

MATTHEW ZAPATA

Data Scientist



210-663-4979



matthew.zapata21@gmail.com



in/matthew-zapata



/matthewzapata

I enjoy the creative aspect of data science and how there are a multitude of ways to solve a problem. Similar to playing an instrument or editing a video, every project is a chance to be creative, whether that's with solving a problem or creating amazing visualizations. Being able to make each project my own is something I love. Data science allows me to be creative while also providing valuable insights.

TECHNICAL SKILLS

Applied Statistics - SQL - Python - pandas - Matplotlib - Machine Learning - Natural Language Processing - Distributed Data - Data Storytelling - Git - Jupyter Notebooks - Anaconda - PySpark - Tableau

EDUCATION

Codeup

June 2019

Fully-immersive, project-based 18-week Data Science career accelerator that provides students with 600+ 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.

Texas A&M University

2014

Graduated in 2014 with a Bachelor's Degree in Economics. My courses included econometrics, accounting, marketing, microeconomics, and macroeconomics. These courses gave me a good understanding of both economic and business principles.

PROFESSIONAL EXPERIENCE

Alliance Data Card Services

2017 - 2018

Collections Associate

Collected on delinquent accounts and reached a mutually beneficial agreement with the customer, even in high stress situations. Utilized computer programs to reach out to customers, look up account information, and answer questions. Used my customer service experience to ensure that each interaction was professional and pleasant, which made me into one of the top performers on the team.

Vetco Clinics

2016 - 2017

Clinic Lead

Ensured the clinic ran smoothly and handled customer relations. Accurately completed the end of clinic paperwork. Prepared my team for the day by making them aware of any changes in products/services, what we could improve on, and any changes to the flow of the clinic.

Customer Advisor

Educated clients on the best preventative care options for their pets as well as ensured they received great customer service. Collected information about both the owner and pet then determined which products/services were best based on lifestyle. Collected payment, assisted the veterinarian with vaccinating and examining pets, and filled out the vaccination certificates.

Macy's

2015 - 2016

Sales Associate

To reach my sales goal, I applied a method of asking questions and active listening to determine customer needs. Other duties included completing transactions and identifying opportunities for new credit accounts.

DEVELOPMENT PROJECTS

San Antonio Drought Predictions

June 2019

Along with two other group members, we combined statistical analysis with a Support Vector Regression model to correctly predict drought 4 out of 6 times. This first involved researching the topic to gain domain knowledge, as well as cleaning incredibly messy datasets. We then created and tuned our SVR model using cross validation, and finally presented our findings at Data Scientist Day. A huge takeaway was the importance of working as a team and splitting up the work to get it done.

Fitbit Data Time Series Analysis

May 2019

With another group member, I analyzed an individual's fitbit data to determine if they were an employee or an experiment participant at company X. I cleaned and structured several messy csv files into a Pandas dataframe, then we used time series models to correctly identify the individual as well as recognize trends and periods of abnormality. These models included Prophet and Holt Linear Trend. A presentation was created and delivered with a general audience in mind.

Customer Churn Classification

April 2019

Two group members and I created and evaluated several classification models, including Decision Trees, Random Forests, and K-Nearest Neighbors, to determine whether a customer would churn or not and their likelihood to churn. Having an extra member taught us coordination with our GitHub pushes and Jupyter Notebook additions. We also created and presented a single slide that could be easily consumed by a stakeholder.

Zillow Regression

March 2019

Using linear regression along with other modeling techniques, I created a model that would predict the log error for the Zillow estimate, with the purpose being to identify the drivers of error. I cleaned up the data, including what to do with large amounts of null values, and feature engineered, one example being clusters I created through K means.

PASSION PROJECTS

Trigger Warnings With Tweets

July 2019

Self-harm scenes in movies or shows can bring up traumatic memories for those who have a history of self-harm. The goal of this project was to create a classification model that can predict if a movie has a self-harm scene based on tweets. The necessary data was scraped from Twitter, IMDb, and Tumblr using Selenium and BeautifulSoup. The text data was cleaned and lemmatized, then various classification models were trained to determine the best performing model.

Animal Shelter Outcome Predictions

June 2019

Using real data from Animal Care Services and a Decision Tree Classifier, I was able to predict if an animal would be euthanized or adopted with an accuracy of over 90%. To achieve this result I engineered new features by thinking about what factors contribute to an animal being adopted and also used undersampling to balance the classes. I then used cross-validation and tried several classification models to determine which performed the best.