Project report on

**Customer Behaviour Prediction using Logistic Regression.**

horizontal line



## Introduction

A dataset given contains the information of 400 customers of a particular product in the market. Each customer is associated with 5 features in the dataset. Namely, UserId, Gender, Age, Salary and the last feature indicates if the customer went on to make the purchase or not.

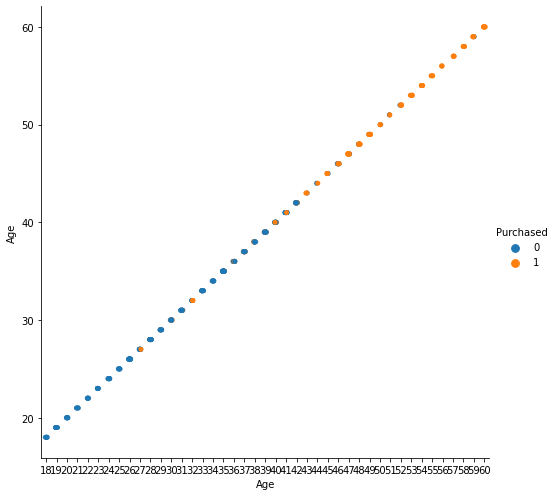
## Problem Statement

Given the dataset containing the information of a customer’s gender, age and salary. Predict whether he/she would make the purchase of a particular product in the market or not.

## 

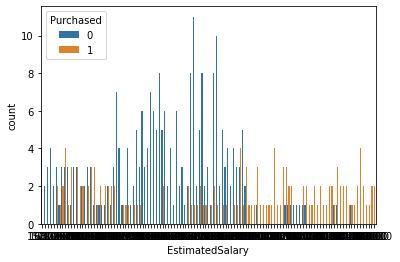
## Data Visualization

1. **A** **Categorical Plot** to find out what the majority Age group is that made the purchase

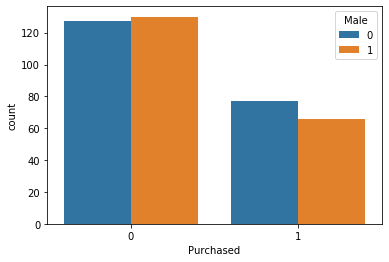


From the graph above, we can reason that most of the people who made the purchase were older in age (above 40 years).

1. **A Count plot** for Salary to know what the majority salary group was that made the purchase



1. **A Count plot** to find out whether men or women made most of the purchase



From the above graph, we can arrive at two conclusions. First one being that there were significantly more women customers considered than men customers in the dataset. And the second one being that a higher percentage of women made the purchase than men.

## Why Logistic Regression ?

We can clearly reason that predicting whether a customer purchased a particular product or not, is a **classification** problem, where the dependent variable ‘purchased’ is **binary** in nature. It’s value is either 0, which indicates that the customer did not make the purchase, or 1, which indicates that the customer made the purchase.

In such scenarios, where we have 1 or more independent variables and one dependent categorical variable, solving the problem using **Logistic Regression** is a feasible choice.

## 

## 

## Accuracy and Error Metrics

|  |  |
| --- | --- |
| Accuracy ( R2 score) in percentage | 91.25 |
| Mean squared error | 0.0875 |
| Root mean squared error | 0.2958 |

## Confusion Matrix

|  |  |  |
| --- | --- | --- |
|  | PREDICTED NO | PREDICTED YES |
| ACTUAL NO | 56 | 2 |
| ACTUAL YES | 5 | 17 |

## Classification Report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 score | Support |
| 0 | 0.92 | 0.97 | 0.94 | 58 |
| 1 | 0.89 | 0.77 | 0.83 | 22 |
| accuracy |  |  | 0.91 | 80 |
| Macro avg | 0.91 | 0.87 | 0.89 | 80 |
| Weighted avg | 0.91 | 0.91 | 0.91 | 80 |

***Note****: All the above results are for a test dataset which is of size 20% of the original dataset*

*Link to the Project Notebook:*

[*https://github.com/BharathMKulkarni/Customer-Behaviour-Prediction-Using-Logistic-Regression.git*](https://github.com/BharathMKulkarni/Customer-Behaviour-Prediction-Using-Logistic-Regression.git)