Mustapha Ayle

mustapha.ayle@gmail.com | (714) 222-9954 | Orange, CA | github.com/mhayle

EDUCATION

M.S. in Statistics   **March 2023**
> University of California Irvine

B.S. in Applied Mathematics and Statistics   **May 2021**
> California State Polytechnic University Pomona

SKILLS

- *Languages:* SQL, Python (Pandas, NumPy, Scikit-Learn, PyTorch), R (tidyverse, Shiny), VBA, SAS, MATLAB
- *Statistics*: Linear Regression, Logistic Regression, Time Series Models, Survival Analysis, Random Forests, Nearest Neighbors, Natural Language Processing (NLP), Neural Networks, A/B Testing
- *Tools*: Microsoft Excel, Microsoft PowerPoint, Looker, Tableau, Power BI, AWS Redshift, GitHub

EXPERIENCE

Data Analyst   **05/2023 – Present**
Big Blue Marble Academy – Atlanta, GA

- Developed ETL scripts using Python to enhance data collection, analysis, and reporting across a network of 50+ schools in 10+ states, leading to a 70% improvement in finance and business team efficiency.
- Leveraged Excel to analyze large datasets, utilizing pivot tables for multi-dimensional data analysis, VLOOKUP/HLOOKUP for efficient data retrieval, and complex formulas for advanced calculations and data manipulation, enhancing the accuracy and insights of business reporting tools.
- Used Python and Windows Task Scheduler to develop automated data pipelines, optimizing various workflows, reducing manual intervention, and enhancing overall data accuracy for multiple company-wide projects.
- Successfully condensed a 6-month data collection and reconciliation project into a 1-month timeframe using innovative automation strategies, saving hundreds of man-hours and accelerating stakeholder decision making.
- Integrated, processed, and cleaned diverse data sets from multiple data sources, ensuring a 95% data accuracy rate and facilitating a seamless workflow for downstream analysis.

Data Science Intern   **06/2020 – 09/2022**
GoFundMe – Redwood City, CA

- Built a logistic regression model using R with 20+ predictors to identify what where the leading forces driving users to leave a tip for the site after donating and finding places to improve user engagement.
- Developed complex SQL queries on AWS Redshift using window functions and joins to efficiently extract and transform large datasets, resulting in faster, more precise analysis that informed key business strategies.
- Utilized time series modeling in R to refine the accuracy of monthly user count forecasts, optimizing customer service staffing needs and aiding in marketing campaigns adjustments during low and high periods.
- Created and implemented an NLP in Python model using LDA and tf-idf to categorize ~10,000 survey responses, gaining actionable insights from users and leading to more targeted communication and campaign strategies.
- Designed engaging and interactive dashboards using Looker to share and visualize KPIs and make them easily understandable and actionable for stakeholders.

PROJECTS

arXiv Publication Analysis

- Created a database using arxiv.org's open-access API containing over 2.3 million research papers in STEM fields and ensured its constant updating with automated scripts run daily through Windows Task Scheduler.
- Performed extensive analysis to identify popular research areas, study trends, analyze overlaps between subjects, and create visualizations such as word clouds and network graphs, using Jupyter notebooks.
- Constructing a dashboard using Python Dash to display findings in an interactive webpage hosted on GitHub.
- Full description and code for the project are available at github.com/mhayle/arxiv-analysis.