

Car Price Prediction Model

Submitted by:

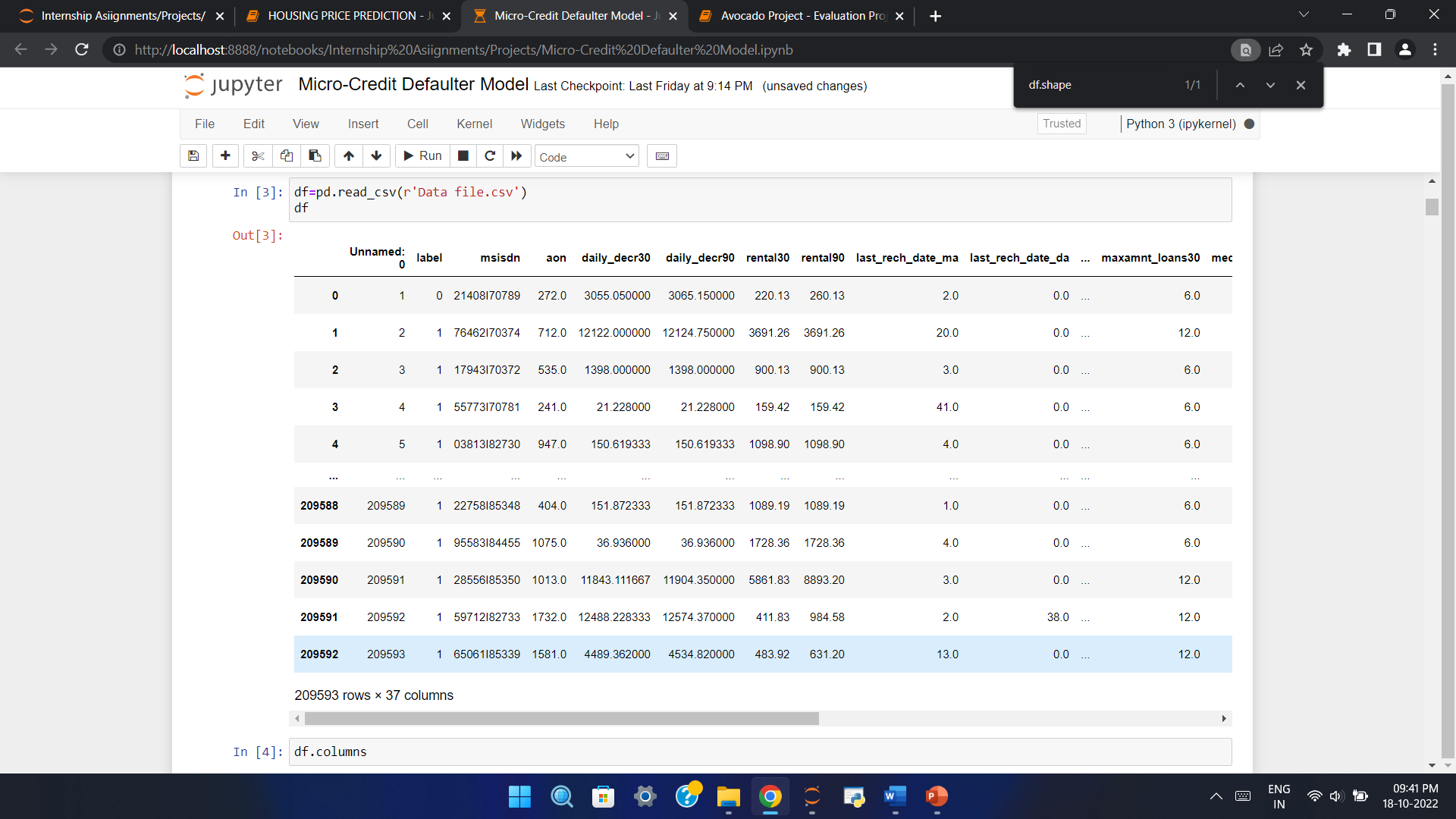
Parija Deshmukh

**INTRODUCTION**

With the covid 19 impact in the market, we have seen lot of changes in the car market. Now some cars are in demand hence making them costly and some are not in demand hence cheaper. One of our clients works with small traders, who sell used cars. With the change in market due to covid 19 impact, our client is facing problems with their previous car price valuation machine learning models. So, they are looking for new machine learning models from new data.

**Analytical Problem Framing**

* Data Sources and their formats



Dataset is contains 6000 rows and 6 columns. Data set has different features like Brand, model, kilometres, year of manufacturing and Fuel type. There is target columns as Sales Price having continuous numeric values.

* Data Pre-processing Done

1. Checked null values and no null values found.
2. Deleted Unnamed: 0 column as it wasn’t significantly contributing in data analysis.
3. Converted object type columns into numeric columns using LabelEncoder.
4. Using astype() converted object type Kilometres column into float data type.

* Data Inputs- Logic- Output Relationships

1. Brand of car is highly negatively correlated with target column – sales price with correlation of -0. 584247 value.
2. Kilometers is highly positively correlated to Sales price column with value: 0. 408952.

* Hardware and Software Requirements and Tools Used

The preliminary step involved in devising a model is loading the required libraries. In this case, we mainly load four libraries namely pandas, numpy, scipy, sklearn and seaborn.

1. Pandas: Pandas is a python package which is quite quick, easy to use and structured in nature. Pandas data-frame is mainly used for data analysis purposes. Pandas also helps us to handle missing data, data to be reshaped and data transformation methodologies.
2. Numpy: Numpy is essentially used for creating very powerful and intuitive n-dimensional arrays. It offers various mathematical functions and also supports various kinds of computing hardware and software requirements. It is an open-source project and also contains various array-objects which are generally quick for usage.
3. Sci-kit: Sci-Kit is one of the most efficient and useful machine learning libraries in python. It provides various statistical and mathematical tools. It also has various techniques like regression, clustering etc. in it.
4. Seaborn: Seaborn helps us to create plots and live-interactive statistical plots and images. It also contains matplotlib which helps us to view our data in a much more intuitive manner.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

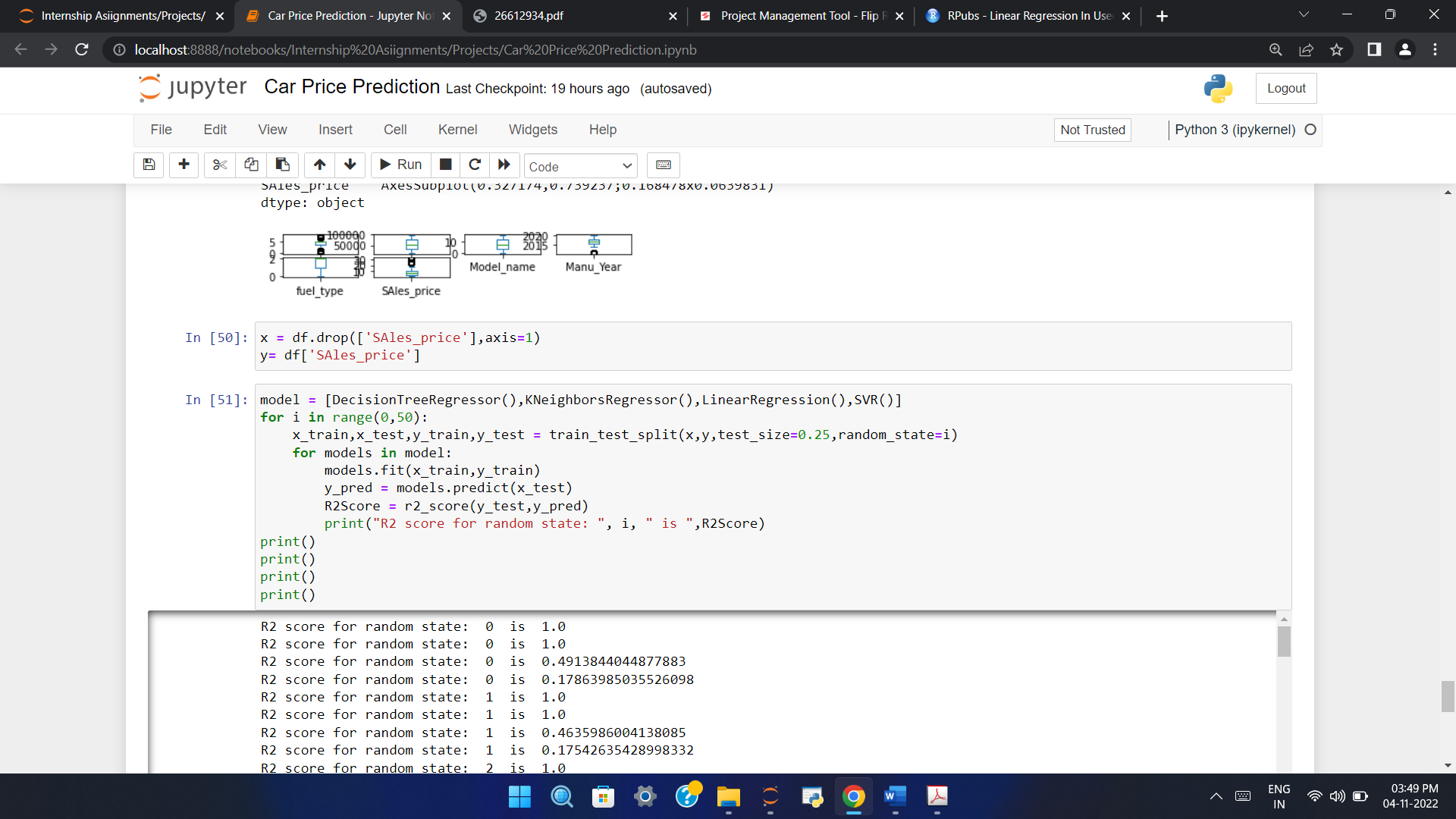
As target column is not a categorical column hence we can select linear regression or other regression models.

In this case we have used Decision Tree Regressor, SVR, KNeighbours Regressor and Linear Regression.

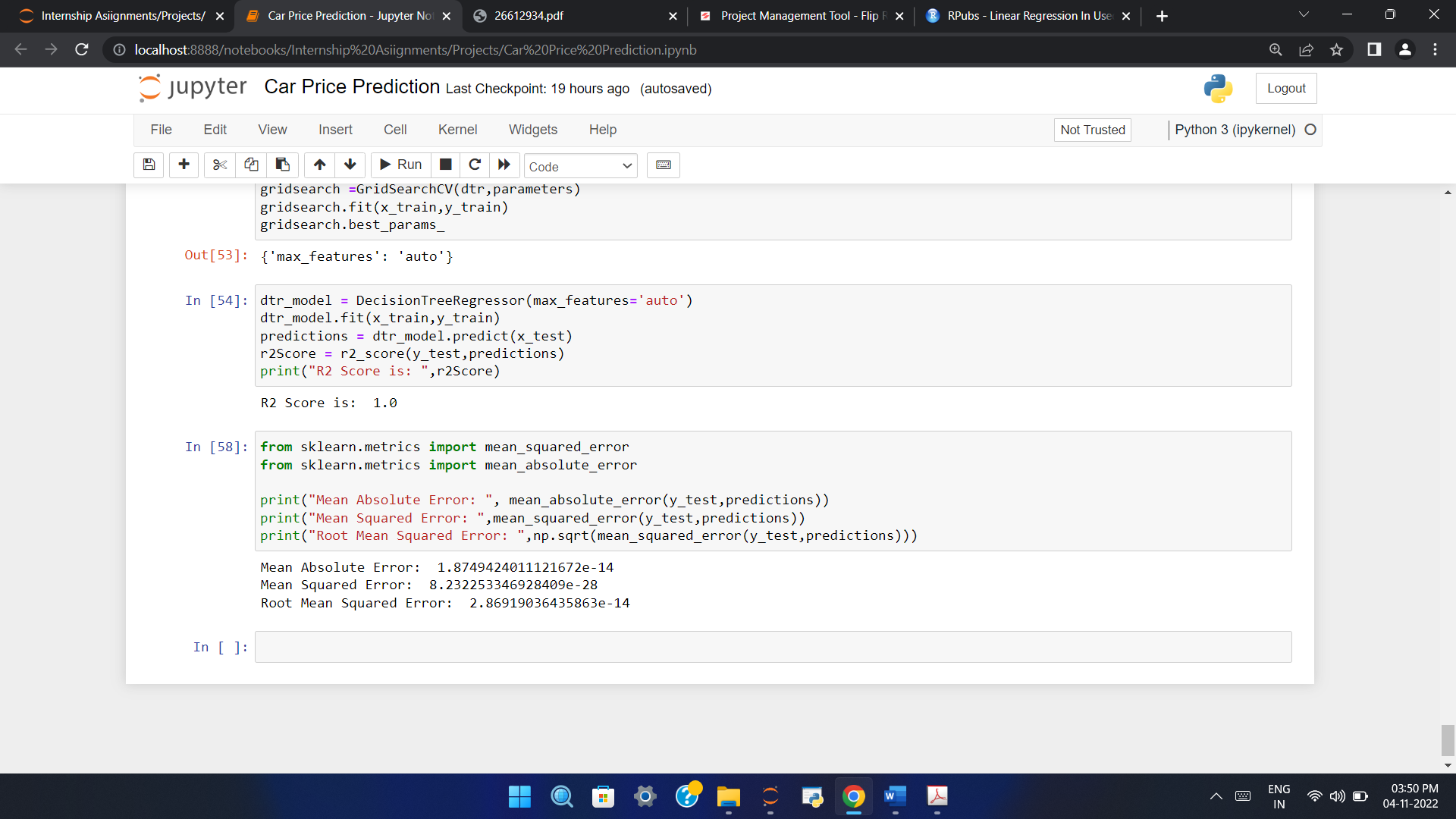
* Testing of Identified Approaches (Algorithms)

1. Decision Tree Regressor
2. SVR
3. KNeighbours Regressor
4. Linear Regression

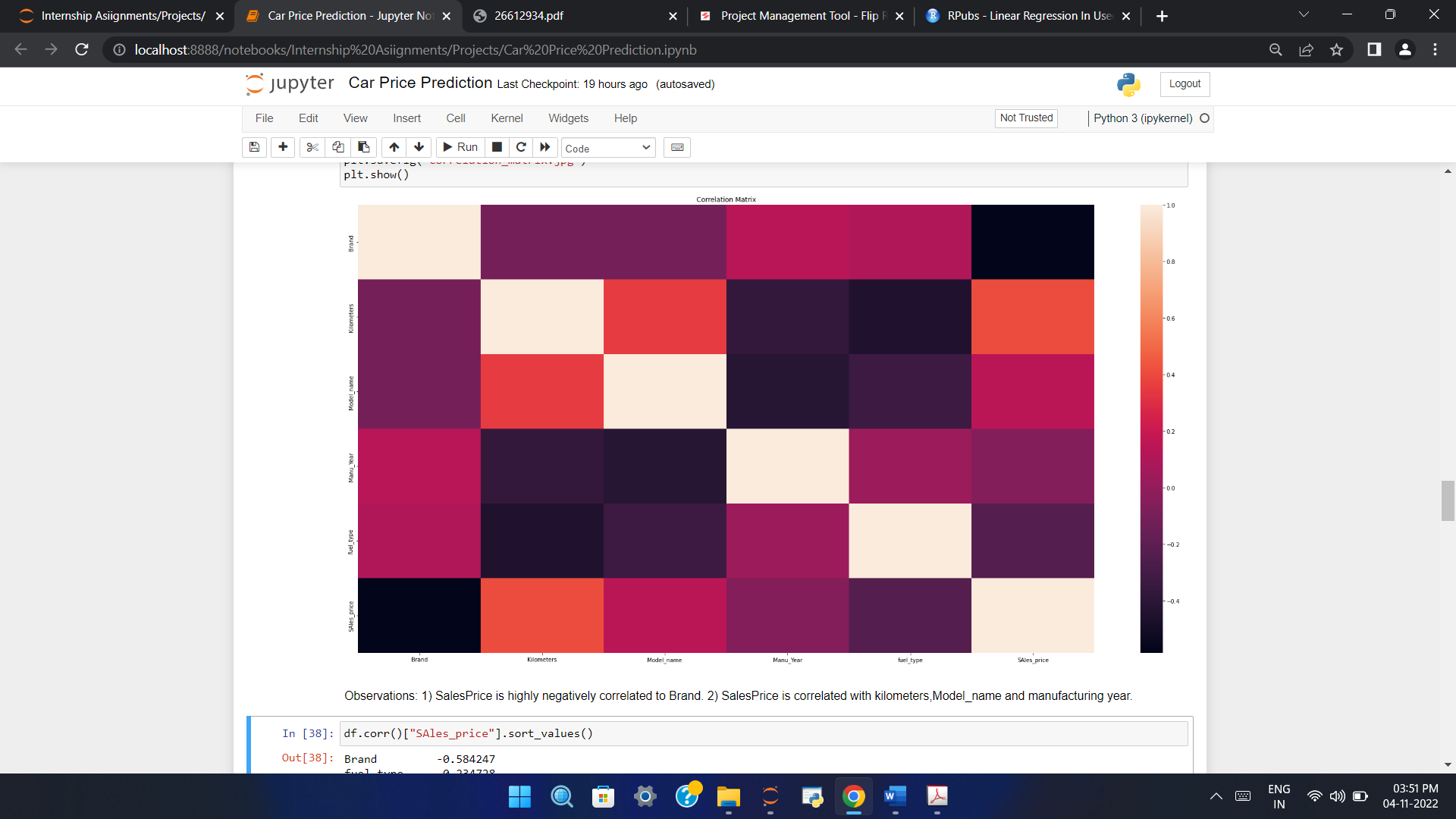
* Run and evaluate selected models

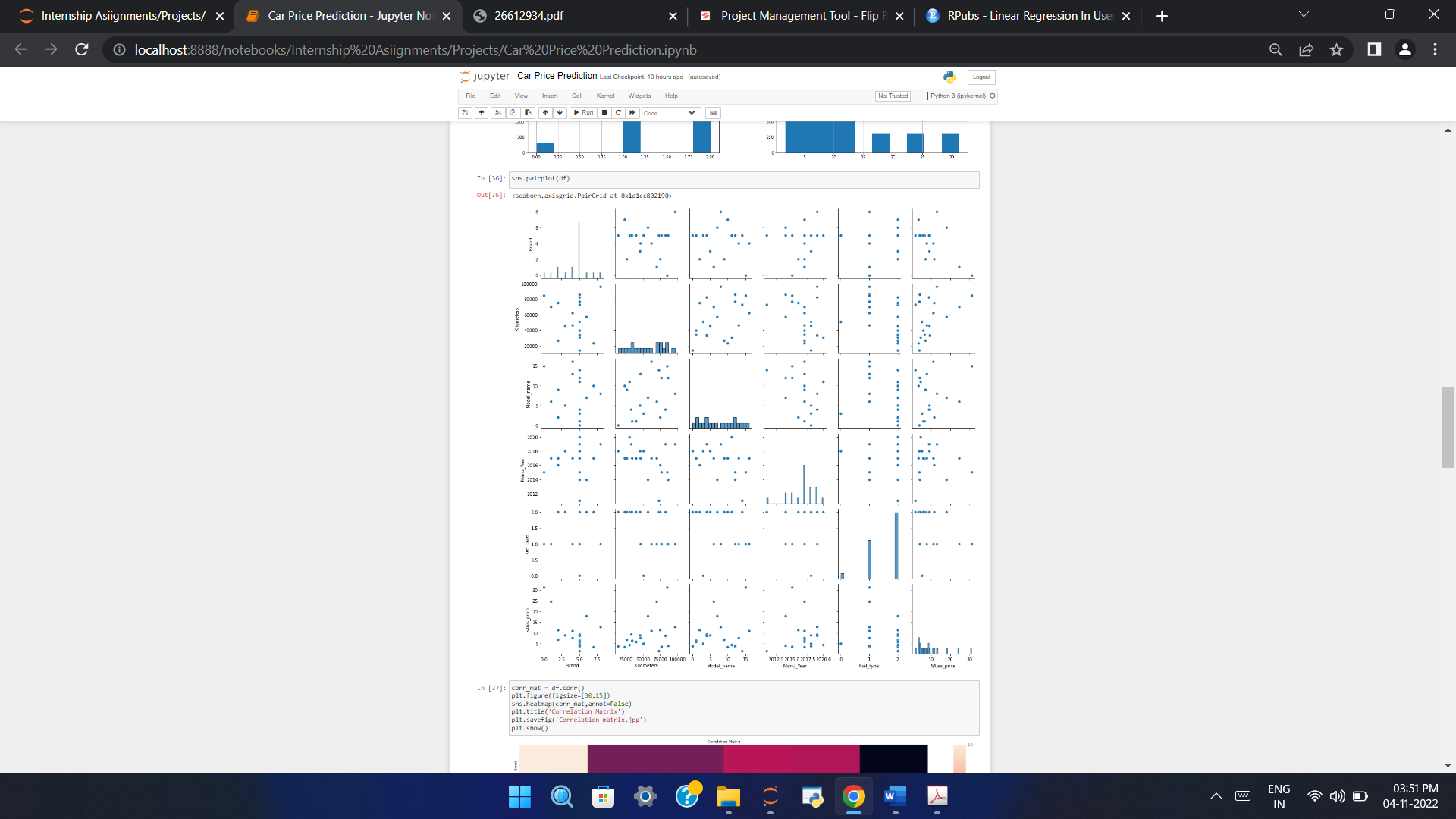


* Key Metrics for success in solving problem under consideration



* Visualizations





* Interpretation of the Results

1. Decision Tree Regressor model is giving best R2 score : 1.0

**CONCLUSION**

Decision Tree regressor Model is the best model out of all model tested above.