# SANKET CHAVAN

Jacksonville, FL (Open to Relocation) chavansanket193@gmail.com
+1 716-578-9525 **US Visa:** H1B

## 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, Glue), Azure (Synapse, Data Factory), 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, Dimensionality Reduction, 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:

### Senior Data Analyst – Deutsche Bank — Nov 20 – Present

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

- Managing two major compliance applications as **Client Service Manager (CSM-1)**, overseeing releases, capacity planning, and infrastructure updates to ensure seamless operations.
- Acting as the primary contact for Compliance Officers and senior stakeholders, **delivering BI solutions** and generating ad-hoc compliance data reports.
- Implemented **data integrity controls** on **131 feeds from 52 brokers**, reducing regulatory risk from incomplete or missing feeds by 85%. Enabled **100% near real-time visibility** for compliance officers, ensuring timely detection of data issues. Automated incident creation via ServiceNow, cutting investigation timelines from **4-5 days to 1-2 hours**.
- Deployed a shell script + Oracle-based solution, **automating 70–80%** of recurring compliance reports and dropping L2 weekly requests from 150+ to around 40. Standardized reporting formats and speed up delivery from days to minutes.
- Defined **13 KPIs** for tracking L2 team performance, cutting average incident resolution **from 5 days to ~24 hours**. Led and trained a 16-member global team, raising user satisfaction scores and scaling the solution to **18+ compliance applications**.
- Drove a **92% reduction** in privileged access over 2 years by implementing Ansible-based automation. Achieved consistently low risk scores and full audit traceability, with every production action now fully logged and compliant.
- Built the initial data model and pipeline in Azure Synapse Analytics to centralize compliance data for advanced analytics. Enhanced KPI reporting and facilitated future integrations with BI tools like Power BI

### Junior Data Scientist – People Tech Group — Sep 19 – Oct 20

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

- Built 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

*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 built a functional prototype capable of understanding user messages, identifying intent and entities, and delivering context-appropriate responses

### Data Science Research Assistant – University at Buffalo (R, Python, Tableau, SSIS) — Dec 18 – Sep 19

- 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.

# SANKET CHAVAN

Jacksonville, FL (Open to Relocation) chavansanket193@gmail.com
+1 716-578-9525 **US Visa:** H1B


- 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**

**Data Analyst Intern - Niagara Falls Bridge Commission (MS excel, R)** **Sep 17 – Dec 17**

- 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 17 – Feb 19**

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 12 – May 16**

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

## RELEVANT PROJECTS:

**NYC Taxi Data Analytics with Azure Synapse (Python, SQL)** **Jan 24**

- Developed and optimized SQL scripts and Spark notebooks to process three years of NYC Yellow Taxi data (2021-2023) in Azure Synapse Analytics
- Configured dedicated SQL pools and Spark pools to manage and analyze over 100 million records efficiently.
- Ingested, transformed, and loaded over 1TB of data using Serverless SQL Pool, Spark Pool, and automated pipelines in Synapse.
- Enabled real-time analytics by integrating Synapse Link with Cosmos DB and visualized data through Power BI.
- Analyzed payment behaviors, identifying that 60% of trips were paid by credit card, with Queens showing a unique pattern of higher cash transactions.
- Developed Power BI dashboards reveal that Manhattan accounted for 40% of overall taxi demand, with demand peaking on Fridays and lowest on Sundays across all boroughs.

**Movie Recommender System (Python)** **Jan 19**

- Developed Item and User-Based Collaborative Models with Cosine, Pearson, and MSD similarity metrics, tuning and evaluating by measuring Hit Rates for recommended movies
- Trained KNN and SVD++ models to predict user ratings for unseen movies in the test set
- Achieved 75% accuracy in user preference prediction using a Restricted Boltzmann Machine
- Enhanced recommendations with an Autoencoder model, maintaining an avg difference of 1-star rating between predicted and true ratings

**Movie Revenue Prediction (Python)** **Oct 19**

- Eliminated Anomalies using EDA, created 39 new features, imputed missing values using Prediction from BO XGBoost model
- Built Linear Regression, Bayesian Optimized Random Forest, XGBoost, LightGBM models to predict the Revenue
- Achieved an RMSLE score of 0.96