



STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

BSc. Statistics and Data Science

**SEMESTER PROJECT (Group)**

**OBJECT ORIENTED PROGRAMMING**

**Award: 25 Marks**

**Due : Week 13**

---

**INSTRUCTIONS**

- i. Each group has been given a zip folder containing csv files with a task they should analyze. The allocation for each group is indicated on the online groups excel.
  - ii. Attempt this work in groups.
  - iii. Ensure you backup your work (e.g. using Git – see [How to Use Git](#))
- 

**Project Description:**

Each group is required to develop and implement a Java-based Data Analysis Application (GUI Based on Swing) that demonstrates the complete workflow of working with real-world datasets. Each group has been assigned a unique dataset (corresponding to their group number), as a folder with csv files, and the required analysis.

**You are to:**

1. Design a normalized database schema based on the CSV files provided, either in MySQL, MariaDB, or PostgreSQL.
2. Import the data into the relational tables while ensuring proper data integrity (primary and foreign keys).
3. Develop a Java Program (with GUI) that connects to the database, runs queries to import data from the csv files, and provides meaningful statistical analysis. The analysis will be based on the provided text file in the folder.
4. Visualize the results using appropriate charts and graphs.
5. Document your process with explanations of the schema, queries, results and insight. Ensure to highlight the challenges experienced.

**Tasks:****1. Database Design and Setup**

- Study the dataset assigned to your group (2–4 CSV files).
- Design an ER diagram to capture entities and relationships.
- Create relational tables in SQL with correct data types, primary keys, and foreign keys.
- Import the CSV data into the database.

**2. Java Application Development - Write a Java program to:**

- Upload the CSV files into the database (optional if you imported via SQL first).
- Establish a database connection using JDBC.
- Run queries to extract and aggregate data.
- Generate and display statistical results.
- Produce charts and graphs (e.g., bar charts, line charts, pie charts).

**3. Statistical Analysis**

Each group must complete the core analysis tasks listed in your group's README file (examples: admission rates by program, sales by product, patient diagnoses by severity, etc.).

- Write SQL queries to support the analysis.
- Present numerical results and visualizations.
- Provide a written interpretation of the results.

**4. Documentation & Presentation**

Submit a short report detailing:

- Screenshots/exports of charts and graphs.
- A summary of findings and insights.
- Challenges faced.
- Ensure your Java code is well-documented with comments.
- Prepare to demonstrate your system in class.

---

**Learning Outcomes:**

This project aims to:-

- i. Gain experience in relational database design and implementation.
- ii. Learn to import, query, and analyze data in SQL.
- iii. Apply Java programming skills to handle real-world datasets.
- iv. Practice data visualization and interpretation.
- v. Strengthen teamwork, problem-solving, and presentation skills.