College of Applied Computer Sciences

King Saud University

Big Data ISY 356

Mini Project(Using R Language)

Dead line for the submission the project 20th November 2025

Project Title

Analyzing Retail Sales Data to Identify Customer Purchase Patterns

Objective

- Analyze retail sales data to understand customer buying patterns.
- Identify top-selling products, seasonal trends, and high-value customers.
- Apply R tools for Big Data processing, visualization, and analysis.

Tools & Technology

- R Programming Language
- RStudio – IDE for R
- tidyverse – For data manipulation (dplyr, tidyr)
- data.table – Efficient handling of large datasets
- ggplot2 – For data visualization
- lubridate – For date handling

Dataset Description

Data set uploaded on LMS

- `CustomerID` – Unique identifier for customers
- `ProductID` – Unique identifier for products
- `Category` – Product category
- `Quantity` – Number of items purchased
- `Price` – Price per item
- `PurchaseDate` – Date of purchase
- `Region` – Customer location

Students Analysis Phase (Student Task) :-

Students should mentioned the following phases in details in the project .

- Define Problem Statement/ Business Requirement/Abstract
- Data Collection (load Data)
- Data Cleaning
- Compute Total Sales
- Top Product Analysis
- Sales by Category
- Regional Sales Analysis
- Monthly Sales Trend
- Data Exploration & Analysis (Means to visualize the data according to your understanding)
- Conclusions

Students Responsibility

- Making Group(Maximum 2 students or can do the single student if he wants to do this project own)
- If students wish to work on this project in a group, please send me the group name by next week via email. Both group members must have an equal understanding of the project.

Marks Distribution :-

The Project has total 10 Marks . 6 Marks for submission and 4 Marks for Project Viva.

In Submission you have to take care of format , deadline ,each member contribution of the project .

Project Viva – Every one should have over all knowledge about the project and understanding of R code.