README: eCommerce Transactions Analysis Assignment

Overview

This project focuses on analyzing an eCommerce transactions dataset to uncover valuable insights, develop a customer lookalike model, and perform customer segmentation. The objective is to demonstrate proficiency in exploratory data analysis (EDA), predictive modeling, and clustering techniques, while delivering actionable business recommendations.

Dataset Details

The analysis uses the following three datasets:

1. **Customers.csv**

    o CustomerID: Unique identifier for customers.

    o CustomerName: Name of the customer.

    o Region: Continent where the customer resides.

    o SignupDate: Date of customer signup.

2. **Products.csv**

    o ProductID: Unique identifier for products.

    o ProductName: Name of the product.

    o Category: Product category.

    o Price: Product price (USD).

3. **Transactions.csv**

    o TransactionID: Unique identifier for transactions.

    o CustomerID: Associated customer.

    o ProductID: Associated product.

    o TransactionDate: Date of the transaction.

    o Quantity: Quantity purchased.

    o TotalValue: Transaction total value.

    o Price: Product price in the transaction.

Tasks and Deliverables

1. Exploratory Data Analysis (EDA)

- Merge and clean the datasets to create a unified view.

- Visualize key metrics, such as revenue trends, regional contributions, and product performance.

- Generate actionable insights to guide marketing, inventory management, and customer engagement strategies.

- Deliverables:

- Jupyter Notebook with EDA code and visualizations.

- PDF report summarizing business insights.

2. Lookalike Model

- Build a machine learning model to recommend three similar customers based on profiles and transaction histories.

- Use customer and product data to calculate similarity scores.

- Deliverables:

- Lookalike.csv: Contains similarity scores for the first 20 customers.

- Jupyter Notebook with model development and results.

3. Customer Segmentation

- Perform clustering using customer profiles and transaction data.

- Evaluate clusters using metrics such as the Davies-Bouldin Index (DB Index).

- Visualize clusters and provide recommendations for targeting each segment effectively.

- Deliverables:

- Jupyter Notebook with clustering analysis and visualizations.

- Report detailing the number of clusters, DB Index value, and actionable insights.

Key Tools and Technologies

- **Python Libraries:**

  - Pandas, NumPy for data processing and cleaning.

  - Matplotlib, Seaborn for data visualization.

  - Scikit-learn for machine learning and clustering.

- **Jupyter Notebook:**

  - Used for analysis, coding, and visualization.

Expected Outcomes

- In-depth understanding of revenue drivers, customer behavior, and sales trends.

- A functional lookalike model providing meaningful customer recommendations.

- Customer segments with actionable insights for targeted marketing and product strategies.

How to Run

- Clone this repository.

- Install required libraries with pip install -r requirements.txt.

- Open the Jupyter Notebooks (EDA.ipynb, Lookalike_Model.ipynb, Clustering.ipynb) to explore the analyses.

- Review the generated reports for business insights and recommendations.

Potential Applications

- Optimize marketing campaigns by focusing on high-value customers.

- Personalize product recommendations using the lookalike model.

- Improve operational efficiency by addressing seasonal trends and return rates.

- Expand into untapped regional markets and enhance sales of underperforming products.