**Qatar World Cup Data**

Joseph Brown | 19334845 | 17/12/2022

Tools/technology used:

The programming language used to create the visualisation for this assignment was Python. Python is a popular language for data visualisation because it has a wide variety of libraries, including matplotlib, seaborn and plotly, that make it easy to create plots and graphs. It also has a very large and active community of developers and users, meaning that the resources available online are vast. (Ravikiran, 2022)

Plotly is one of Python’s many graphing libraries, but it offers something that no other graphing library does; interactivity. This allows users to interact with graphs on display, allowing for a better storytelling experience, with out-of-the-box features such as zoom, scroll and pan. It is also optimised for displaying graphs in a browser, which lends itself to creating an interactive dashboard.

Unlike traditional visualization tools like Tableau, Plotly allows users to fully customize their plots, including the number of graphs displayed at a time, the interactive tools like dropdown menus, titles, axis, and data sources for each plot. Plotly is based on Pandas, a data analysis library, so users can perform complex transformations on their data before visualizing it. To create plots, Plotly uses three main objects: data, layout, and figure. The data object is usually a Pandas data frame and can be fully customized. The layout object defines the titles, axis labels, and other details of the plot. The figure object defines how the data should be plotted according to the layout.

In addition to being able to create the same types of plots and charts as Matplotlib and Seaborn, Plotly offers a wide range of additional graph types, including statistical charts like parallel categories and probability tree plots, scientific charts like network graphs and radar charts, financial charts for time-series analysis like candlesticks and bullet charts, and geological maps and 3D plots. In addition to being used for data visualization, Plotly is also frequently used for machine learning, and it offers a variety of AI and ML charts. When used for creating dashboards, Plotly is often supported by the Dash python framework. (Leong, 2021)

Python Dash is a framework for building web applications with Python, and was used with plotly to create the dashboard in this assignment. It is built on top of JavaScript and it is designed to be used with the Flask web framework. One of the main features of Dash is that it allows users to create interactive, data-driven applications with little or no JavaScript knowledge. Dash uses a declarative syntax, which means that users define the layout and behaviour of their application using Python code, rather than writing HTML, CSS, or JavaScript. This makes it easy for Python developers to build web applications without needing to learn web development technologies. (Kelley, 2017)

Pandas is a Python library that is used for data manipulation and analysis. It provides data structures for efficiently storing large amounts of data, as well as tools for working with that data in a variety of ways. One of the main data structures in Pandas is the DataFrame, which is a two-dimensional table of data with rows and columns. DataFrames can be created from a variety of sources, including CSV files, Excel sheets, and databases. Pandas is optimised to work with the plotly library, so all the data that was read by the program was then converted to a dataframe for ease of use within the code.

The data set:

The data set used for this assignment was the ‘Fifa World Cup 2022: Complete Dataset’, found on Kaggle. The dataset includes information about all the matches in the 2022 Qatar FIFA World Cup. The data includes the scores and the football teams involved, as well as various statistics for each match, such as the number of assists, possession, crosses, red and yellow cards, passes, fouls, attempts, switches of play, offsides, and the number of times different areas of the pitch were crossed.

The data set was of the type table, and included eighty eight different attributes of various types:

* Categorical Data:
  + The names of the teams playing in each match was included in the data set. This attribute can be classed as categorical, since there is no implicit ordering.
  + Similarly, the stage at which each match took place is given in the data set, and is also catergorical.
* Quantitative – Discrete:
  + The ranking of the team from the data set can be thought of as discrete as there is very definitely an order to the data.
* Quantitative – Continuous:
  + Match stats, such as possession, goals scored, free kicks etc. can be classed as continuous data, since they can be arithmetically compared, and there exists an absolute zero

The variety of data types present in the data set, together with the volume of data, provides enough complexity that an interesting and engaging visualisation is warranted. (Iron486, 2022)

Tasks supported

The main task of this visualisation was to analyse the given data in an attempt to discover the underlying trends and traits that cause a team to succeed in the World Cup. This data visualisation may be used by teams and coaches to examine what needs to be done to improve or may simply be enjoyed by an average soccer fan.

One of the task actions was to analyse trends; this was done by consuming the dataset taken from Kaggle and visualising it to discover trends. The bottom right graph in the visualisation, for example, looks at how various stats such as possession, goals scored etc. affected a teams overall ranking in the world cup, and a trendline shows the overall trend. The top left graph, shows in the form of a bar chart, the ranking of teams in the world cup based on various stats.

Encoding channels and idioms

Colour, position, and size were the three visual encoding channels that are used to convey information in these data visualisations. These channels are effective because they allow users to quickly and easily understand and interpret data, making it easier to extract insights and draw conclusions.

Colour is a powerful visual encoding channel because it is highly salient and can be used to highlight important information. Different colours can be used to represent different categories or values in a visualization, making it easy to distinguish between different data points. For example, the top right graph of the dashboard, colour is used to distinguish between the team values and the average values; the same id achieved in the bottom left graph.

Position is another effective visual encoding channel because it allows users to easily compare the relative sizes or values of different data points. For example, in the line chart, position allows the reader to understand the effect of a given stat on the ranking.

Size is a third visual encoding channel that can be used to convey information in a visualization. Larger sizes can be used to represent larger values or to highlight important data points, while smaller sizes can be used to represent smaller values or to de-emphasize less important data points. For example, in a bar chart, the size of the bar was used to indicate the magnitude of each data point.

The idioms used in these visualisations include line charts and bar charts, combined with interactive elements which allow the user to easily see a variety of different data.

Line charts are effective at showing trends, and they are particularly well-suited for visualizing data in which there is a direct link between two attributes. They can be used to show the changes in a single variable or to compare the changes in multiple variables. This is effective in the case of the country ranking vs. metric graph in the bottom right of the dashboard, as there is clearly a trend between some of the metrics and the country ranking

Bar charts are effective at comparing the relative sizes or values of different data points. They are particularly useful for comparing data points within a single category or for comparing data across multiple categories. Bar charts are easy to read and interpret, and they can be used to effectively convey information about the relationships between different data points, which is why they were used in this visualisation. Two of the charts in this graph are showing the relationship/comparison between two data sets, for which the bar chart is ideal. (Sisense Team, n.d.)

Novelty of visualisation

Although the idioms used and encoding channels used are not particularly novel, the way in which the statistics are displayed relative to each other on the dashboard, the manner in which a user can interact with the data and the novelty of that data set itself – its only a few days old at time of writing – means that the visualisation as a whole is a moderately novel approach to the task.

The visualisation is complex in that it is entirely interactive. Almost any statistic that a user could want about the Qatar 2022 Fifa World Cup can be examined on the dashboard, while still keeping the visualisations concise as to not over stimulate the viewer.

Critical analysis

Although the encoding channels and visual idioms used are not the most complex, the dashboard that has been created makes the data accessible to most viewers. This is ideal, as the data presented could be of real use to those managing soccer clubs or those who just enjoy soccer, and the visualisation holds no value if it is not accessible to those who need/want it.

# References

ActiveState. (2022, August 9). *What Is Pandas In Python? Everything You Need To Know*. Retrieved from ActiveState: https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/

Iron486. (2022, December 17). *Fifa World Cup 2022: Complete Dataset*. Retrieved from Kaggle: https://www.kaggle.com/datasets/die9origephit/fifa-world-cup-2022-complete-dataset

Kelley, K. (2017, June 21). *Introducing Dash*. Retrieved from Medium: https://medium.com/plotly/introducing-dash-5ecf7191b503

Leong, N. (2021, July 27). *4 Reasons Why Plotly Is The Best Visualization Library*. Retrieved from Towards Data Science: https://towardsdatascience.com/4-reasons-why-plotly-is-the-best-visualization-library-18c27de05b95

Ravikiran, A. S. (2022, October 25). *A Complete Guide to Data Visualization in Python With Libraries, Chart, Graphs & More*. Retrieved from simplilearn: https://www.simplilearn.com/tutorials/python-tutorial/data-visualization-in-python#:~:text=Python%20offers%20several%20plotting%20libraries,most%20simple%20and%20effective%20way.

Sisense Team. (n.d.). *13 Powerful Ways to Visualize Your Data (with Examples)*. Retrieved from Sisense: https://www.sisense.com/blog/10-useful-ways-visualize-data-examples/