The Impact of Bike-Sharing Ridership on Air Quality: A Scalable Data Science Framework

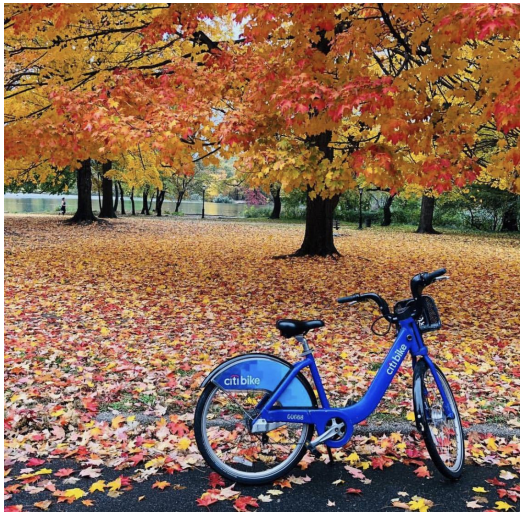
Nina Hua, Rebecca Reilly, Victoria Suarez, Philip Trinh

Mentors: Paul Intrevado, Diane Myung-kyung Woodbridge

USF
Master of Science in Data Science

Agenda

- Intro
- Workflow
- Data
- Experiment Output
- Conclusion



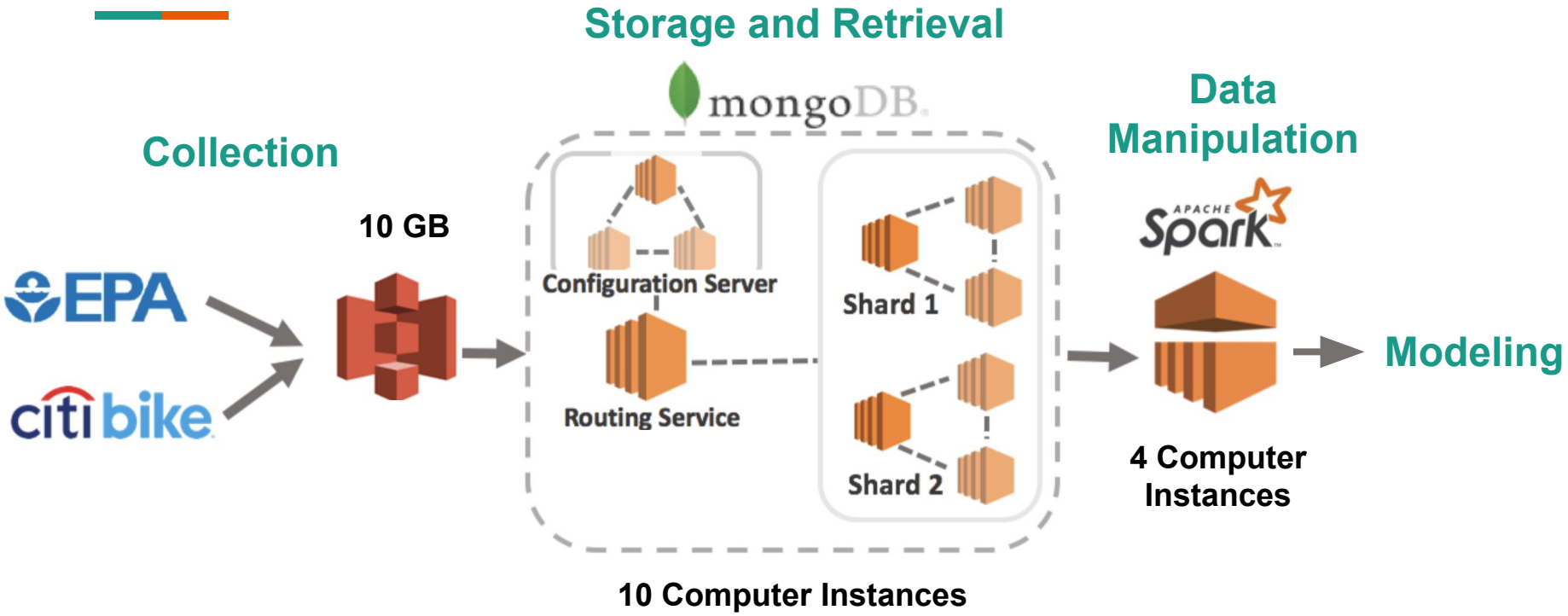
Intro

- We explored the relationship between daily air quality indicator (AQI) values and the daily intensity of bike-share ridership
- Vehicle emissions are a main cause of increased atmospheric CO2
- A rider on a bicycle will generate 80% less emissions per kilometer than a passenger car

Workflow

Amazon Web Services	**MongoDB**	**Apache Spark**	**Python**
Create Computer Instances	Distributed Database	Distributed Computing Framework	Popular Data Science Programming Language
Ease of System Setup	Improved Query Performance	Divide, Compute, Collect	Orchestrate Processes

Workflow


Storage and Retrieval
mongoDB.
Collection
Data Manipulation
10 GB
EPA
citi bike
Configuration Server
Routing Service
Shard 1
Shard 2
APACHE Spark
Modeling
4 Computer Instances
10 Computer Instances

Data

CitiBike (9GB)

- 12,000 bicycle
- 750 stations

EPA (1GB)

- Air Quality Index (AQI)
- 3 years (2016 - 2018)

Air Quality Index - Particulate Matter	
301 – 500	Hazardous
201 – 300	Very Unhealthy
151 – 200	Unhealthy
101 – 150	Unhealthy for Sensitive Groups
51 – 100	Moderate
0 – 50	Good

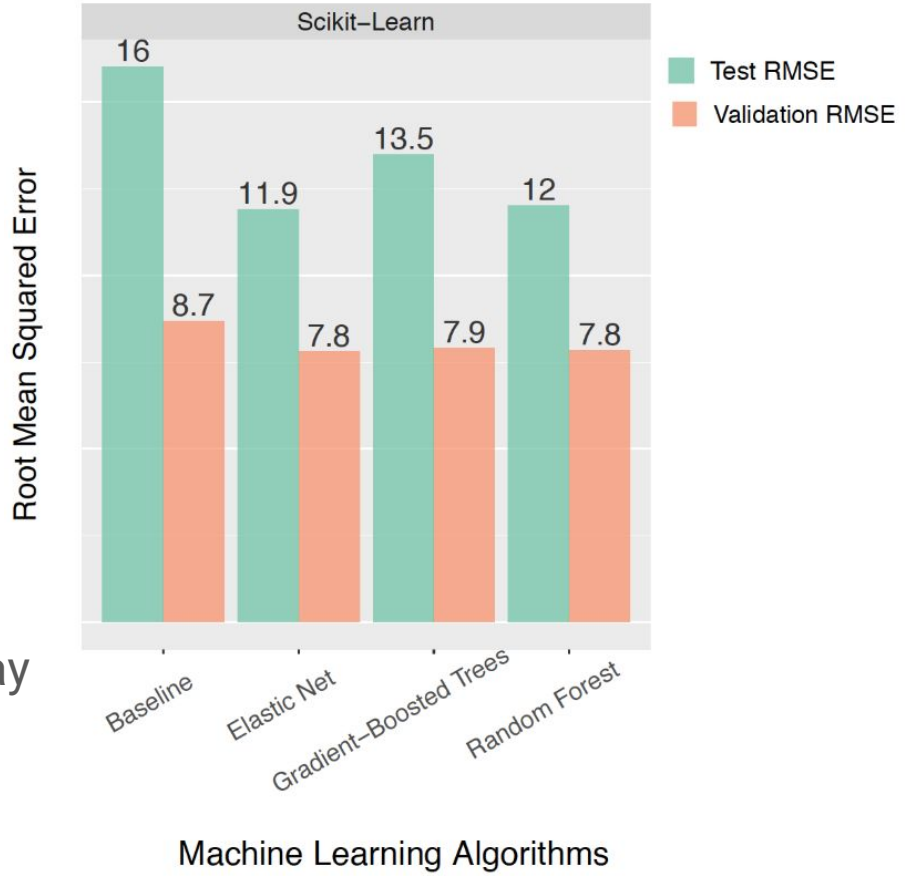
Experiment Output

Algorithms:

- Elastic Net
- Gradient-Boosted Trees
- Random Forest

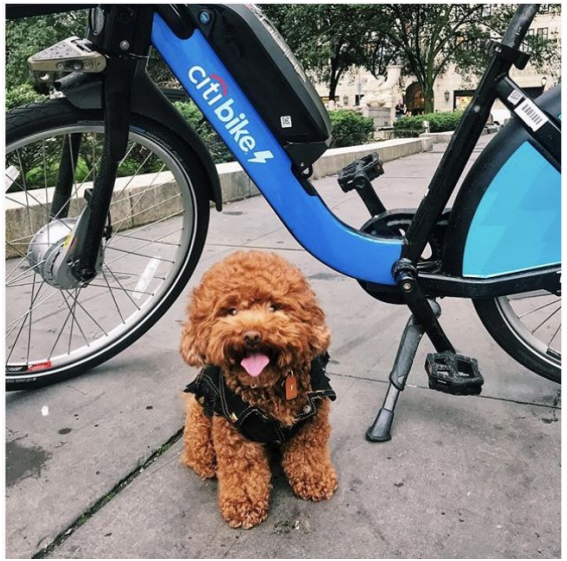
Features:

- Seasonal Indicator
- Previous day's AQI
- Total # of bike-sharing rides per day



Conclusion

- Established inverse relationship between daily AQI and bike-share ridership
- Recommend distributed platforms for pre-processing data
- Encourage continuation of research on the relationship between modes of transportation and AQI



Thank You