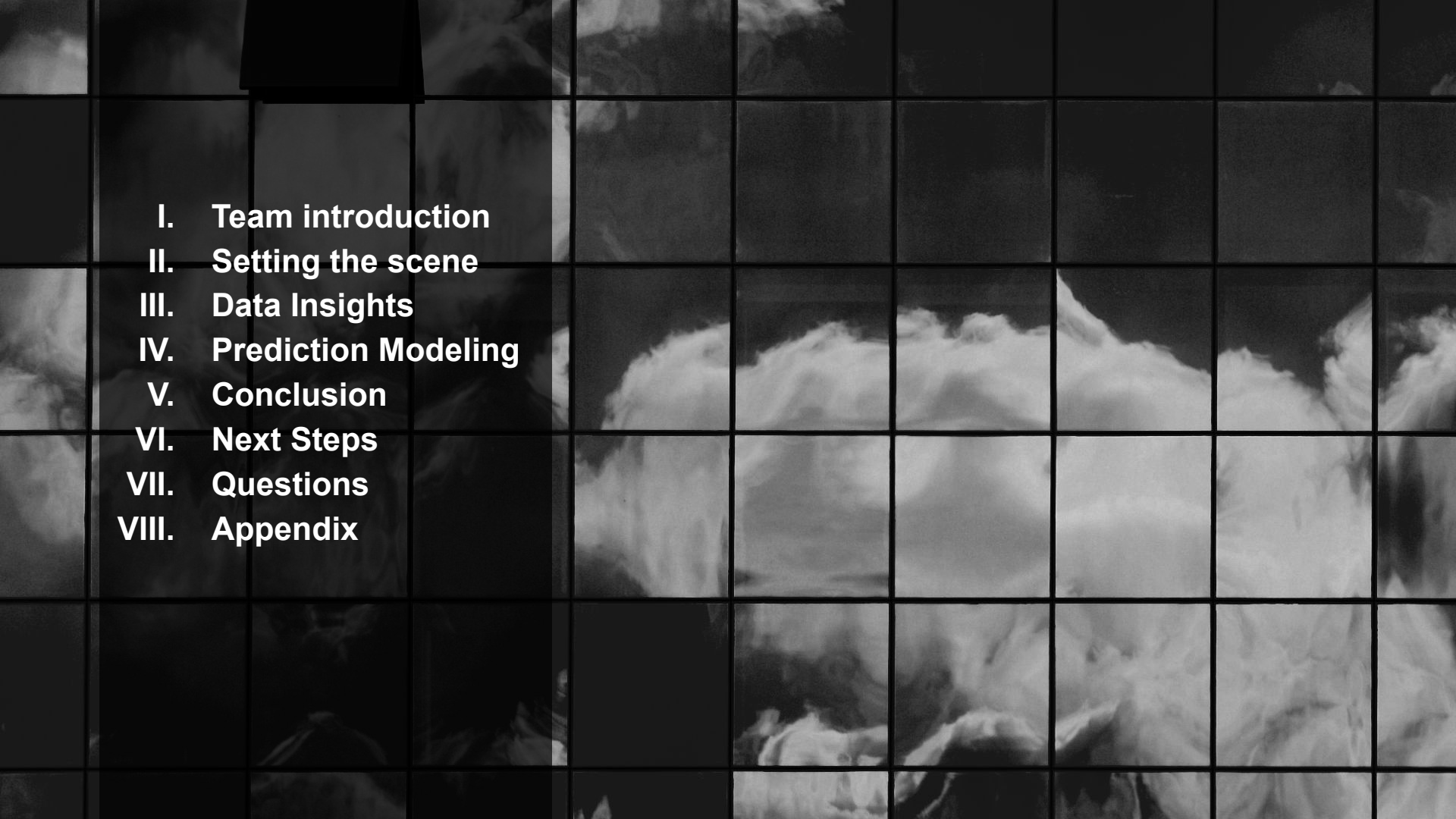
AIR QUALITY FORECAST



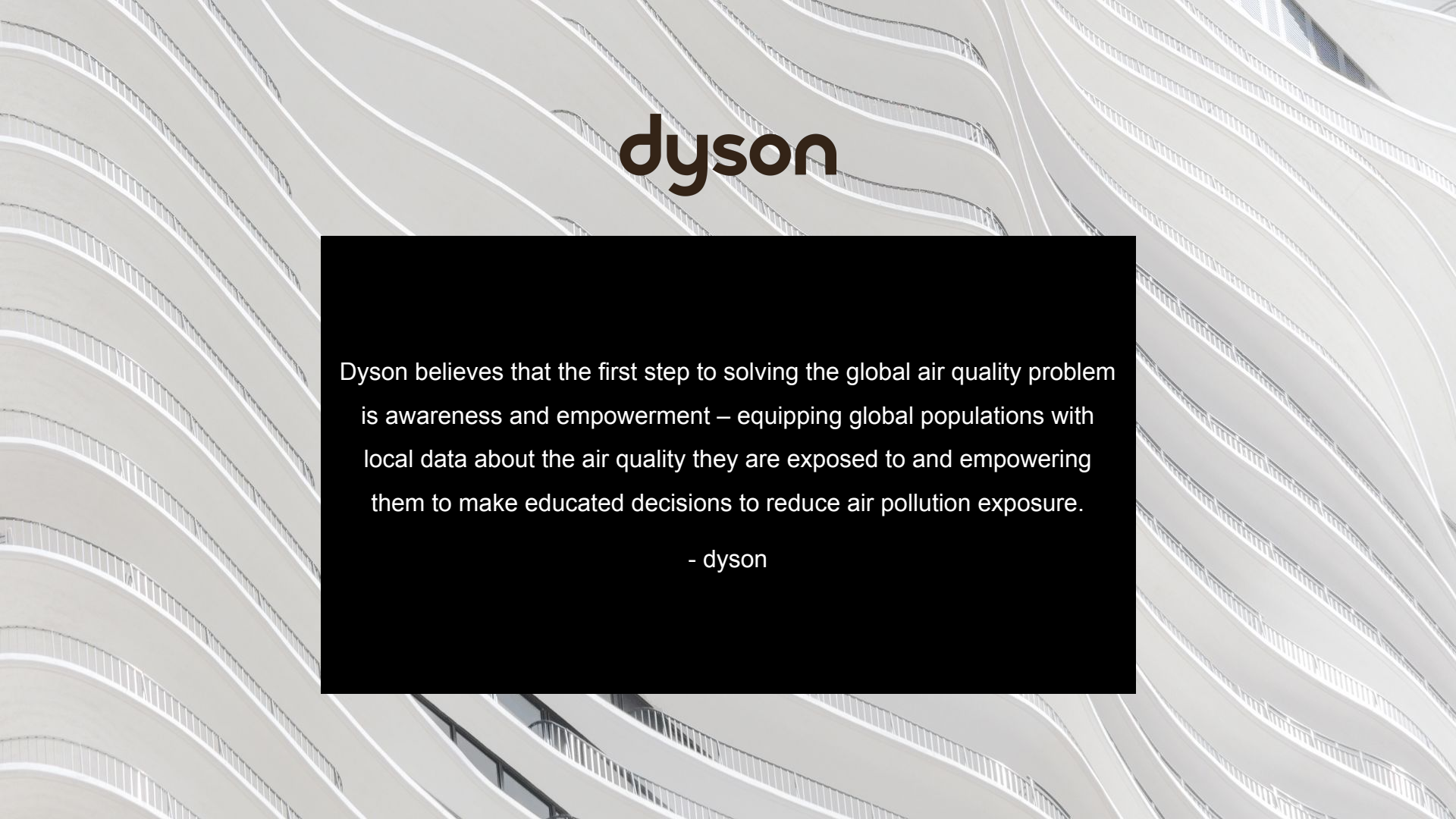
I. Team introduction
II. Setting the scene
III. Data Insights
IV. Prediction Modeling
V. Conclusion
VI. Next Steps
VII. Questions
VIII. Appendix

TEAM

Artem Lukinov
Data Scientist

Mackenzie Dowling
Data Scientist

Q Ghaemi
Data Scientist


dyson
Dyson believes that the first step to solving the global air quality problem is awareness and empowerment – equipping global populations with local data about the air quality they are exposed to and empowering them to make educated decisions to reduce air pollution exposure.
- dyson

The next era of air purification systems

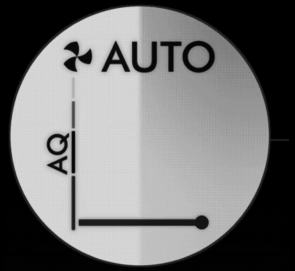
Dyson Pure Cool 007

- Predicts next 7 days of air-quality
- Empowers consumer with knowledge to prepare
- Acts as built-in marketing initiative to increase Dyson filtration sales
- Increase global awareness of air-quality levels

The Golden State is the most highly populated state in the U.S. with the fifth largest economy in the world.

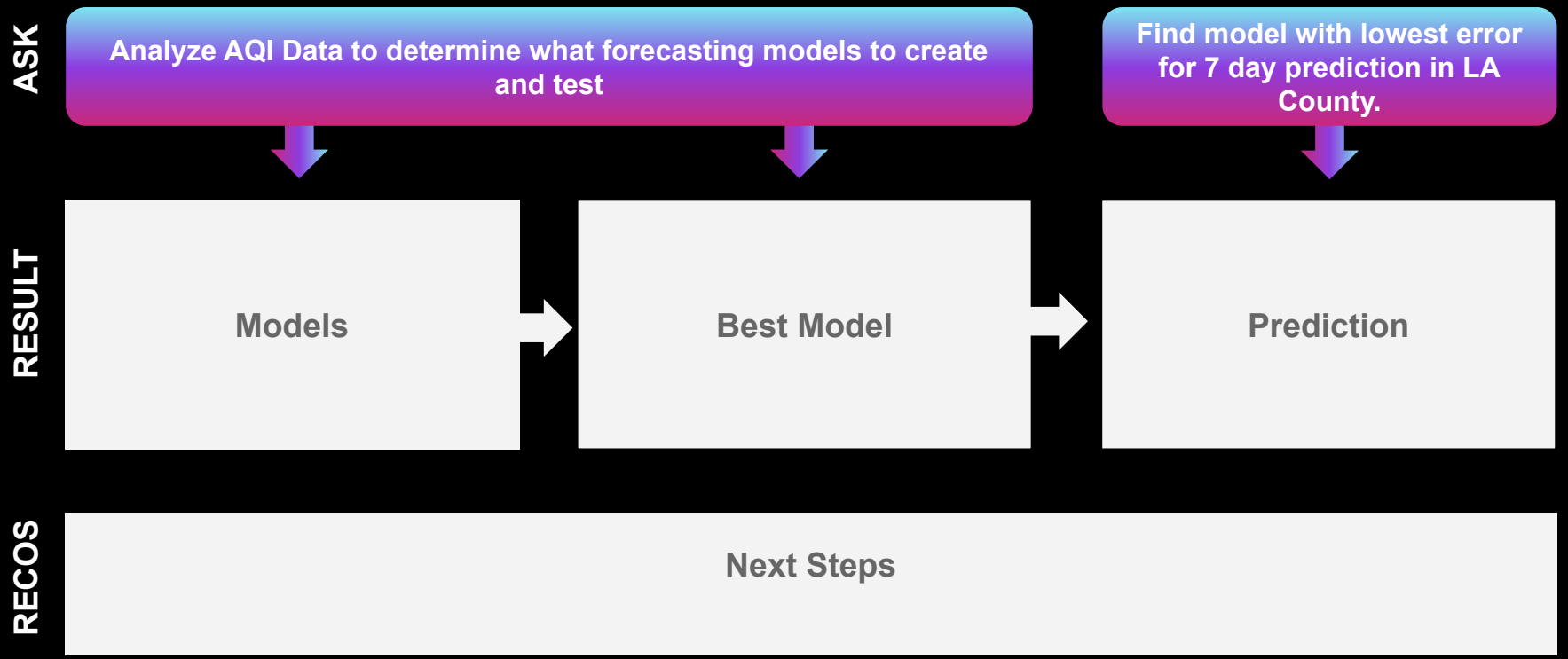
6/10

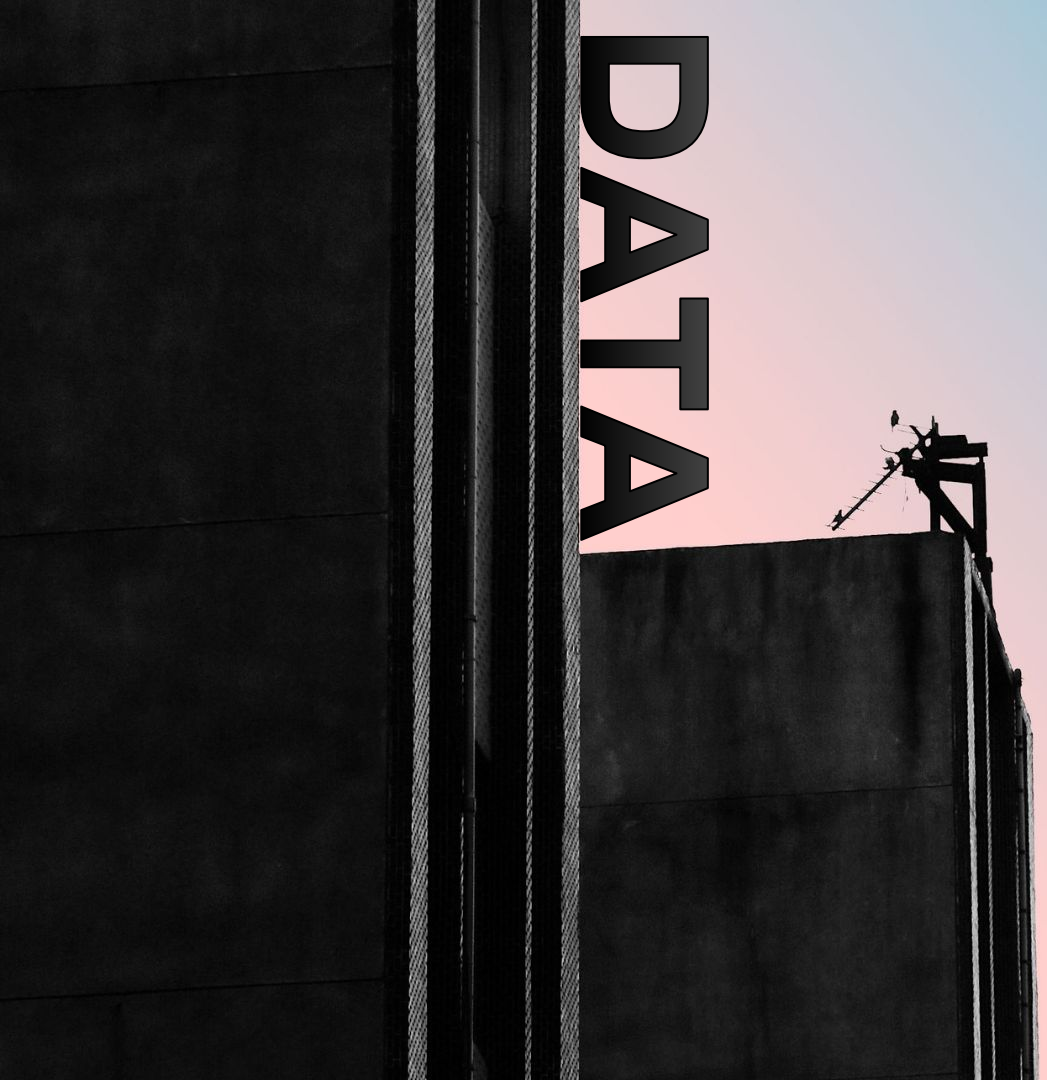
Six of the top 10 most polluted cities (for year-round particle pollution) in the US are from California.



In order to integrate this technology into the **Dyson Pure Cool 007** product we have analyzed the largest county relative to population and AQI index in California to create a model that forecasts AQI with the lowest error. This model will provide updates notifying consumers in the county 7 days out, should it forecast an AQI increase in order to help consumers better prepare for poor air quality.

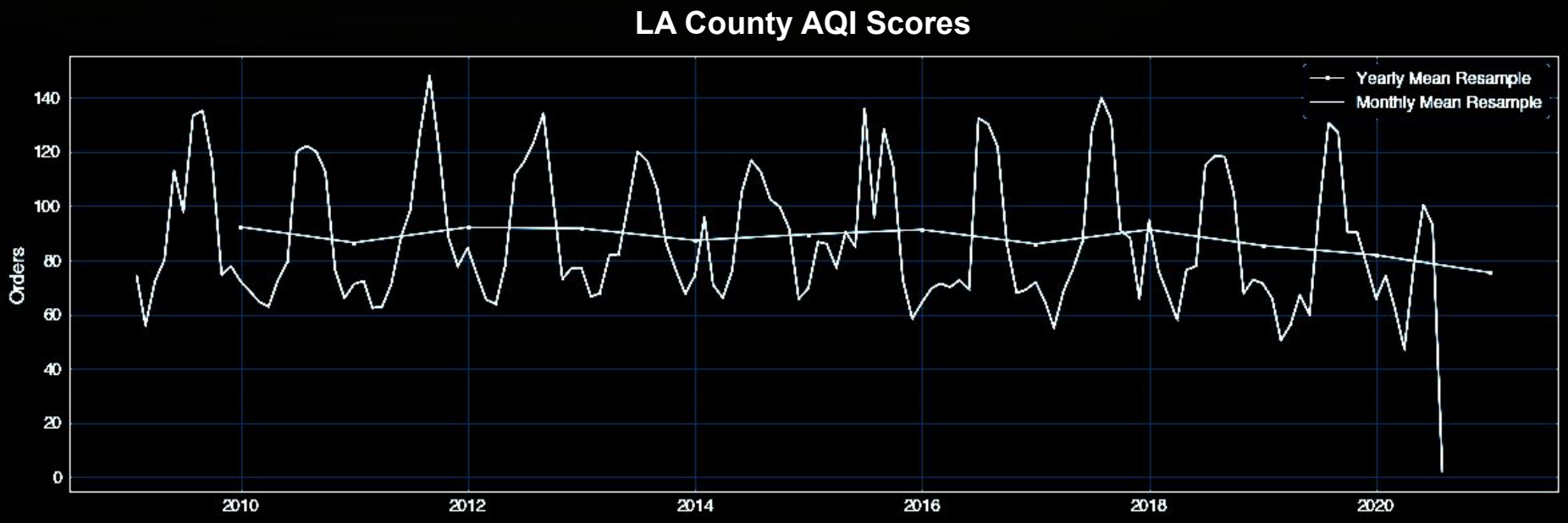
Objectives


ASK
Analyze AQI Data to determine what forecasting models to create and test
Find model with lowest error for 7 day prediction in LA County.
RESULT
Models
Best Model
Prediction
RECOS
Next Steps

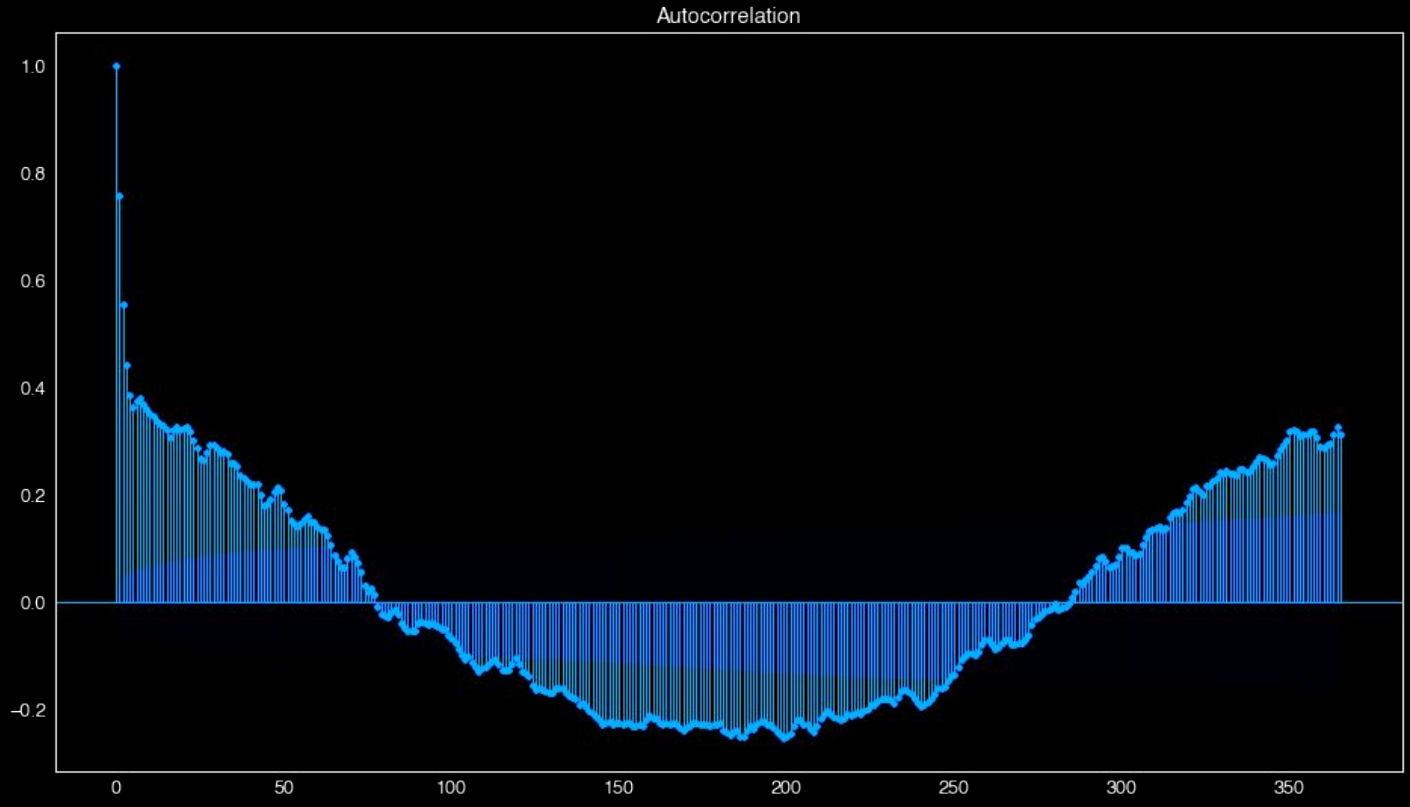


- The data source - EPA's web portal
- AQI score by county for state of California
- daily AQI measurements spanning across January 2009 to July 2020
- 50% of California population live in 5 counties and those were identified as the target for our models

AQI Scores Data is not trending



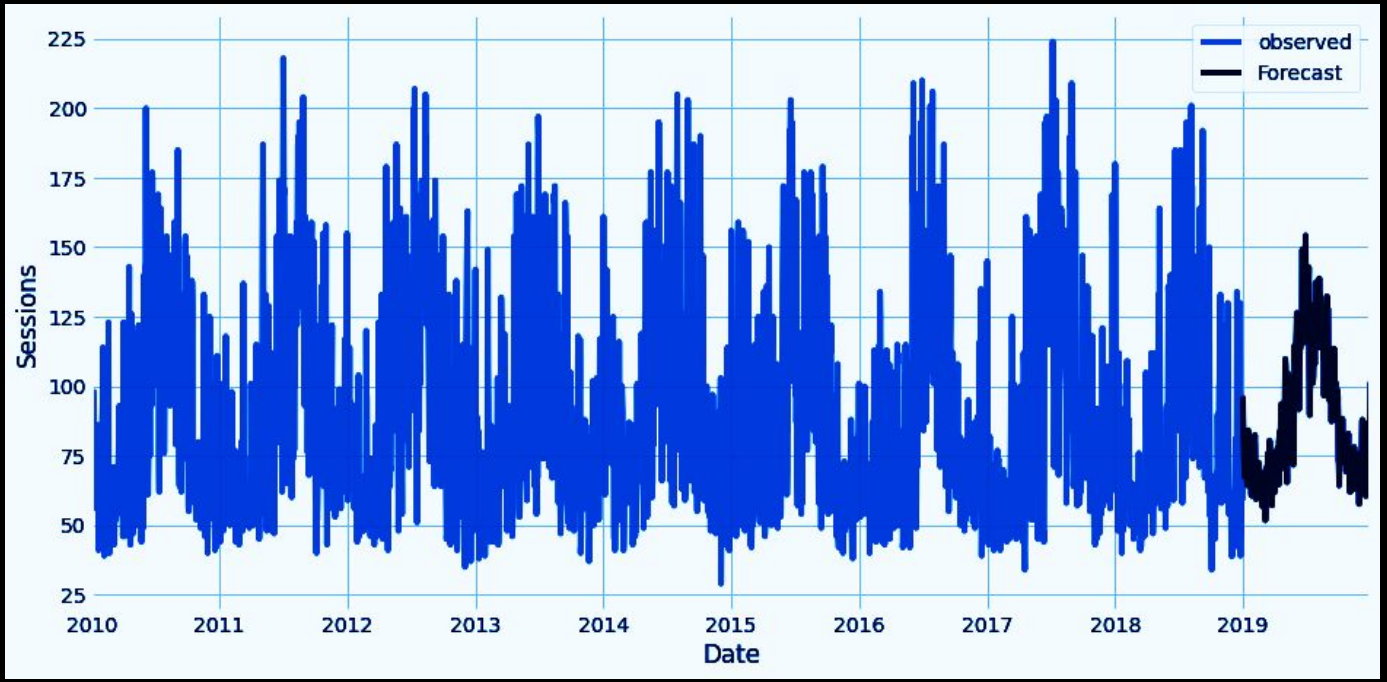
Seasonality detected annually



MODEL ERROR SCORES

Model	7-day error score (RMSE)	Overall score (RMSE)
AR	16.70	36.42
ARIMA	19.15	37.49
SARIMA	17.09	31.90

SARIMA FORECAST



Executive Summary

ASK

Analyze AQI Data to determine what forecasting models to create and test

Find model with lowest error for 7 day prediction in LA County.

RESULT

The AQI Data showed no trend, seasonality, and autocorrelation leading us to three initial model to test.

Best Performing Model based on error: SARIMA

Using SARIMA model, with a 95% confidence we can predict AQI in Los Angeles for 7 days out.

NEXT STEPS

1. **Tune SARIMA model for best RMSE**
2. **Build out individual models for each county in CA**
3. **Build out individual models for each county in the country**


QUESTIONS