Report of internship in DataScience, AI&ML

At

Sprout Knowledge Solutions Pvt. Ltd.

24.06.2022 to 12.07.2022

Submitted by:

1.Kavyanjali.S S-(20CS087)
2.Lakshana. R-(20CS089)
3.Manjusree.SV-(20CS093)
4.Prakalya.M-(20CS121)

Department

CSE

College

KPR INSTITUTE OF ENGINEERING AND TECHNOLOGY

Table of contents

- Acknowledgement
- Information about the company
- Information about the internship position
- Description of internship experience
- Project Work
- Conclusion

Acknowledgement

I undertook this internship project and completed the internship report under the guidance of Mr Vignesh, Associate Professor Department of CSE and Mr Narendiranath Academy head of Sprout Knowledge solutions Pvt. Ltd. I'm grateful to the Sprout Knowledge solutions Pvt. Ltd. staff for their patience and assistance during my training at their company. It was a good learning experience for me to work with them, as the project involved real-time scenarios and got an idea of how to deal with real-time projects.

Information about the company

An EdTech Firm that provides Upskilling Courses for Students in Software Coding & Development Programmes across India & various other countries like Australia, the United States, France, Saudi Arabia & Far East. With our Global Experience very soon we are stepping into 11 countries of the African continent.

We Develop our propriety content that suits the current & future trends, which will help the student community navigate the everchanging future. Sprout's Courses are designed to achieve the following for the students - Future Ready, Develop Problem Solving Skills, Enhance Collaboration & Creativity, and Develop Communication.

We aim to create engaging and enriching learning experiences that empower the next generation of computer coders, developers, and makers of the future. Our Industry ready courses with 100% Hands-on methodology and our 4C approach (Critical thinking, Creativity, Collaboration and Communication) make us unique in the tech market. We are also working with DataScience projects for our clients, which makes us unique in this domain.

Description of the internship experience

To be honest, when we initially started, we had no idea what data science or machine learning was. but now we're ready to describe it and carry out some simple jobs. We were mostly assisted by the way Sir taught. Whatever the misunderstanding, he never hesitated to clarify and resolve it. Thank you for your excellent support.

The tasks I undertook included:

1. DataScience Project Lifecycle
2. How to interact with clients and project manager
3. How to identify client requirements
4. Data Collection
5. Data Analysis using Python libraries such as NumPy, Pandas, Matplotlib, Seaborn
6. AI and ML algorithm using Python library Scikit Learn
7. Realtime projects
8. How to document client requirements
9. Creating Insights and presenting them to the project manager as well as the client

Project work

Title: Trip Duration prediction algorithm

Problem Statement:

Happy Ride is a company, they are providing 2-wheeler rides and taxi services at a low cost. They are now working on a mobile application for their booking services. In that, they are working on a feature that will display the trip duration based on the distance (kilometre). Develop an ML model which gets distance as input and shows the trip duration.

Data set: Prepare your own based on the Google map duration and also check OLA, UBER, and RedTaxi fare details.
The process to be followed:

Understanding the business and client requirements:

We, the company Happy ride, have considered developing a mobile application for the business. The happy ride company offers inexpensive bike and car taxi services. In which the business offers a service that allows customers to compute a duration in minutes by entering a distance

BRIGHTER TOMORROW

Sprout Knowledge Solutions Pvt. Ltd.

Data Analysis:

ML model:

Machine learning, a subset of data science, is the scientific study of computational algorithms and statistical models to perform specific tasks through patterns and inference instead of explicit instructions. Machine learning can be described as a set of tools to build models on data. Data scientists explore data, select and build models (machine), tune parameters such that a model fits observations (learning), and then use the model to predict and understand aspects of new unseen data.

```python
from sklearn import linear_model
prediction=reg.predict(x)
plt.scatter(x,y,marker='d',s=100,color='red')
plt.plot(x,prediction)
```
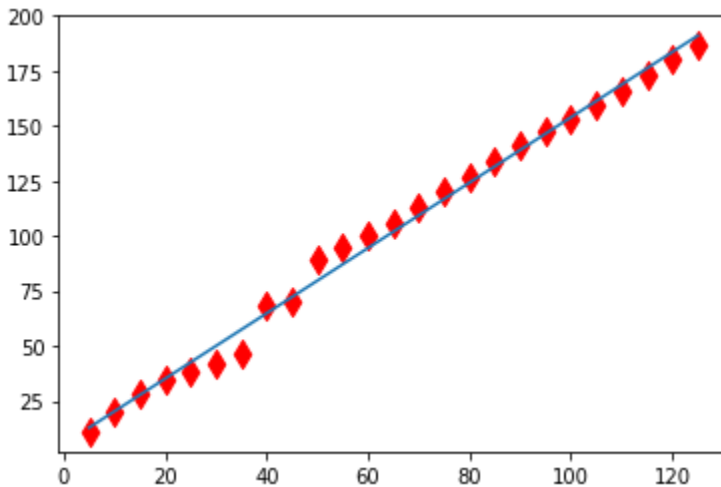
Model Accuracy

Score is used to calculate the accuracy

In multilabel classification, this function computes subset accuracy: the set of labels predicted for a sample must exactly match the corresponding set of labels in y_true.

```
#accuracy
reg.score(x,y)
```

```
 ACCURACY:0.9924632865430826
reg.score(clientdf,resultdf)
```

Library used:

- numpy
- Pandas
- Sklearning
- Matplotlib
- Pickle

Client requirement:

Computing the duration by entering the distance

Input files:

distance in km	duration in min
5	10
10	19
15	27
20	34
25	37
30	42
35	45
40	67
45	69
50	90
55	94
60	99
65	105
70	112
75	119
80	126
85	133
90	140
95	147
100	152
105	160
110	165
115	172
120	179
125	186

Output files:

Client file(car)

distance in	duration in min
5	12
10	19.43
15	26.86
20	34.29
25	41.72
30	49.15
35	56.58
40	64.01
45	71.44
50	78.87
55	86.3
60	93.73
65	101.16
70	108.59
75	116.02
80	123.45
85	130.88
90	138.31
95	145.74
100	153.17
105	160.6
110	168.03
115	175.46
120	182.89
125	190.32

Client file(bike)

distance in	duration in min
5	12
10	19.43
15	26.86
20	34.29
25	41.72
30	49.15
35	56.58
40	64.01
45	71.44
50	78.87
55	86.3
60	93.73
65	101.16
70	108.59
75	116.02
80	123.45
85	130.88
90	138.31
95	145.74
100	153.17
105	160.6
110	168.03
115	175.46
120	182.89
125	190.32

Source code:

Drive Link:

https://drive.google.com/drive/folders/1mw9TNY2AIGzvrfQp51OkYNCmKEfsvr-F
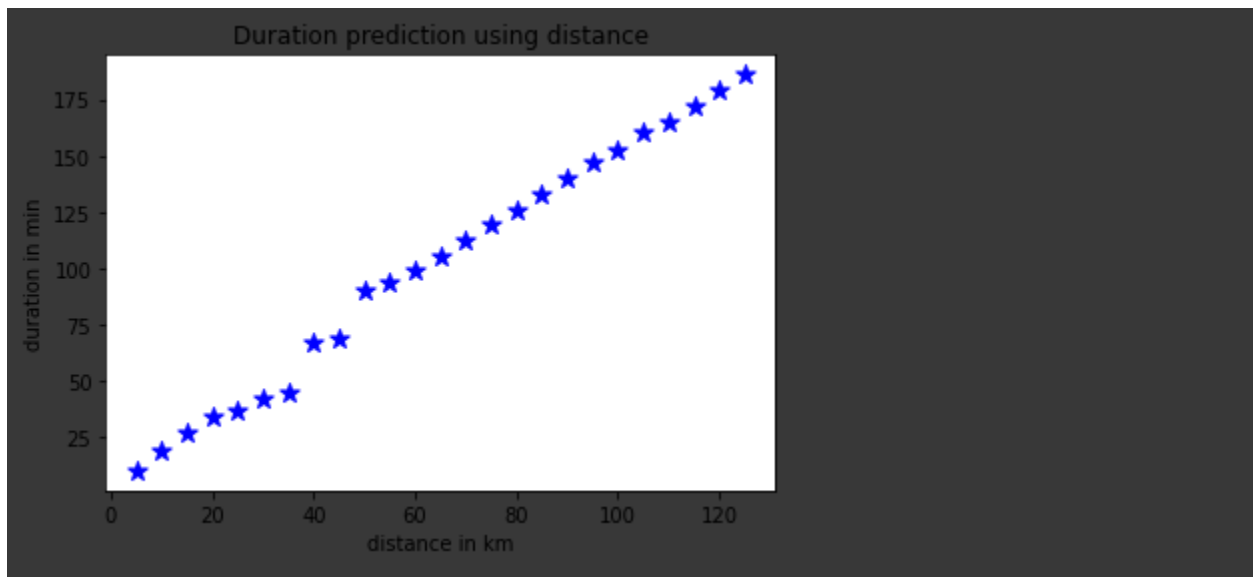
```
bikedf=pd.read_csv("/content/bike.csv")
bikedf
```

```python
#scatter
plt.scatter(x,y,marker='*',s=100,color='blue')
plt.xlabel('distance in km')
plt.ylabel('duration in min')
plt.title('Duration prediction using distance')
```
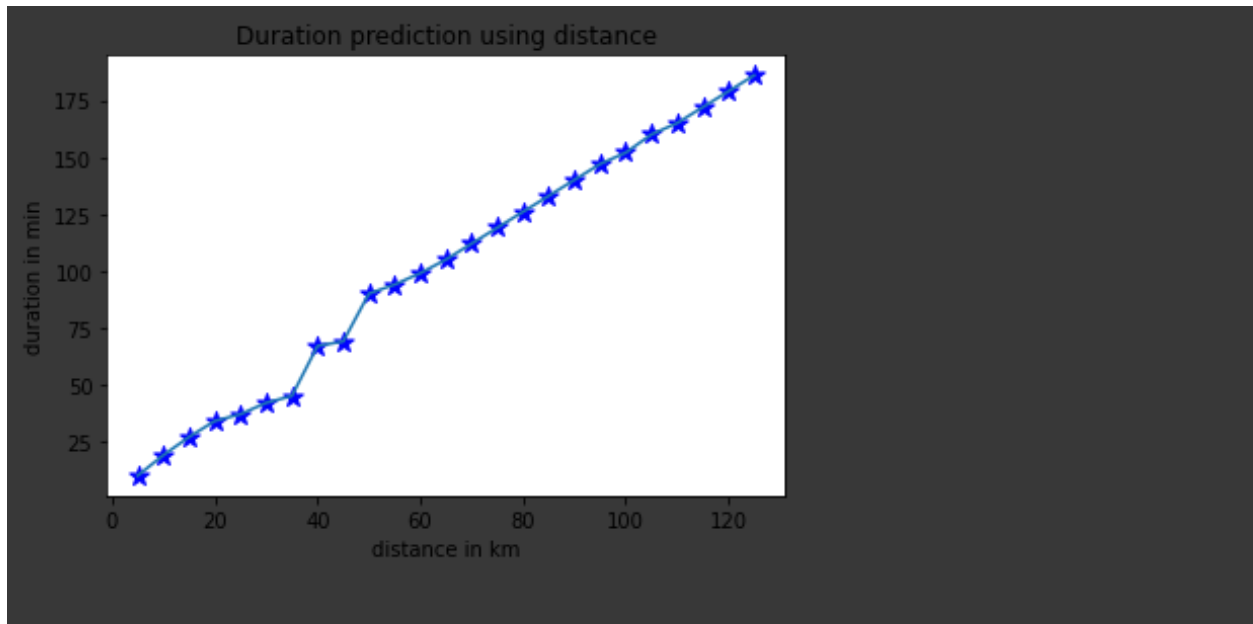


```python
#plotting
plt.scatter(x,y,marker='*',s=100,color='blue')
plt.xlabel('distance in km')
plt.ylabel('duration in min')
plt.title('Duration prediction using distance')
plt.plot(x,y)
```
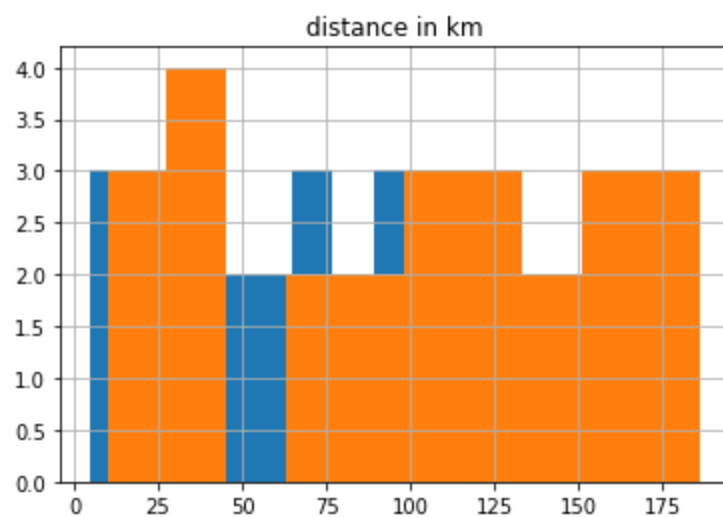
```
#fitting
reg.fit(x,y)
```
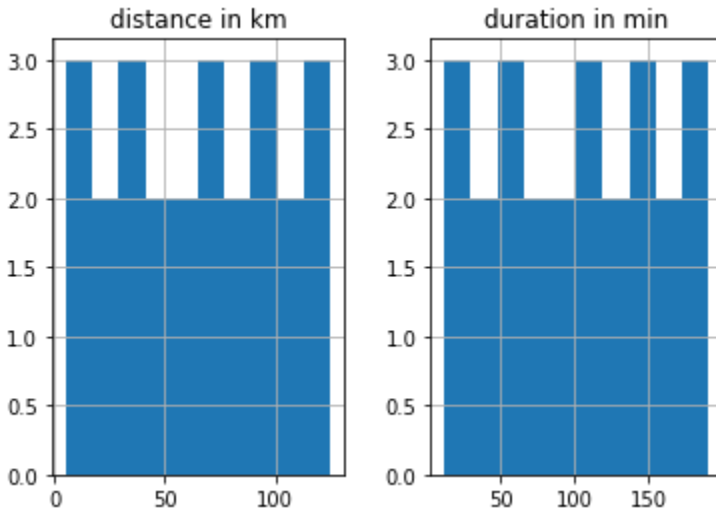
```
#testing
reg.predict([[7]])
ytest=reg.predict(x)
```



```
x.hist()
y.hist()
plt.show()
```

```
clientdf.hist()
plt.show()
```


**Model:**

Trip Duration prediction model

Happy Ride is a company, they are providing 2-wheeler rides and taxi services at a low cost. They are now working on a mobile application for their booking services. In that, they are working on a feature that will display the trip duration based on the distance (kilometre). Develop an ML model which gets distance as input and shows the trip duration.

Screenshot
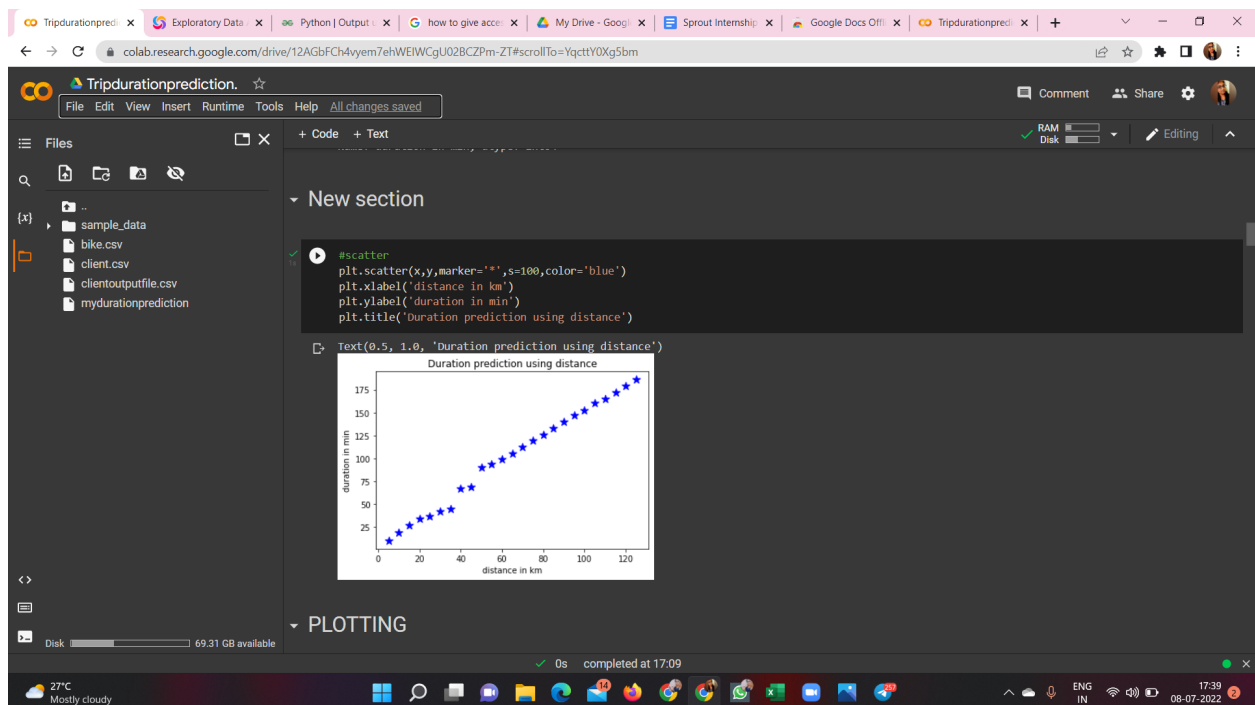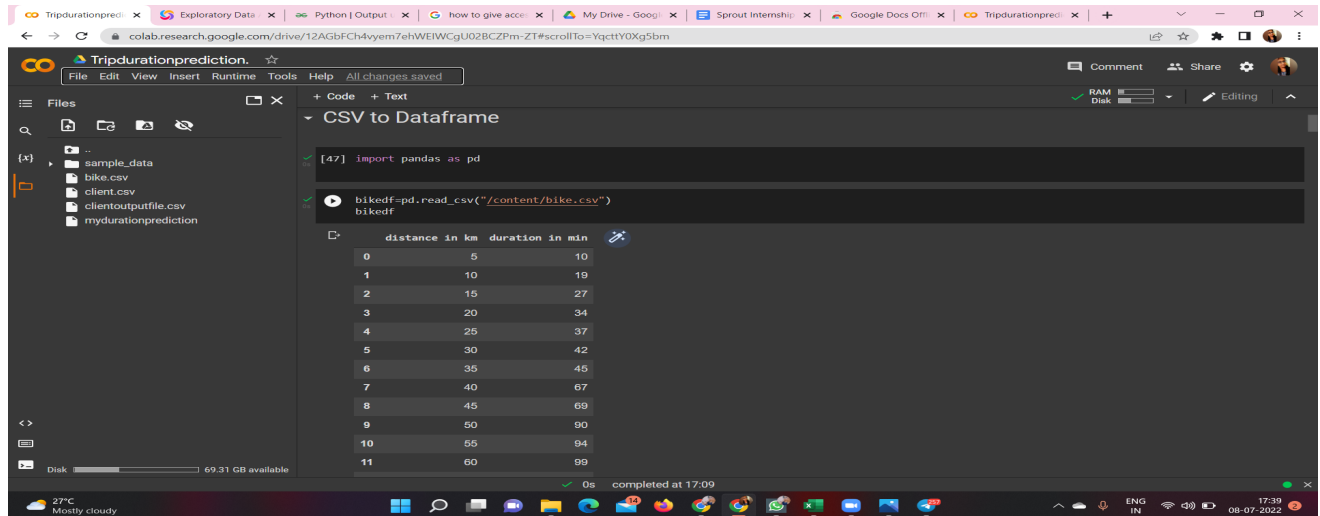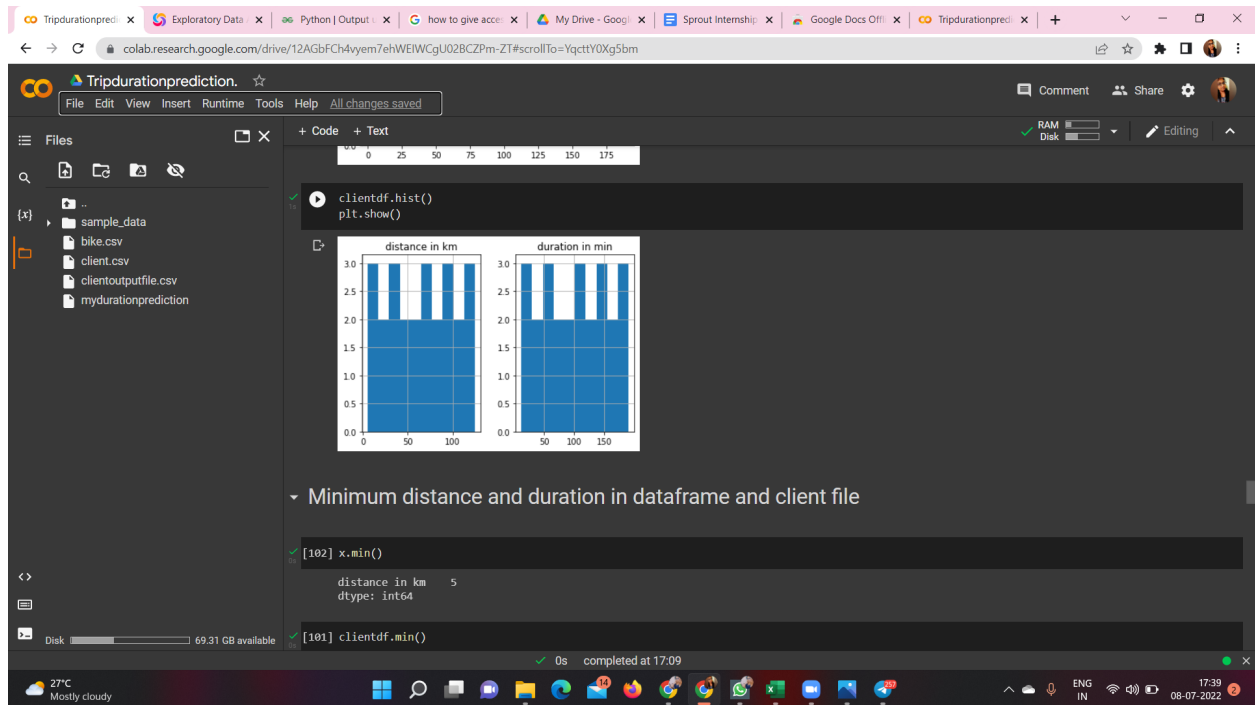
Sample output:

```
clientdf['duration in min']=resultdf
clientdf
```

```python
#Getting an input(distance in km) from the user/client and producing the
output(duration in min)dis=input("Enter a distance in kilometer ")
print("The Duration taken is",reg.predict([[dis]]))
```

OUTPUT:
```
Enter a distance in kilometre 234
The Duration taken is [352.294]
```
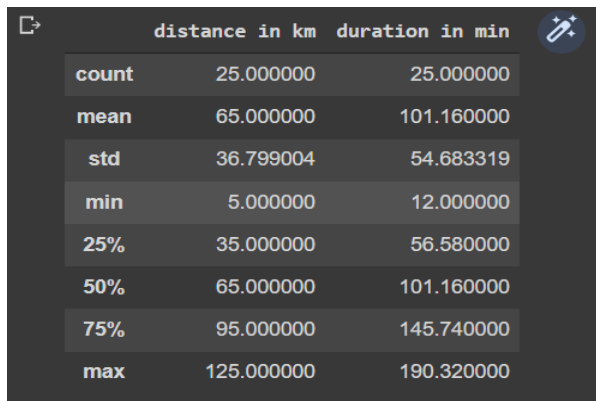
```
clientdf.describe()
```

	distance in km	duration in min
count	25.000000	25.000000
mean	65.000000	101.160000
std	36.799004	54.683319
min	5.000000	12.000000
25%	35.000000	56.580000
50%	65.000000	101.160000
75%	95.000000	145.740000
max	125.000000	190.320000

Drive link for Car Project:

https://drive.google.com/drive/folders/1spyRdT3AkTwSxy5b8fzjpFHqxjLDEHcm

Conclusion

During my Internship, at Sprout Knowledge Solutions Pvt. Ltd. I understood the actual workflow of how Data Science projects will look like and had hands-on experience with some projects. However, I need to learn a lot in this domain to excel.

Overall I got Hands-on experience in DataScience, AI and ML.