Ayaan Omair

480-570-4945 | aomair1@tamu.edu | LinkedIn | Portfolio

Education

Texas A&M University — College Station, TX
Master of Science in Data Science — *Expected Graduation: Dec. 2025*
- **GPA:** 3.71/4.00

Arizona State University — Tempe, AZ
Bachelors of Science in Mathematics (Statistics) — *Graduation: May 2024*
- **GPA:** 3.90/4.00 *(summa cum laude)*

Skills

Languages & Tools: Python, R, MySQL, PostgreSQL, SAS, JavaScript, HTML/CSS, Linux, Power BI, Excel, Tableau, Alteryx, Jupyter, LaTeX
Libraries & Machine Learning Models: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Decision Trees, Regression (Linear, Logistic, Lasso, Ridge, Poisson, etc.), K-Means, Mean-Shift, PCA, KNN, SVM, Neural Networks
Coursework: Data Mining and Analysis, Applied Linear Regression, Mathematical Statistics, Scientific Computing, Data Wrangling with SQL, Data Analysis with Python, Exploring Data in R/Python, Applied Analytics

Professional Experience

Sports Data Operator — Various Locations
SportRadar — *May 2023 – Present*
- Watch major sporting events and record key actions in real-time using a mobile app for live statistics gathering
- Attend live sporting events to collect on-site data
- Remain knowledgeable of sport rules, teams, and players to ensure accurate data reporting

Data Analyst Intern — Phoenix, AZ
Arizona State University — *January 2024 – May 2024*
- Designed and implemented an ***ETL*** process to support youth soccer coaches by extracting player data from over **1,000** Excel files, combining and preprocessing it with ***Python (Pandas)***, and loading results into Excel
- Developed a user-friendly interactive *dashboard* in ***Excel*** to visualize comprehensive player insights and statistics
- Coordinated with coaching staff to identify important metrics and adjust how data was shown to meet their needs
- **Tech Stack:** Python, Pandas, Excel

Relevant Experience

Global Tech Experience — Remote
SQL & Python Trainee — *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***
- **Tech Stack:** Python, Matplotlib, SQL, Tableau

Projects

Disease Prediction Model | *Python, Scikit-Learn, Pandas, NumPy, Matplotlib, Jupyter Notebook*
- Constructed *logistic regression* and *decision tree* models in ***Python*** to predict Heart, Kidney, and Skin Disease
- Achieved up to **75%** accuracy and AUC scores as high as **0.84** across all models
- Identified high recall (**78%**) with low precision due to class imbalance between diseased and undiseased cases
- Compared model performance using F1-score and AUC; verified no overfitting by analyzing training vs. test accuracy

Loan Default Prediction | *Python, Scikit-Learn, Imbalanced-learn, Pandas, NumPy, Jupyter Notebook, Matplotlib*
- 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 performance