

Gretta Grettas

Team 8: M. Yarkın Ergin, Plamen Kamenov, Ghanashyam Khanal, Duygu Yılmaz

A quick look at pollutants and temperature since the lockdown

<https://github.com/yilmazduygu/gretta-grettas>

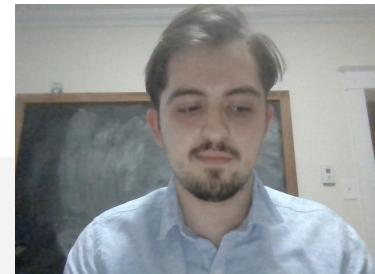


Was it chillier? (then, not now)

Observation: After the start of the lockdown procedures throughout the world due to COVID-19, posts showing natural life slowly returning to environments where humans usually frequent, began to crop up on social media. Some people claimed that this could be because of quarantining which led to a decline in activities that pollute the environment.

Questions: Did the decreased levels of pollutants in the air affect the temperatures we feel and could the current levels of pollution be used to better predict the current temperature?

Goal: Conduct a time series analysis of historical temperature and pollution data to see how the two relate to each other.



Our Approach and Teamwork

Data gathering and cleaning: We pulled temperature data from Kaggle using SQL, and pollution data from EPA [PK, GK, DY]. We manually cleaned duplicates and filled in the missing data with naïve method [GK, PK].

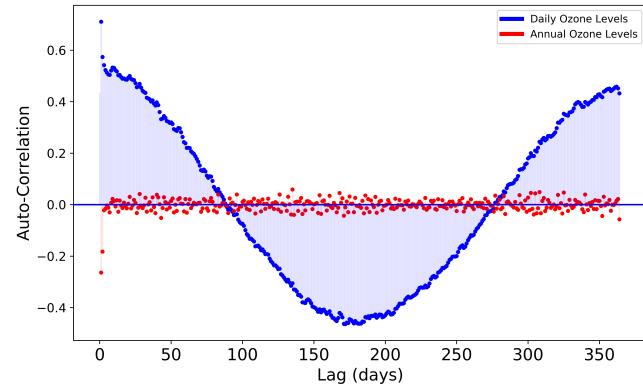
Exploratory analyses: We did autocorrelation plots to determine the period of the data [PK, GK, DY, MYE]. We tried classification analyses for different pollutants and temperature measures from various cities [PK, GK]

Time series analysis: We modeled the temperatures after lockdown with seasonal naïve method [GK, PK, MYE] and a seasonal ARIMA. [MYE, DY]

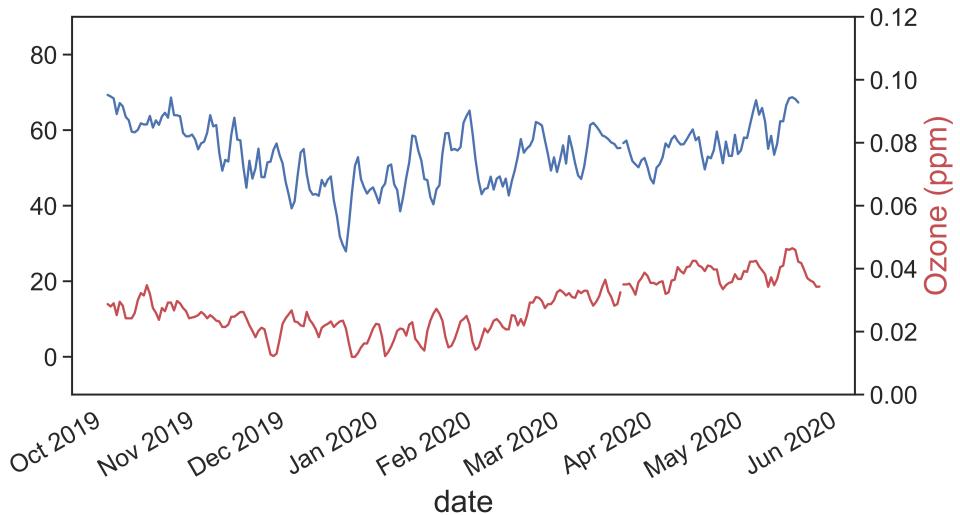
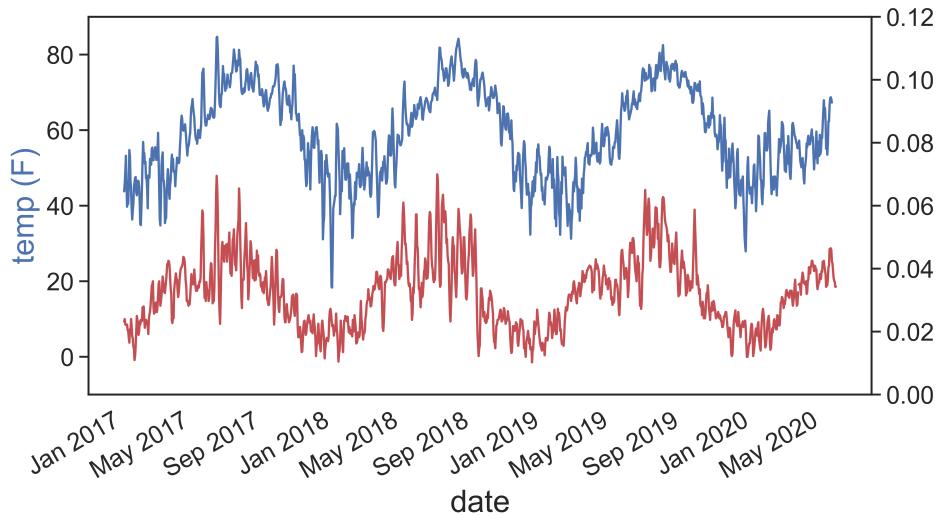


Methods/Packages Used

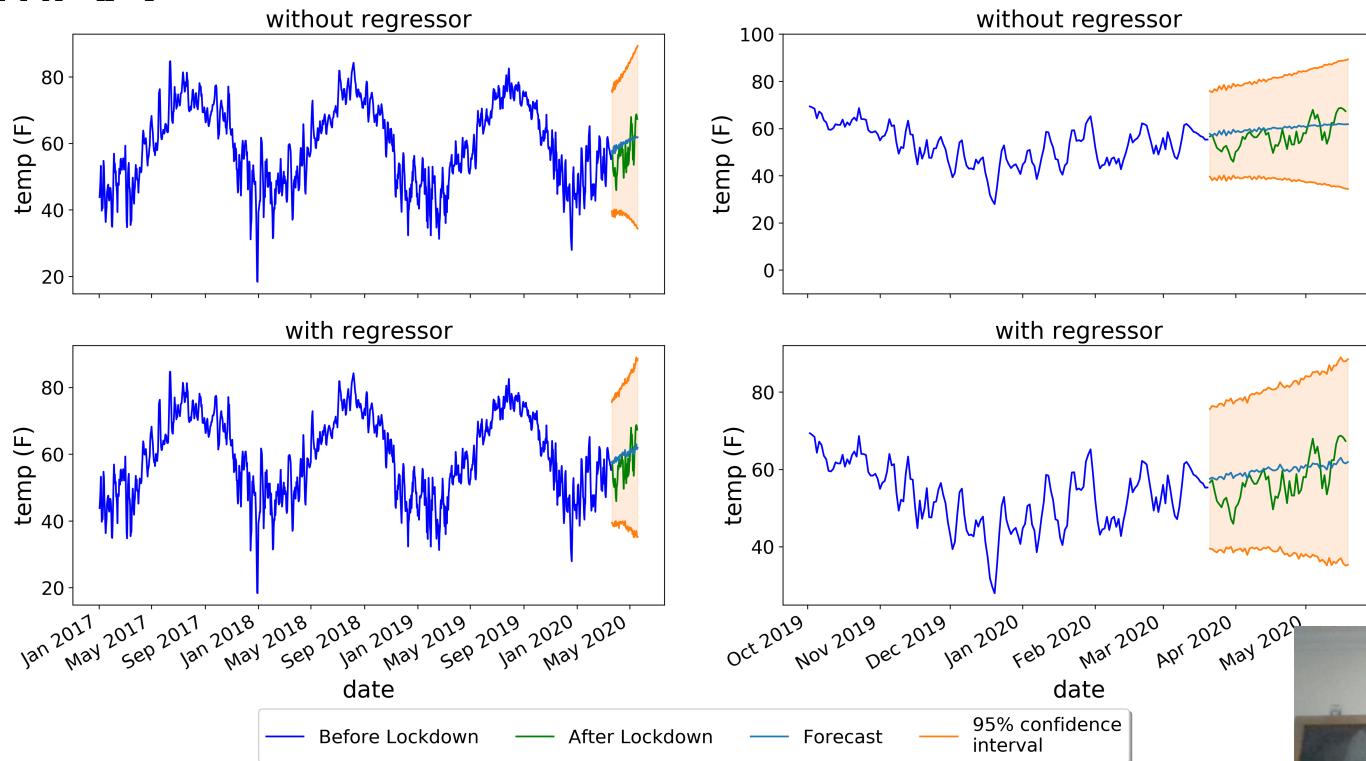
- BigQuery dataset retrieved using SQL
- Duplicates/Missing data were removed/filled manually with naive method
- Time series analysis using R packages: *tseries* & *fpp2* & *forecast*
- Python modules: *seaborn*, *matplotlib*, *pandas*, *sklearn*, *statsmodels* (statistical analysis)
- ARIMAX and Exponential Smoothing methods to identify the underlying pattern of temperature data and its association with pollutant gases



Results



SARIMAX



Challenges

- Our data had many missing data points and duplicates, for time periods and locations

Next Steps

- Using machine learning techniques with a cleaner and bigger dataset to identify broader patterns
- Using other pollutants such as NO₂, CO to improve the model.



Thank You!

TEAM 8: Gretta Grettas



Ghanashyam Khanal
khanalg44@gmail.com



Güzin Duygu Yılmaz
[duyguyilmaz@window
slive.com](mailto:duyguyilmaz@windowslive.com)



Plamen Kamenov
plamen0606@gmail.com



M. Yarkin Ergin
m.yarkin.ergin@gmail.com

<https://github.com/yilmazduygu/gretta-grettas>