

# SAKSHI MENGHANI

## Data Analyst

+1 (667) 369-8506 | [sakshi.menghani98@gmail.com](mailto:sakshi.menghani98@gmail.com) | [www.linkedin.com/in/sakshi-menghani](https://www.linkedin.com/in/sakshi-menghani) | [github.com/SakshiMenghani](https://github.com/SakshiMenghani)

### SUMMARY

- **3+ years** experienced **Data Analyst** and proficiently advised and coached in **programming languages** such as **Python, SQL** and possibly others like **MATLAB, R, Java, C** and **C++** for data manipulation, analysis and development of data pipelines.
- Skilled in using essential data analysis libraries such as **NumPy, Pandas, Seaborn, Matplotlib** and potentially others like **SciPy, Scikit-Learn** and **TensorFlow** for advanced data processing and analysis, effectively demonstrating insights.
- Demonstrated expertise in databases like **MySQL, PostgreSQL, Oracle** and **MongoDB** for data storage, retrieval and management, efficiently organizing data.
- Competently directed and managed data visualization tools such as **Tableau, Power BI** and **Excel** for effective communication of insights, stimulating collaboration.
- Familiarity with cloud platforms such as **AWS, Azure** and **Google Cloud Platform** for scalable and reliable data solutions, facilitating efficient deployment.
- Proficiently coordinated and developed tools like **Git/GitHub, Docker, Jupyter Notebook** and potentially others like **Spark** and **API** for efficient development, collaboration and deployment, accelerating project timelines.
- Adept at creating comprehensive **documentation** for data analysis processes and tools to ensure clarity, reproducibility, and ease of understanding for stakeholders and team members. Experience with **SAP** for integrating and managing enterprise data, enhancing data-driven decision-making across organizational processes.

### EDUCATION

**University of Maryland**, Baltimore MD - USA

Masters of Science in Information Systems with Specialization in Data Science and AI

**Aug 2022 – May 2024**

**Bhilai Institute of Technology**

Bachelors of Technology - Computer Science and Engineering

**Aug 2016 – Sep 2020**

### TECHNICAL SKILLS

<b>Languages:</b>	Python, SQL, Matlab , R, Java, C, C++
<b>Libraries:</b>	NumPy, Pandas, Seaborn, Matplotlib, SciPy, Scikit-Learn, TensorFlow
<b>Databases:</b>	MySQL, PostgreSQL, Oracle, MongoDB
<b>Visualization Tools:</b>	Tableau, Power BI, Excel (VLOOKUP, Macros)
<b>Data Analysis skills:</b>	Data cleaning, Data Modelling, Data warehousing, ETL, Hypothesis Testing, Regression Analysis, A/B Testing, Clustering, Forecasting
<b>Machine Learning:</b>	Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), Principal Component Analysis
<b>Methodologies:</b>	Agile/SCRUM, SDLC
<b>Cloud Platform:</b>	AWS (EC2, S3, Lambda, AWS Glue, Redshift), Azure (Blob Storage, Azure Synapse Analytics, Azure Data Factory, Azure Databricks, Azure Functions), Google Cloud Platform (Google Cloud Storage, BigQuery, Cloud Dataflow, Compute Engine)
<b>Tools:</b>	Git/GitHub, Docker, Kubernetes, MS Office, Airflow, Jupyter Notebook, Locker, Pyspark, Weka, Postman, VS Studio, Spark, API, Web Services, Microsoft Power Automate, SAP, Salesforce, ServiceNow

### EXPERIENCE

**Data Management Analyst | CRISP Shared Services, USA**

**Jun 2023 – Current**

- Automated data ingestion and transformation processes using Python and Airflow, orchestrating data workflows and scheduling dataprocessing jobs for timely delivery of insights to stakeholders.
- Worked on Databricks SQL (SQL Analytics) by leveraging proprietary methods to retrieve over 500 protected health information (PHI) from numerous databases and prepare massive data from public, private, and other hospitals in Washington, DC, Maryland and Virginia.
- Performed data integration and synchronization on IBM InfoSphere to efficiently manage false positives/negatives, overlays, and duplications by achieving 98% accuracy and maintaining the data quality of over 1500 patient records.
- Initiated data visualization tasks, using Agile approach to develop engaging dashboards and graphs, resulting in a 20% increase in dataunderstanding and a 15% boost in decision-making efficiency, providing significant insights for the Insights team.
- Reviewed over 300 reports from diverse clinical datasets by enhancing data precision, achieving a 30% improvement in data quality.
- Collaborated closely with implementation managers for Alaska & DC, prioritizing ENS panels and ensuring data integrity.
- Conducted performance optimization exercises on Power BI workbooks, identifying and resolving bottlenecks in data extraction, processing, and visualization to enhance dashboard responsiveness and user experience.

**System Data Engineer | Infosys Pvt. Ltd**

**Jun 2021 – Aug 2022**

- Analyzed Google Fitbit user data for over 200 users on a weekly basis, producing insights that resulted in a 22% increase in user retentionand actively contributing to app testing while improving iOS and Android platforms.
- Build a cutting edge real-time monitoring utility system that efficiently tracks user feedback, leading to the identification and mitigation

of duplicate bugs by 35% in the system.

- Developed automated data validation scripts and processes using Python and SQL, reducing manual data cleaning efforts by 50% and ensuring data integrity across multiple systems.
- Actively contributed to the development and deployment of an embedded data system, enabling real-time test data capture on smartwatches and smartphones, which led to a 20% reduction in data acquisition latency and significantly enhanced data analysis capabilities.
- Architected and implemented scalable ETL pipelines using Apache Spark and AWS Glue, processing terabytes of data daily and reducing data processing time by 30%.
- Defined and enforced data governance policies and standards across ETL processes, ensuring compliance with regulatory requirements and best practices for data security, privacy, and confidentiality.
- Collaborated with cross-functional teams to gather requirements and translate business needs into interactive reports and dashboards in Tableau ensuring alignment with organizational objectives and KPIs.

#### **Data Analyst | Think and Learn Pvt Ltd**

**Nov 2020 – Apr 2021**

- Spearheaded data analysis efforts to identify client needs and collaborated cross-functionally to develop data-driven marketing and sales.
- Achieved a remarkable 15% increase in conversion rates, directly attributable to data insights-driven strategies. This translated to a substantial boost in revenue of approximately \$2.5 million over the course of the project, highlighting the tangible impact of data-driven decision-making on overall business success.
- Designed and implemented complex SQL queries, stored procedures, and views to extract, transform, and analyze large-scale datasets, optimizing query performance and reducing processing time by 40%.
- Developed and implemented advanced statistical models and machine learning algorithms, including logistic regression, decision trees, and random forests, to forecast trends and identify patterns in large-scale datasets.
- Conducted performance tuning and optimization exercises on ETL workflows, utilizing techniques such as parallel processing, partitioning, and indexing to enhance data processing speed and efficiency.
- Collaborated with cross-functional teams to define KPIs and metrics, and developed interactive dashboards and visualizations using Tableau and Power BI to track performance and monitor key business metrics.
- Leveraged advanced Excel functionalities such as VLOOKUP, INDEX-MATCH, and Macros to automate data processing tasks, streamline reporting workflows, and improve overall productivity.

#### **Data Analyst Intern | Philips Healthcare**

**Dec 2019 – Sep 2020**

- Developed a predictive model using Python to analyze patient data and forecast potential health issues, improving early intervention measures.
- Implemented data cleaning and preprocessing techniques in SQL and Pandas to ensure the accuracy and reliability of the dataset.
- Created interactive visualizations in Tableau to present key insights and trends to stakeholders, enhancing data-driven decision-making.
- Conducted hypothesis testing and regression analysis to identify critical factors affecting patient health outcomes.
- Collaborated with cross-functional teams using Agile/SCRUM methodologies to integrate the analytics model into the existing patient monitoring system.

### **PROJECTS**

---

#### **Covid - 19 Data Analysis (Python, JavaScript, ReactJS, Tableau) - [Covid 19 Data Analysis](#)**

- Engineered and designed a comprehensive COVID-19 data analysis system utilizing Tableau, Python, and React JS, spotlighting recovery rates, country-wide statistics, and the profound impact on populous states, to analyze a dataset comprising over 12,000 data points.
- Orchestrated strategic data collection efforts in high-impact regions of Asia, covering 75% of the affected populace, resulting in optimized resource allocation and contributing to a 25% improvement in pandemic response efficiency.

#### **Credit card fraud detection (SQL, Python, Jupyter notebook) - [Credit card fraud detection](#)**

- Applied expertise in model development to seamlessly integrate algorithms (like logistic regression) into our credit card fraud detection system using Numpy, Pandas, and Matplotlib for performance analysis, achieving a 25% efficiency boost and accuracy.
- Developed a predictive model leveraging ANN, SVM, and K-Nearest Neighbors (KNN) algorithms to classify transactions as genuine (0) or fraudulent (1), significantly enhancing fraud detection capability, which accurately predicted 492 frauds out of 284,807 credit cards, achieving an accuracy of 94% and precision of 99%.

#### **Bitcoin Price Prediction using XGBoost**

- Designed a predictive analytics tool to forecast Bitcoin prices using XGBoost, MapReduce, a powerful gradient boosting algorithm, while also performing EDA and feature engineering and achieving an 80% feature selection rate.
- Evaluated the model using ROC-AUC curve, which helped in measuring the accuracy of the predictions. Plotted the confusion matrix in the final steps, which helped in observing the predictions.