SUMIT KUMAR

sumit.csudh@gmail.com | +1(714) 257-5339 | Los Angeles, CA | wulfric06.github.io/portfolio2 | Linkedin.com/in/sumit2306

Data Analyst with 2+ years of experience in SQL, Python, and BI tools (Power BI), & Tableau, passionate about leveraging analytics to drive decisions in the entertainment industry. Proven track record in ETL, dashboarding, and ML-driven insights.

SKILLS

Languages: Python, SQL, Java, JavaScript, TypeScript
Analytics Tools: Power BI, Tableau, Looker, Jupyter Notebook
Frame Works & Libraries: Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch, Matplotlib, Seaborn
ETL & Data Tools: AP BODS, MySQL, PostgreSQL, MongoDB, Redshift
Cloud & DevOps: AWS, Google Colab, Docker, Kubernetes, CI/CD, Git
Operating Systems: Microsoft Windows, Linux.

WORK EXPERIENCE

Research Assistant | CSUDH – IT Department   Carson, CA, USA | Mar 2023 – Present
- **Fine-tuned and deployed** Large Language Models (LLMs) including LLaMA, T5, and BERT using LoRA/QLoRA and prompt engineering, boosting model performance by 30%.
- **Developed and validated** ML models using statistical analysis to improve predictive accuracy by 30%.
- **Designed** a multi-stage RAG pipeline with BM25 and Cross-Encoder re-ranking to optimize query results.
- **Processed and analyzed** large-scale datasets (500K+ rows) for fine-tuning NLP models and enhancing inference quality.

ETL Developer | IBM India Pvt Ltd   Bengaluru, India | Jun 2021 – Nov 2022
- **Built and optimized** ETL workflows using **SAP BODS** and **complex SQL queries**, cutting manual intervention by 40%.
- **Automated** repetitive reporting tasks with Python scripts and Power BI, saving 30% in effort
- **Developed** SCD Type-2 logic for historical data tracking, ensuring high data integrity.
- **Created** executive dashboards in Power BI to monitor KPIs, increasing business insights accuracy by 35%.
- **Monitored** ETL pipelines and enhanced reliability through KPI tracking (data freshness, load failures).
- **Defined & tracked** KPIs for ETL workflows (e.g., data freshness, load time, failure rates), improving pipeline reliability by 30%.

Analyst | Enuke Software Pvt Ltd   Gurugram, India | Jul 2020 – Jun 2021
- **Extracted and manipulated** data using SQL (MySQL, PostgreSQL) and MongoDB for deep analysis.
- **Designed and maintained** ETL pipelines to ensure clean, accessible, and consistent data flow.
- **Developed** interactive dashboards in Power BI and Looker for both technical and non-technical stakeholders.
- **Led** ad-hoc analytics tasks, providing rapid insights for urgent business needs.

Student Intern | Nagarro Software Pvt Ltd   Bengaluru, India | Jan 2020 – Jun 2020
- **Performed** data cleaning and transformation using Pandas and NumPy for structured reporting.
- Queried large datasets from MySQL/PostgreSQL with performance-optimized SQL scripts.
- **Built** dashboards and trend visualizations in Power BI, enhancing decision-making.
- **Used** Matplotlib and Seaborn to create visual reports for data storytelling.

PROJECTS

Sales Performance Dashboard   **Python, SQL, Power BI**
- **Built** an interactive dashboard to analyze regional sales, track revenue trends, and identify top-performing products.
- **Preprocessed** raw sales data using Pandas, handled missing values and outliers.
- **Designed** Power BI visuals with slicers, KPI cards, and drill-downs for easy user interaction.

Fraud Detection System   **Python, Scikit-Learn, TensorFlow, PyTorch**
- **Developed** a fraud detection pipeline using Random Forest, XGBoost, and Neural Networks.
- **Implemented** feature engineering and anomaly detection to identify suspicious financial activities.
- Achieved high precision and recall in experimental evaluation of the models.

Financial Data Analysis   **SQL, Excel, Power BI**
- Analyzed 1M+ World Bank financial transactions to identify top borrowing countries, interest rate trends, investment impact.
- Correlation between transaction count and total debt; identified India, Pakistan, and Bangladesh as highest debt holders.

EDUCATION

Masters in Computer Science | Cal State – Dominguez Hills   Jan 2023 – May 2025
- **Relevant Courses:** Data Analysis, Software Engineering, AI, Data Structure and Algorithms, Operating Systems, Programming Languages (Advance Python)