



SUBMISSION BY:

- AVINASH GOEN (23PGAI0036)
- MINAL SETH (23PGAI0079)
- PUNIT DROLIA (23PGAI0113)
- SYED ABBAS RAZA (23PGAI0129)

BEIJING AIR QUALITY

PROJECT

The goal of this project is to analyse data on particulate matter in Beijing in order to understand trends and predict levels of PM 2.5, a type of particulate matter that can remain suspended in the air for long periods of time. The data comes from 12 testing sites and was combined into a single dataset, with missing values interpolated. The team then checked for stationarity and correlations between features and time periods in the dataset. They also fit models to predict the last 20% of the data, which had been set aside as a test case. However, due to time and hardware constraints, they decided to focus on a univariate time series forecast instead of a multivariate forecast.

BUSINESS CASE:

1. Marking a Target/Identifying Stakeholders

To whom does it matter if we can predict the amount of particulate matter in the air?

- **Residents of an Area and Tourists:**

Will they be able to enjoy time outdoors?

- **People Who Suffer From Respiratory Illnesses:**

Will their conditions worsen? Do they have to worry about loved ones contracting the same illness if not already predisposed? Will this contribute to premature death?

2. Understanding how different countries and regions may battle against the proliferation of particulate matter

Are some regions experiencing a rise in PM 2.5 while others are experiencing a decrease?

3. Understanding the viability in different models

Is it better to use a Facebook Prophet or a basic ARIMA?

Will SARIMAX models work with this data set?

Does the best model do a good job of predicting the target levels?

AFFECTS OF AIR POLLUTION



20 million tonnes of failed crops in 2018

Images Green Peace and Time

In 2017, air pollution was responsible for 5 million global deaths.

pollution problem

In 2018 China, reported 1.1 million premature deaths

linked to the air

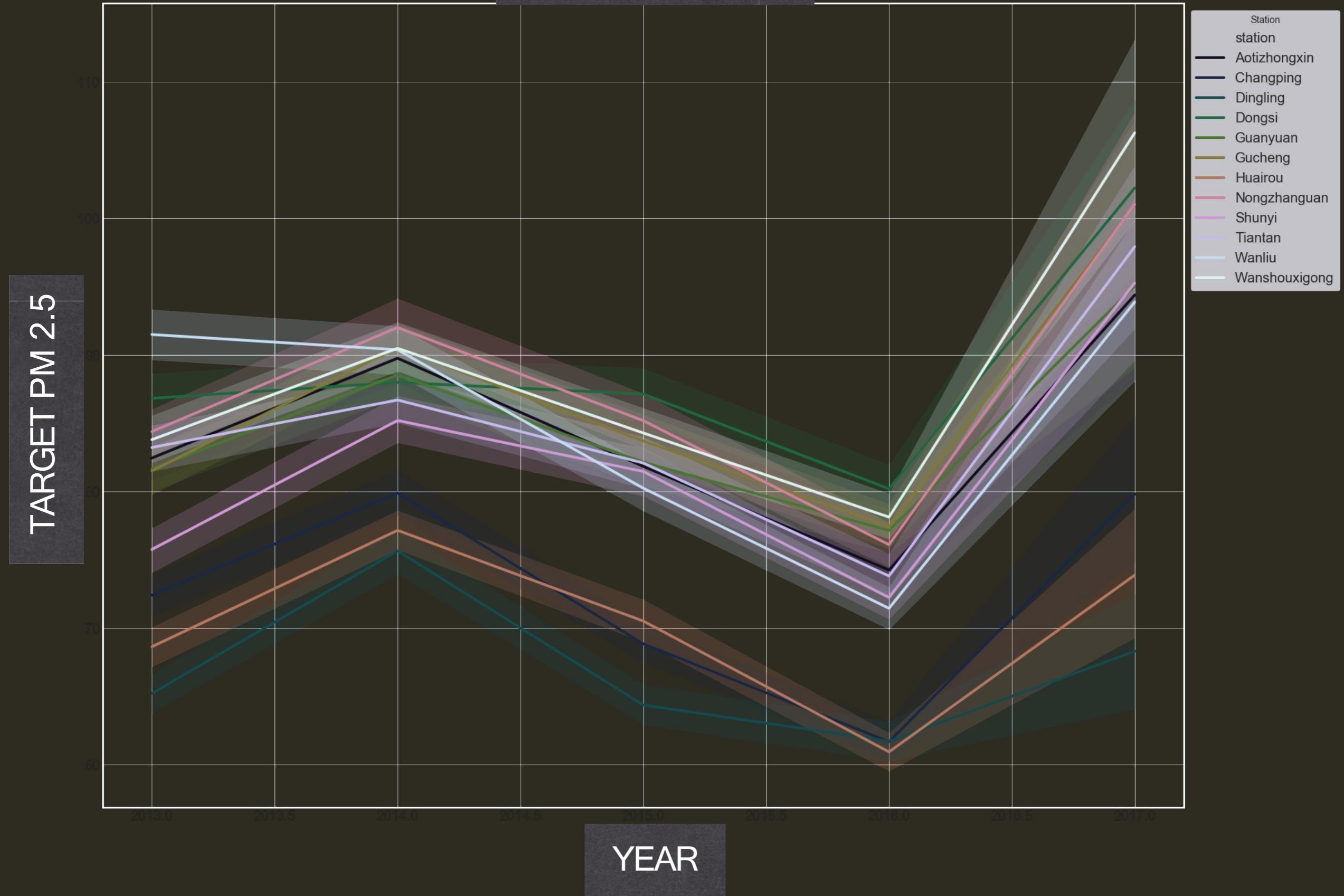
OUR DATA IS
PROVIDED BY:

UCI MACHINE
LEARNING LIBRARY:

BEIJING MULTI SITE AIR
QUALITY DATA SET



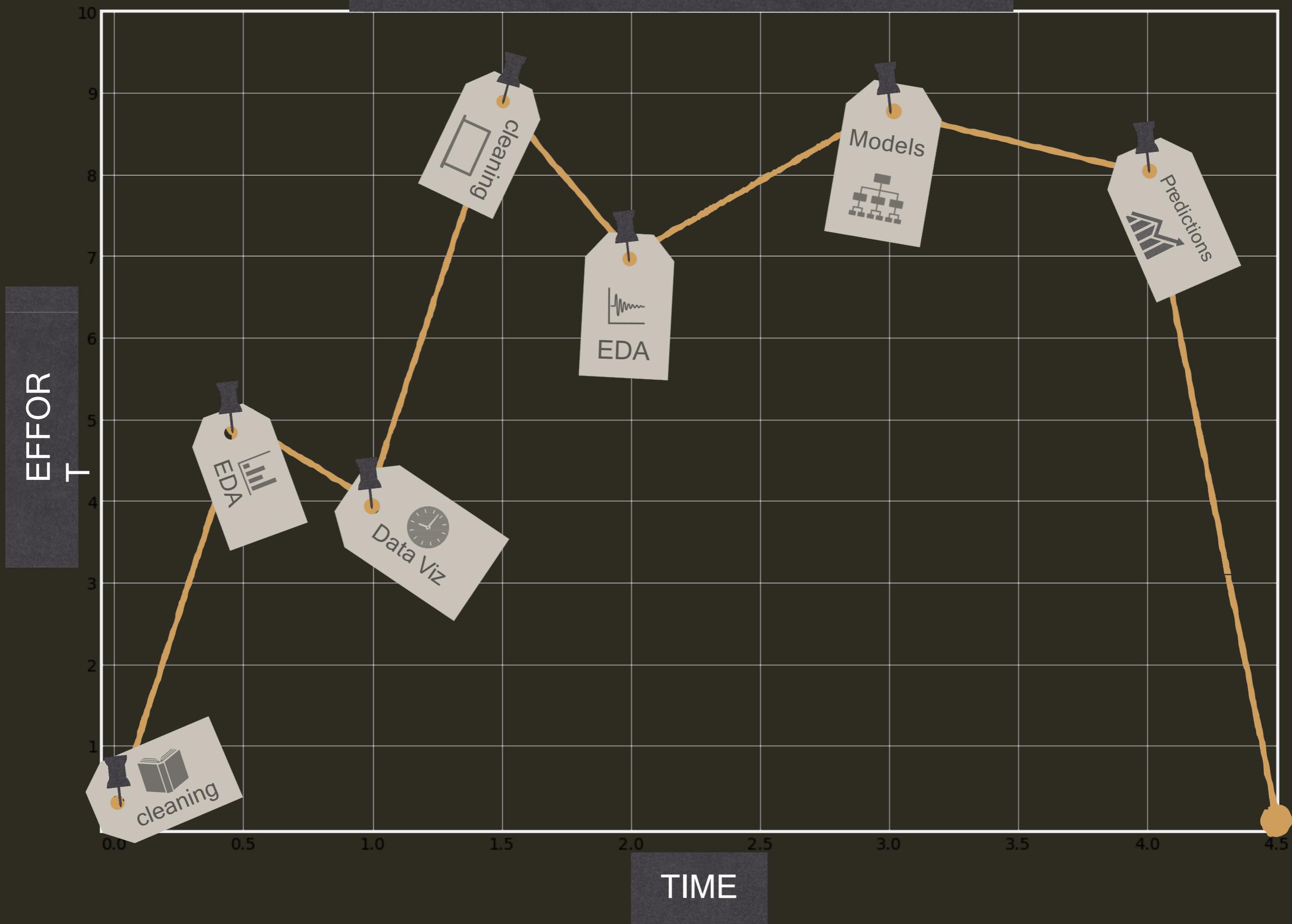
PM 2.5 by YEAR

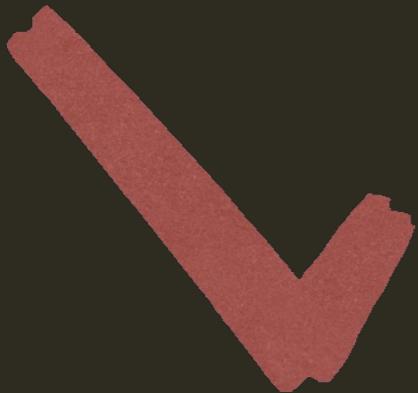


PM 2.5 by YEAR



JOURNEY THROUGH TIME SERIES





CHECKING FOR
NAN VALUES

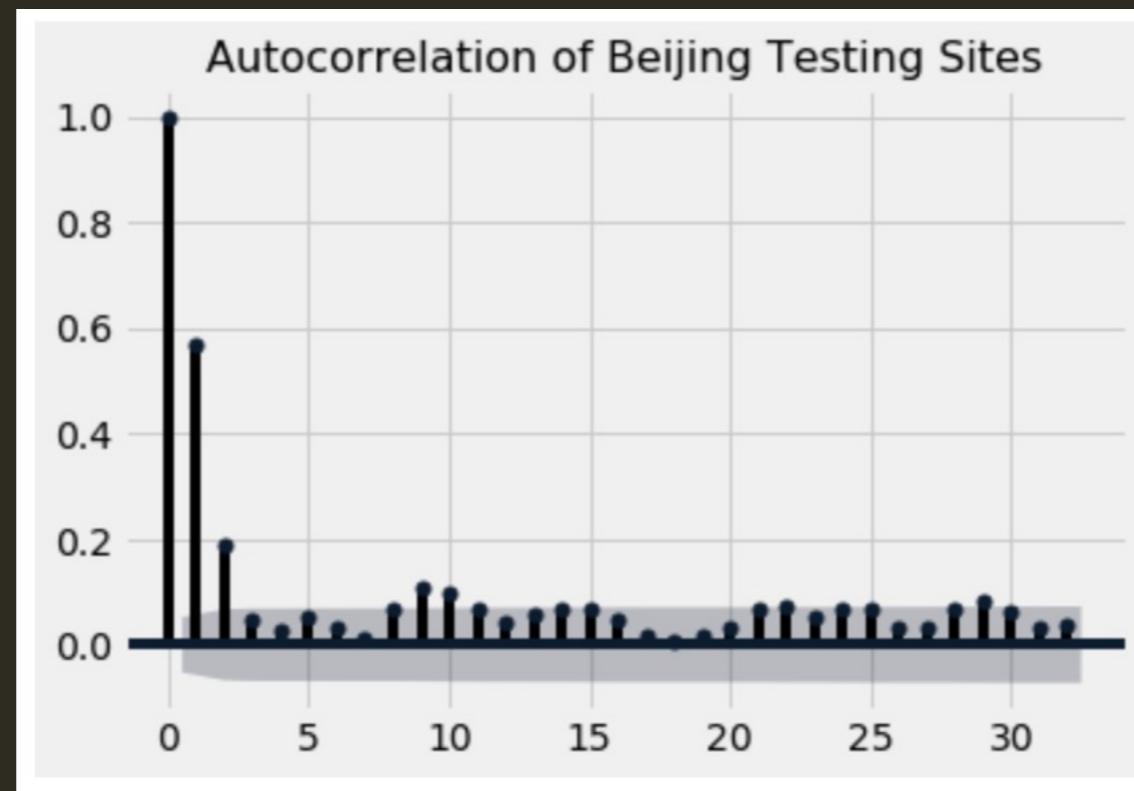
INTERPOLATION METHOD = TIME



STATIONARITY

DICKEY FULLER TEST
NP.LOG for Daily

ACF



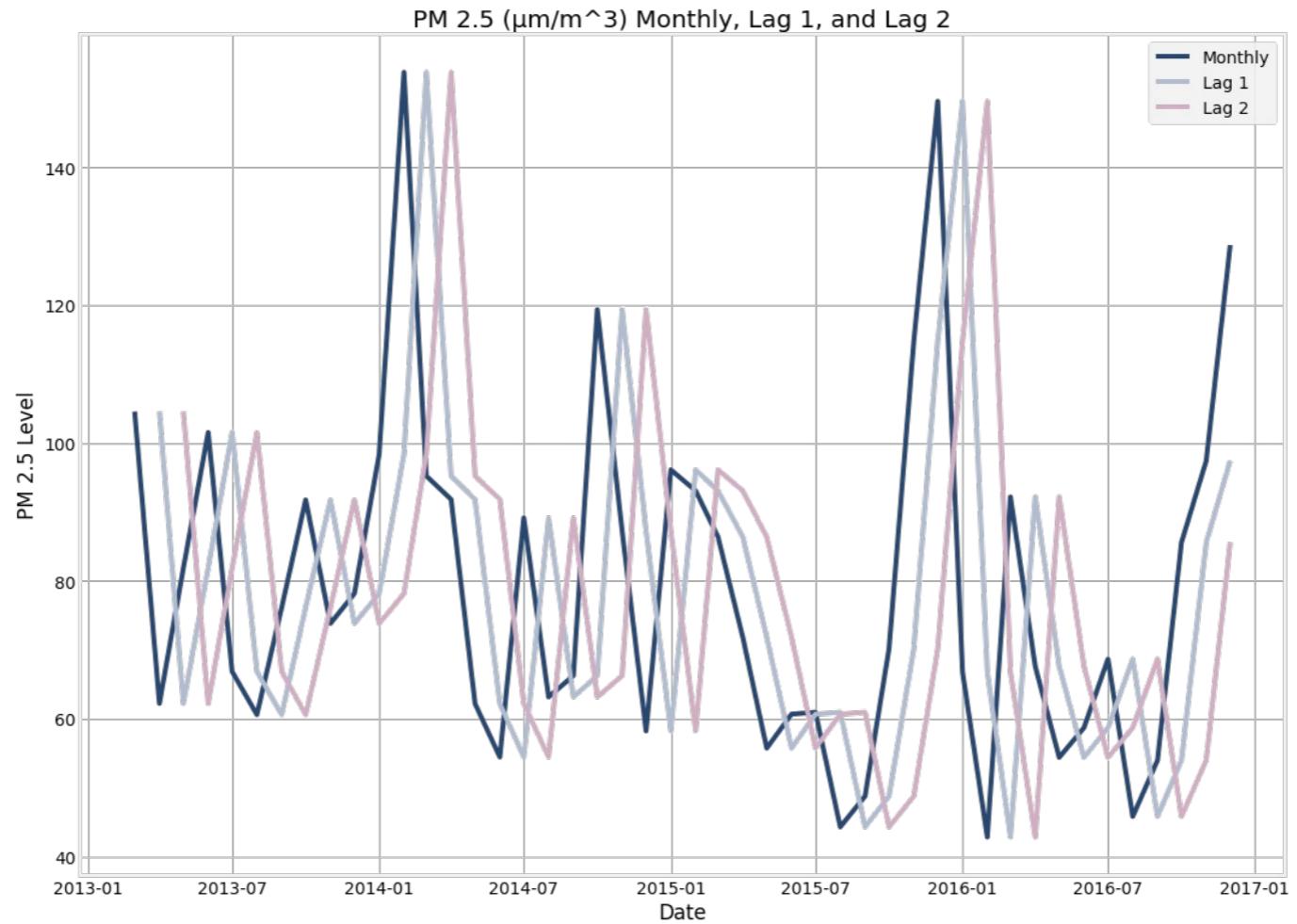
- The correlation of time series observations calculated with values of the same series at previous times
- Measures both direct and indirect correlations
- This graph shows there is correlation between the first three lags.

PACF



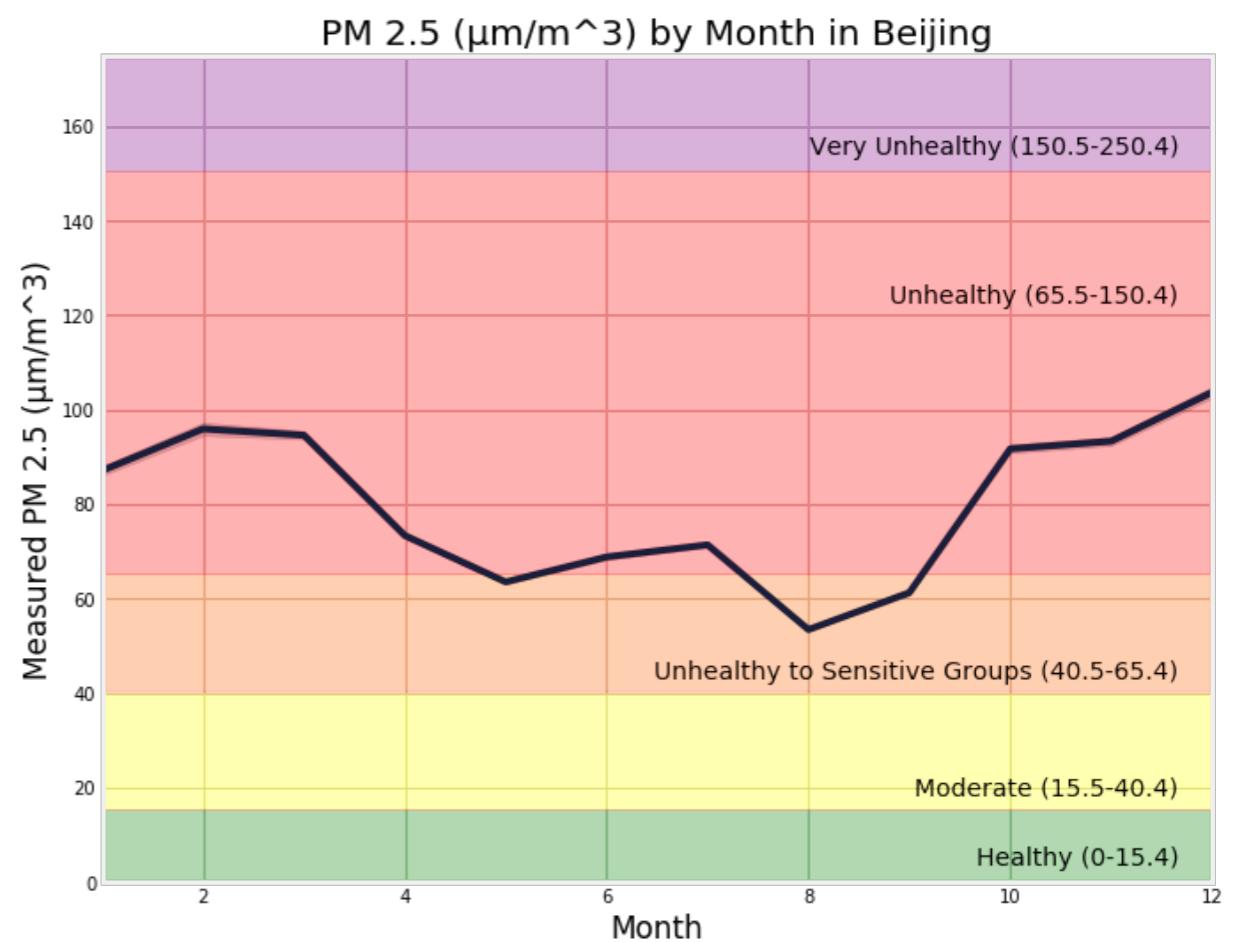
- The partial auto correlation removes the indirect correlations
- There is really no correlation after the first two lags.

WE THEN MOVED TO FINDING IDEAL PARAMETERS (p,d,q)
AND THEN RAN THE MODEL.



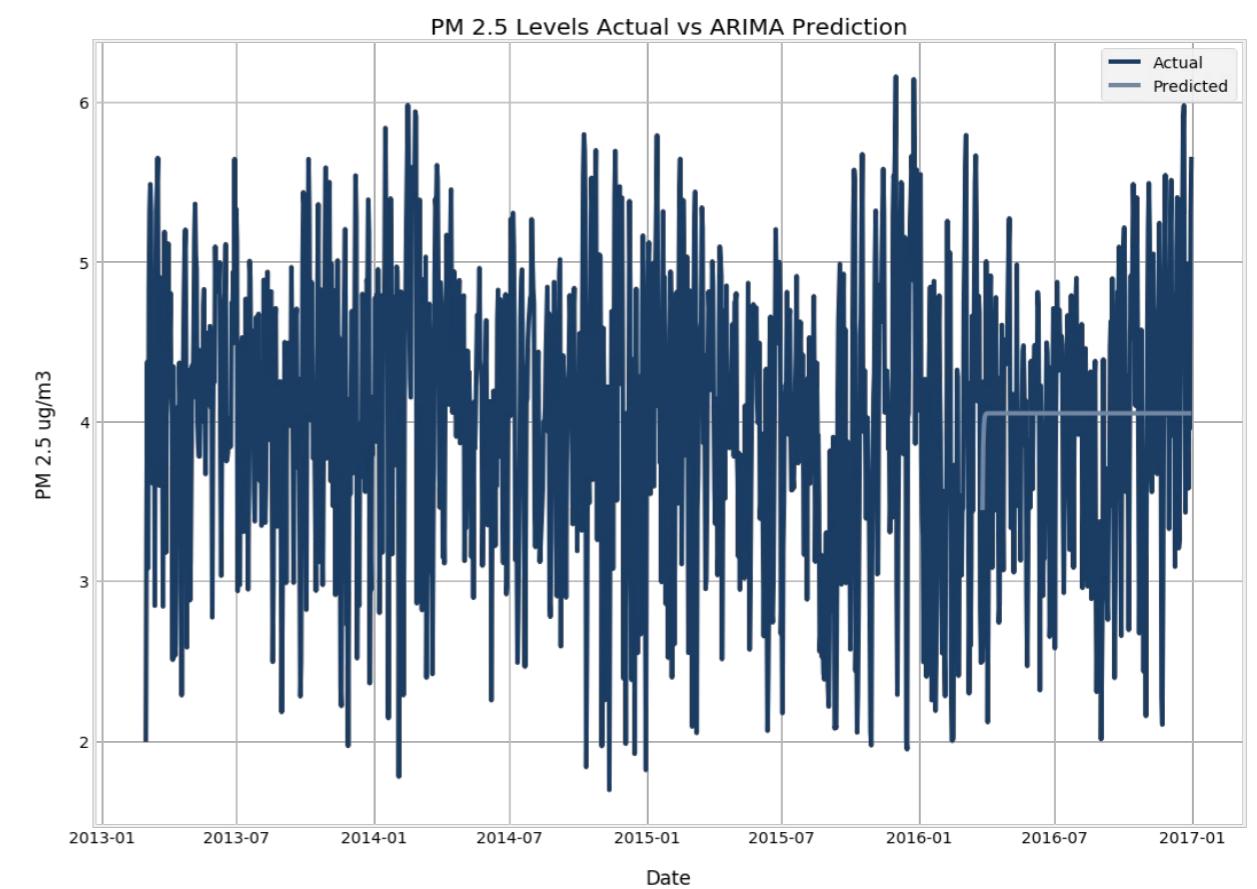
IS IT STATIONARY?

VISUALIZING BEIJING'S AIR QUALITY PROBLEMS



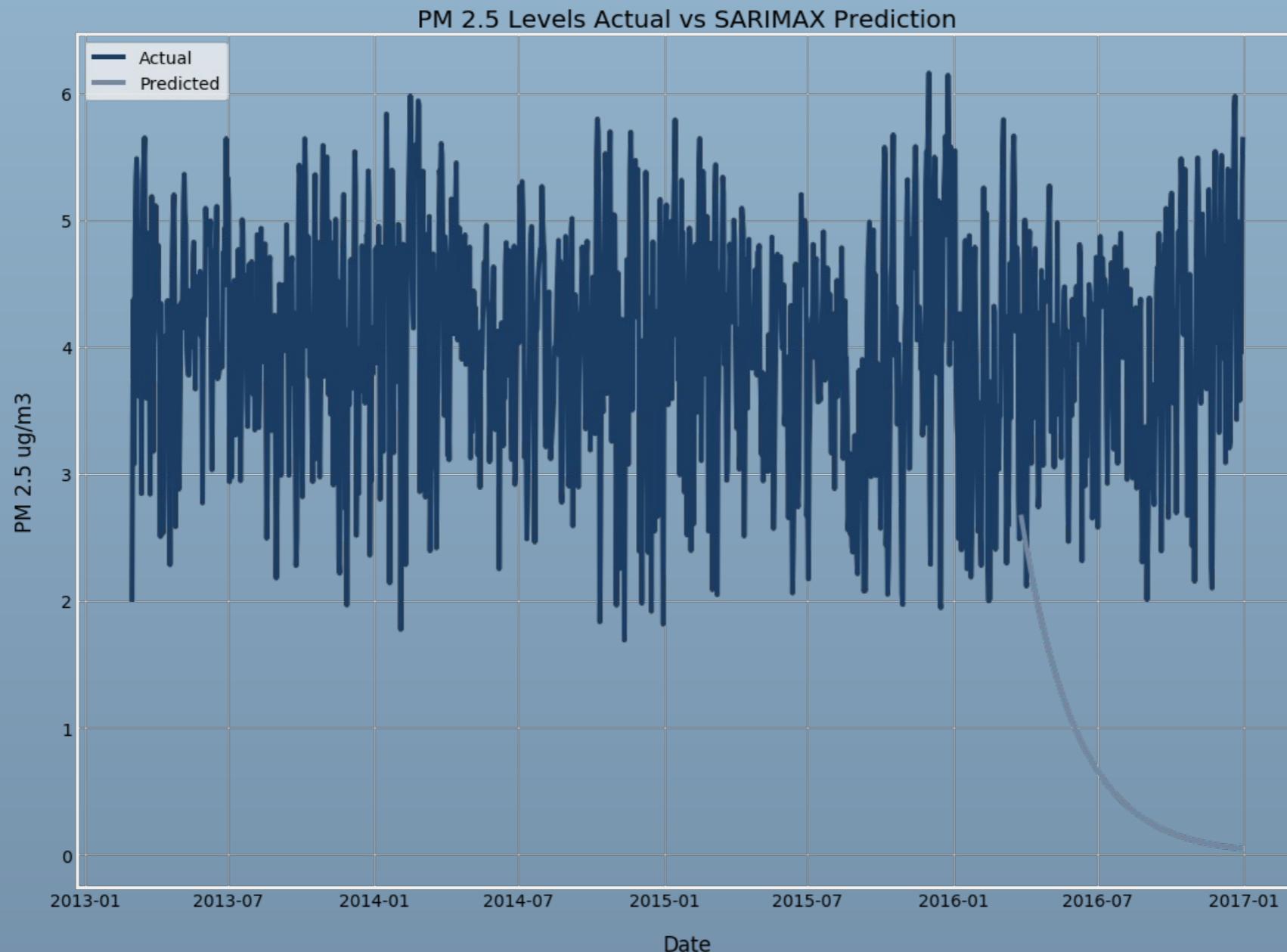
MODELS

ARIMA

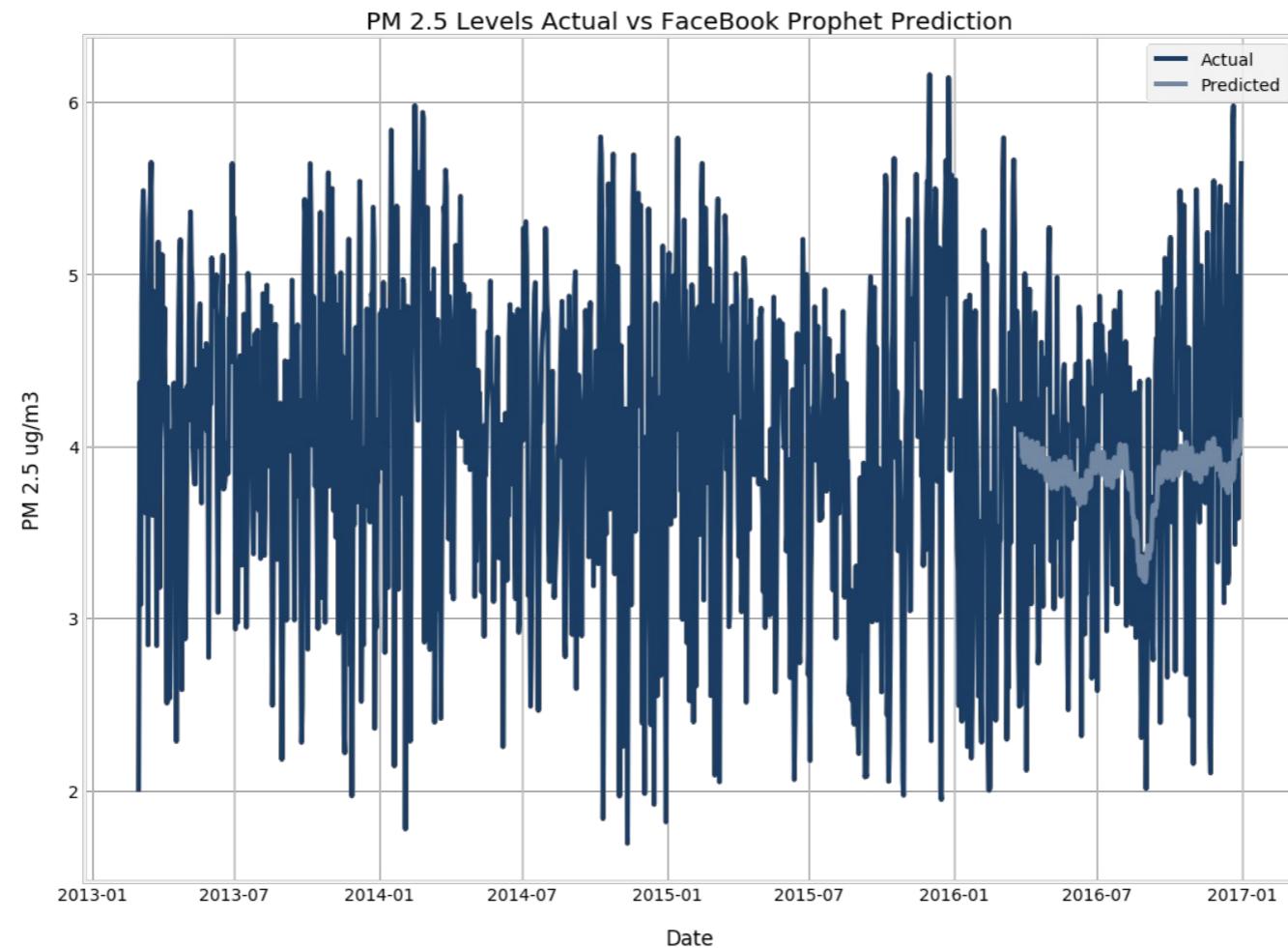


SARIMAX

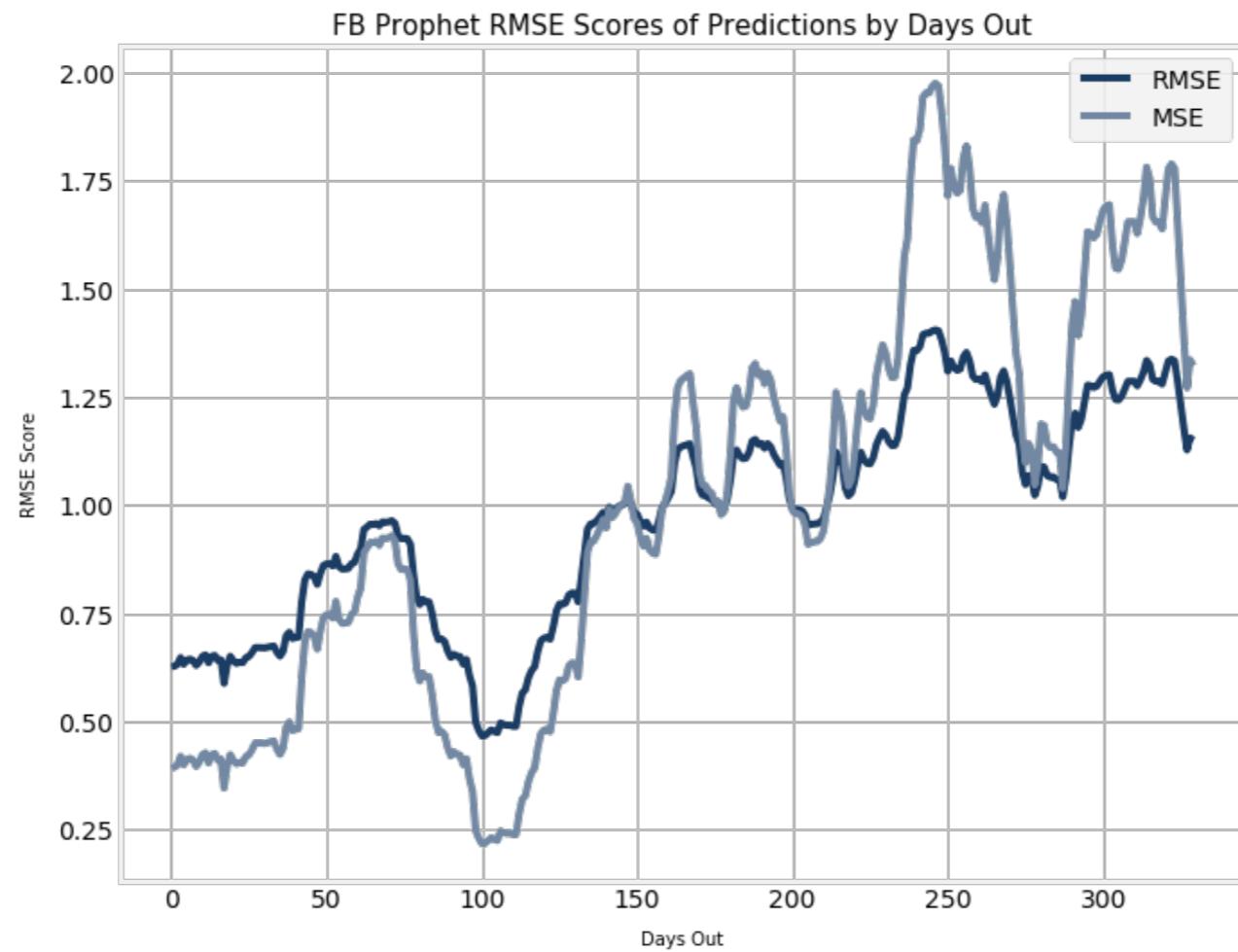
SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE WITH EXOGENOUS REGRESSORS MODEL



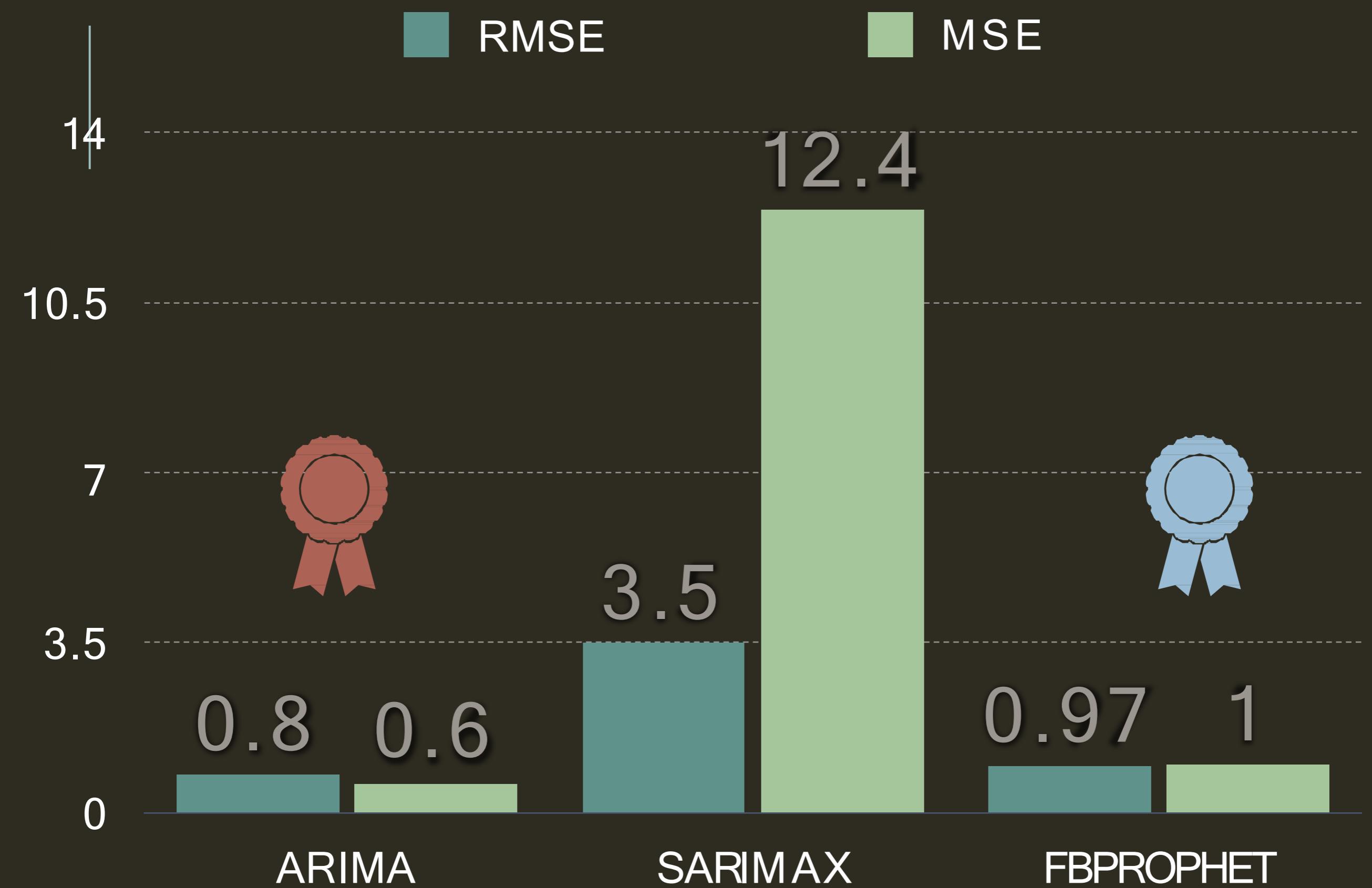
WOULD LIKELY HAVE BEEN BETTER FOR A MULTIVARIATE APPROACH.



FACEBOOK PROPHET



FACEBOOK PROPHET



THERE IS HOPE

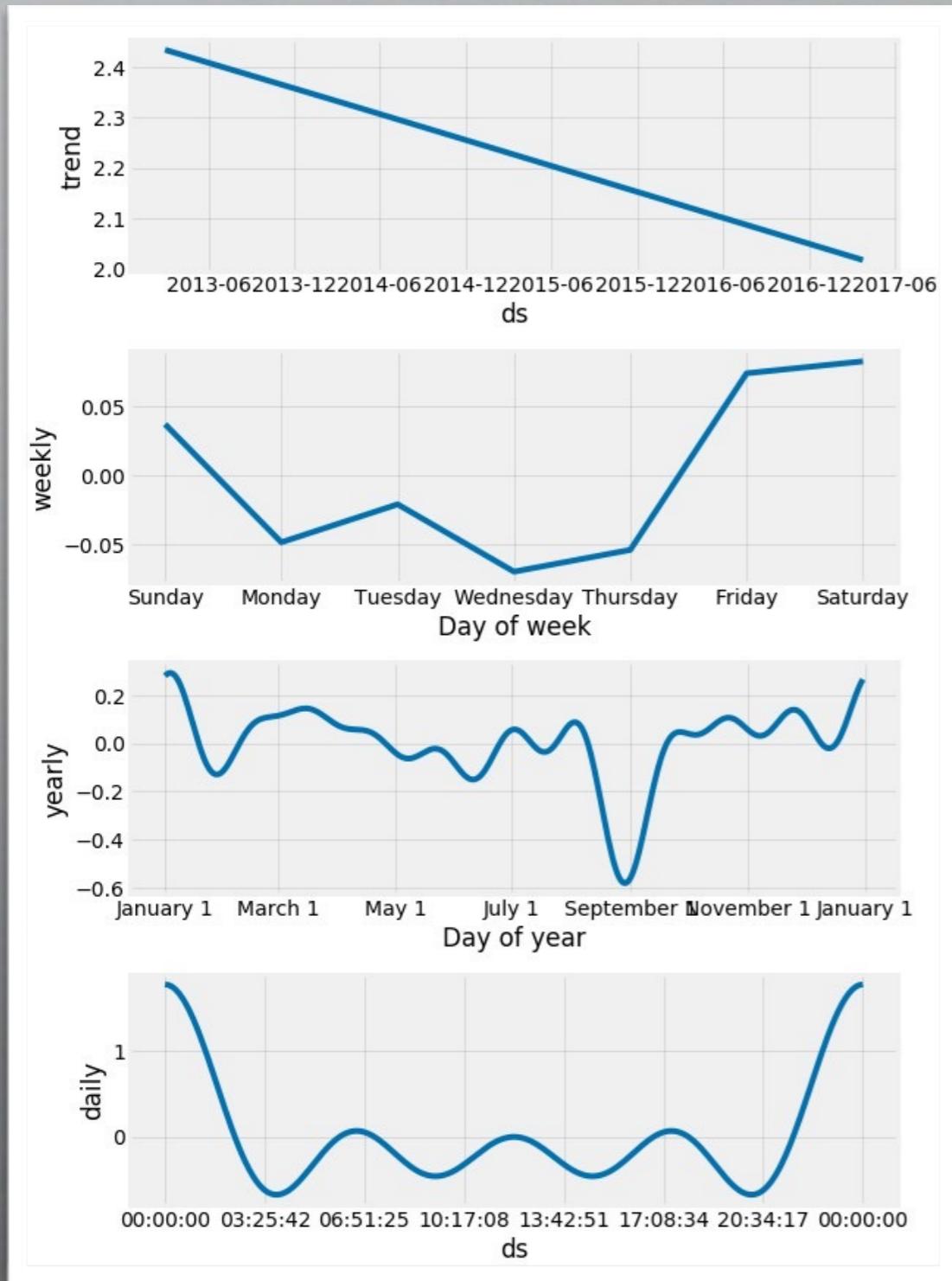


Image Ohio State



FURTHER STEPS

- EXPAND USE OF FACEBOOK PROPHET'S FEATURES
- INTRODUCE RNN AND LSTM AS MODELS TO ENHANCE PREDICTIVE POWERS

Image South China Morning Post

RECOMMENDATIONS FOR SUCCESS

- INCREASING TRANSPARENCY OF TESTING SITE
- KEEPING RESIDENTS INFORMED ON PROGRESS SHOWING AIR QUALITY IS A VIABLE SHORT TERM GOAL
- PUBLIC HEALTH CAMPAIGNS TO ENCOURAGE LESS USAGE OF TRANSPORTATION THAT RELIES ON FOSSIL FUELS
- AFFORDABLE PUBLIC TRANSPORTATION OPTIONS THAT RUN ON CLEAN ENERGY
- USING THE DATA PROVIDED BY THE SITES TO TRACK INCREASES IN SMOG CAUSING POLLUTANTS WOULD HELPING ENVIRONMENTALISTS RECOGNIZE INCREASING POLLUTION TRENDS AND RESOLVING PROBLEMS QUICKLY.

SOURCES

SOUTH CHINA MORNING POST

UCI MACHINE LEARNING LIBRARY OUR WORLD IN DATA