Data Analyst

WILL MIZER

(941) - 812 - 7787 | will.mizer@gmail.com
Portfolio Website
Will Mizer | LinkedIn
willmizer | GitHub
Sarasota, FL
GPA: 3.81

Education

BACHELOR OF SCIENCE | DATA SCIENCE – Florida Polytechnic University – Lakeland, FL Expected 5/2027

ASSOCIATE OF SCIENCE | COMPUTER PROGRAMMING & ANALYSIS – State College of Florida – Venice, FL

Relevant Skills

Programing: Python (pandas, matplotlib, seaborn, scikit-learn), R (dplyr, ggplot, tidyr), SQL

Data Visualization: Tableau and Power BI (interactive dashboards, data storytelling), Excel, Office 365

Big Data: Hadoop, PySpark, MapReduce

Soft Skills: Leadership and effective communication

Projects

PYTHON | SQL: VIRGINIA HOUSING ANALYSIS – Personal Project (2025)

- **Improved price prediction accuracy by 50%** (from ~140K to ~70K) using a tuned **XGBoost Random Forest** model.
- Built a **tool to estimate house prices** based on square footage, acreage, location, and amenities**, aiding build vs. buy decisions.**
- **Empowered family relocation** with data-driven insights, **boosting confidence in Virginia home buying**.
- Enabled builders **to forecast future home values** based on lot size and locations for **market-aligned pricing**.

PYTHON: MOVIE-MATCH AI MOVIE RECOMMENDATION SYSTEM – Personal Project (2025)

- **Increased recommendation accuracy from 10% to 35%** using **SentenceTransformer** embeddings and **Optuna-tuned weights**.
- **Developed a hybrid similarity model** for movie recommendations based on cast, director, genre and themes
- **Created an interactive tool** for personalized movie suggestions, **solving group decision-making challenges**
- **Managed full ML pipeline**, from data cleaning to model tuning, using **Python and scikit-learn**

PYSPARK: END TO END DATA MOVIE PERFORMANCE PREDICTION – University Project (2025)

- **Led end-to-end development** of a scalable PySpark pipeline processing over 70M records to predict movie performance.
- **Enabled producers to simulate and assess** the commercial viability of movie ideas by inputting key features into a trained linear regression model.
- **Processed 70M+ IMDb entries** using distributed computing in PySpark, demonstrating scalability and real-world applicability.
- **Visualized prediction trends** and explained model outcomes with interpretable metrics, supporting data-backed greenlighting decisions.