**Project Report: Unlocking Customer Insights – RFM-Based Segmentation Using Retail Transaction Data**

**Team Members and Roles**

* **Harsimranjit Singh** – *Team Lead*  
  Oversees project execution, assigns responsibilities and ensures timely delivery of milestones.
* **Anhadpreet Singh** – *Research & Data Engineer*  
  Responsible for data acquisition, cleaning, transformation, and implementation of RFM (Recency, Frequency, Monetary) logic.
* **Sandeep Kaur** – *DevOps / Git Manager*  
  Manages version control (Git), automates deployment pipelines and ensures continuous integration and collaboration.
* **Gurpreet Kaur** – *Documentation Manager*  
  Handles project documentation, meeting notes, final report writing, and formatting of insights for stakeholders.

**Project Overview**

**Title:** *Unlocking Customer Insights: RFM-Based Segmentation Using Retail Transaction Data*

**Objective:**  
To uncover hidden patterns in customer behavior through RFM (Recency, Frequency, Monetary) analysis, enabling businesses to:

* Enhance customer targeting strategies
* Improve retention and loyalty
* Drive personalized marketing campaigns

**Datasets Used**

1. **Online Retail Dataset – UCI Repository**  
   🔗 <https://archive.ics.uci.edu/dataset/352/online+retail>
2. **Online Retail II Dataset – Kaggle**  
   🔗 <https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci>

**Workflow Summary**

**1. Raw Data Acquisition**

* Imported both datasets for comparative exploration.
* Verified data integrity (e.g., duplicates, nulls, data type mismatches).

**2. Data Cleaning**

* Removed incomplete and canceled transactions.
* Normalized column formats (e.g., date-time parsing, numeric conversions).

**3. Data Validation**

* Validated each column based on:
  + Expected **data type** (e.g., InvoiceDate as datetime, UnitPrice as float)
  + Value ranges and logical consistency
* Ensured all transformations preserved data quality.

**4. Data Warehousing**

* Data ingested and stored into **Google BigQuery** using Google SDK.
* Set up separate environments:
  + **Database 1** for the first dataset
  + **Database 2** for the second dataset

Screenshot of the BigQuery schema setup was captured and added to project documentation as proof of successful ingestion.

**Tools & Technologies**

| **Area** | **Tools Used** |
| --- | --- |
| Programming | Python (Pandas) |
| Cloud | Google Cloud Platform (BigQuery, SDK) |
| Version Control | Git & GitHub |
| Data Cleaning | pandas, NumPy |
|  |  |