**Big Data - Case Study:**

**Coffee sales Analysis**

**1. Introduction**

In today’s data-driven world, understanding customer preferences and sales trends is essential for business growth. This project focuses on Coffee Sales Analysis using MySQL, Hive, and Sqoop.  
The objective is to analyze coffee sales performance, identify top-performing products, and gain insights into consumer buying patterns.

MySQL serves as the foundation for structured data management and querying. Hive, a Hadoop-based tool, enables large-scale data analysis, while Sqoop bridges data transfer between MySQL and Hadoop for scalable computation.

Through this case study, we aim to demonstrate the power of integrated data technologies to perform efficient, insightful, and data-driven decision-making in the retail coffee business.

**2. Description of the Dataset**

**Dataset Name: Coffee\_Sales.csv**

This dataset records sales transactions for different coffee products across multiple dates and customer types.

**Attributes:**

1. **transaction\_id** – Unique identifier for each transaction.
2. **date** – Date of the coffee sale.
3. **time** – Time of the transaction.
4. **day\_of\_week** – Day name (e.g., Monday, Tuesday).
5. **customer\_type** – Indicates whether the customer is a member or non-member.
6. **coffee\_name** – Type of coffee sold (e.g., Latte, Espresso, Cappuccino).
7. **size** – Coffee size (Small, Medium, Large).
8. **quantity** – Number of cups sold.
9. **unit\_price** – Price per cup.
10. **money** – Total amount of the sale (quantity × unit\_price).
11. **payment\_mode** – Payment method used (e.g., Cash, Card, UPI).
12. **branch** – Store location or branch where the sale took place.

This structured dataset enables a detailed exploration of customer behavior, sales performance by coffee type, and revenue analysis across time periods.

**3. Project Scope**

The scope of this project is to conduct a comprehensive analysis of coffee sales data to uncover business insights.  
Using **MySQL**, **Hive**, and **Sqoop**, we will integrate and process the dataset to understand:

* Daily and weekly sales trends
* Top-selling coffee types and branches
* Customer purchase behavior (members vs. non-members)
* Revenue contribution by size and payment mode

By combining SQL and big data tools, the project aims to build a robust analytical framework to help coffee businesses improve marketing, stock management, and overall sales strategy.

**4. Goals**

1. **Sales Trend Analysis:** Examine daily, weekly, and monthly sales patterns.
2. **Top Product Identification:** Identify best-selling coffee products and their revenue contribution.
3. **Customer Insights:** Compare member vs. non-member buying behaviors.
4. **Branch Performance:** Evaluate revenue performance across different branches.
5. **Payment Method Analysis:** Understand customer preferences in payment modes.
6. **Integration:** Implement data transfer between MySQL and Hadoop using Sqoop.
7. **Scalable Analytics:** Use Hive for efficient querying on large-scale datasets.
8. **Visualization:** Create visual dashboards for better insight communication.

**5. Tools and Working Environment**

**1. MySQL**

* **Description:**MySQL is an open-source relational database management system used to store and manage structured data efficiently.
* **Working Environment:**In this project, MySQL was used to import the Coffee\_Sales.csv dataset, create tables, and perform SQL queries for preliminary data analysis — such as total sales, popular coffee types, and weekday vs weekend trends.

**2. Python (for Visualization)**

* **Description:**  
  Python, with libraries such as **Matplotlib** and **Seaborn**, was used to visualize key insights from the dataset.
* **Working Environment:**  
  Visualization scripts were executed in Jupyter Notebook to create bar charts and graphs for:
  + Top 3 best-selling coffees
  + Most profitable month
  + Peak sales hours
  + Day-wise and branch-wise revenue comparison

**Performing Analysis on MySQL**

**To load data –**

A computer screen with white text

AI-generated content may be incorrect.

**Then Verify-**

A screenshot of a computer

AI-generated content may be incorrect.

1️⃣ **Total Sales Revenue-**

A black screen with white text

AI-generated content may be incorrect.

2️⃣ **Most Popular Coffee-**

A screen shot of a computer

AI-generated content may be incorrect.

3️⃣ **Average Sale Value-**

**A black screen with white text

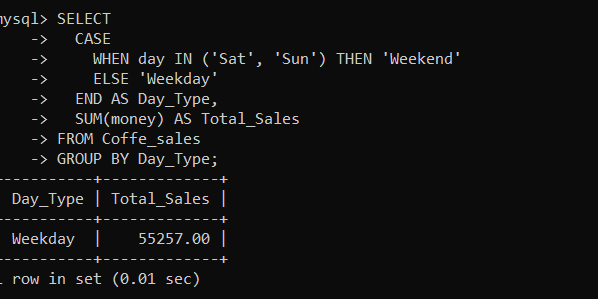
AI-generated content may be incorrect.**

**4️⃣ Top 3 Best-Selling Coffee Types-**

**A computer screen shot of a number

AI-generated content may be incorrect.**

**5️⃣ Weekend vs Weekday Sales-**

****

**6️⃣ Most Profitable Month-**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**7️⃣ Most Popular Cup Size-**

**A screenshot of a computer

AI-generated content may be incorrect.**

**8️⃣ Peak Hour of Sales-**

**A screen shot of a computer

AI-generated content may be incorrect.**

**9️⃣ Average Quantity Sold per Transaction-**

**A computer screen with white text

AI-generated content may be incorrect.**

**🔟 Day with Highest Total Sales-**

**A computer screen with white text

AI-generated content may be incorrect.**

**Data Visulization**

**Step 1: Import Libraries  
A screenshot of a computer

AI-generated content may be incorrect.**

**Total Sales Revenue**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Most Popular Coffee**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Top 3 Best-Selling Coffee Types**

**A computer screen shot of a computer code

AI-generated content may be incorrect.**

A graph of different colored bars

AI-generated content may be incorrect.

**Most Profitable Month**

**A screen shot of a computer

AI-generated content may be incorrect.**

A graph of a number of months

AI-generated content may be incorrect.

**Peak Hour of Sales**

**A screenshot of a computer code

AI-generated content may be incorrect.**

A graph with a line and a line

AI-generated content may be incorrect.

**Day with Highest Total Sales**

**A graph on a white sheet

AI-generated content may be incorrect.**