

# SOURAB SAKLECHA

San Jose, CA | +1 408-581-4470 | saklechasourab@gmail.com | LinkedIn: [/sourabsaklecha](#) | GitHub : [/SSaklecha](#) | [Data Portfolio](#)

## EDUCATION

- **M.S., Data Analytics**, San Jose State University, San Jose, CA, USA | *GPA: 3.77* **Graduating in Dec 2024**
- **B.E., Computer Science**, Visvesvaraya Technological University, India | *GPA: 3.51* **Aug 2016-Sep 2020**

## TECHNICAL SKILLS

**Coding Languages:** SQL, Python, R, DAX

**Databases/Datawarehouse:** MySQL, MongoDB, AWS Redshift, BigQuery, Snowflake, SAS

**Data Visualization:** PowerBI, Tableau, Microsoft Excel, Python (Matplotlib, Seaborn)

**Cloud Platforms:** Amazon Web Services (AWS), Google Cloud Platform (GCP), Azure, Microsoft Fabric

**ETL and Other Tools:** AWS S3, AWS Glue, Azure Data Factory, Power Query, SSIS, PySpark, Apache Spark, Apache Hadoop, Git, JIRA

**General Expertise:** ETL, Data Modelling, Data Querying, Data Cleaning, Statistics, Statistical Modelling, ML, NLP, Agile, A/B Testing.

**Soft Skills:** Analytical Problem Solving, Driving Business Impact, Problem Solving Skills, Cross-Functional Team Collaboration.

## WORK EXPERIENCE

**Teaching Assistantships (TA) – Machine Learning, Mathematics | San Jose State University - California, USA** **Jun 2023-Nov 2024**

- Delivered weekly instruction to 80+ students with practical demos, boosting class enrollments by 17% in the second semester.
- Graded assignments 30% ahead of schedule. Established an on-demand office-hour model for greater student convenience.

**Data Engineer and Data Analyst Consultant | Core Molding Technologies Inc - Ohio, USA (Remote)** **Jul 2022-Jan 2023**

- Built an end-to-end data warehouse infrastructure with SSIS, enhancing data integration performance by 30%.
- Deployed 3 ad-hoc PowerBI dashboards using data from the established warehouse, driving a 30% increase in decision-making speed.
- Transitioned 6 Tableau finance reports consumed by the upper management to PowerBI, boosting daily report consumption by 25%.

**Business Analyst and Power BI Developer | JoulestoWatts - Bangalore, India** **Apr 2022-Jul 2022**

- Overhauled reporting and analytics infrastructure to Power BI. Created 4 automated dashboards with 2 KPIs, delivered to stakeholders.
- Implemented role-based access control to enhance data security for 200+ employees, reducing unauthorized access incidents by 50%.
- Boosted report consumption by 12% QoQ within three months of joining with an optimized report distribution strategy.

**Systems Engineer-Azure Data Factory and SSIS | Infosys Limited - Bangalore, India** **Oct 2020-Apr 2022**

- Piloted ETL operations on parquet files using Azure Data Factory, ensuring higher data integrity by reducing transfer errors by 5%.
- Optimized data pipelines using Azure DevOps Services, boosting performance of overnight refreshes by 20%.
- Minimized downtime by 23% by eliminating refresh bottlenecks and enhanced data loading with shorter and wider pipeline.

**Business Intelligence Analyst Intern | Infosys Limited - Mysore, India** **Jan 2020-Mar 2020**

- Automated four everyday reporting tasks using Power Flow, saving the team 15 minutes daily.

## PROJECTS

**LexLLM: Personal AI Legal Assistant, Capstone Project | Vector Database, Python, Agile, Finetuned LLMs** **Jan 2024-Nov 2024**

- Led a five-person team to develop a RAG-based PDF chat assistant to help comprehend complex legal codes.
- Finetuned Llama-3 on a 52 GB text data corpus stored in a vector database to reach an accuracy of 68% in legal query answering.

**Subscription Churn Rate Prediction for OTT Platforms, Data Mining | SQL, Python, Tableau** **Jan 2024-May 2024**

- Achieved 97% accuracy in catching possible subscription churn with the Random Forest predictive modeling.
- Conducted data cleaning, transformation and feature engineering to enhance the performance of SVM models by 12%.

**Optimizing Ride-Hailing, Big Data Project | GCP, Dataflow, BigQuery, Colab, Looker Studio** **Aug 2023-Dec 2023**

- Created a data pipeline for 84.5 million trip records sourced from NYC taxis. Discovered trends to improve ride-hailing services.
- Leveraged GCP and BigQuery for scalable storage and fast query processing, resulting in a 25% reduction in operational costs.

**Cricket Database Management System, Sports Analytics | Database, SQL, ETL Pipeline, PowerBI, AWS** **Jan 2023-May 2023**

- Implemented an end-to-end data pipeline, automating data handling, pre-processing, and storage on MySQL and MongoDB databases.
- Examined performance of over 830 players on 6 KPIs and designed 2 dashboards on PowerBI to help improve real-time analysis.

## RESEARCH AND PUBLISHED WORK

**Mitigated Risk in P2P Lending Network: Enhanced Loan Default Predictions | Machine Learning, OpenAI Research** **Aug 2023-Dec 2023**

- Leveraged SVM, Random Forest and 2 other Machine Learning models to predict loan defaults in a peer-to-peer lending platform.
- Compared the impact of SMOTE and Synthetic data on the accuracy of predictions. Benchmarked 99.4% accuracy and 97.4% F1-score.  
[Published in Journal of Networking and Network Applications, Jun 2024](#)

## LICENSES, CERTIFICATIONS AND AWARDS

- Microsoft Certified: Power BI Data Analyst ([PL-300](#)) and Azure Data Fundamentals ([AZ-900](#)) among [others](#).
- NPTEL Certified: Programming, Data Structures and Algorithms using Python (*From IIT, Madras*).