# Time series revenue analysis sample project

## Task formulation

Predict metric (revenue) for one fleet (BER) for 1 week ahead.

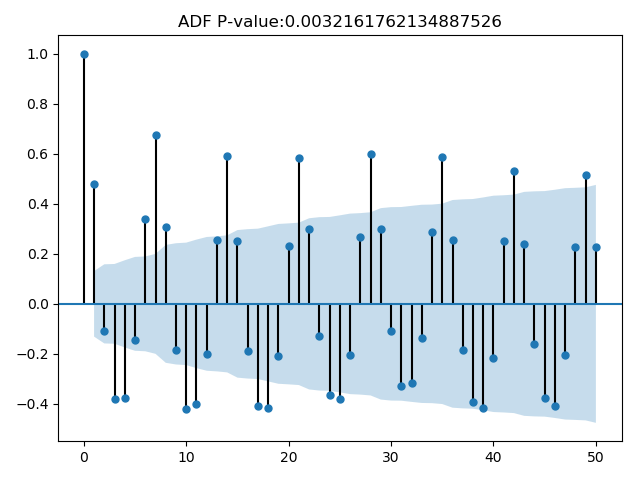
## Test data

Data was taken for 6 months, last 7 days were used as test data, the remaining data was used for training model and prediction. After that predicted data was compared with test data to determine deviation.

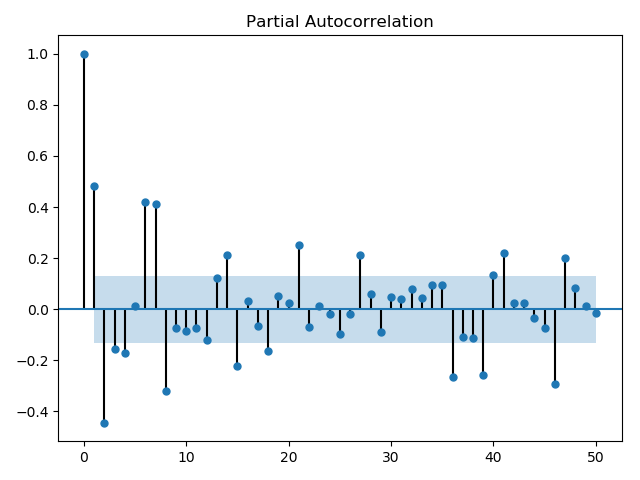
Prediction error formula is mean absolute error divided by mean fact value.

## Predictions via time series

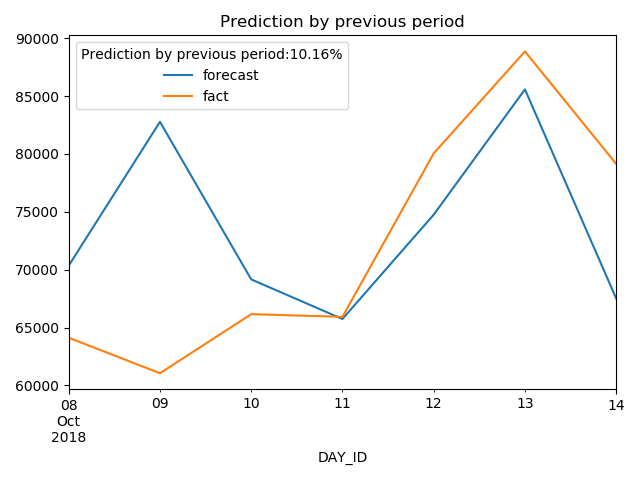
Initial data analysis:



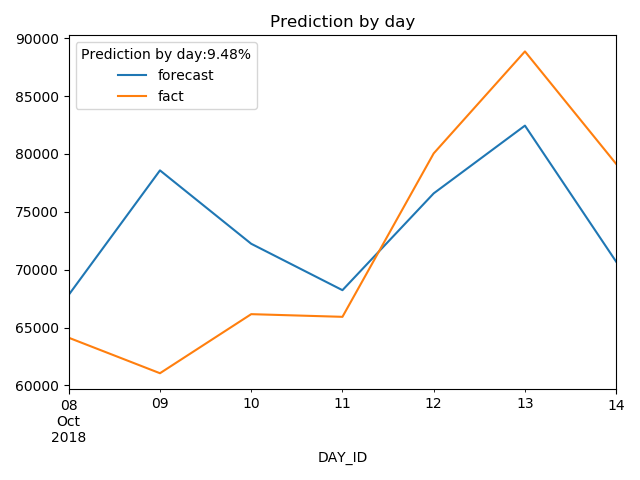
ADF P-value <0.05 which means we can reject hypothesis that the process is a random walk, possible to predict.



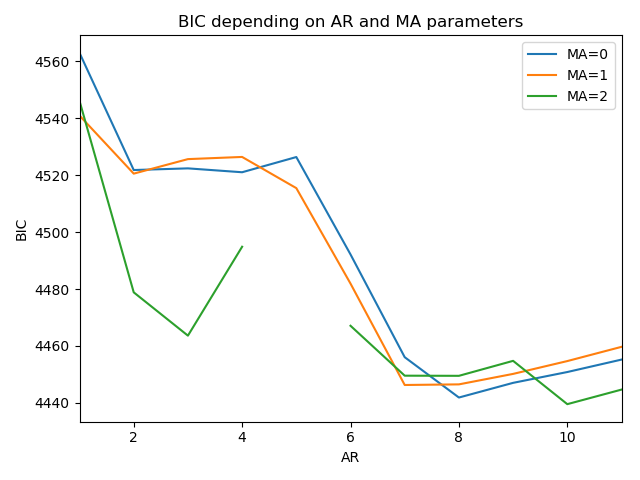
Simple prediction by using previous week data (predicted value for Monday is previous Monday, for Tuesday is previous Tuesday, etc):



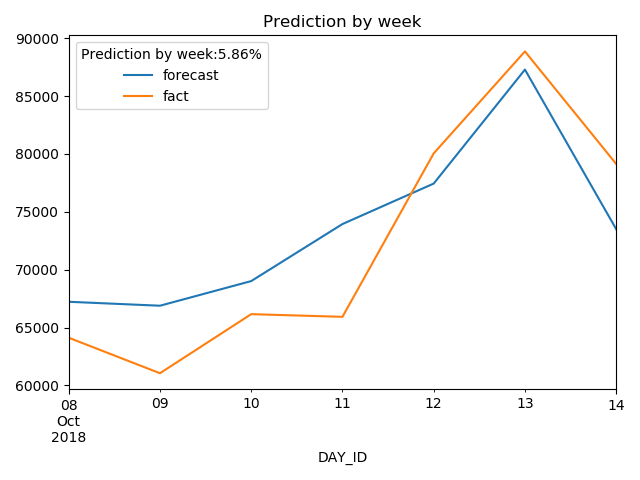
Prediction from daily data with (AR, MA) = (7, 2):



Tuning daily parameters result:

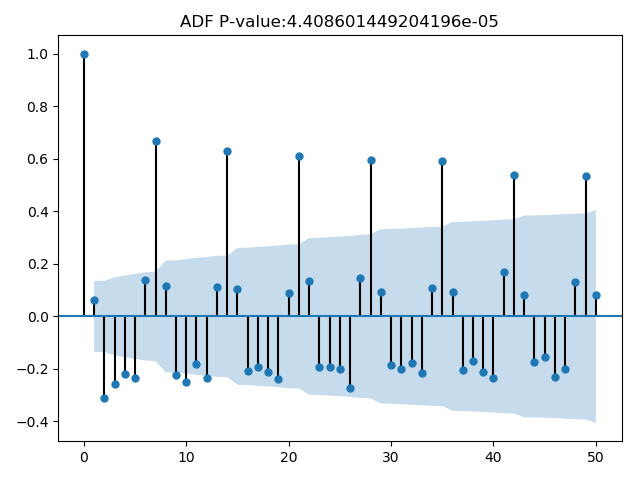


Prediction from weekly data with (AR, MA) = (2, 0) (data is resampled by week and each week day is modeled separately):



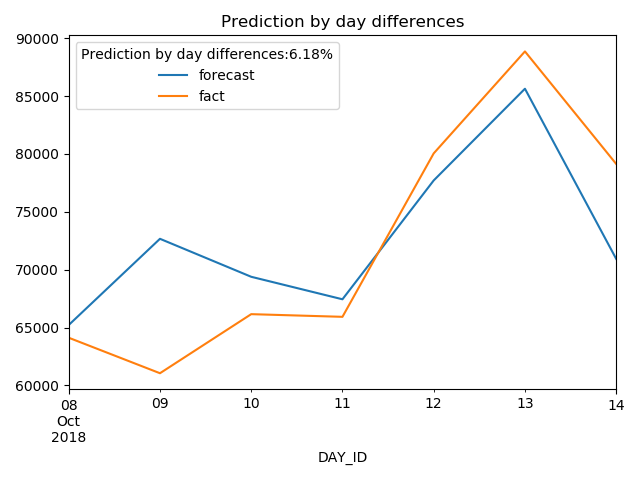
## Analysis of daily data via differences

Transformation: difference between current and previous value

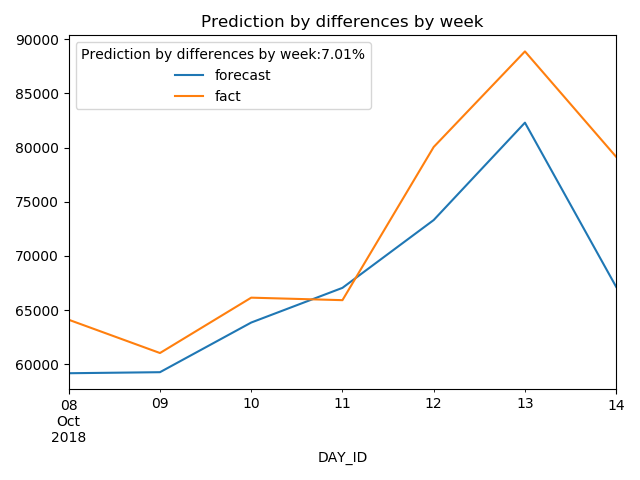


ADF P-value is much less than 0.05 which means that the process is more stationary comparing with pure values

Prediction from daily data with (AR, MA) = (14, 1)



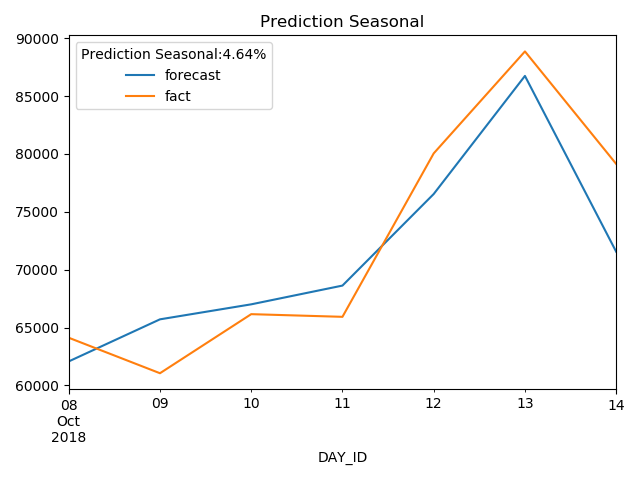
Prediction from weekly data with (AR, MA) = (2, 0)



## Prediction using autocorrelation model with seasonal component

Initial unprocessed data was used.

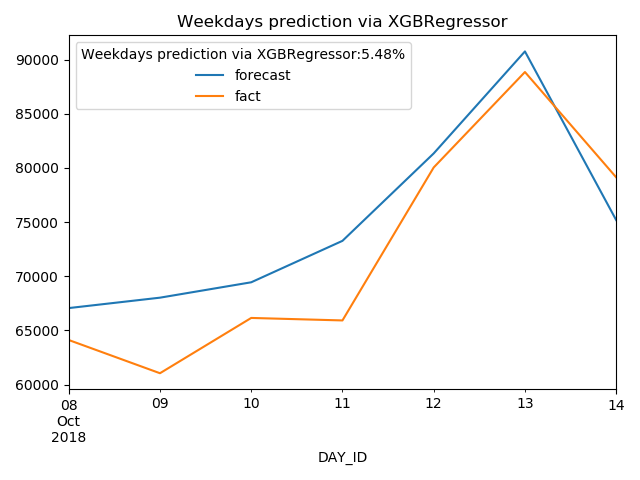
Prediction result with (AR, DIFF, MA) = (14, 0, 2), (SAR, SDIFF, SMA, S) = (3, 0, 1, 7):



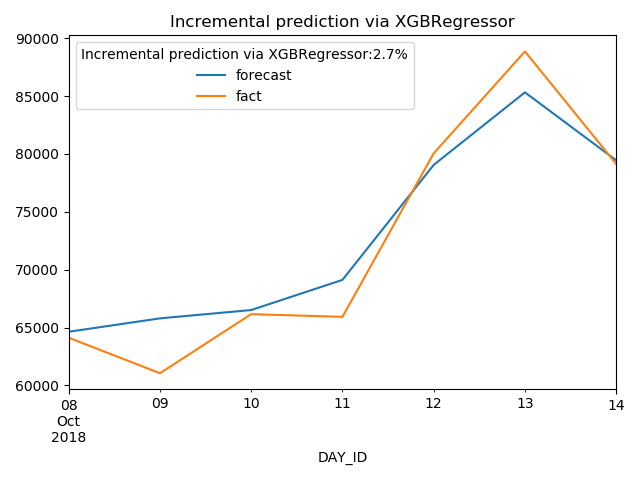
## Prediction using supervised learning

Initial data was enriched by shifted attributes from previous periods from the same dataset.

Prediction result with day of week only features using XGBOOST parameters (n\_estimators=10000, max\_depth=3)



Prediction with day of week features and previous days labels using XGBOOST parameters (n\_estimators=10000, max\_depth=3)



## Summary

* A few approaches based on time series data analysis to predict revenue for a fleet for a week were studied.
* Simple prediction using just previous period data shows deviation of more than 10% from real data.
* The studied models showed 30% - 40% improvement in precision comparing with previous period data analysis
* There is a possibility for further models parameters tuning which can lead to further prediction quality improvement
* It is also possible to use other approaches to predict revenue based on supervised learning techniques