

# BIG MART

## SYNOPSIS

E-commerce or electronic commerce is a process of Buying, selling, transferring, or exchanging products, services, and/or information via electronic networks and computers. Commerce is a fundamental to the success of any business. To streamline trading operations and maintain profits, the industry must focus on commerce, which deals with a lot more than just buying and selling. As the world is digitalizing, ecommerce solutions are increasingly becoming common. The advent of machine learning and artificial intelligence has further enhanced the effectiveness of ecommerce. The examples of ecommerce include eBay, Amazon, Upwork, Olx, etc.

The advantages of ecommerce shopping are Lower price for products compared to traditional shopping, Time saving, we can avoid crowds in shopping, Available wide range and variety of products, shipping/delivery options and also it provides Feedback option from customers.

The ecommerce provides two options for selling products they are **Ecommerce Website** and **Ecommerce Mobile Application**. Some customers choose to buy goods from the website and others from the mobile app. The company needs to make sure which one gets the most revenue/annual income for future benefits. The annual income of a company is the total money it earns from its business operation during one financial year.

The company is trying to decide whether to focus their efforts on their mobile app experience or their website.

I worked with the Ecommerce Customers csv file from the company. It has Customer information.

Then it also has numerical value columns:

- Avg. Session Length: Average session of in-store style advice sessions.
- Time on App: Average time spent on App in minutes
- Time on Website: Average time spent on Website in minutes
- Length of Membership: How many years the customer has been a member.
- Yearly Amount Spent: The total amount the customer is spending.

- 📊 Business problem: Interpret which variables are contributing towards the more annual income prediction

## **Contents of the PROJECT:**

### 1. Business Understanding

- Description
- Objective

### 2. Data Understanding

- Import Libraries
- Load Data
- Statistical summaries and visualisations

### 3. Data Preparation

- Missing values imputation

### 4. Exploratory Data Analysis

- Features Exploration

### 5. Data Splitting

- Train test split

### 6. Modeling

- Model1: Modeling with the linear model and tree based model technique with Standard Scalar.
- Model2: Modeling with the linear model and tree based model technique by adjusting outliers & Standard Scalar.
- Model3: Modeling with the linear model and tree based model technique by adjusting outliers & Log Transformation.

### 7. Finalizing Model Choice

- Model performance
- Feature importance

## 8. Deployment

## 9. Conclusion & Recommendation

Firstly I use Exploratory Data Analysis (EDA), it is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

The next step is Data Preparation, which includes

- Data cleaning:- cleaning of data by checking for null values/missing values (if we have null values we need to impute mean, median or mode values)
- checking for duplicates and conversion of data types
- Feature engineering: - elimination of unwanted features , feature addition(if needed) and feature transformation
- Data transformation: - I need to transform the data to get good model (with low errors)

Then I test the data by using various Testing strategies, they are Normality Testing, Multicollinearity testing and also checking the outliers and Relationship between features and target variables. Finally I realize which variables are contributing towards the more annual income prediction. And split the data set for building models from it.

Then I build 4 models (Linear Regression,, Lasso, ElasticnetCV, Ridge Regression ) from it, and find out the errors ( MAE - Mean Absolute Error, RMSE - Root Mean Square Error ) from these models And also again build these same model and find out errors after the aggregation log transformation are applied to these models for getting the accurate result.

Then I finally get the best model out of these models with low MAE & RMSE on normal models and other transformation models. And build a page for annual income prediction for a new record (Page↗Python raw file, working through Sreamlit ) So that I can predict the annual income of a new record by giving the values to the variables of the dataset.