

SANKALP BISWAL

✉ sankalpbiswal99@gmail.com | 📞 +1 (617) 256 1543 | 🔗 linkedin.com/in/sankalp-biswal | 🐙 github.com/Sankalp20487

SUMMARY

Data Scientist with a Master's in Data Analytics (4.00 GPA) and experience at EY Consulting. Skilled in Python, SQL, and machine learning for extracting insights from complex data. Experienced in statistical methods like hypothesis testing and regression, and in working with both tabular data and natural language inputs to build AI-driven solutions.

TECHNICAL SKILLS

Python (Pandas, NumPy, scikit-learn, Keras, PyTorch, TensorFlow), SQL, MySQL, Athena, A/B Testing, Tableau, Power BI, Grafana, Databricks, Apache Spark, Azure, AWS, SageMaker, GCP, VertexAI, Flask, FastAPI, Streamlit, Docker, Git, GitHub, GitHub Actions, CI/CD Pipelines, Snowflake, Pinecone, Excel, Powerpoint, MLFlow

EXPERIENCE

Allagash Brewing Company | Data Scientist Intern Jan 2025 - Mar 2025
Boston, MA

- Developed a real-time truck recommendation Streamlit app using live weather and temperature data, reducing outbound shipping costs by 10% (\$150K annually).
- Applied geospatial DBSCAN clustering on route coordinates to uncover delivery patterns and support logistics optimization.
- Built an interactive Power BI dashboard to visualize route inefficiencies, truck type recommendations, and cost savings.

DataWorksAI | Data Scientist Intern - Conversational AI Apr 2024 - Jun 2024
Boston, MA

- Collaboratively built a RAG-based LLM chatbot using ChatGPT-3.5, LangChain, and advanced prompting (chain-of-thought, few-shot/zero-shot), improving query accuracy by 14%.
- Deployed Dockerized microservices on OpenShift with GitHub Actions CI testing 200+ SQL queries.

Falcon Cables | Data Scientist Feb 2022 - Aug 2023
New Delhi, India

- Built ELT pipelines in Databricks to ingest ERP and flat-file data into S3; cleaned and standardized it into raw, cleansed, and curated tables with daily schedules and automated data-quality checks.
- Published datasets via Databricks SQL to Power BI (Incremental Refresh) and designed lightweight operational dashboards for production, sales, and inventory, enabling reliable daily reporting.

Ernst & Young LLP | Analyst - Technology Consulting (Financial Services) Jun 2021 - Jan 2022
Gurugram, India

- Led vendor evaluation for \$11M digital transformation at Samunnati (NBFC), creating Excel-based comparison matrices for LOS, LMS, and CRM systems while gathering requirements from C-suite stakeholders.
- Developed industry research and market analysis for NBFC sector, creating knowledge resources and presentation decks that informed EY's fintech consulting approach and were presented at banking summits by senior partners.

PROJECTS

Text-to-SQL Agentic Healthcare Chatbot

- Built a RAG-based LLM assistant using LangChain, Pinecone, and Snowflake to generate insights and visualizations from structured healthcare data via SQL generation with Gemini.
- Performed ETL on 2M+ rows of NY SPARCS inpatient discharge data (HIPAA-compliant) and loaded into Snowflake.
- Engineered structured prompts for a zero-shot ReAct agent, creating precise step-by-step instructions for LLM reasoning.
- Enabled multi-tool agent workflows with memory and web search fallback for resilient, context-aware responses.

Customer Churn Prediction with AWS and MLops

- Built a churn prediction pipeline on 500K+ customer records using PySpark (AWS Glue), stored cleaned data in Parquet on S3, and validated outputs via Athena; integrated schema using Glue Crawler and Data Catalog.
- Trained and tracked classifiers (LogReg, RF, XGBoost) in SageMaker with MLflow, automated training pipeline with GitHub Actions, and monitored performance in Grafana.

Ad Campaign A/B Testing & Optimization

- Applied causal inference via A/B testing on 580K+ users, using Chi-Squared and Mann-Whitney U tests to uncover statistically significant conversion drivers.
- Measured 43% relative uplift in conversion for the test group and identified key impact from timing and ad volume.

EDUCATION

Northeastern University (Boston, MA) Sep 2023 - Mar 2025

Master of Science in Data Analytics (Applied Machine Intelligence) ; GPA: 4.00/4.00

Relevant Courses: Probability and Statistics, Data Mining Applications, Data Visualization, Data Management and Big Data, AI System Technologies, Predictive Analytics, Healthcare Data and Applications

Jaypee Institute of Information Technology (Delhi, India) Jul 2017 - Aug 2021

Bachelor of Technology in Computer Science Engineering ; CGPA: 7.3/10

Publication: Co-authored a research paper on diabetic retinopathy detection using CNN and KNN, achieving 89.72% accuracy with KNN and applying computer vision (cv2) for enhanced diagnostic precision.