BROOKE HOLYOAK

DATA SCIENTIST

brooke.holyoak24@gmail.com Brooke-Holyoak in/Brooke-Holyoak 210.247.8522

TECHNICAL SKILLS

Python - SQL - Tableau - Applied Statistics - Pandas - Matplotlib - Seaborn - Plotly - Machine Learning - Natural Language Processing - Apache Spark - Data Storytelling - Git - Jupyter Notebooks - Anaconda

EDUCATION

Codeup- Data Science
Certificate of Completion, Dec 2021
Fully-immersive, project-based 22-week career accelerator that provides 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 data to deliver actionable insights to diverse stakeholders.

University of Texas at San Antonio
Bachelors of Music, Music Studies, May 2012

PROFESSIONAL EXPERIENCE

Head Choral Director
Northeast ISD, San Antonio *Aug 2017 to June 2021*
Directed 4-6 ensembles per year consisting of 20-80 members each. Using data I collected, maintained and analyzed, I was able to identify and improve education and performance gaps. I maintained a >90% retention rate, and received consistent highest performance ratings.

Head Choral Director
Harlandale ISD, San Antonio *Aug 2012 to June 2017*
I rebuilt a choral program with over 300% growth. My program achieved historical success, earning the highest ratings in choral competition at district, region and state levels.

ABOUT ME

A creative data scientist with over ten year's people leadership experience. Expert in cultivating excellence through vision, interpretation, and the commitment to improvement. I have a hunger for drawing unique insights, and the drive to always make an organization better.

PROJECTS

MUSIC AND EMOTION, October 2021
Utilized the full data science pipeline to create a classification model that predicts musical features that induce energized emotions. I combined an existing dataset with new features I acquired using a web application. I used pandas, matplotlib, and seaborn to explore and visualize the data, and used SciKit Learn to prepare and model. The Random Forest model performed with 76%accuracy.

ZILLOW CLUSTERING, September 2021 Acquiring Zillow data from a SQL database, I used the full data science pipeline to find drivers of logerror in for 2017 Single Unit Properties. After exploring using clustering model technique, I created a linear Regression Model (Tweedie Regressor Model) which performed 1.04% better than the baseline when using select features.

ZILLOW REGRESSION, August 2021
Using data from Zillow, and going through the data science pipeline, I was able to determine the features that are the biggest predictors of specific property values in LA. I acquired and cleaned Zillow data. I explored data via individual distributions and correlated distributions. Finally I created an OLS model that performed 13% better than the baseline prediction.

PREDICTING CUSTOMER CHURN, August 2021
Constructed a ML classification model that predicts with 89% accuracy what features are drivers in Telco Customer Churn.