**Project Proposal (DS5003 Stat. & Math Methods for DS** **MS(DS)-1A)**

|  |  |
| --- | --- |
| Submission, title and duration | |
| Date of submission | 25-10-2024 |
| Project title | **ANALYZING & PREDICTING Sales Based on Customer Demographics USING DATA SCIENCE TECHNIQUES** |
| Start date | 25-10-2024 |
| End date | 1-12-2024 |
| Submitted By | MUZAFFAR SHABIR KHAN ROLL NO: F24-8003 |
| Lead Academic/Research Institution and Lead Researcher\* | |
| Name of research institution/university | Fast National University |
| Name of Lead Advisor | Dr. Sana Shahid |

PROJECT TITEL :

ANALYZING & PREDICTING Sales Based on Customer Demographics USING DATA SCIENCE TECHNIQUES

INTRODUCTION : **About Dataset:- Dataset History**

A retail company “ABC Private Limited” wants to understand the customer purchase behavior (specifically, purchase amount) against various products of different categories. They have shared purchase summaries of various customers for selected high-volume products from last month.  
The data set also contains customer demographics (age, gender, marital status, city type, product details (productid and product category), and Total purchase amount from last month.

Now, they want to build a model to predict the purchase amount of customers against various products which will help them to create a personalized offer for customers against different products.

**Dependent Variable :** Product ID

**Independent Variables:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | Occupation | City\_Category | Product\_Category\_1 | Purchase |

**No of Observations:** 400

Tasks to perform

**The purchase col column is the Target Variable, perform Univariate Analysis and Bivariate Analysis w.r.t the Purchase.**

**Masked in the column description means already converted from categorical value to numerical column.**

**Below mentioned points are just given to get you started with the dataset, not mandatory to follow the same sequence.**

# Aim and Objectives

DATA PREPROCESSING

* **Check basic statistics of the dataset**
* **Check for missing values in the data**
* **Check for unique values in data**

Perform EDA

* **Purchase Distribution**
* **Check for outliers**
* **Analysis by Gender, Marital Status, occupation, occupation vs purchase, purchase by city, purchase by age group, etc**
* **Drop unnecessary fields**
* **Convert categorical data into integer using map function (e.g 'Gender' column)**
* **Missing value treatment**
* **Rename columns**
* **Fill nan values**
* **Map range variables into integers (e.g 'Age' column)**

Data Visualization

* **visualize an individual column**
* **Age vs Purchased**
* **Occupation vs Purchased**
* **Productcategory1 vs Purchased**
* **City category pie chart**
* **Check for more possible plot**

SOURCE OF DATA:

 **Dataset**: **ANALYZING & PREDICTING Sales Based on Customer Demographics USING DATA SCIENCE TECHNIQUES**

<https://www.kaggle.com/datasets/rajeshrampure/black-friday-sale>