Aveg Ganorkar
Washington, D.C.| +1(240)398-1650 | avegganorkar@gmail.com | avegganorkar.com| github.com/DarakhTech

PROFESSIONAL SUMMARY

Data Analyst and Data Science professional with 4+ years of experience in financial analytics, fraud detection, data engineering, and business intelligence. Proficient in Python, SQL, PostgreSQL, Snowflake, Power BI, Apache Airflow, and dbt. Experienced in developing ETL/ELT pipelines, interactive dashboards, SQL queries, data models, and automated reporting solutions. Skilled in financial risk analysis, KPI development, data quality validation, and workflow automation to improve reporting accuracy, operational efficiency, and data-driven decision-making across industries.

TECHNICAL SKILLS

Programming & Query Languages: Python, SQL, PostgreSQL, Advanced SQL, Pandas, NumPy
Data Analytics & Business Intelligence: Tableau, Power BI, DAX, Power Query, KPI Development, Financial Reporting, Dashboard Development, Data Visualization, Funnel Analysis, Behavioral Analytics
Data Engineering: Apache Airflow, dbt, Snowflake, Spark, Kafka, ETL/ELT Pipeline Development, Workflow Orchestration, Data Warehousing, SQL Data Modeling, Incremental Data Loading, Data Quality Validation, Data Deduplication
Cloud Technologies & Automation: AWS Lambda, S3, Redshift, Glue, CI/CD, REST APIs, API Integration, Workflow Automation, GitHub, Slack Integrations
Data Science & Generative AI: Retrieval-Augmented Generation (RAG), pgvector, Vector Embeddings, Semantic Search, Knowledge Graphs, OpenAI API, Anthropic API, Data Preprocessing
Financial & Risk Analytics: Credit Card Fraud Detection, Credit Risk Analysis, Loan Portfolio Analytics, Financial Reconciliation, Anomaly Detection, Charge-Off Analysis, Transaction Monitoring
Data Quality & SQL Techniques: Data Cleaning, Data Validation, Data Quality Checks, CTEs, Window Functions, SQL Joins, CASE Statements, Aggregations, Schema Mapping
Product & Research Analytics: PostHog, Event Tracking, User Engagement Analytics, Data Collection, Research Data Analysis

PROFESSIONAL EXPERIENCE

Data Analyst
Capital One | USA | September 2025 - Present

- Built SQL and Python fraud-monitoring reports on 2.4M monthly card transactions, flagging high-risk merchant and time-of-day segments and cutting false-positive alerts by 14% through threshold tuning.
- Developed Snowflake pipelines using CTEs and window functions to track 30/60/90-day delinquency roll rates across 180K loan accounts, surfacing early-warning segments for the collections team.
- Designed a 4-page Tableau loan portfolio dashboard with drill-through covering charge-offs, repayment trends, and outstanding principal, replacing 5 manual weekly Excel reports and saving ~8 analyst hours per week..
- Standardized 12 risk KPIs (fraud rate, charge-off rate, net loss rate) across SQL and Tableau calculated fields, ending metric discrepancies between risk and finance teams.
- Automated Python (Pandas) data quality checks on daily transaction feeds, catching duplicates and null spikes and reducing data errors in daily reports by 27%.

Data Science & Analytics Intern
Drug Hunter | USA | June 2025 - August 2025

- Engineered Apache Airflow pipelines ingesting PubChem and ClinicalTrials.gov API data, integrating compounds, clinical trials, and publications into a normalized PostgreSQL database containing over one million biomedical records.
- Designed PostgreSQL schemas and SQL transformations to link biomedical entities across sources, enabling researchers to analyze relationships between compounds, clinical trials, and scientific publications consistently.
- Implemented incremental ingestion and deduplication so datasets refresh without full reprocessing, cutting pipeline runtime and manual curation effort by ~50%.
- Developed Retrieval-Augmented Generation (RAG) workflows using pgvector, OpenAI APIs, and Anthropic APIs to support semantic search, evidence retrieval, and automated summarization of biomedical research content.

Data Science Graduate Assistant
University Of Maryland | USA | January 2025 – September 2025

- Built a Retrieval-Augmented Generation (RAG) pipeline with pgvector embeddings, document chunking, and semantic retrieval to make foundational data science course materials easier to search and explore.
- Connected course content to a searchable knowledge graph, organizing relationships among learning resources and supporting contextual information retrieval for students and instructors across educational materials.
- Implemented PostHog event tracking and supporting data pipelines to capture student interactions, establish engagement metrics, and provide instructors with previously unavailable learning platform usage visibility.
- Analyzed student interaction events through behavioral and funnel analytics, identifying usage patterns and actionable opportunities for instructors to assess course engagement and learning platform adoption.

Data Analyst
Smart Tech Software | India | August 2023 – July 2024

- Developed Apache Airflow DAGs to orchestrate ELT workflows into Snowflake, consolidating restaurant-client point-of-sale, billing, and accounting data while improving preprocessing efficiency by 42%.
- Investigated financial anomalies using Python and advanced SQL reconciliation queries, identifying approximately $300,000 in billing discrepancies and automatically escalating exceptions to finance stakeholders for review.
- Built an AWS Lambda-powered Slack bot integrating GitHub, QuickBooks, and Toast POS APIs, giving stakeholders timely pipeline status updates and improving operational issue visibility.
- Created dbt models and advanced SQL transformations using CTEs, window functions, and aggregations to standardize business metrics and deliver consistent financial reporting datasets in Snowflake.

Data Engineer
White Heaven Entertainment| India | July 2022 – July 2023

- Owned data engineering infrastructure for an event management business, designing ingestion pipelines, API integrations, backend services, and warehouse structures supporting ticket sales and operational reporting.
- Integrated ticketing and sales data from Ticketmaster, BookMyShow, Eventbrite, StubHub, and Paytm Insider, mapping five incompatible platform schemas into a consistent analytical reporting data structure.
- Built SQL validation checks to reconcile ticketing records across platforms, investigate inconsistent transaction fields, and strengthen data reliability for consolidated event sales and revenue reporting.

EDUCATION

Masters of Science in Data Science| August 2024 - May 2026
University of Maryland, College Park | USA
Bachelor of Engineering (B.E) in Electronics and Communication | Aug 2019 - May 2023
Ramdeobaba University, Nagpur | India

PROJECTS

Pharma M&A Data Platform (Molecule to Mergers) | Python, PostgreSQL, Docker, Oracle Cloud, Azure, AWS - Link

- Built a distributed ingestion system scraping 31,533 pharma organizations across 15 cloud VMs, coordinated by a fault-tolerant PostgreSQL work queue running 50+ parallel workers with automatic retry. The normalized warehouse yields 3,761 verified M&A events via two-stage fuzzy entity resolution across 35,277 clinical trial records.

Real-Time Weather ETL Pipeline | Python, Airflow, dbt, PostgreSQL, Docker Compose, Tableau - Link

- Built a fully Dockerized ETL pipeline: API ingestion, Airflow DAG orchestration with retries and dependency management, dbt transformations with staging and deduplication layers, incremental loading, health checks, and live Tableau dashboards on PostgreSQL.

Automated Financial Transaction Pipeline | n8n, Node.js, Redis, Docker Compose, Oracle Cloud- Link

- Deployed a self-hosted automation platform parsing bank transaction emails (Chase, Amex, Discover) into structured datasets with Redis deduplication and real-time Discord alerts, running 24/7 at zero recurring cost.