

VERONICA F. REYES

DATA SCIENTIST



(210) 386-8563



veronica.fierro.reyes1@gmail.com



/veronica-f-reyes



/veronica-f-reyes

TECHNICAL SKILLS

- SQL
- Python
- Pandas
- Numpy
- Matplotlib
- Seaborn
- Applied Statistics
- Machine Learning
- Jupyter Notebooks
- Natural Language Processing
- Git Hub
- Tableau
- Storytelling
- Microsoft Excel
- SDLC

EDUCATION

CODEUP

Certificate of Completion December 2021
Fully-immersive, project-based 22-week career accelerator that provides students with 670+ hours of expert instruction in applied data science. Developed expertise across the full data science pipeline (planning, acquisition, preparation, exploration, modeling, delivery), and became comfortable working with data to deliver actionable insights to diverse stakeholders.

UNIVERSITY OF TEXAS AT AUSTIN May 2003
B.S. Mathematics - Concentration in Computer Science

EXPERIENCE

DATA ANALYST | E-CONTROLS 2020

Responsible for providing the daily metrics report for manufacturing engineers, I collected data using SQL queries and created Microsoft Excel dashboards using Power Pivot. I also prepared technical documentation and provided technical support fielding ad-hoc data requests.

SUBSTITUTE TEACHER | NISD 2016-2019

Fulfilled substitute teaching assignments across all subject areas, teaching students of various age levels in elementary and middle schools, instructing classes of as many as 24 students, and communicating with teachers and administrators

TEACHER | CROSSROADS BAPTIST CHURCH 2013-2016

Planned and implemented lesson plans according to Wee class curriculum for children. Led play-based lessons and skill building activities while fostering a nurturing, fun environment and building a strong working relationship with parents and administrators

SOFTWARE ENGINEER | USAA 2006 - 2009

Conducted analysis for business requests to create system design enhancements with business users, developers and testers by gathering business requirements and working thru workflows and business scenarios.

Enhanced usaa.com functionality by coding components, providing production support, testing code and fixing defects in Java. Prepared technical documentation for existing processes and disaster recovery scenarios.

PROFESSIONAL BIO

Driven by a natural curiosity and the drive to learn and explore, I am eager to ask questions and seek solutions. I am a quick learner with experience as a data analyst, business analyst, software developer, and teacher. With the recent immersive training in data science, I am equipped to take on roles in the data science field tackling questions with the data, math, and technical tools provided.

DEVELOPMENT PROJECTS

PREDICTING MAJOR LEAGUE BASEBALL (MLB) OUTCOMES

Classification & Linear Regression Modeling

Using MLB data, I was able to create a machine learning model using classification algorithms to predict whether a team would win or lose a game using a team's previous three games average hitting statistics. I was also able to create a regression model to predict the number of runs a team would have in a game using the team's previous three games average hitting statistics. The model created was able to predict a baseball games outcome with an accuracy of 58% on out of sample data beating the baseline of 50%. Another model for predicting the number of runs a baseball team scored also performed better than the baseline accounting for a 35% improvement over the baseline error and predicting the score within 2.43 points.

GUESSING GITHUB

Natural Language Processing

Given the text of the README file in GitHub, we built a model that predicted the primary programming language for a GitHub repository. Because Microsoft has a large number of repos on its GitHub site with a wide variety of coding languages, we pulled the README and coding language used for that repository. This required the use of a variety of web scraping and Natural Language Processing tools as well as use of GitHub's API. In the end we acquired 1500 READMEs along with their associated primary coding language from Microsoft's GitHub page. We used the classification modeling Decision Tree classifier to predict whether the repos used TypeScript, the target language, based on features from the READMEs. We were able to predict with an 84% accuracy whether a repo used TypeScript based on words used and the word length of a README.

ZILLOW: ESTIMATING HOME VALUE

Linear Regression Modeling

Using Zillow data, I created regression models to predict the values of single unit properties. Using the Polynomial Regression Model, I was able to beat the baseline prediction. The model explained 35.6% of the variance on out of sample data, predicting over \$60,000 better than the baseline.

WHY ARE CUSTOMERS LEAVING?

Classification Modeling

The goal of this project was to determine why customers are churning by developing machine learning classification models to predict customer churn based on the customer data available. I created different classification models using Python, Pandas, Matplot, Seaborn, and Scikit-Learn libraries. Using the Random Forest classifier, I was successful in predicting on out of sample data with an accuracy of 81%.