

DISHA GOYAL

+1 (213) 783-4130 | dishagoyal2001@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

USC Computer Science graduate with hands-on experience analyzing and transforming complex datasets into actionable business insights using SQL, Python, Databricks, Tableau, and Power BI. Experienced in data cleaning, exploratory data analysis, KPI development, dashboard reporting, and business intelligence, with projects spanning healthcare, telecommunications, and environmental analytics. Passionate about leveraging data to solve business problems, identify trends, and support data-driven decision making.

EDUCATION

University of Southern California

Aug 2024 - May 2026

Master of Science in Computer Science - GPA: 3.75/4.0

Los Angeles, United States

Vellore Institute of Technology

Sep 2020 - May 2024

Bachelor of Technology in Computer Science - GPA: 9.29/10

Vellore, India

TECHNICAL SKILLS

Programming & Querying: SQL, Python, PySpark, Spark SQL, pandas

Databases: Databricks SQL, MySQL

Data Analysis: Data Cleaning, Data Validation, Exploratory Data Analysis (EDA), Statistical Analysis, Trend Analysis, Root Cause Analysis, Ad Hoc Analysis, KPI Development

Business Intelligence: Power BI, Tableau, Excel, Dashboard Development, Data Visualization, Pivot Tables, XLOOKUP/VLOOKUP

Data Preparation & Processing: ETL/ELT, Data Transformation, Batch & Incremental Processing, JDBC Ingestion, Data Quality Monitoring

Tools: Git, GitHub, Jupyter Notebook

PROFESSIONAL EXPERIENCE

Institute for Creative Technologies, USC

Jan 2026 - May 2026

Data Analyst

Los Angeles, United States

- Cleaned and validated 100+ WhisperX-generated negotiation transcripts, improving data quality for conversational AI research.
- Reviewed speaker attribution, transcript accuracy, dialogue completeness, and formatting inconsistencies to ensure reliable research datasets.
- Structured transcript data into standardized datasets supporting behavioral analysis, conversational AI research, and reporting.

Cosasco

Sep 2025 - Jan 2026

Data Infrastructure Intern

Santa Fe Springs, United States

- Analyzed engineering test data stored in MariaDB databases to identify reliability trends and support engineering investigations.
- Standardized repeatable database export procedures and structured engineering datasets, improving data consistency and traceability.
- Maintained test tracking documentation and data records supporting engineering reporting and troubleshooting.

Blackcoffer Pvt. Ltd.

Jan 2024 - Jul 2024

Backend and Data Analytics Intern

Delhi, India

- Supported development of client-facing analytics dashboards using the MERN stack, contributing to backend and data-handling workflows for dashboard applications.
- Worked with development teams to prepare and organize application data used in client-facing dashboard interfaces.

PROJECTS

Telecom Churn and Retention Analytics | Python, SQL, Machine Learning, SQLite, Tableau

- Analyzed **1,000** telecom customer records using Python and SQL to identify high-risk segments, quantify **\$805K** in annualized revenue exposure, and uncover churn patterns across contract type, tenure, pricing, and tech support.
- Developed customer value, churn-risk, and campaign-priority metrics, identifying **750** actionable customers and measuring campaign cost, expected retention benefit, and revenue at risk.
- Built an interactive Tableau dashboard and budget-aware campaign framework that ranked retention candidates and evaluated expected outcomes across multiple campaign budgets.

U.S. EPA Emissions Analytics Dashboard | Databricks SQL, Data Visualization

- Built a Databricks SQL workflow to ingest, clean, and transform 2023 U.S. EPA emissions data, converting string-based numeric fields and aggregating emissions across state and county levels.
- Calculated population-normalized emissions metrics and prepared dashboard-ready datasets to compare geographic emissions patterns beyond total emissions.

Healthcare Analytics & Patient Monitoring Platform | Databricks, PySpark, Spark SQL, Delta Lake, MLflow

- Built a healthcare analytics platform processing **60K+** patient, encounter, admission, clinical, and monitoring records, transforming raw healthcare data into analytics-ready datasets.
- Cleaned, validated, and integrated healthcare data using automated quality checks, deduplication, and referential-integrity validation to improve reporting accuracy.
- Developed Patient 360, hospital operations, and monitoring KPI datasets for dashboard reporting and readmission-risk analysis.

CERTIFICATIONS AND PROFESSIONAL DEVELOPMENT

- Databricks Certified Data Engineer Associate
- Databricks Academy - Data Analyst Learning Path