PIZON SHETU

[**pizon.skhan@gmail.com**](mailto:pizon.skhan@gmail.com)[**www.linkedin.com/in/pizon-shetu/**](http://www.linkedin.com/in/pizon-shetu/)[**www.github.com/izinex/**](http://www.github.com/izinex/)

# PROFESSIONAL SUMMARY

Highly motivated and skilled data scientist with extensive experience in utilizing data-driven solutions to drive business growth and decision-making. Proven ability to work with large datasets, extract meaningful insights, and build predictive models. Experienced in leading teams, conducting technical discussions, and presenting data visualization to stakeholders.

# PROFESSIONAL EXPERIENCE

**Data Scientist/Quant– Webster Bank (June, 2022 – Present Stamford, CT):**

* Conducted Quantitative analysis on various financial data to facilitate model implementation in credit risk management under CECL guidelines.
* Utilized machine learning and OCR technologies to text mine data and process images for model building data.
* Parsed and cleaned large data sets which consist of files in CDR using regex and NLP, leading to the development of data set for model development.
* Extract/Scrape unstructured data from over 1 million PDF files to structured CSV tables for model development by training ABBYY software using OCR’s image recognition and NLP.
* Feature engineered predictor variables and response variables to build predictive models from unstructured data set.
* Trained and managed team members, leading the project from proof-of-concept to development and taking ownership of project outcomes.

**Data Scientist – Whiterock.ai (Jan, 2022 – May, 2022 Manhattan, NY):**

* Conducted ETL and EDA on large real-estate datasets to identify key insights and features on past and present markets.
* Automated incoming data from various sources using Apache Airflow, GCP BigQuery, and Google Cloud Storage
* Coordinated with CTO to maintain daily model clusters, build predictive models for real-estate products, and research different models such as Neural Networks.

**Junior Data Scientist – ProMarketingHub (2020-2021 Queens, NY):**

* Managed and stored user data on cloud database, and performed daily ETL, data cleaning, and preprocessing.
* Worked closely with Sr. and Lead Data Scientists to generate and test hypotheses related to product engagement.
* Defined real-time customer data needs, evaluated data quality, and determined suitability for use.

**Data Analyst – Centerplate (2016-2020 Elmont, NY):**

* Conducted data entries, sorting, and analysis of over 10,000 client data to improve customer engagement.
* Increased customer orders by 13% by incentivizing coupons and combo deals.
* Analyzed food data to optimize pricing and maximize company profits.
* Presented data visualization of customer habits and findings to stakeholders, driving further growth.

# EDUCATION

**Springboard Data Science Bootcamp –** O**nline (2021):**

## Completed a comprehensive program in the full Python Data Science Stack, including Data Wrangling, Statistical Inference, Supervised and Unsupervised Machine Learning, Deep Learning, SQL, A/B Testing, etc.

**Queens College - NY, Queens (2015** – **2020):**

## Double Major: Bachelor’s in Computer Science and Applied Mathematics

## Relevant Coursework:

Object-Oriented Programming, Data Structures and Algorithms, Database Systems, Computer Architecture, Software Engineering, Internet/Web Technologies, Theory of Computation, Probability and Statistics, Bayesian Modeling, Linear Algebra, Linear Programming, Advanced Calculus, Machine Learning in R, Blockchain Mathematics.

# PROJECT EXPERIENCE

## Designed and built Convolutional Neural Network for Image Recognition – Classification

* Built a complex neural network using the Keras API, capable of classifying 315 species of birds with a 94% accuracy.
* Utilized Transfer-Learning with VGG16 and hyper-parameter tuning to achieve a further 4% improvement, resulting in a 98% accuracy in predictions.

## New York Housing Price Prediction – XGBoost Decision Tree

* Implemented advanced imputation techniques such as MICE to clean over 75K invalid and missing data points in Zillow's housing data, leading to a more accurate representation of the NYC housing market.
* Conducted extensive data analysis to provide insights into key factors impacting the housing market, such as wealth gap, cost of homes, and migration patterns of native residents.
* Evaluated multiple predictive models including Linear Regression, RandomForest, and KNN, using Mean Absolute Error as the evaluation metric. Conducted hyper-parameter tuning on each model to find the best result, ultimately achieving the best performance using XGBoost Gradient Boosting.

# TECH STACK

Python: Pandas, NumPy, OOP, SciPy, Statsmodel, Tensorflow, Keras, Flask, requests, PyTorch

Machine Learning: Neural Networks, Regression, Decision Trees, RandomForest, Classification, RandomForest, NLP, Git Scikit-Learn, OpenCV, R, ABBYY.

SQL: NoSQL, sql-lite, Microsoft SQL Server, DBMS, PostgressSQL, DML

Visualization: Tableau, Seaborn, Matplotlib.

Cloud: Google Cloud Platform, BigQuery, AWS, MongoDB, Databricks