This document summarizes the approach used in performing the prediction tasks, communicating the observations.

1. Data Preprocessing and feature engineering:

The excel file was converted into CSV and preprocessed using Pandas data science library, in order to generate the data frames of the from the CSV file. Further, feature engineering was performed on the date-time column of the data-frame and obtaining following attributes “**'Month','weekoftheyear','Dayofyear','Dayinmonth', 'Weekday', 'Hours', 'Quarterinhour', 'total\_calls'**”.

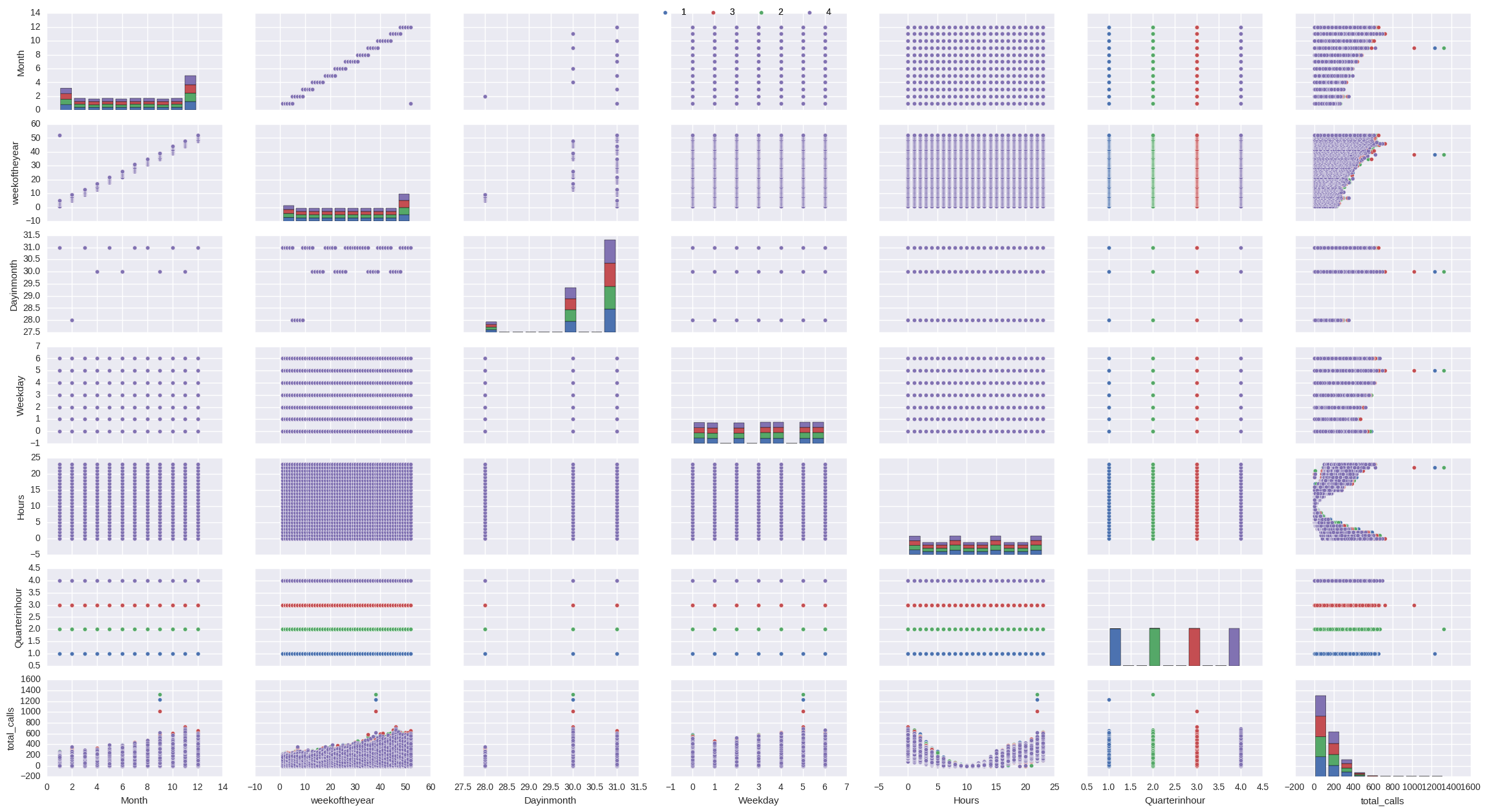
1. The entire dataset was used for data visualizations using Seaborn library. 

Figure 1 pair plot of the feature

The Figure 1 shows the pair plots obtained for seven features of the dataset. The plot of “total\_calls” with other feature are of more interest. The Figure 2 shows the pair plot of the total\_calls attribute with other attributes. In Figure 2 we can observe “weekoftheyear” and “Hours” attributes are more informative than other pair plots. Figure 3 shows the heat map of the seven attributes depicting higher correlation between attribute “total\_calls” and “Month”, “weekoftheyear” and “Hours”, and lesser correlation between the remaining three attributes. Figure 4 shows the plot of paired attributed along with their best fit line obtained. In Figure 4 we can observe that the best fit was obtained for the pair of “total\_calls” with “Month” and “weekoftheyear”.

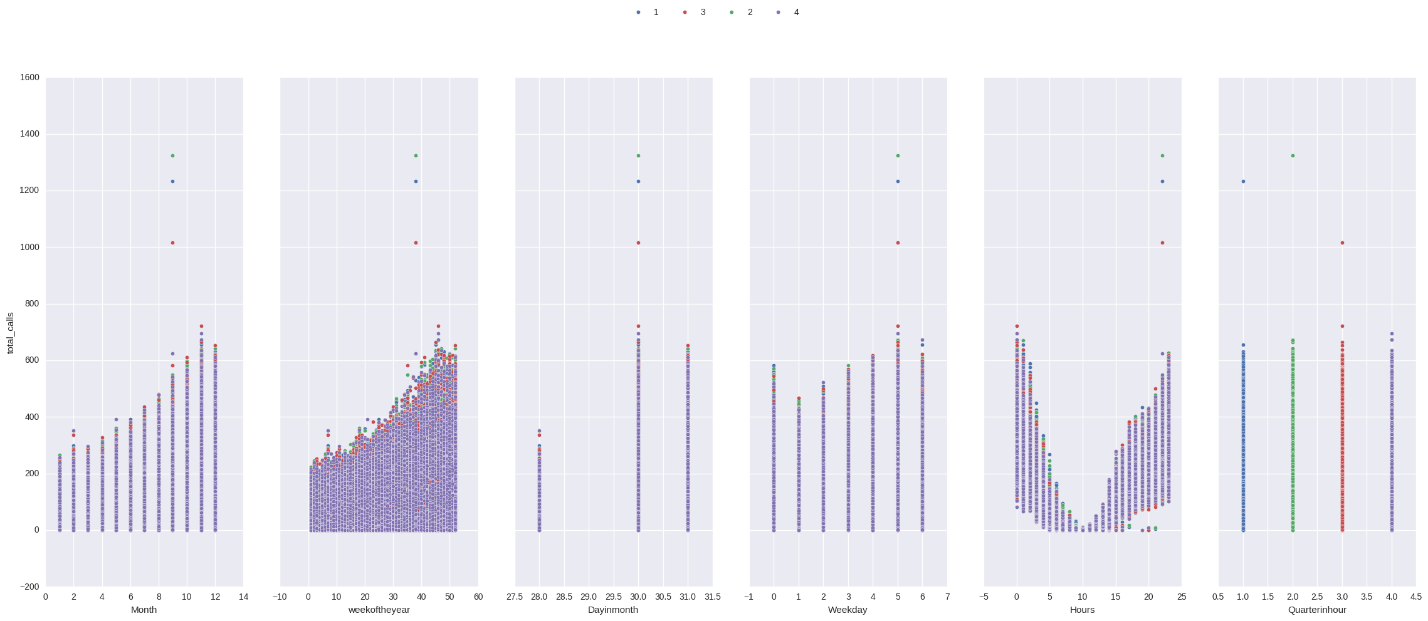


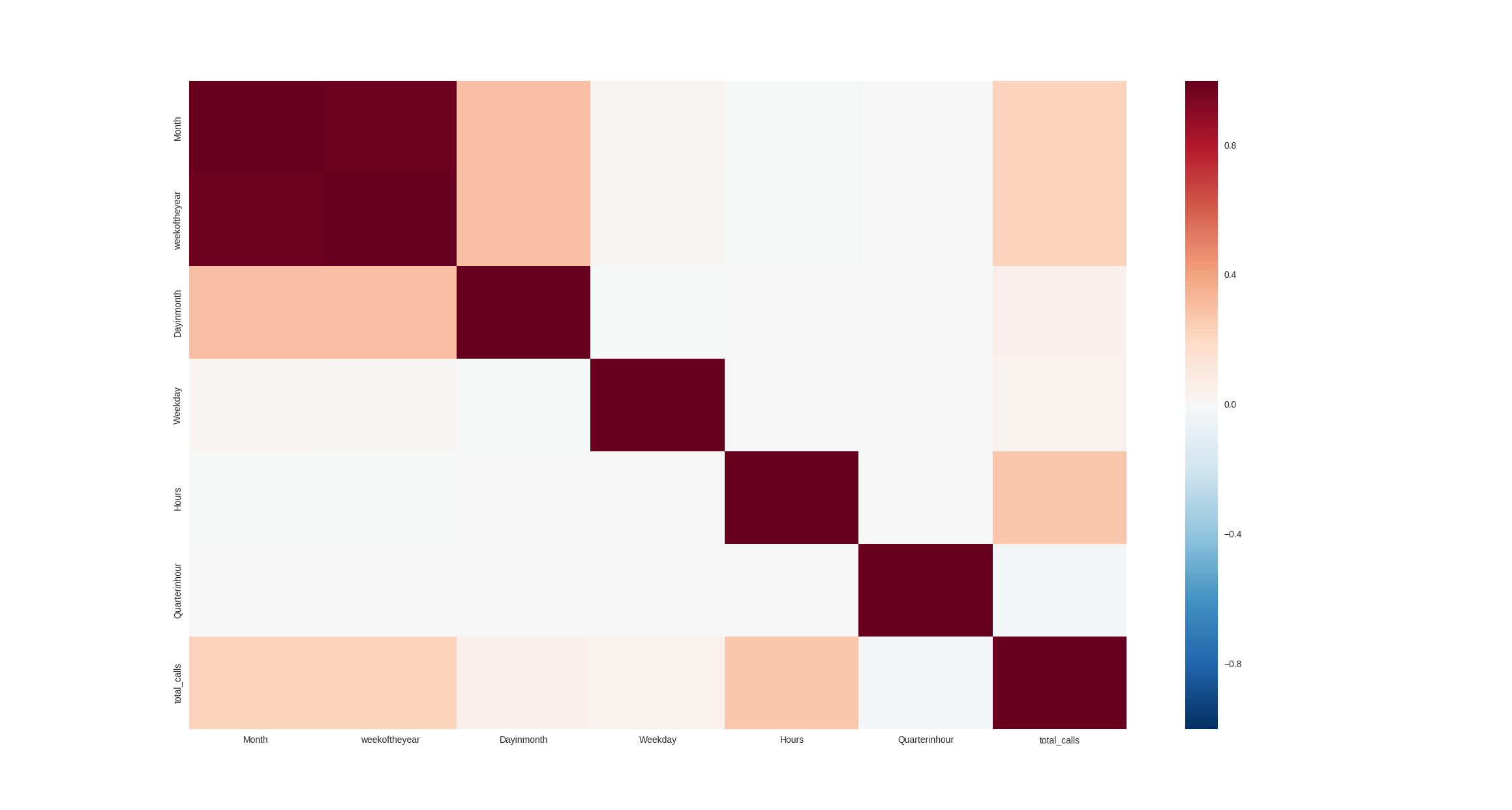
Figure 2 Pair plot of total\_calls 

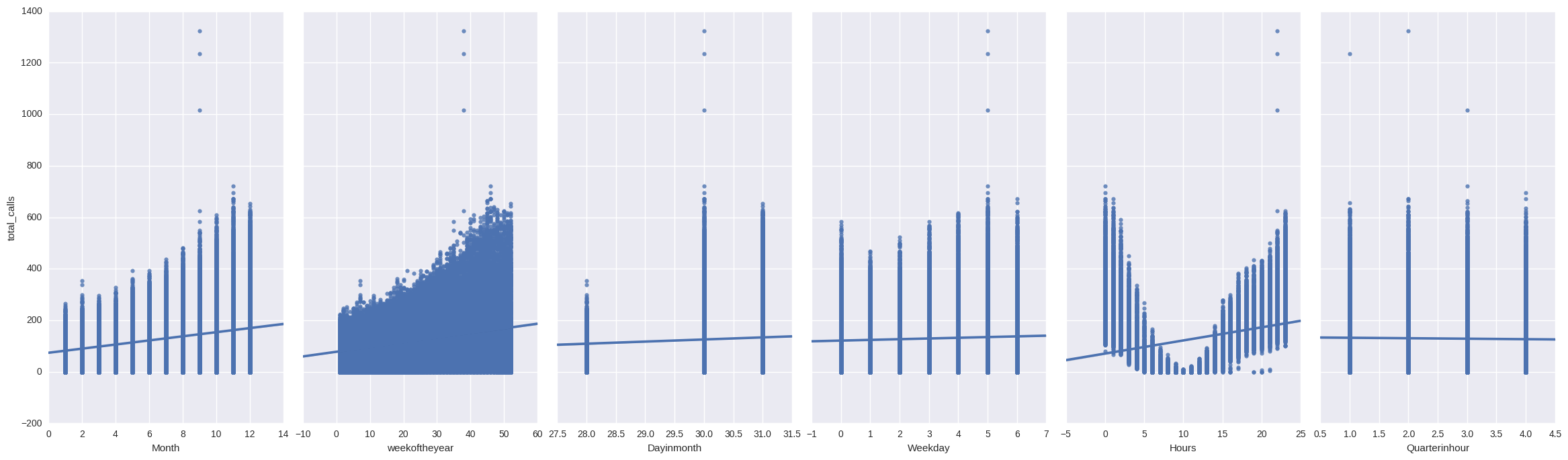
Figure 3 Heat map of seven attributes

Figure 4 Linear regression of paired attributes

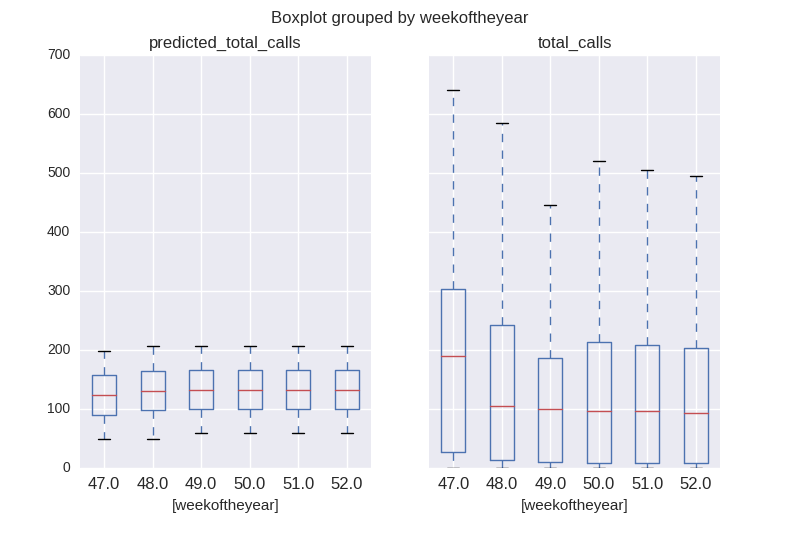
1. The attributes “weekofday” and “Month” are important but not relevant for the current task because the data is for one year and the predictions are to be made of last 6 weeks of the year based upon the historical 46 weeks of the year. Therefore, the attribute “Month” and “weekoftheyear” are left out of prediction task.
2. For prediction only four attributes were used “Dayinmont”, “Weekday”, “Hour”, “Quarterinhour”:
3. Observation with attributes “Dayinmont”, “Weekday”, “Hour” and “Quarter

Figure 5 Observations with Dayinmont”, “Weekday”, “Hour” and “Quarterinhour attributes

Prediction with attributes : ['Dayinmonth', 'Weekday', 'Hours', 'Quarterinhour']

intercept = -219.558489

coefficent : [ [ 9.17952931 2.63340849 5.54954442 -1.82397774] ]

Root Mean Square Error: 148.162767

Mean Absolute Error: 117.188654

Mean Square Error: 21952.205463

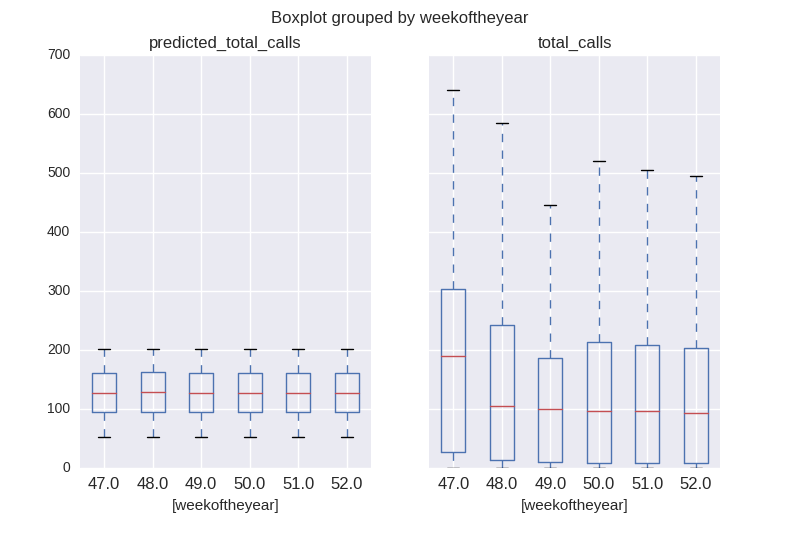


Figure 6 Observations with attributes Weekday, Hours, and Quarterinhour

Prediction with attributes : ['Weekday', 'Hours', 'Quarterinhour']

intercept = 59.804056

coefficent : [ [ 2.533631 5.54956009 -1.82397774] ]

Root Mean Square Error: 148.094189

Mean Absolute Error: 115.848343

Mean Square Error: 21931.888740

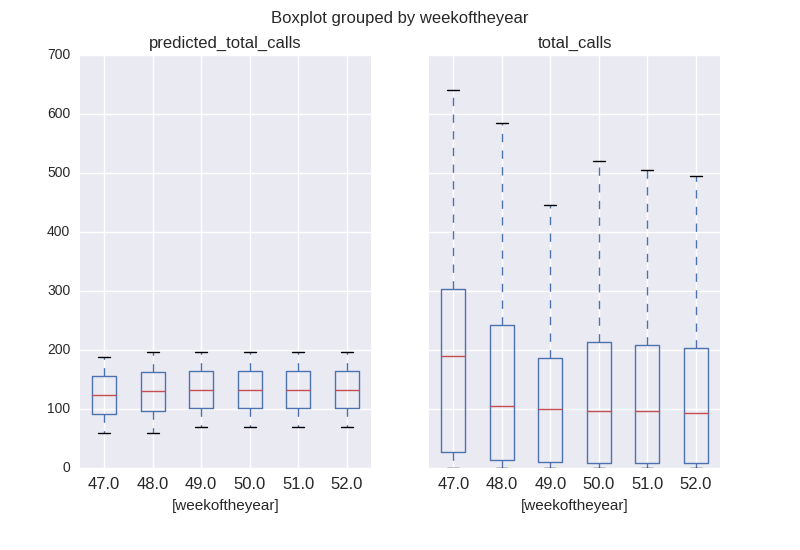


Figure 7 Observations with attributes Dayinmonth and Hours

Prediction with attributes : ['Dayinmonth', 'Hours']

intercept = -211.649721

coefficent : [ [ 9.02924504 5.549566 ] ]

Root Mean Square Error: 148.321260

Mean Absolute Error: 117.135713

Mean Square Error: 21999.196180

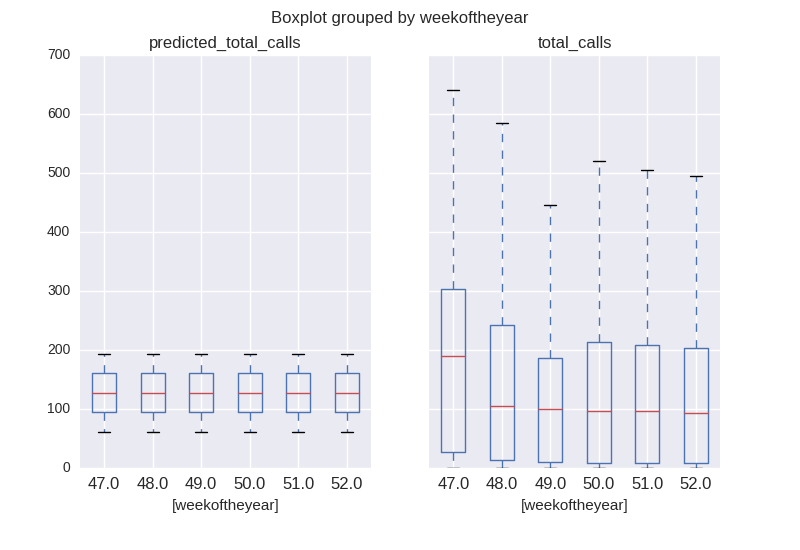


Figure 8 Observation with attributes Hours and Quarteinhour

Prediction with attributes : ['Hours', 'Quarterinhour']

intercept = 67.404713

coefficent : [ [ 5.54958061 -1.82397774] ]

Root Mean Square Error: 148.256826

Mean Absolute Error: 115.798845

Mean Square Error: 21980.086344

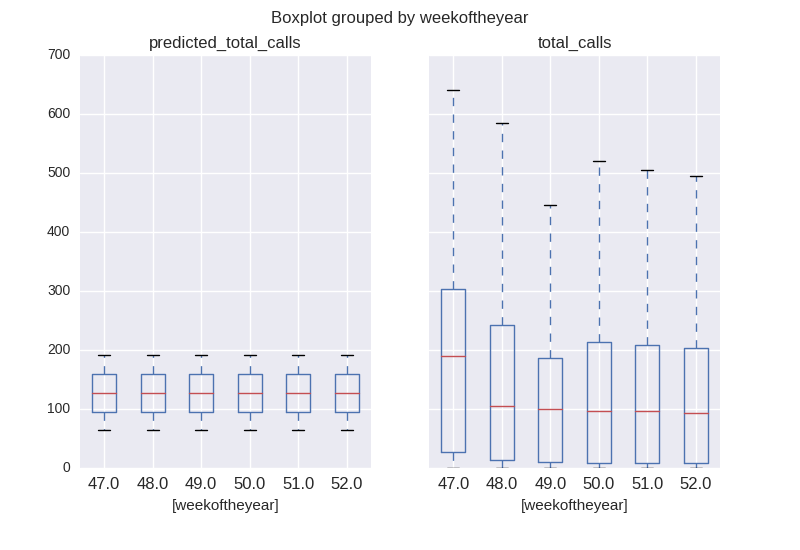


Figure 9 Observation with attribute Hour

Prediction with attributes : ['Hours']

intercept = 62.844769

coefficent : [ [ 5.54958061] ]

Root Mean Square Error: 148.273350

Mean Absolute Error: 115.818698

Mean Square Error: 21984.986250

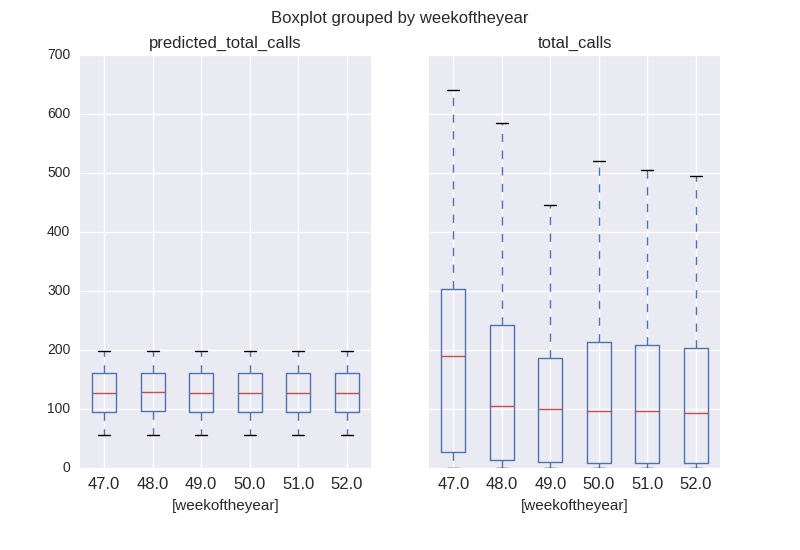


Figure 10 Observation with attributes Weekday and Hours

Prediction with attributes : ['Weekday', 'Hours']

intercept = 55.244112

coefficent : [ [ 2.533631 5.54956009] ]

Root Mean Square Error: 148.110731

Mean Absolute Error: 115.865856

Mean Square Error: 21936.788646

In all the experiments the Root Mean Square Error is 148 (ca.), Mean Absolute Error is 115 (ca.) and Mean Square Error is 21900 (ca.). This depicts that generated model does not predict well. This inference can also be observed in the Box-plots where the means are similar in all the cases with a huge variance in the prediction values.