as plt

plt.plot(rules.index, rules['support'], marker='^', label='support',
color='red')
plt.plot(rules.index, rules['lift'], marker='o', label='lift',
color='blue')
plt.plot(rules.index, rules['confidence'], marker='+', label='confidence',
color='green')

plt.xlabel("Rule Index")
plt.ylabel("Metric Value")
plt.title("Association Rule Metrics")
plt.legend()
plt.grid(True)
plt.show()
plt.hist(rules['confidence'],color='red')
plt.title("Confidence")
```

```
plt.xlabel("confidence")
plt.ylabel("Frequency")
plt.tight_layout()
plt.hist(rules['lift'],color='red')
plt.xlabel("lift")
plt.ylabel("frequency")
plt.title("Lift")
plt.tight_layout()
import seaborn as sns
import matplotlib.pyplot as plt

plot_data = rules[['lift', 'confidence', 'support']]
sns.pairplot(data=plot_data)
plt.suptitle("Pairplot of Association Rule Metrics", y=1)
plt.show()
sns.heatmap(plot_data.corr(),annot=True)
```

Screenshots:

DM_lab4.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Share Gemini Connect v2-8 TPU

69 rows × 2 columns

```
rules = association_rules(freq_items, metric='confidence', min_threshold=0.01)
display(rules)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction	zhangs_metric	jaccard	certainty	kulczynski
0	(other vegetables)	(rolls/buns)	0.122101	0.110005	0.010559	0.086481	0.786154	1.0	-0.002872	0.974249	-0.236553	0.047662	-0.026432	0.091235
1	(rolls/buns)	(other vegetables)	0.110005	0.122101	0.010559	0.095990	0.786154	1.0	-0.002872	0.971117	-0.234091	0.047662	-0.029742	0.091235
2	(other vegetables)	(whole milk)	0.122101	0.157923	0.014837	0.121511	0.769430	1.0	-0.004446	0.958551	-0.254477	0.055948	-0.043241	0.107730
3	(whole milk)	(other vegetables)	0.157923	0.122101	0.014837	0.093948	0.769430	1.0	-0.004446	0.968928	-0.262461	0.055948	-0.032068	0.107730
4	(whole milk)	(rolls/buns)	0.157923	0.110005	0.013968	0.088447	0.804028	1.0	-0.003404	0.976350	-0.224474	0.055000	-0.024222	0.107711
5	(rolls/buns)	(whole milk)	0.110005	0.157923	0.013968	0.126974	0.804028	1.0	-0.003404	0.964550	-0.214986	0.055000	-0.036752	0.107711
6	(soda)	(whole milk)	0.097106	0.157923	0.011629	0.119752	0.758296	1.0	-0.003707	0.956636	-0.260917	0.047776	-0.045329	0.096694
7	(whole milk)	(soda)	0.157923	0.097106	0.011629	0.073635	0.758296	1.0	-0.003707	0.974663	-0.274587	0.047776	-0.025995	0.096694
8	(yogurt)	(whole milk)	0.085879	0.157923	0.011161	0.129961	0.822940	1.0	-0.002401	0.967861	-0.190525	0.047975	-0.033206	0.100317
9	(whole milk)	(yogurt)	0.157923	0.085879	0.011161	0.070673	0.822940	1.0	-0.002401	0.983638	-0.203508	0.047975	-0.016634	0.100317

LINE PLOT

Variables Terminal

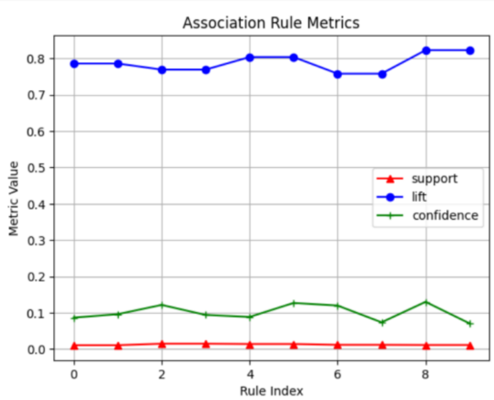
DM_lab4.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Share Gemini Connect v2-8 TPU

```
plt.title("Association Rule Metrics")
plt.legend()
plt.grid(True)
plt.show()
```



Histogram

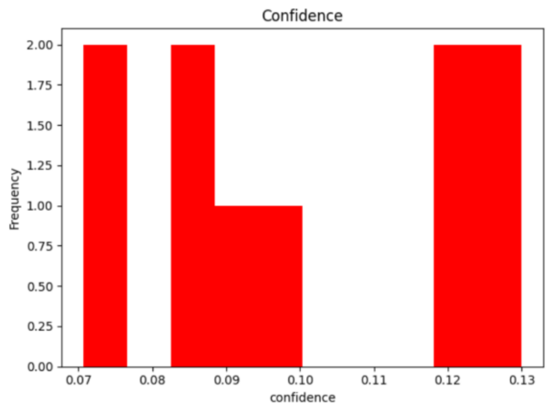
Variables Terminal

DM_lab4.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all Share Gemini Connect v2-8 TPU

```
plt.ylabel("Frequency")
plt.tight_layout()
```




```
plt.hist(rules['lift'],color='red')
plt.xlabel("lift")
plt.ylabel("frequency")
plt.title("Lift")
```

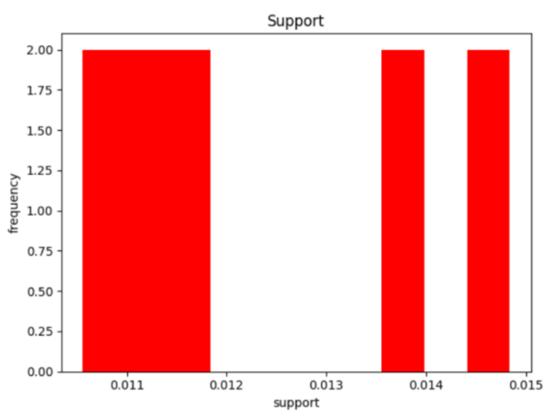
Variables Terminal

DM_lab4.ipynb

File Edit View Insert Runtime Tools Help

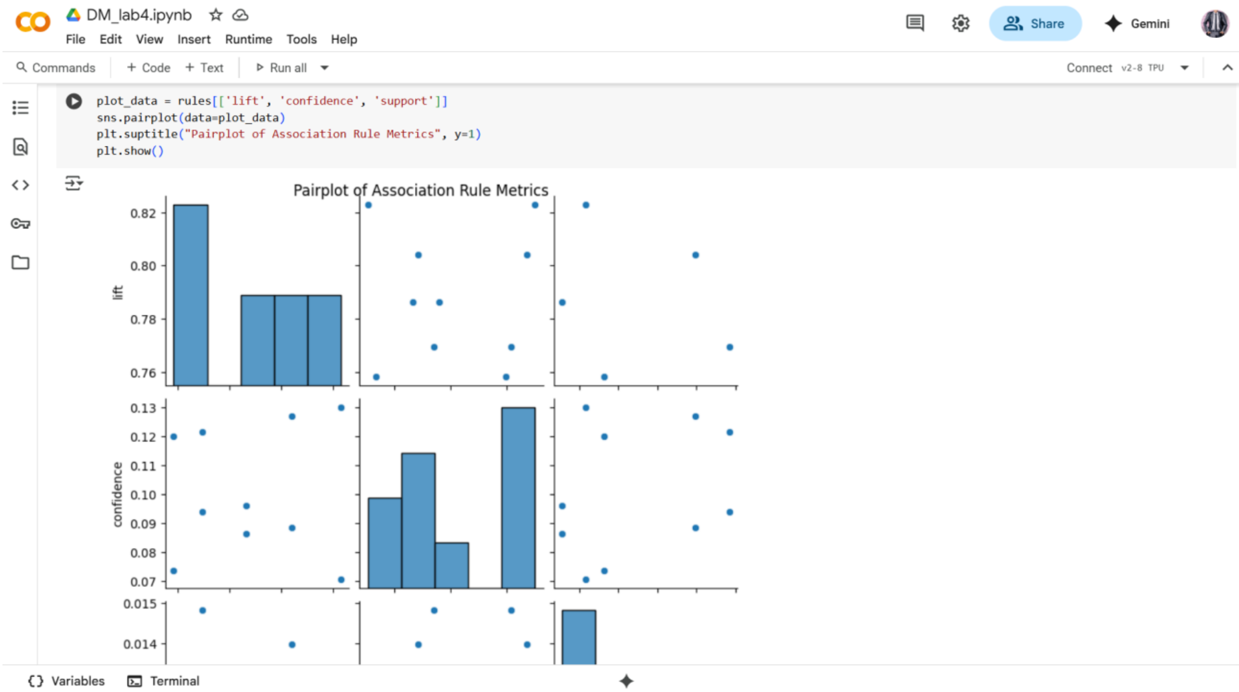
Commands + Code + Text Run all Share Gemini Connect v2-8 TPU

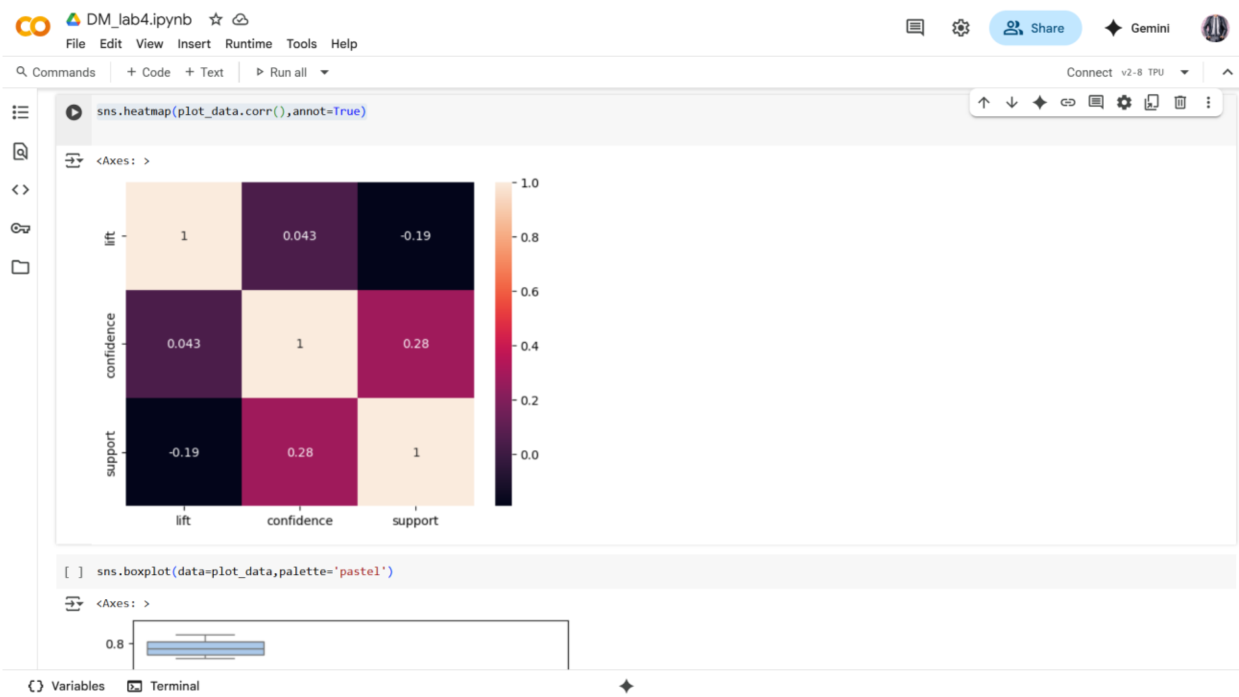
```
plt.hist(rules['support'],color='red')
plt.xlabel("support")
plt.ylabel("frequency")
plt.title("Support")
plt.tight_layout()
```




```
import seaborn as sns
import matplotlib.pyplot as plt
```

Variables Terminal





Observations:

After applying the Apriori algorithm and analyzing the results, a few interesting patterns were noticed:

Commonly Purchased Items
Some items appear very frequently in transactions. Products like *whole milk, bread, yogurt,* and *rolls/buns* are among the most purchased items. This shows that customers tend to buy these as their daily essentials.

Frequent Itemsets and Associations
When we look at combinations of items, we see that certain products are often bought together. For example, customers who buy *yogurt* are also quite likely to buy *whole milk*. These associations highlight natural pairings in people's shopping habits.

Strength of Rules (Support, Confidence, Lift)

Many rules have **low support**, meaning they happen rarely, but a few rules occur often and are more reliable.

Some rules show **high confidence**, which means if a customer buys one item, there is a good chance they will also buy the associated item.

Rules with **lift greater than 1** show strong and useful relationships. For instance, `{yogurt} → {milk}` with a high lift suggests that this is more than just a coincidence.

Patterns from Graphs

From the histograms, most rules had **moderate confidence** but only a few stood out with very high values.

The lift histogram shows that only a handful of rules have a lift well above 1, meaning only some rules truly reflect strong product connections.

The pairplot revealed a clear trend: rules with high support usually had lower lift (because common items are purchased with many different things), while rare combinations sometimes gave very high lift.

The heatmap further confirmed this by showing that support and lift are not strongly correlated.

Practical Insights
The most valuable rules are those with **moderate support, high confidence, and high lift**. These can be used for store promotions or product bundling. For example, pairing *yogurt and milk* or *bread and rolls* in offers could encourage more sales.