

# Susmitha Arikatla

+1(281)-750-6048|[susmitha.usk7@gmail.com](mailto:susmitha.usk7@gmail.com) | [GitHub](#) | [LinkedIn](#) | Houston, Texas

---

## SUMMARY

Data Analyst with 2 years of experience specializing in data analysis, visualization, and reporting. Proficient in SQL, Power BI, and Excel, with a strong background in data modeling, statistical analysis, and machine learning. Adept at using various programming languages and databases to derive actionable insights and optimize business processes.

## PROFESSIONAL EXPERIENCE

---

### Data Visualization Teaching Assistant | University of Houston | Houston, Texas

Aug 2022 – May 2023

- Demonstrated the implementation of data science techniques and machine learning models using Excel, SAS, R, and Tableau to enhance students' understanding.
- Generated visualizations for professors, resulting in a 17% increase in student comprehension within 4 months.

### Data Analyst | Freelance | Remote|India

June 2020 – Apr 2022

- Designed custom dashboards daily using Power BI, providing valuable insights into customer behavior, and saving the company \$34,000 annually.
- Collaborated with cross-functional teams to analyze key metrics and KPIs, identifying opportunities for improvement in user acquisition.
- Automated several Excel data tools (pivot tables, VBA, and Excel Formulas) and reports into Tableau dashboards, saving 117 hours per month.

## ACADEMIC PROJECTS

---

### Construction Safety Analysis using OSHA Dataset

[Kaggle Link](#)

**Tech Stack:** Numpy, Google Collab, PowerBI, Excel

- Applied web scraping to collect Construction Accidents records from OSHA Website.
- Preprocessed data using advanced NLP techniques, such as stemming and lemmatization, to enhance data quality.
- Implemented Principal Component Analysis (PCA) for dataset dimensionality reduction and utilized TF-IDF Vectorization for feature representation.
- Employed K-means Clustering technique to identify the optimal number of clusters (K) and centroids.

### Seoul Bike Sharing Demand

[Kaggle Link](#)

**Tech Stack:** Regression, Google Collab, MySQL, Python, Tableau

- Collected and aggregated data from various sources to create a comprehensive dataset for analysis.
- Cleaned, normalized, and engineered features to ensure data integrity and suitability for modeling.
- Developed predictive models, including regression and machine learning algorithms, achieving 80% accuracy in forecasting bike sharing demand in Seoul.

### Comparative study of supervised learning algorithms for Intrusion Detection

[Kaggle Link](#)

**Tech Stack:** Numpy, Algorithms, Python, Data Analysis

- Gathered external data and applied techniques like PCA, Label encoding, and Normalization for preprocessing.
- Trained and evaluated Decision Tree, Random Forest, Support Vector Machines, Boosting, Bootstrap Aggregation, Naive Bayes, and Advanced Neural Networks algorithms.
- Implemented the best algorithm based on performance metrics like Accuracy Score, Execution Time, and F1 score.

## EDUCATION

---

### Master of Science in Data Science | University of Houston | Houston, Texas | GPA: 3.6/4.0

May 2023

- **Relevant Coursework:** Machine Learning, Statistics, Data Analysis, Database Management, Big Data, Time Series Forecasting, Artificial Intelligence, Data Science for Security, Data Visualization, Time Series Analysis

## SKILLS

- 
- **Programming Languages:** Python, R, SQL
  - **Databases:** Oracle, MongoDB, AWS S3, Microsoft Office MYSQL, Snowflake
  - **DW Skills:** Apache Spark, Apache Kafka, Data Pipelines, Hadoop
  - **Python libraries:** Pandas, Numpy, TensorFlow, Keras, Seaborn, Matplotlib, Scikit-learn, PyTorch, NLTK
  - **Data Modeling:** Logistic Regressions, Decision Trees, Multivariate Regression, Neural Networks, Random Forests
  - **Tools:** Data Mining, Statistical Analyses, ETL, Hypothesis Testing, Qlik Sense, Alteryx, DAX, Informatica, PowerBI, Tableau, SSRS, Agile, SAP, Financial Analyses