SINDHUJA REDDY
SENIOR DATA ANALYST
EMAIL: sindhujay1005@gmail.com **CONTACT NO: +1(940) 242-3869**
LINKEDIN: linkedin.com/in/sindhuja-reddie-

PROFESSIONAL SUMMARY:

- Overall **9+** years of experience as a **Data Analyst**, specializing in data transformation, statistical modeling, and data pipeline optimization to deliver actionable business insights.
- Proficient in integrating **Python**-based applications with external data sources, APIs, and web scraping tools for real-time data acquisition and analysis.
- Extensive experience in ETL workflow design and automation using **Azure Data Factory, Azure Synapse Analytics, and Azure Stream Analytics** for efficient data extraction, transformation, and loading (ETL).
- Advanced expertise in statistical modeling using **R** including Linear Regression, Time Series Forecasting, Clustering to drive predictive analytics and business insights.
- Skilled in designing and optimizing **SQL queries** for real-time reporting and dashboards, supporting data-driven decision-making across organizations.
- Proficient in **Python** such as Pandas, NumPy, Matplotlib for data analysis, automation, and pipeline optimization across platforms.
- Extensive experience with **SAS** for data manipulation, statistical analysis, and reporting, including **SAS Base**, **SAS Macro**, and **SAS SQL** procedures.
- Expert in **SQL, MySQL, and PostgreSQL** for data cleansing, transformation, aggregation, and optimization **t**o support business intelligence.
- Strong proficiency in **AWS** services, including **AWS Glue**, **Lambda**, **Redshift**, **S3**, and **AWS Step Function**s, to design and automate scalable data extraction, transformation, and loading workflows.
- Expertise in **big data technologies** including Hadoop, HDFS, Hive, Pig for large-scale data ingestion, processing, and analytics.
- Skilled in leveraging cloud-based tools such as **Jupyter Notebooks** for collaborative data analysis and machine learning model development.
- Highly skilled in managing and optimizing **PostgreSQL databases** for performance, scalability, and business insights.
- Proficient in building interactive dashboards using **Tableau, Power BI, and Google Analytics**, enabling effective visualization of KPIs and business trends.
- **Experienced in advanced data modeling and optimization** for **Power BI and Oracle ETL workflows**, ensuring efficient data relationships and query performance.
- Advanced proficiency in **MS Excel** such as Pivot Tables, Power Query, Data Analysis techniques for business intelligence and reporting.

TECHNICAL SKILLS:

Data Analysis & Modeling	Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-learn, IBM SPSS, , Tableau, Shiny, Synapse, Data Factory, SQL, MySQL, PostgreSQL, Oracle
Clouds	AWS, Azure
Big Data Technologies	Hadoop, HDFS, Hive, Pig, Map Reduce, Flume, Spark, Athena, Blob Storage
Database Management	MySQL, PostgreSQL, Oracle, SQL, SAS, Redshift, Azure SQL Database
Machine Learning/AI	Random Forest, Decision Trees, Predictive Models (R), Clustering, Time Series Forecasting, Statistical Modeling, Regression Analysis
Data Visualization	Tableau, Power BI, Google Analytics, Excel Quick Sight, SAS Visual Analytics
Data Transformation	Data Cleansing, Data Manipulation, Data Pre-processing, Data Profiling
Data Security	Database Security (SQL, MySQL), Role-based Security (Tableau), Security Audits (SQL), Privacy Regulations
Version Control	Git, GitHub
Automation & Scripting	AWS Step Functions, Python (Automation), SQL Queries, Jupyter Notebooks

PROFESSIONAL EXPERIENCE:

Client: Wintrust Financial Corporation, Rosemont, IL **Feb 2023 – Present**
Role: Sr. Data Analyst
Responsibilities:

- Integrated data from on-premises sources into **Azure SQL Database** and **Azure Data Lake** for further analysis and reporting using **SQL** and **Azure Data Explorer.**
- Integrated **Python** applications with external data sources, **APIs**, and **web scraping** tools to acquire real-time data for analysis.
- Applied advanced statistical modeling techniques in **R**, including linear regression, time series forecasting, and clustering, to drive business insights.
- **Designed and optimized SQL queries and database schemas** in **MySQL, PostgreSQL, and Oracle** for real-time reporting, data retrieval, and performance optimization.
- Conducted data integration tasks in **SAS**, merging data from disparate systems and transforming it into a unified, actionable dataset for analysis.
- Leveraged **Azure Blob Storage** for secure, scalable, and cost-effective storage of large datasets and integrated it with other **Azure services** for data processing.
- Created and optimized data pipelines using **Python** for **ETL** processes, ensuring seamless data movement across platforms.
- **Developed and optimized Tableau and Power BI dashboards**, improving data visualization, scalability, and business insights for stakeholders.
- **Leveraged big data technologies** (Hadoop, HDFS, Hive, Spark) to process and analyse large datasets, ensuring efficient data handling.
- Designed and implemented budgeting and forecasting tools in **Excel** to support financial planning activities.
- Conducted **Google Analytics** tracking and segmentation to analyse customer behaviour, website traffic patterns, and business metrics.
- Developed and executed complex data preprocessing pipelines using **Jupyter Notebook** to clean, transform, and validate datasets for actionable insights.
- Analyzed large datasets using **IBM SPSS** to identify trends, correlations, and patterns, delivering actionable business recommendations.
- **Performed comprehensive data quality checks** within **Hadoop and SQL databases**, ensuring data accuracy, consistency, and reliability.
- Enhanced data processing capabilities by leveraging **SAS Macro** programming, improving code reusability and reducing development time for new projects.
- Automated data workflows with **Azure Logic Apps**, streamlining data movement and operational tasks between on-premises and cloud systems.
- Utilized **Jupyter Notebook's** ability to combine code, visualizations, and narrative text to present data findings in a clear, understandable format for non-technical stakeholders.

Environment: Azure SQL Database, Data Lake, Data Explorer, Blob Storage, Logic Apps, Data Pipelines, SQL, Python, APIs, Web Scraping, R, Linear Regression, Time Series Forecasting, Clustering, MySQL, PostgreSQL, Tableau, Hadoop, Google Analytics, Excel, Jupyter Notebook, IBM SPSS, SAS, SAS Macro Programming, Oracle.

Client: McKesson, Irving, TX **Oct 2019 – Jan 2023**
Role: Sr. Data Analyst
Responsibilities:

- **Developed interactive data visualizations** using **Python (Matplotlib, Seaborn), R (Shiny), and Power BI**, enhancing business insights and stakeholder decision-making.
- **Utilized AWS Athena to run SQL queries on AWS S3 data**, streamlining data exploration without additional ETL processes.
- **Designed and optimized SQL queries and database schemas** in **MySQL, PostgreSQL, and Oracle** for efficient data retrieval, reporting, and performance tuning.
- Built and managed scalable ETL pipelines using **AWS Glue, Lambda, S3**, and **Hadoop-based tools (Flume, Spark)** to handle **real-time and batch data processing**.
- Automated recurring reporting and data transformation workflows using **SAS (Base, Macro), Python scripting, and Jupyter Notebooks**, reducing manual intervention.
- Migrated legacy ETL processes to Hadoop, enhancing data scalability, efficiency, and cost-effectiveness.

- Created executive-level dashboards and reports in **Power BI, Excel, Tableau, and SAS Visual Analytics**, providing real-time tracking of **KPIs and business trends**.
- Applied statistical modelling and **exploratory data analysis (EDA)** using **R and IBM SPSS**, uncovering **trends, correlations, and anomalies**.
- Optimized **Jupyter Notebooks** for performance, improving execution speed and handling of large datasets.
- Utilized **Python** is multi-threading and parallel computing capabilities to handle large-scale datasets efficiently.
- Implemented **PostgreSQL** data models and schema designs that ensured data integrity and optimal performance.
- Designed and optimized **Oracle** database schemas and tables to support high-performance data storage and retrieval.
- Established best practices for data modelling and query optimization in **Power BI and SQL**, improving data relationships and report performance.
- Performed advanced data quality checks within **SQL and Hadoop databases**, improving data accuracy and consistency.
- Conducted customer segmentation and cohort analysis in **Google Analytics**, generating real-time reports to identify user behaviour trends, conversion patterns, and support data-driven decision-making.

Environment: AWS Athena, Lambda, S3, Glue, Python, Matplotlib, Seaborn, R, Shiny, SQL, MySQL, Power BI, Hadoop, Flume, SAS, SAS Visual Analytics, Google Analytics, Excel, Jupyter Notebook, IBM SPSS, PostgreSQL, Oracle, Power BI, PostgreSQL, Oracle Database, Jupyter Notebooks.

Client: Bharti AXA Life Insurance, Mumbai **Apr 2017 – May 2019**
Role: Data Analyst
Responsibilities:

- Developed and maintained **Python**-based dashboards and reporting solutions to deliver real-time business intelligence.
- Developed predictive models using **machine-learning** algorithms such as **random forests, decision trees**, in **R** to optimize business operations.
- Applied **Hadoop's** parallel processing capabilities to analyze big data in real-time, enabling faster decision-making and operational improvements.
- Created **macros** in **SAS** to simplify repetitive coding tasks and improve programming efficiency.
- Leveraged **Excel's Power Query** and **Power Pivo**t tools for advanced data modeling and transformation.
- Automated data reporting processes by scripting reusable **Jupyter Notebook** templates for periodic data analyses and visual summaries.
- Conducted segmentation and clustering analysis in **IBM SPSS** to identify distinct customer profiles and behaviors.
- Deployed and monitored **Azure Stream Analytics** jobs for processing real-time streaming data from IoT devices and other sources.
- Developed batch-processing jobs using **Hadoop Map Reduce** and **HiveQL** to process large datasets and generate detailed reports for business stakeholders.
- Monitored **Tableau** usage metrics to optimize user engagement and implemented role-based security to control data visibility based on user permissions.
- Developed automated data pipelines using **Python**, integrating libraries such as **Pandas** and **NumPy** for data cleaning and transformation, significantly improving process efficiency.
- Ensured data accuracy by implementing data validation rules and constraints within **Oracle** databases.
- Managed **ETL pipelines** by leveraging **SQL** and **MySQL** for seamless integration and transformation of data from diverse sources into centralized databases.
- Leveraged **Azure Synapse Analytics** to manage and analyze large datasets, optimizing performance for real-time reporting and data-driven decision-making.

Environment: Azure Stream Analytics, Synapse Analytics Python, R, SQL, MySQL, Tableau, Hadoop, SAS, Excel, Power Query, Power Pivot, Jupyter Notebook, IBM SPSS, Hadoop Map Reduce, HiveQL, Pandas, NumPy, Oracle,.

Client: Asian Paints, Mumbai, India **Jul 2015 – Mar 2017**
Role: Data Analyst
Responsibilities:

- Developed scalable data pipelines using **AWS services** (**Redshift, S3, Step Functions, RDS**) to automate data extraction, transformation, and loading (ETL) processes.

- Integrated Python scripts with relational databases **SQL, MySQL, Oracle)** to optimize data retrieval and enhance analysis efficiency.
- Conducted security audits of **SQL** databases to ensure compliance with privacy regulations and organizational security standards.
- Applied **R** programming to clean, transform, and analyze complex datasets, providing actionable insights for business intelligence and decision-making.
- Created custom reporting systems and dashboards using **MySQL** queries, delivering real-time insights to leadership teams.
- Utilized **Hadoop ecosystem** tools such as **Hive** and **HDFS** to process and analyze large datasets, ensuring efficient data handling and storage in a distributed environment.
- Conducted data profiling, cleansing, and validation using **SAS** to ensure the integrity and consistency of analytical datasets.
- Implemented advanced visualizations such as heat maps scatter plots, and **KPI** indicators in **Power BI to** communicate complex data trends effectively.
- Designed and implemented budgeting and forecasting tools in **Excel** to support financial planning activities.
- Created user-friendly dashboards and reports in **Oracle BI** to present insights to stakeholders in an accessible format.
- Automated ETL workflows using **AWS Step Functions**, ensuring seamless data movement between relational databases and data lakes**.**

Environment: AWS Redshift, S3, Step Functions, RDS, Python, R, SQL, MySQL, Hadoop, Hive, HDFS, SAS, Power BI, SAS, Excel, Oracle BI.

EDUCATION: B.Tech in Computer Science and Engineering, June 2011 - May 2015 at JNTUH, India.