

Vedant Shah

(551) 358-4846 | imshah.ved2604@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Analytical and technically proficient Data Analyst with a strong background in leveraging **Python, SQL, and BI tools** to translate raw data into business intelligence. Skilled at collaborating with **cross-functional teams** to deliver scalable data solutions that improve process efficiency, customer experience, and operational performance. Proficient in cloud-based analytics using **Azure and GCP**, with a focus on creating data-driven value across domains such as **logistics, healthcare, and digital services**.

TECHNICAL SKILLS

Programming Languages: R, Python, SQL, HTML, CSS, JavaScript, NodeJS, C, Java, Git

Data Science Libraries: NumPy, Pandas, PySpark, Matplotlib, Scikit-Learn, TensorFlow, Seaborn, ggplot2

Databases: MySQL, PL/SQL, Redis, Neo4j, SQLite

Cloud: Azure, GCP

Tools: Power BI, Tableau, Looker, SAS

WORK EXPERIENCE

OnTrackNorthAmerica, Data Analyst, Philadelphia

Feb 2025 – Present

- Structured scalable ETL pipelines using **Python and SQL Alchemy** to ingest and integrate large volumes of logistics and regulatory data from APIs and relational databases, significantly reducing manual intervention and improving processing efficiency by **70%**.
- Formulated comprehensive data transformation workflows with **Pandas** to clean, normalize, and validate multi-source datasets, enhancing data reliability for transportation policy analytics.
- Performed in-depth **exploratory data analysis** using Python (NumPy, Pandas) to identify inefficiencies in supply chain routes, which led to optimized transportation planning and a **30%** improvement in report turnaround time.
- Assembled lively **Power BI dashboards** to monitor KPIs such as emissions, route delays, and regulatory compliance, enabling real-time stakeholder insights and contributing to a **20%** increase in project ROI.

CSS CORP PVT LTD, Data Analyst, Chennai

Jun 2021 – Jul 2022

- Outlined and optimized advanced **SQL queries** to cleanse, aggregate, and normalize operational and customer support data within SQL Server, improving reporting accuracy and reducing query execution times by **40%** across business units.
- Implemented automated data validation workflows leveraging SQL rule engines to ensure regulatory compliance and data integrity, resulting in a **35%** increase in ETL process accuracy and eliminating over **10 hours** of manual auditing weekly.
- Conducted statistical analysis on **A/B testing** datasets using Python to evaluate web engagement metrics such as bounce rate, CTR, and session duration, generating insights that drove a **20% uplift in user retention**.
- Created insightful **Tableau dashboards** to deliver real-time visibility to project manager into product KPIs and experimentation results, facilitating agile decision-making for cross-functional product and UX teams.

ACADEMIC PROJECTS

E-Commerce Customer Analytics & Predictive Modeling | STATSMODELS, MYSQL, PYTHON, PANDAS, NUMPY, SCIKIT-LEARN, MATPLOTLIB, SEABORN

- Curated and preprocessed over **1 million** customer records from multiple relational tables using Python and SQL, resulting in a clean, analysis-ready dataset for segmentation and forecasting.
- Applied **K-Means clustering** on RFM (recency, frequency, monetary) metrics to segment users into five distinct personas, enabling targeted marketing and campaign personalization.
- Constructed a sales forecasting model using **Holt-Winters exponential smoothing** with a **10% Mean Absolute Percentage Error (MAPE)**, supporting inventory optimization and short-term demand planning.

EHR Data Visualization | STATISTICAL MODELING, PYTHON, REDIS, ODBC DRIVER, POWER BI

- Established a scalable analytics system to process **100,000+** Electronic Health Records (EHR), integrating **Redis** for real-time data streaming and caching to boost processing speeds by **60%**.
- Trained **predictive models** to identify key cardiovascular risk factors (e.g., cholesterol levels, BMI, blood pressure and others) from EHR data, achieving **88%** model accuracy that enhanced early risk detection.
- Crafted six Power BI dashboards with advanced filtering and real-time KPI tracking, optimizing data accessibility for clinical staff and administrators with a zero-learning-curve interface.

Custom Chatbot with language Translation using RAG | STREAMLIT, LANGCHAIN, HUGGING FACE, FAISS, RAG, GOOGLE GENERATIVE GEMINI AI

- Built a full-stack chatbot using **RAG and Google's Gemini AI** to enable multilingual question answering, providing **97%** accurate translation across **10+ languages**.
- Enhanced document retrieval performance by integrating **FAISS with HNSW indexing**, reducing query response times by **40%**.
- Developed a **BERT-based** context-aware tokenization pipeline using a sliding window method to maintain document understanding in multilingual outputs with **92%** context retention.

EDUCATION

Rowan University, Master of Science in Data Science

Jan 2023 – Dec 2024

Jain University, Bachelor of Technology in Computer Science and System Engineering specialization in IoT

Jul 2018 – Jul 2022