**🏅 Final Report: Olympics Data Analysis Project**

**👨‍💻 Project Title:**

**Summer Olympics Medal Analysis and Prediction (1976–2008)**

**📌 Objective:**

To perform in-depth data analysis on historical Summer Olympic medal data and build a predictive model to identify patterns in athlete performance, gender trends, and country dominance across years.

**🛠 Tools & Technologies Used:**

* **Python:** Pandas, Seaborn, Matplotlib, Scikit-learn
* **Excel:** Data cleaning, basic exploration, pivot charts
* **SQL:** Data querying and summarization
* **Power BI (Optional):** Visual dashboard layout planning
* **Machine Learning:** Logistic Regression model for medal prediction

**📊 Key Visual Insights:**

1. **Top 10 Countries by Medal Count**  
   → USA dominates across Olympic years
2. **Gender Participation Distribution**  
   → Clear visualization of gender disparity and trends
3. **Medals Awarded Over Time**  
   → Highlights medal trends and fluctuations over years
4. **Top 10 Athletes**  
   → Names like Michael Phelps and Venus Williams stood out
5. **Top Sports by Medals**  
   → Aquatics, Gymnastics, and Athletics lead the chart

**🤖 Machine Learning Output:**

* **Model Type:** Logistic Regression
* **Goal:** Predict if an athlete would win a medal or not
* **Features Used:** Year, Sport, Discipline, Country, Gender
* **Result:**
  + Accuracy: ~83%
  + Output: Confusion Matrix, Classification Report

**🧾 SQL Analysis:**

Queries included:

* Top 5 countries by medal count
* Medals by year
* Gender participation count
* Medal type breakdown (Gold, Silver, Bronze)

**💡 Key Learnings:**

* Effective use of real-world sports data for analysis
* Applying ML to historical event data
* Visual storytelling using Python and Excel
* Understanding how to work with cross-functional tools (SQL + ML + Viz)

**✅ Outcome:**

This project gave me hands-on experience in handling real datasets, understanding domain-specific KPIs, performing predictive analysis, and presenting business insights visually and statistically.

**📍 Remarks:**

“Completed solo within 28 days as part of my **Data Analyst Internship at Unified Mentor**. This project simulates a real-world end-to-end data workflow combining analytics, visualization, machine learning, and reporting.”