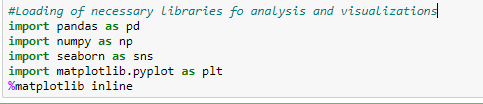
Project ID:2019\_12\_4039 EG\_2020\_3037| 07/01/2020

Problem: To develop a Machine Learning model for predicting crop based on the weather variables using KNN. Merge the two data sets based on conditions you will prepare the predicative analysis for crop & season, City by applying KNN algorithm with visualized graphs. Finally conclude with other algo’s like Gradient Descent Regression, Elastic Net Regression, Ridge Regression, produce your output compare the above 3 algo’s

Solution

* Loading of necessaries libraries



Pandas and Numpy -> For Data Pre-Processing and Data Analysis

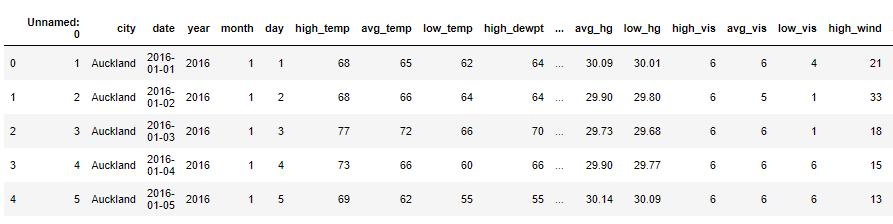
Matplotlib and Seaborn -> For Data Visualization

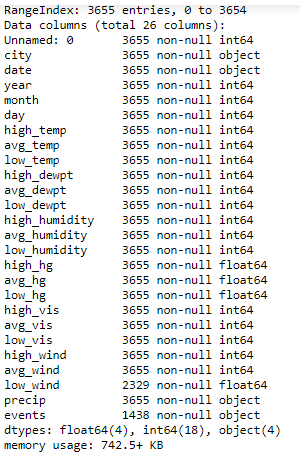
* Data Pre-Processing like loading the datasets

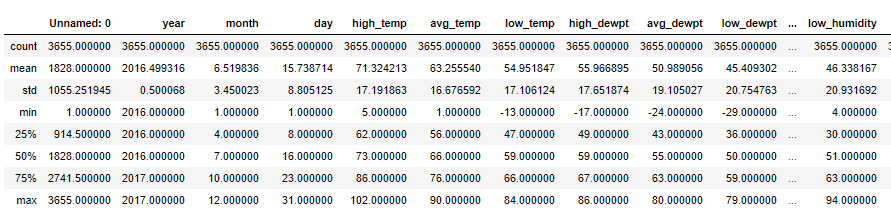


* Datasets info:

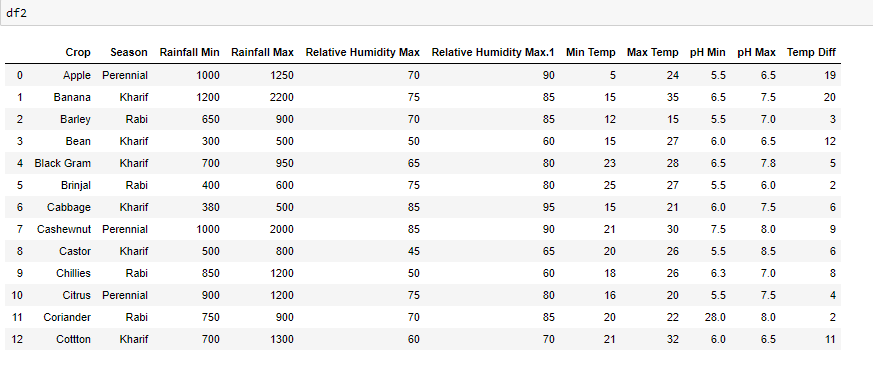
Weather Data:

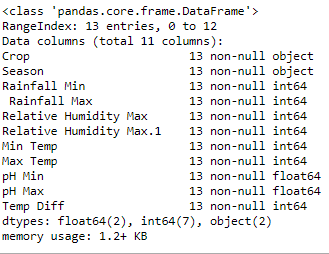


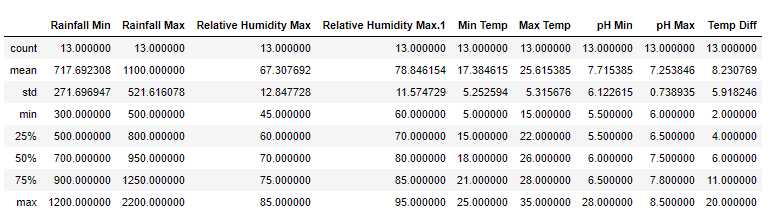


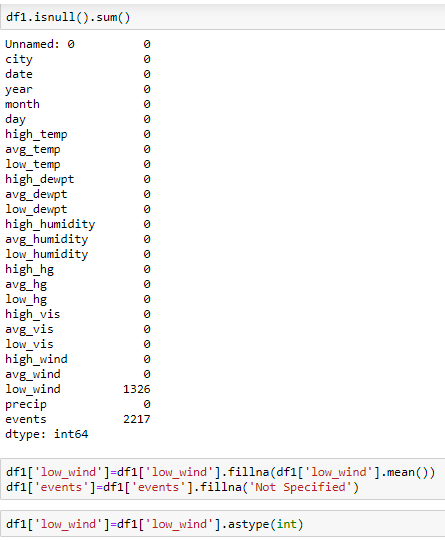


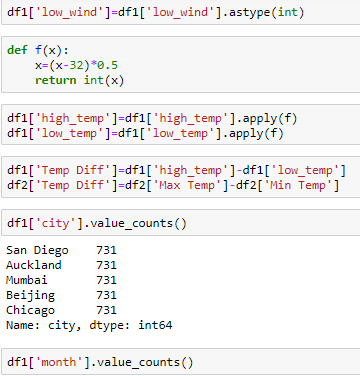
Crop Data:

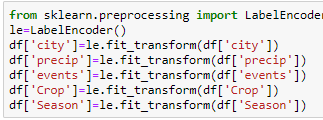






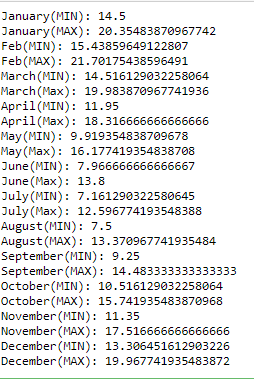
* Data Cleaning and modification of both datasets:
* 

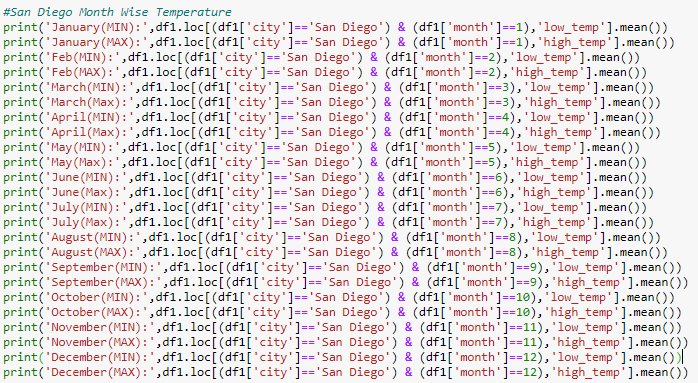


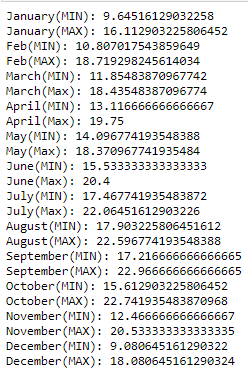


* Data Analysis of Weather and Crop datasets and build a condition on which they can be merged
* We have to analyse the variables of weather dataset and find out suitable crop for that location and for that particular time. We analyse the maximum and minimum temperature of all location month wise and tries to match it with rainfall requirement for each crop. Location and time satisfy the condition can produce that crop

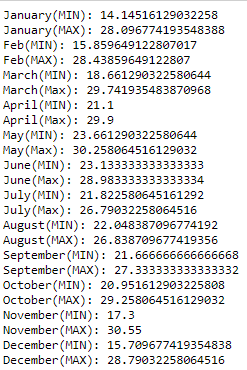


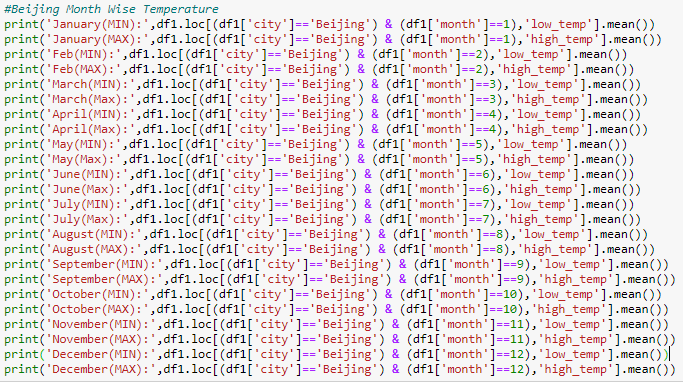


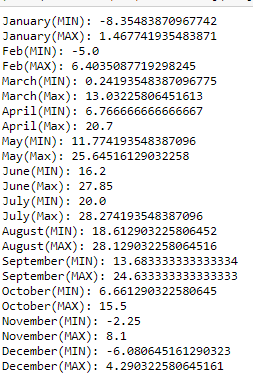


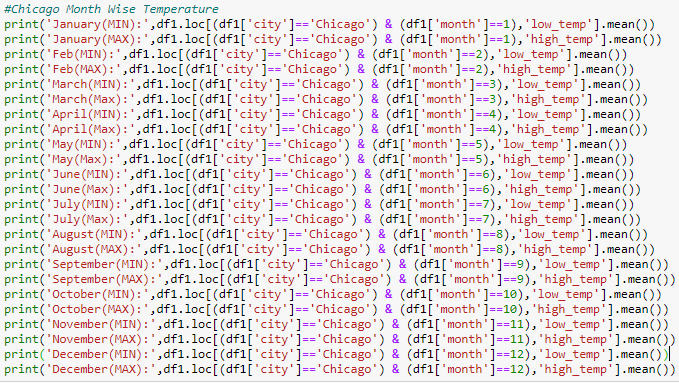


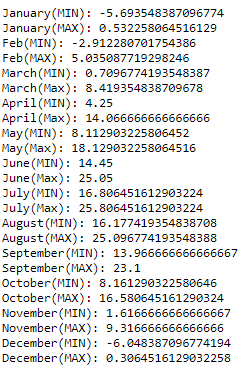




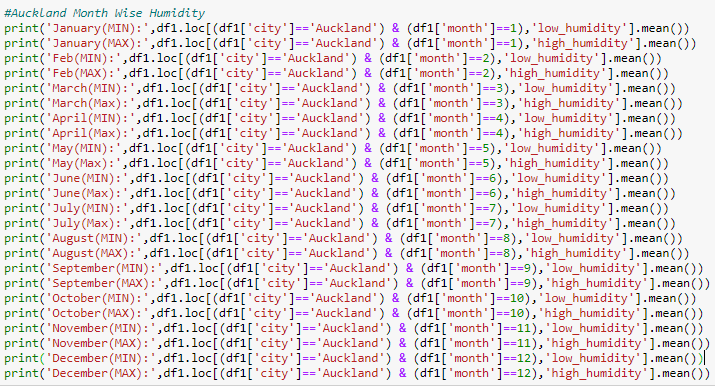


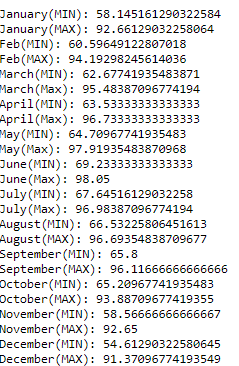


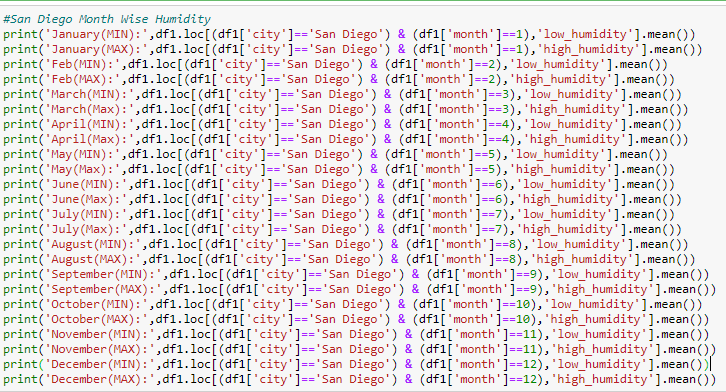


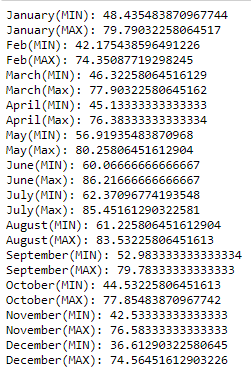


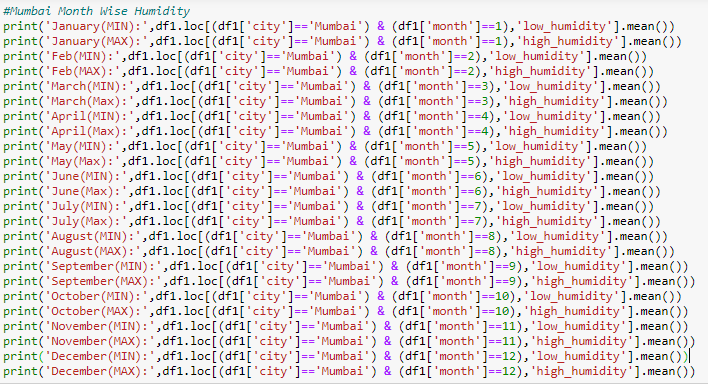
* After analysing the minimum and maximum temperature of city month wise we compared it to crop’s temperature requirement and shortlist crop basis on that. We further filter out crop by comparing the humidity of regions and crops as to get better results

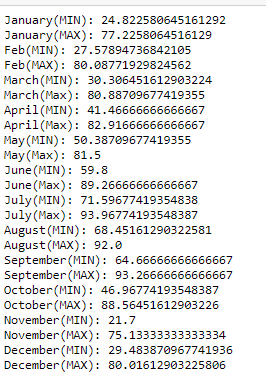


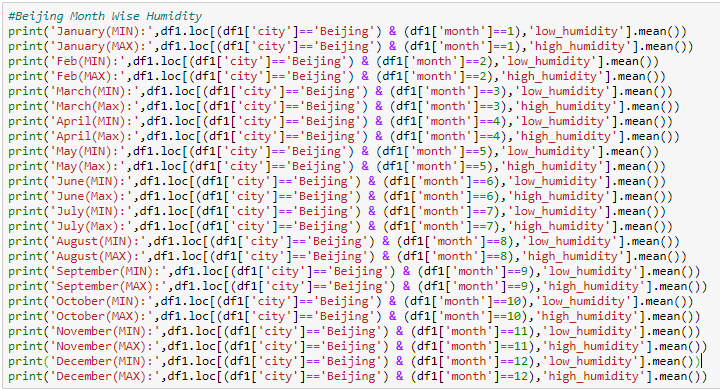


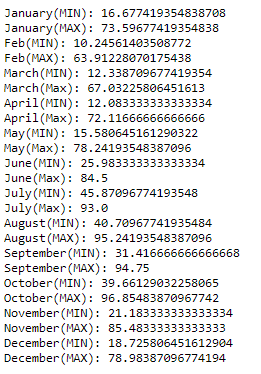


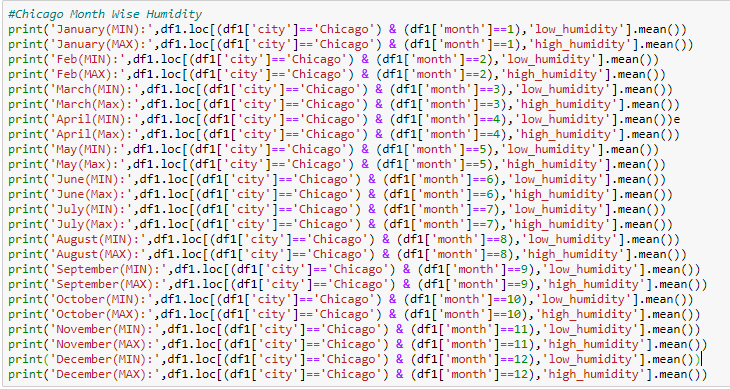


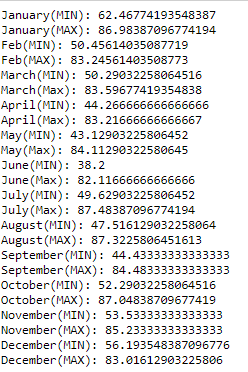










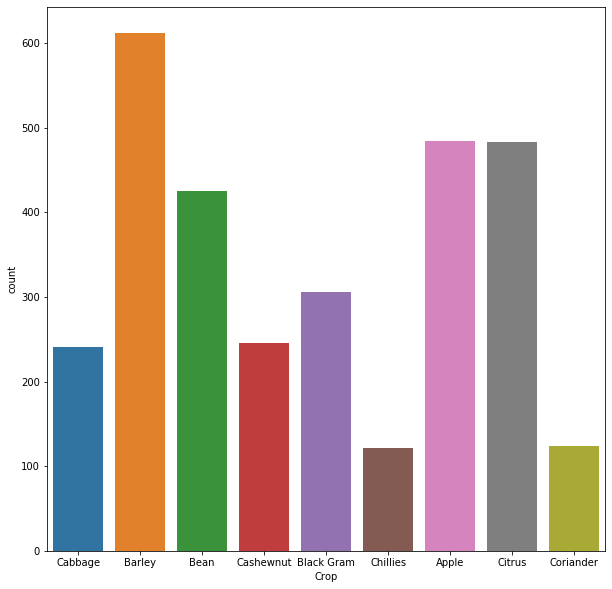


* After analysing the humidity and temperature we assigned the crops to each city depending on the month
* Merging of datasets based on common column (Crop)



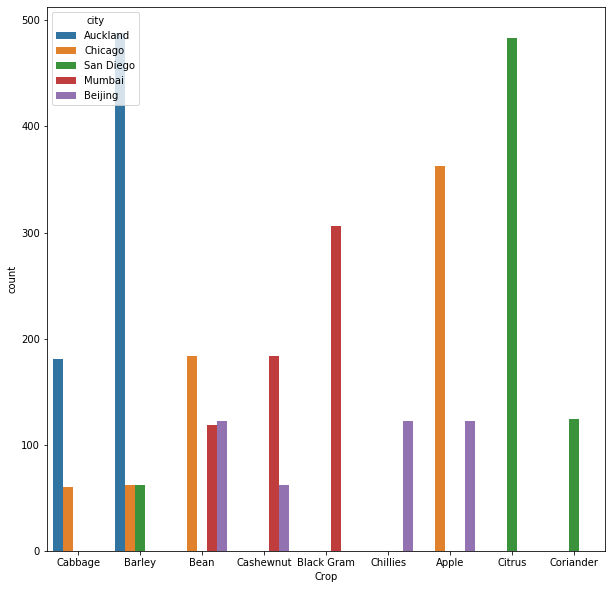
* Data Visualization



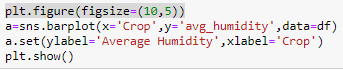


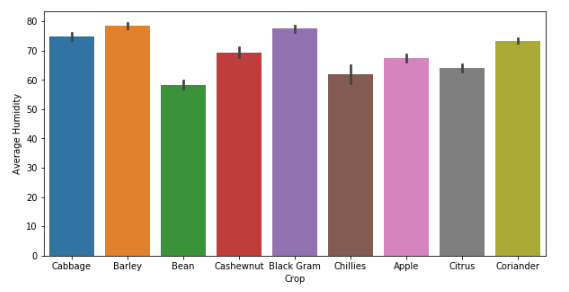
Barley is produced the most of all crops. Banana is not in plot as it none of region has suitable condition to produce it



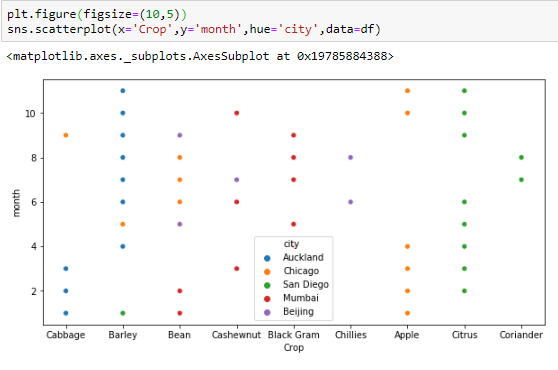


This plot which crop produced in which region. Some of the crops are grown in only one region due to climate condition.

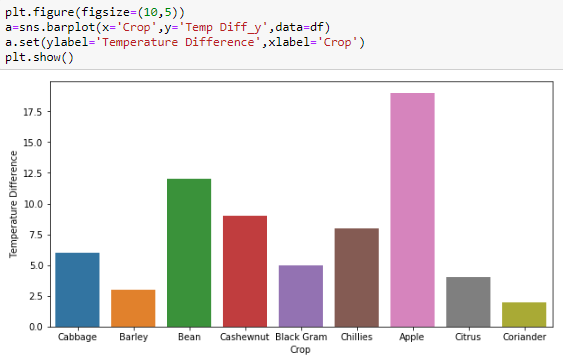




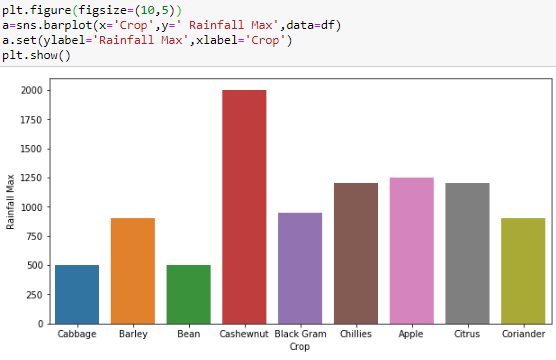
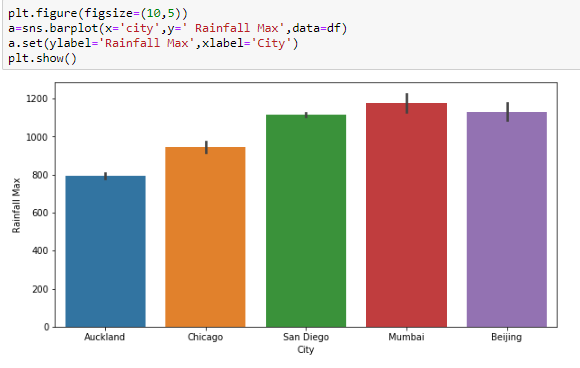
Barley requires high level of humidity along with Black Gram. Hence they are produced in regions like Mumbai and Auckland where rainfall is high



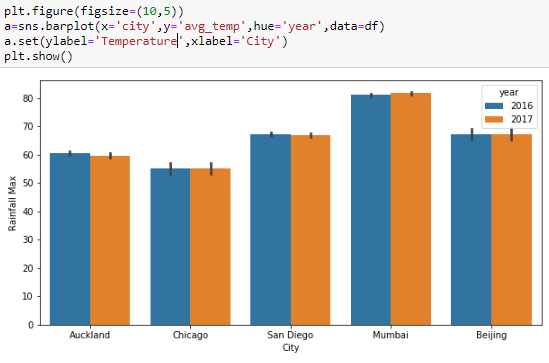
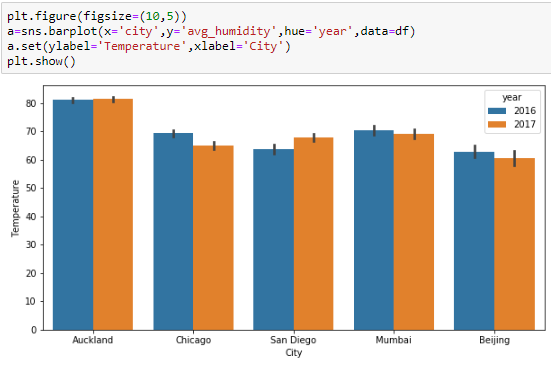
This plot specifies in which month of all region which specific crops are produced

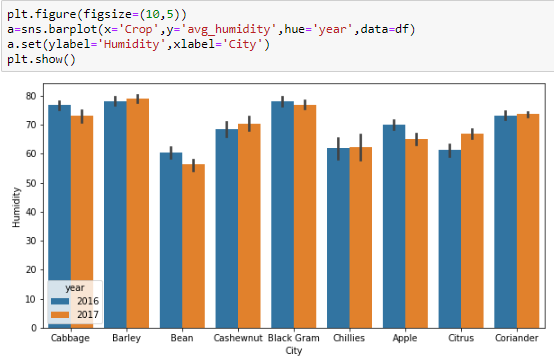
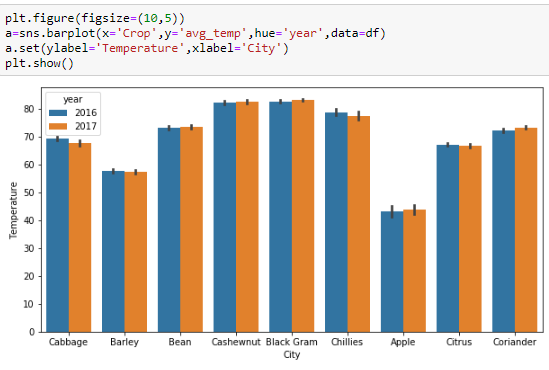


This plot specifies how much difference between and max and min temperature of crops as it specifies their adaptation. Apple has wide range of temperature

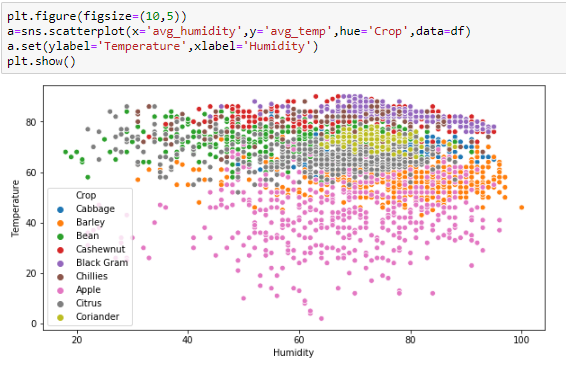
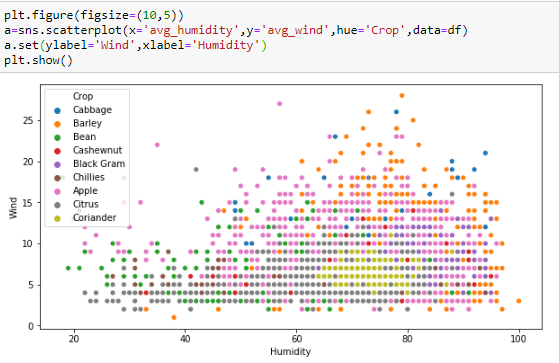
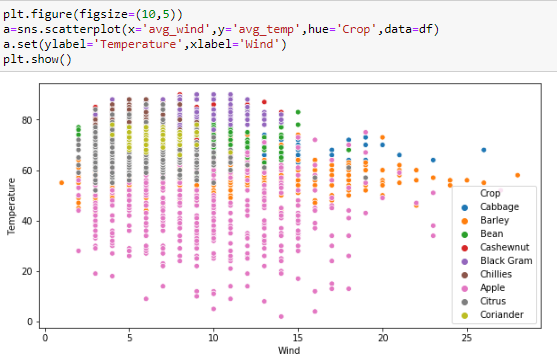
 

Both plots shows rainfall required by crop and rainfall provided by city

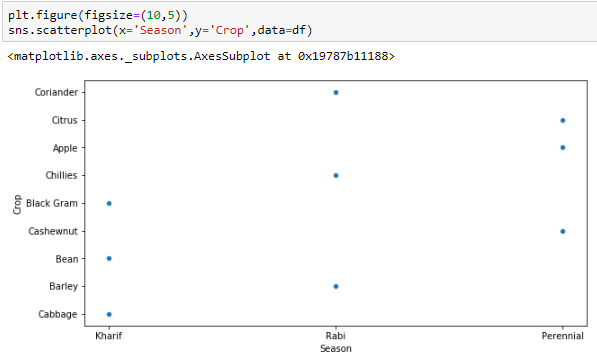
 Both plots shows there is very minor change in rainfall and humidity from each city from both years (2016-2017)

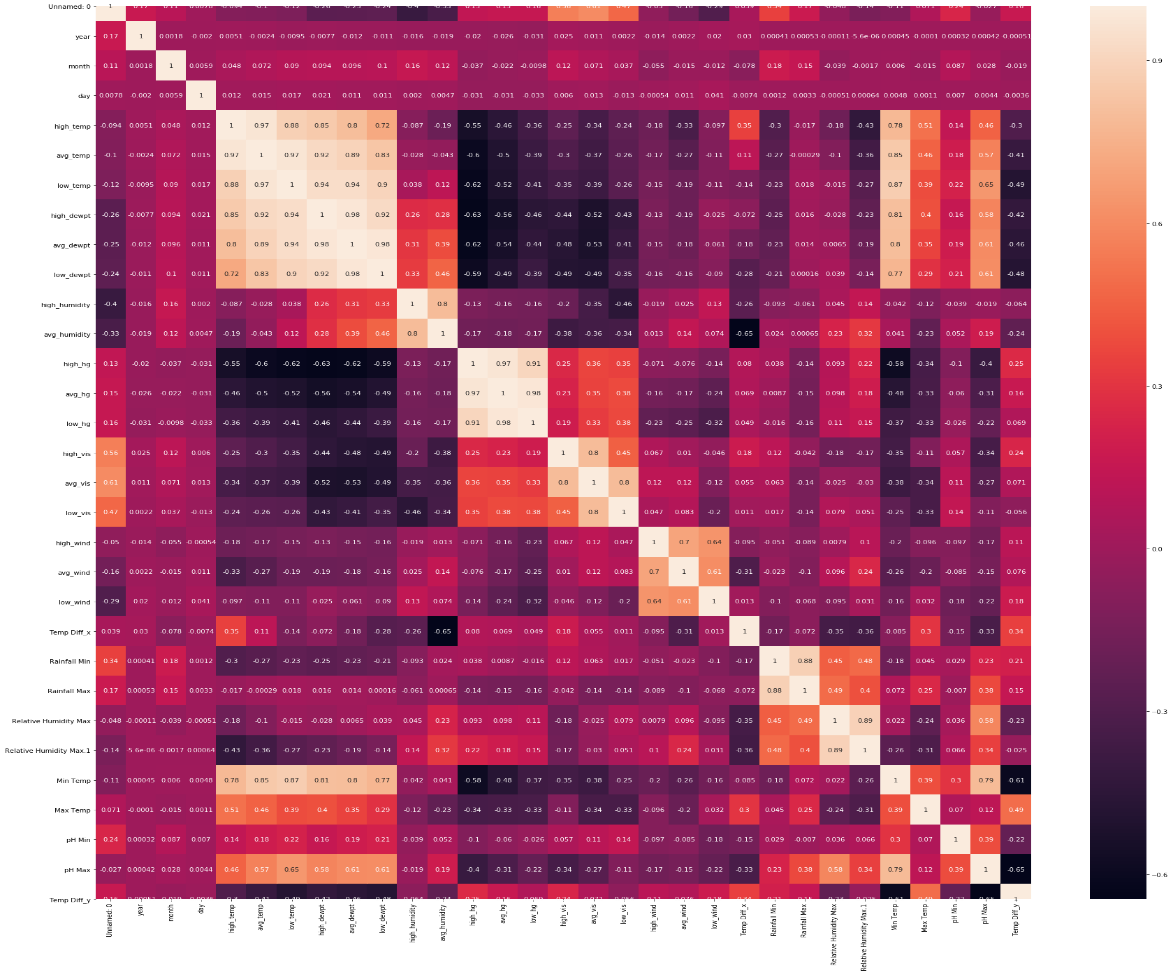
Both plots shows there is very minor change in rainfall and humidity from each crop from both years (2016-2017)

This 3 plots shows correlation between 3 factors such as humidity, temperature and wind. Plots shows how much each increases and decreases on regulating the other factor and how it affects on each crop



This plot specifies different crops produced in different season

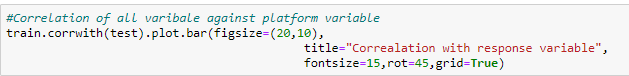


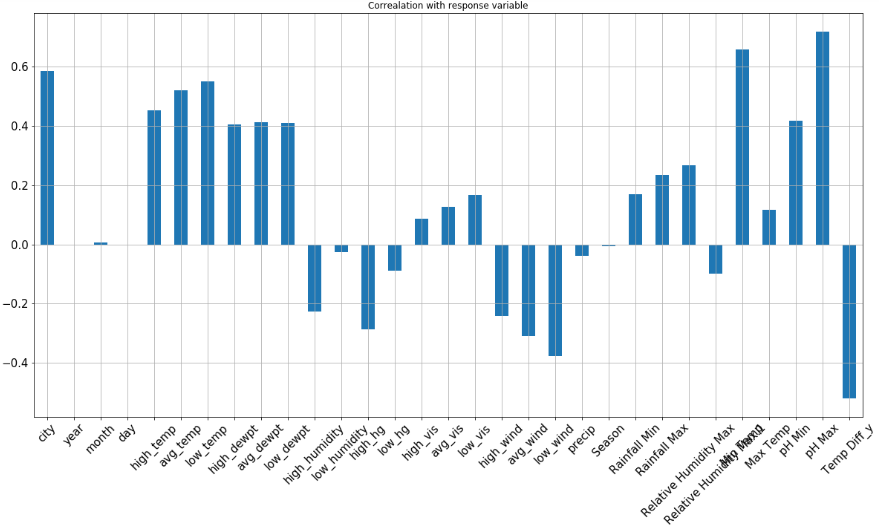
Correlation matrix specifying how much variables are correlated with each other

* Splitting the data into train data which contain all columns except the response variable and test data which contain the response variable



* Correlation of columns with response variable

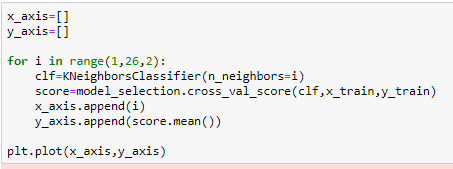


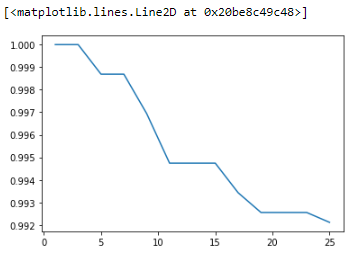






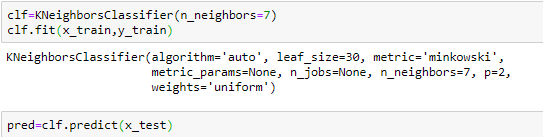
* Implementing the KNN machine learning model
* Finding optimal K value



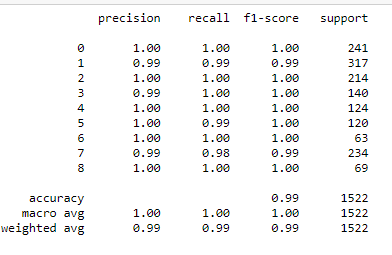


Optimal Value of K is 7 for current data

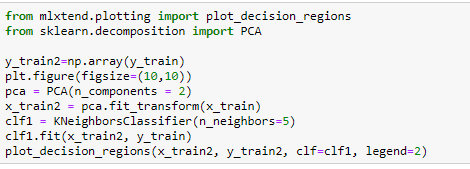
* Implementing the KNN model with 7 K value

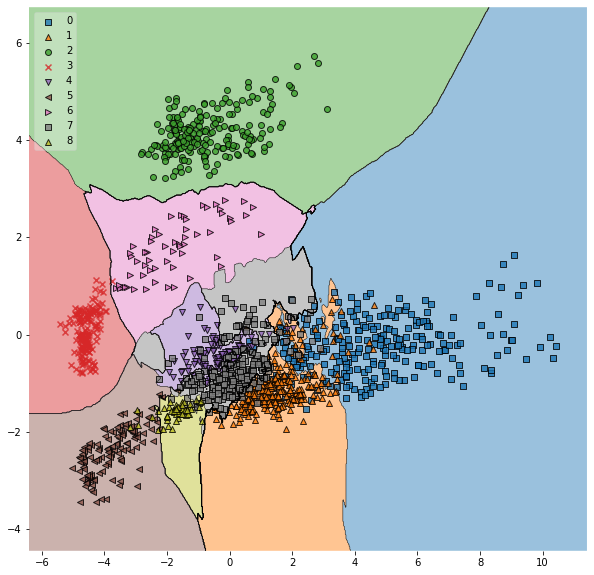


* Evaluating the model

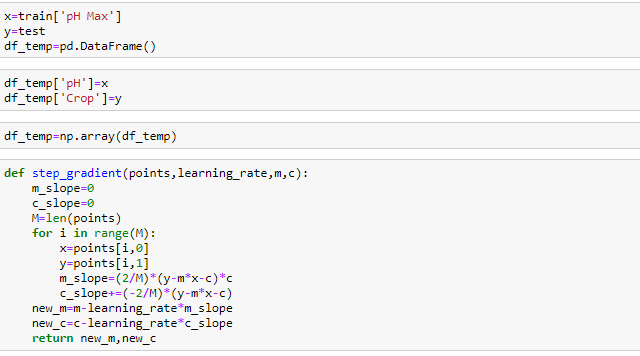


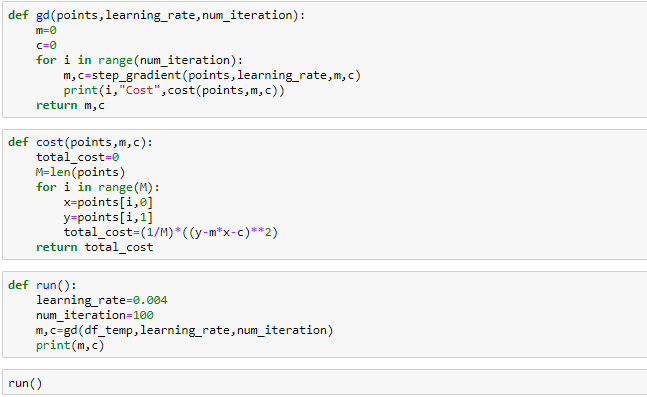
Model is giving nearly perfect results because test data is a subset of train data and model learns very well from train and able to predict test data very accurately

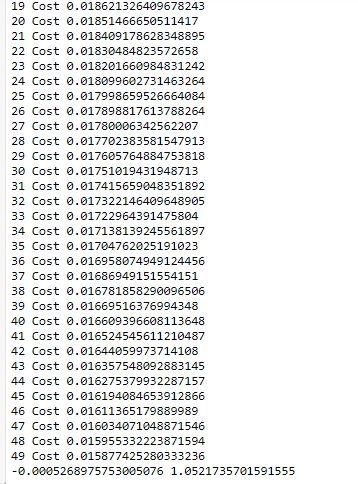




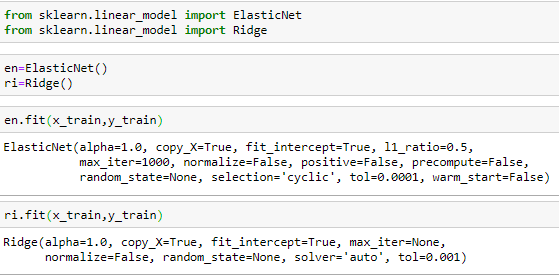
* Implementing Gradient Descenet Regression, Elastic Net Regression, Ridge Regression
* Gradient Descent Regression





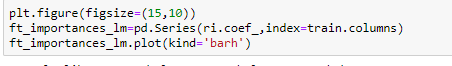


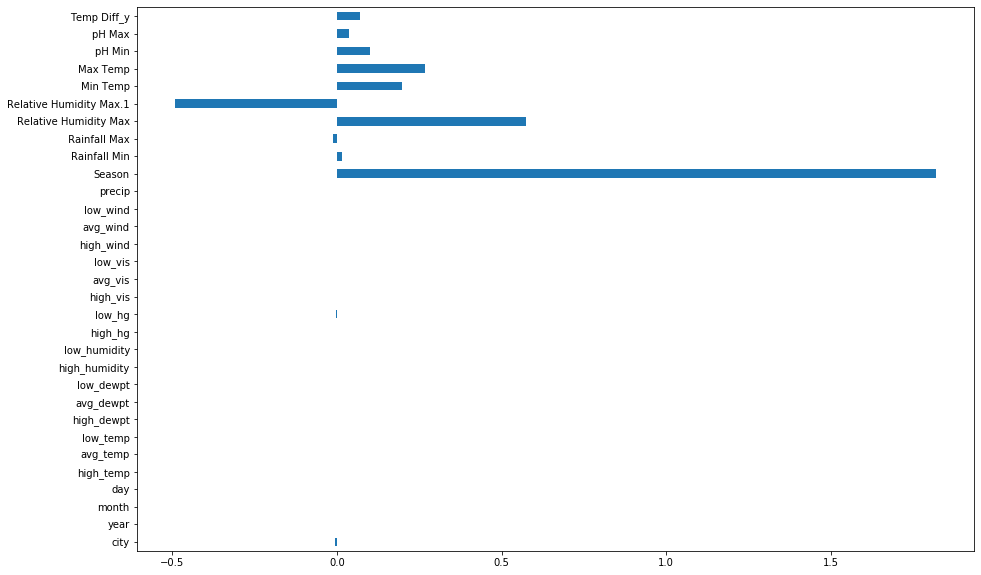
* Ridge and Elastic Net Regression

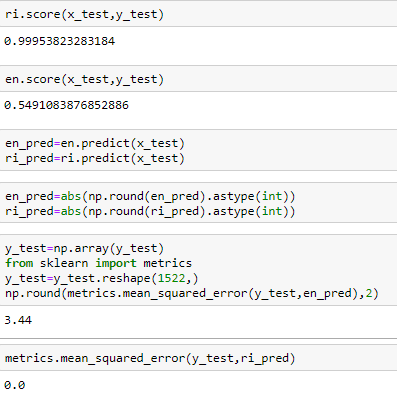












|  |  |  |  |
| --- | --- | --- | --- |
|  | Gradient Descent Regression | Elastic Net Regression | Ridge Regression |
| Cost | 0.015877425280333236 | 3.44 | 0.000000001 |