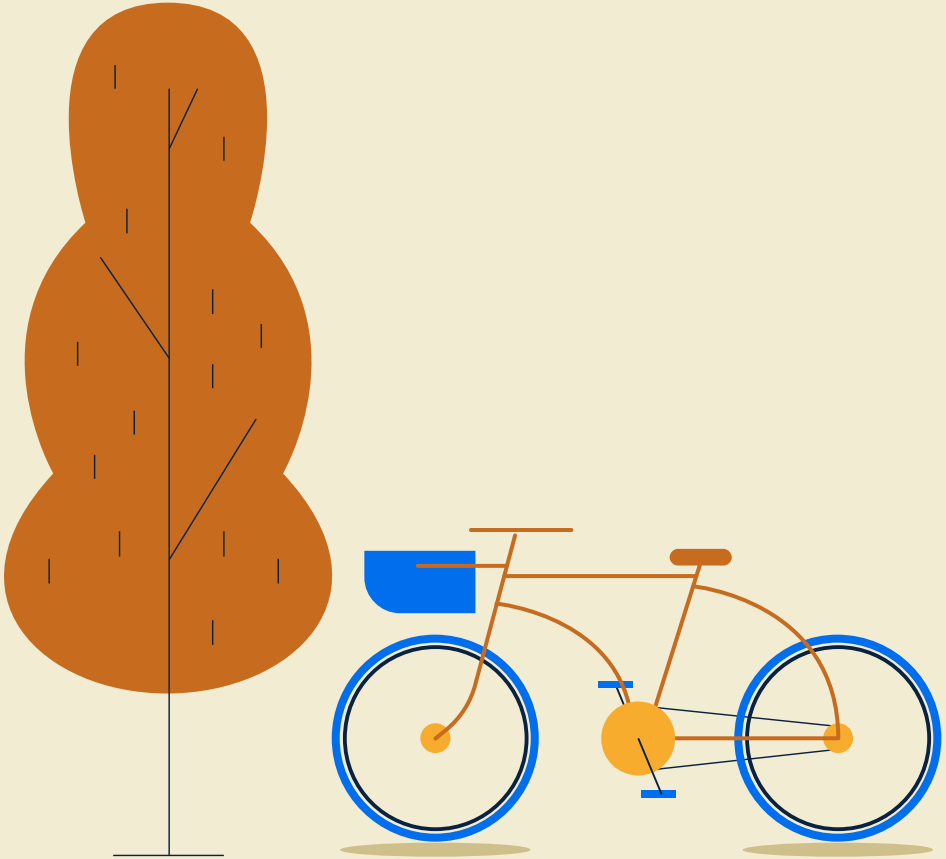
Capital Bikeshare Revenue Prediction

Linear Regression

Presented by
Mai & Basma

Table of contents



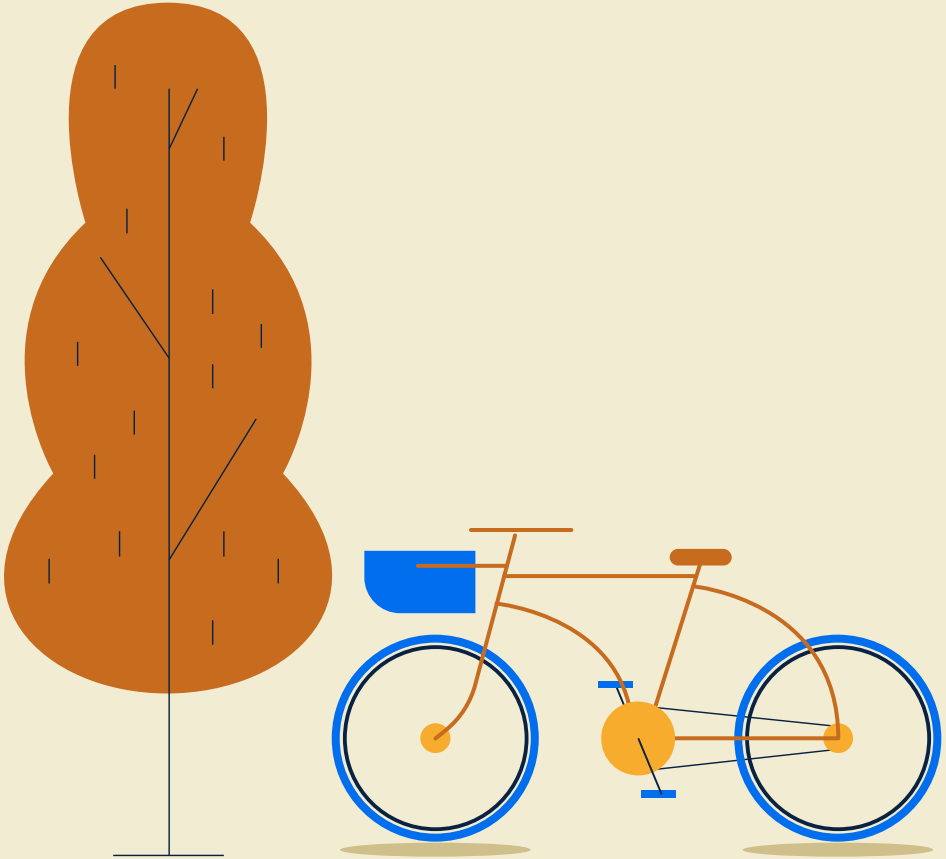
1

Introduction

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.

Capital Bikeshare Location

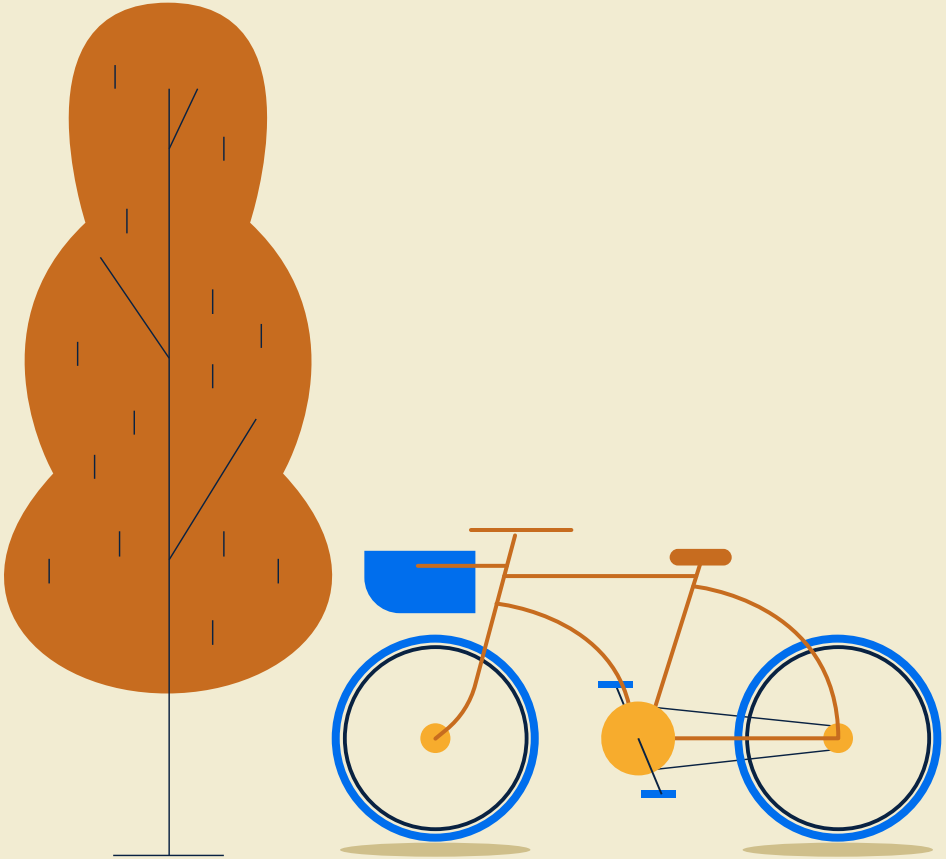


2

Work Flow

Work Flow

1. Idea
2. Web Scraping
3. Data Cleaning
4. EDA
5. Checking Assumptions
6. Building Models
7. Best Model
8. Testing Model



3

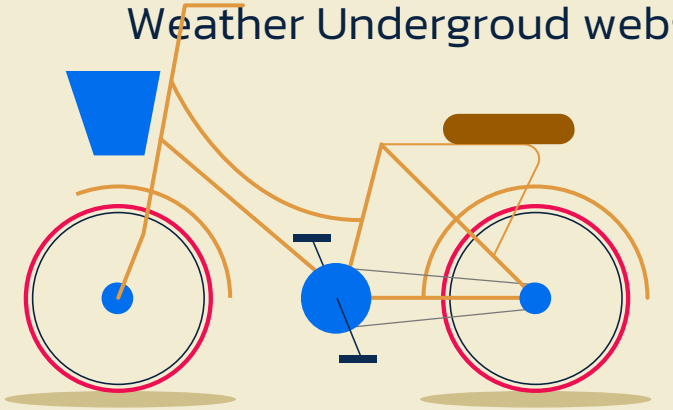
Data

Scraped Data

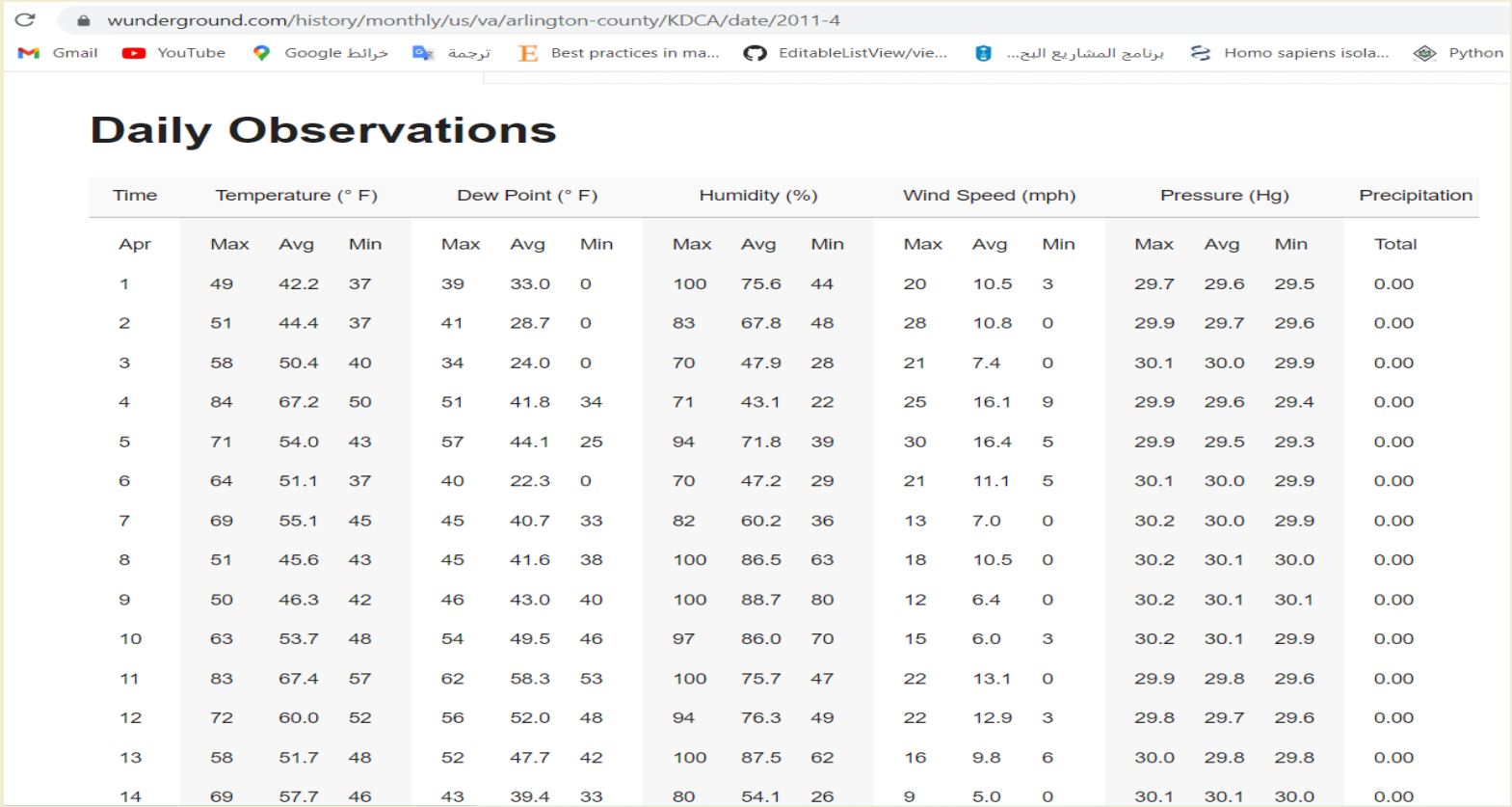
2011,2012

Scrapped the daily weather condtions from

Weather Undergroud website using BeautifulSoup



Scraped Data



Daily Observations

Time	Temperature (° F)			Dew Point (° F)			Humidity (%)			Wind Speed (mph)			Pressure (Hg)			Precipitation
Apr	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Total
1	49	42.2	37	39	33.0	0	100	75.6	44	20	10.5	3	29.7	29.6	29.5	0.00
2	51	44.4	37	41	28.7	0	83	67.8	48	28	10.8	0	29.9	29.7	29.6	0.00
3	58	50.4	40	34	24.0	0	70	47.9	28	21	7.4	0	30.1	30.0	29.9	0.00
4	84	67.2	50	51	41.8	34	71	43.1	22	25	16.1	9	29.9	29.6	29.4	0.00
5	71	54.0	43	57	44.1	25	94	71.8	39	30	16.4	5	29.9	29.5	29.3	0.00
6	64	51.1	37	40	22.3	0	70	47.2	29	21	11.1	5	30.1	30.0	29.9	0.00
7	69	55.1	45	45	40.7	33	82	60.2	36	13	7.0	0	30.2	30.0	29.9	0.00
8	51	45.6	43	45	41.6	38	100	86.5	63	18	10.5	0	30.2	30.1	30.0	0.00
9	50	46.3	42	46	43.0	40	100	88.7	80	12	6.4	0	30.2	30.1	30.1	0.00
10	63	53.7	48	54	49.5	46	97	86.0	70	15	6.0	3	30.2	30.1	29.9	0.00
11	83	67.4	57	62	58.3	53	100	75.7	47	22	13.1	0	29.9	29.8	29.6	0.00
12	72	60.0	52	56	52.0	48	94	76.3	49	22	12.9	3	29.8	29.7	29.6	0.00
13	58	51.7	48	52	47.7	42	100	87.5	62	16	9.8	6	30.0	29.8	29.8	0.00
14	69	57.7	46	43	39.4	33	80	54.1	26	9	5.0	0	30.1	30.1	30.0	0.00

Data Features

Temperature	Dew Point	Wind Speed
Pressure	Humidity	Precipitation
Working Day	Weather Condition	Season

Target

The target of the model is the total revenue of bikeshare company per day.

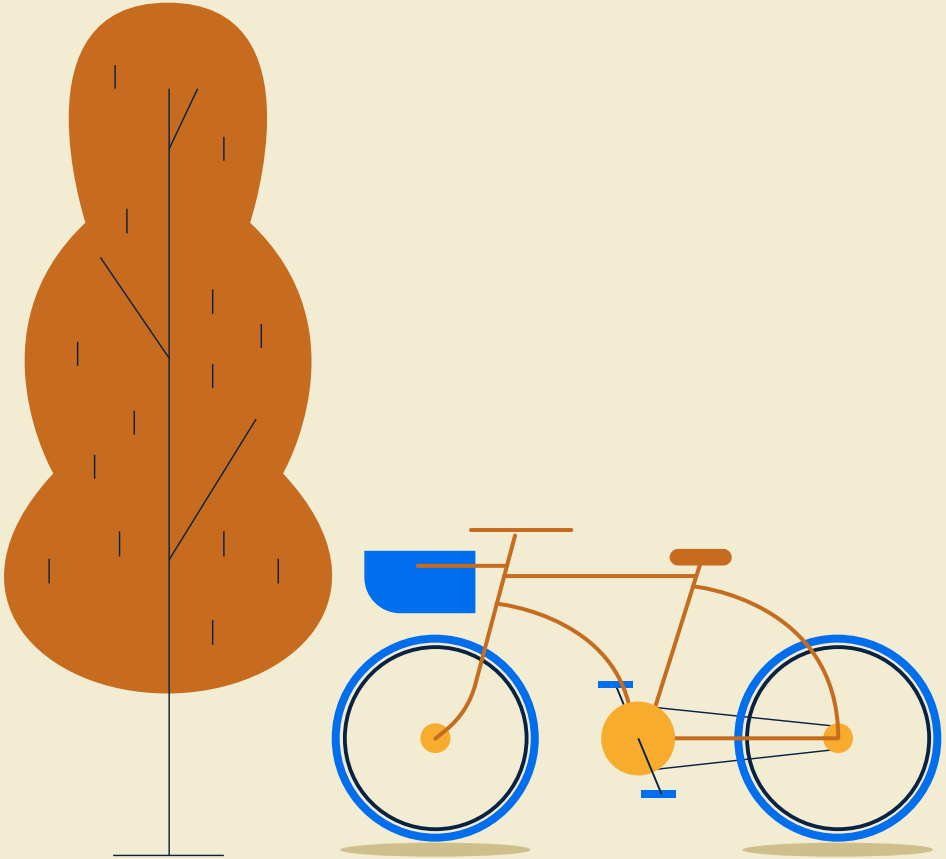
Data

Exploring Data

Explore data after **merging** both datasets

Cleaning Data

By removing **Null** and **Dublicate** values.

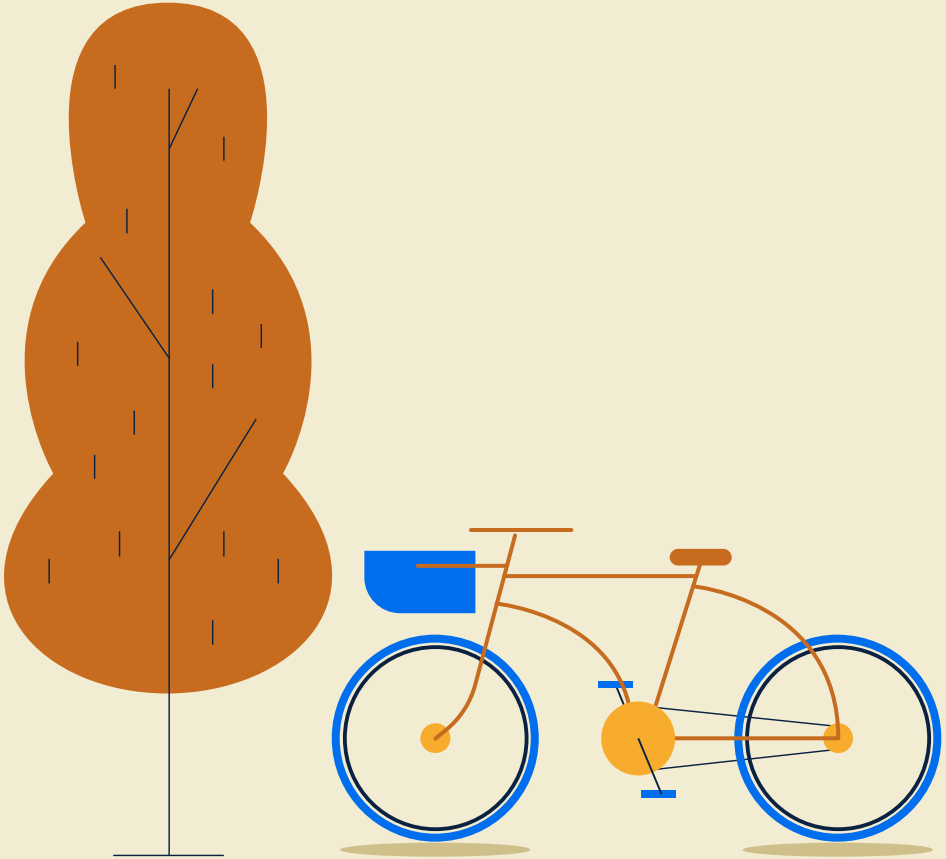


4

Desgin

Design

This project aims to predict the total revenue for capital bikeshare by building a linear regression model. The final data set presents the complete status of the weather in Washington DC and bikeshare information.



5

Tools

Tools

BeautifulSoup

Collect data from website

Pandas

Explor & clean data

Matplotlib, Seaborn

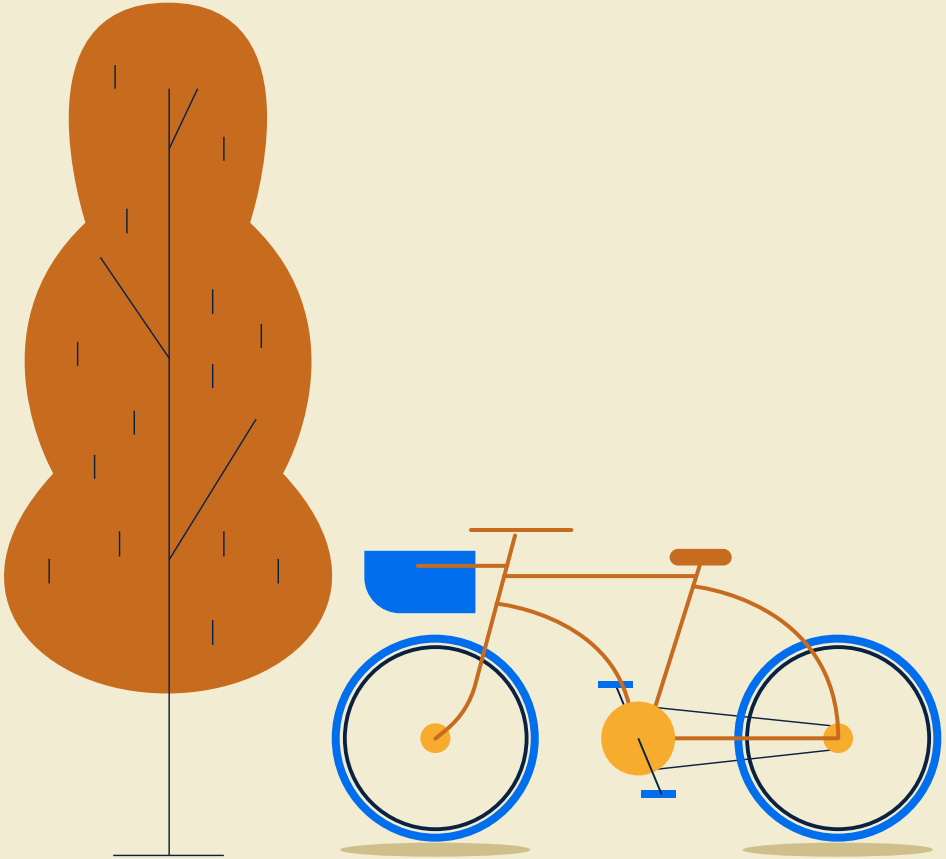
Visualise data and models

sklearn

Models traning

Excel

Download data as csv file



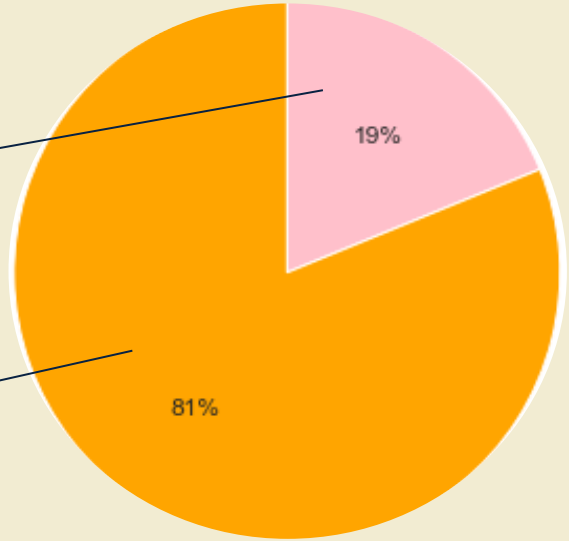
6

EDA - Analysis

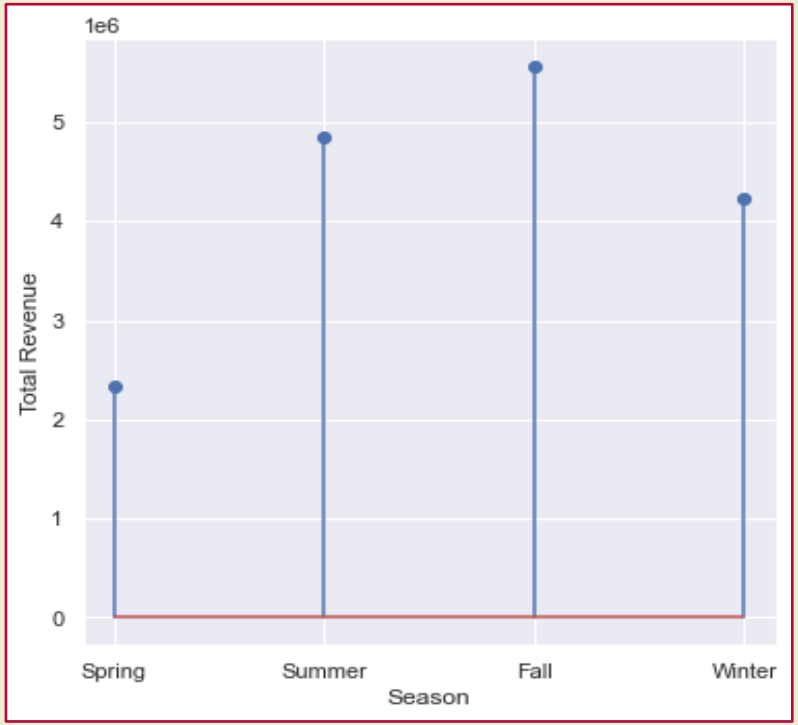
EDA Pie chart

Casual Riders

Registered Riders

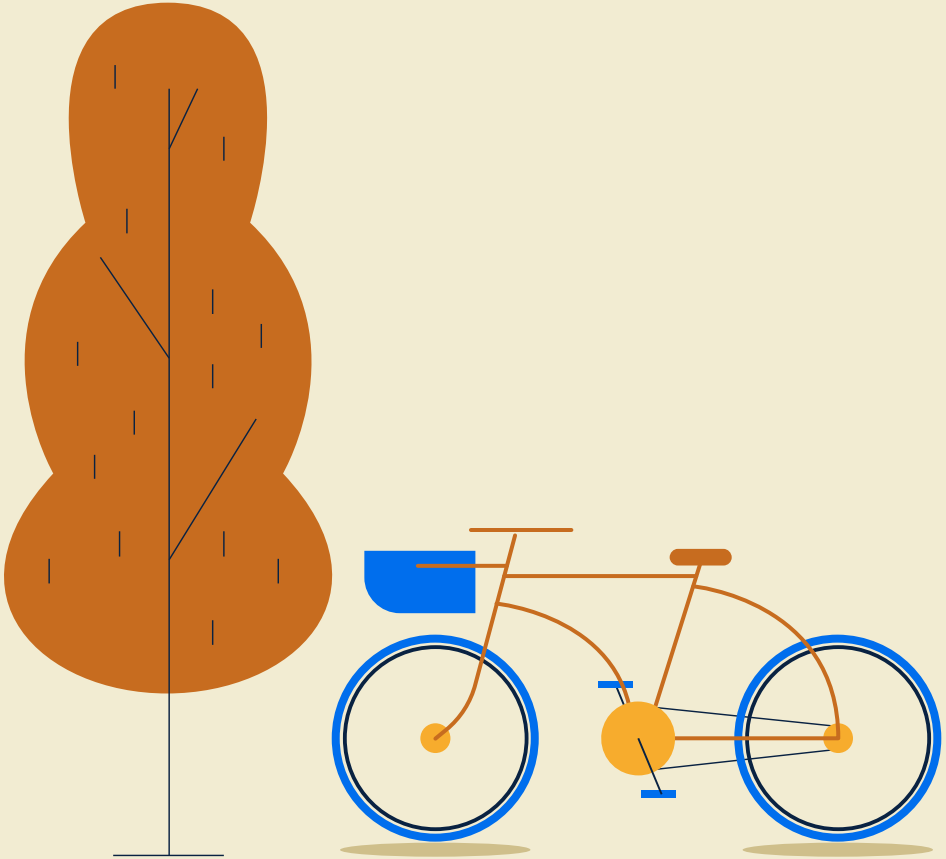


EDA Lollipop graph



Correlation

	temperature	dew_point	humidity	wind_speed	pressure	season	year	month	weekday	bikes_count	total_revenue
temperature	1.000000	0.923648	0.066277	-0.125530	-0.302314	0.335210	0.047242	0.222659	0.001824	0.629202	0.647721
dew_point	0.923648	1.000000	0.428297	-0.189849	-0.343361	0.358848	-0.007092	0.267570	-0.025208	0.507101	0.523259
humidity	0.066277	0.428297	1.000000	-0.200623	-0.218173	0.150917	-0.122031	0.178983	-0.041748	-0.165105	-0.165797
wind_speed	-0.125530	-0.189849	-0.200623	1.000000	-0.315738	-0.210581	-0.012161	-0.193456	0.018161	-0.239832	-0.236982
pressure	-0.302314	-0.343361	-0.218173	-0.315738	1.000000	0.020412	0.046417	0.054921	-0.041393	0.020073	0.018402
season	0.335210	0.358848	0.150917	-0.210581	0.020412	1.000000	-0.001844	0.831440	-0.003080	0.406100	0.385822
year	0.047242	-0.007092	-0.122031	-0.012161	0.046417	-0.001844	1.000000	-0.001792	-0.005461	0.566710	0.528084
month	0.222659	0.267570	0.178983	-0.193456	0.054921	0.831440	-0.001792	1.000000	0.009509	0.279977	0.260944
weekday	0.001824	-0.025208	-0.041748	0.018161	-0.041393	-0.003080	-0.005461	0.009509	1.000000	0.067443	0.069801
bikes_count	0.629202	0.507101	-0.165105	-0.239832	0.020073	0.406100	0.566710	0.279977	0.067443	1.000000	0.985522
total_revenue	0.647721	0.523259	-0.165797	-0.236982	0.018402	0.385822	0.528084	0.260944	0.069801	0.985522	1.000000
holiday_holiday	-0.027812	-0.026257	-0.013989	0.004532	0.063848	-0.010537	0.007954	0.019191	-0.101960	-0.068348	-0.044381
holiday_not a holiday	0.027812	0.026257	0.013989	-0.004532	-0.063848	0.010537	-0.007954	-0.019191	0.101960	0.068348	0.044381
working_day_weekend	-0.053169	-0.055003	-0.035133	0.016849	0.119158	-0.012485	0.002013	0.005901	-0.035790	-0.061156	0.067883
working_day_working day	0.053169	0.055003	0.035133	-0.016849	-0.119158	0.012485	-0.002013	-0.005901	0.035790	0.061156	-0.067883
her_condition_Clear_Partly cloudy	0.120942	-0.119244	-0.620334	-0.004430	0.179764	0.002771	0.029429	-0.021960	-0.029379	0.252870	0.261079
ther_condition_Light Snow, Light Rain	-0.058538	0.047185	0.315434	0.151185	-0.143736	0.070618	-0.073941	0.078543	0.016586	-0.239958	-0.237551
her_condition_Mist_Cloudy	-0.102542	0.104821	0.520595	-0.046879	-0.132381	-0.027762	-0.003869	-0.005365	0.024074	-0.172879	-0.182093



7

Models

Assumptions

1

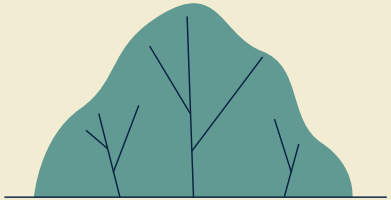
Linear Relastionship between some featurs and target

2

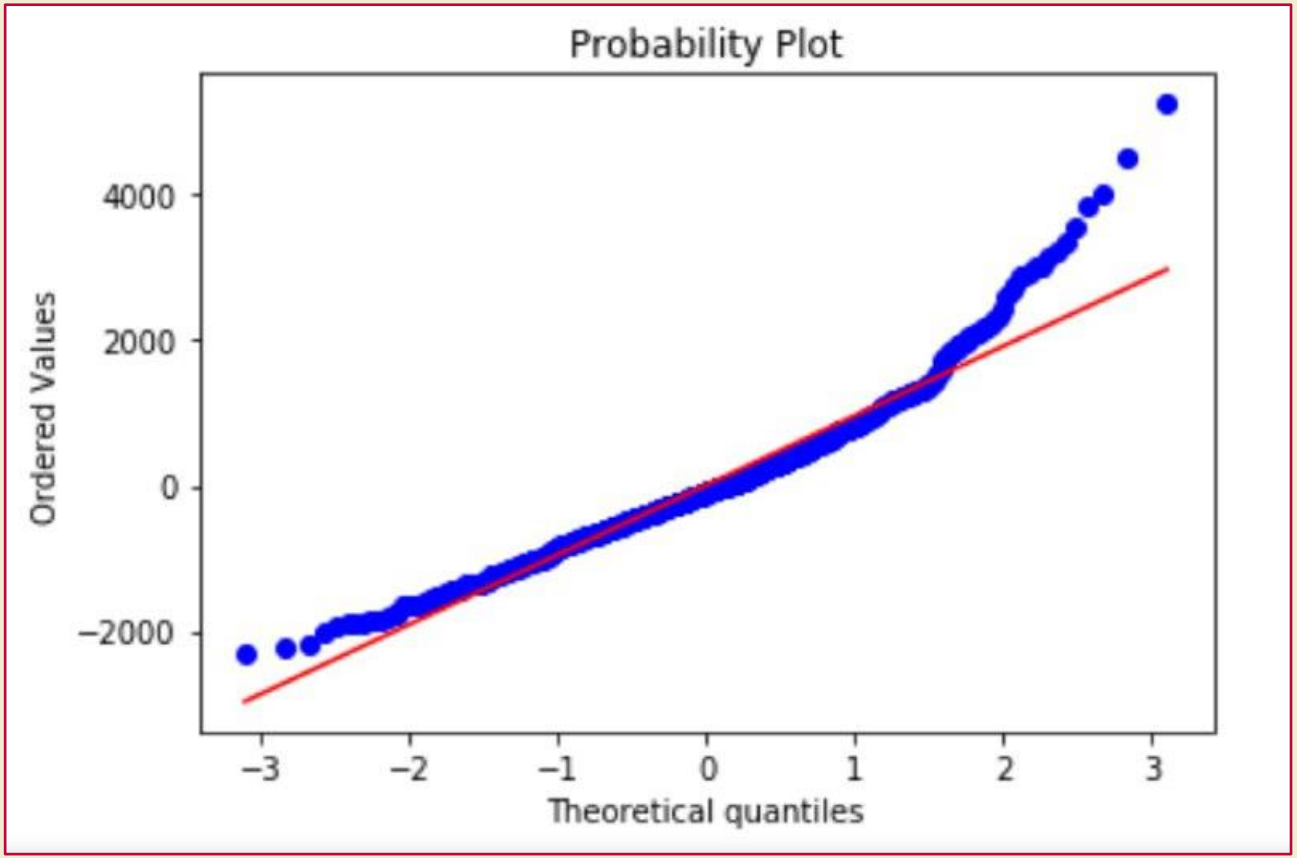
Error is following the **Normal Distribution**

3

There is **No** Strong Multicolinearty

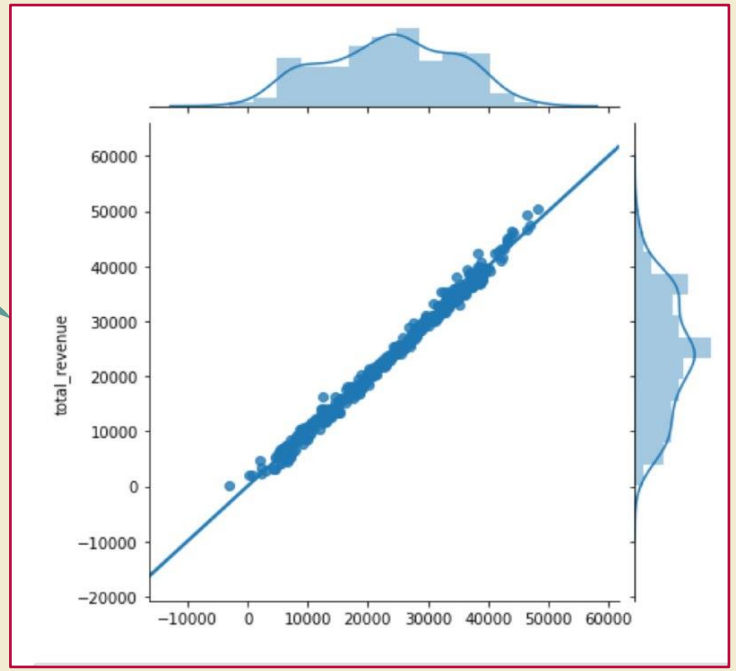


Error is following the **Normal Distribution**



Models

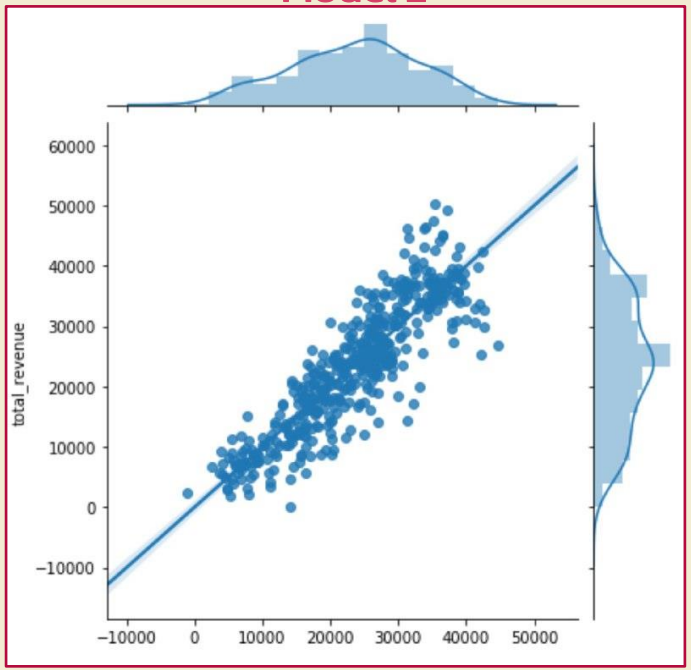
Model 1



Linear Regression train R^2: **0.991**

Linear Regression validation R^2: **0.989**

Model 2

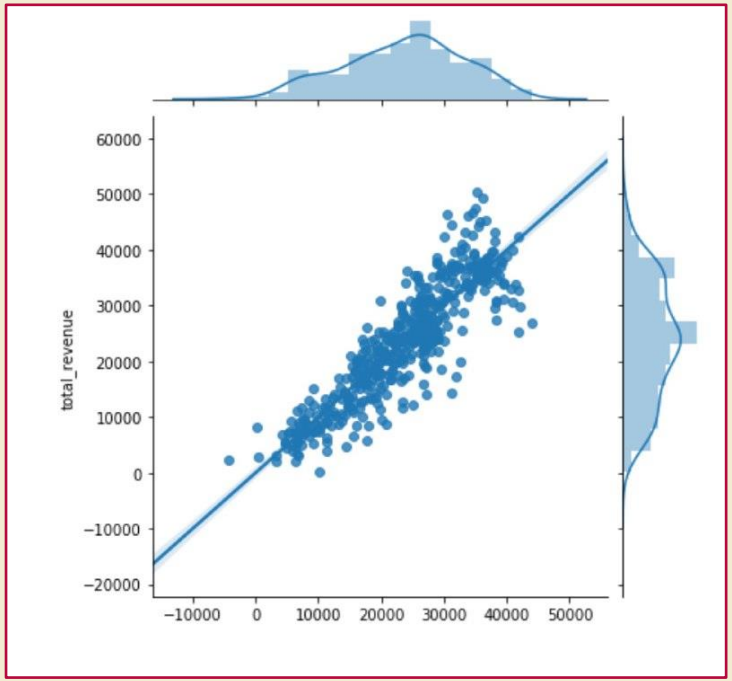


Linear Regression train R^2: **0.771**

Linear Regression validation R^2: **0.757**

Models

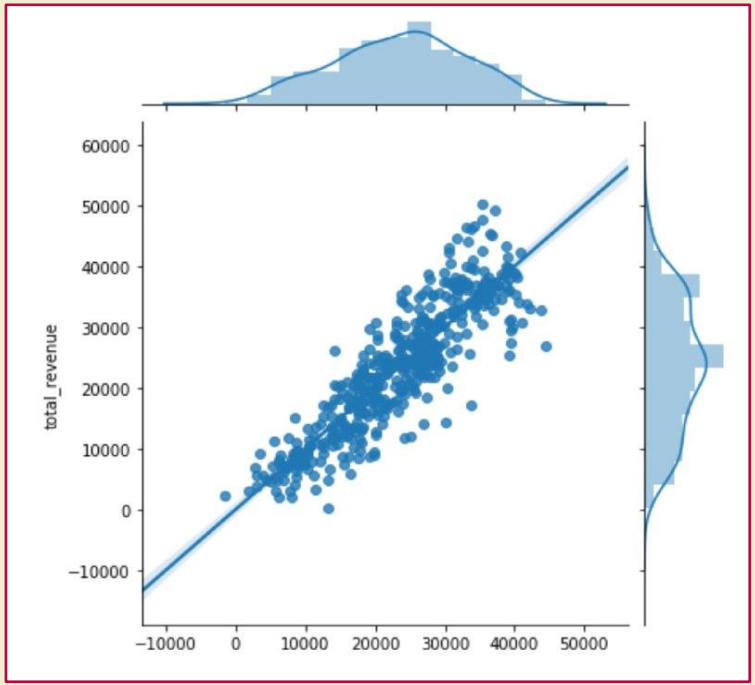
Model 3



Linear Regression train R^2: **0.782**

Linear Regression validation R^2: **0.785**

Model 4

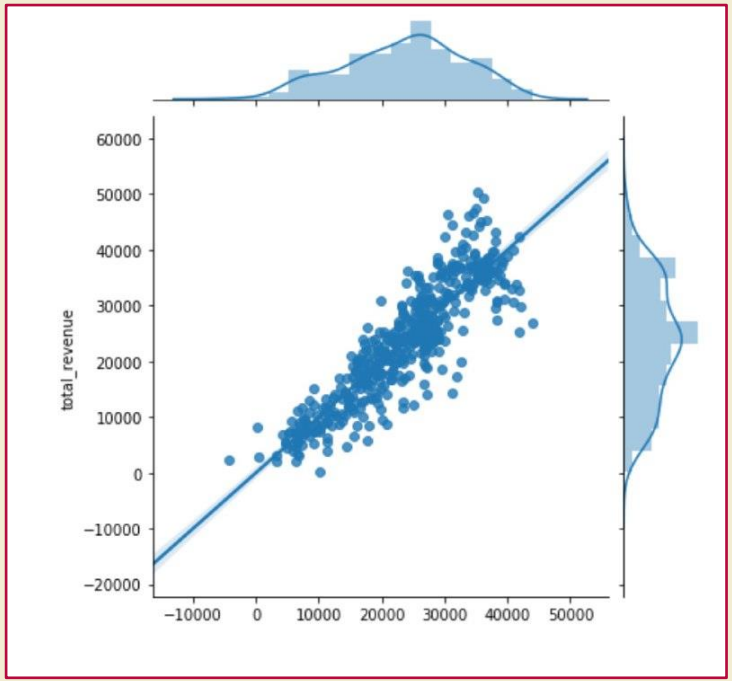


Linear Regression train R^2: **0.772**

Linear Regression validation R^2: **0.766**

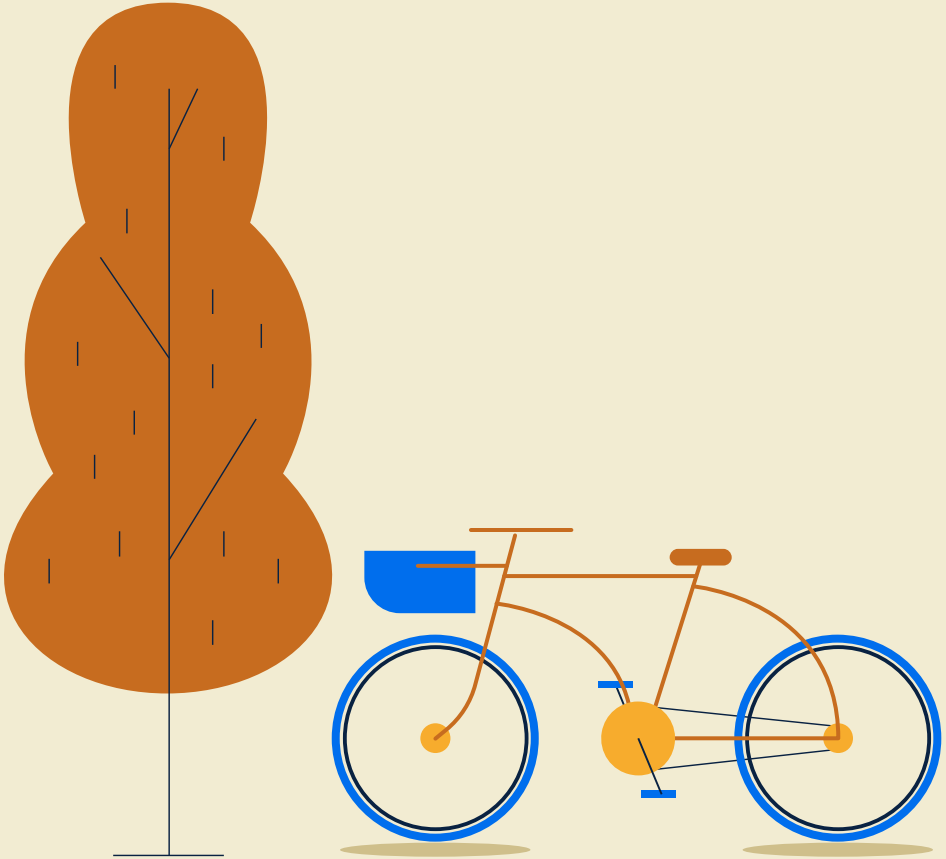
The chosen Model

Model 3



Linear Regression train and validation R^2: 0.784

Linear Regression test R^2: 0.789



8

Colnclusion

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.

Thanks

☺