

# ANALYSING CUSTOMER BEHAVIOUR IN E-COMMERCE USING SEARCH AND SORT ALGORITHMS

(Working Duration: Monday of Week 5 to Monday of Week 10 – 30 Marks)

## Background:

E-commerce has revolutionized the retail industry, enabling customers to purchase products online with ease. With millions of transactions happening daily, understanding customer behavior is essential for businesses to improve user experience, predict sales trends, and detect fraudulent activities.

One of the key challenges in e-commerce is **efficiently managing and processing vast amounts of customer data**, including purchase histories, product reviews, and browsing behavior. The ability to **search for specific transactions and sort large datasets efficiently** is crucial for optimizing business operations.

Data structures play a vital role in these processes. **Arrays and linked lists** are fundamental data structures that can be used to store and manage e-commerce data, while different **searching and sorting algorithms** determine the efficiency of retrieving and analyzing information.

In this assignment, your goal is to evaluate how different search and sorting algorithms perform when processing e-commerce customer data and explore how the choice of data structure, such as array or linked list, affects performance in terms of time and memory efficiency.

## Question:

In this assignment, you are provided with **two datasets**:

- **transactions.csv** – Contains customer transaction details (Customer ID, Product, Price, Date, Category, Payment Method).
- **reviews.csv** – Contains customer reviews for various products (Product ID, Customer ID, Rating, Review Text).

You are required to apply various search and sort algorithms to analyse these datasets and answer the following:

1. How can you efficiently sort customer transactions by date and display the total number of transactions in both datasets?
2. What percentage of purchases in the “Electronics” category were made using Credit Card payments?  
(*Hint: You will need to search and filter transactions based on category and payment method.*)
3. Which words are most frequently used in product reviews rated 1-star?  
(*Hint: Extract the most common words, sort them by frequency, and present the results.*)

## Minimum Requirements for Lab Work #1

1. Each group may have up to **FOUR (4) members**.
2. This assignment requires the development of **two separate programs**:
  - a) **Array Implementation**
    - a) Implement an **array-based** data structure to store and process the customer data.
    - b) Perform operations like sorting, searching, and filtering transactions based on certain attributes (e.g. data, category).
  - b) **Linked List Implementation**
    - a) Implement a **linked list-based** data structure to store the same data.
    - b) Similar to the array-based approach, allow efficient operations like sorting and searching.
3. **Sorting Requirement:**
  - Sort customer transactions by **date** using suitable **sorting algorithm** for both data structures.
4. **Search Requirement:**
  - Implement **search algorithms** to filter and retrieve relevant transactions based on specific conditions (e.g. category, payment method).
5. **Review Analysis:**
  - Process **text-based customer reviews** and identify the most frequently used words in **negative reviews** (1-star ratings).
  - Sort and display the top frequent words.
6. **Performance Comparison:**
  - Compare the time and memory efficiency of search and sorting algorithms for **Arrays vs Linked Lists**.
7. **Error Handling:**
  - Implement **robust error handling and data validation** to address unexpected formats or missing data.

**Submission Guidelines #1: Program and Video Submission – Lab Work #1 (15 Marks)**

1. **Programming Language:** Use C++ to develop both programs.
2. **No built-in containers** like `<vector>` or `<list>`, you must create your own data structures.
3. **ZIP Submission:**
  - Include only .cpp, .hpp, and .csv/text files.
  - Follow this file naming format:  
`<GroupNo>_<TeamLeaderID>_<1stMemberID>_<2ndMemberID>_<3rdMemberID>.zip`
  - For example, "G1\_TP012345\_TP014556\_TP067554\_TP034325.zip"
4. **Video Recording:**
  - Each team leader must also upload **ONE (1)** video recording and the maximum video duration is **30 minutes**.
  - All members must participate (each member speaks maximum of 5 minutes).
  - Each member must relate their explanation to the workload matrix distribution table provided in the Word document.
  - Compress the final video recording to **under 200MB** before submitting.
  - If the video exceeds the specified time limit (30 minutes), it will only be assessed up to the specified duration.
  - Videos must be recorded at normal speed (1x) and cannot be speed up or adjusted to meet the demo video duration requirements.
  - The video recording file must adhere to the following name format:  
`<GroupNo>_<TeamLeaderID>_<1stMemberID>_<2ndMemberID>_<3rdMemberID>.mp4`
  - For example, "G1\_TP012345\_TP014556\_TP067554\_TP034325.mp4"
5. Refer to the **Page 5** for marking criteria of this Lab Evaluation Work #1 submission.

**Submission Guidelines #2: Documentation Submission – Solution Work (15 Marks)**

What to include in your documentation:

1. **Cover Page**

2. **Workload Matrix Table with Signature**

*Note that this table will impact each member's personal final mark in Lab Evaluation Work #1 based on their stated contribution percentage.*

3. **Theoretical Explanation**

- Explain data manipulations in your chosen data structures.

4. **Input-Output Screenshots**

- Include system input and output screenshots.

5. **Summary Discussions**

- Analyze system efficiency (**execution time, time efficiency, space efficiency, etc.**)
- Summarize and discuss observations made during development.
- Critically evaluate strengths and weaknesses of the code.

6. **Conclusion and Reflection**

- Summarize key findings.
- Discuss potential improvements and system weaknesses.
- Share personal thoughts on the assignment.

7. **References (if applicable)**

- Properly cite external sources (APA format).  
*Failure to reference code will be treated as plagiarism.*

8. **Appendix (if applicable)**

**Submission & Formatting:**

- File Naming Format:  
<GroupNo>\_<TeamLeaderID>\_<Member1ID>\_<Member2ID>\_<Member3ID>.docx
- For example, G1\_TP012345\_TP014556\_TP067554\_TP034325.docx

Deadline:	<b>Monday of Week 10, before 5:00PM</b>
Max Pages:	<b>30 pages</b>
Max Words:	<b>4500 words</b>

## MARKING CRITERIA

**(Lab Evaluation Work #1 - 15 MARKS)**

This Lab Evaluation Work #1 will be evaluated according to the following performance criteria:

Assessment Components	Inclusive	15 Marks
<i>CLO3: Lab Evaluation Task #1 – 30-Minute Video Recording</i> <i>(Assessment will be based on individual performance)</i>		
Practical Skills: Use of Data Structures & Algorithms + Personal Understanding		
Utilization of data structures	Technical Proficiency	
Implementation of relevant algorithms	Technical Proficiency	
Demonstrates understanding of data structures/algorithms used	Comprehensive Understanding	
Justifies choices of structures/algorithms	Insightful Justification	

## MARKING CRITERIA

**(Solution Work - 15 MARKS)**

This solution work will be evaluated according to the following performance criteria:

Assessment Components	Inclusive	15 Marks
<i>CLO2: Solution Work - Documentation</i> <i>(Assessment will be based on group component)</i>		
<b>Theoretical Explanation</b> (e.g., Data Structures, Algorithms)	Clear explanation	
<b>Input Output Screenshots</b>	Adequate Screenshots	
<b>Summary Discussions</b> (Inclusive Time and Space Complexity)	Clear and insightful analysis	
<b>Conclusion &amp; Reflection</b> (Other Relevant / Importance Discussions)	Clearly highlighted and insightful	
<b>Content Organization</b>	Well-structured and logical flow	

Approximation of Total Pages for the documentation: **30 (max).**

Approximation of Words for the documentation: **4500 words (max)**