

Mansi Bhadani

201-630-1123 | mansi.bhadani31@gmail.com | [LinkedIn](#) | [Github](#)

Summary

Data Analytics Engineer with 2+ years of experience building reliable ETL pipelines, data models, and cloud data solutions using Python, SQL, Airflow, Spark, and Snowflake. Skilled in data quality and observability, large-scale migrations, real-time streaming, and integrating structured and unstructured data to support accurate analytics and operational decision-making across consumer, financial, and transit domains.

EDUCATION

Pace University

Master of Science, Computer Science

• GPA: 3.82/4.0

Aug 2024 - May 2026

PROFESSIONAL EXPERIENCE

Beats by Dre

Data Analytics Extern

- Built automated ETL pipelines using Python, SQL, Oracle, and Airflow to integrate sales, customer behavior, and social media datasets; improved CRM data quality through validation and standardized data models for marketing performance dashboards and trend analysis.
- Processed large volumes of structured and unstructured data to support consumer segmentation and product feedback analysis.
- Applied NLP pipelines to transform raw reviews and social discussions into analytical features for business insights. Collaborated with cross-functional teams to translate business questions into scalable data workflows.

Mar 2025 - Jun 2025

Techcolabs Softwares Inc

Data Scientist Intern

- Built production-ready data preprocessing and feature engineering pipelines in Python and SQL on 1M+ mortgage records, applying relational data concepts and validation checks to ensure data integrity and scalability; developed machine learning models with 92% accuracy for prepayment risk prediction.
- Implemented ensemble models (Random Forest, Gradient Boosting) and optimized performance through hyperparameter tuning.
- Delivered analytical datasets and insights that improved risk assessment workflows by 30%, collaborating directly with stakeholders to capture business requirements, refine data models, and improve the quality of inputs used for analytical recommendations.
- Contributed to migration of large-scale datasets (~10TB) from Snowflake to an Apache Iceberg data lake on S3 using Spark for orchestration, integrating storage systems to support time-travel queries and reducing long-term storage costs by 60%.
- Built and deployed a real-time AI pipeline using Kafka Streams and a fine-tuned BERT model to classify and redact PII from 1M+ daily user-generated reviews and salary reports, optimizing performance and resource usage to ensure full CCPA/GDPR compliance.

Dec 2022 - Apr 2023

Innovative Research Labs

Data Engineer Intern

- Developed ETL pipelines for 100K+ user ratings and metadata using Python and SQL workflows, transforming source data into structured analytical datasets and validating records before downstream use.
- Developed Python scripts for RDS and Redshift migrations, cutting database update time by 98% through efficient SQL generation and reliable system integration.
- Built an LLM-powered automation framework for a drag-and-drop interface to create Airflow DAGs, using transformer-based models to parse natural language inputs into executable DAG configurations for AI agents development.
- Optimized dataset joins and aggregations, reducing feature generation runtime by 25%. Implemented evaluation pipelines using Precision@K and RMSE to continuously monitor model performance.

Jul 2022 - Oct 2022

NJ Transit Smart Management & Optimization System

Data Engineer Project

- Designed and implemented a real-time transit monitoring system with a high-throughput data pipeline using Kafka, Spark Streaming, and Snowflake, processing millions of transit events daily for operational decision-making.
- Implemented predictive analytics for delay and congestion forecasting, improving route optimization and operational efficiency by 30%.
- Built a data quality and observability framework with Great Expectations and OpenLineage, reducing debugging and anomaly detection time by 50% through automated validation and monitoring.

Jul 2025 - Oct 2025

Technical Skills

- **Programming Languages:** Python, Java, Scala, SQL
- **Cloud & Databases:** AWS, GCP, Snowflake, BigQuery, Redshift, PostgreSQL, Cassandra, Oracle
- **Big Data & Streaming:** Apache Spark, PySpark, Apache Kafka (Confluent), Apache Flink, Apache Iceberg, Hadoop, Hive, Avro
- **Orchestration & Tools:** Airflow (Astronomer), Kubernetes, Docker, Tableau, dbt, Great Expectations, AI Agents Development-LLM-Powered Applications, Model Context Protocol (MCP)
- **Data Engineering:** ETL Processes, Pipeline Development, Data Migration, Data Modeling, Data Warehousing, OLAP Technologies, CRM Data Quality Improvement
- **Integration & Automation:** API Design, Webhooks, System Integration, Sales and Revenue Operations Automation