

Final Project Proposal: An analytical look at equestrian sport

Gabriela Jiménez Conde

10/05/2025

Business Case Description

The numbers tell the stories behind the sport. In competition, data analysis allows us to understand athlete performance, anticipate results, and improve decision making. However, in the equestrian world, we run into a problem: there is no accurate and meaningful measurement of the performance of the horse-rider combination in events.

Horse riding is not a linear sport, but each competition is influenced by multiple factors, from the level of the competitors, the difficulty of the course, the conditions of the terrain, the physical condition of the horse and rider, among other things. All this makes the equestrian competition a unique sport, in which even the best riders and horses have all the odds stacked against them.

This complexity makes horse riding a difficult sport to understand for those who are unfamiliar with it. In addition, the limited use and accessibility of data makes it difficult to make informed decisions both for users within the sport, such as riders, trainers, or event organizers, and also for users outside the sport, the fans. This project is born with the aim of changing this reality, making use of data as a strategic and informative tool, and to promote a better understanding of the sport.

Objectives and expected impact

The main objective of the project is to unveil the story behind competition results. Through data analysis, it seeks to add value to equestrian competition, improving understanding, decision-making, and communication in the equestrian sport environment.

It is a project focused on different audiences. For fans, it seeks to improve the understanding of the sport from the outside. For riders, there will be a historical analysis of their performance, focused on providing personalized feedback to optimize training and set goals based on past results. Event organizers and brands will have at their disposal a way to tell the story of the sport that will improve storytelling, leading to better audience engagement. The analysis will be focused on two of the Olympic disciplines: show jumping and eventing.

For this purpose, an interactive dashboard will be developed to show general insights of the equestrian competition, as well as explanations of past results. On the other hand, a results prediction system will be generated to anticipate the performance of the horse-rider combinations in future competitions.

Technologies and tools

The project is divided into several phases, for which different tools will be used:

- **ETL process:** information extraction tools will be used, such as APIs or existing databases, as well as SQL with Python for loading the data to the database, using tools such as Psycopg2.
- **EDA process:** SQL and Python tools, such as Pandas, Matplotlib or Seaborn, will be used in order to obtain relevant insights.
- **Predictive system:** Python prediction tools will be used.
- **Dashboard and final report:** Power BI and Streamlit will be used to create insights in a visual and understandable way.

Data sources

The data sources used in the project are listed below:

- **FEI Database/API internal to the FEI** (<https://data.fei.org>): this is the primary source for the project, from which information on competitions, riders and horses will be obtained in detail.
- **FEI website** (<https://www.fei.org/>): this source provides information on regulations and on the general functioning and rules of the tests and disciplines analyzed.
- **size Equiratings** (<https://www.equiratings.com/>): this source provides information on the predictive model and on the performance of data analysis applied to equestrian competition.

Next Steps

Once the main objectives of the project have been achieved, a series of lines of work to be developed in the future are set out below:

- Revenue/cost information through the FEI Annual Reports of each equestrian competition, so that a revenue/cost analysis of the competitions can be performed.
- Perform analysis for other equestrian disciplines outside of the Olympic disciplines.
- Create dashboards and predictions specifically for federations, initially for the FHE (Spanish Equestrian Federation) or the FDHM (Madrid Equestrian Federation). In this case, the data would be obtained from the official websites of the federations (<https://rfhe.com> or <https://www.fhdm.es>) or from equestrian magazines (Oxersport, Avanza Eventos, Ecuestre).