README for eCommerce Transactions Data Analysis

Overview

This project involves analyzing an eCommerce transactions dataset to derive business insights, build a loo

1. Task 1: Exploratory Data Analysis (EDA) and Business Insights
2. Task 2: Lookalike Model
3. Task 3: Customer Segmentation / Clustering

Task 1: Exploratory Data Analysis (EDA) and Business Insights

Description

Performed EDA on the dataset to identify trends, patterns, and anomalies. Derived actionable business ins

Steps

1. Loaded and merged the Customers.csv, Products.csv, and Transactions.csv files.
2. Examined data quality (e.g., null values, data types, duplicates).
3. Derived key metrics:
   - Unique customers and products.
   - Total sales by region.
   - Most popular products.
4. Visualized data using bar plots and count plots.

Deliverables

- Code: Akshay_GS_EDA.ipynb

- Report: Akshay_GS_EDA.pdf

Task 2: Lookalike Model

Description

Developed a lookalike model to recommend similar customers based on their profile and transaction histor

Steps

1. Used features such as total spending, transaction frequency, and product preferences.

2. Preprocessed data by aggregating transaction-level information to customer-level.

3. Built a similarity-based model using cosine similarity.

4. Generated recommendations for the first 20 customers.

Deliverables

- Code: Akshay_GS_Lookalike.ipynb

- Output: Akshay_GS_Lookalike.csv

Task 3: Customer Segmentation / Clustering

Description

Performed customer segmentation using clustering techniques to group customers based on similar behav

Steps

1. Prepared features combining profile information (Customers.csv) and transaction data (Transactions.csv

2. Scaled the data for clustering.

3. Used the DBSCAN algorithm to create clusters.

4. Evaluated cluster quality using Davies-Bouldin Index and visualized clusters.

Deliverables

- Code: Akshay_GS_Clustering.ipynb

- Report: Akshay_GS_Clustering.pdf

Instructions for Execution

1. Ensure all required libraries (e.g., pandas, numpy, seaborn, matplotlib, sklearn) are installed.

2. Place the dataset files (Customers.csv, Products.csv, Transactions.csv) in the working directory.

3. Execute the Jupyter notebooks in the following order:

   - Akshay_GS_EDA.ipynb

   - Akshay_GS_Lookalike.ipynb

   - Akshay_GS_Clustering.ipynb

4. Refer to the generated reports (PDFs) for summaries and insights.

Acknowledgments

This project demonstrates skills in data analysis, machine learning, and clustering to solve real-world busin