

SUMMARY

4 years of industry experience in data analytics and machine learning, specialized in patient journey and customer engagement analytics. Expert in developing **advanced SQL queries (Snowflake and Redshift)**, conducting data manipulation with **Python**, and generating story-telling visualizations with **Tableau**. Proven ability to handle large datasets and proactively excel in fast-paced, dynamic environments.

WORK EXPERIENCE

LuxCouture Wellesley, MA

May. 2024 – Present

Data Consultant

- Deployed an automated **full-stack analytics pipeline** using **Shopify API**, **GA4**, and **Python** to ingest, preprocess, and visualize real-time sales and customer behavior, driving a 49% increase in sessions and 43.1% revenue growth YoY.
- Developed a **K-means segmentation model** and **GPT-driven** interpretation engine to recommend cluster-specific products and seasonal campaign strategies.
- Built a sales & customer insights dashboard with **Streamlit** to visualize sales performance, customer characteristics, and traffic attribution.

Putnam Associates Boston, MA

Apr. 2022 – May. 2024

Junior Data Scientist

- Developed and optimized **SQL queries** (in **Snowflake**, **AWS Redshift**) to streamline treatment journey analysis by integrating multi-source claims data. Generated key patient features and provided actionable insights through **Tableau** dashboards, supporting patient profiling and market mapping.
- Automated monthly analytics processes, reducing manual effort by 50%; enhanced treatment identification algorithms, improving accuracy by 5%.
- Implemented data cleansing, imputation techniques (**K-NN**, **Random Forest**), and advanced machine learning algorithms (**Random Forest**, **XGBoost**, **H2O AutoML**) to predict physician prescribing behavior, achieving 89% accuracy.
- Implemented statistical models (**K-means**, **Bayesian networks**, **Random Forest**) to segment patient journeys, optimizing patient opportunity for client's product portfolio.
- Spearheaded and contributed to 12 analytics projects on patient journeys and market opportunities in oncology and rare diseases by leveraging large-scale datasets, which demonstrated strong multi-tasking and collaborative ability.

Vee New York, NY

Dec. 2021 – Apr. 2022

Data Science Intern

- Built an ETL pipeline for candidate evaluation data in **AWS Lambda**, and established **API Gateway** to **Power BI** dashboards for automated updates, reducing 20% recruiting time.

PathosAI, Medford, MA – [Github Link](#)

Feb. 2021 – May 2021

Data Science Intern

- Developed an LSTM-based **deep learning** model with **Keras**, improving emotion detection accuracy by 16%.
- Designed **BI** dashboards showcasing model's result, including experimental tracking and attention score analysis.

EDUCATION

Northeastern University *Operational Research* (GPA 4.0/4.0)

May 2021

Shandong Agricultural University *Information and Computing Science* (GPA 3.5/4.0)

June 2019

TECHNICAL SKILLS

- Language and Script: Python, SQL, R, Jupyter Notebook
- Database: Snowflake, Redshift, PostgreSQL
- Data Science Libraries: NumPy, Pandas, Scikit-learn, Plotly, TensorFlow, Keras, SciPy, NLTK, H2O
- Data Visualization Tools: Tableau, Streamlit, Power BI, MS Excel, Matplotlib, Seaborn
- Cloud-based Technologies: AWS (Redshift, S3, Lambda, API Gateway, Athena), Google Cloud Platform