

Mirav Parekh

miravparekhedu@gmail.com | +1 (857) 424-9418 | [LinkedIn](#) | [Tableau](#)

Summary

Data engineer with experience in ETL processes, cloud-based data infrastructure, and large-scale data pipelines using SQL, Python, and AWS. Progressed from analytics to engineering roles across healthcare and technology sectors, improving data performance and driving data-based decisions. Bridges technical teams and business stakeholders to deliver scalable solutions.

Professional Experience

Zenith Technologies

Jan 2024 - Current

Data Engineer

Boston, MA

- Architected data models and dashboards with product managers to track **KPIs**.
- Engineered automated **data ingestion** and transformation workflows, enhancing efficiency by **20%**.
- Developed and tuned complex **SQL** queries for database management, reducing query execution time to seconds.
- Leveraged **AWS Glue** to automate **ETL** processes for data feeds, reducing processing time by approximately 40%.
- Constructed logging and monitoring systems across five data **pipelines** to ensure data reliability.

Beth Israel Deaconess Medical Center

Jul 2023 - Dec 2023

Healthcare Data Analyst

Boston, MA

- Coordinated with senior investigators in cardiovascular research studies, leveraging **R & SAS** for analytical support.
- Improved data cleaning and pre-processing strategies in **R & Python**, improving accuracy by **50%**.
- Applied statistical methods, including **t-tests**, **Chi-squared** tests, and **correlation** analysis to analyze clinical data.
- Executed ETL of **clinical** data handling **1M+** patient records, shaping methodology for publication and grants.
- Pioneered **NLP** techniques to extract, analyze, and classify over 1M+ clinical data and create structured cohorts for research and clinical insights.

Voda.ai

Sep 2022 - Dec 2022

Data Scientist Intern

Boston, MA

- Formulated a **GitHub** workflow for code checks, streamlining the development process, improving efficiency by 25%.
- Optimized 10+ Python scripts and Jupyter notebooks for data cleaning and analysis using **PyCharm**.
- Conducted statistical analysis and **A/B testing** to validate data models and data accuracy.
- Integrated **APIs** with Python libraries for data manipulation and retrieval in time series and tabular formats.

ImEx Cargo LLC

Apr 2022 - Jul 2022

Business Data Analyst

Peabody, MA

- Implemented **forecasting** model and dashboards in Power BI to analyze sales performance for four business units.
- Streamlined data mining processes with **Alteryx**, improving efficiency by 30%, deriving insights from customer data.
- Performed analysis of 3 years of sales data utilizing **Power BI**, supporting strategies to reduce operational costs.
- Utilized **Salesforce CRM** to manage data and sales pipeline, optimizing and improving sales reporting accuracy.

Projects

Airbnb: Barcelona Price Prediction

- Conducted data collection, preprocessing, and model evaluation, achieving high prediction performance through hyperparameter tuning and **cross-validation**.
- Built and optimized prediction models, **XGBoost**, **Logistic Regression**, **Random Forest**, **KNN**, and Linear Regression in Python, achieving 84% accuracy in predicting Airbnb prices in Barcelona.

Stock Trend & Feed Analysis

- Built supervised **regression algorithms** (decision trees, random forest) and performed sentiment analysis on stock feeds to predict returns and assess price changes.
- Developed **ML** model achieving 82% accuracy in predicting customer churn, resulting in 15% reduction in customer attrition.

Education

Northeastern University

Sep 2021 - Jul 2023

Master's in Analytics

Boston, MA

Coursework: Probability Statistics, Big data, Fundamentals of AI, Predictive Analytics, Risk Analysis, Data Science

Skills

Languages	: Python, R, Git, MySQL, SAS, Linux, NoSQL
Tools	: Tableau, Power BI, Alteryx, Looker, dbt, Jira, Redshift, S3, Quicksight, BigQuery, Salesforce CRM, Oracle
Certifications	: AWS, Jira, SQL, dbt, Python
ML Concepts	: Classification, Regression, Clustering, Hypothesis testing, Predictive analytics, A/B testing
Libraries	: Scikit-Learn, Keras, TensorFlow, PySpark, PyCharm, OpenCV, Pandas, NumPy, Matplotlib, Seaborn, PyTorch