Ayaan Omair

omairayaan@gmail.com ❖ (480) 570-4945
LinkedIn: www.linkedin.com/in/ayaan-omair-166b6b223
Personal Website https://balleromair12.github.io/my-portfolio/

EDUCATION

Master of Science, Data Science **December 2025**
Texas A&M University, College Station, TX

Bachelors, Mathematics (Statistics); Psychology minor **August 2020 - May 2024**
Arizona State University**,** Tempe, AZ 3.90 GPA *(summa cum laude)*

TECHNICAL SKILLS

Languages and Tools:

- **Python, R, SQL (MySQL, PostgreSQL)**, SAS, JavaScript, HTML/CSS, Linux, Excel, Tableau, Alteryx, Jupyter, LaTeX

Libraries & Machine Learning Methods:

- Pandas, NumPy, Scikit-learn,Matplotlib, **Decision Trees, Regression (Linear, Logistic, Lasso, Ridge, Poisson, etc.)**, K-Means, Gaussian Mixture Models, Mean-Shift, DBSCAN, **PCA, KNN, SVM**

Coursework:

- Data Mining and Analysis**,** Linear Algebra, **Applied Linear Regression**, Mathematical Statistics, Scientific Computing, **Data Wrangling with SQL, Data Analysis with Python, Exploring Data in R/Python,** Applied Analytics

RELEVANT EXPERIENCE

SQL and Python Trainee - Global Tech Experience **May 2023 - July 2023**

- Devised ***SQL*** queries to efficiently extract, analyze, and manipulate complex datasets for actionable insights
- Analyzed and visualized data findings using ***Jupyter Notebook*** and ***Python***
- Gained insights into global business strategy through EDA, summary statistics, and visualization
- Collaborated with a global team to perform different tasks using SQL and Python
- Obtained experience using Python and SQL by creating various personal and professional projects

PROJECTS

Disease Prediction Model **April 2025**

- Constructed ***logistic regression*** and ***decision tree*** models in ***Python*** to predict Heart Disease, Kidney Disease, and Skin Cancer
- Achieved up to *75% accuracy* and *AUC scores* as high as *0.84* across all models, demonstrating strong model performance
- Identified high recall (e.g., *78%*) but low precision for positive cases due to class imbalance, highlighting the challenge of predicting rare disease occurrences
- Compared and evaluated model performance using *F1 score* and *AUC*, ensuring a comprehensive assessment while confirming model generalizability by observing no overfitting through training and testing accuracy comparison

Loan Default Prediction **March 2025**

- Built a ***logistic regression*** model in ***Python*** to predict loan default using LendingClub data
- Addressed class imbalance using sampling techniques such as ***SMOTE, ADASYN, and RandomOverSampler***
- Achieved up to *70% recall, 0.70 AUC,* and *0.35 F1-score* on unseen test data
- Conducted data preprocessing and model evaluation to compare resampling strategies and enhance prediction performance

Grammy Awards Project **July 2023**

- Leveraged ***Python*** to perform data analysis and visualization of real website data used by the Recording Academy
- Examined the impact of splitting up a website into two separate websites (grammy.com and recordingacademy.com) by analyzing different variables (number of visitors per day, average session time, user interaction, etc.)
- Assessed and visualized website data for Grammy.com to identify key trends and provided data-driven recommendations aligned with business goals

PROFESSIONAL EXPERIENCE

Sports Data Operator *(Part-time)* **May 2023 – *Present***
SportRadar

- Watch major sporting events and record events in real-time using a mobile app for statistic gathering
- Attend live sporting events
- Remain knowledgeable of the rules of the sport and the teams/players involved

Data Analyst Intern **January 2024 - May 2024**
Internship *(for Credit)*

- Designed and implemented an ***ETL process*** to support SC Del Sol youth soccer coaches by extracting player data from multiple Excel sources, transforming it with ***Python (Pandas)*** for cleaning and formatting, and loading it into an interactive ***Excel*** dashboard
- Developed a user-friendly dashboard interface in Excel to visualize comprehensive player insights and statistics
- Collaborated with coaching staff to identify key metrics and tailor data views to their decision-making needs