SANCHI GUPTA

✉ gupta.sanch99@gmail.com | 📞 +1 617 251 5460 | ○ github.com/sanchigupta12 | 🔗 linkedin.com/in/sanchi-gupta99

SUMMARY

Data professional with experience in building ML models, data pipelines, and AI-powered systems. Skilled in SQL, Python, Tableau, Power BI, and cloud tools. Adept at stakeholder reporting, dashboarding, and translating technical data into business insights. Familiar with LLMs, AI compliance, and governance workflows

WORK EXPERIENCE

Data Analyst Intern - Capstone Project | Allagash Brewing Company   01/2025 - 03/2025
Boston, MA
- Developed a real-time truck recommendation system using weather API to reduce spoilage and ensure shipping quality
- Applied DBSCAN clustering to identify high-demand zones and optimize delivery routes
- Built a Power BI dashboard to monitor shipping trends, cost metrics, and delivery performance for cross-team visibility
- Presented actionable insights to logistics and operations teams, improving delivery performance

Data Analyst | Falcon Cables   09/2022 - 08/2023
New Delhi, India
- Analyzed sales and production trends to enhance inventory planning and targeted marketing
- Optimized inventory management using data-driven insights and interactive dashboards in Power BI, minimizing stock shortages and excess inventory

PROJECTS

Chicago Traffic Crash Analysis
Tools used: Azure Blob Storage, Databricks (PySpark), Snowflake, SQL, Tableau
- Built an ETL pipeline to ingest, clean, and merge 2M+ records using PySpark, with data sourced from Azure Blob Storage
- Conducted EDA to uncover trends in crash severity, time of day, lighting conditions, and driver behavior using Spark SQL
- Enforced schema validation and data normalization before storing transformed data in Snowflake for optimized querying
- Designed interactive Tableau dashboards to analyze crash trends, enhancing city traffic reporting and safety insights

Emergency Care Analytics Dashboard (Healthcare)
Tools used: Power BI, DAX
- Developed an ETL pipeline in Power BI Query Editor to process patient admission data from a pre-existing dataset
- Created calculated metrics with DAX to track wait times, admission rates, and departmental referrals
- Interpreted patient flow trends and provided insights to improve emergency care efficiency

University Chatbot for Academic Q&A
Tools used: OpenAI API, LangChain, OpenShift, Docker
- Contributed to an ongoing NLP-based AI chatbot using OpenAI and LangChain to convert natural language queries into SQL, increasing response accuracy by 20
- Implemented a microservices architecture with cross-functional teams and containerized services using Docker
- Documented limitations and collaborated with stakeholders to align with privacy and compliance standards

Auto Insurance Fraud Detection
Tools used: Python, XGBoost, AWS SageMaker, AWS EC2
- Built an XGBoost model on AWS SageMaker to detect auto insurance fraud, improving claim processing efficiency
- Enhanced model performance through targeted feature engineering and automated hyperparameter tuning
- Deployed the model on AWS EC2 for real-time inference and registered it in SageMaker Model Registry

A/B Testing – Simulated Clickstream Analysis
Tools used: Python (Pandas, NumPy, SciPy, Matplotlib, Seaborn)
- Simulated 20K+ user records to model marketing attribution via A/B testing on control vs. variant webpage designs
- Conducted Z-test with p-value and confidence intervals to assess statistical significance of conversion differences
- Visualized click distributions and rejection regions to effectively communicate experiment results

EDUCATION

Northeastern University (Boston, MA)   09/2023 - 03/2025
Master of Analytics; GPA: 3.97/ 4.00
Relevant Courses: Predictive Analytics, Data Mining, Database Management, AI Technologies, Data Visualization, Big Data

Guru Gobind Singh Indraprastha University (Delhi, India)   08/2017 - 05/2020
Bachelor of Business Administration; CGPA: 3.23/ 4.00
Relevant Courses: Business Analytics, Computer Applications, Project Management

TECHNICAL SKILLS

Python (Pandas, NumPy, SciPy, scikit-learn, XGBoost), SQL, Snowflake, Databricks, Apache Spark, Tableau, Power BI, AWS (SageMaker, EC2), OpenAI API, Docker, FastAPI, A/B Testing, Hypothesis Testing, PySpark, Git, GitHub