CZ4124/SC4024 Assignment #1 - Summary Report

Name: Juan Felix Pangestu **Title**: Exploratory Data Analysis on Powerlifting

1. Objectives

Describe clearly the objectives of the project and the messages the presentation was meant to communicate.

The essence of this project was to analyze the powerlifting records dataset by openpowerlifting to answer key questions that enhance our understanding of the sport. We investigated where and when the sport's popularity such as from Figure 1 & 2. Additionally, we focused on analysing trends that we could find from the various columns from the dataset. Such included figuring out the how different equipment affected athletes in their total weight lifted shown in Figure 3 & 4

Furthermore, we compared male and female athletes in terms of absolute and relative strength to recognize gender differences in performance and promote inclusivity within the sport. Finally, we aimed to identify the best measurement of strength in powerlifting by evaluating various metrics, thus providing a nuanced understanding of effective strength assessment in competitive contexts.

In all, the objective of the project is to showcase powerlifting itself as a sport and increase awareness of it.

2. Novelty

Highlight what you think were original contributions in your data visualisation that you are particularly proud of. What new visualisation techniques did you use in creating the video presentation? If some of the major reference sources [2] have contributed preliminary ideas and rendered technical help during design and implementation of the project, you should describe what improvements you applied to these ideas to make your contribution original.

In my data visualization project, I take particular pride in my innovative use of a stacked histogram to analyze the absolute strength differences between male and female powerlifters. This technique provided a clear visual representation of strength distribution across genders, not only highlighting the differences in absolute strength but also offering valuable insights into the overlap between these distributions. By effectively illustrating these nuances as can be seen in Figure 5, the visualization enhances understanding of how gender influences performance in powerlifting.

Additionally, I successfully classified the weight and age categories within the dataset provided by OpenPowerlifting, which was crucial for accurate analysis and interpretation within the context of powerlifting competitions. By adhering to the established classifications defined by the International Powerlifting Federation (IPF), I ensured that the data was processed systematically.

3. Technical Challenges and Innovation

What visualisation tools did you employed in exploring your data and creating your visualisation and presentation? Describe any noteworthy technical contributions done during the design and implementation of the visualisation. What were the technically challenging aspects in creating the visualisation?

I utilized the Matplotlib and Seaborn libraries to explore and visualize the data, which enabled me to identify key trends and derive meaningful interpretations. This initial analysis significantly deepened my understanding of the dataset. Afterward, I transitioned to Canva for video editing and to enhance my data visualizations. This step not only elevated the aesthetic quality of my visuals but also ensured that the information was presented in a clear and engaging manner.

One of the standout features of my visualizations was their animation, which made it easier for viewers to follow the data being showcased. However, I faced challenges in creating dynamic visualizations for certain graphs within Canva. The platform's limitations hindered my ability to achieve the desired level of interactivity. I think with this I learned how important it is to have video editing skills to tell my story better.

Additionally, I encountered complications due to a substantial number of null values across various columns in the dataset. This issue stemmed from sourcing data from multiple powerlifting federations, each adhering to different record-keeping standards. Consequently, ensuring data consistency and completeness became a crucial aspect of my analysis.

4. References

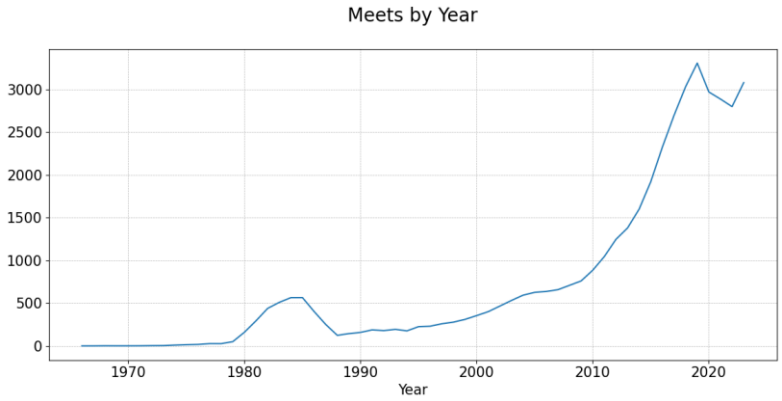
List the main reference sources that have contributed preliminary ideas and technical help during design and implementation of the project. The source and weblink where the datasets were taken from should be listed at the start of your list of references.

[1] Dataset used and weblink.
Dataset used: https://www.kaggle.com/datasets/open-powerlifting/powerlifting-database/data

[2] References and weblinks that have contributed preliminary ideas to your visualisation (if available).

[3] Other references and weblinks (if relevant) ….
Wilks Coefficient: https://en.wikipedia.org/wiki/Wilks_coefficient
Video reference to showcase powerlifting: https://www.youtube.com/watch?v=FjBavOW-mbA
Definition of Powerlifting: https://dictionary.cambridge.org/dictionary/english/powerlifting

Figures and Tables



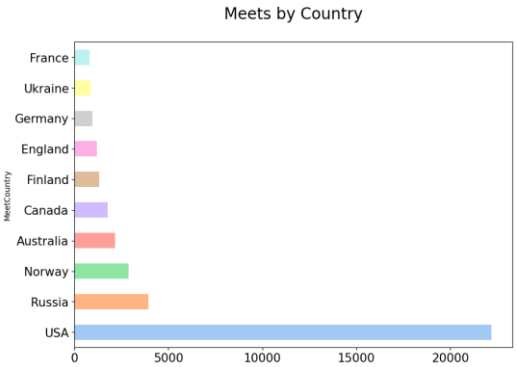
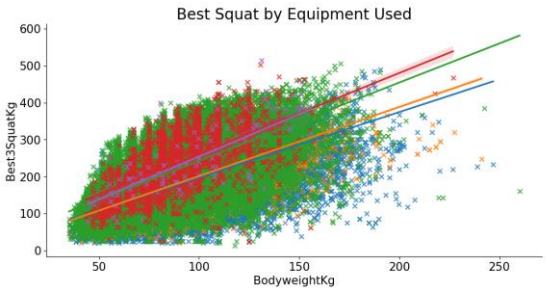


Figure 1 & 2



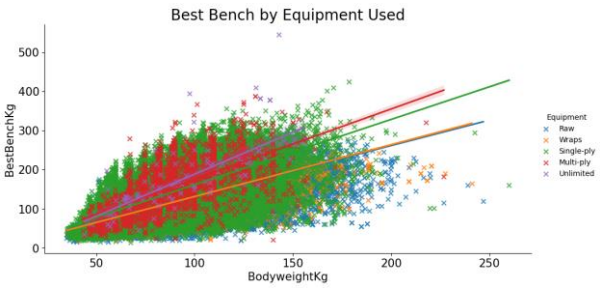


Figure 3 & 4

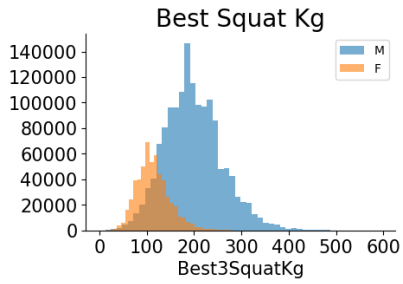


Figure 5