Tae Hyon Lee

Data Scientist

(415) 832 - 9291

leeth7830.github.io

leeth7830@ucla.edu

/in/taehyonlee

leeth7830

Skills

Languages

Java, SQL, Python, R, Javascript

Tools

Jupyter, Tableau, Linux

Other Skills

Communication, Decision making, Machine Learning

Research

Spatio-temporal Visualization Research

Researched and wrote a literature survey on modern methodologies of visualizing spatio-temporal data such as traffic to identify human activity and geographic space.

Designed a visualization technique in D3 to help uncover patterns and anomalies in human activity that would normally be hard to recognize through statistical analysis.

Experience

July 2017 - Feb 2018 **Data Science Intern** MonJa

- **(Data Transformation, Data Mining, Data Modeling)** Built, extended and maintained MonJa's data stores for market data using Postgresql and Hiveql and proposed new data structures for data storage and code that will lead to more effective and efficient use of market data by MonJa's product software and modeling toolkit
- **(Statistical Predictive Modeling)** Researched, experimented, and trained a logistic regression machine learning model using loan data that identifies significant variables which highly impact delinquency.
- **(Report Generation, Analysis)** Prepared slice-and-dice Tableau reports of modeled data to support investors and senior research members' deep analysis of Marketplace Lending borrower credit behavior, loan delinquency, default, and investor expected return

May 2017 - July 2017 **Database Administrator Intern** Moneycorp

- **(Data Management)** Cleaned and managed the database to omit or fix bad data and maintain high-quality data.
- **(Marketing Analysis)** Analyzed market data to identify behaviors to assist the marketing team with the email campaigns

Education

2018 - 2019 **MS, Computer Science** **University of San Francisco**
Specialization in Data Science
Coursework: Data Visualization Research, Data Structure/Algorithm

2016 - 2017 **MBA, Information Systems** **San Francisco State University**
Emphasis in Information Systems
Coursework: Data Analysis, Managerial Decision Making, Business Intelligence

2012 - 2015 **BA, Business Economics** **University of California, Los Angeles**
Specialization in Programming in Computing
Coursework: Statistics for Economists, Econometrics

Projects

P2P Lending Platform (Avant) Analysis
Tools: Python, SQL, Tableau, Spark, AWS, Machine Learning

- An analysis of peer-to-peer loan made with logistic regression model built with R and visualizations created with Tableau. The data is imported and mapped to Postgres database on AWS using Spark and Hiveql and exported for analysis.

Data Hub
Tools: Python, SQL, Spark

- A web server built with Flask and Postgres that can store publicly available data in a common and organized format and allow users to access the data through API requests.

User Study - Error Bars
Tools: Python(Pandas, Statsmode, SciPy)

- A user study that evaluates the performance of different type of error bars (bar chart, box plot, gradient plot, and violin plot). The goal of the analysis is to investigate the drawbacks with this standard encoding and consider a set of alternatives designed to more effectively communicate the implications of mean and error data to a general audience.