Report of internship in DataScience, AI&ML

# At

# Sprout Knowledge Solutions Pvt. Ltd.

# 

# 11.07.2022 to 11.08.2022

## 

## Submitted by

## Student Name: Seetharaman S

## Department: AI & DS

## College: Karpagam Institute of Technology

## 

# Table of contents

* Acknowledgement
* Information about the company
* Information about the internship
* 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 proprietary content that suits the current & future trends, which will help the student community navigate the ever changing 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

***Student’s experience with the internship***

I learnt a lot about the many Machine Learning Techniques in the Internship. Gained Knowledge about the Missing Data and Filling the Missing Data in the Data frame. I got more insight about the Linear Regression and Logistic Regression. And learnt more about using the data efficiently. The Internship helps me to gain knowledge about the Data Science and Machine Learning with the Hands-on training and via. Mini Projects.

The Mini Project is very Useful to understand the Problem in Data Analysis and Data Science Project Life Cycle. Mr. Narendiranath Academy head of Sprout Knowledge Solutions Pvt. Ltd was Guided us throughout the Internship and Taught every nook and corner of every Libraries used in the project.

It helps me to gain the Insight about the Data Science and Machine Learning Technique. The Internship Helps to Know more about the Data Science Project Lifecycle.

## 

## 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 (based on the project given)***

**Title:**

**Trip budget prediction algorithm**

**Problem Statement:**

Arjun, Darsh and Harsha are planning to go for a few day's trips. They need to know the budget for different numbers of days. For example, they want to go to OOTY for 3 days. Develop an ML algorithm to help find the best budget trip based on the distance (kilometer) and the number of days using the DataScience.

**Data set:**

Prepare on your own

**Process:**

Data Science LifeCycle

**Data Science Lifecycle:**

* The same survey shows that 53% of enterprises using AI today recognize unexpected outcomes and predictions as the greatest risk when building and deploying machine learning (ML) models this can be solved by the Data Science Life cycle.
* A data science lifecycle indicates the iterative steps taken to build, deliver and maintain any data science product. All data science projects are not built the same, so their life cycle varies as well.
* “Exploring the Data” is the most important step in the Data Science Project Life Cycle. • Data Science Life Cycle have many steps
* Gathering Data
* Cleaning Data
* Exploring Data
* Modeling Data
* Interpreting Data
* A Data Science life cycle is nothing but a repetitive set of steps that you need to take to complete and deliver a project/product to your client.
* A General data science lifecycle process includes the use of machine learning algorithms and statistical practices that result in better prediction models.

• It helps to convert the big quantity of uncooked and unstructured records into significant insights.

**Understanding the Business and Client requirement:**

* Knowing and Understanding Customer needs is at the center of every successful business.
* Know who your clients are they will make your business more successful.
* Understanding your Client Requirements allows you to better market your business.
* Our Knowledge of the Client’s business also helps us to develop and maintain a positive relationship with the client.
* It gives you an opportunity to make right change at the right time to improve upon the customer journey and meet their needs.
* Analyze your customer feedback to understand the Requirements of the client.
* Getting Personal feedbacks is also useful in understanding the Business and client Requirement.
* In our case, Happy Ride is a company (The Client), they are providing 2-wheeler rides and taxi services. They are working on a feature that will display the trip duration based on the distance.
* The Client wants to get the Estimated Time taken in Hours to the given Distance as input in the Kilo Meters.

**Data Collection**

* Data Collection is the process of the gathering and Measuring information from countless different sources.
* In order to use the data, we collect to develop practical projects solutions, it must be collected and stored in a way that makes sense for the business problem at hard.
* The data collected using Interviews, Questionaries and surveys, Observation, Document and Records, Focus groups and Oral Histories.
* The Main purpose of data collection is to gather information in a measured and systematic manner to ensure accuracy and facilitate data analysis.
* The Data Collection is the content for the Data Analysis in the process.
* In our case, The Data Set is created by ourself based on the Google Map Duration and Also referred in Ola, Uber and Red Taxi Data.
* The Data Set is Consists of the two variables Distance and Time Taken.
* The Good data is required to get the Good Prediction. The Data should be collected in the Proper Manner and Good Source.

**Data Analysis**

* Data Analysis is the process of inspecting, Cleaning, Transforming and Modeling data with the goal of the discovering useful information by informing conclusion and supporting decision making.
* Machine Learning automates the entire Data Analysis workflow to provide Deeper, Faster and more Comprehensive Insights.
* The Data analysis is useful to the clean the data and get the full efficacy of the data.
* It also helps to find possible solutions for a business problem.
* It is the practice of working with data to glean useful information, which can then be used to make informed decisions.
* The Data Set should not contain the null values. If it contains the Null values, it will affect the Accuracy of the Prediction. If the null values are present in the Data Set the missing data are to be filled using the required Methods.
* The Data Analysis must be done to make the further process of the project.

**ML Model**

* A machine learning model is defined as a mathematical representation of the output of the training process.
* Machine Learning models can be understood as a program that has been trained to find patterns within new data and make predictions.
* These models are represented as a mathematical function that takes requests in the form of input data, makes predictions on input data, and then provides an output in response.
* Machine Learning Model is a Binary File which contains the Trained data from the Data Set.
* It can be Downloaded using the Pickle Library.
* And the Model File can be reused many times whenever the trained data want to be used.
* This Method is used to reuse the trained model in the various other project. It will reduce the time taken to reduce the time to write the long code.
* Based on different business goals and data sets, there are three learning models for algorithms. Each machine learning algorithm settles into one of the three models:
* Supervised Learning
* Unsupervised Learning
* Reinforcement Learning

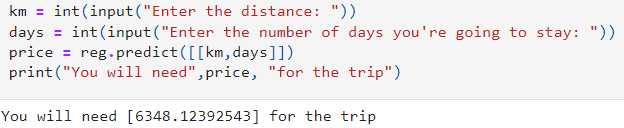
**Insight (Final Output)**

* Data insights refers to the deep understanding an individual or organization gains from analyzing information on a particular issue.
* The Output Show the Estimated the Time Taken in Hours to Travel the given Distance in the Kilometer.
* The Output should be clear and want to be Well Defined.
* The Output should be easily readable but the client.
* Output should be accurate and the inaccurate output should not be accepted but the client.

The Output that gets two input and

display the Predicted budget for the

distance for days and distance

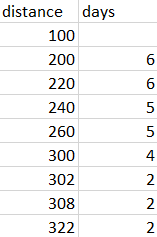
****

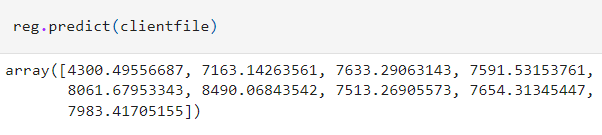
The Output when the client gives the

input as a .csv file and want to

get the output as .csv file

**Client file**



****

**Data set:**

* The Data Set should be Cleaned and Collected from the Good Source. It will be useful in the Process of the Prediction of the output.
* The Data Set should be Visualized to get the better insights of the collected Data.
* Data Set plays a major role in the Data Science Lifecycle. The Data Set should be trained and fitted to the Data Model that can be Downloaded using the Pickle Library.
* The Data Set file is loaded into the Data Frame using the Pandas Library.

**Input files:**

**https://github.com/seetharaman52/MiniProjects-/tree/main/Project%205%20(Trip%20Budget)**

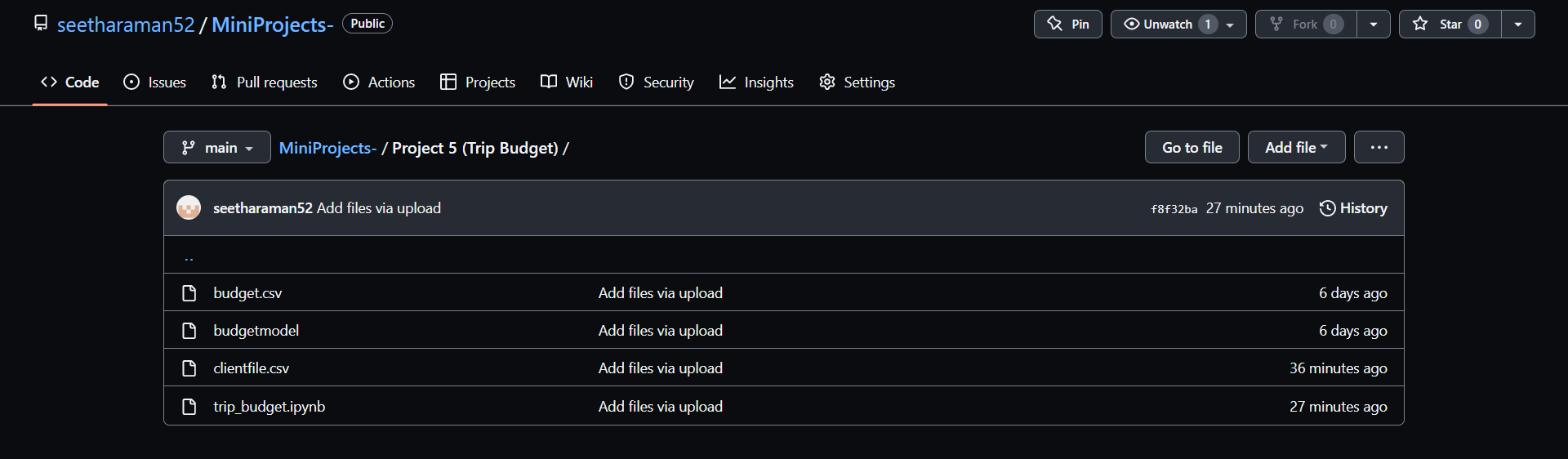
**Output files:**

**https://github.com/seetharaman52/MiniProjects-/tree/main/Project%205%20(Trip%20Budget)**

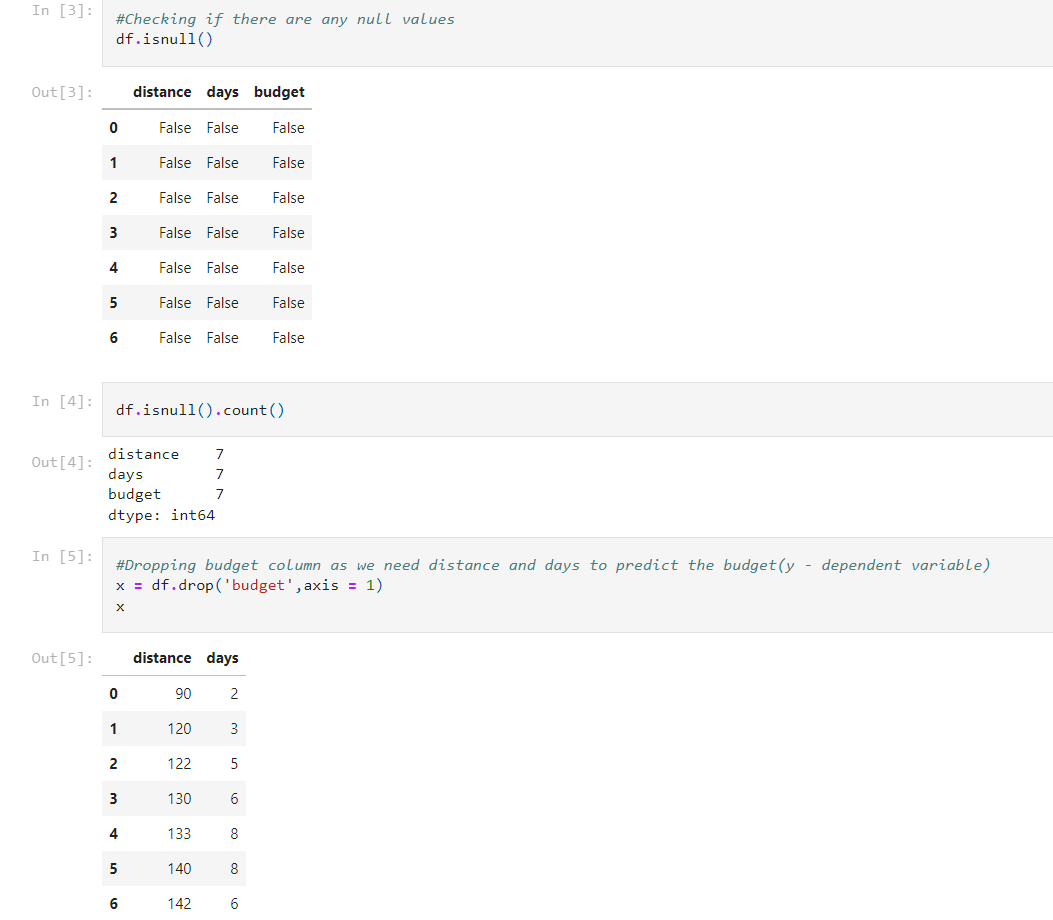
**Source code:**

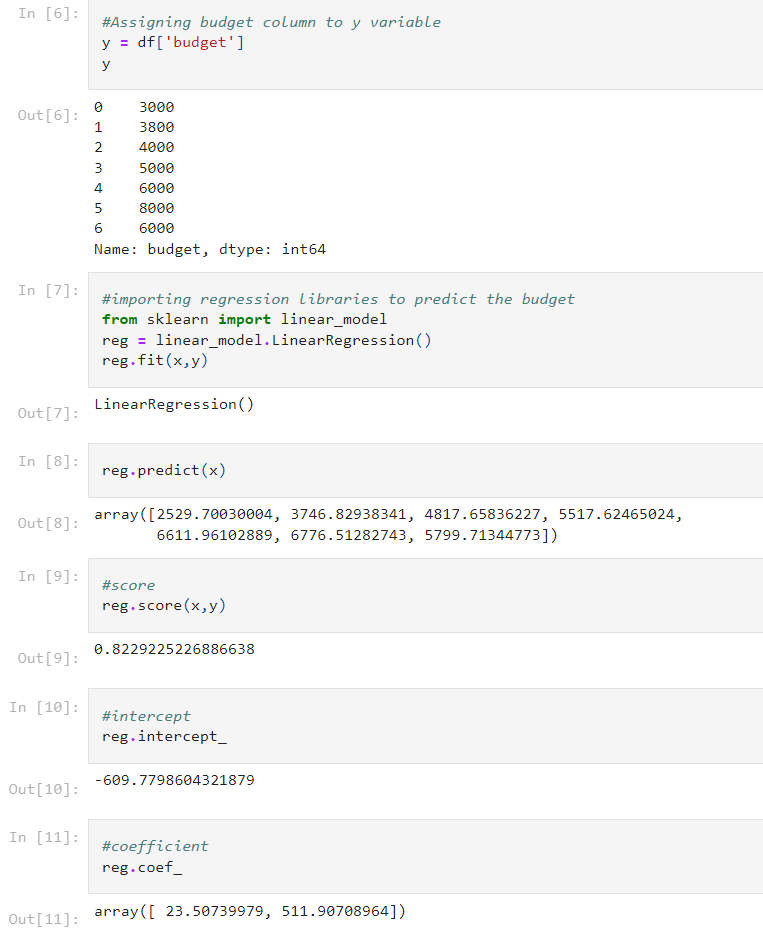
**https://github.com/seetharaman52/MiniProjects-/blob/main/Project%205%20(Trip%20Budget)/trip\_budget.ipynb**

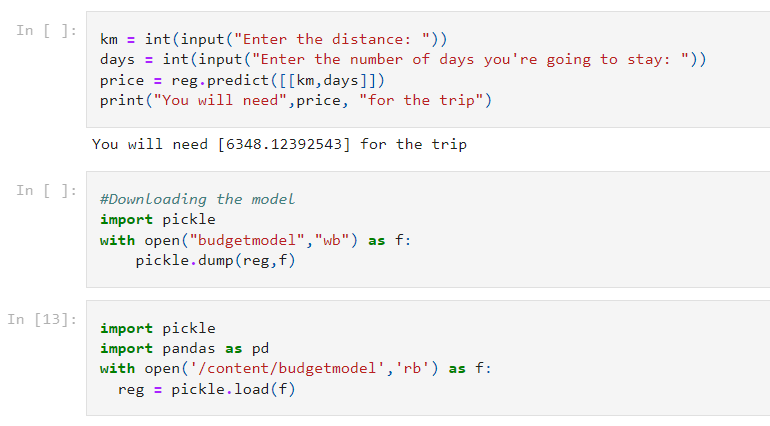
**Screenshot:**

****

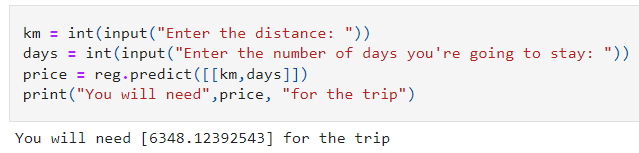
****

****

****

****

**Sample output:**

****

**Project Explanation:**

The project is to predict the Trip Budget (price) with help of kilometers (distance) of the place and days (Number of days i.e., staying period). For the prediction and make the Data Science Life Cycle should be followed. First step to the prediction is to Understand the Problem Properly because the misunderstanding the problem may cause the errors in the prediction. Understanding the problem is the most important and have some Domain knowledge. If we understood the problem the problem can be solved easily and efficiently. The Data should be gathered from the good resources because the data set plays the major role in the machine Learning part. The data should be properly arranged. The Data Set should be Visualized to get the better insights of the collected Data. We Collected and Created the Project’s Data Set by ourself. The Next part is the Data Analysis. Once the data is collected the data should be analyzed and get rid of the null values in the data set. The null values must be filled by the different methods. If the Data Set has the null values unfilled the data set in incomplete and results the low accuracy of Sprout Knowledge Solutions Pvt. Ltd. the Prediction. So, Data set must be properly analyzed and cleaned the null Values of the data set. The Project’s Data Set is crested by ourself without any null values. Then, we should make the regression model and fit the data set in the ML Model to get the Prediction.­­­­

# 

# 

# 

# Conclusion

From my Internship at Sprout Knowledge Solutions Pvt. Ltd. I was able to understand 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.

Submit your projects: [narain2829@gmail.com](mailto:narai2829@gmail.com)