Paris Olympics 2024: GDP, Athlete Participation & Medal Outcomes

Data Analytics Case Study

Executive Summary

This case study analyzes the relationship between **national economic strength**, **athlete participation**, and **Olympic medal outcomes** using data from the Paris 2024 Summer Olympics. The objective is to evaluate whether **structural economic advantages** and **delegation size translate into measurable competitive success.**

Business & Analytical Questions

• Does a **country's GDP** correlate with the **total number of medals won**?
• Does sending **more athletes** increase **medal outcomes**?

Data Sources

• GDP by country (Wikipedia – **web scraping**)
• Medals by country (Codante.io **Olympic Games API**)
• Number of athletes per country (Olympic participation pages – **web scraping**)
• **Normalization dataset** for country codes and **naming consistency**

Methodology

An **end-to-end** exploratory data analysis **(EDA)** pipeline was implemented, including **API consumption**, **web scraping**, **data cleaning**, **normalization**, **dataset integration**, **correlation analysis**, and **linear regression modeling**. **Matplotlib** and **Seaborn** were used to generate **statistical visualizations** supporting the findings.

Key Findings

• A **positive correlation** exists between **GDP** and total **medals won**.
• Countries with **larger athlete delegations** consistently achieve **higher medal counts**.
• **Economic capacity** and **participation scale** act as **performance multipliers**.

Business & Policy Implications

The findings support **data-driven sports funding strategies**. Investments in **athlete development**, **infrastructure**, and broader participation are associated with **improved competitive outcomes** and higher return on investment **(ROI)**.

Tools & Technologies

Python, Pandas, NumPy, BeautifulSoup, REST APIs, Matplotlib, Seaborn, Jupyter / Google Colab

Conclusion

This project demonstrates **applied data analytics skills** in real-world **data acquisition**, **multi-source data integration**, **statistical analysis**, and **executive-level insight communication**. It is directly applicable to Data Analyst, Business Intelligence, and Applied Data Science roles.