Unit – 4: Association Analysis	
Problem Defecation, Frequent Item Set generation, Rule generation, compact representation of frequent item sets, FP-Growth Algorithm.	10

Assignment problem-1

Traffic Congestion Prediction

Title: "Predicting Urban Traffic Congestion Using Data Mining Techniques"

Problem Definition

Assignment Problem: Identify and define a data mining problem related to predicting traffic congestion in urban areas. **Bloom's Taxonomy Level:** Analyze **Real-World Issue:** Traffic congestion in cities leads to increased pollution and wasted time. Students will analyze traffic data to identify patterns and predict congestion.

Suggested datasets for this assignment problem.

Dataset: Traffic Prediction Dataset on Kaggle **Description:** This dataset contains hourly traffic data from four different junctions, which can be used to analyze and predict traffic congestion.

https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset

Assignment problem-2

Frequent Item Set Generation

Title: "Analyzing Grocery Store Transactions to Identify Frequent Item Sets"

Assignment Problem: Generate frequent item sets from a dataset of grocery store transactions to identify commonly bought items. **Bloom's Taxonomy Level:** Apply **Real-World Issue:** Understanding customer buying patterns can help stores optimize inventory and improve customer satisfaction.

Dataset: Online Retail Dataset from the UCI Machine Learning Repository **Description:** This dataset contains transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based online retail. It is ideal for generating frequent item sets.

Assignment problem-3

Rule Generation

Title: "Developing Association Rules for Product Recommendations in E-Commerce"

Assignment Problem: Develop association rules from a dataset of online retail transactions to recommend products to customers. **Bloom's Taxonomy Level:** Create **Real-World Issue:** E-commerce platforms can enhance user experience and increase sales by recommending relevant products.

Dataset: Instacart Market Basket Analysis on Kaggle **Description:** This dataset includes over 3 million grocery orders from more than 200,000 Instacart users, providing a rich source for generating association rules.

Assignment problem-4

Compact Representation of Frequent Item Sets

Title: "Efficient Representation of Frequent Item Sets in Medical Prescription Data Using Apriori Algorithm"

Assignment Problem: Use the Apriori algorithm to find and represent frequent item sets compactly in a dataset of medical prescriptions. **Bloom's Taxonomy Level:** Evaluate **Real-World Issue:** Identifying common prescription patterns can help in managing drug inventory and understanding prescription trends.

Dataset: Retail Market Basket Data on Kaggle **Description:** This dataset contains retail market basket data, which can be used to find and represent frequent item sets compactly.

Assignment problem-5

FP-Growth Algorithm

Title: "Identifying Frequent Patterns in Social Media Interactions Using FP-Growth Algorithm"

Assignment Problem: Implement the FP-Growth algorithm to analyze a dataset of social media interactions and identify frequent patterns. **Bloom's Taxonomy Level:** Apply **Real-World Issue:** Social media platforms can use this analysis to understand user behavior and improve content recommendations.

Dataset: Groceries Dataset on Kaggle **Description:** This dataset contains transactions from a grocery store, which can be used to implement the FP-Growth algorithm and analyze frequent patterns.

Assignment problem-6

Environmental Sensor Readings for Air Quality Monitoring

Title: "Monitoring and Analyzing Air Quality Using Environmental Sensor Data"

Assignment Problem: Define a problem and generate frequent item sets from a dataset of environmental sensor readings to monitor air quality. **Bloom's Taxonomy Level:** Analyze and Apply **Real-World Issue:** Monitoring air quality helps in taking timely actions to reduce pollution and protect public health.

Dataset: Air Quality Data Set from the UCI Machine Learning Repository **Description:** This dataset contains sensor readings of air quality, which can be used to monitor and analyze air pollution levels.

Assignment problem-7

Energy Consumption in Smart Homes

Title: "Developing Energy-Saving Tips from Smart Home Energy Consumption Data Using FP-Growth Algorithm"

Assignment Problem: Develop rules using the FP-Growth algorithm from a dataset of energy consumption in smart homes to suggest energy-saving tips. **Bloom's Taxonomy Level:** Create and Apply **Real-World Issue:** Reducing energy consumption in homes can lead to significant cost savings and environmental benefits.

Dataset: Smart Home Dataset with Weather Information on Kaggle **Description:** This dataset includes energy consumption data from smart homes along with weather information, which can be used to develop energy-saving tips.

For each topic, you can structure the document with the following sections to ensure a comprehensive and well-organized report:

1. Abstract

- **Purpose:** Provide a brief summary of the entire document, including the problem statement, methodology, key findings, and conclusions.
- **Length:** Typically 150-250 words.

2. Introduction

- **Purpose:** Introduce the topic, provide background information, and state the objectives of the study.
- **Content:**
    - Background and context of the problem.
    - Importance and relevance of the study.
    - Objectives and scope of the assignment.

3. Literature Review

- **Purpose:** Review existing research and studies related to the topic.
- **Content:**
    - Summary of relevant studies and findings.
    - Identification of gaps in the existing literature.
    - How the current study addresses these gaps.

4. Methodology

- **Purpose:** Describe the methods and techniques used to conduct the study.
- **Content:**
    - Data collection methods.
    - Data preprocessing steps.
    - Algorithms and tools used (e.g., Apriori, FP-Growth).
    - Justification for the chosen methods.

5. Data Analysis

- **Purpose:** Present the analysis of the data.
- **Content:**
    - Description of the dataset.
    - Steps taken to analyze the data.
    - Generation of frequent item sets, association rules, etc.
    - Use of visualizations to support the analysis.

6. Results

- **Purpose:** Present the findings of the study.
- **Content:**
  - Key results and patterns identified.
  - Interpretation of the results.
  - Comparison with expected outcomes or hypotheses.

7. Discussion

- **Purpose:** Discuss the implications of the findings.
- **Content:**
  - Significance of the results.
  - How the findings address the problem statement.
  - Limitations of the study.
  - Suggestions for future research.

8. Conclusion

- **Purpose:** Summarize the key points of the study.
- **Content:**
  - Recap of the objectives and findings.
  - Final thoughts on the study's impact.
  - Recommendations based on the findings.

9. References

- **Purpose:** Cite all the sources and references used in the document.
- **Content:**
  - List of all academic papers, books, websites, and other resources referenced.
  - Follow a consistent citation style (e.g., APA, MLA).

10. Appendices (if applicable)

- **Purpose:** Provide additional material that supports the document.
- **Content:**
  - Supplementary data, charts, or tables.
  - Detailed calculations or code snippets.