

SANKET CHAVAN

Austin, TX
+1(716)578-9525

chavansanket193@gmail.com
www.linkedin.com/in/sanketchavan08/

SUMMARY

Data Analyst/Engineer with **5 years** of experience in banking and IT, specializing in SQL, Python, Data Warehousing, and BI tools including Tableau and Power BI. Designed and implemented a fully automated compliance monitoring solution, tracking **130+ broker feeds in real time**, eliminating **100% of manual processes**, and reducing compliance officer data-processing time by **80%**. Skilled at translating business requirements into scalable, data-driven solutions that enhance performance and inform strategic decision-making.

SKILLS

Programming:	Python, R, PySpark, Shell
Databases & Warehousing:	Oracle, MongoDB, PostgreSQL, Snowflake, Databricks
Cloud & Big Data Tools:	AWS (Lambda, Step Functions, S3, DynamoDB), Azure (Synapse), Spark
Data Engineering & ETL:	PySpark, SSIS, AWS Glue, Azure Data Factory, Airflow
Data Modelling & Orchestration:	Star/Snowflake Schema, Normalization, DAGs, Workflow Automation
Business Intelligence & Visualization:	Power BI, Tableau
Statistical & ML Methods:	Regression, Classification, Clustering, NLP, Statistical Analysis
Data Analysis & ML Libraries:	Pandas, NumPy, Scikit-learn, NLTK, Matplotlib, Plotly, Leaflet
Other Tools:	Git, Jupyter, VS Code, Postman, Hue, Docker, JIRA

WORK EXPERIENCE

Deutsche Bank | Senior Data Analyst

Nov 2020 – Present

Project: Employee Trading Compliance Applications (SQL, Python, Power BI, Snowflake)

- Oversaw releases, capacity planning, and infrastructure updates for two enterprise-grade compliance applications as the Client Service Manager (CSM-1), ensuring high availability and operational continuity.
- Served as the key liaison for Compliance Officers and executive stakeholders, delivering BI dashboards in Power BI and producing ad-hoc data reports to support regulatory investigations.
- Designed and implemented SQL-based data validation scripts to monitor 131 feeds from 52 brokers, reducing incomplete feed risk by **85%**. Integrated the process with ServiceNow APIs to enable real-time incident generation, accelerating resolution time from **5 days to 2 hours**.
- Developed an automated reporting framework using shell scripting and Oracle stored procedures, eliminating **70–80%** of repetitive compliance reports and reducing weekly L2 support requests from 150+ to ~40. Standardized output formats to improve downstream data handling.
- Created **13 KPI** metrics in collaboration with compliance stakeholders to measure L2 team performance. Trained a globally distributed team of 16 analysts and scaled the dashboard across 18+ applications, reducing average incident resolution time from **5 days to under 24 hours**.
- Engineered an Ansible-based access control automation pipeline that reduced privileged access incidents by **92%** over two years. Logged all production activity to ensure full compliance and audit traceability.
- Designed and implemented a Snowflake-based data warehouse to centralize compliance datasets, currently in early phase. Built ETL pipelines and developed **20+** near real-time KPIs in Power BI to accelerate reporting and enhance visibility into employee compliance activities.

People Tech Group | Junior Data Scientist

Sep 2019 – Oct 2020

Product: Elliptica Data Platform (Python, Kafka, SQL, Tableau, MongoDB, MS-SQL)

- Engineered fault-tolerant data connectors for relational and non-relational databases to capture real-time changes
- Tracked and streamed the data changes from database to the Tableau dashboard through Redshift DW.
- Performed stress testing on data pipeline for multiple use cases
- Coordinated with cross-functional remote teams for system design and sprint planning

SANKET CHAVAN

Austin, TX
+1(716)578-9525

chavansanket193@gmail.com
www.linkedin.com/in/sanketchavan08/

Project: Candidate Screening Chatbot (Rasa, Python, NLP)

- Led backend development efforts for an advanced AI chatbot designed to conduct preliminary interview rounds, utilizing the Rasa open-source ML framework, NLP, integrated with a PostgreSQL database.
- In a rapid, two-month timeframe, successfully developed a functional prototype capable of understanding user messages, identifying intent and entities, and delivering context-appropriate responses

University at Buffalo | Data Science Research Assistant (R, Python, Tableau) Dec 2018 – Sep 2019

- Spatial-Temporal Analysis on a Data collected using survey which is funded by "National Institutes of Health".
- Build ETL Pipeline to cleaned and ingest transformed data in MS-SQL using SSIS.
- Constructed Interactive Maps for Visualizing trajectories of users and clustered those **GPS** points to locate the most visited places in R
- Analyzed and visualized social influence on travel behavior of a single user and single household in R
- Built the model to predict the next Activity of a user on a specific day & time using **Markov Model** and **Conditional Probabilistic Model**.
- Designed a program to find Co-location between users using **python** and Visualize that Social Network in **Tableau Dashboard**

Niagara Falls Bridge Commission | Data Analyst Intern (MS excel, R) Sep 2017 – Dec 2017

- Performed EDA on Bridge Crossing Data to discover Patterns and Trends in traffic
- Conducted the Survey to identify the characteristics and travellers' behaviour on change in toll cost
- Achieved 13.11% traffic reduction during peak hours using the Multinomial Logistic Regression model

CERTIFICATIONS

- Microsoft Certified: Power BI Data Analyst Associate
- Certified Tableau Desktop Specialist
- AWS Certified Machine Learning – Specialty – Amazon

EDUCATION

University at Buffalo, State University of New York Aug 2017 – Feb 2019

Master of Science in Industrial Engineering - **GPA: 3.73/4.0**

Relevant Coursework: Machine Learning, Transportation Analytics, Data Mining, Data Intensive Computing, DOE

North Maharashtra University, India Aug 2012 – May 2016

Bachelor of Engineering in Mechanical Engineering - **GPA: 8.10/10.0**

RELEVANT PROJECTS ([GitHub](#))

NYC Taxi Data Analytics with Azure Synapse (Python, SQL) Jan 2024

- Analyzed 3 years of NYC Yellow Taxi data (2021–2023) to uncover rider behavior trends and inform fleet operations for high-traffic urban zones (100+ million records)
- Designed and developed scalable ETL pipelines using Serverless SQL and Spark Pools in Azure Synapse to ingest and process over 1TB of trip data
- Integrated Cosmos DB with Synapse Link for real-time analytics and built Power BI dashboards to deliver on-demand insights to stakeholders
- Revealed that 60% of rides used credit cards and Queens had the highest share of cash trips, enabling borough-level optimization of payment infrastructure and resource allocation.