Data Science Project - Darts

Emilie Terhaar, Sara Rolfs, Tyra Kausch

Christian-Albrechts-Universität zu Kiel

Introduction

This project dives into the world of darts using data science techniques to uncover fascinating insights. It aims to transform raw data into structured datasets suitable for visualization and analysis. The goal is to answer predefined research questions by gathering data from multiple sources, processing it, and generating analytical datasets.

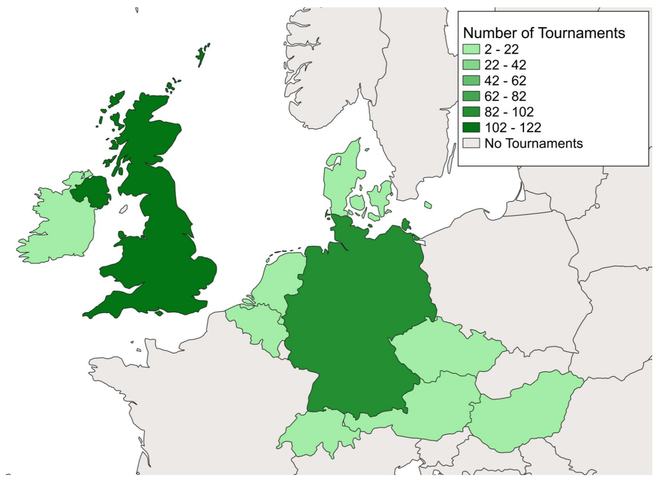
Method

The research questions – of which a few are represented on this poster – were formulated while looking ahead at available data sources. Web scraping scripts were then developed and executed to extract structured data. This data consisted of player statistics, match results, and other relevant information. The extracted data was validated, cleaned, and stored in organized CSV files. In the next step, datasets were transformed to ensure correctness and consistency, addressing any missing values or errors manually. The final datasets were then analysed and visualized using appropriate graphical representations, forming the basis for answering the research questions.

Results

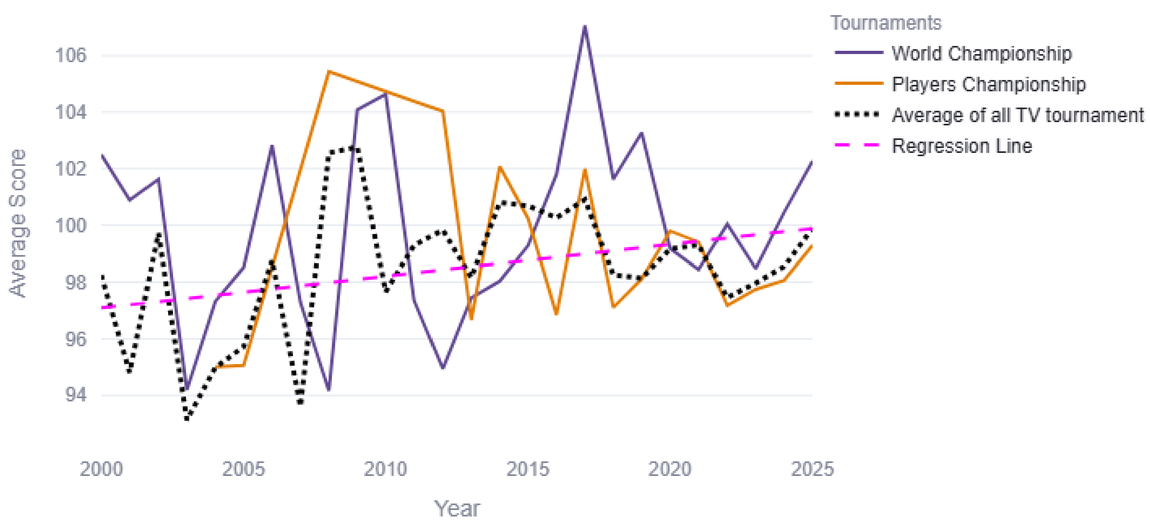
How do the tournament's locations are distributed?

There is a big discrepancy of tournaments held in the United Kingdom and Germany in comparison to the other countries. This shows that there is a much bigger influence on darts in Europe than in the rest of the world.



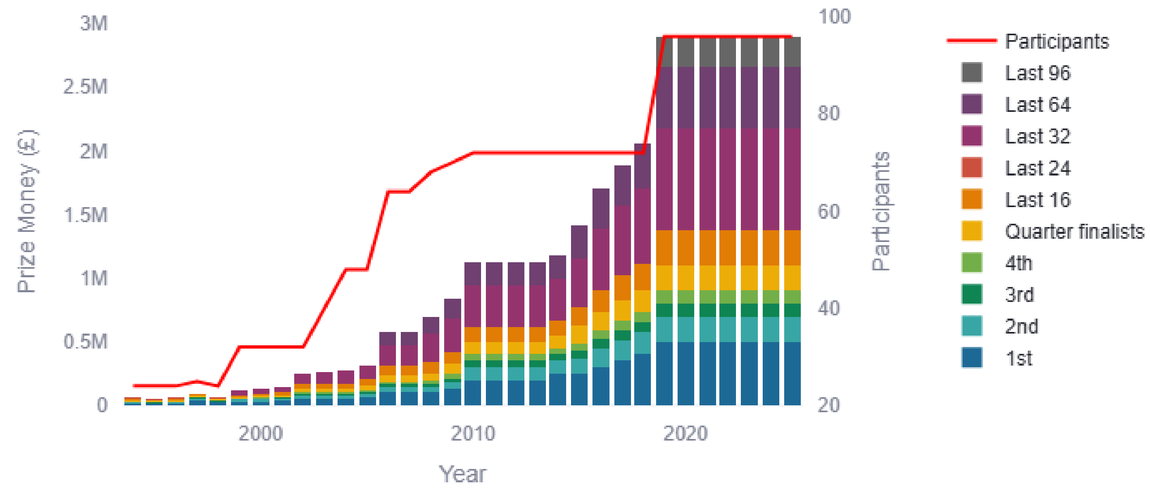
How do the averages of tournaments vary over time?

This question was asked, assuming that over time, the sport have reached a higher maturity level. Meaning the standards for the skill of players participating in the tournaments will rise with better strategies and better coaching. In this diagram, only the averages of the tournament's winners between 2000 and 2025 were looked at.
The World Championship scores demonstrate greater volatility, with notable peaks around 2010 and 2017-2018 where an average of 106 exceeded. All lines generally show an upward trend over time, which could also be seen because of the increasing regression line. The average scores generally range between 94 and 105. This validates the assumption the research came into existence with.



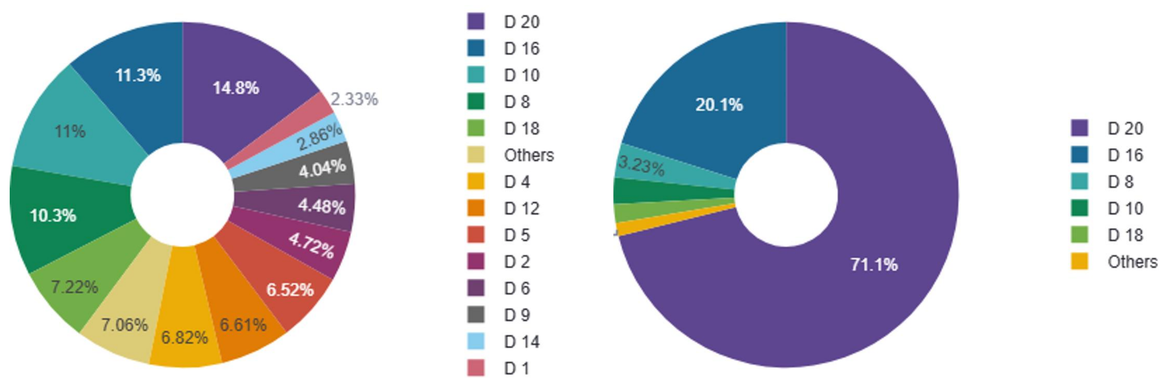
How does the price money and number of participants vary over time?

For this question, only the data for the PDC World Championship was analysed. In this graph, the prize money and number of participants is rising. The coloured stacked bars indicate prize allocation by tournament stage, while the red line shows the increasing participant numbers. There's a clear correlation between participant numbers and prize money: Major increases in participants typically coincided with significant prize money increases. The graphs show that while the total prize pool has increased, the distribution maintains a similar structure.



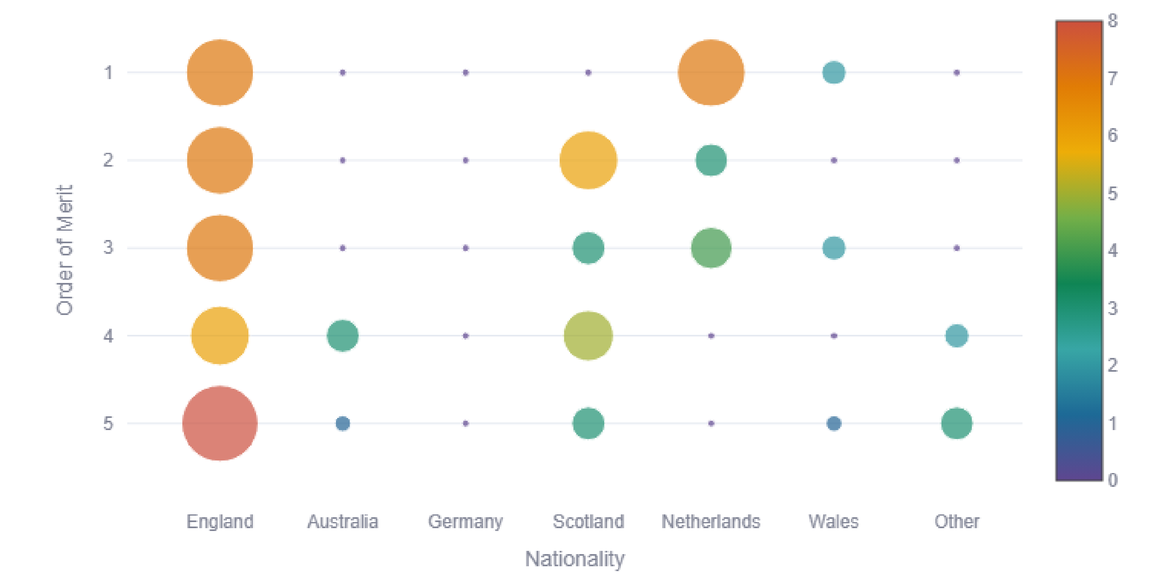
What are most popular double fields and their corresponding check out quotes?

A leg ends by hitting a double field, called a check out. With these come the statistics of the check out percentage. Other than in the check outs, doubles are usually not targeted. In this context, the distribution of throws on the different double fields was analysed.
The upper pie chart shows that D20, D16, D10, and D8 are the most popular double fields. This makes sense, since if a throw on D20 or D16 fails, the player can target the double field with the value half as big as D10 or D8 to finish the leg with a check out. This can also be seen in the lower chart, where the most player's frequently thrown double field is the D20 for 71.1% and the D16 for 20.1%.



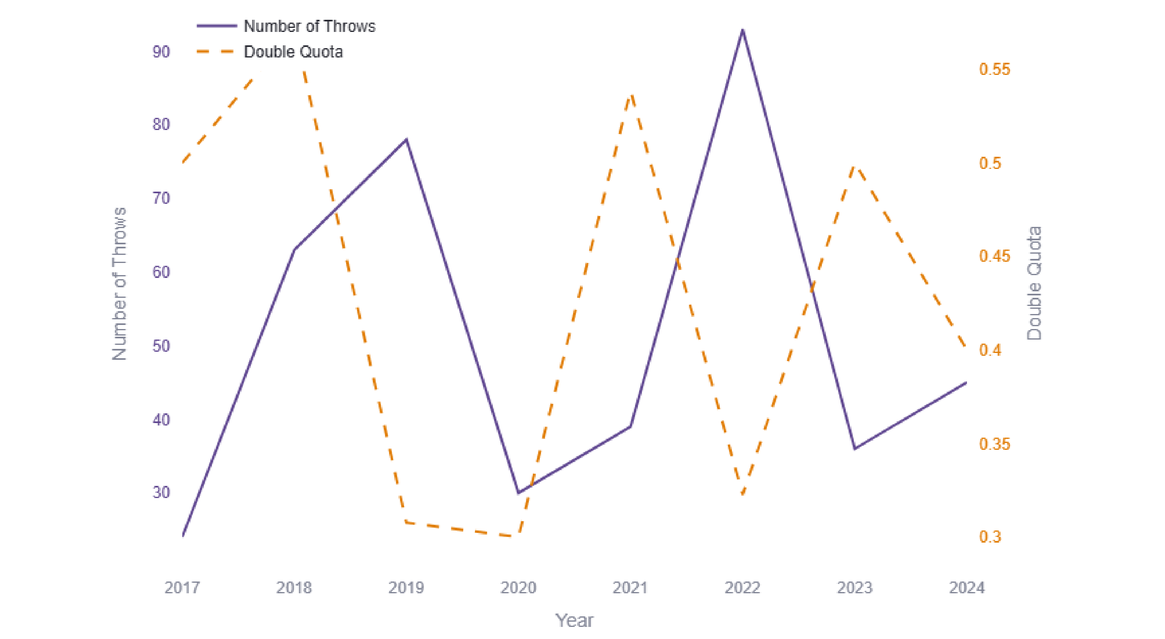
How does nationality effect the rankings?

The graph below shows the player's nationality demographics related to their rankings. The bigger the bubble, the more players belong to that demographic. As we seen, there are a lot of English players which are also very high ranked, suggesting that the sport might be better developed there.
It is also seen that there is one player in the Netherlands (Michael van Gerwen) which dominated the darts for a few years and is now also on a high-ranking position. Other countries like Germany, Japan, and Australia have representation, but their players are less frequent and frequent lower ranks.



How does the performance of individual players change over time?

The graphs below shows the player Rob Cross' performance from 2017 to 2024. On the left is the number of throws and check out percentage respective of the double 1. This player is a good example of the phenomenon, where their check out percentage drops and therefore the number of throws on this double drops in a time-shifted manner, too. This is because the player noticed that this double field does not work that well for them, so they adjust their throws so that they didn't always end up with the previously bad double at the end of a leg anymore. On the right is the gradual average of Rob Cross.



Bibliography

https://app.dartsorakel.com/
https://mastercaller.com/tournaments/pdc-world-championship/
https://www.dartn.de/Dart-Profis
https://de.wikipedia.org/wiki/Professional_Darts_Corporation
https://de.wikipedia.org/wiki/PDC_World_Darts_Championship
https://www.flashscore.de/