Soham Ghosh

soham7704 | Soham Ghosh | My Portfolio | sohamghosh7704@gmail.com

Summary

B.Tech Computer Science student with a strong interest in turning data into meaningful insights. Skilled in **Python**, **SQL**, **Advanced Excel**, and **Power BI** for analyzing and visualizing data effectively. Exploring the world of **Generative AI** and building smart applications using frameworks like **LangChain** and **Streamlit**. Passionate about combining analytics and AI to solve real-world problems in creative ways. Eager to begin a career as a Data Analyst, bringing curiosity, adaptability, and a drive to learn.

Technical Skills

Key Skills: *Gen AI, Python, SQL, Java, HTML, CSS, Excel, Power BI*

Frameworks: *Langchain, Streamlit, NumPy, Pandas, Matplotlib, Seaborn*

Tools & Platforms: *Jupyter Notebook, Git, VS Code, Google Colab, DBeaver, Netlify*

Data Engineering: *ETL, Data Cleaning*

Projects

RepoSage: AI-powered tool designed to analyze GitHub Repos — *(Streamlit, Python, LangChain, Groq API, FAISS, HuggingFace Sentence Transformers)*

- Built a Streamlit-based UI that **auto-fetches repository content** and supports **memory-enabled follow-up** queries for seamless developer learning.
- Leveraged **retrieval-augmented generation (RAG) techniques** with **HuggingFace embeddings** to deliver fast, accurate insights into public codebases—ideal for educational and analytical use cases.
- Integrated **natural language processing (NLP)** capabilities to interpret developer queries and deliver precise, code-aware answers in real time.

AI-Career-Advisor: Resume and a Job Description Parser — *(Streamlit, Python, LangChain, Groq LLM, Streamlit, PyMuPDF, dotenv)*

- Achieved a **95% match-score prediction accuracy**, empowering users to quickly assess how well their resume aligns with target job descriptions.
- Delivered **85% precise skill-gap identification** to highlight areas for improvement and enhance resume relevance.
- Generated personalized resume recommendations that **improved candidate keyword alignment by 70%**, boosting their chances of standing out.

Financial_analysis: Stock Price & Volatility Forecasting — *(Pandas, NumPy, statsmodels, yfinance, plotly, ARIMA, SARIMA, GARCH, XGBoost, LSTM)*

- Implemented **time-series forecasting, volatility estimation, and deep learning techniques** to achieve **85%+ accuracy in predicting Tata Motors stock trends**, enabling data-driven investment decisions.
- **Integrated real-time data pipelines with yfinance**, enabling instant **7-day forward forecasting** for dynamic investment decision support.
- **Conducted rigorous model diagnostics (ADF, ACF/PACF, Ljung-Box) for robust validation** and enhanced model trustworthiness.

Cafe Sales Analysis: SQL-based exploration of a retail cafe's sales dataset — *(MySQL, IDE: DBeaver)*

- **Elevated business intelligence through 30% faster trend detection**, using advanced aggregations, window functions *(e.g., LAG, RANK), and time-series analysis*.
- **Enhanced decision-making by forecasting 30% monthly revenue growth** insights and peak sales hours across branches.

Education

University of Engineering & Management, Kolkata — *2022–2026 (Expected)*
B.Tech in Computer Science Engineering, GPA: 7.77/10 (Till 5th Semester)

Julien Day School, Howrah, ISC — *2022*
Score: 74.66%

Certifications & Achievements

- HackerRank SQL Basic Skill Certification Test
- HackerRank 5 Star Gold Badge in SQL
- German Language A1 Certificate