Forecasting Footfalls for a National Park Based on weather

Business Forecasting project

Chinmayee Thakur

Riya Joy

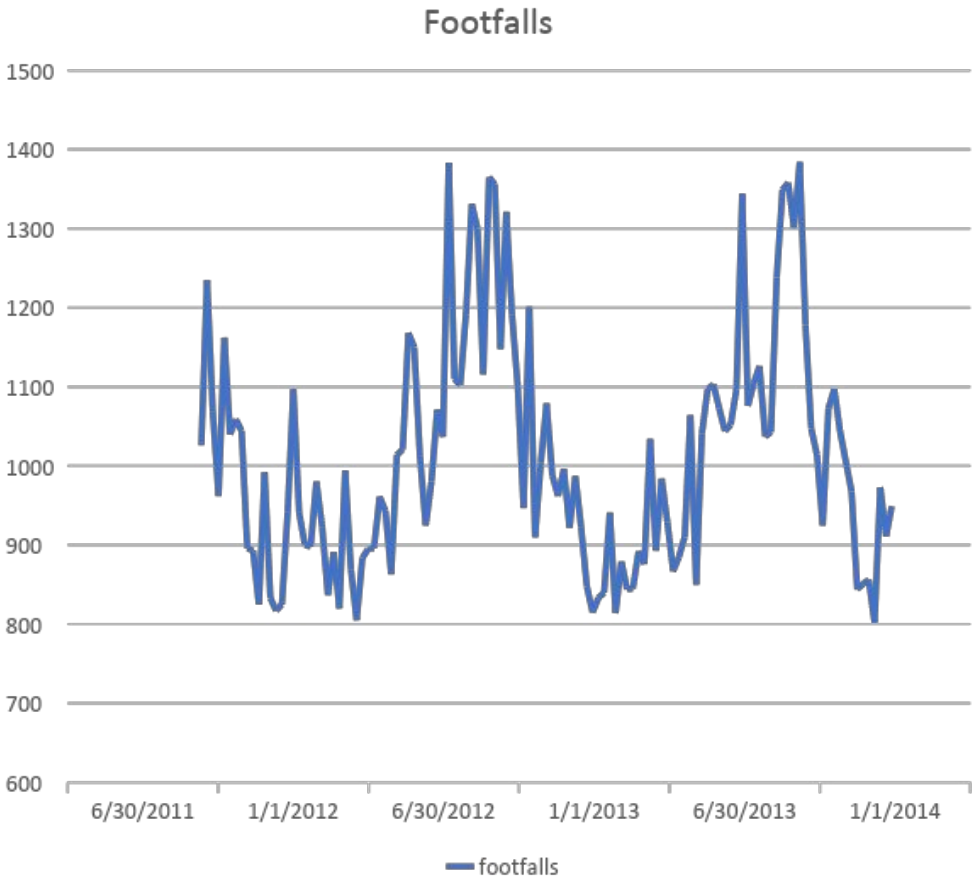
Jash Shah

WHY?

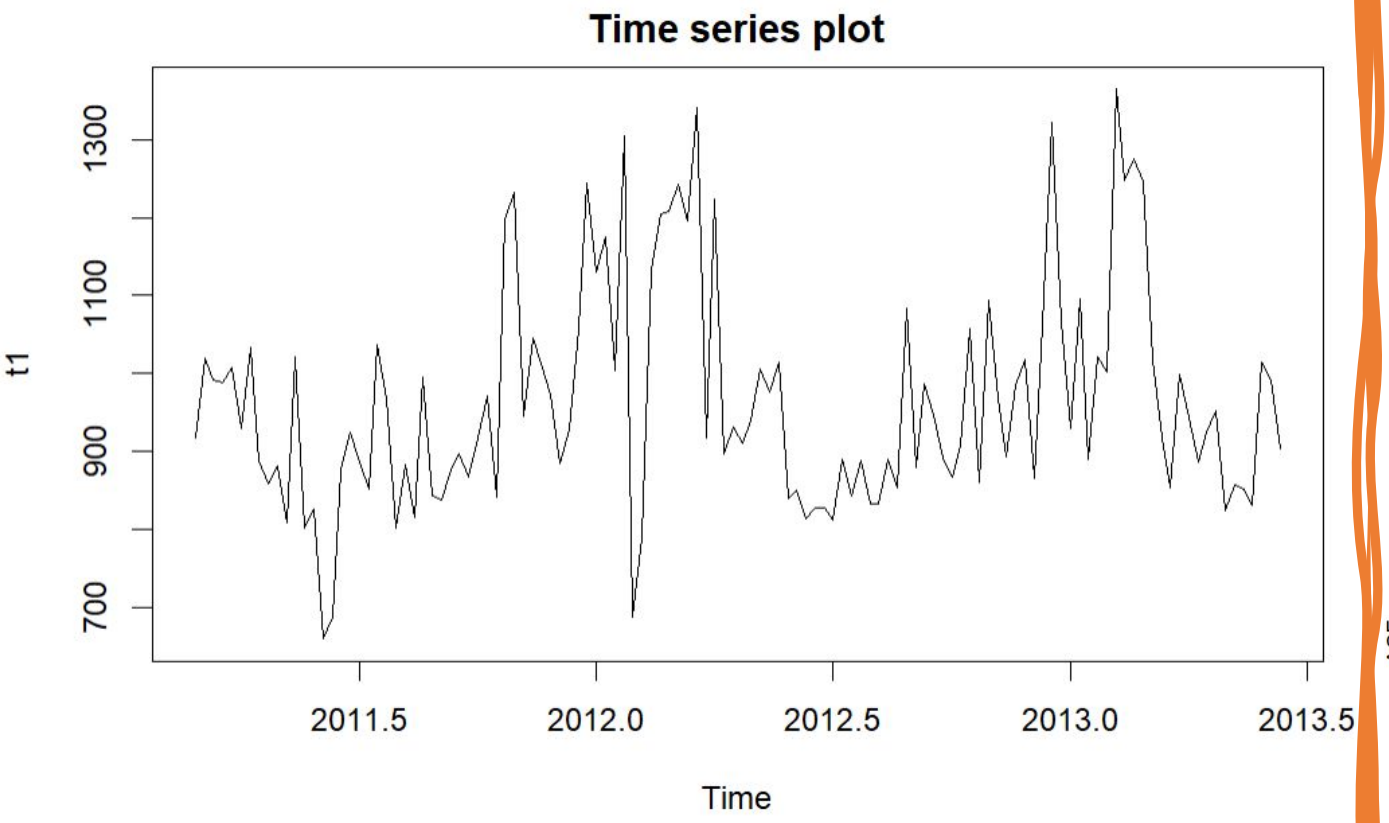
- National park authorities must start preparing for season with high foot traffic
- Forest officials may have to start patrolling more often.
- Bus and other public transport services can start more buses on the route
- Also, tourism industries can use this to stock supplies in preparation

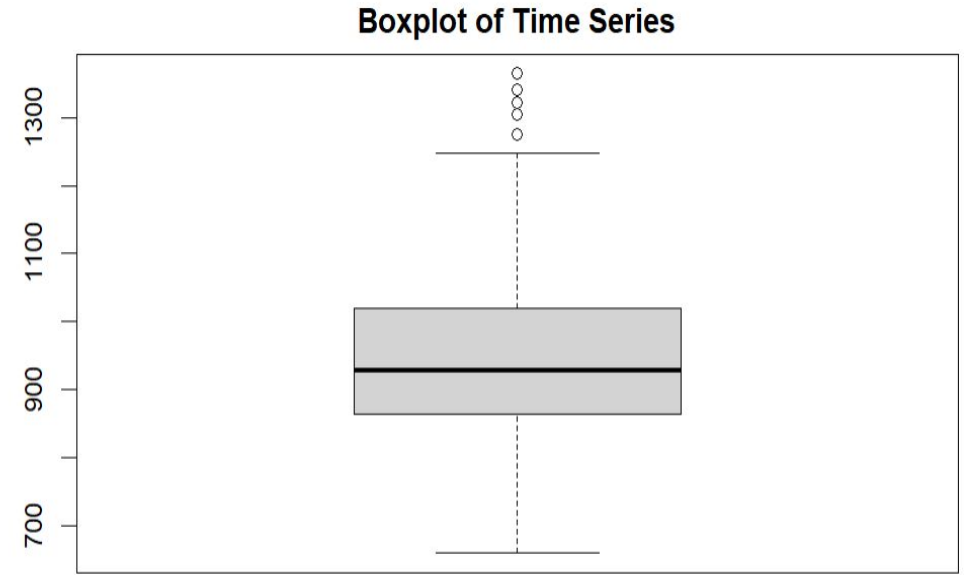
DATA

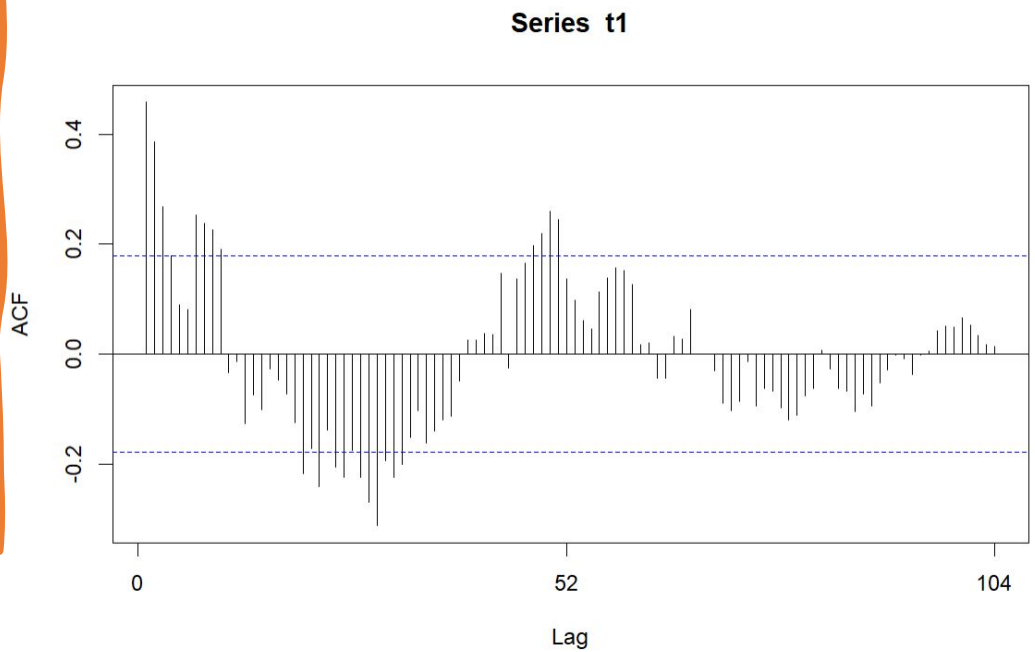
- The Dataset consists of temperature(in Fahrenheit) and footfalls of a national park in California.
- The Dataset holds data points from 2011 to 2013
- It is a weekly data each starting on a Monday.
- Our data is aggregated to weekly values as the original dataset was on an hourly basis.



Time Series Plots







Accuracy

- We considered Mean Percentage Error (MPE) accuracy measure to predict the accuracy of our models in the project

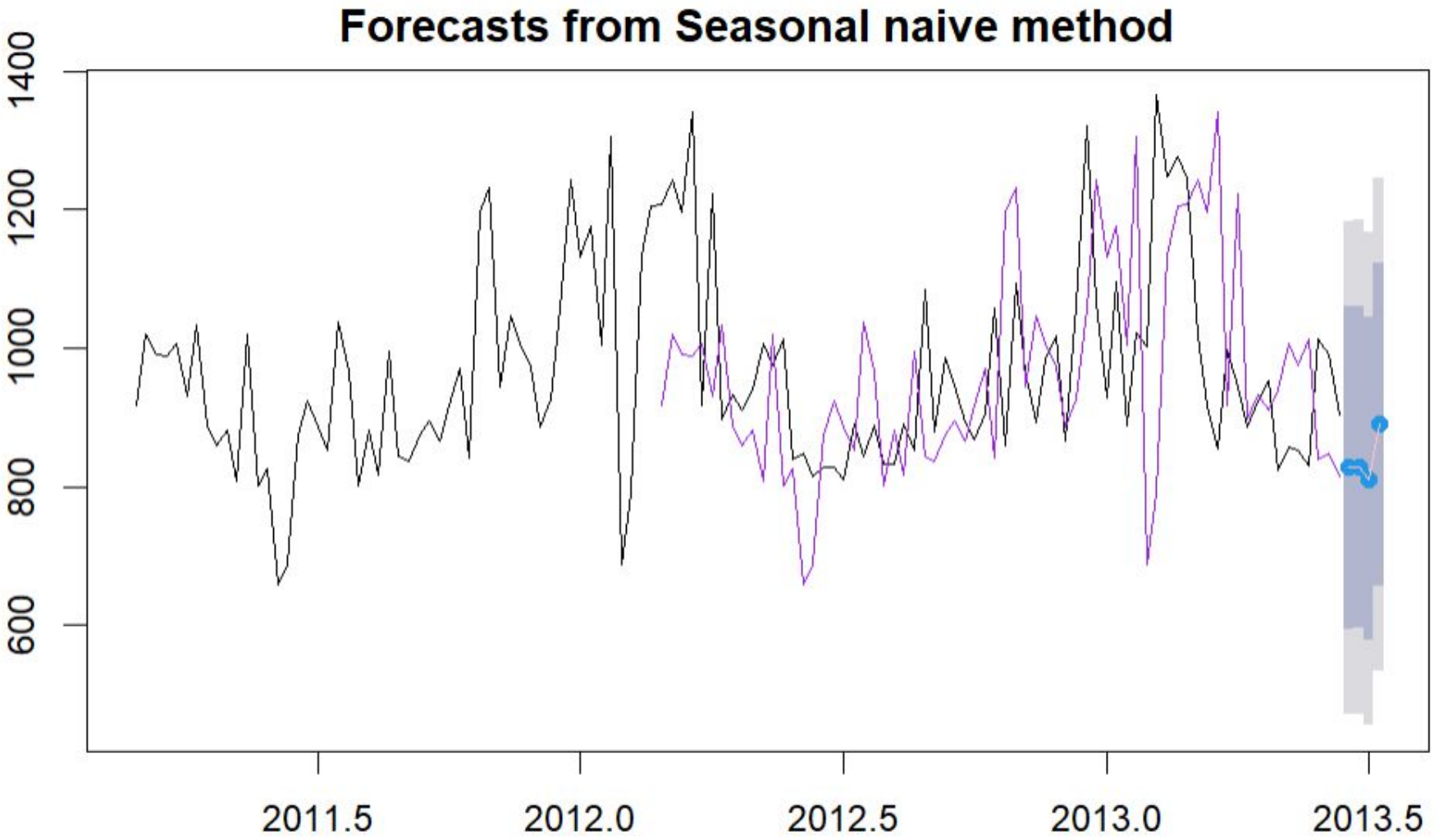
Zero
The forecast is unbiased

Positive
Large positive value signifies that the forecast is underestimated

Negative
Large negative value signifies that the forecast is overestimated

Comparison of Accuracy for Different Models

Model Name	MPE	Point Forecast
Mean Forecast	-2.057414	965
Random Walk Forest	-1.163738	903
Naïve Forecast	-1.163738	903
Snaive Forecast	-0.2849747	827
Moving Averages	0.04691367	802
Holtswinter Forecast	-2.131463	834
ARIMA	-1.43963	975



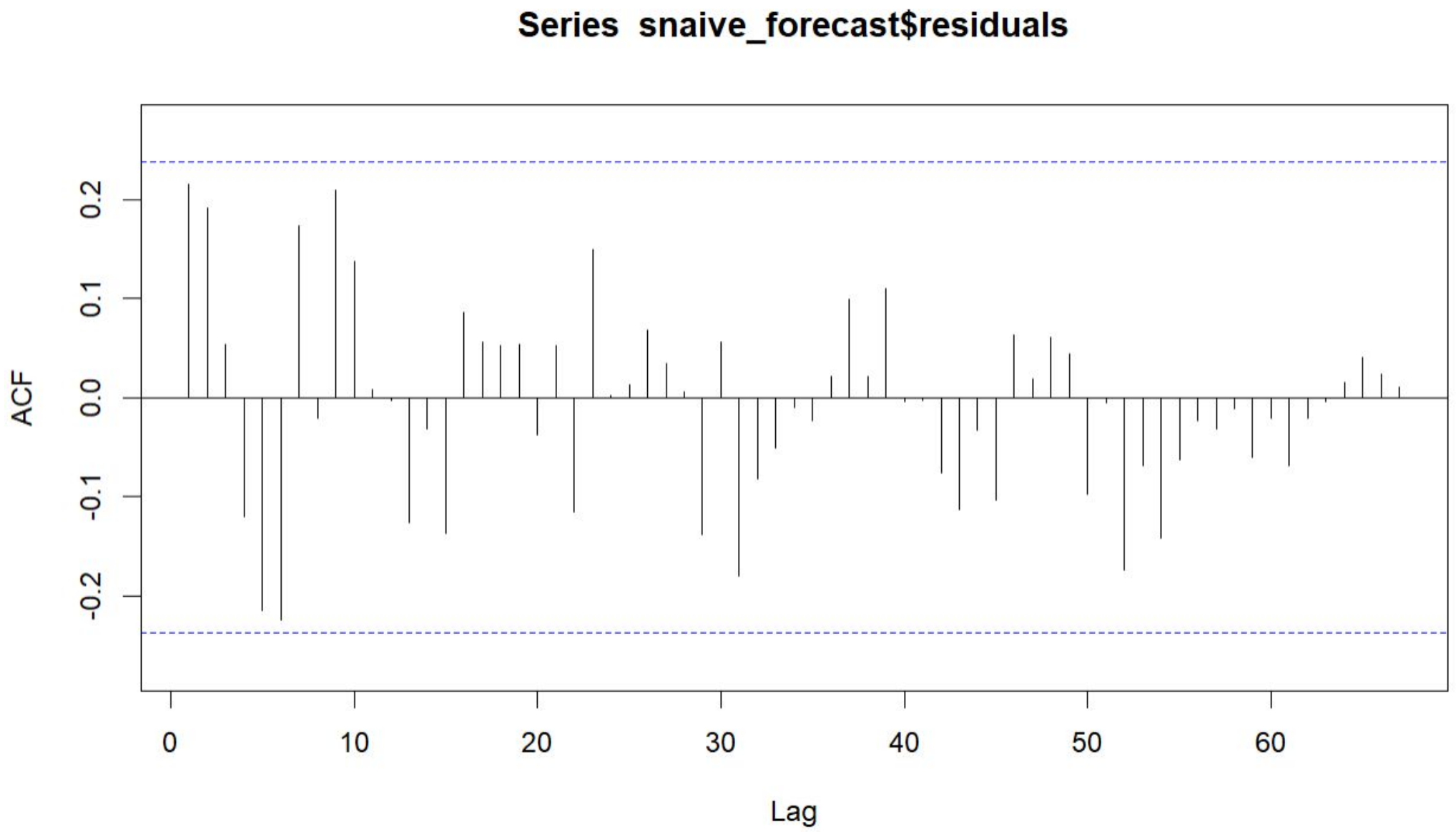
MPE

-0.2849747

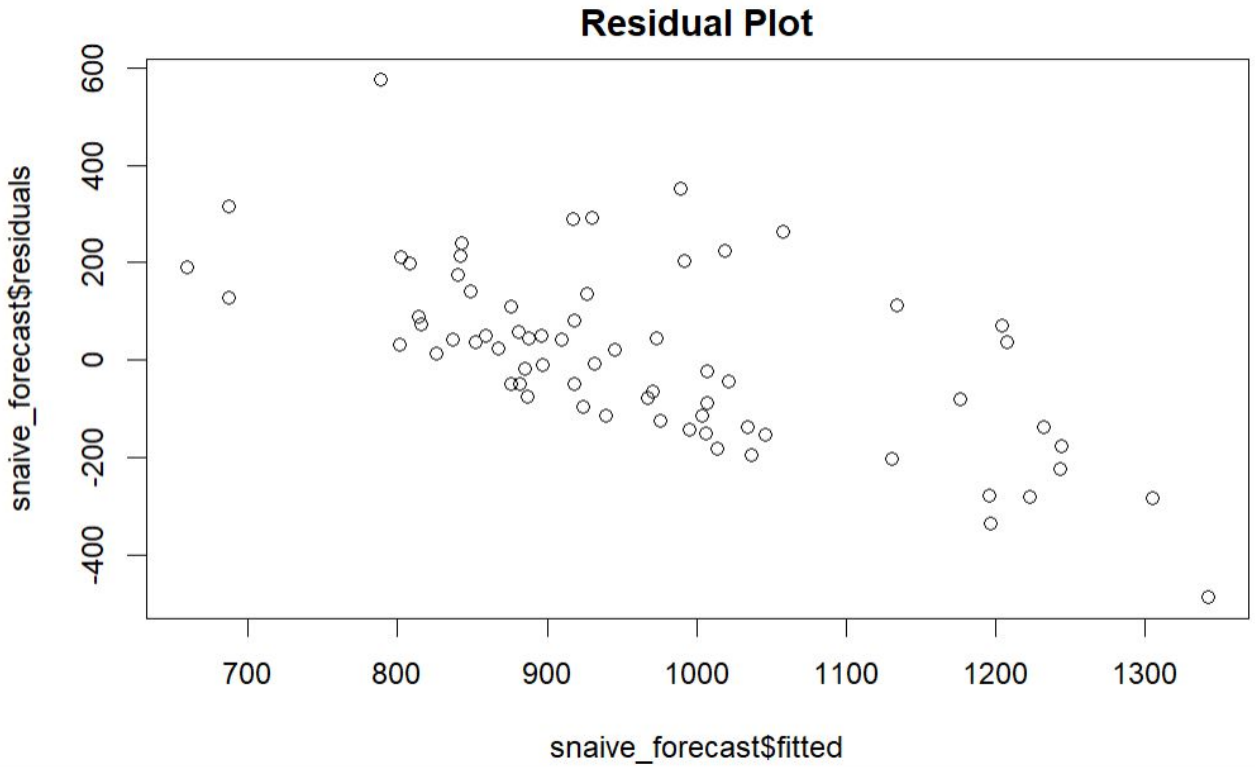
	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
2013.462	827	594.1069	1059.893	470.8208	1183.179
2013.481	828	595.1069	1060.893	471.8208	1184.179
2013.500	811	578.1069	1043.893	454.8208	1167.179
2013.519	890	657.1069	1122.893	533.8208	1246.179

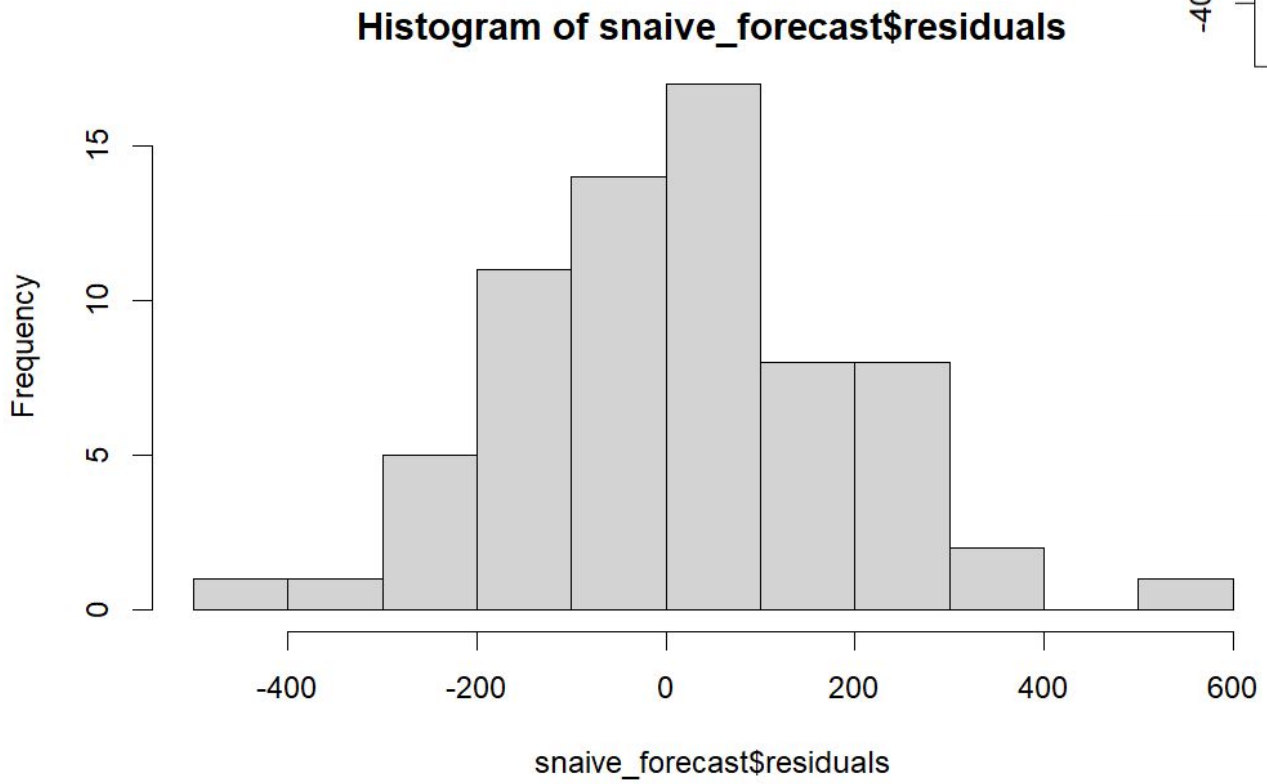
4 rows

ACF Plot for Snaive Forecasting



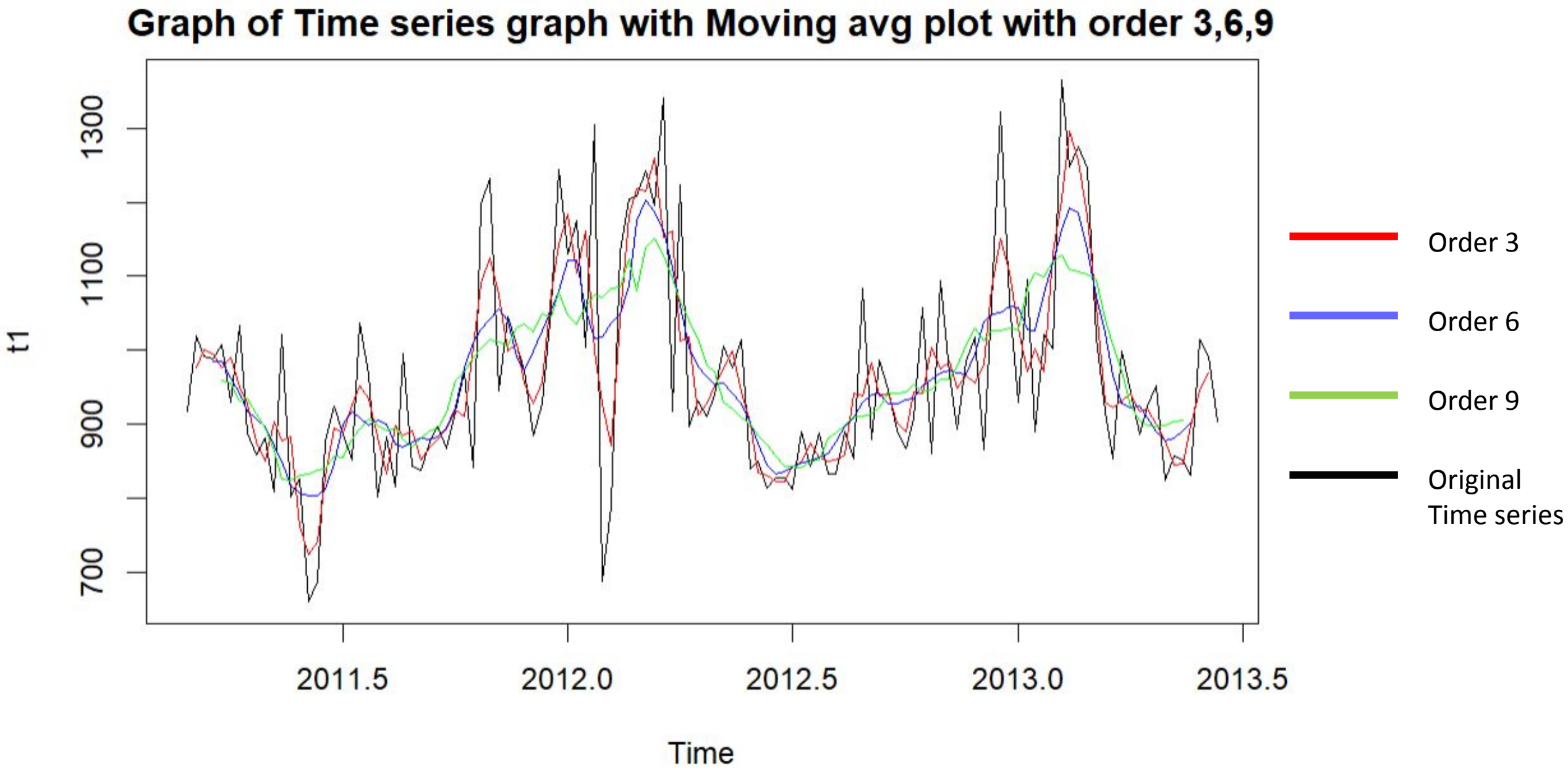
Residual Plot for Snaive

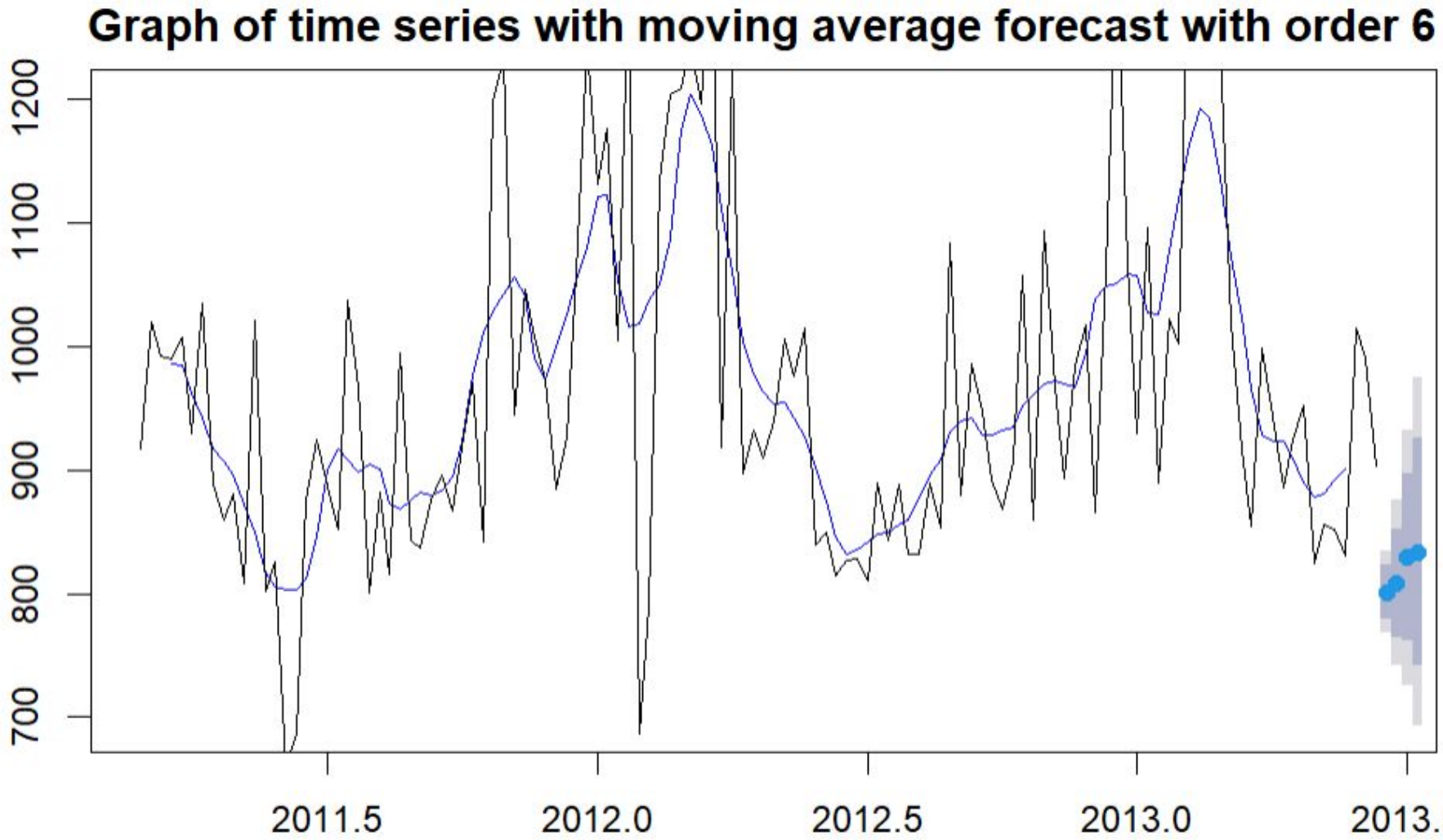




Histogram of Residuals for Snaive

Moving Averages Graph





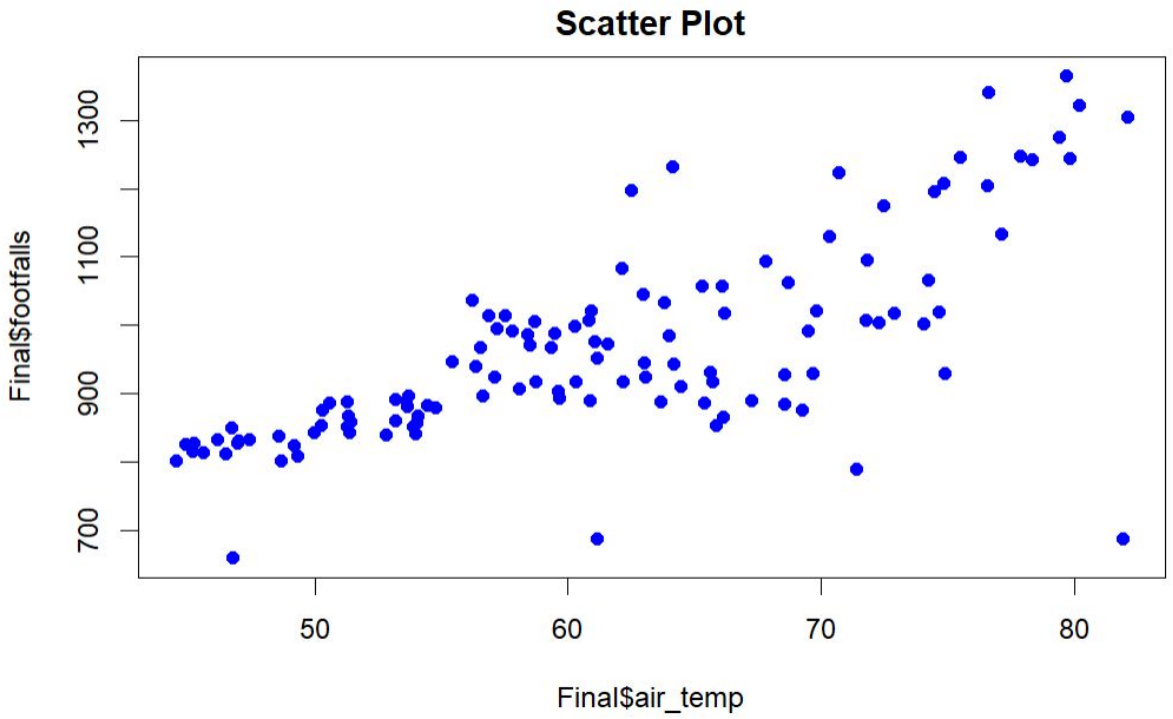
MPE=

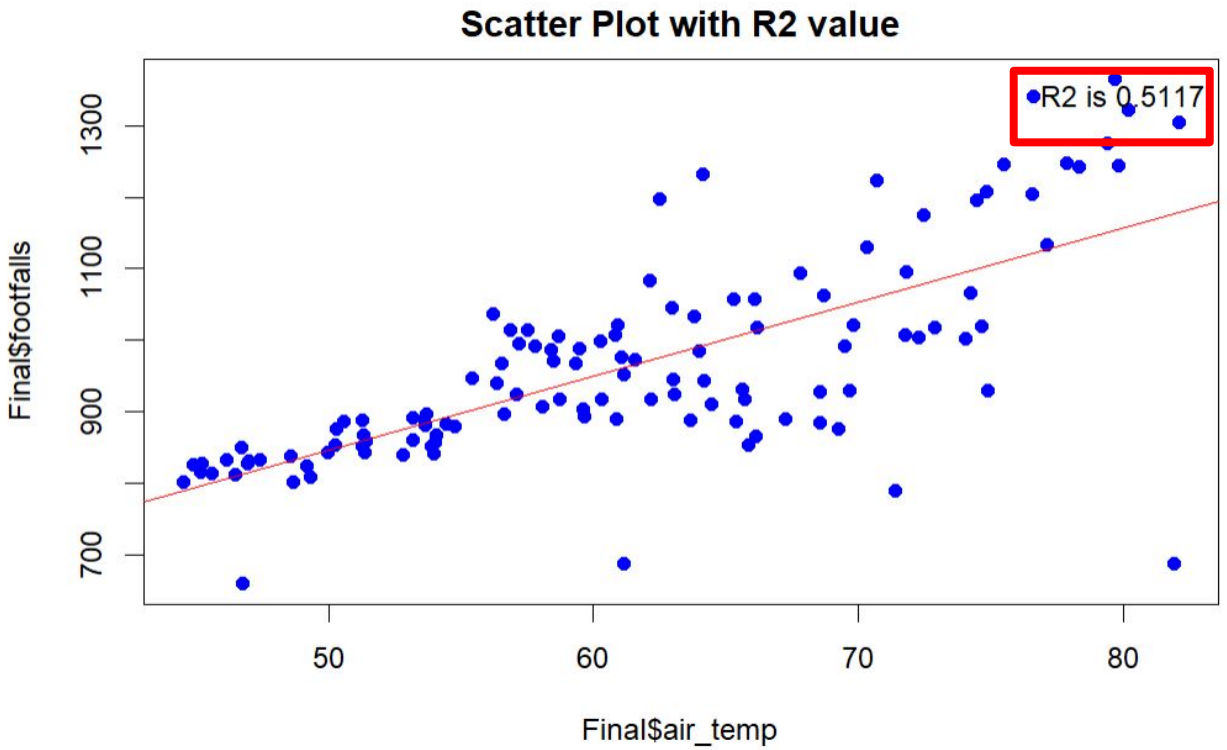
0.04691367

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
2013.462	801.2051	779.6269	822.7833	768.2041	834.2061
2013.481	808.2107	764.4580	851.9634	741.2968	875.1246
2013.500	828.9135	761.2653	896.5618	725.4545	932.3726
2013.519	833.4274	741.3805	925.4744	692.6537	974.2011

4 rows

Linear Regression





Linear Regression Model

```
Call:
lm(formula = Final$footfalls ~ Final$air_temp)

Residuals:
    Min      1Q  Median      3Q     Max
-490.57  -37.65    6.39   49.56  239.37

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    326.7243    57.6119   5.671 1.03e-07 ***
Final$air_temp  10.3867     0.9264  11.212  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 100.2 on 118 degrees of freedom
Multiple R-squared:  0.5158,    Adjusted R-squared:  0.5117
F-statistic: 125.7 on 1 and 118 DF,  p-value: < 2.2e-16
```

P value=

	Pr(>|t|)
(Intercept)	1.03e-07 ***
Final$air_temp	< 2e-16 ***

T value=

	t value
(Intercept)	5.671
Final$air_temp	11.212

Adjusted R2= **0.5117**

Implementation for Forecasting

```
Call:
lm(formula = Final$footfalls ~ Final$air_temp)

Coefficients:
   (Intercept)  Final$air_temp
        326.72           10.39
```

The formula for our model is ,

Y= mx+C ,
where y = average footfalls in a week,
m= slope or coefficient of x,
x= average air temperature in a given week,
C= intercept or the constant.

So for example in a given week the average temperature was 78F then the prediction of average footfalls for that week will be,

Y= 10.39(78)+326.72

which is 1137.14 which will be considered as 1137 or 1138 people.

Conclusions

- As the adjusted R2 value is 0.5117 ,51% variance in the footfalls can be attributed by the change in the temperature.
- We realise that the R2 value is very less for our model, but we do not discard the model as hiring of park labor does not happen on a weekly basis, so the forecast is good enough for them to be prepared.
- However, if we consider some more relevant variables in regards to the weather we can establish a better model with improved significance.
- Hence we also chose Moving averages order 6 and Snaive to be our best models as the accuracy helped us decide that the forecast is not biased and that should give the officials or authorities a fair chance to prepare for stocking or hiring in advance.


THANK.YOU