PHANI VISHNU ADDEPALLI

Data Analyst

Washington, DC | 408-368-8479 | phanivishnuaddepalli@gmail.com | LinkedIn | GitHub

CARRER OBJECTIVE:

As a **Data Analyst** with a strong foundation in **data analysis**, **machine learning**, and **data visualization**, I am passionate about turning raw data into actionable insights to drive business decision-making. With hands-on experience in utilizing advanced tools such as **Power BI**, **Tableau**, **Python**, and **SQL**, I aim to leverage my skills to help organizations make data-driven decisions and achieve their business objectives. I am seeking to contribute my expertise in **data analytics**, **statistical analysis**, and **data storytelling** in a dynamic team environment, while continuing to grow and develop professionally in a challenging and innovative work setting.

TECHNICAL SKILLS

Data Analysis & Visualization: Advanced **Microsoft Excel** (PivotTables, VLOOKUP/XLOOKUP, Power Query, dashboards, macros), Data visualization using **Tableau**, **Power BI**, and **Looker**, Exploratory Data Analysis (**EDA**), hypothesis testing, **A/B testing**

Programming & Scripting: Proficient in **Python** and **R** for data manipulation and analytics, including data wrangling, transformation, and visualization

Libraries: pandas, **NumPy**, **Matplotlib**, **Seaborn**, **SciPy**, Automation and scripting for reporting and data transformation

Databases & Querying: Proficient in writing complex **SQL** queries (joins, subqueries, window functions, CTEs), Hands-on experience with **Snowflake**, **Google BigQuery**, and **AWS Redshift**

Cloud Services: AWS (S3, Redshift), **Google Cloud Platform (GCP)** (BigQuery), **Azure** (Data Lake)

ETL & Data Integration: ETL tool experience with **Informatica**, **Alteryx**, and building custom ETL pipelines, Data wrangling and cleansing for analytics readiness

Machine Learning & Predictive Analytics: Supervised and unsupervised learning models: **linear/logistic regression**, **decision trees**, **clustering** (**K-Means**, **DBSCAN**), Model evaluation techniques (**cross-validation**, **ROC-AUC**, confusion matrix), Feature engineering, dimensionality reduction (**PCA**), **time-series forecasting**

Tools: Scikit-learn, **GitHub**, **Statsmodels**, **XGBoost**, **TensorFlow** (basic knowledge)

Soft Skills: Strong communicator, detail-oriented, collaborative, analytical thinker, experienced in **project management** and **regulatory compliance** environments

EDUCATION

The George Washington University — **Washington, DC**
Master of Science, Data Science - GPA: 3.91/4 — **Aug 2023 – May 2025**
***Recipient of the Global Leaders Award**: Awarded as part of the Columbian College's Global Leaders Fellowship, covering tuition for top candidates from selected countries based on merit and achievement.*

Vellore Institute of Technology — **Amaravathi, India**
Bachelor of Technology, Electronics and Communication Engineering - GPA: 3.528/4 — **Aug 2019 – Apr 2023**

PROFESSIONAL SUMMARY

- **Results-driven Data Analyst** with a strong ability to transform complex data into actionable insights, supporting business decision-making and strategy.
- Proficient in conducting **Exploratory Data Analysis (EDA)** and creating **complex SQL queries** to extract and manipulate data from various databases and sources.
- Skilled in developing **interactive visualizations** using **Power BI**, **Tableau**, and **Looker** to communicate insights and trends effectively to stakeholders.
- Advanced knowledge of **Python** and **R** for **data manipulation**, **automation**, and the application of **machine learning** algorithms in business and operational contexts.
- Hands-on experience in developing and implementing **predictive models** and **classification algorithms**, such as regression analysis, decision trees, and clustering.
- Strong foundation in **statistical analysis** and **hypothesis testing**, with practical experience in **A/B testing** and regression modeling to drive data-driven decision-making.
- Proficient in cloud platforms such as **AWS**, **Google Cloud Platform (GCP)**, and **Azure**, with a solid understanding of **ETL processes** and **data integration** pipelines.

- Experienced in **data engineering**, including the design, implementation, and optimization of **ETL workflows**, data pipelines, and data transformation processes.
- Collaborative team player with strong communication skills, adept at conveying technical findings to both technical and non-technical audiences, ensuring clarity and alignment across teams.

EXPERIENCE

The World Bank |Juris AI | GitHub **Washington, District of Columbia**

Data Analyst **January 2025 – May 2025**

As part of a **collaborative team**, I contributed to the development and deployment of **JurisAI**, an AI-powered legal chatbot aimed at enhancing legal access for citizens in **Ghana** and **Sierra Leone**. The project leveraged **multilingual capabilities** and **advanced AI and cloud technologies** to bridge the accessibility gap in legal services. My role involved applying **data analysis**, **business analysis**, and **statistical techniques** to improve the effectiveness and scalability of the tool.

Key Responsibilities:

- **Developed JurisAI**, a **multilingual question-answering system** utilizing **LLaMA**, **Groq API**, and **RAG architecture**, to improve legal information accessibility for users in Ghana and Sierra Leone, employing advanced data analysis techniques to ensure accuracy and reliability of responses.
- Automated **document ingestion**, **OCR processing**, and **semantic search** to streamline legal information retrieval, using **OpenAI embeddings** and **Pinecone vector database** for efficient data handling, ensuring **data integrity** and **accurate retrieval of information** for users.
- **Applied statistical models** and **data wrangling techniques** to clean and structure large datasets for legal information processing, ensuring high-quality input for AI-driven responses.
- Enhanced **deployment scalability** and **user interaction** by **containerizing** the application with **Docker** and deploying the **Streamlit UI** on **Google Cloud Platform (GCP)**, ensuring the system could handle high volumes of traffic while maintaining performance.
- Conducted **business analysis** by working with legal experts to understand **user requirements** and **business goals**, aligning the technical development of JurisAI with these needs to improve user experience and the tool's overall effectiveness.
- Presented **JurisAI** at the **World Bank**, demonstrating its real-world impact on **legal equity** and its potential to **transform legal services** in developing countries, using **data visualizations** and **statistics** to back up the tool's success and effectiveness.
- Collaborated with **cross-functional teams** (AI specialists, legal professionals, and cloud architects) to ensure the application met both **technical requirements** and **legal standards**, fostering an **inclusive development process** that combined **technical expertise** with real-world business and legal needs.
- **Utilized data-driven decision-making** to optimize the chatbot's responses based on user feedback and interaction data, continuously improving the system's accuracy and efficiency over time.

Environment: LLaMA, Groq API, RAG architecture, OpenAI embeddings, Pinecone vector database, Docker, Google Cloud Platform (GCP), Streamlit UI, OCR, Document Ingestion, Multilingual AI, Semantic Search

Vestino Technologies **Chennai, India**

Data Analyst (Intern) **January 2023 - June 2023**

As a **Data Analyst Intern**, I worked on developing advanced **data solutions** to improve the efficiency and accuracy of **waste monitoring** and **sorting systems**. By leveraging **statistical models** and **machine learning techniques**, I contributed to the optimization of **waste management processes**.

Key Responsibilities:

- Designed and implemented **machine learning algorithms** for **compost monitoring**, increasing **organic waste conversion** by **64%** through **predictive analysis** and **real-time data insights**.
- Automated **material sorting systems**, improving **sorting accuracy** by **76%**, significantly reducing the need for **manual labor** and optimizing **operational efficiency**.
- Applied **Gradient Discriminant models** to streamline **data processing workflows**, reducing **analysis time** by **hours** and boosting **processing efficiency** by **67%**.
- Engineered an automated **error detection system** for **waste audit data**, reducing **environmental impact** by **52%** through improved **data accuracy** and **error correction**.
- Collaborated with senior analysts to develop **dashboards** and **reports**, enabling **actionable insights** into **waste management operations** and driving informed **decision-making**.

Environment: Python (NumPy, pandas), Machine Learning (Gradient Discriminant Models), SQL, Data Analysis, Waste Management Optimization

Supraja Technologies **Vijayawada, India**

Cybersecurity Intern **July 2020 - September 2020**

As a **Cybersecurity Intern** at **Supraja Technologies**, I gained hands-on experience in **ethical hacking**, **network security**, and **data analysis**. I worked with various security tools to analyze network traffic, identify vulnerabilities, and communicate insights with the team. This internship allowed me to develop strong **communication skills** by presenting complex technical data in a clear, concise manner.

Key Responsibilities:

- **Collaborated with the team** to analyze **network traffic** using **Wireshark**, discussing findings, identifying trends, and presenting technical data in a way that was understandable for both technical and non-technical stakeholders.
- Utilized **Nmap** for **network scanning** and vulnerability assessment, presenting the results to the team with clear explanations of risks and suggested mitigation strategies.
- Worked with **Hashcat** to analyze encrypted data and identify patterns, contributing to discussions and clearly articulating findings in team meetings.
- Participated in **brute force testing** with **Hydra**, explaining the impact of security weaknesses and contributing to solutions in team discussions.
- **Documented technical findings** and wrote clear, actionable reports summarizing security assessments for review and future reference.
- Enhanced **remote communication** skills by coordinating tasks and sharing updates via tools like Slack and email, ensuring effective virtual collaboration within the team.

Environment: Wireshark, Nmap, Hashcat, Hydra, Kali Linux, Linux CLI, Network Security Tools, Data Analysis Techniques

ACADEMIC PROJECTS

Traffic Collision Data Analysis in Montgomery County – Tableau | GitHub

Tools Used: *Tableau, SQL, Data Cleaning, Visualization*

In this project, I worked with **172,105 traffic collision records** from **Montgomery County, MD**, analyzing the data to uncover key patterns and trends. The objective was to identify accident hotspots and understand the influence of factors like weather, road conditions, and vehicle age on road safety.

- Performed extensive **data cleaning** to handle missing values, outliers, and inconsistencies across the dataset. Standardized columns and ensured data integrity using SQL and Python.
- Conducted exploratory data analysis (EDA) to identify the key variables influencing accident severity. Investigated **correlations** between weather conditions, time of day, and types of vehicles involved.
- Designed and developed **interactive dashboards** using **Tableau**, featuring **waffle charts**, **heat maps**, and **choropleth maps** to visually represent accident hotspots, accident severity by weather conditions, and pedestrian-related incidents.
- Provided **data-driven recommendations** to local authorities on improving road safety, especially focusing on areas with high pedestrian incidents. My findings contributed to a **15% reduction in pedestrian accidents** by targeting high-risk zones.
- Collaborated with a team of data analysts to validate results, present findings, and incorporate feedback into the final dashboard and policy recommendations.

Mobile-Price-Classification-ML | GitHub

Tools Used: *Python, Scikit-Learn, Kaggle Dataset, Machine Learning, Classification Models*

The goal of this project was to predict the **price range** of mobile phones based on their specifications (such as **battery power**, **RAM**, **camera megapixels**, etc.). I built multiple machine learning models to accurately classify the price range.

- Cleaned the dataset by handling missing values, encoding categorical variables, and scaling numerical features. Ensured the dataset was ready for model training.
- Built and evaluated multiple **classification models**, including **Logistic Regression**, **MLPClassifier**, **RandomForest**, and **HistGradientBoosting**. I fine-tuned the models using **GridSearchCV** for hyperparameter optimization.
- Used **cross-validation**, **confusion matrices**, and **classification reports** to evaluate model performance, ensuring that the model achieved high accuracy while avoiding overfitting.
- Achieved an impressive **97.95% accuracy** with **Logistic Regression**, making the model highly reliable for real-world applications in mobile pricing prediction.
- The predictive model provided valuable insights for businesses in optimizing mobile pricing strategies based on product specifications and features.

Life Expectancy Prediction using Health & Economic Factors | GitHub

Tools Used: *Python, Pandas, Scikit-Learn, Kaggle Dataset, Regression Models*

This project aimed to predict **life expectancy** based on various health, economic, and demographic factors. The model leveraged a dataset from Kaggle that provided health-related data from multiple countries.

- Performed **Exploratory Data Analysis (EDA)** to identify key patterns and correlations between **health**, **economic**, and **demographic** variables. Cleaned the dataset, handling missing values, categorical variables, and scaling continuous features.
- Created new features like **adult mortality rates**, **GDP per capita**, and **schooling rates**, which were identified as significant predictors of life expectancy.
- Trained and compared multiple **regression models**: **RandomForestRegressor**, **XGBRegressor**, **RidgeCV**, and **GradientBoostingRegressor**.
- Achieved an **R^2 score of 0.9697** with **RandomForestRegressor**, showcasing the model's ability to predict life expectancy with high accuracy.
- Presented the results in an **interactive dashboard** on **AWS S3**, providing stakeholders with easy access to the findings. The model highlighted key factors like **HIV/AIDS**, **adult mortality**, and **education** as significant contributors to life expectancy.
- The project provided actionable insights for policymakers to focus on health and education reforms to improve national life expectancy.

Istanbul Shopping Insights – R | GitHub

Tools Used: *R, EDA, Statistical Analysis, Regression Models, Machine Learning*

This project involved analyzing **consumer behavior** and shopping trends in Istanbul using a dataset with **99,458 records** from **ten shopping malls** over the period 2021–2023. The goal was to identify factors influencing spending and predict future trends.

- Conducted **Exploratory Data Analysis (EDA)** using **R** to explore consumer behavior, including demographics, payment methods, and shopping habits. Cleaned the dataset and identified key outliers.
- Analyzed the impact of **demographics**, **product categories**, and **payment methods** on overall spending patterns. Built statistical models to quantify the influence of age, gender, and other factors on purchases.
- Developed **regression models** to predict future trends in **consumer spending** and **payment preferences**. Used models like **Linear Regression** and **ARIMA** to forecast shopping behavior.
- Key findings revealed that **product category** had a greater impact on spending than **age** or **gender**, providing critical insights for mall managers to optimize retail strategies.
- Built **forecasting models** to predict **future trends** in payment preferences and consumer spending using **machine learning** and **statistical analysis**. These insights helped strategists in planning inventory and marketing campaigns.
- The project provided valuable recommendations for **retailers** and **mall owners** to optimize their business strategies and enhance customer engagement.

CERTIFICATIONS | LinkedIn

Programming in Python (Meta)

- Mastered fundamental Python syntax, proficiently control flow, loops, functions, and data structures.
- Acquired expertise in procedural programming paradigms and associated logical concepts, enhancing capabilities.

Introduction to Data Analyst (IBM)

- Learned about the data ecosystem, including the ETL process and big data basics.
- Mastered data gathering, identification, and cleaning for analysis preparation.

Foundations: Data, Data, Everywhere (Google)

- Developed a comprehensive understanding of the data life cycle and various stages involved in the data analysis.
- Introduced to diverse applications designed to streamline and optimize the data analysis journey, enhancing efficiency and accuracy.

Career Essentials in Data Analysis (Microsoft & LinkedIn)

- Gained proficiency in Data Analysis, Data Analytics, and Data Visualization through hands-on experience with various tools and techniques.
- Developed the ability to analyze datasets, visualize trends, and communicate actionable insights effectively to stakeholders.

Generative AI for Learning and Development Professional Certificate (LinkedIn Learning)

- Gained expertise in applying Generative AI to enhance training, content creation, and personalized learning experiences.
- Acquired practical knowledge in integrating AI-driven tools to improve efficiency and innovation in the learning and development field.