**Machine Learning Project:**

The purpose of the project is to predict median house values in Californian districts, given many features from these districts.

The project also aims at building a model of housing prices in California using the California census data. The data has metrics such as the population, median income, median housing price, and so on for each block group in California. This model should learn from the data and be able to predict the median housing price in any district, given all the other metrics.

Handle missing values Fill the missing values with “mean” of the respective column.

3 Encode categorical data Convert categorical column in the dataset to numerical data.

4 Split the dataset Split the data into 80% training dataset and 20% test dataset.

5 Standardize **data Standardize training and test datasets.** technique

6 **Perform Linear Regression** Perform Linear Regression **on training data**. Predict output for test dataset using the fitted model. Print root mean squared error (RMSE) from Linear Regression

7) **Perform Decision Tree Regression**

Perform Decision Tree Regression on training data. Predict output for test dataset using the fitted model. Print root mean squared error from Decision Tree Regression.

8) **Perform Random Forest Regression**

Perform Random Forest Regression on training data. Predict output for test dataset using the fitted model.

**Data Science with Python Project:**

used Python libraries such as NumPy, Pandas, scikit-learn, matplotlib, and so on.

 Perform a service request **data analysis of New York City 311 calls**. You will focus on the **data wrangling techniques to understand the pattern in the data** and also visualize the major complaint types.

* Import a 311 NYC service request
* Basic data exploratory analysis
  + Explore data
  + Find patterns
  + Display the complaint type and city together
* Find major complaint types
  + Find the top 10 complaint types
  + Plot a bar graph of count vs. complaint types
* Visualize the complaint types
  + Display the major complaint types and their count

**Deep Learning with Tensorflow project:**

Build a CNN model that classifies the given pet images correctly into dog and cat  images.   The project scope document specifies the requirements for the project “Pet  Classification Model Using CNN.” Apart from specifying the functional and nonfunctional  requirements for the project, it also serves as an input for project scoping.

**Data Science with R**

Project:

Analyze the Healthcare cost and Utilization in Wisconsin hospitals

A nationwide survey of hospital costs conducted by the US Agency for Healthcare consists of hospital records of inpatient samples. The given data is restricted to the city of Wisconsin and relates to patients in the age group 0-17 years. The agency wants to analyze the data to research on the healthcare costs and their utilization.

Project:

An education department in the US needs to analyze the factors that influence the admission of a student into a college.

Analyze the historical data and determine the key drivers. Analysis information:

Predictive • Run logistic model to determine the factors that influence the admission process of a student (Drop insignificant variables) • Transform variables to factors wherever required • Calculate accuracy of the model • Try other modeling techniques like decision tree and SVM and select a champion model • Determine the accuracy rates for each model • Select the most accurate model • Identify other Machine learning or statistical techniques that can be used