Use data for visualization

Load CSV data to Jupiter notebook / Google Colab from AWS database

Push cleaned data frame to AWS

(Are we able to do this??)

Make Connection to Postgres and AWS using Spark

Join the tables

Load tables in Postgres

Clean data using Spark

Pull data from AWS to Jupyter notebook

Use data for Machine learning:

Linear regression

Random forest regressor

Xxx

Xxx

Link data to Tableau