

VATSAL KADAKIA

Email:

vatsalkadokia20@gmail.com

Phone: +1 813-843-9863

LinkedIn:

<https://www.linkedin.com/in/vatsalkadokia/>

Github Portfolio: <https://vatsalkadokia.github.io/>

Data Scientist and analyst **with 2 years of** deep expertise in data engineering, machine learning, and data visualization. Proven ability to effectively translate business needs into technical solutions, optimize processes for maximum efficiency, and deliver projects on time and within budget. Skilled in handling large-scale data pipelines and proficient in a wide range of data engineering tools and platforms.

PROFESSIONAL EXPERIENCE:

Data Science Intern, Changing the Present, New York

(Aug'24 – Current)

- Developed engagement-insights ETL pipelines using SQL (AWS Redshift) for data pipelines and QuickSight for visualization, enabling daily tracking of metrics across various business domains and locales, resulting in 3 FTE saves and reducing manual intervention by 25%.
- Facilitated BI requirements, and synchronized efforts with the Central Data team, leading to a 50% increase in data accessibility and 30% enhancement in decision-making efficiency.

IT Intern, Minerals Technologies Inc, Pennsylvania

(Jun'23 – Dec'23)

- Established predictive ML models, Pandas and NumPy for data processing and compared accuracy of Regression and Classification models; analyzed real-time customer sentiment, increasing actionable insight by 40%.
- Engineered geospatial data visualization in Oracle Analytics Cloud (OAC) to visualize Pet Care product distribution, tool is now utilized by 4+ analysts for strategic resource allocation streamlining product management.
- Revamped data project with OAC's ML module, slashing duplicate data by 30%, facilitating report and dashboard creation.

Data Analyst Intern, Edvicer Pvt Ltd, Mumbai

(Jan'22 – May'22)

- Collaborated in data transformation project involving over 3800+ student organizations using Google Sheets. Extracted, cleaned and analyzed data using Matrixify(Shopify app), VLOOKUP and Pivot Tables leading to better data management.
- Improvise ARIMA and GARCH (Time Series Analysis) models using R to forecast student's spending power on private courses, significantly improving accuracy score by 46%.
- Performed EDA and residual analysis on student financial dataset to assess seasonality and stationarity using ACF, PACF, QQ plots achieving p-value <0.05.

Web Developer Intern, Thakur College of Engineering and Technology, Mumbai

(May'20 – Jun'20)

- Inhouse opportunity to develop university website using HTML, CSS, JavaScript, NodeJS, Express and MongoDB for database.

EDUCATIONAL QUALIFICATION:

M.S in Data Science	Stevens Institute of Technology, Hoboken, New Jersey	Aug'22 – May'24
B.E in Computer Engineering	Thakur College of Engineering and Technology, Mumbai	Aug'18 – May'22

SKILLS:

Programming Languages: Python, SQL, PL/SQL, R, PySpark, MongoDB

Data Engineering Tools: Snowflake, Pandas, NumPy

Machine Learning Models: Decision Trees, Naive-Bayes, Linear Regression

Machine Learning Frameworks: TensorFlow, PyTorch, Scikit-learn

Cloud Platforms: AWS - GLUE, Lambda, EMR, RDS, S3, Redshift, Python
Lambda, GCP - BigQuery, VertexAI

Visualization Tools: Tableau, Power BI, QuickSight, Elastic-Kibana

Data Science Techniques: Data Modeling (Relational Models, Star and Snow Schema), Data Preparation, Transformation, and Processing

Big Data Technologies: Spark

Machine Learning Operations: AWS environment, Generative AI techniques

RECOGNITIONS:

Published 2 papers in De Gruyter and Springer with undergrad final year project team. Project Title: Automated Healthcare System using Gamification and AI based Chatbot.

PROJECTS:

- "You are What You Eat": Customer segmentation by developing Regression algorithms.
- Compressing Feature Space: Developed a classification model using PCA to reduce 100 features to 24.
- Quantifying Sales Uplift: Built counterfactual model using Causal Impact Analysis.
- Alcohol Product Relationship: Analyzed 3,500 transactions using Apriori algorithm