PROJECT PROPOSAL

Arafat Okino Sadiq

Monday 22nd October 2024

1. PROJECT TITLE

Analyzing the Impact of Customer Demographics on Product Category Preferences

2. DESCRIPTION OF DATA SET

The dataset, sourced from Kaggle's **Retail Sales Customer Behavior Analysis**, contains **1 million observations** with **78 variables**, including customer demographics, behavior, and product information. Key variables include **age, gender, income**, and **product categories** (e.g., Electronics, Groceries). The dataset will be narrowed down to **1,000 observations** for faster analysis, to understand how demographics influence product category preferences.

3. DESCRIPTION OF THE PROBLEM

The primary goal of this project is to analyze how **customer demographics** impact **purchase behavior**, specifically focusing on **product category selection**. The project aims to:

- Identify significant **demographic factors** (e.g., age, gender, income) that influence product category choices.
- Build a **predictive model** that classifies customers into specific product categories based on their demographic data.
- Provide **actionable insights** to help **retailers** segment and target customers more effectively.

Key questions include identifying the most influential demographic factors, predicting product categories using machine learning, and uncovering trends or patterns across demographic groups.

4. ANALYSIS PLAN

The analysis will apply **regression models** and **clustering** to quantify relationships between demographic factors and product categories. Techniques like **Logistic Regression** and **Multinomial Logistic Regression** will be used for categorical outcomes, while **visualizations** such as heatmaps will uncover trends in customer behavior. **Exploratory data analysis (EDA)** with bar charts and heatmaps will highlight demographic patterns, and models like **Random Forest** will predict product categories based on demographics. Model performance will be assessed using metrics such as **accuracy** and **F1-score**, with key demographic predictors identified to provide actionable insights for retailers.

5. REFERENCES https://www.kaggle.com/datasets/utkalk/large-retail-data-set-for-eda/data