PARKER R. VOIT

DATA SCIENTIST

210-591-3605   parker.r.voit@gmail.com   in/parkervoit/   github.com/parkervoit

TECHNICAL SKILLS

- Python
- SQL
- Natural Language Toolkit
- Scikit-learn
- SciPy
- Pandas
- Matplotlib
- Natural Language Processing
- Data Storytelling
- Jupyter Lab
- Anaconda
- Tableau
- Seaborn
- BeautifulSoup

EDUCATION

CODEUP, CERTIFICATE OF COMPLETION

March 2021 - September 2021
Fully-immersive, project-based 22-week career accelerator providing students with 670+ hours of expert instruction in applied data science. Students develop expertise across the full data science pipeline (planning, acquisition, preparation, exploration, modeling, delivery), and become comfortable working with real, messy data to deliver actionable insights to diverse stakeholders.

TRINITY UNIVERSITY, B.S. NEUROSCIENCE

San Antonio, TX | August 2014 - May 2018
Honors Thesis: Cocaine-mediated changes in localization of glutamatergic inputs in murine nigral dopamine cells

EXPERIENCE

SERVER

The Hoppy Monk | September 2019 - March 2021
Provided excellent customer service in a restaurant with the largest taproom in San Antonio. Leveraged personal communication and multi-tasking skills to ensure customer satisfaction.

FINISHED BEER SPECIALIST

BJ's Restaraunt and Brewhaus | February 2019-September 2019
Maintained draught beer systems at a location that averaged $13,000 in sales per day. Ensured beer quality and system performance to support waitstaff in providing a seamless customer service experience.

PROFESSIONAL SUMMARY

Leveraging technological solutions to develop actionable insights that aid in problem solving is what drives my interest in data science. I am inquisitive by nature and enjoy learning new ways to find answers to difficult questions. I am always looking to improve myself and the world around me.

DATA SCIENCE PROJECTS

TECH BLUES: IMPROVING MENTAL HEALTH IN TECH   AUGUST 2021
Classification Model
Used mental health survey data from workers in the tech industry to identify factors that impact mental health and work performance. Used Pandas and Seaborn for data cleaning and exploration. Cross-validation and feature engineering were implemented to create a multi-layer perceptron model to predict impacts to work performance. A feature analysis was performed to identify the main factors that contribute to mental health in the workplace. Results and business solutions were communicated in the form of a Jupyter Notebook and presentation.

GITHUB LANGUAGE PREDICTION   AUGUST 2021
Natural Language Classification
Used a K-Nearest Neighbors classifier to predict coding languages from GitHub repository README.md files. A web scraping tool was developed to acquire the data and was cleaned using BeautifulSoup and Regex. Scikit-Learn was used for preprocessing and for model development. The model improved on baseline performance by 15% and was able to accurately predict Python and Shell langugages from README.md text alone. Results were communicated to peers in the form of a presentation and Jupyter Notebook.

ANOMALOUS ACCESS REQUEST IDENTIFICATION   JULY 2021
Anomaly Detection
Used server access logs to identify anomalous access requests to the Codeup curriculum. Used Pandas, Seaborn, and Time Series Analysis to visualize and explore the data. Git was used for version control and team collaboration. Was able to identify anomalous user requests and reported findings to the Codeup team in the form of a Jupyter Notebook.

PREDICTING FATAL CASES OF HEART FAILURE   JULY 2021
Classification Model
Used clinical patient data to predict instances of death in cases of heart failure. Utilized data manipulation methods such as SMOTE oversampling, min-max scaling, and feature engineering to create a random forest classification model. The model was able to correctly identify fatal cases of heart failure 95% of the time. Findings were shared with peers in the form of a Jupyter Notebook.

ZILLOW TAX VALUE PREDICTION   JUNE 2021
Regression Model
Used Zillow property data to predict tax valuations of properties for sale on the website. Data was obtained using MySQL and manipulated with Pandas and NumPy. Scikit-learn was used to create a polynomial linear regression model that predicted valuation within a range of $31,000 of the actual price. Results and takeaways were communicated to peers in the form of a report and presentation.