# PAIGE GUAJARDO
## Data Scientist

#: 210-383-2533 | paigeeguajardo@gmail.com | in/paige-guajardo | github.com/PGuajardo

Driven by improving in the field of technology, I've always kept a keen eye to data science throughout my life. I wanted to better my life so I've now stepped forward in the realm of data science applying my previous technical skills, learning and applying techniques to new domains and applications. Now my skills and experience will showcase how versatile and resourceful a data scientist I'll be!

## TECHNICAL SKILLS

Python - SQL - Pandas - Numpy - Scikit-learn - Matplotlib Seaborn - Jupyter Lab - Anaconda - Git - Applied Statistics Machine Learning - Classification - Regression - Clustering Time Series - Anomaly Detection - Natural Language Processing - Data Analysis - Data Storytelling - Tableau - C# - CSS - HTML - Java - C - Swift

## EDUCATIONAL BACKGROUND

**Codeup 2021-2022**
**Certificate of Completion**

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.

**St Mary's University** **2012-2016**
Pursued Bachelors Computer Science
Completed 90hrs

## PROFESSIONAL EXPERIENCE

**Cenveo** **2018-2021**
Assistant Manager: Shipping/Receiving

Helped procure shipments of paper via ordering through paper / truck companies. Delivered products to customers, made sure each item met inspection guidelines. Attended team meeting for company agendas on products/training. Became comfortable working at meeting deadlines and delivering products at a high standard.

## DATA SCIENCE PROJECTS

**Capstone- Enron Emails | NLP/Time-Series | March 2022**
Worked with a team and helped create time series analysis on sentimental trends within persons of interest in the data set. Developed and used BERTopic model to develop topics based off emails from persons of interest in 2000 and compared that with teams 2001 model.

**Brazil Cancer Data Analysis | Classification | Jan 2022**
Created Decision Tree, Random Forest, KNN, and used SMOTE for models to help predict those with no cancer with a 69% F1-score. Acquired data from National Cancer Institute (INCA) based on all cities in Brazil from 2000 to 2019.

**Finding Error Drivers | Clustering/Regression | Jan 2021**
Implemented K-means clustering models as part of exploratory data analysis of single-unit property Zillow data. Utilized OLS, GLM, LassoLars, and polynomial linear regression models to identify potential drivers error in the preexisting model. Compiled information gained to provide recommendations for improvement of current model.

**Predicting Assessed Value | Regression | Dec 2021**
Developed regression models to predict assessed value of single-unit homes using Zillow data. Best-performing model significantly outperformed baseline and increased explained variance by 20% on out-of-sample data. Employed recursive feature elimination in concert with exploratory data analysis to select best features from large feature set.

**Predicting Customer Churn | Classification | Nov 2021**
Created Decision Tree, Random Forest, Logistic Regression, and KNN models to predict if teleco customers would churn with a 78% out-of-sample accuracy. Owned each phase of the data science pipeline including acquisition, preparation, EDA, modeling, and delivery. Leveraged skills in Python, SQL, Pandas, Scikit-learn, Seaborn, Numpy, and Matplotlib to complete project