أكاديمية سدايا
SDAIA Academy

2011–2016

BIKES SALES

Data set

Table of Contents

introduction

This data contains bike sales in Europe and America and the relationship between leading profits with age groups of people and according to males and females and needs that can be purchased when buying a bike and we can predict future prices using machine learning once the product is developed so the machine will make the decision in setting the price.

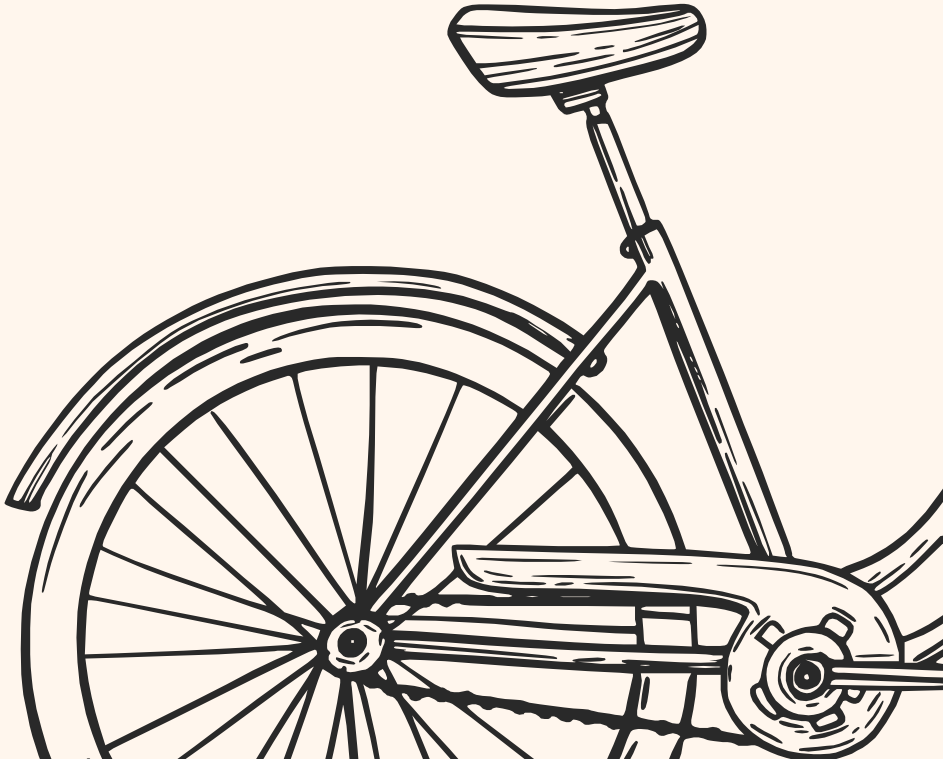
Data Source

we used the Bike Sales dataset. This data has been published on "kaggle.com". The data set contains 113036 rows X 18 columns a number of duplicate data and has no missing values. The numbers for all valid data can be found on the Kaggle source page.

https://www.kaggle.com/sadiqshah/bike-sales-in-europe

dataset description

This data is characterized by a lot of diversity in bicycles and their accessories and clothes. We also noticed the large demand for them of all ages, their purchase and use on a daily basis, which led to an increase in their revenues in the countries selected in this data

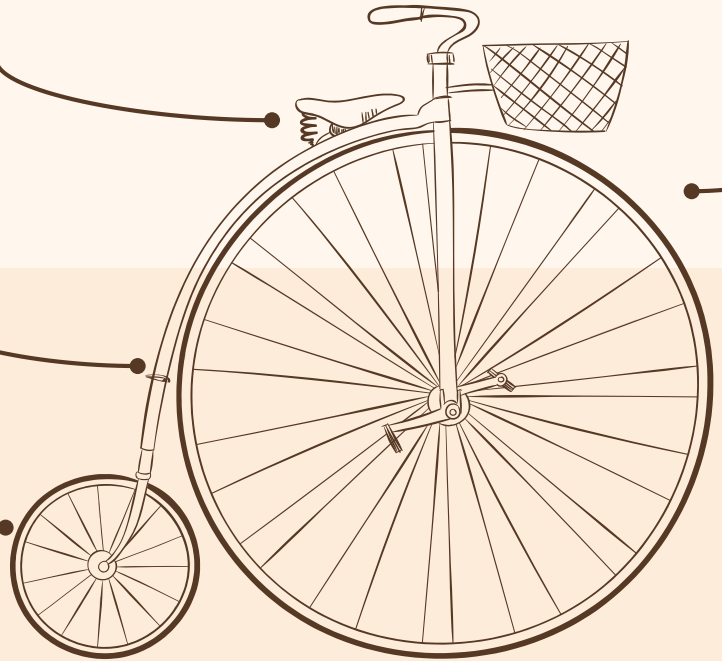


preprocessing

Import and read data

Remove duplicate

Delete some coloumns Because it represents a strong relationship

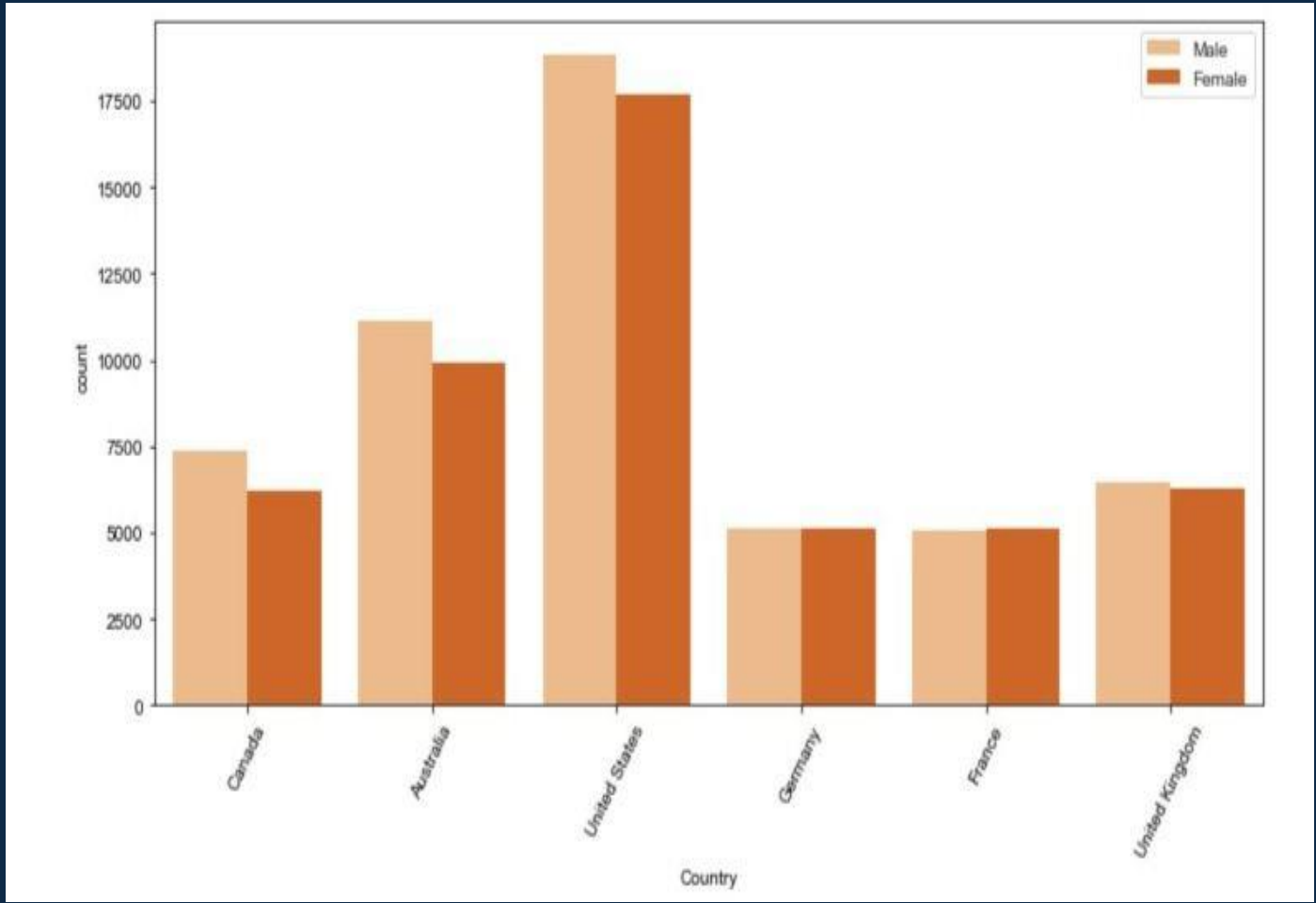


Remove outliers

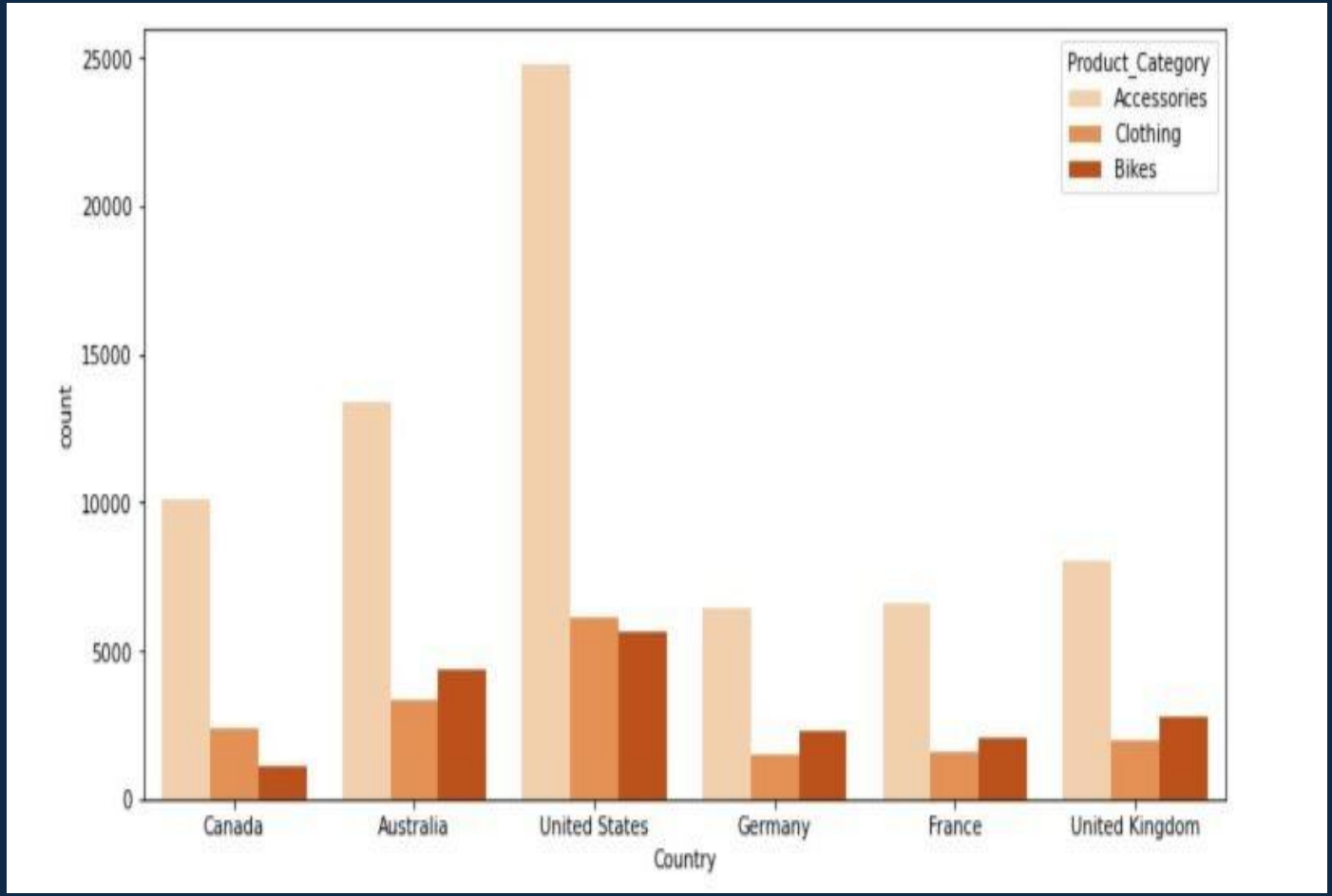
EDA for data set


Visualization
Results

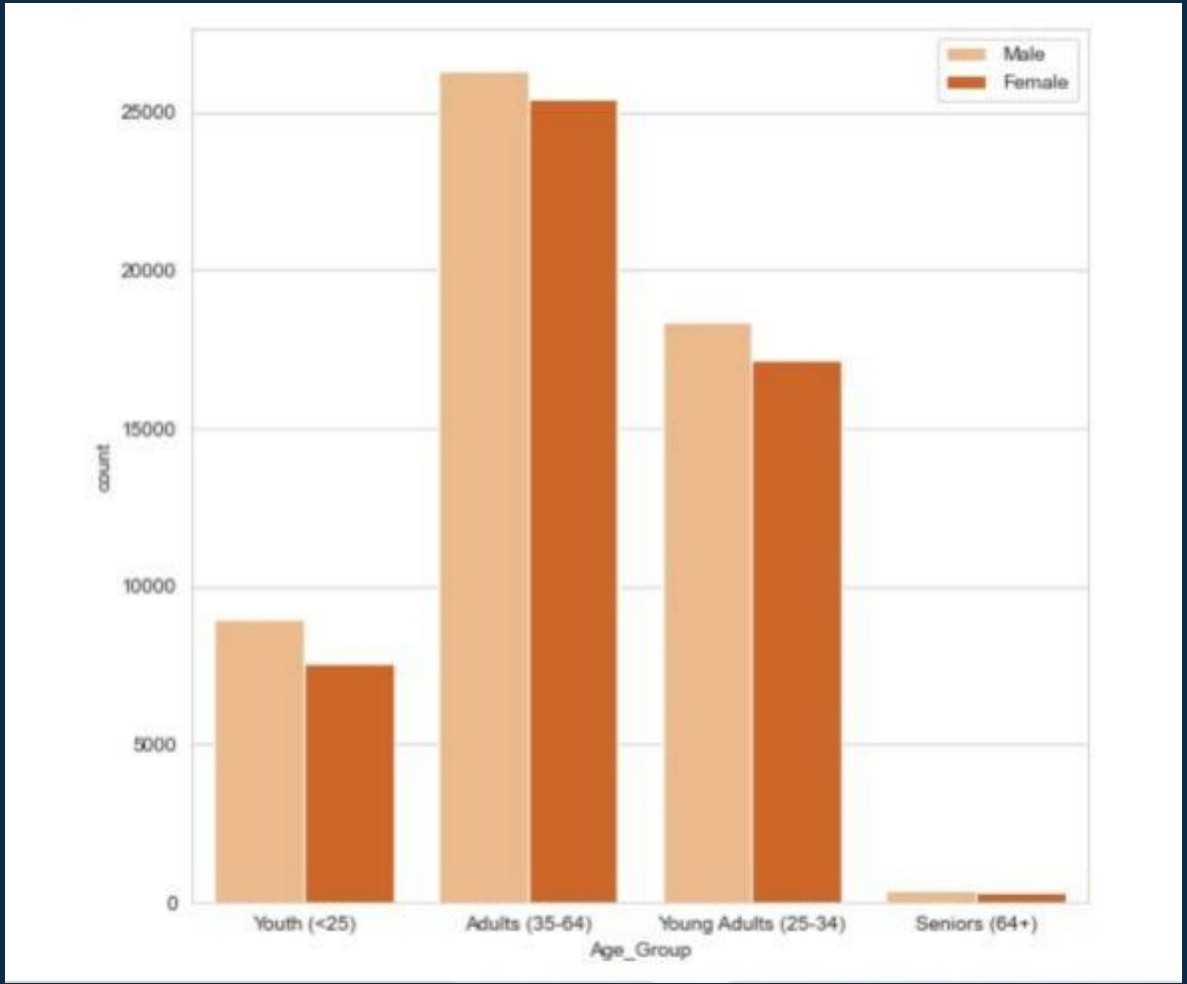
Most buy bikes and accessories for both gender .



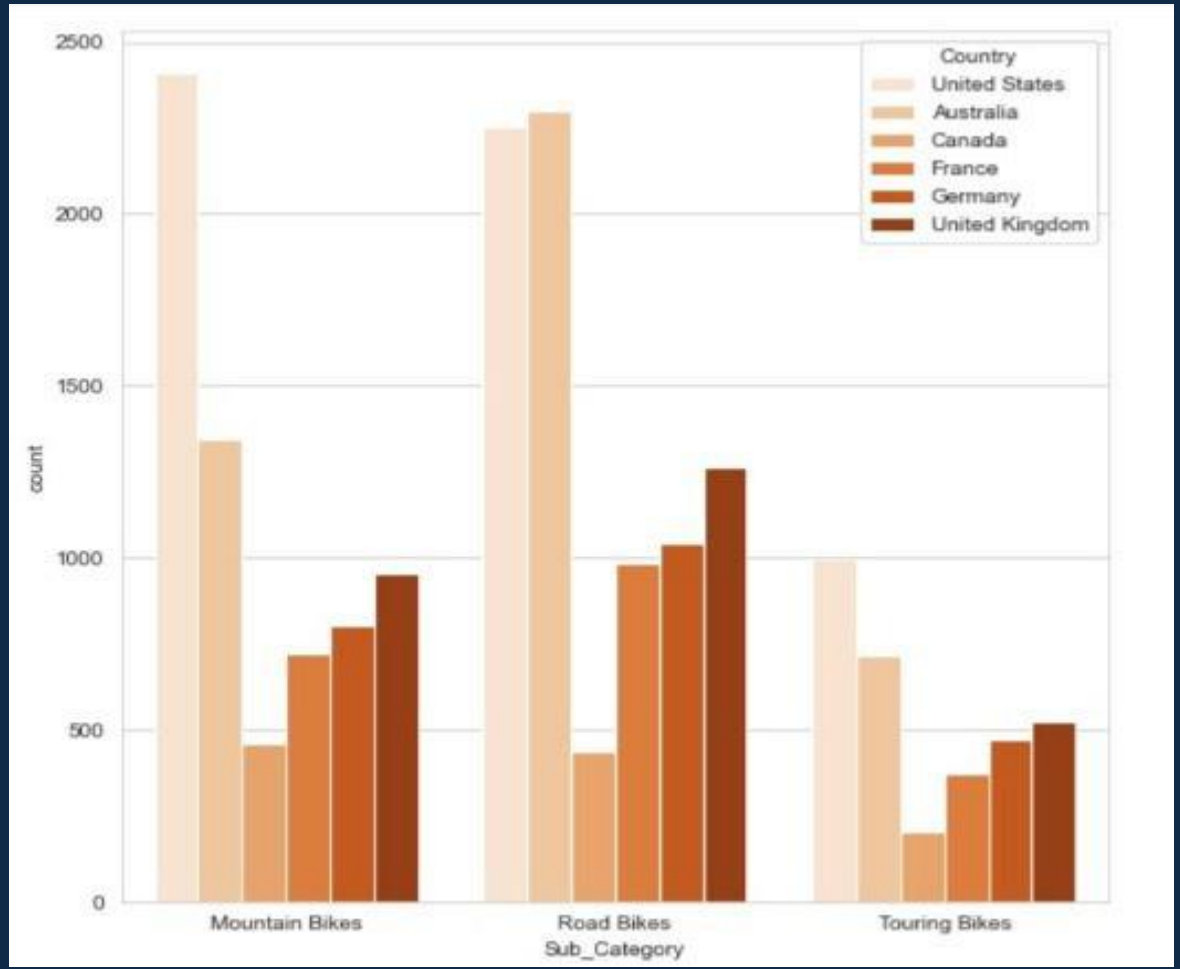
We noticed that accessories are sold more than both gender



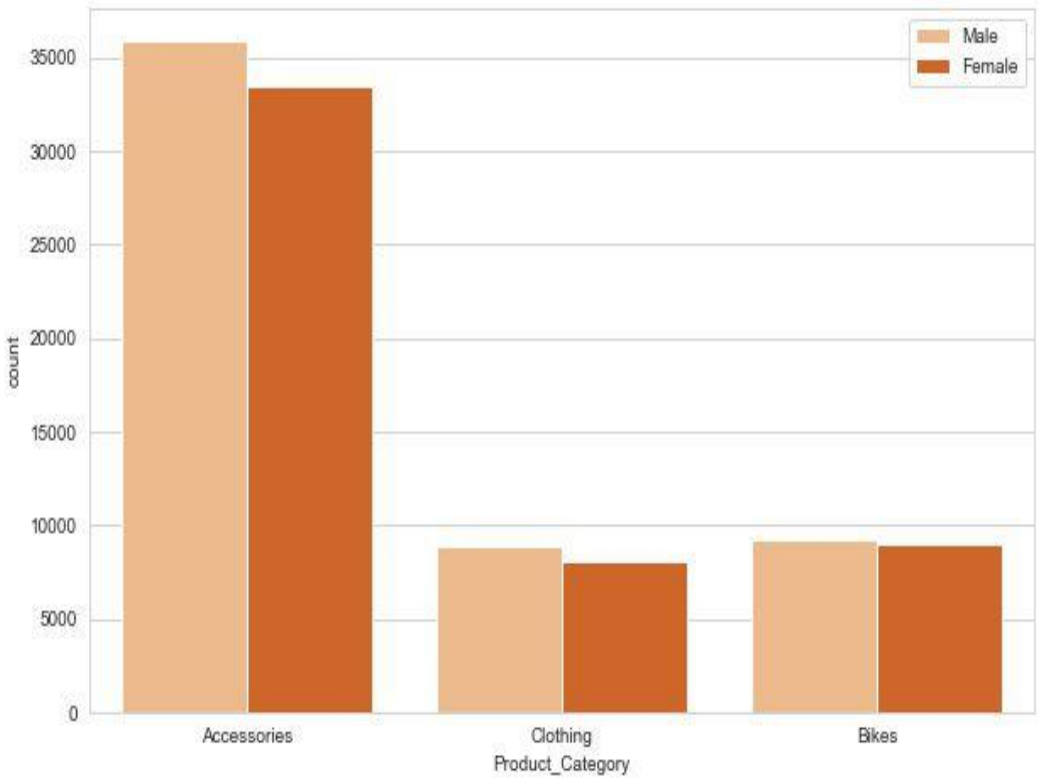
The category for his age most buys bikes, accessories and clothes



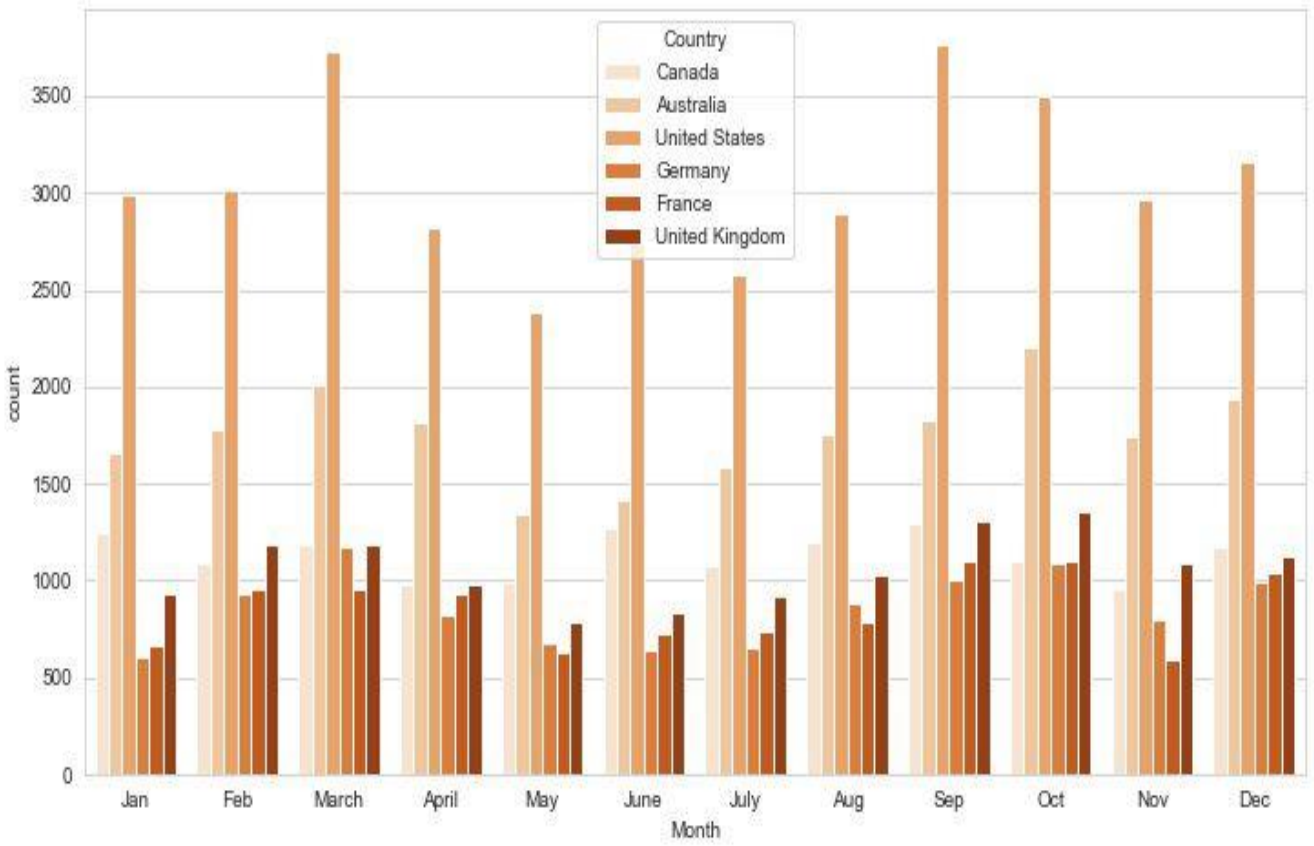
The rate of use of each type of bicycle in each country



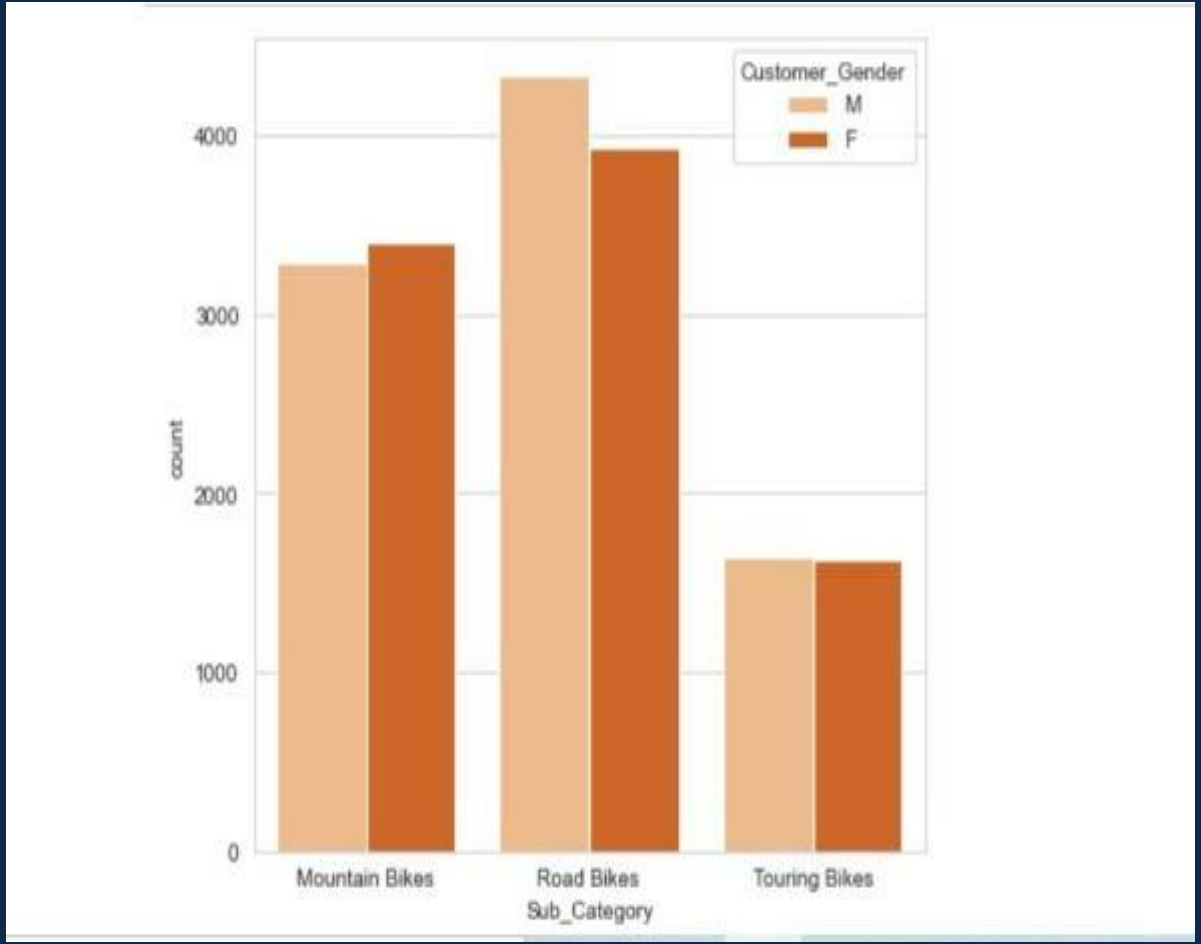
The rate of purchase of each product by males and females



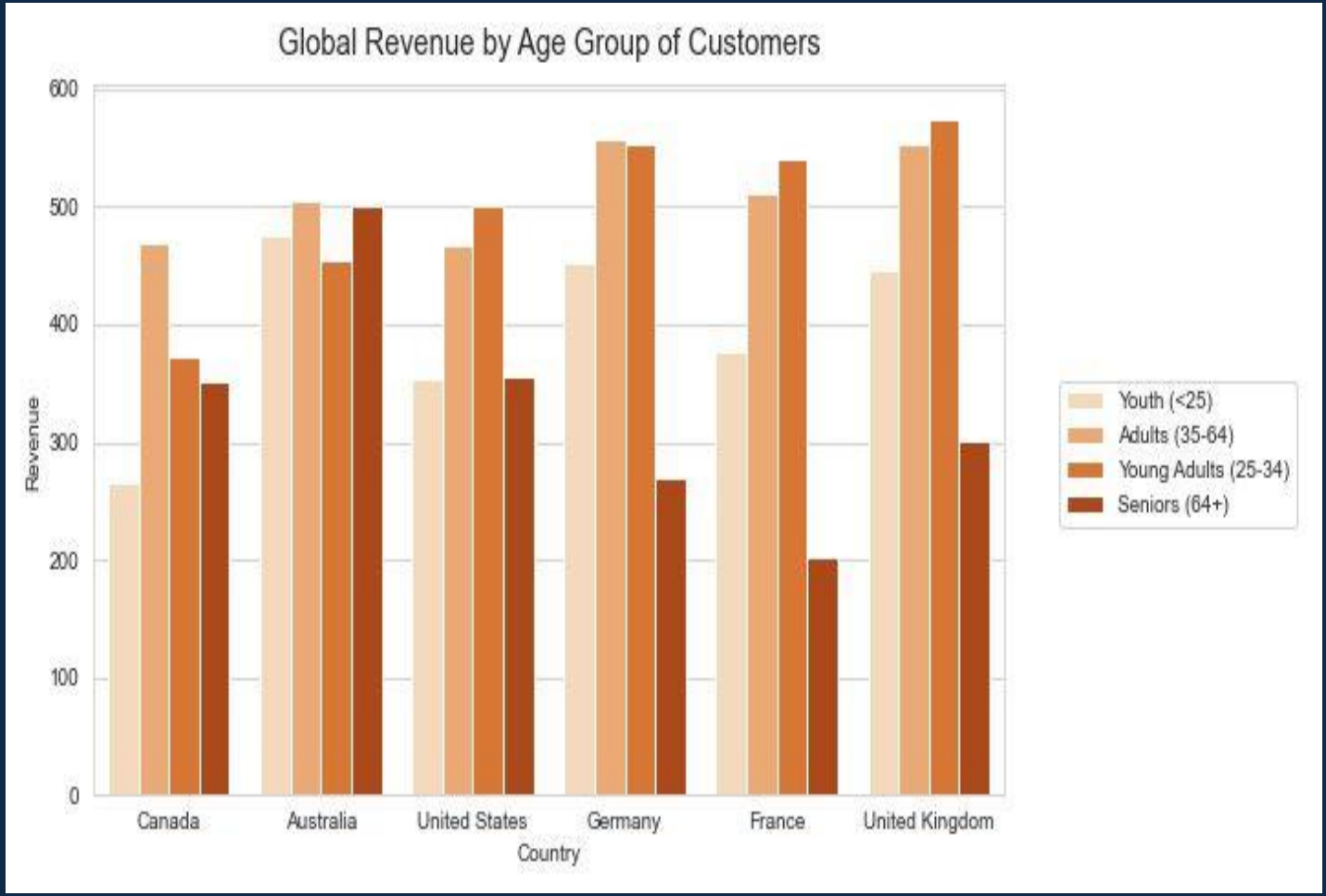
sales per month each country



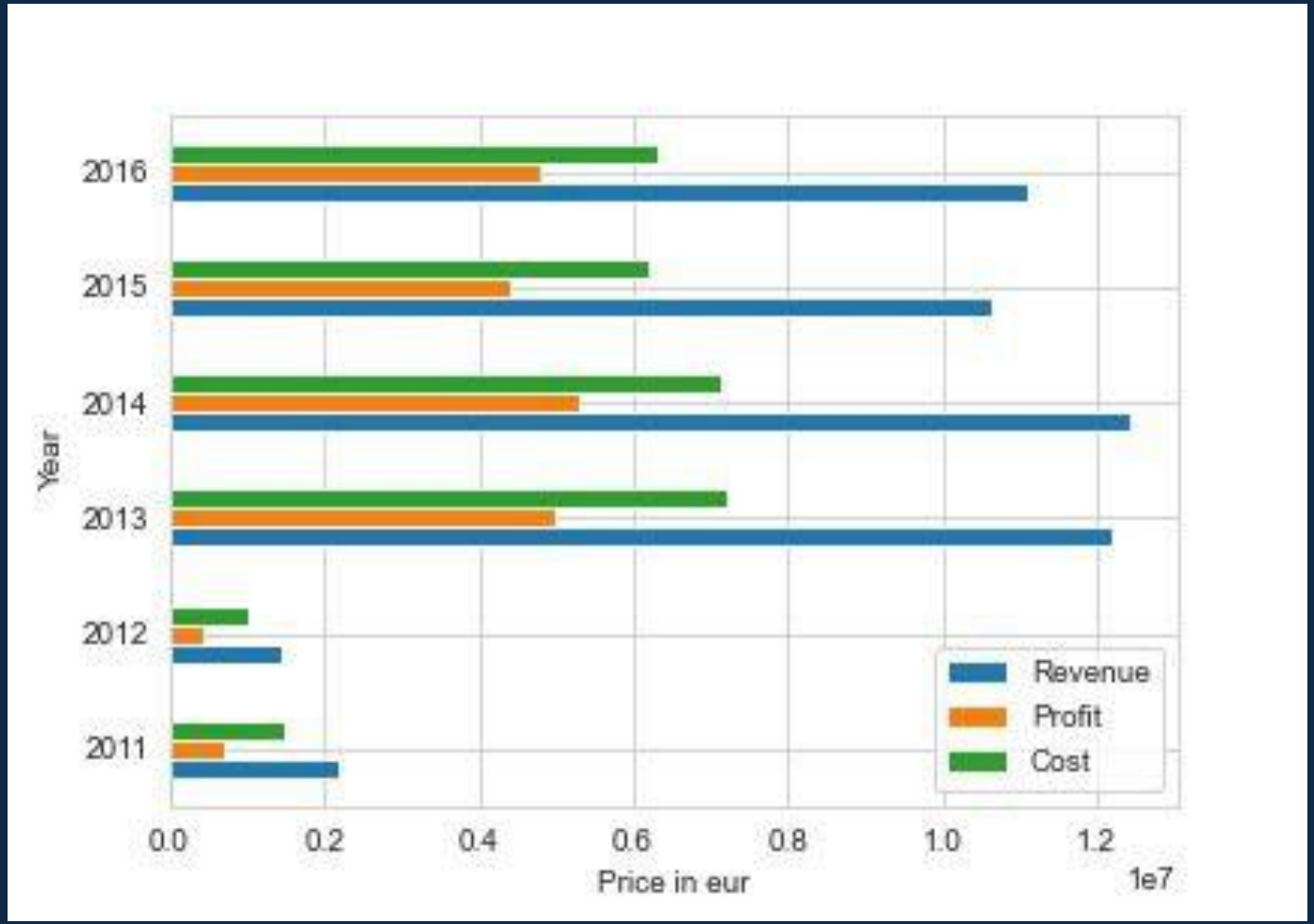
We made a filter on the types of bikes, and it is the most used for each type of male and female



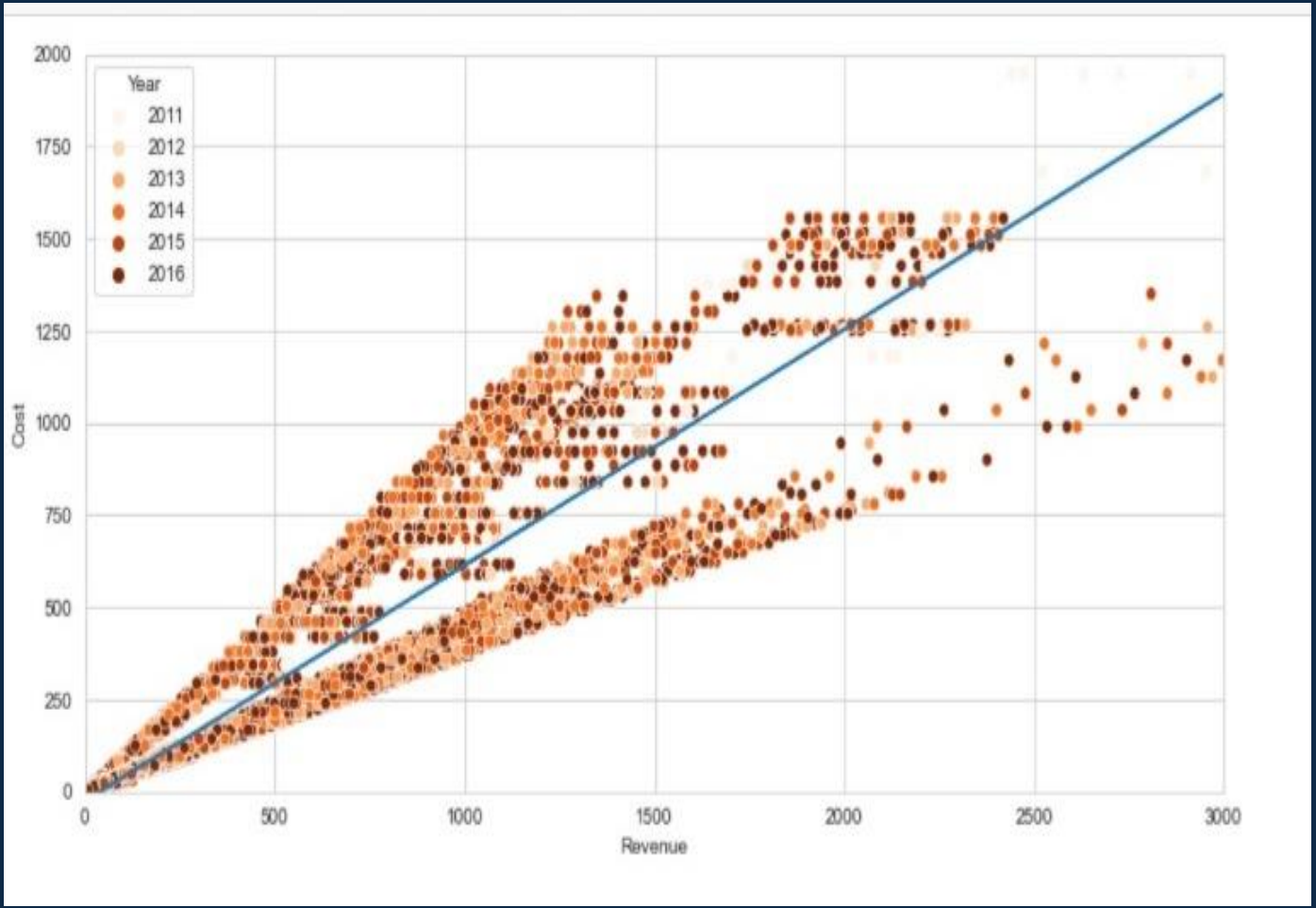
We note that the ages of 25 and over are the most buying bikes



The relationship of revenue increases with cost (a positive relationship)



The relationship of revenue increases with cost (a positive relationship)

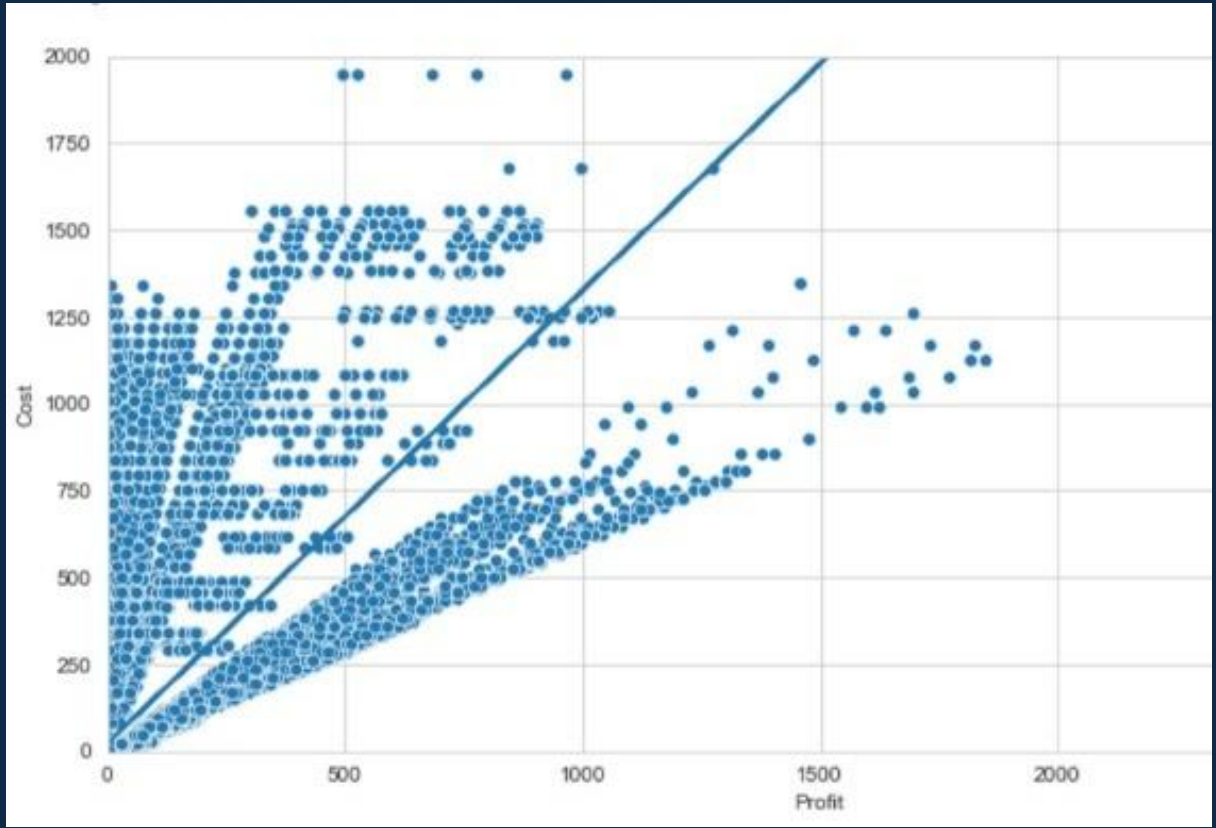


Comparison of the model and the best result

```
Linear Regression train R^2: 0.648
Ridge Regression train R^2: 0.648
Degree 2 polynomial regression train R^2: 0.814
Degree 4 polynomial regression train R^2: 0.839
```

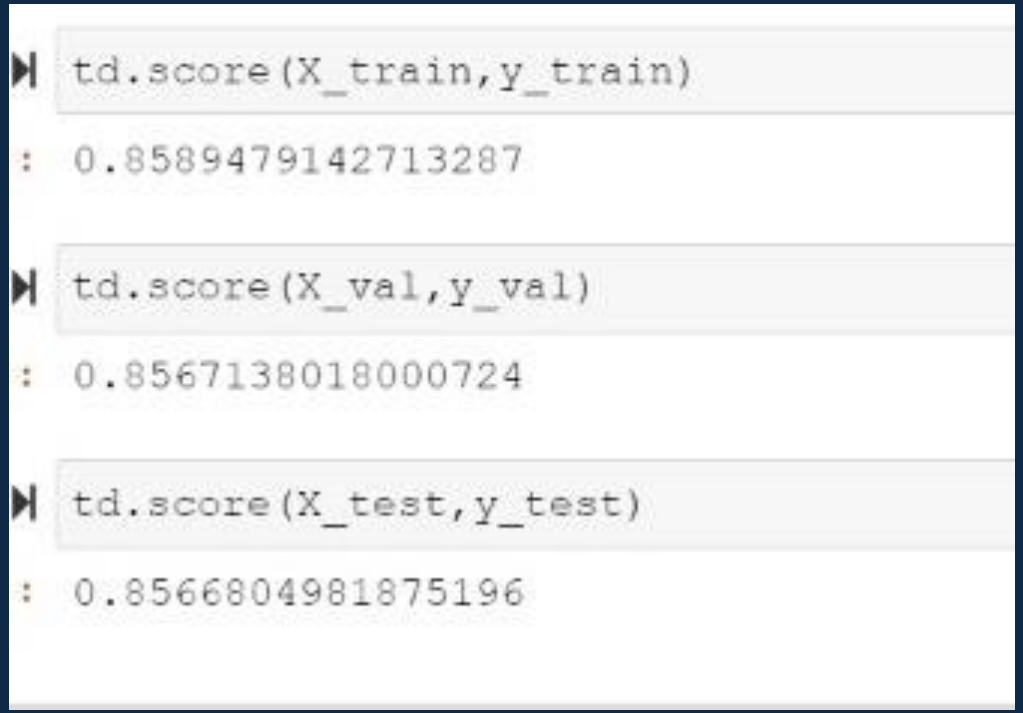
```
Linear Regression val R^2: 0.643
Ridge Regression val R^2: 0.643
Degree 2 polynomial regression val R^2: 0.818
Degree 4 polynomial regression val R^2: 0.840
```

```
Linear Regression test R^2: 0.645
Ridge Regression test R^2: 0.645
Degree 2 polynomial regression test R^2: 0.819
Degree 4 polynomial regression test R^2: 0.842
```


2000
1750
1500
1250
1000
750
500
250
0
Cost
0
500
1000
1500
2000
Profit

Decision Tree Regressor

We searched for a model that showed us the best ratio, and we tried a tree regressor and read and achieved a higher rate



conclusion

- Reasons for the picking this data .
- The types of bicycles are due to the nature of each country, and therefore the percentage of purchase of each type varies from each country
- First test results were promising .
- Supervised .

Thanks for listeneng

Alaa alghamdi
Yara aldossari
Shiekah saud