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
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

<table>
<thead>
<tr>
<th>Amazon Web Services</th>
<th>MongoDB</th>
<th>Apache Spark</th>
<th>Python</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Computer Instances</td>
<td>Distributed Database</td>
<td>Distributed Computing Framework</td>
<td>Popular Data Science Programming Language</td>
</tr>
<tr>
<td>Ease of System Setup</td>
<td>Improved Query Performance</td>
<td>Divide, Compute, Collect</td>
<td>Orchestrate Processes</td>
</tr>
</tbody>
</table>
Workflow

Collection

Storage and Retrieval

Data Manipulation

Modeling

10 GB

EPA
citibike

Configuration Server
Routing Service
Shard 1
Shard 2

mongoDB

Apache Spark

4 Computer Instances

10 Computer Instances
Data

CitiBike (9GB)
- 12,000 bicycle
- 750 stations

EPA (1GB)
- Air Quality Index (AQI)
- 3 years (2016 - 2018)
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