**Demographic and Behavioral Analysis for Bike Purchase Insights**

Arman Tamim

Student ID: 16  
Batch No: 54

Mobile Number: 01718183478

Department of English

An Assignment

Submitted to the EDGE

In Partial Fulfillment of the Requirements for the Course

Computer Fundamental and Office Applications   
(Microsoft Word, PowerPoint, Excel)

Submitted to

Md Mahbub E Noor

Lecturer

Assistant Professor

Department of CSE



UNIVERSITY OF BARISHAL, BARISHAL-8200

On 26 November 2024

Contents

[Project Objective: 3](#_Toc185071358)

[Scope of Work: 3](#_Toc185071359)

[Dataset Description: 3](#_Toc185071360)

[Methodology: 4](#_Toc185071361)

[1**.** Data Preprocessing and Cleaning: 4](#_Toc185071362)

[2. Descriptive Analytics: 4](#_Toc185071363)

[3. Exploratory Data Analysis (EDA): 4](#_Toc185071364)

[4. Predictive Modeling: 4](#_Toc185071365)

[5. Visualization and Reporting: 4](#_Toc185071366)

[6. Insights and Recommendations: 4](#_Toc185071367)

[Deliverables: 4](#_Toc185071368)

[Detailed Analytical Report: 4](#_Toc185071369)

[Predictive Model: 5](#_Toc185071370)

[Visual Dashboards: 5](#_Toc185071371)

[Executive Presentation: 5](#_Toc185071372)

[Project Timeline: 5](#_Toc185071373)

[Tools and Technologies: 5](#_Toc185071374)

[Expected Outcomes: 5](#_Toc185071375)

# Project Objective:

This project aims to analyze a dataset comprising demographic, socioeconomic, and behavioral characteristics to uncover key factors driving bike purchases. Insights from this analysis will be utilized to develop predictive models, enabling informed decision-making for targeted marketing and strategic customer segmentation.

# Scope of Work:

The project will involve a comprehensive analysis of the dataset to achieve the following:

1. Identify patterns and trends in bike purchases across different demographic groups.
2. Determine correlations between socioeconomic factors and purchasing decisions.
3. Develop predictive models to classify and forecast the likelihood of bike purchases.
4. Provide actionable recommendations for customer-focused marketing strategies.

# Dataset Description:

The dataset consists of 30 records and 13 attributes, as outlined below:

| **Variable** | **Description** |
| --- | --- |
| **ID** | Unique identifier for each individual. |
| **Marital Status** | Marital status of the individual (Married M, Single S). |
| **Gender** | Gender of the individual (M for Male, F for Female). |
| **Income** | Annual income in USD. |
| **Children** | Number of children in the household. |
| **Education** | Education level (e.g., High School, Partial College, Graduate Degree). |
| **Occupation** | Employment category (e.g., Professional, Clerical, Skilled Manual, Manual). |
| **Home Owner** | Indicates home ownership status (Yes, No). |
| **Cars** | Number of cars owned by the individual. |
| **Commute Distance** | Typical daily commute distance (e.g., 0-1 Miles, 10+ Miles). |
| **Region** | Geographic region (e.g., Europe, Pacific). |
| **Age** | Age of the individual in years. |
| **Purchased Bike** | Indicates whether the individual purchased a bike (Yes, No). |

# Methodology:

## 1**.** Data Preprocessing and Cleaning:

* **Data Integrity Check**: Ensure completeness and consistency of the data.
* **Outlier Detection**: Identify and address anomalies in variables like income and age.
* **Feature Engineering**: Transform categorical variables (e.g., Gender, Marital Status) into numerical encodings for analysis.

## 2. Descriptive Analytics:

* Generate summary statistics for income, age, education levels, and commute distances.
* Explore distributions of categorical variables such as gender, marital status, and occupation.
* Visualize the data through histograms, bar charts, and box plots.

## 3. Exploratory Data Analysis (EDA):

* **Correlation Analysis**: Examine relationships between variables (e.g., income vs. bike purchase, commute distance vs. home ownership).
* **Segment Analysis**: Investigate bike purchases across demographic groups (e.g., age brackets, education levels, and regions).
* **Behavioral Trends**: Analyze the impact of commute distance, occupation, and income on purchasing decisions.

## 4. Predictive Modeling:

* **Classification Models**: Use machine learning algorithms (e.g., Logistic Regression, Random Forest, Decision Trees) to predict bike purchase likelihood.
* **Feature Importance**: Identify the most influential factors driving bike purchases.
* **Model Validation**: Evaluate model performance using metrics such as accuracy, precision, recall, and F1 score.

## 5. Visualization and Reporting:

* Develop dashboards and visual reports for presenting findings.
* Highlight key patterns using heatmaps, scatter plots, and correlation matrices.

## 6. Insights and Recommendations:

* Summarize key takeaways and trends from the analysis.
* Provide tailored recommendations for marketing campaigns, including target demographics, pricing strategies, and regional focus.

# Deliverables:

## Detailed Analytical Report:

* + Comprehensive summary of the data.
  + Key patterns and insights derived from EDA.
  + Recommendations for marketing and sales optimization.

## Predictive Model:

* + Machine learning models with clear documentation of methodology and results.
  + Codebase (in Python or R) for future use.

## Visual Dashboards:

* + Interactive visualizations highlighting trends and insights.

## Executive Presentation:

* + Slide deck summarizing the project objectives, methodology, findings, and recommendations.

# Project Timeline:

| **Phase** | **Duration** | **Activities** |
| --- | --- | --- |
| **Phase 1: Data Preparation** | 1 Week | Data cleaning, preprocessing, and feature engineering. |
| **Phase 2: EDA** | 2 Weeks | Descriptive analytics, correlation analysis, and visualizations. |
| **Phase 3: Modeling** | 2 Weeks | Build and validate predictive models. |
| **Phase 4: Reporting** | 1 Week | Develop final reports, dashboards, and presentations. |

# Tools and Technologies:

* **Programming Languages**: Python (pandas, matplotlib, seaborn, scikit-learn) or R.
* **Visualization Tools**: Tableau, Power BI, or Matplotlib for dashboards.
* **Data Processing**: Microsoft Excel, SQL for initial exploration.

# Expected Outcomes:

1. Identification of key demographic and behavioral factors influencing bike purchases.
2. A robust predictive model for classifying individuals based on bike purchase likelihood.
3. Actionable insights to enhance customer engagement and improve marketing ROI.

Would you like further assistance in implementing this project?