

Summer McGrogan

Thousand Oaks, CA • summermcgrogan@gmail.com • (408) 806-0065 • www.linkedin.com/in/summermcgrogan
https://github.com/smcgrogan

EDUCATION

University of California, Berkeley

Berkeley, CA

Master of Data Science

GPA 4.0

December 2024

Loyola Marymount University

Los Angeles, CA

Bachelor of Science, Applied Information Management Systems

Minor, Environmental Studies

GPA 3.8

May 2021

HIGH LEVEL SUMMARY

- Working background in data analytics and my previous jobs have focused on SQL and python for data cleaning and modeling.
- Graduate experience gave me hands on experience with LLMs and feature extraction techniques, especially for Natural Language Processing and Computer Vision.

INDUSTRY EXPERIENCE

CalAmp

Richardson, TX

Data Analyst

October 2021 – Present

- Developed machine learning models to automate car accident predictions to save 5% of employee time per year
- Created and optimized SQL queries and dashboards in Power BI to provide insights into driver behavior that could reduce accident rates of large transport vehicles by 6.5% in the United States
- Designed an automated ETL pipeline to process semi-structured data from AWS buckets to store in Azure SQL database
- Gathered, validated, and analyzed all data for bi-weekly C-suite meeting with comprehensive analysis of team progress
- Collaborated with multiple teams of project managers and engineers to provide critical data analysis for sales team to retain lucrative customer base in a platform transition

FactSet Research Systems

Remote (San Francisco, CA)

Client Solutions Intern

June 2020 – July 2020

- Created financial dashboards for clients and coworkers highlighting specific KPI's important to the client
- Performed qualitative and quantitative analysis of financial and market data through data acquisition and database querying
- Executed queries of financial information using FQL and presented detailed analysis of query findings
- Mastered FactSet products with an in-depth review of functionality as applied to select client data

Nextflex

San Jose, CA

Marketing Intern

May 2019 – August 2019

- Distributed survey to external companies to collect data about current company permissions and satisfaction with access
- Presented data collected in meetings with marketing, communication, and executive department heads to modify external company permissions to improve the membership experience
- Executed all promotional material for company events, lectures, conferences, and seminars

RELEVANT COURSE PROJECTS

- **Abstractive Text Summarization for Domain-Specific Articles:** This project focuses on creating a high-performance abstractive summarization model for environmental science articles using three state-of-the-art language models, PEGASUS, FLAN-T5, and BART, fine-tuning them on a curated subset of environmental science papers from the Semantic Scholar Open Research Corpus (S2ORC). We evaluate the models using both automated metrics (ROUGE, BERTScore) and human assessments, noting that, while traditional metrics may not fully capture summary quality, human evaluations reveal the strengths of our fine-tuned models.
- **Artist Image Classification:** Team project focusing on developing an image classification model to categorize images of paintings into ten artist categories using the WikiArt dataset. We explore a comprehensive set of features, ranging from simple features such as Color Histograms and Histogram Oriented Gradients (HOG), to more complex features, such as Bag of Visual Words (BOVW) and ResNet18. We perform Principal Component Analysis to reduce the dimensionality of our dataset, which reduces overfitting, increases generalizability, and increases model efficiency. We capture our results in two classification models, Support Vector Machines (SVM), and Linear Regression (LR).
- **Impossible Foods Product Improvement:** Using python and SQL, created a market segment analysis for Impossible Foods to improve their marketing strategy and increase the size of their customer base.

SKILLS & AWARDS

SOFTWARE SKILLS

Power BI, Athena, Data Warehousing, AWS RDS, API's, Tableau, Microsoft Office (Excel, Word, PowerPoint), Jupyter, Anaconda, A/B testing, Microsoft Azure, Hadoop, Spark, MySQL, EC2, Google Analytics

PROGRAMMING SKILLS

SQL, Python(Pandas, NumPy, scikit-learn), TensorFlow, GIT, Bash, Java, JavaScript, HTML