

# Matthew Huh | Data Scientist

503-505-6344 | [mhuh1993@gmail.com](mailto:mhuh1993@gmail.com) | [linkedin.com/in/mhuh/](https://www.linkedin.com/in/mhuh/) | <https://mhuh22.github.io/>

## Skills

---

**Programming:** SQL, git, shell/unix, spark, python (numpy, pandas, scikit-learn, scipy, NLTK, matplotlib, seaborn)

**Business Intelligence:** Tableau, Salesforce, Google Analytics

**Data Science:** Linear/Logistic Regression, Time series analysis, Feature Engineering, Natural Language Processing (NLP), Principal Components Analysis (PCA), Random Forest, GBM, classification, clustering

**Big Data:** AWS (S3, EC2, Lambda, Data Lakes), Splunk

## Experience

---

**Senior Data Analyst | Capital One** 05/2020 - Current

- Develop end-to-end machine learning models to assess individual customer financial health and risk tolerance
- Developed automated processes and reporting to transform and prepare data in snowflake / AWS
- Analyzed potential losses from removing overdraft fees, potential disparate impact to customers
- Explored, validated, and performed feature engineering for data in preparation for machine learning models
- Evaluated financial and security impact of limiting access through malicious devices
- Developed clustering models to group customer chatbot interactions

**Inventory Analyst | Mt Adams Fruit** 04/2019 - 05/2020

- Subject Matter Expert (SME) on Salesforce, Tableau, Python, & Github - main point of contact for expanding platforms, creating reports, dashboards, scripts, and processes
- Modernized data validation with automated processes, and ETL for live dashboard updates
- Reported directly to COO to integrate Python, Github, & Tableau into our technology stack

**Data Science Program | Thinkful** 03/2018 - 12-2018

- Intensive program focused on python, statistics, and big data techniques including machine learning.
- Learned industry best practices and standards by collaborating with a senior data scientist

## Projects

- **Predicting the Outcome of Kickstarter Projects** - created machine learning models to determine what features are most important for determining the success / failure of a project
- **Sentiment Analysis using Twitter Data** - trained a dataset using validated tweet sentiments using NLP to determine the sentiment on a large collection of untrained tweets

**GIS Technician (Contract) | Apple** 10/2016 - 10-2017

- Validated data for their GIS mapping platform .

## Education

---

BA in Geography (GIS), University of Washington, Seattle WA

## Professional Development

Data Mining Specialization (University of Illinois), Coursera, 2022

K-Means Clustering in Python (University of London), Coursera, 2021