Data Analyst

WILL MIZER

(941) - 812 - 7787 | will.mizer@gmail.com
Will Mizer | LinkedIn
willmizer | GitHub
Portfolio Website
Sarasota, FL 34275
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 12/2023

Skills

Python
Libraries: pandas, matplotlib, seaborn, sci-kit learn

Tableau
Interactive dashboards and visualizations
Storytelling and filtering

R Programming
Packages: dplyr, ggplot, tidyr

SQL
Data cleaning, joins, query optimization
Aggregation functions, filtering and group by

Microsoft Office
Emphasis on Excel, VLOOKUP, pivot tables, basic formulas

Hadoop
Tools: PySpark, MapReduce

Projects

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

- **Improved price prediction accuracy by 50%** (from ±$140K to ±$70K) by switching from basic regression to a tuned **XGBoost Random Forest** model.
- Built a **predictive tool to estimate house prices** based on square footage, acreage, location, and amenities**, helping users evaluate build vs. buy options.**
- **Empowered family members with data-driven insights** on home buying decisions in Virginia, **increasing confidence in relocation choices.**
- **Enabled users to estimate future home values,** based on lot size and location, helping prospective builders forecast market-aligned pricing in target areas.

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

- **Boosted recommendation accuracy from 10% to 35%** by replacing TF-IDF with **SentenceTransformer** embeddings and tuning feature weights via Optuna.
- **Engineered a hybrid similarity model** that recommends similar movies based on cast, director, overview, genre, and themes using weighted cosine similarity.
- **Developed an interactive tool** that allows users to input a movie and receive personalized recommendations, solving a real-world group decision problem.
- **Led full ML pipeline development** from data cleaning and vectorization to model evaluation and fine-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 to predict movie success using runtime, genre, budget, cast/director and various other metrics.
- **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.