Project Requirements Design Document
MSBX5420
Team La Plata Peak
Spring 2020

Goals

1. Ingest the taxi data to S3 bucket - Kaegan
2. Use Jupyter notebook and Amazon Athena to visualize the data in tabular formats - Cassidy
3. Visualize data using ElasticSearch and/or seaborn and matplotlib - Binod
4. Machine learning - Priyanka

Milestones

1. Ingest data to dev cluster
    a. Ingest all 2019 data from both the green and yellow taxi service to S3 bucket
2. Optimize the data & streamline ETL process using AWS Glue
3. Create graphical visualizations of data using seaborn and matplotlib
    a. i.e. heatmap, line graph, histograms, etc.
4. Develop machine learning models
    a. Regression and/or decision tree
5. Push Athena queries, visualizations and machine learning code to dev cluster
6. Deploy to production cluster
7. Design report and presentation