

**1.Improving student retention with data driven analysis [Use Excel with graph] – Refer analytics in action3.1.**

**Step 1: Prepare Your Data**

Open **MS Excel** and type this data:

Student	Attendance (%)	Marks (%)	Retained (Yes/No)
S1	90	85	Yes
S2	70	60	No
S3	95	80	Yes
S4	60	50	No
S5	85	75	Yes

**Step 2: Create a Pivot Table**

1. Select the whole table (all rows and columns).
2. Go to the **Insert tab** → Click **PivotTable**.
3. A pop-up will ask:
  - *Where do you want the PivotTable to be placed?*
  - Choose **New Worksheet** → Click **OK**.

Now a blank PivotTable appears.

**Step 3: Arrange the Fields**

On the right, you'll see “PivotTable Fields”.

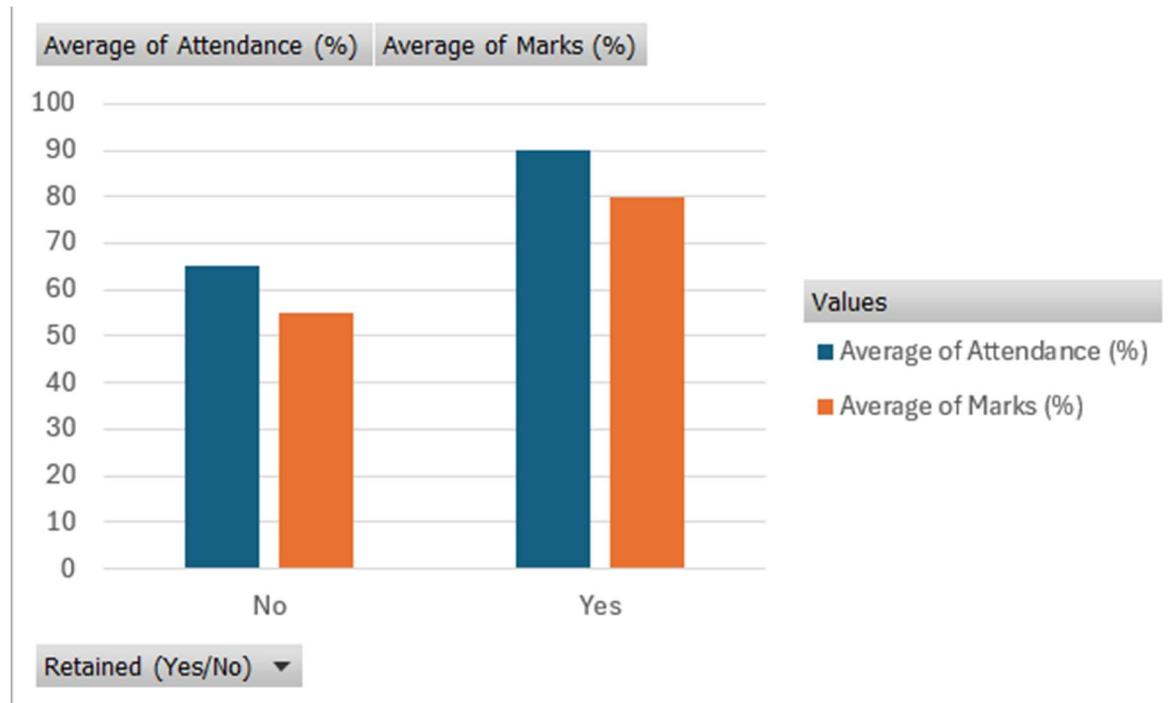
- Drag **Retained (Yes/No)** → **Rows**.
- Drag **Attendance (%)** → **Values**.
- Drag **Marks (%)** → **Values**.

👉 Excel will automatically calculate **average attendance and marks** for “Yes” and “No”.

Row Labels	Average of Attendance (%)	Average of Marks (%)
No	65	55
Yes	90	80
<b>Grand Total</b>	<b>80</b>	<b>70</b>

#### Step 4: Create a Chart

1. Click anywhere inside the Pivot Table.
2. Go to **Insert → Column Chart (Bar Chart)**.
3. Now you see a chart:
  - o Blue bar for Attendance.
  - o Orange bar for Marks.
  - o X-axis: Retained = Yes / No.



#### Step 5: Understand the Results

- Students who are **Retained (Yes)** have **higher attendance (90%) and marks (80%)**.
- Students who are **Not Retained (No)** have **lower attendance (65%) and marks (55%)**.

#### Conclusion:

Students with poor attendance and marks are more likely to drop out.

## **2. Executive dashboard design for a given business analytics scenario using Tableau Public**

### **Step 1: Install Tableau Public**

1. Go to <https://www.tableau.com/products/public/download>
2. Download and install **Tableau Public** (free).

### **Step 2: Prepare Sample Dataset & save it**

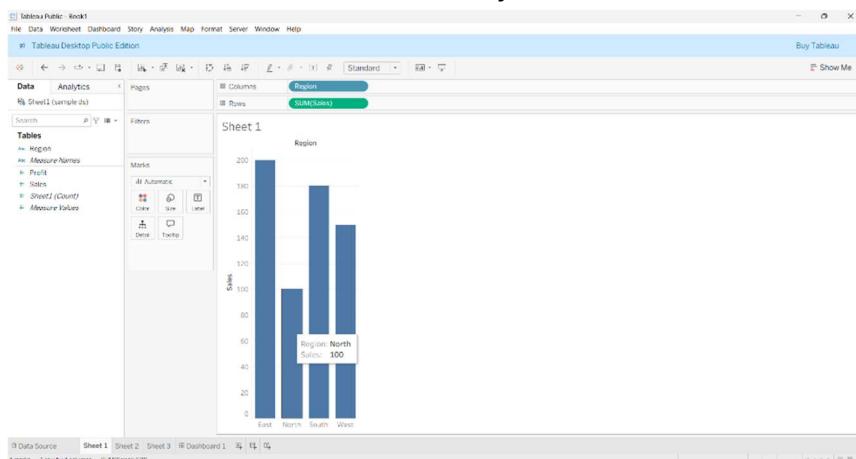
Region	Sales	Profit
East	200	50
West	150	30
North	100	20
South	180	40

### **Step 3: Import Data into Tableau**

1. Open **Tableau Public**.
2. On the start page → Click **Connect** → **Microsoft Excel**.
3. Browse and select **Saved file**
4. Drag the sheet into the **workspace area** (Tableau will read the data).

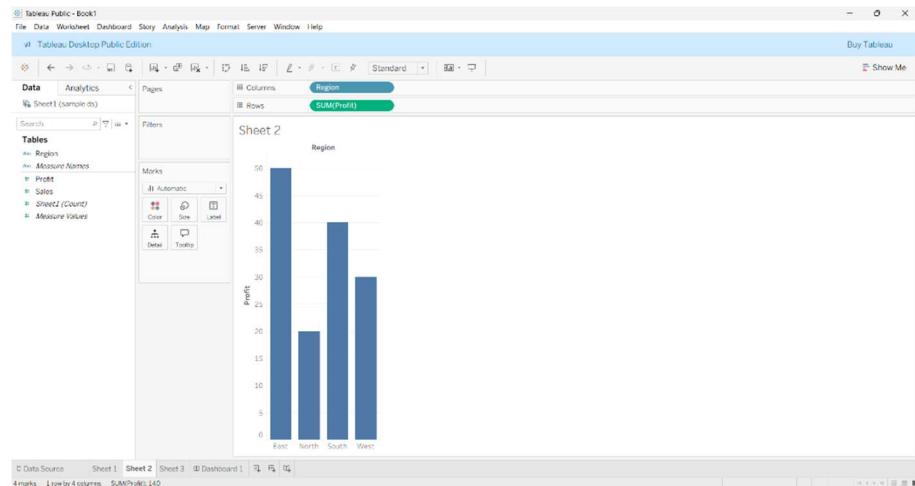
### **Step 4: Create First Chart (Sales by Region)**

1. Go to **Worksheet 1**.
2. Drag **Region** → Columns.
3. Drag **Sales** → Rows.
4. Tableau creates a **Bar Chart** automatically, this shows Sales for each Region.



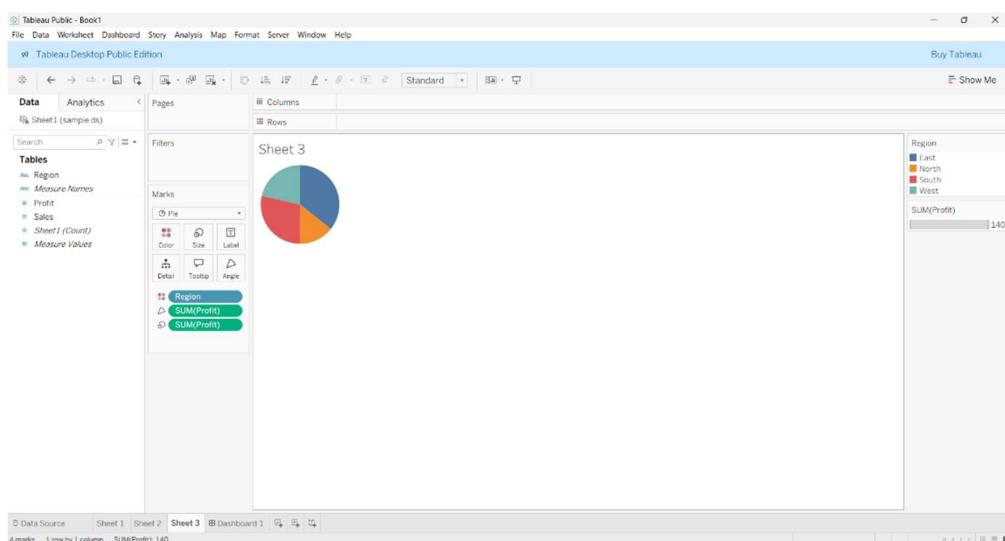
## Step 5: Create Second Chart (Profit by Region)

1. Open a new **Worksheet 2**.
2. Drag **Region** → Columns.
3. Drag **Profit** → Rows.
4. You get a **Bar Chart for Profit**.



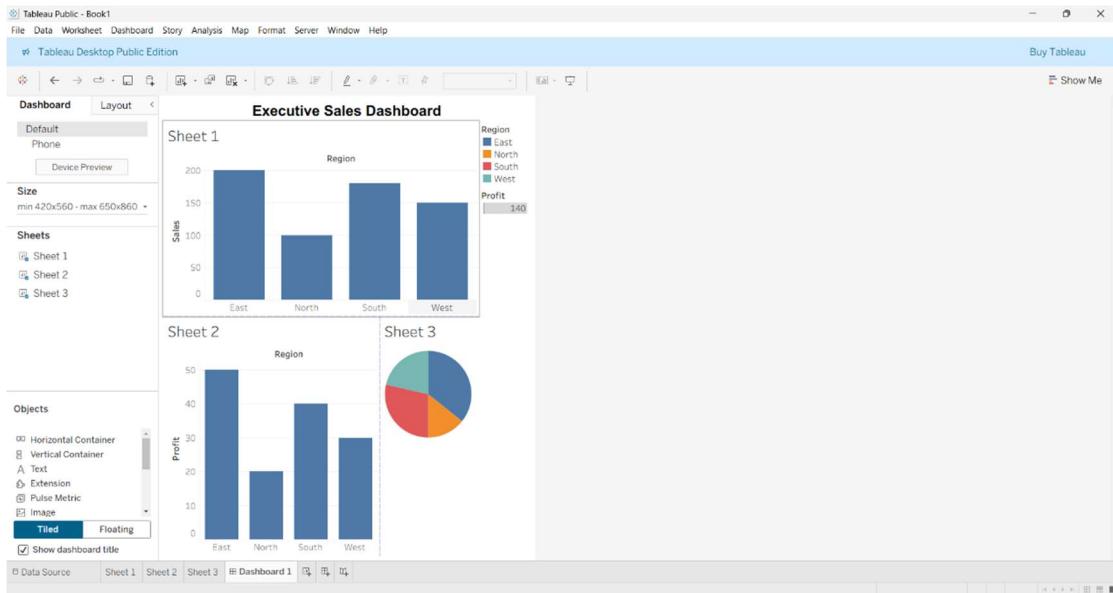
## Step 6: Create a Pie Chart (Profit Contribution)

1. Open **Worksheet 3**.
2. Drag **Profit** → Rows.
3. Click on the chart type → Choose **Pie Chart**.
4. Drag **Region** → Color (each region gets a different color).



## Step 7: Build the Executive Dashboard

1. Click **Dashboard** → **New Dashboard**.
2. A blank dashboard appears.
3. Drag **Worksheet 1**, **Worksheet 2**, and **Worksheet 3** into the dashboard.
4. Arrange them neatly (Bar Chart, Profit Bar, Pie Chart).
5. Add a **Title**: “Executive Sales Dashboard”.



## Step 8: Save and Share

1. Go to **File** → **Save to Tableau Public** (you'll need a free account).
2. Your dashboard is now **online and interactive**.

### **3. Generate visual analytics for a given business tasks and data using Tableau Public.**

#### **Step 1: Prepare a Sample Dataset**

Let's use a **Sales dataset** (you can make in Excel/CSV):

<b>Category</b>	<b>Sales</b>	<b>Profit</b>
-----------------	--------------	---------------

Furniture	300	60
-----------	-----	----

Electronics	500	120
-------------	-----	-----

Clothing	400	90
----------	-----	----

#### **Step 2: Import Data into Tableau**

1. Open Tableau → **Connect** → **Text File / Excel**.
2. Select sales.csv (or Excel file).
3. Tableau will show a preview of your data.

#### **Step 3: Create a Bar Chart (Category vs Sales)**

1. Drag **Category** → Columns.
2. Drag **Sales** → Rows.
3. Tableau automatically creates a **Bar Chart**.
4. Format colors, labels if you like.

#### **Step 4: Create a Side by side Chart (Category vs Profit)**

1. Create a **New Worksheet**.
2. Drag **Category** → Columns.
3. Drag **Profit** → Rows.
4. From the chart options, select **side by side chart**

#### **Step 5: Create a Pie Chart (Category Contribution)**

1. Create another **New Worksheet**.
2. Drag **Sales** → Rows.
3. Drag **Category** → Color (in Marks panel).
4. Change chart type to **Pie**.

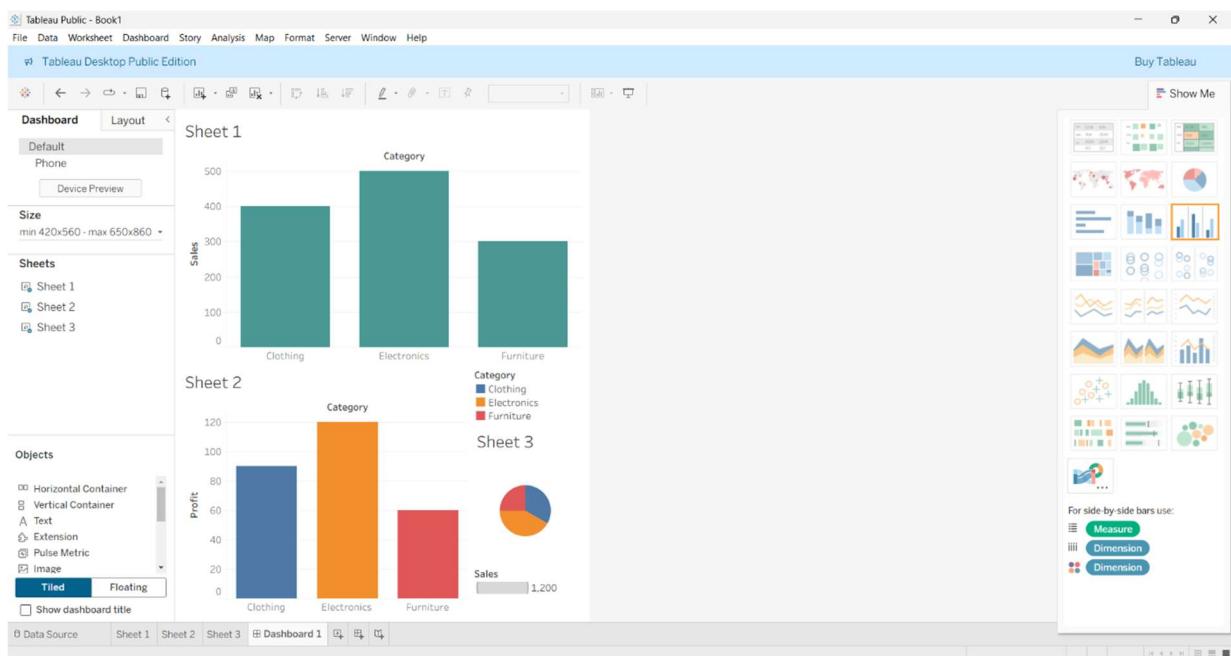
- Now you have a pie chart showing share of each category.

### Step 6: Combine into a Dashboard

- Click **Dashboard** → **New Dashboard**.
- Drag your three charts (Bar, Line, Pie) into the dashboard area.
- Adjust layout → Arrange neatly.
- Add a **Filter (Category)** so you can interact with data.

### Step 7: Save & Share

- Save your Tableau Public workbook.
- (Optional) Click **File** → **Save to Tableau Public** to upload online.



**4. Enhancing customer experience with predictive analytics and data mining by taking suitable business scenario (Use Weka/RapidMiner/Spark/R/Microsoft power BI).**

**Step 1: Prepare Dataset**

In Excel (customer\_data.xlsx), update the table like this:

**Customer Age Income Purchased**

C1	22	20000	No
C2	25	25000	No
C3	28	30000	Yes
C4	35	50000	Yes
C5	40	60000	Yes
C6	30	35000	(blank)

**Step 2: Load Data into Power BI**

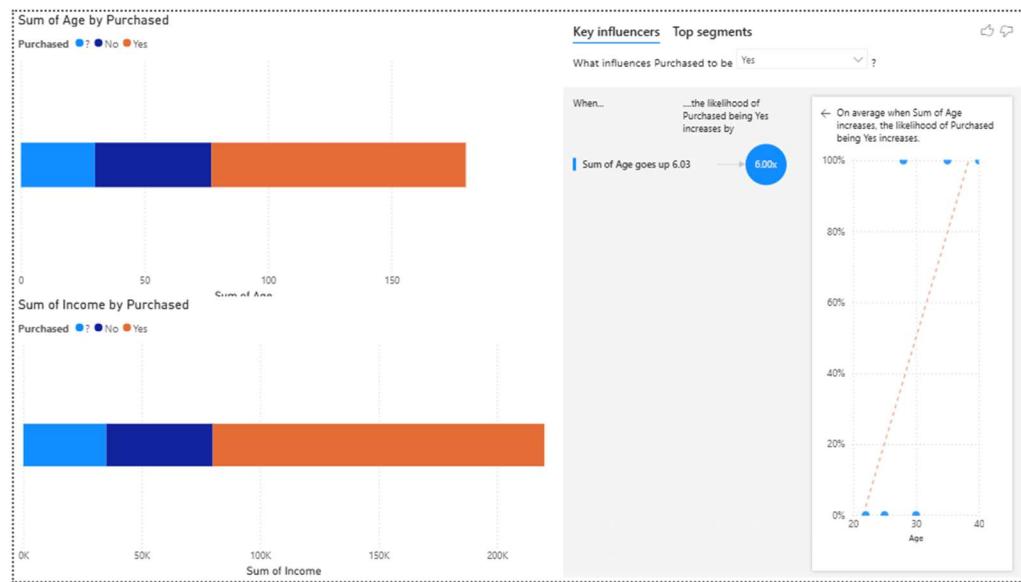
1. Open **Power BI Desktop**.
2. **Home → Get Data → Excel → customer\_data.xlsx → Load.**
3. In **Data View**, check:
  - Age → Whole Number
  - Income → Whole Number
  - Purchased → Text (Yes/No/blank).

**Step 3: Create Simple Charts**

1. In **Report View**, click on a **Bar Chart** visual.
2. Drag **Age** to **X-axis**, and **Purchased** to **Legend**.
  - This will show who purchased vs not by age.
3. Add another chart:
  - Use **Income** on **X-axis** and **Purchased** on **Legend**.
  - This shows purchase pattern by income.

## Step 4: Use “Key Influencers” for Prediction-Like Insight

1. Insert **Key Influencers** visual.
2. In **Analyze** → drag Purchased.
3. In **Explain By** → drag Age and Income.



Now Power BI will show rules like:

- **Income > 28,000 → More likely Yes**
- **Age < 26 → More likely No**

Since **C6 (Age=30, Income=35000)** fits the “Income > 28,000” rule, we can **predict C6 as “Yes”** (likely to purchase).

## Step 5: Interpret Result

- Even though C6 had no Purchased value, Power BI analysis suggests it belongs to the **Yes group**.
- This is how we simulate predictive analytics in a simple way without advanced ML.

**5. Cluster analysis using k-means algorithm for a given customer data set (use Python/R/any other tool).**

Pgm:

```
import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

data = pd.read_csv("customers.csv")
X = data[["Age", "Income"]]

kmeans = KMeans(n_clusters=2, random_state=0)
data["Cluster"] = kmeans.fit_predict(X)

print(data)

plt.scatter(X["Age"], X["Income"], c=data["Cluster"], cmap="rainbow")
plt.xlabel("Age")
plt.ylabel("Income")
plt.title("Customer Clusters")
plt.show()
```

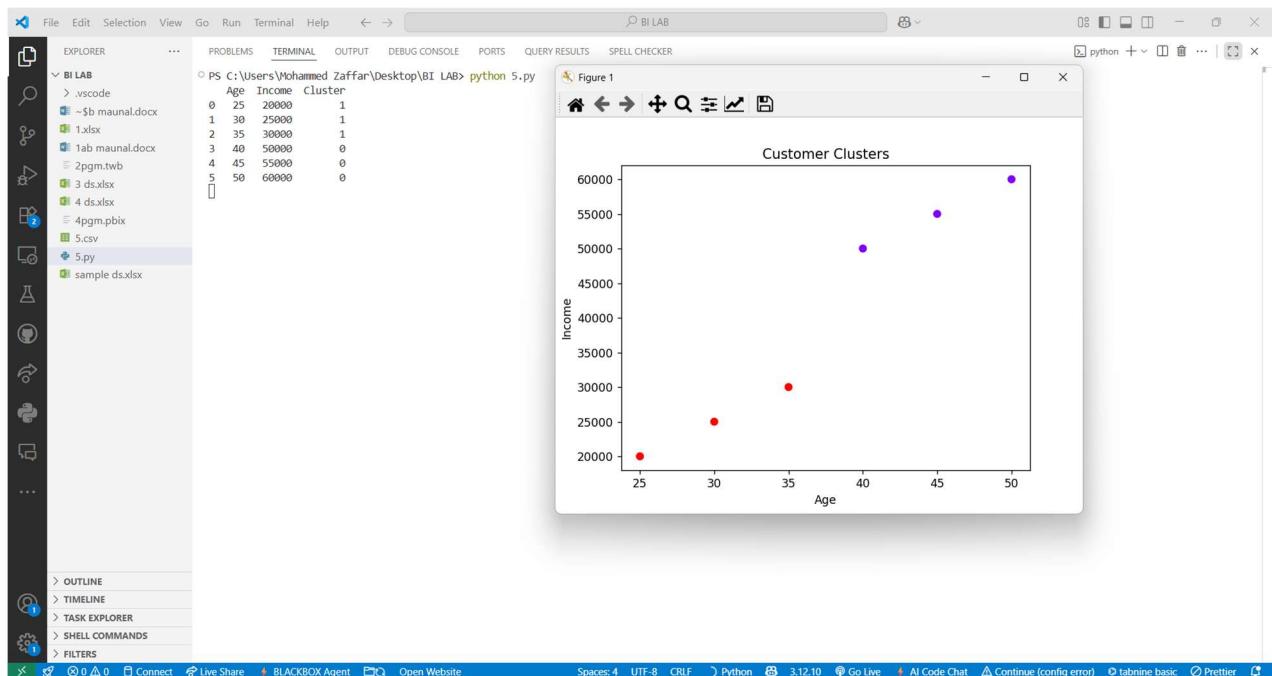
**Csv file:**

Create a simple dataset `customers.csv` with two columns:

```
CSV

Age,Income
25,20000
30,25000
35,30000
40,50000
45,55000
50,60000
```

## Output:



**6. Identify frequent item sets using the Apriori algorithm for a given transaction data set (use Python/R/any other tool).**

Pgm:

```
 6.py > ...
1 import pandas as pd
2 from mlxtend.frequent_patterns import apriori, association_rules
3
4 data = pd.read_csv("6.csv")
5 basket = data.groupby(['TransactionID','Item'])['Item'].count().unstack().fillna(0)
6 basket = basket.applymap(lambda x: 1 if x > 0 else 0)
7
8 rules = association_rules(apriori(basket, min_support=0.2, use_colnames=True),
9                           metric="lift", min_threshold=1)
10 print(rules[['antecedents','consequents','support','confidence']])
11 |
```

Csv file:

TransactionID, Item

1, Bread  
1, Milk  
2, Milk  
2, Eggs

Output:

	antecedents	consequents	support	confidence
0	(Milk)	(Bread)	0.5	0.5
1	(Bread)	(Milk)	0.5	1.0
2	(Eggs)	(Milk)	0.5	1.0
3	(Milk)	(Eggs)	0.5	<u>0.5</u>

**7. Use a dataset of customer product reviews (e.g., Amazon reviews) to classify the sentiment of each review as positive, negative, or neutral using a pre-trained machine learning model (e.g., Naïve Bayes). Evaluate the accuracy of your sentiment classifier. (use Python/R/any other tool)**

```
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB

data = pd.read_csv("7.csv")
X = CountVectorizer().fit_transform(data['text'])
y = data['label']

X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.3)
model = MultinomialNB().fit(X_train,y_train)

print("Accuracy:", model.score(X_test,y_test))
```

Csv file:

```
text,label
"I love this product",positive
"Terrible experience",negative
"It was okay, not great",neutral
"Absolutely fantastic quality",positive
"Very bad experience",negative
"Decent for the price",neutral
"Exceeded my expectations",positive
"Terrible, would not recommend",negative
"Not worth the money",negative
```

Output:

- PS C:\Users\Mohammed Zaffar\Desktop\BI LAB> python 7.py  
Accuracy: 0.3333333333333333

**8. Use text mining techniques to analyse a collection of news articles. Identify the most frequent terms and perform topic modelling using Latent Dirichlet Allocation (LDA) to find hidden topics within the articles.(use Python/R/any other tool).**

```
import pandas as pd
from sklearn.decomposition import LatentDirichletAllocation
from sklearn.feature_extraction.text import CountVectorizer

docs = pd.read_csv("8.csv")['article']
X = CountVectorizer(stop_words='english').fit_transform(docs)

lda = LatentDirichletAllocation(n_components=2, random_state=0).fit(X)
terms = CountVectorizer(stop_words='english').fit(docs).get_feature_names_out()

for i, topic in enumerate(lda.components_):
    print("Topic", i, ":", [terms[j] for j in topic.argsort()[-5:]])
```

Csv file:

```
article
"The stock market rose today due to tech shares."
"The economy is showing signs of recovery."
"The new policy has been met with mixed reactions."
```

O/p:

```
Topic 0 : ['new', 'economy', 'recovery', 'showing', 'signs']
Topic 1 : ['today', 'tech', 'shares', 'market', 'stock']
```

**9. Given a dataset representing a social network (e.g., Twitter follower data), create a graph and perform Social Network Analysis (SNA) to find the most influential users using centrality measures like degree, closeness, and betweenness centrality. (use Python/R/any other tool).**

```
import matplotlib.pyplot as plt
import networkx as nx
import pandas as pd
data = pd.read_csv("9.csv")

G = nx.from_pandas_edgelist(data, 'source', 'target', create_using=nx.DiGraph())

print("Degree:", nx.degree_centrality(G))
print("Closeness:", nx.closeness_centrality(G))
print("Betweenness:", nx.betweenness_centrality(G))

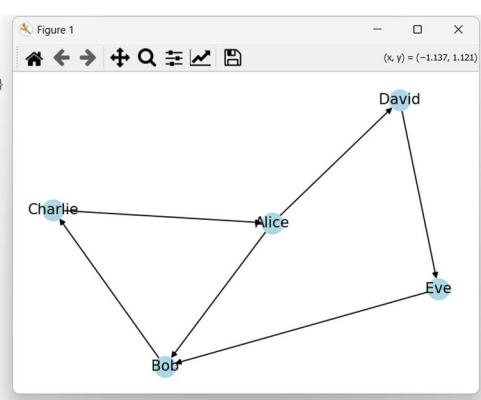
nx.draw(G, with_labels=True, node_color="lightblue", arrows=True)
plt.show()
```

**Csv file:**

```
source,target
Alice,Bob
Bob,Charlie
Charlie,Alice
Alice,David
David,Eve
Eve,Bob
```

**O/p:**

```
PROBLEMS TERMINAL OUTPUT DEBUG CONSOLE PORTS QUERY RESULTS SPELL CHECKER
PS C:\Users\Mohammed Zaffar\Desktop\BI LAB> python 9.py
Degree: {'Alice': 0.75, 'Bob': 0.75, 'Charlie': 0.5, 'David': 0.5, 'Eve': 0.5}
Closeness: {'Alice': 0.4, 'Bob': 0.6666666666666666, 'Charlie': 0.5, 'David': 0.4, 'Eve': 0.4}
Betweenness: {'Alice': 0.5, 'Bob': 0.5, 'Charlie': 0.5, 'David': 0.25, 'Eve': 0.25}
```



**10. Using web scraping tools, extract the content from a set of web pages (e.g., news, blogs). Analyse the text to identify trends and topics. (use Python/R/any other tool).**

```
import requests
from bs4 import BeautifulSoup

url = "https://news.ycombinator.com/"
soup = BeautifulSoup(requests.get(url).text, "html.parser")

for t in soup.select(".titleline a")[:5]:
    print(t.text)
```

O/p:

GrapheneOS and Forensic Extraction of Data  
grapheneos.org  
Gregg Kellogg has passed away  
w3.org  
Reshaped is now open source

**11. Extract user comments from a web forum or blog post using web scraping.**  
**Perform sentiment analysis on the extracted data and identify the overall tone of the discussion. Visualize the sentiment distribution using a pie chart or bar graph. (use Python/R/any other tool).**

```

import matplotlib.pyplot as plt
import requests
from bs4 import BeautifulSoup
from textblob import TextBlob

url = "https://www.bbc.com/news/world"
soup = BeautifulSoup(requests.get(url).text, "html.parser")

comments = [c.text for c in soup.find_all('p')]

sentiments = {"positive":0, "negative":0, "neutral":0}

for c in comments[:10]:
    polarity = TextBlob(c).sentiment.polarity
    if polarity > 0:
        sentiments["positive"] += 1
    elif polarity < 0:
        sentiments["negative"] += 1
    else:
        sentiments["neutral"] += 1
    print(c, "=>", polarity)

plt.pie(sentiments.values(), labels=sentiments.keys(), autopct='%1.1f%%', colors=['green','red','gray'])
plt.title("Sentiment Distribution of Comments")
plt.show()

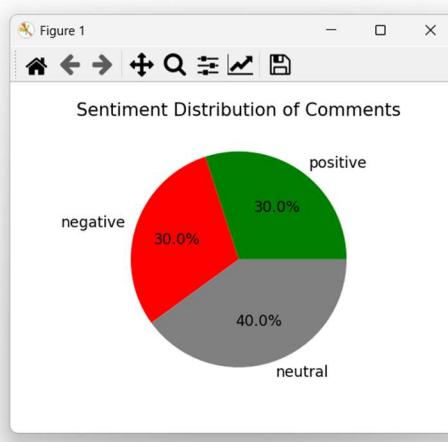
```

**O/p:**

```

PS C:\Users\Mohammed Zaffar\Desktop\BI LAB> python 11.py
Witnesses describe screaming and running to flee the scene after hearing the fatal gunshot that killed Kirk. => -0.2
Charlie Kirk's killing is another episode of gun violence in America and the latest in a line of recent political violence. => 0.1666666666666666
The Irish broadcaster said that Ireland's participation in the contest "would be unconscionable given the ongoing and appalling loss of lives in Gaza". => -0.175
Whether the drone incursion was deliberate or accidental, Moscow will be watching Nato's response to it closely. => 0.0
Earlier this month, Chinese nationals working with Nasa suddenly found themselves denied access. => 0.0
Charlie Kirk's killing is another episode of gun violence in America and the latest in a line of recent political violence. => 0.1666666666666666
Earlier this month, Chinese nationals working with Nasa suddenly found themselves denied access. => 0.0625
Foreign nationals are blamed for the incident, which has striking similarities with other recent provocations. => 0.0
The highest court rules that a law barring men from taking the surnames of their wives is discriminatory. => 0.0
The Israeli military says it hit military targets, but the Houthis say only civilian sites were targeted. => -0.0666666666666667

```



**12. Using web crawling techniques, analyse the internal and external link structure of a website. Create a graph of the website's hyperlink structure and identify key pages using Page Rank or HITS algorithm. (use Python/R/any other tool).**

```
from urllib.parse import urljoin
import matplotlib.pyplot as plt
import networkx as nx
import requests
from bs4 import BeautifulSoup

start_url = "https://books.toscrape.com/

page = requests.get(start_url)
soup = BeautifulSoup(page.text, "html.parser")

G = nx.DiGraph()

for link in soup.find_all("a", href=True)[:5]:
    full_url = urljoin(start_url, link["href"])
    G.add_edge(start_url, full_url)

pr = nx.pagerank(G)
print("PageRank:", pr)

nx.draw(G, with_labels=True, node_color="lightblue", arrows=True)
plt.show()
```

o/p:

```
PageRank: {'https://books.toscrape.com/': 0.17094005169645404, 'https://books.toscrape.com/index.html': 0.20726498707588653, 'https://books.toscrape.com/catalogue/category/books_1/index.html': 0.20726498707588653, 'https://books.toscrape.com/catalogue/category/books/travel_2/index.html': 0.20726498707588653, 'https://books.toscrape.com/catalogue/category/books/mystery_3/index.html': 0.20726498707588653}
```

