

Kaushal Joshi

[LinkedIn](#) | [GitHub](#) | [Portfolio Website](#) |
kaushalvjoshi37@gmail.com | +91 8104942129 | Mumbai, India

EDUCATION

Nilkamal School of Mathematics and Applied Statistics-NMIMS, Mumbai, INDIA

May 2023-Aug 2025

Master of Science, Statistics and Data Science

CGPA-3.4/4.0

Courses: Machine Learning, Tableau, Cloud Computing, Big Data, Data Visualization, NLP, Python, SQL, AWS

D.G. Ruparel College of Arts, Science and Commerce, Mumbai, INDIA

Aug 2020-April 2023

Bachelor of Science, Statistics

CGPA- 9.09/10.0

Courses: Statistics, Java, Python, R programming, SQL, Statistical modelling, Hypothesis testing

TECHNICAL SKILLS

- **Interests and Domain:** Deep Learning, Machine Learning, NLP, Data Analysis, Data Science, Generative AI, LLM
- **Languages and Frameworks:** Python, R, MySQL, SAS, Power BI, Excel, MS Power-Point, Git, Tableau
- **Libraries:** Pandas, NumPy, Stats Models, Sklearn, Matplotlib, Seaborn, Scikit, Beautiful Soup, Keras, TensorFlow, Langchain
- **Tools and Infrastructure:** Power BI, AWS, Tableau, Streamlit, ChatGPT, GitHub Copilot, GitHub

WORK EXPERIENCE

• **Data Science/AI Intern at Coding Jr (Planto.ai), Mumbai, India**

June 2024–Aug 2024

- Explored **10+ open-source LLMs (LLaMA, GPT, etc.)** for fine tuning to build **copilot** for developers, while considering costs.
- Learning to develop custom copilots for specific domains by applying advanced model **fine-tuning** and **RAG** methods.

ACADEMIC PROJECTS

• **Resume analyzer for Internship/job recommendation using NLP | Python, Excel, Prompt Engineering, Streamlit, BERT**

An automated job recommendation system for data science roles which recommend job/internship links based on the similarity between the job description and candidates resume and integrated an LLM using Gemini API and prompt engineering for providing suggestions to improve the resume score.

• **Sentiment analysis for British Airways | Python (Pandas, NumPy, Scikit-learn), NLTK, NLP, Web Scraping**

This project categorized **50000 + web scraped** Skytrax reviews into **positive, neutral, and negative sentiments** using **NLP** and **Random Forest**, outperforming other **machine learning models** (i.e. Logistic Regression, Naïve Bayes, etc.) achieving an average **F1 score** of **84%**.

• **Credit Risk Modelling by Logistic regression | Python (Pandas, NumPy, Scikit-learn), Matplotlib, Supervised ML Metrics**

Developed a **Probability of Default** model using logistic regression on **466000** data rows with feature engineering (**WoE, IV**) and evaluated model performance using **ROC, AUC, PR, and KS test**.

• **Enhanced Brain Signal Classification with Advanced Feature Extraction | Python, Excel, 3D Image data**

Improved right vs. left hand brain signal classification by 50% using Whitening Transform with Gram-Schmidt Orthogonalization, enhancing feature extraction and model accuracy with Eigen Face Analysis.

• **Power BI Dashboard for Hardware Company | Python, SQL, Power Query, Power BI, DAX, Excel, Lean 6 Sigma, KPI**

Develop and maintain **custom reports, dashboards** and use **visualizations**, automated **ETL** processes to **optimize sales reporting**, enhancing data accuracy and decision-making speed using **KPIs**. Employed **DMAIC** for optimization and to improve and address inconsistencies and designed a **Star Schema**, uncovering a **56%** revenue drop with actionable regional insights.

• **EDA for Bank Marketing Dataset | Python (Pandas, Matplotlib), Streamlit, Data Preprocessing, Statistical Analysis**

Conducted comprehensive exploratory data analysis (EDA) on a bank marketing dataset, uncovering key insights and trends. Developed an interactive GUI to effectively showcase findings and facilitate data-driven decision-making.

• **Airbnb Dataset EDA and Clustering | Python, Pandas, Scikit-learn, Clustering algorithms, Reporting and Story telling**

Conducted **exploratory data analysis** on the Airbnb dataset to assess business trends across New York City locations. Implemented clustering techniques to identify latent features, **enhanced feature engineering**, and created a **report summarizing the clustering analysis**.

• **People Analytics: Employee Attendance and Leave Trends | Tableau, Excel, SQL, Reporting and Story telling**

Analyzed attendance data for **200+** employees using **SQL** and **Python**, tracking patterns in leaves, **work-from-home**, and **hybrid models**, and created a **Tableau dashboard** that identified a **15%** increase in **leave requests** on Mondays. Provided data-driven recommendations to **HR**, improving **workforce planning** and which can reduce absenteeism by **10%**, while ensuring data confidentiality.

LEADERSHIPS AND ACHIEVEMENTS

- Achieved **2nd Rank** in PPT competition for the topic: "**How Statistics revolutionized Sports Industry**".
- Got **SAS Certified specialist: Base Programming Using SAS 9.4** with a high score of **945 out of 1000**.
- Got **SQL Advanced Certificate** from Hacker Rank for completing the SQL advance questions.
- **My hobbies** are Chess, Gamer, Statistics, Astrophile, Philosophy, Psychology, Personal finance, Spirituality, Cosmology