

# Susmitha Arikatla

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

## WORK EXPERIENCE

### Data Visualization Teaching Assistant | University of Houston

Aug 2022 – May 2023

- Showcased the utilization of various software tools including Excel, SAS, R, Tableau to demonstrate the implementation of data science techniques and machine learning models.
- Generated visualizations for professors to use in class, increasing student comprehension by 17% in 4 months.

### Data Analyst | Freelance | Remote India

June 2020 – Apr 2022

- Implemented data visualization techniques for creating Dashboards and reports using Python and Tableau to present complex data insights to stakeholders, resulting in improved decision-making and a 20% reduction in time spent on data analysis.
- Collaborated with cross-functional teams to develop and implement machine learning algorithms for predictive maintenance and collect and report key metrics.
- Utilized advanced statistical techniques and predictive modeling to analyze a large dataset of customer behavior, resulting in a 15% increase in customer retention.

## PROJECT EXPERIENCE

### Construction Safety Analysis using OSHA Dataset

**Tech Stack:** Numpy, Google Collab, Power BI, Excel

- Performed Data Extracted using web-scraping from OSHA Website and preprocessed using advanced NLP techniques, including stemming and lemmatization, to enhance data quality and trends.
- Applied Principal Component Analysis (PCA) to reduce dataset dimensionality and Conducted TF-IDF Vectorization for feature representation.
- Employed the K-means Clustering technique to identify the optimal number of clusters (K), centroids.

### Visualization of Data Scientist Job Salaries

**Tech Stack:** Pandas, Tableau, Glyphs, Power BI

- Performed data cleaning and filtering techniques in Excel to prepare dataset for analysis.
- Conducted Exploratory Data Analysis (EDA) using Python Pandas to identify patterns in the dataset.
- Utilized Tableau Desktop to create visually impactful representations of Data Science job salaries based on experience levels and job titles for 2023.
- Designed and implemented unique visual representations using Glyphs to enhance the understanding of various experience levels in the data.

### Seoul Bike Sharing Demand

**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 in the dataset to ensure data integrity and suitability for modeling.
- Developed predictive models, including regression and machine learning algorithms, to accurately forecast the demand for bikes at different stations. Achieved 80% accuracy rate, significantly improving bike sharing demand forecasting in Seoul.

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

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

- Gathered Data from external sources and processed it using techniques PCA, Label encoding, and Normalization.
- Trained and evaluated Decision Tree, Random Forest, Support Vector Machines, Boosting, Bootstrap Aggregation, Naive Bayes, and Advanced Neural Networks algorithms.
- Assessed the best algorithm based on performance metrics like Accuracy Score, Execution Time, F1 score.

## EDUCATION

**Master of Science in Data Science** | University of Houston | Houston, Texas | GPA: 3.6

May 2023

- **Relevant Coursework:** Machine Learning, Statistics, Data Analytics, Database Management, Big Data, Time Series Forecasting, Artificial Intelligence, Data Visualization, Time Series Analysis
- **Certifications:** Google Data Analytics Professional Certificate, Tableau Desktop Specialist (Tableau Software, LLC), Machine Learning-HarvardX, Python & Machine Learning for Financial Analysis

## SKILLS

- **Programming Languages:** Python, R, SQL, SAS
- **Databases:** Oracle, MongoDB, AWS S3, MYSQL, Snowflake
- **Data warehouse:** Apache Spark, Apache Kafka, Data Pipelines, Hadoop
- **Python libraries:** Pandas, Numpy, TensorFlow, Keras, Seaborn, Matplotlib, Scikit-learn, PyTorch.
- **Data Modeling:** Logistic Regressions, Decision Trees, Regression, Neural Networks, Random Forests
- **Tools:** Looker, ETL, Hypothesis Testing, Qlik Sense, DAX, Informatica, Power BI, Tableau, Microsoft Excel, PowerPoint