

AIR QUALITY INDEX (AQI) PREDICTOR

SPRINT 2 - PRELIMINARY
MODELING

BROOKE HALL

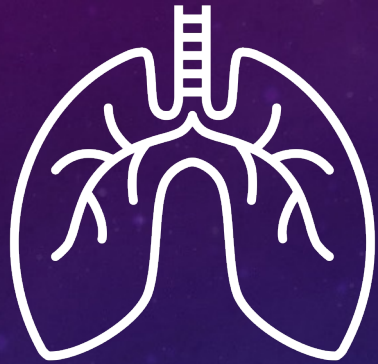
APRIL 2, 2024

AIR QUALITY MONITORING



- 15-20 breaths per minute
- 20,000 breaths per day
- 8,000,000 breaths per year

AQI MONITORING



Problem

Respiratory Sensitivities are difficult to manage



Solution

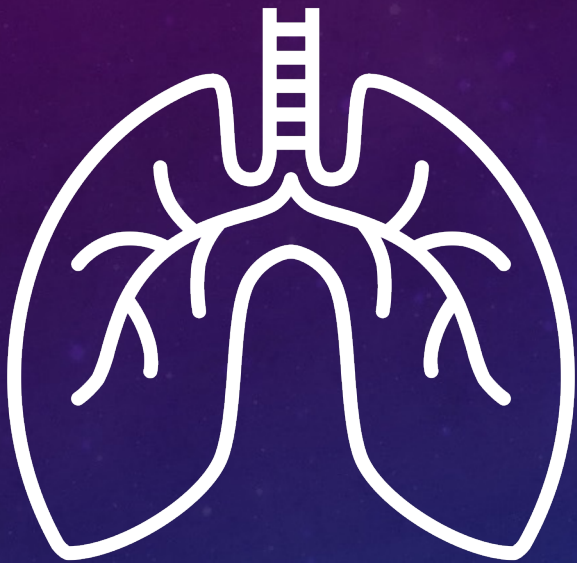
US State AQI Predictor



Impact

Freedom, control

DATA PREPROCESSING



Additional:

- Scope – United States
- Monthly



AQI Monthly Dataset:

- Daily Sentiment
- Largest parameter contribution (CO₂, NO, Ozone, PM)
- Variance Threshold

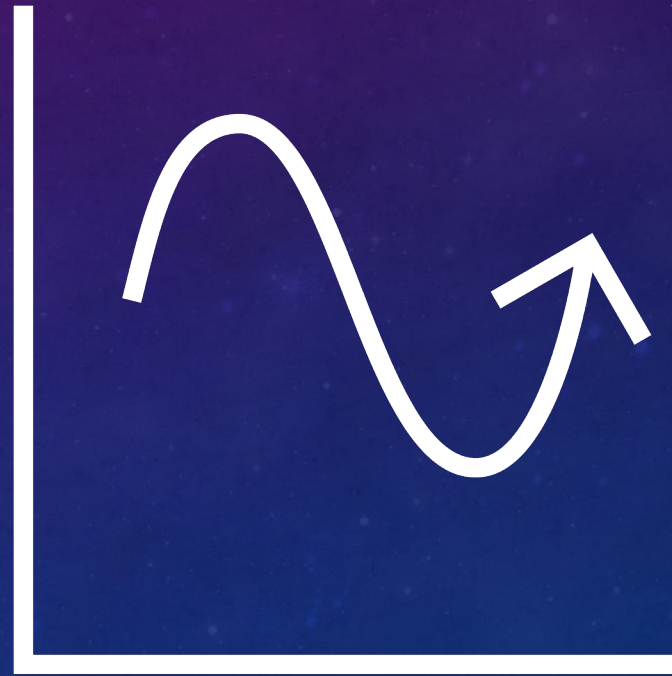


Parameter Monthly Dataset:

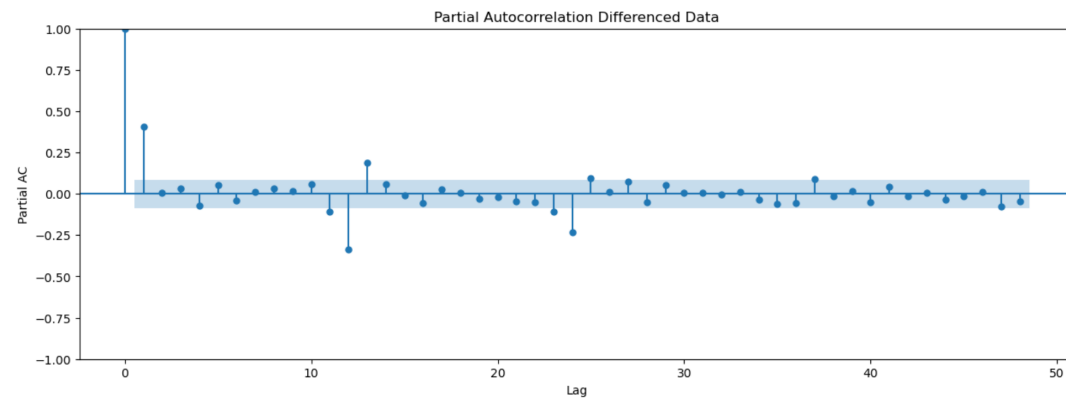
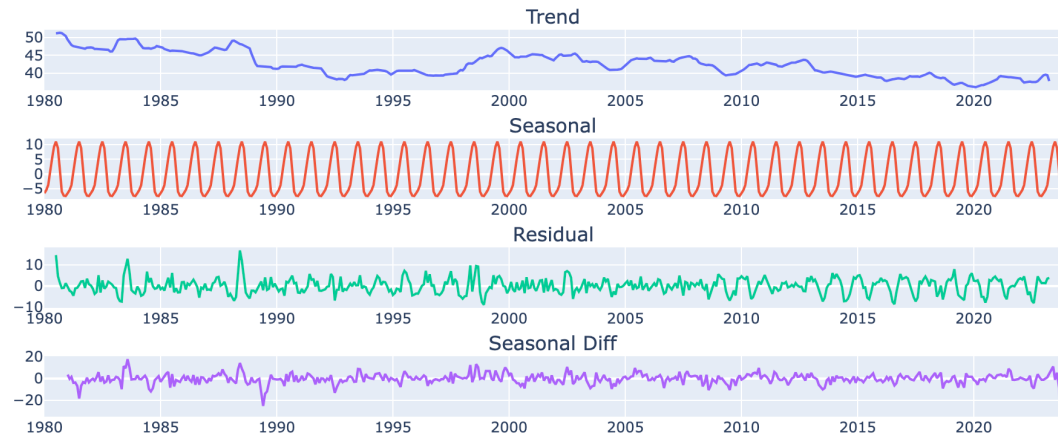
- 200+ Parameters
- fillna()
- Parameter Reduction

EDA

- Correlation
- Monthly Behavior
- Decomposition Seasonality Review
- Changepoints
- By State



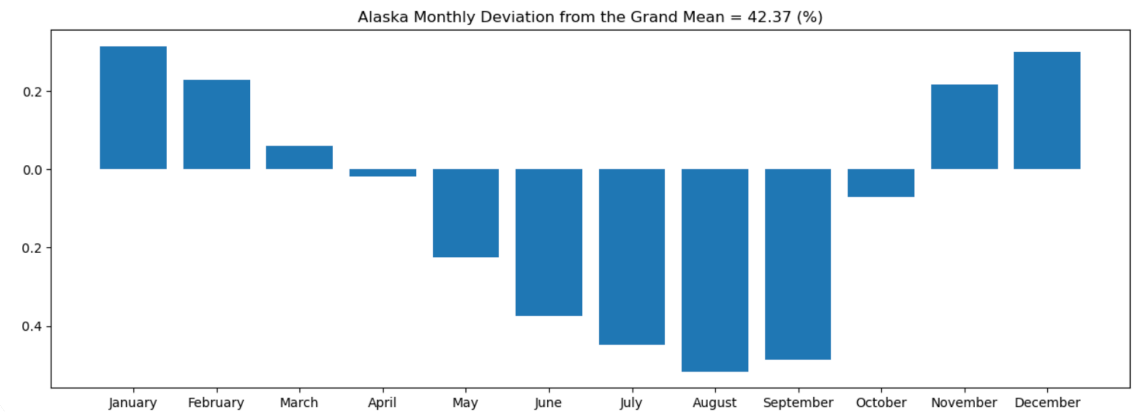
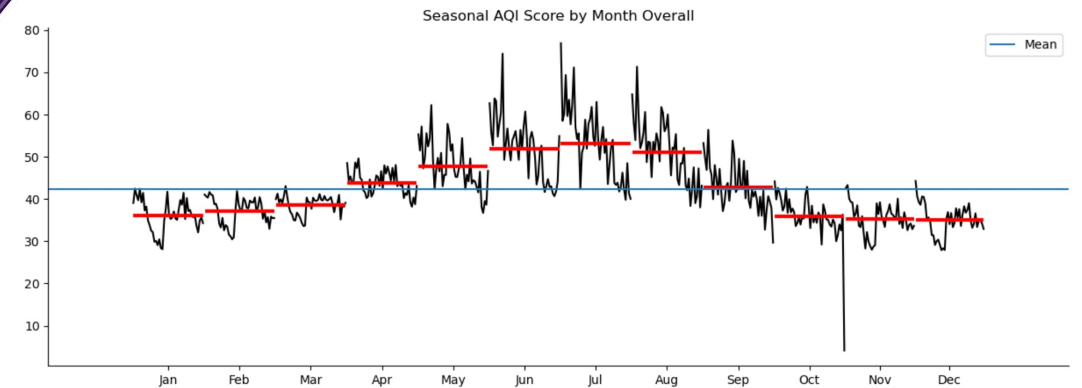
EDA



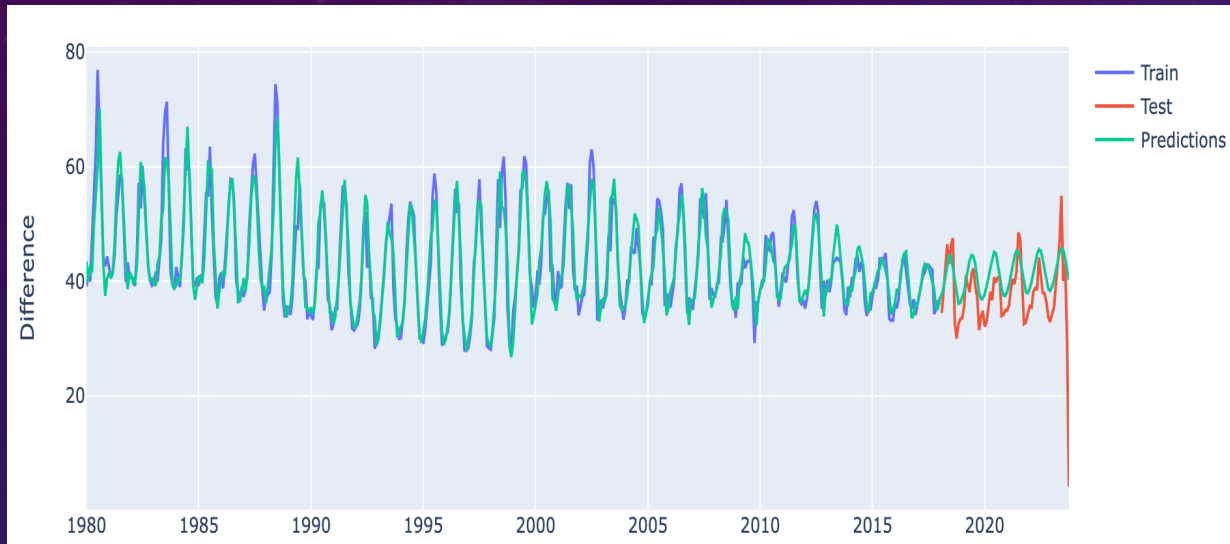
- Univariate Seasonality Review
- Stationary Difference Data
- Significance 1, 2, 12, 13, 14, 24, 25, 26

EDA

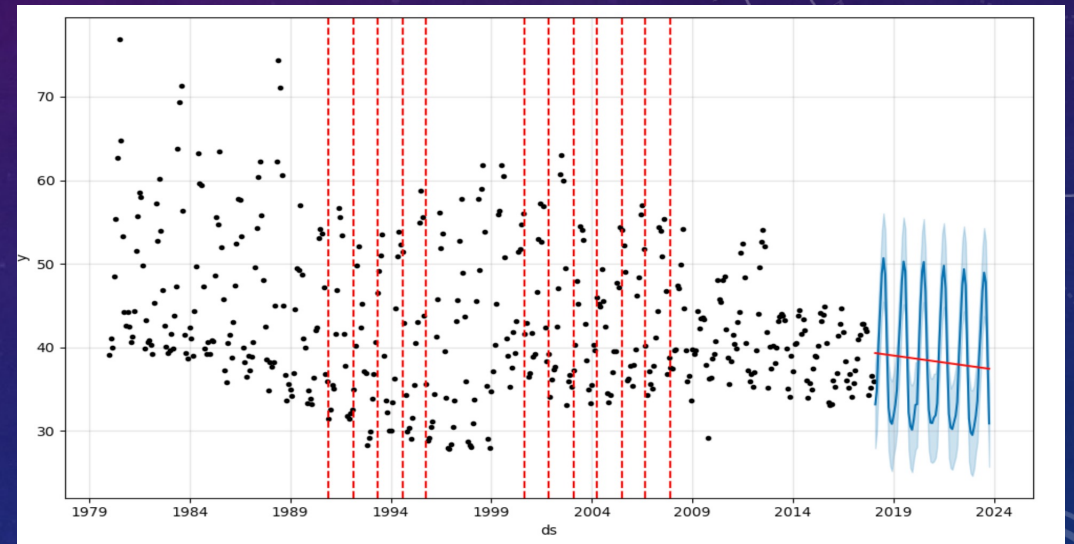
- Average AQI Higher in summer
- Inverted for some states
- Cyclical in all cases



MODEL



SARIMAX



Prophet

HYPERPARAMETER OPTIMIZATION

```
param_grid = {  
    'seasonality_mode': ['additive', 'multiplicative'],  
    'changepoint_prior_scale': [0.01, 0.1, 0.5],  
    'seasonality_prior_scale': [1, 10, 30]}
```

state	r2	mae	mse	seasonality_mode	changepoint_prior_scale	seasonality_prior_scale	r2_optimized	mae_optimized	mse_optimized
Overall	0.164	4.39	33.58	multiplicative	0.01	1.0	0.231	3.85	30.89
Alabama	-1.24	7.13	69.17	multiplicative	0.01	1.0	-0.750	5.87	53.95
Alaska	-0.827	8.88	105.52	multiplicative	0.10	1.0	-1.66	10.47	153.41

NEXT STEPS

- Feature Engineering – More parameters
- Multivariate – Vector Autoregressive Model (VAR)**
- Random Forrest/XG Boost/Tensor Flow

