

Prediction and Visualization of Air Quality Index for US Regions

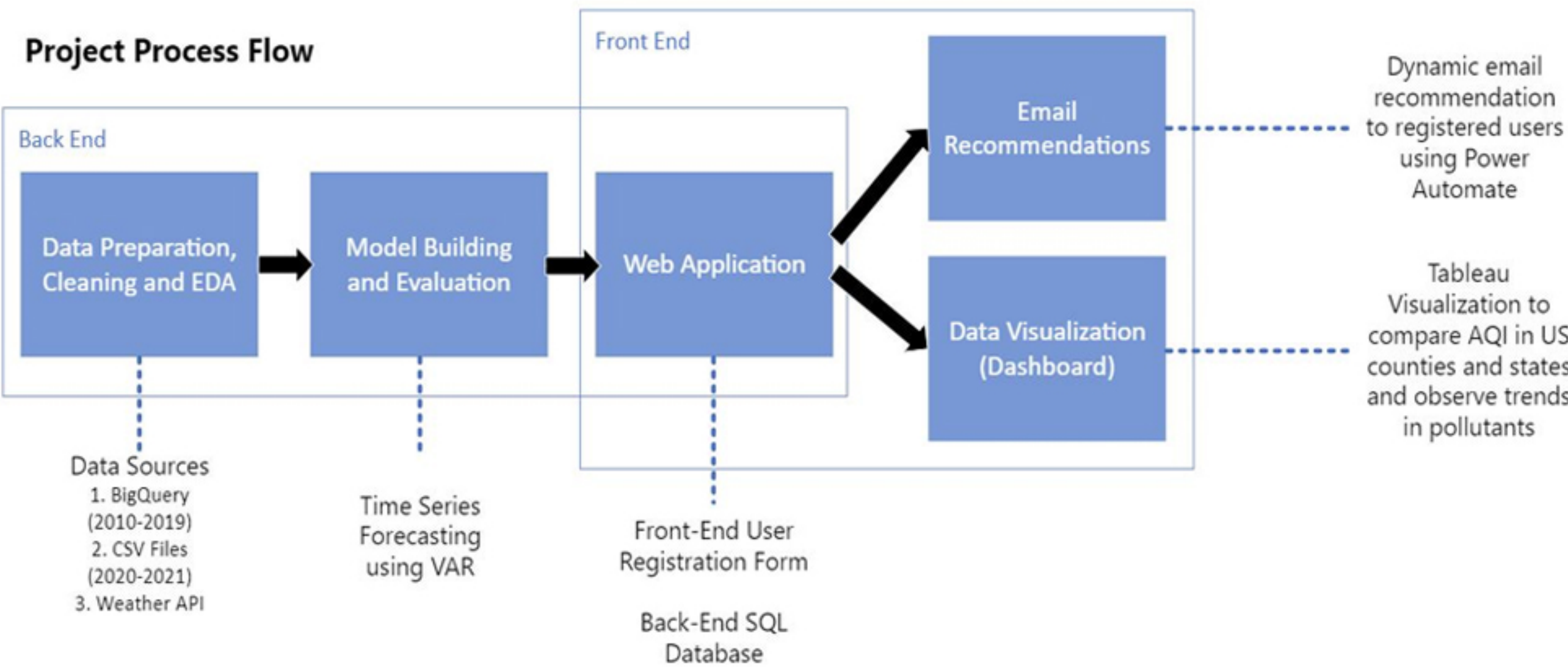
Abhishek Gawande, Ayush Gordhan Agarwal, Ajay Krishna Manoj, Harshal Mittal, Nicholas Ramallo, Sandeep Sainath

Project Summary

The aim of our project is to predict the concentration of different pollutants in the atmosphere, compute AQI across different US regions, and present the trends using interactive visualizations. We aim to provide personal health recommendations directly to the user's inbox.

We see two main benefits of this: users can take preventive measures to stay healthy (prevent allergy and asthma attacks) based on adequate preparation through our recommendations, and government bodies can take effective and quicker decisions based on predicted air quality and pollutants in the atmosphere.

Approach



List of Innovations

- End-to-end web app/visualization product that deliver personalized recommendations to individuals.
- Time series forecasting model that uses K-nearest neighbors to understand geospatial effects in conjunction with time series modeling on pollutant concentrations to predict AQI for each region.

I. Data Sources and Preparation

We are using two main sources of data to compile the last 11 years of data for six key atmospheric pollutants and a weather API:

- BigQuery database (2010-2019)
- Another official dataset (2020)
- Weather API (OpenWeatherMap)

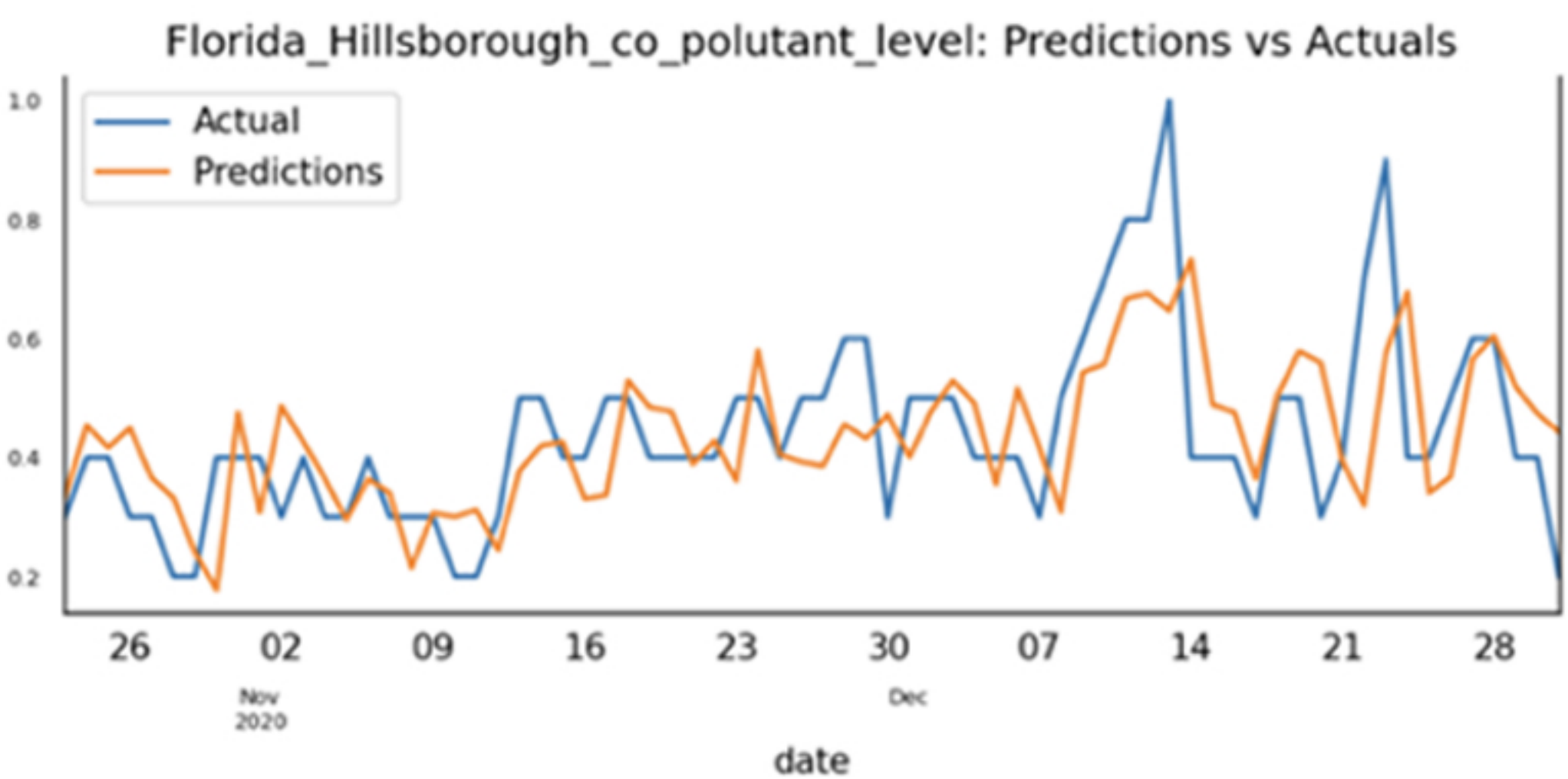
Data preparation included the following tasks:

- Data cleaning and merging
- Interpolation of missing values
- Rolling up data to county level from site level
- Establishing data quality metrics

II. Modeling

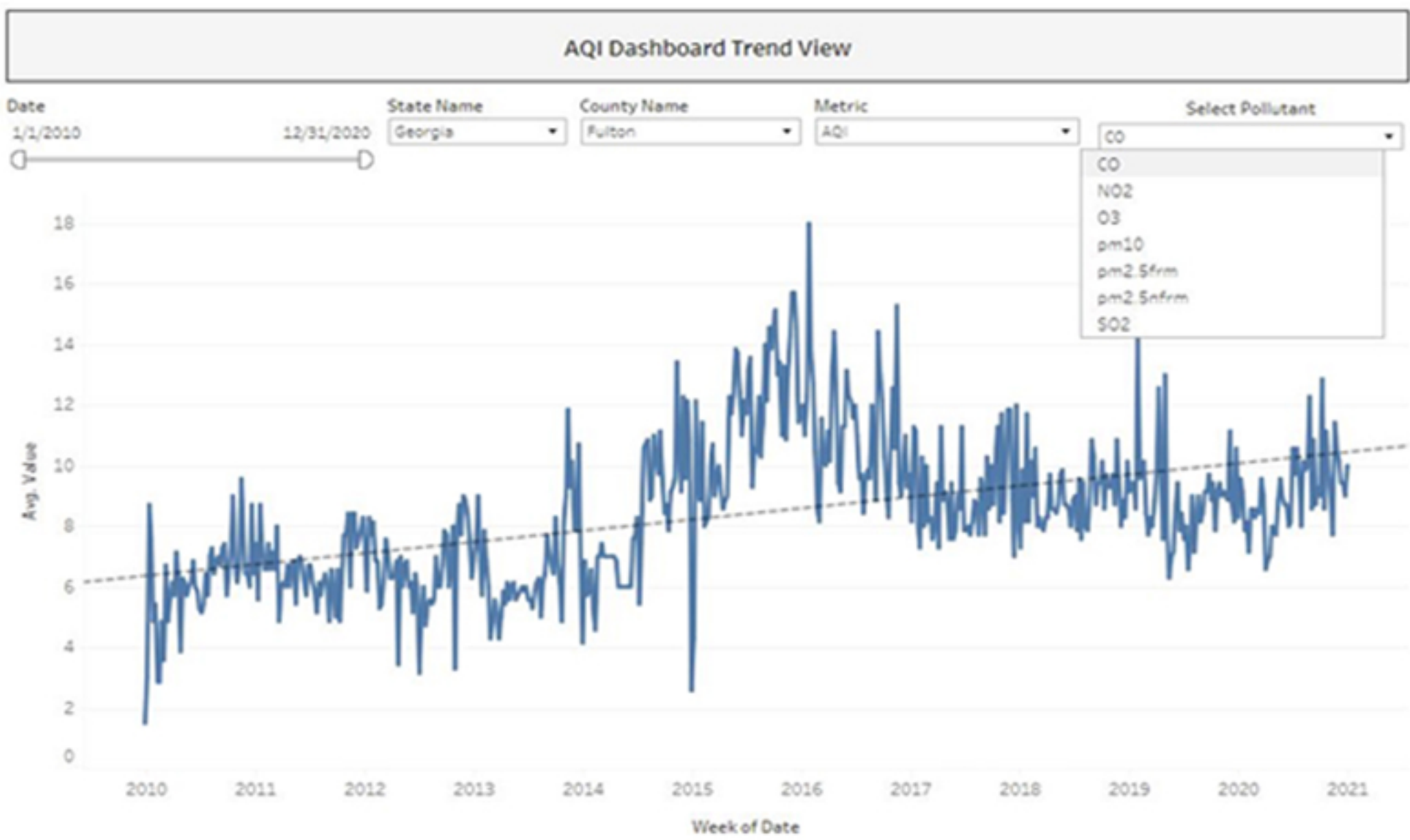
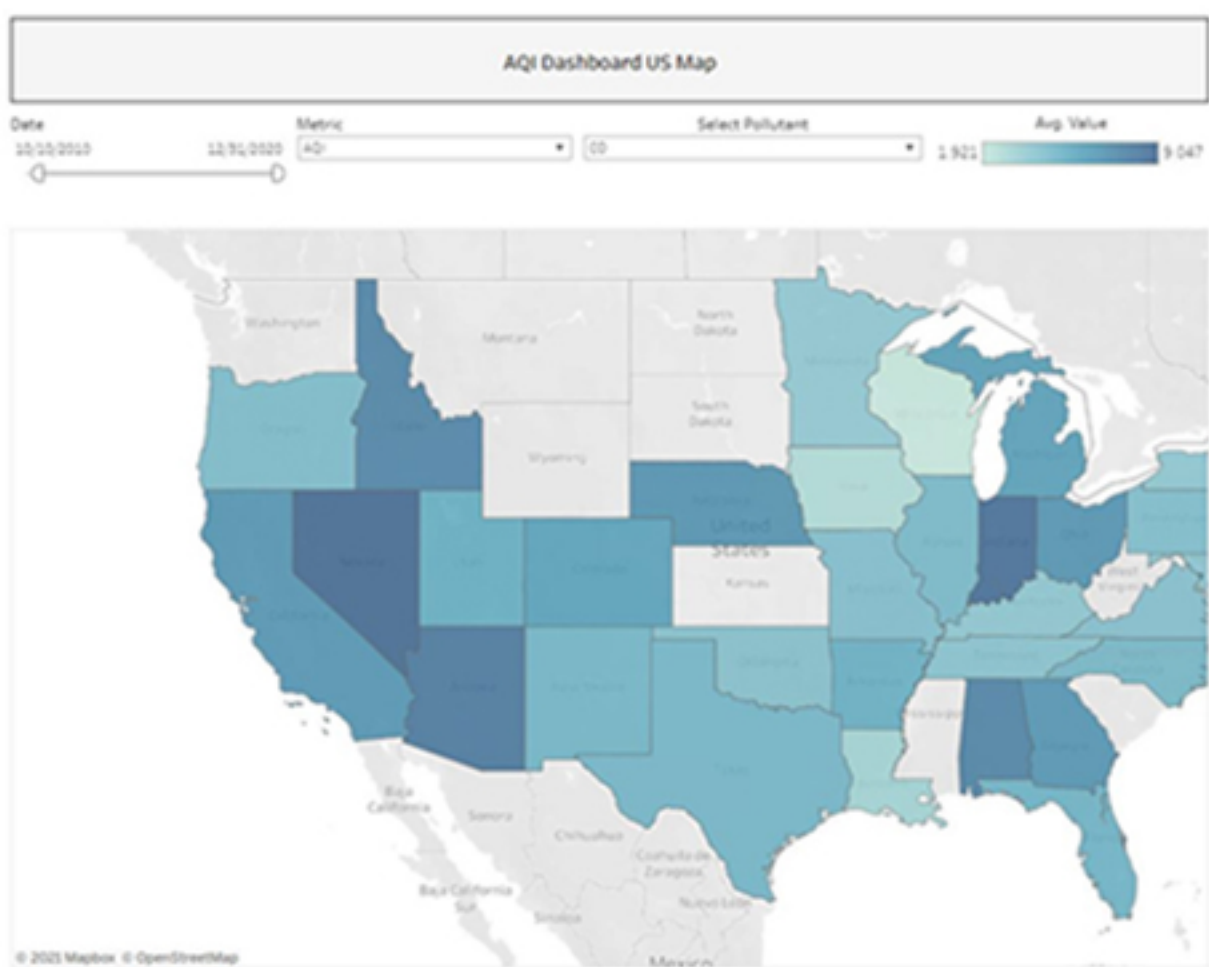
We used a novel time series approach where we considered multiple features such as pollutants, pressure, humidity, wind and temperature for 5 nearest neighbours for each county.

We developed 100 models, one for each county using VAR with optimized lag. These models forecasted AQI values for December 2020.



III. Data Visualization

We are using Tableau to build a dashboard that shows the pollutant levels across the states as well as the predicted values and trends of pollutants across the years. The user can filter the pollutant, the date range and county name as per the user requirement.



IV. Email Recommendation System

We are leveraging the Microsoft Power Automate platform to send out daily recurring emails to users based on the location they have signed up for and the existing health conditions they have.

Your 2020-12-2 Personal Recommendation is here!



Your personal health recommendation is here!

Today, weather is Clouds and the AQI is 66 in Fulton, Georgia. Recommendations based on your pre-existing health conditions is shared as follows:

Pollution Level :	Moderate
Recommendation:	Unusually sensitive people should consider reducing prolonged or heavy exertion.
Worst Pollutant :	pm25

V. Web App and User Registration

We are using Flask as our framework for the web app which consists of one page containing the registration form and a "Register" button prompting the user to fill in their registration information for email updates.

The Tableau dashboard is embedded into the web app.

The link to the web app is:
<https://dva-app-test.onrender.com/>

WELCOME TO OUR APP! PLEASE REGISTER

Registration form fields:

- Username
- Email
- State (Alabama)
- County (Ada)
- Do you have any existing chronic/serious respiratory conditions? (Yes)
- Do you have any existing chronic/serious heart conditions? (Yes)
- Do you have children or elderly people around you on a regular basis? (Yes)
- REGISTER button

Conclusion

Our novel time series forecasting models predict the pollutant concentration and AQI accurately for most counties and across various pollutants. This project thus provides an end-to-end scalable framework for a personalized health recommendation system, delivered to registered users via email. The companion web app includes interactive dashboards to visualize AQI forecasts and trends by region to further enforce the urgency and importance of monitoring air quality.