

# VIVEK BASAVANTH HANAGOJI

Boston, MA (Open to Relocate) | +1 (617)-368 0288 | [hanagojivivek@gmail.com](mailto:hanagojivivek@gmail.com) | [LinkedIn](#) | [GitHub](#)

## SUMMARY

Experienced Data Analyst with a solid 3 years background in IT. Proven expertise in Python for data ingestion, transformation, and analysis. Eager to apply analytical skills and technical knowledge in a Data Analytics and Engineering role starting May 2024.

## EDUCATION

### Northeastern University

*Master of Science in Information Systems*

**Expected May 2024**

*Boston, MA*

### Savitribai Phule Pune University

*Bachelors in Engineering in Electronics & Tele-Communication*

**Jul 2015 – Jun 2019**

*Pune, India*

## PROFESSIONAL EXPERIENCE

### Data Analyst

**Jul 2020 – Jun 2022**

*Vodafone Intelligent Solutions*

*Pune, India*

- Engineered and deployed custom ETL solutions for Salesforce, efficiently transferring over 5 TB of critical customer data into Salesforce, with increased data processing speed by 40% and enhancing data quality for CRM activities.
- Developed Business Intelligence solutions by building scalable ETL data pipelines on Snowflake, AWS S3, and Salesforce databases to automate month-end reports & provide ad-hoc data solutions reducing turnaround time by 45%.
- Administered Salesforce Tableau CRM, which entailed extraction of data from Salesforce, development of datasets with dataflows and dataflow builder, and execution, scheduling, and tracking of dataflows.
- Demonstrated success in leading and participating in 30+ Agile sprints, ensuring timely delivery of project milestones and fostering a collaborative team environment for productivity, attaining a 100% on-time delivery rate.
- Utilized Workbench and Salesforce Object Query Language (SOQL) for advanced data patching, correcting over 5,000 records with discrepancies during year-end activities, enhancing data reliability by 45%.

### Salesforce Administrator

**Jul 2019 – Jun 2020**

*Vodafone Intelligent Solutions*

*Pune, India*

- Streamlined data entry processes by 20% by implementing data validation rules and custom fields in Salesforce.
- Utilized tools like Process Builder, Email Templates, Workflow Rules, Workflow Actions, and Approval Flows to improve system functionality, ensuring overall Salesforce efficiency.
- Automated complex data patching processes using Apex Data Loader, processing over 10,000 records monthly, resulting in a 50% reduction in manual data entry errors and a 30% increase in process efficiency.

## TECHNICAL PROJECTS

### GenAI Chatbot | Python, FastAPI, AWS, Docker, Airflow, OpenAI, Pinecone DB

**Nov 2023 – Dec 2023**

- Crafted an OpenAI-driven chatbot for SEC documents, enabling precise query resolution for over 75 forms with 95% accuracy in similarity searches.
- Engineered 2 Airflow ETL pipelines for SEC document processing, boosting efficiency by 40% and enhancing data storage.
- Containerized application deployment using Docker on a GCP VM Instance.

### Text to SQL with Snowpark | Snowflake, Python, Snowpark, OpenAI, LangChain, Streamlit

**Oct 2023 – Nov 2023**

- Streamlined Snowflake data views by 15% with Python UDFs and EDA, boosting efficiency through Snowpark.
- Built a Streamlit application with LangChain for text-to-SQL in Snowflake, improving query efficiency by 20%.
- Designed Streamlit application with Pyplot for Snowflake data analysis, boosting visualization and analytical insights by 25%.

### New York City Motor Vehicle (MV) Collision | ER/Studio, Talend, SQL Server, Power BI

**Aug 2023 – Sept 2023**

- Migrated 10 million data records from BigQuery to MySQL Server using Talend, optimizing data migration procedures.
- Engineered ETL workflows in Talend for NYC Motor Collision Dataset, orchestrating data merging and cleansing processes.
- Leveraged Tableau and PowerBI to analyse trends, patterns, and insights, facilitating year-over-year (YOY) performance tracking for 30+ key performance indicators (KPIs).

## TECHNICAL SKILLS

### Programming Languages

: Python, R, SQL, PySpark

### Orchestration Tools

: Apache Airflow, Talend, dbt

### Data Science Libraries

: Pandas, Sci-kit learn, NumPy, TensorFlow, Matplotlib, Seaborn

### Databases

: MySQL, SQL Server, PostgreSQL, Pinecone Vector DB

### Data Cloud Services

: BigQuery, Amazon Redshift, Amazon S3, Snowflake, Amazon EC2, Google Cloud Storage

### Visualization Tools

: PowerBI, Tableau, Looker