

NADIYA PAZ

Authorized to work in the USA

seeking entry-level roles as

Data Scientist / Data Analyst / Data Engineer

📞 (512) 844-2633

✉️ nadina.abouhamdan@gmail.com

🌐 [LinkedIn/NadiaPaz](https://www.linkedin.com/in/NadiaPaz)

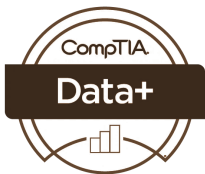
🐙 [GitHub/nadia-paz](https://github.com/nadia-paz)

Highly motivated and skilled Data Professional with a versatile background in linguistics, marketing, teaching, tourism, and customer service. My passion for technology, mathematics, and problem-solving allows me to bring meticulous attention to detail to my work. I am currently seeking a career in Data Science and Analytics, and I'm open to on-site or hybrid roles in Austin, TX, remote positions nationwide, and relocation opportunities.

TECHNICAL SKILLS

Python - Pandas - NumPy - SQL - Tableau - Matplotlib - MySQL - Git - Linux - Applied Statistics - Machine Learning - Natural Language Processing - Data Storytelling - Presentation - Microsoft Excel - PowerPoint

CERTIFICATIONS



EDUCATION

Codeup - Data Science program · San Antonio, TX

Aug, 2022 - Feb, 2023

Fully immersive, project-based 20-week career accelerator that provided me with 670+ hours of expert instruction in applied data science. Developed expertise across the full pipeline (planning, acquisition, preparation, exploration, modeling, storytelling), and become comfortable working with real, messy data to deliver actionable insights to diverse stakeholders.

Ukrainian Academy of Sciences · Kyiv, Ukraine

2001 - 2005

Post-graduate student, junior research worker

Kyiv National Linguistic University · Kyiv, Ukraine

1995 - 2000

Master of Arts

Philologist (Arabic language)

DATA SCIENCE PROJECTS

Currently working on Epilepsy trigger detection - Anomaly detection, Video Analysis

Paired project. We downloaded videos in *.mp4 format and converted them into NumPy arrays using the OpenCV library. During the exploration and data analysis, we confirmed one of the reasons that trigger seizures is strobing light flashes. My partner in the project used the OOP approach and created Python classes for analyzing videos. I created an algorithm that identifies the parts of the video that are potentially dangerous for people with epilepsy. Outcome: The algorithm that detects one of the biggest epilepsy triggers in videos.

Technologies: *Python, Numpy, OpenCV*

[Cisco Sales Forecast](#) - Time Series Analysis, Regression

Group project. We used the data available on the Texas Open Data portal to analyze the transactions between Cisco Systems, Inc and Texas public agencies. We took a closer look at how the pandemic affected the seasonality and trends of sales and implemented statistics and machine learning techniques to make our forecasts. I mostly contributed to the cleaning and exploration part of this project.

Outcome: Our best model improves the baseline model's results by 8%.

Technologies: *Python, Pandas, Numpy, Sklearn, XGBoost, Statsmodels, Matplotlib, Seaborn, Tableau, Canvas*

[NLP project - GitHub scrape](#) - Natural Language Processing, Classification

Group project. Utilized BeautifulSoup to scrape GitHub repositories for Readme texts. We explored the data using insights discovered and feature engineering to develop the most accurate model for production. My role in this project was data cleaning, preparation, and creating machine learning models.

Outcome: We created the function that accepts text from the GitHub repository and predicts the programming language with an accuracy of 59%.

Technologies: *Python, BeautifulSoup, NLTK package, Regex, Pandas, Numpy, Sklearn, XGBoost, Canvas*

[Austin Animal Center](#) - Classification

Individual project. Used classification algorithms to predict shelter animal adoption rates. This classification helps to identify animals at high risk of euthanasia and save their lives by putting more effort into their adoption process.

Outcome: My best model achieved an accuracy score of 80.3%, a 10% increase from the baseline.

Technologies: *Python, Pandas, Numpy, Matplotlib, Seaborn, Sklearn, SciPy Stats.*

[Zillow Zestimate's error](#) - Clustering, Regression

Paired project. By estimating the errors in the model, we wanted to present advice about which model features should be incorporated into or modified in Zillow's price-determining model. In this project, I worked on creating clusters and developing regression models.

Outcome: the best model outperformed Zillow's estimate by 0.01%

Technologies: *SQL, Python, Pandas, Numpy, Matplotlib, Seaborn, Plotly, Sklearn, SciPy Stats.*

[Single Family Residence Price Predictions](#) - Regression

Individual project. This project analyzed data from the Zillow database to identify significant features in the prediction of home value in Los Angeles, Ventura, and Orange counties in California, USA. Built a predictive model.

Outcome: the best model beats the baseline by 23.5%

Technologies: *SQL, Python, Pandas, Numpy, Matplotlib, Seaborn, Sklearn*

WORK EXPERIENCE

Career break - Austin, TX

Stay-at-home mom, education 2019 - present

Gigaaa International S.A. - remote

Software Tester / Mar 2019 - Jun 2019

- Trained the AI personal assistant by building questions and answers on a wide range of topics in Russian.
- Ran test cases to make sure the application would pick up on keywords and provide the correct results
- Provided the results and suggested improvements in team meetings with other testers, developers, and managers