

# Jacqueline Shay

Data Scientist

469.408.4121 

jacqueline.a.shay@gmail.com 

Jacqueline-A-Shay 

in/jacqueline-a-shay 

## TECHNICAL SKILLS

Applied Statistics

Machine Learning

NLP. Apache Spark

SQL. Python. pandas

Matplotlib. Seaborn

Jupyter Notebook. PySpark

Git. Tableau

Data Storytelling

## EDUCATION

**Codeup Data Science Certificate**  
Aug 2019 – Jan 2020

Full-immersive, project-based 20-week program. Developed expertise across the full data science pipeline and become comfortable working with messy data to deliver and communicate actionable insights to diverse stakeholders.

**UTSA Ph.D. Candidate  
Cell & Molecular Biology**  
Aug 2015 – Present

Proposed, and defended scientific hypotheses and research. Managed NIH fund of USD600K budget with cross-functional project management skills. Refined experimental and procurement protocols and workflow to improve project efficiency and reduce material waste.

**Virginia Tech B.S.  
Biological Sciences**  
Aug 204 – May 2008

A data practitioner with over 8 years of analytical experience in the field of cancer and stem cell biology. Driven by the compassion for humans in need of care, I seek to rely on computational power to uncover valuable data in transforming human life.

## PASSION PROJECT

### Shelter Animal Outcome Prediction Project (Individual)

Austin Animal Shelter & Kaggle

Present

Ongoing project utilizing data from Austin Animal Shelter, Python/ pandas and various machine learning algorithms, I am hoping to predict outcomes for animals.

Goal: To promote animal adoption rate by constructing a recommender and notification system for potential adopters.

## DEVELOPMENT PROJECTS

### Natural Language Processing Project (Group)

GitHub Repository Scrape and Learn

Dec 2019

Constructed models to predict programming language used for each GitHub repository. Methods: (1) web-scrape repo on GitHub (2) process data by natural language methods and regular expression (3) construct prediction model e.g. linear regression.

### Time Series Analysis Project (Group)

Fitbit Activity Prediction and Persona Analysis

Nov 2019

Explored and analyzed Fitbit data to predict the persona of the original Fitbit owner. Used multiple modeling methods, and validated model performance with statistics. Holt's exponential smoothing method was the best performing method for the prediction of future activities for the Fitbit owner.

### Cluster Project (Group)

Zillow How to Improve Property Value Prediction?

Nov 2019

Used clustering methods, e.g. KMeans, DBScan, to identify potential drivers of prediction error and as an alternative feature engineering method. Several prediction models were constructed and compared. The decision tree regressor method was reported to the stakeholders as the best performing model.

### Classification Project (Individual)

Telco How to Improve Customer Retention?

Oct 2019

Improved customer retention by identifying the signatures underlying churning customers and propose rooms for improvement. Curated 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.

### Regression Project (Individual)

Zillow How to Improve Property Value Prediction?

Oct 2019

Polynomial regression was identified as the best method to predict property value without prior clustering or too much feature engineering.