STAT 790 Third Meeting Notes

DL

2/25/2022

Things to consider:

- US Census Bureau or ACS data is good, but for now, focus on getting NYC data as general population
    - If we need to go through the community districts, then brush up on the ACS queries
    - NYC Population FactFinder [https://popfactfinder.planning.nyc.gov/#11.59/40.7239/-73.8894]

Goal for next meeting:

- Look for seasonality with a further plot for each borough

    - look for trends
    - reference ffps [https://otexts.com/fpp3/seasonal-plots.html]

- Begin a preliminary forecast

    - Create a subset of the DSNY_thrid dataframe, to include the Bronx
    - Make sure your dataframe is turned into a tibble
    - Create a model with auto.arima()
    - Compare the metrics (MAE, RMSE, MAPE)
    - This is done with auto.arima()

- Summarize your results
- Then, if time permits, explore an ARIMAX model

    - For example, regress the Bronx refuse onto economic index, total population, etc. . .
    - Explore adjustments: [https://otexts.com/fpp3/transformations.html]
    - Explore models with adjustments and without adjustments, both for population and economic data

- Grab the data-set for 2021

    - To compare the forecast values with the actual collected tonnage

- Send the professor your summary (copy+past results) by next Fri morning