RYVYN YOUNG

Data Scientist

I am an inquisitive results driven professional with the data science skills to interrogate data and find real solutions. My background in industry leadership taught me flexible decision making, excellent analytical and programming skills, and allowed me to further develop my creative and problem-solving abilities.

Ryvyn.Young@gmail.com   210.745.1195   github.com/RyvynYoung   linkedin.com/in/ryvynyoung

SKILLS

Data Storytelling - Data Analysis- Pandas - SQL - Python - Tableau - Matplotlib - Machine Learning/AI - Natural Language Processing - Time Series - Applied Statistics - Git - Jupyter Notebooks - Anaconda - Seaborn - Spark

EDUCATION

Codeup
Fully-immersive, project-based 22-week career accelerator that provided 670+ hours of expert instruction. I developed expertise across the full data science pipeline (planning, acquisition, preparation, exploration, modeling, delivery), and became comfortable working with real, messy data to deliver actionable insights to diverse stakeholders.

University of Colorado
BFA Design and Technical Theatre

EXPERIENCE

The Dentist's Choice
Franchise owner/operator   2014-2018
- Serviced dental hand tools for dental offices in the San Antonio area
- Over $75K in gross sales annually
- Analyzed customer data, created geographical based marketing plan, updated and created marketing tracking system

The Home Depot
General Manager Distribution   2007-2013
- Created outstanding associate environment, achieved cost and productivity goals through data analysis and project implementation
- Facility scope: serviced 49 stores in south Texas area, 17-acre site with 250K covered sq. ft., $100M in annual volume
- Reduced overall facility cost per unit by 35% while volume increased 38%

DATA SCIENCE PROJECTS

A Tale of Two Cities   January 2021
The SVI is a scale that predicts a population's vulnerability in the event of a natural disaster. COVID is the first global pandemic since the development of this measure. We evaluated the effectiveness of utilizing SVI as a tool to predict COVID case count per 100,000 individuals in San Antonio and Dallas, Texas. We found that the best indicator of case count in both cities was total socioeconomic need. Over-reliance on SVI during pandemic type natural disasters may lead to vulnerable populations missing out on critical resources.

Supply Chain Units per Hour   December 2020
"We manage what we measure, but frequently we measure what is easy" inspired by this quote I found a Kaggle dataset with units and pick time to evaluate. Exploring the data I found that there was minimal variance in many features I expected to be predictive. I found using a polynomial features model with order complexity was 65% more accurate in predicting the time needed to pick an order.

Natural Language Processing Project   November 2020
For this project my partner and I scraped 1K GitHub repository urls related to "environmental" to create a dataset for analysis. We used NLP techniques to explore the data and build a model that would predict the programming language of the repository based on the text in the README.

Zillow Project   October 2020
This project was done in two iterations. In the first iteration my partner and I created a regression model built on property data between May and June 2017 to predict home value. Our best performing model was a polynomial linear regression model which provided a 31% improvement over baseline. In the second iteration I worked with the KMeans clustering algorithm to build a model that would better predicting the Zestimate error in the Zillow data.

PASSION PROJECTS

Natural Language Processing - Harry Potter   November 2020
Using my NLP project as a starting point I improved the model working with a new dataset. Resampling provided the greatest improvement in prediction while accuracy increased only 3%, the average F1 score increased 14%.

Fitbit Analysis   October 2020
Used time series analysis to explore Fitbit data. Best predictor was 7 day rolling average and the model improved prediction by 28% over baseline.

Market Basket Analysis   September 2020
Inspired by a grocery dataset in Kaggle I found an algorithm to assess frequency of basket items. Appriori is frequently used for itemset mining and association rule learning over relational databases.