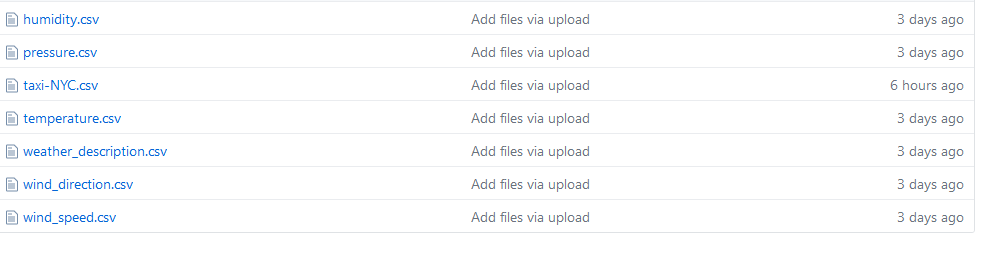
**Techincal Blog Assignment**

Goal: Finding a model to predict the number of the taxi rides given time date and temperature.

Tools:

* Python
* Random Forest of predictive regression modeling.
* Multiple relatinal data bases.

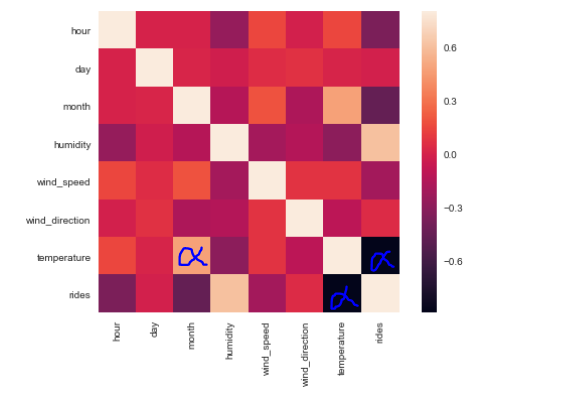
The datasets had information from different date and time.



After cleaning, converting and aggregating data, I achieved to my processed database.

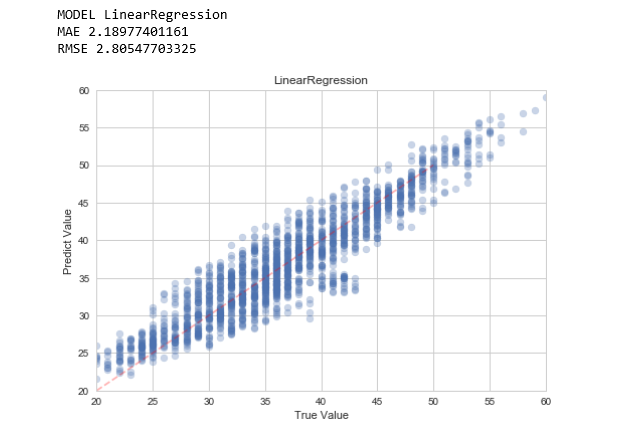
I have to deal with different arrangement of data and it took too long to manage them. All data sets are Relational database and I had primary key and foreign key to use them.

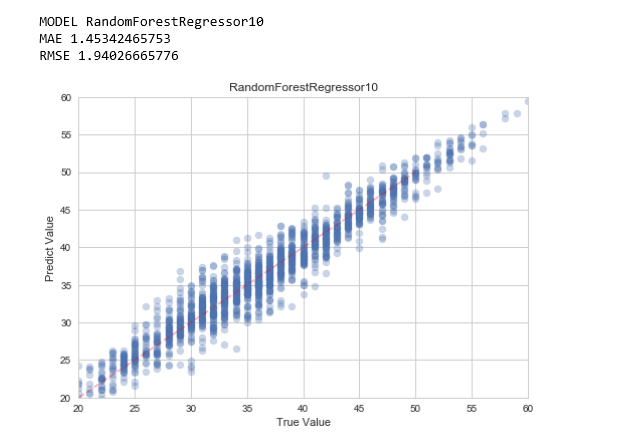
Them I chose the feature .The correlation between the feature has been examined.

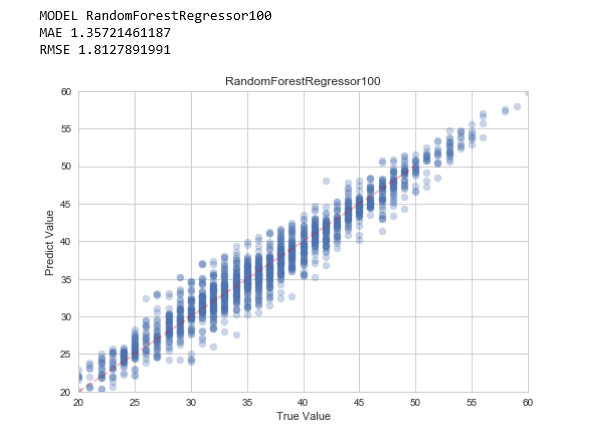


There is a high negative correlation between temperature and taxi rides. In addition of day hour and month I chose temperature to use in model. On the other hand, **There is a high correlation between month and temperature, So it could be enough to use one of them**. But I tried both of them knowing there is no need.

So I chose data related to 2015 to build my models. I tried 3 models :







The best result has been provided by RandonForestRegressor100. Here the ranking table of the models is available.

