

# ELIHEZER LOPEZ

## DATA ENGINEER



(210) 309 - 2669



lopezelihezer@gmail.com



in/lopezelihezer/



lopezelihezer/

## OBJECTIVE

Seeking position as a Data Engineer Consultant to supply critical and meaningful insights that provide solutions to business challenges. I'm a passionate individual, interested in a company that values continued education and growth.

## SKILLS

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

## EDUCATION

### Codeup

Dec 2021

*Data Science*

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.

### Open Cloud Academy

Dec 2018

*System Administration*

Earned CompTIA Net+ and RHCSA certifications. Received hands on experience, working with routers and switches. Excellent faculty and staff shared critical and resourceful input of best practices learned throughout their careers in the IT field. Great place to come learn and grow in preparation for a career in Information Technology.

### Palo Alto College

May 2015

*Liberal Arts Associates Degree*

Founded in 1985 from the desire to provide higher education to residents of south San Antonio, Palo Alto College has spent more than 35 years serving over 150,000 individuals throughout San Antonio, Bexar County, and surrounding counties.

## WORK EXPERIENCE

### System Administrator

COLUMBIA IT, INC.

MAR 2019 - JUN 2021

Worked with small-to-medium sized businesses as a one-stop shop for all their IT needs. Collaborated directly with ownership/executives to setup, deploy, and maintain working IT solutions including VOIP, Backup&Replication, Network Infrastructure, Virtualization, CyberSecurity and other managed IT services. Gained experience working with non-profits, healthcare clinics, and veteran-owned businesses.

## PROJECTS

### Predicting Stroke - Capstone Project

DEC 2021

The CDC cited in 2018, that 1 in every 6 deaths from cardiovascular disease was due to stroke. In the United States, this means every 40 seconds, someone has a stroke. This shocking statistic was the drive for our team. We found drivers of stroke to include advanced age; hyperglycemia; hypertension; heart disease; marriage under 55 years old. We built a classification model based on these drivers. Our best model's performance (Naive-Bayes) had a Recall of 83% and ROC AUC of 85%. This model was then used to produce a stroke-risk calculator that can be found on our team's GitHub.

### GitHub Repos - NLP Project

NOV 2021

Web scrapped README files of trending repos on GitHub. Used BeautifulSoup to parse the HTML and acquire desired text. Gathered the repo name, language, readme text, word count, and language frequency in a tabular format. Built a Random Forest model with a depth of 20 that was able to predict the appropriate language with 75% accuracy.

### SuperStore - Individual Project

OCT 2021

Using a SuperStore dataset, I built an OLS regression model that beat the baseline in predicting Profit. I tested by Category, Sub-Category, and Segment, and found that overall SuperStore is losing money in Texas, with Office Supplies consisting of 61% of sales. A T-test suggested there is no difference between profit generated from office supplies and the 'Texas Mean Profit'.

### Weather - Time Series Project

OCT 2021

With earth surface temperature data offered by Berkeley Earth, I focused on San Antonio and analyzed patterns in temperature over time. I then modeled those patterns to forecast temperatures. The models were built using the last observed value, simple average, moving average, holt's linear trend, and year-over-year differential (yoy-diff). My best model used yoy-diff and was within 3 degrees of the average temperature for my test data split.

### Zillow - Regression Project

SEP 2021

The goal of this project was to find drivers that affect tax value using the Zillow data set. Bedroom count, bathroom count, and area(sqft) all had a linear correlation with taxvalue. Using this data, I created a ML Regression Model that predicts tax value. I documented the entire process to be presented or read like a report. My best model (Quadratic Polynomial) beat the baseline (mean) RMSE value by 12,259 points.