Logo

Description automatically generated with medium confidence

SDAIA T5 bootcamp

Linear Regression Project

Bike Model Regression

Final Report

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Basma Ghazi Abdullah Alduaiji

Mai Abdullah Fahad Aljuaid

14/10/2021

Contents

[Table of tables 2](#_Toc85081030)

[Introduction 3](#_Toc85081031)

[Design 3](#_Toc85081032)

[Algorithm 3](#_Toc85081033)

[Used data set 3](#_Toc85081034)

[Used tools 5](#_Toc85081035)

[Conclusion 5](#_Toc85081036)

# Table of tables

[Table 1 Scraped Weather Dataset 4](#_Toc85081156)

[Table 2 Predefined Dataset 4](#_Toc85081157)

# Introduction

Washington DC Weather data will be used in this project to build a linear regression model that would predict the revenue of bikes rented from bikeshare stations based on weather conditions. This model will provide insights about their sales performance.

# Design

The data set in this project is taken from the Wunder Weather Online website and predefined data frame. This data set presents the complete status of the weather in Sydney and the daily number of bikes along with revenues. However, data analysis was used to predict.

# Algorithm

The steps to analyse the data set are gathering data from the website within the needed URL. After that, Request the URLs to return the HTML code and save the collective data from the website within a CSV file. In addition, exploring data by using all the functions like info and describe. Then cleaning data by removing null values and duplicates as well as, removing any irrelevant data. After that, plot the graphs using seaborn and matplotlib modules from python. Finally, find the correlation between features and fit the linear regression model in order to find the best model that fits our data and also generalize on unseen data.

# Used data set

This project will use a scrapped data set from the Wunder Weather Online website and predefined data as a data frame. Moreover, the data set contains multiple fields with different data types as shown in table 1.

Table Scraped Weather Dataset

|  |  |
| --- | --- |
| Date | Represents the date when data acquiring. |
| Temperature | The measure of the warmth or coldness when data acquiring. |
| Humidity | the concentration of water vapour present in the air when data acquiring. |
| Wind Speed | Represents the atmospheric quantity caused by air moving from high to low pressure when data acquiring. |
| Pressure | Represents the pressure when data acquiring. |
| Precipitation | Representing the precipitation |

Table Predefined Dataset

|  |
| --- |
| Date |
| Season |
| Weather Condition |
| year |
| month |
| weekday |
| Casual riders |
| Registered riders |
| Revenue |

However, The Individual sample from the Weather data set are two years: 2011 and 2012. Moreover, the expected characteristics used to predict the revenue are the attributes from Weather data such as Temperature, Humidity and Wind Speed. Furthermore, the predicted target is to predict the total revenue per day according to a specific weather status

.

# Used tools

The company will utilize the tools provided as a module in pythons such as Pandas, Matplotlib, Seaborn NumPy, Beautifulsoup, and Excel. The python libraries support multiple data analysis and cleaning methods to ensure the data is clean and ready to visualize.

# Conclusion

Since the revenue increase on good weather days with hotter temperature, the company must increase their bike availability and promotions during the summer months to further increase their revenue.