

Benson Duong

323-446-5493 | bensonduong007@gmail.com | [linkedin.com/in/benson-duong-36552a180/](https://www.linkedin.com/in/benson-duong-36552a180/) | benduong2001.github.io
15529 Florwood Ave. Lawndale CA 90260 | Los Angeles Area | US Citizen

EDUCATION

University of California San Diego

Data Science, B.S.

Sept 2019 – May 2023

EXPERIENCE

Business Analyst Intern

June 2022 – Aug. 2022

Avanir Pharmaceuticals

Aliso Viejo, CA

- **Extracted** from **Microsoft SQL** database and **Snowflake cloud data warehouse** with **SQL**.
- **Transformed** and cleaned the extracted SQL data with **Pandas** and **NumPy** for **feature engineering** and **data analysis**. Developed **machine learning models** that predicted sales and client engagement at 78% accuracy with **Sklearn Pipelines**, and ranked influential business factors with **SHAP** and **statistical t-testing**.
- Developed GIS Python scripts for animated, interactive Tableau map dashboards on regional business growth of sales reps, easily updated with new data. Provided my maps to the parent company as consulting visualizations. Showcased and communicated my findings and recommendations to stakeholders and executive leadership.

SCHOOL PROJECTS

[Offer Acceptance Industry Capstone](#) | *Pandas, Sklearn, Geopandas*

Fall 2022 – Winter 2023

- Worked in a 6-month capstone group project mentored under FlockFreight, a logistics delivery company. Optimized their decision-making for accepting and reneging business offers, through feature engineering using GIS methods and data analysis, and developed a final machine learning prediction model reducing costs by 9.8%. Handled nuanced data issues such as imbalanced data, sample-weighting, and log-normalizations
- Showcased the project to faculty and industry professionals and authored its research paper for FlockFreight's use.
- Taught about reinforcement learning techniques leveraged by Flock Freight- multi-armed bandits and Q-learning.

PySpark Text Review Sentiment Classification | *PySpark*

Winter 2021

- Used Pyspark to conduct data analysis and NLP feature engineering on 20 GB+ Amazon review text-data, and train a machine learning model to predict customer satisfaction with Decision Trees, and handled data integrity.

INDEPENDENT PROJECTS

[Food Recommendation](#) | *Recommender System, NLP, CNN, Tensorflow, Pyspark, Unsupervised ML* Winter 2023

- Analyzed restaurant review text data and built ML models that predicted user-restaurant recommendations (collaborative filtering) with 73% accuracy, and customer satisfaction (sentiment analysis) with 80% accuracy.
- Performed customer segmentation dividing restaurant customers into distinct food-based subgroups on review data, using unsupervised ML clustering. Organized code into extensible OOP code, with Pyspark.
- Prepared unstructured review text data with NLP feature engineering (e.g. TF-IDF & custom-trained Word2Vec), and extracted text from review images with a pre-trained, image-labelling neural network.

[NYC Traffic Prediction](#) | *ArcGIS, GeoPandas, Flask, Leaflet.js, Sklearn*

Fall 2021 – Summer 2022

- Developed a Python app that can predict NYC street traffic at any given street and time of day (83% Accuracy), with clickable NYC street-map & clock-input using Flask & leaflet.js (Front-end GIS); used GeoPandas for GIS data preprocessing, & Sklearn for machine learning (Logistic Regression, PCA).
- Authored a Kaggle article tutorial on how to replicate project with ArcGIS using open source NYC government geodata, with matplotlib data visualization and ANOVA testing

TECHNICAL SKILLS

Python: Pandas, NumPy, Sklearn, Keras, Tensorflow, PyTorch, Dask, PySpark, Flask

SQL: Snowflake Data Warehouse, Microsoft SQL Server Management Studios, CTE, Windowing

Data Visualization: Tableau, D3.js, Matplotlib, Seaborn

HTML: CSS, JavaScript, Leaflet.js, BeautifulSoup, Selenium, Webscraping

Others: ArcGIS, QGIS, Geopandas, Microsoft Excel (VLookup, Pivot Table), Word, Powerpoint, R, Docker