Yengkong Vang Sayaovong

414-739-1246• Milwaukee, WI (Open to Relocation) • YSayaovong@gmail.com • https://linkedin.com/in/ysayaovong
https://github.com/YSayaovong • https://ysayaovong.github.io/Portfolio/index.html

Data Analyst

Data Analyst with hands-on experience from a current internship, specializing in Python, SQL, and data visualization. Skilled in analyzing complex datasets, creating data models, and generating actionable insights to support business decision-making. Proficient in data warehousing, ETL processes, and building interactive dashboards that drive strategic growth. Actively seeking a Data nalyst position to utilize my analytical skills and contribute to meaningful data-driven solutions.

Career Highlights

Data Engineer Intern | Refonte Infini – Remote 11/2024 – Present

- Cleaned and preprocessed data by addressing missing values, removing duplicates, and encoding categorical data, ensuring data accuracy and completeness for modeling and analysis.
- Performed exploratory data analysis (EDA) using seaborn and matplotlib to visualize distributions, correlations, and insights related to survival rates by passenger characteristics.
- Engineered new features to enhance model performance, carefully selecting relevant variables that significantly impacted prediction accuracy.
- Developed a predictive model using a Random Forest Classifier in scikit-learn, leveraging ensemble learning for improved robustness and classification accuracy.
- Evaluated model performance and reliability through metrics such as the confusion matrix and classification report, ensuring the model's effectiveness and precision.
- Identified key features influencing survival, providing insights into critical factors like age, fare, and passenger class that impacted survival rates.
- Optimized model parameters with hyperparameter tuning to achieve improved accuracy and balanced performance across evaluation metrics.
- Documented the analysis process and results in Jupyter Notebook, establishing a clear and reproducible workflow for future enhancements.

Mechanical Designer | Prolec-GE Waukesha – Waukesha, WI 10/2020 – Present

- Designed transformers and nameplates, ensuring compliance with industry standards using CREO 8.0.
- Led process optimization initiatives, reducing errors and improving design efficiency by 10%.
- Mentored 6 detailers, providing guidance to improve overall team performance.
- Worked cross-functionally to ensure timely and high-quality project delivery.
- Facilitated production and training sessions to enhance team knowledge and ensure efficient operational workflows.
- Developed an Excel-based dashboard to track project metrics, including error rates and completion percentages, for transformer detailing and nameplate drafting.
- Created automated visualizations (line and pie charts) to analyze error trends over time, improving personal workflow and quality control.
- Enhanced efficiency by implementing data-driven self-monitoring, reducing manual tracking time and proactively addressing recurring errors.

Education

B.S., Information Technology; Minor in Music | Arizona State University, Expected Completion: 5/2025

A.I. & Machine Learning Engineer Career Path | Zero to Mastery, Expected Completion: 5/2025

A.S. in Mechanical Design Technology | Milwaukee Area Technical College, 5/2021

Technical Proficiencies

Languages:	**Python (Advanced):** Object-oriented programming, data structures, algorithms, backend development (Flask, Django), and process automation. **JavaScript (ES6+):** Backend services, asynchronous operations, and event-driven architecture. **C++ (Basic), C# (Basic), Java (Basic):** Foundational programming knowledge
Machine Learning & Data Science:	**Scikit-learn, Pandas:** Building, training, and tuning ML models; efficient data manipulation and preparation; proficient in feature engineering and EDA to uncover patterns and insights. **NLP:** Text preprocessing and sentiment analysis for insights from unstructured data sources.
Data Engineering:	**Apache Kafka, Spark, Hadoop:** Building scalable, high-performance data pipelines and distributed processing of large datasets for real-time and batch ingestion. **ETL & Data Warehousing:** Constructing end-to-end pipelines ensuring data quality and accessibility; designing warehouses that support comprehensive business intelligence.
Cloud & DevOps:	**AWS (EC2, S3, Lambda):** Scalable infrastructure, production deployment, and serverless computing. **Azure Synapse:** Unified analytics for large-scale data exploration and management. **CI/CD & Containers:** Docker, Jenkins, GitHub Actions for automated builds, testing, and deployment.
Database	**PostgreSQL, SQL, NoSQL (MongoDB, Couchbase):** Complex queries, relational and NoSQL databases, executing ETL for large-scale workflows, optimizing data storage and retrieval.

Projects

Predictive Analysis and Data Insights on the Titanic Dataset November 2024

Description: Conducted a comprehensive analysis of the Titanic dataset to uncover factors influencing passenger survival and developed a predictive model to estimate outcomes.

- **Data Preprocessing:** Addressed missing values, removed irrelevant columns, and prepared the dataset for analysis.
- **Exploratory Data Analysis (EDA):** Utilized Python libraries (Seaborn, Matplotlib) to visualize survival trends across variables such as class and age.
- **Feature Engineering:** Selected and prepared key features for optimal model training.
- **Machine Learning Model:** Developed a Random Forest Classifier in scikit-learn, evaluating performance with metrics including confusion matrix and classification report.

Key Skills: Python, Data Analysis, Machine Learning, Data Visualization, EDA, scikit-learn, pandas, Seaborn, Matplotlib.

Github repository: https://github.com/YSayaovong/Titanic-Dataset-Analysis

NBA Game Performance Analytics November 2024

Description: Conducted an in-depth analysis of NBA team performance data for the 2022-23 season, focusing on identifying the top 10 teams by average points scored per game.

- **Data Retrieval**: Retrieved NBA game data for the 2022-23 season using the nba_api library, focusing on official games and excluding non-standard events.
- **Data Filtering**: Filtered the dataset to retain relevant game data, ensuring accuracy by excluding All-Star and other non-official games.
- **Data Analysis**: Calculated average points scored per game for each team, highlighting key trends in scoring performance.
- **Data Visualization**: Developed a bar chart using Matplotlib and Seaborn to visualize the top 10 NBA teams by average points, providing a clear and comparative view of team performances.

Key Skills: Python, Data Analysis, Data Visualization, nba_api, Pandas, Matplotlib, Seaborn
Github repository: https://github.com/YSayaovong/NBA-Game-Performance-Analytics