# **Pravallika Avula Data Analyst**

**Fairfax, VA | (571) 574-1074 |** [**pravallikaavula04@gmail.com**](mailto:pravallikaavula04@gmail.com) **| LinkedIn**

# **Summary**

* Data Analyst with 4+ years of industrial experience in big data analysis, data mining, statistical inference, A/B testing, machine learning, data visualization, and ETL data pipelines.
* Proficient in NoSQL and SQL databases, including PostgreSQL, MS SQL Server, MySQL, Oracle, MongoDB, and DynamoDB. Expertise in writing stored procedures, triggers, and complex queries with multiple joins, sub-queries, and window functions for report creation and analysis.
* Proficient in data visualization tools such as Tableau, and Python to create visually powerful and actionable interactive reports and dashboards
* Excellent Tableau skills, and expertise in building, and publishing customized interactive reports and dashboards with customized parameters and user - filters using Tableau.
* Experience in creating technical metadata for underlying physical data models of Source systems as well as Data warehouse
* Knowledge in documenting requirements changes using a Traceability matrix, good understanding of Business Requirements Documents (BRD), Functional Requirements Documents (FRD), and Software Requirements Specifications (SRS).
* Strong experience using Excel functions to dump the data and analyze based on business needs, proficient in PivotTables, VLOOKUPs, and Power Query
* Proficient in preparing Presentations, Graphs, and Pivot Tables using Microsoft Excel and PowerPoint.
* Working understanding of Tableau, MS Office Suite, MS Visio & Power BI for visualization & presentations Proficiency in Microsoft Office Suite (Excel, PowerPoint, Word, Visio & Outlook)
* Experienced in developing Python scripts for the full data science project lifecycle, including data acquisition, cleaning, exploration, and modeling using libraries like Pandas and Scikit-learn.
* Experienced in building ETL pipelines using Apache Airflow and AWS Glue, supporting data warehousing applications like Snowflake.
* Solid experience with cloud platforms including AWS, Azure, and Google Cloud Platform. Detail-oriented self-starter with strong communication skills, capable of presenting analysis results to technical and non-technical audiences and collaborating effectively within cross-functional teams.
* Able to translate business needs into complex analysis designs and recommendations
* Team player with detail and result-oriented problem-solving, communication, and interpersonal skills and quickly assimilate the latest technologies, skills, and ideas

# **Key Skills**

|  |  |
| --- | --- |
| **Data Analytics Tools:** | Power BI, Tableau, Jupyter Notebook, SQL Developer, Regression Analysis, Classification Modeling, Statistical Inference, Time- Series Analysis, Model Diagnostics |
| **Programming language:** | Python, R, SQL, and SAS |
| **Management Tools:** | JIRA, SharePoint, Confluence, Bitbucket, GitHub, Service Now |
| **Databases:** | PostgreSQL, MS SQL Server, MySQL, Oracle, MongoDB, DynamoDB, and RDS |
| **Methodologies:** | SDLC, Agile, Scrum, and Waterfall |
| **Data Manipulation**  **& Transformation:** | Data Wrangling, ETL & ELT (Extract, Transform, Load), Data Cleaning, Metadata Management |
| **Data Visualization and Reporting:** | Data Visualization, Dashboards & Reporting |
| **Analytics and Decision-Making:** | Risk & Decision Analysis, Analytical Skills |
| **Libraries:** | Pandas, NumPy, SciPy, TensorFlow, Scikit-Learn, and PySpark |
| **Others:** | AWS, Azure, GCP, ETL, Glue, Airflow, and DBT |
| **Operating Systems:** | Windows, Mac, and Linux |

# **Professional Experience**

## DATA ANALYST Jan 2024 – Present

### **Clairvoyant – USA**

* Developed and implemented statistical models and machine learning algorithms using scikit-learn, TensorFlow, and PySpark.
* Designed and implemented ETL processes using AWS Glue, Airflow, Informatica Power Center, and DBT to ensure seamless data flow and integration across systems.
* Managed and optimized data pipelines for efficient data extraction, transformation, and loading.
* Utilized Snowflake for data warehousing solutions, ensuring scalability and performance for large datasets.
* Managed and optimized databases, including PostgreSQL, MS SQL Server, MySQL, Oracle, MongoDB, and DynamoDB.
* Optimized existing SQL queries for faster performance on large datasets.
* Designed and maintained SQL Server databases, creating pipelines with user-defined functions and stored procedures for daily reporting tasks.
* Developed dashboards in Tableau to track KPIs and product performance, integrating data from sources including MS Excel, AWS S3, AWS RDS, AWS Redshift, JSON, and XML.
* Developed and maintained over 25 dashboards and reports using BI tools like Tableau and Power BI.
* Designed A/B testing metrics, created Tableau dashboards to monitor test processes, analyzed results, and provided stakeholder recommendations.
* Developed SQL code in Snowflake, Teradata, and Python scripting to meet evolving file transfer and automation project needs.
* Monitored email delivery statistics and optimized email efficiency to increase customer engagement.
* Analyzed A/B test results to determine optimal email delivery times for different topics using R.
* Developed source-to-target data mappings for reports and create pivot tables in Excel, increasing reporting efficiency by 25%.
* Collaborated with ETL developers to meet new requirements, creating data files and integrating new data into the data warehouse.
* Applied Agile, Scrum, and Waterfall methodologies to ensure timely delivery of projects and adaptability to changing requirements
* Used GitHub/Git for version control, ensuring codebase integrity and collaboration among team members.
* Trained 10+ team members on best data visualization and dashboard creation practices.
* Conducted market research and competitive analysis to identify trends and opportunities for growth.

## DATA ANALYST Sep 2018 – Sep 2022

### **Avenir Technology - INDIA**

* Worked on combined unstructured/structured data from multiple sources, automated cleaning using Python scripts.
* Extensively used Python's multiple analyses and visualization packages like Pandas, NumPy, Matplotlib, and Seaborn.
* Performed Exploratory Data Analysis, Data Wrangling, and finding trends and clusters.
* Implemented predictive modeling algorithms and created SQL queries to extract data from sales data marts, improving forecast accuracy by 20%.
* Designed mapping and identified, and resolved performance bottlenecks in Source to Target, Mappings.
* Gathered data from various external sources into a data warehousing system with the help of the ETL process.
* Leveraged cloud platforms such as AWS, Azure, and GCP for data storage, processing, and analysis.
* Retrieving data from the database through SQL as per business requirements.
* Created SQL tables with referential integrity and developed advanced queries using stored procedures and functions.
* Create, maintain, modify, and optimize MySQL and SQL Server databases.
* Involved in implementing machine learning model with Python’s Scikit-learn & Keras.
* Designed Tableau visualizations using Scatter Plots, Geographic Maps, Pie Charts, Bar Charts, and Density Charts, enhancing stakeholder engagement by 40%.
* Developed various solution-driven views and dashboards by developing different chart types including Pie Charts, Bar Charts, Tree Maps, Circle Views, Line Charts, Area Charts, and Scatter Plots in Power BI.
* Created reports utilizing Excel services and Power BI.
* Developed Power BI graphical/visualization solutions with BRDs & plans for creating interactive dashboards.
* Participated in Business Analysis, talking to business users, and determining the entities and attributes for the Data Model.
* Identified and determined physical attributes and their relationships through cross-analysis of functional areas.

# **Projects**

**Generating Synthetic flight tracks using Generative adversarial networks [Python, Tableau, and AWS]**

* Demonstrated problem-solving skills by implementing GANs with TensorFlow and PyTorch in Python, generating over 1M synthetic flight tracks and improving model performance by 90% through effective data preprocessing with MinMaxScaler and StandardScaler.
* Spearheaded the development of a scalable cloud-based system architecture using AWS (S3, EC2) for data storage and clustering, alongside Jupyter Notebooks for interactive model development, enhancing collaboration efficiency by 25%.

**Fuel consumption and CO2 emissions from passenger cars: Statistics and Analysis [Python, SQL, R]**

* Implemented Python scripts to transform data into ETL logic, performing real-time data cleaning, standardization, quality checks, and profiling, enhancing data efficiency by 56% and preparing datasets for analysis.
* Evaluated car specifications' impact on pollution using R and SQL for statistical and ad-hoc analysis, revealing a 15% reduction in CO2 emissions from 2000 to 2013 in newer car models, suggesting potential emissions reduction measures through technology and regulations.

**Wine Quality-Applied Statistics and Visualizations for analytics [Excel, SQL, R, Python]**

* Compared Regression data models in RStudio, based on R-squared, RMSE, MSE, and MAE values, finding that the random forest model is the best with a 50% accuracy for wine quality prediction and preferences.
* Performed Exploratory Data Analysis, and correlation analysis to locate anomalies, and outliers; classification, regression, and KNN clustering for effective data visualization and preprocessing, resulting in a 25% reduction in processing time.

# **Education**

## Master’s in data Analytics Engineering

## George Mason University, Fairfax, Virginia

## Bachelor in Electronics and communication engineering

## Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana