USC Viterbi School of Engineering

CitiBike Data Analytics

- https://phumphri.github.io/CitiBike/
- https://humphries-citibike.herokuapp.com/

Patrick Humphries, M.B.A.

U.S.C. Graduate Student, Big Data Fundamentals
U.S.C. Graduate Student, Data Informatics

Project Description

CitiBike provides five years of ride history in raw, csv format.
This data is analyzed and presented to the end user in the following steps:

- Data Acquisition
- Data Modeling
- Data Access
- Data Visualization.

Timeframe

The project was completed in two weeks. An existing project was used as a framework, allowing the inclusion of additional features.

Tech Used

- Database Server: Postgres hosted by Amazon Web Services (AWS).
- Application Server: Heroku server hosting a python/flask application.
- Visualization Server: Tableau Public
- Performance: Materialized views used to perform selection, projection, and aggregation.
- Decompose: `statsmodels.tsa.seasonal`
- Regression: `sklearn.linear_model`

Challenges

- The 60 files did not have a consistent naming scheme.
- The 60 files had different structures.
- The 60 files resulted in a database with 60 million rows.