

SAM KEELER

DATA SCIENTIST

I'm a mechanical engineer with a passion for using statistics and performing analysis. I not only enjoy extracting useful information from data but love the challenge of communicating clear and inspiring data stories that motivate people to take action.

CONTACT

832 - 922 - 3064



keelersam@gmail.com



[/Sam-keeler](#)



[/sam-keeler](#)

TECHNICAL SKILLS

Python - SQL - Git - Pandas - Numpy - Seaborn - Scikit-learn - Matplotlib - Statistics - Jupyter Notebooks - Excel - Reliability and Quality Control - Solidworks - MATLAB

EXPERIENCE

HEB Produce Stocker Oct 2019 - Dec 2020

- Stocked produce and ensured produce area and backroom were clean and organized for optimal sales.
- Kept track of current inventory by doing counts at the end of the night.

CVS Shift Manager Jun 2016 - May 2019

- Supervised cashiers and delegated tasks efficiently for completion by the end of business hours.
- Responsible for closing out the registers and readying the deposit at the end of the night.
- Resolved any customer issues regarding cashiers, transactions or store conditions in general.

PROJECTS

Capstone: 311 Response Times

My team and I acquired the data for this project from San Antonio's open records on 311 calls. Seeking to ensure that San Antonio is serving every community equally, we are examining drivers of response time to 311. We also plan to create a classification model to predict how timely a response to a 311 call will be.

Predicting Customer Churn

The data for this project came from a bank's credit card portfolio where customer attrition is on the rise. Seeking to relieve customer attrition and focus efforts towards customers who are more likely to churn, I've made a classification model to help predict customer churn as well as examined the drivers of that churn. The classification model I created predicted churn with 95% accuracy, an improvement from baseline by 11%.

Natural Language Processing

I acquired the data for this project using web scraping techniques to obtain readme contents from different GitHub repositories. The readme was then standardized using various NLP methods before it was finally lemmatized. It was then vectorized using a TF-IDF method. The machine learning model I created predicted the primary coding language used in each repository. My best model was a KNN classifier that beat my baseline by 19%.

EDUCATIONAL TRAINING

Bachelor of Science

University of Texas at San Antonio May 2019
Mechanical Engineering

Certificate of Completion

Codeup June 2021
Data Science Program