SnehaPriya Muppiri 16354863 sm743@umsystem.edu
Sreeja Aluguvelly 16354868 sa8hm@umkc.edu
Roopesh Kesava Reddy 16354885 rk4gb@umsystem.edu
Jaswanth Reddy Narala 16355312 jnmhc@umkc.edu

Project Proposal: Enhancing E-Commerce Insights with Big Data Analytics

Project Title: Advanced Analytics for E-Commerce Growth: A Case Study of Flipkart Sales Data

Dataset: The project will utilize a comprehensive dataset from Flipkart, consisting of 4.5GB of data detailing product sales and customer transactions. This dataset, available on Kaggle, encompasses more than 40 million records across two main data tables: Sales and Products.

Problem Statement: In the burgeoning field of e-commerce, businesses seek innovative strategies to understand consumer behavior, optimize logistics, and enhance profitability. This project aims to leverage big data analytics to uncover insights from Flipkart's sales data, focusing on customer purchase patterns, product preferences, and seasonal buying trends.

Tentative Solution: The proposed methodology involves a blend of big data tools for efficient data processing and analysis. Hadoop Distributed File System (HDFS) will serve as the foundational storage system, while Apache Spark will be employed for its robust data processing capabilities. Preliminary data cleaning and preprocessing will prepare the dataset for in-depth analysis, which includes identifying key trends, customer sentiments, and market dynamics. The use of Matplotlib and Seaborn for data visualization will aid in the interpretation and presentation of findings.

About dataset: We collected our dataset of 4.5GB of data from Kaggle. https://www.kaggle.com/datasets/iyumrahul/flipkartsalesdataset/?select=products.csv.We have two datasets in this project. Sales dataset and Products dataset.

Sales dataset contains the records of all the sales in the month of April in India. It consists of more than 40 million records, and it has 13 columns which include.

- Date
- Order_id,
- Product_id,
- City_name,
- Unit_selling
- Customer_id

and all the required sales columns.

The product dataset contains the data of all the products available for sale in Flipkart, it consists

of more than 32000 products, and it has 12 columns which include.

- Product_id
- Product_name,
- Product_type,
- Brand_name,
- Manufacturer_name

and all the required products columns.

Target Application: Insights derived from this analysis will empower e-commerce platforms to make data-driven decisions in areas such as inventory management, targeted marketing, and customer experience optimization. Ultimately, the project seeks to demonstrate how big data analytics can transform e-commerce strategies, leading to increased sales, customer satisfaction, and competitive advantage.