[Shared with UNSW DataSoc] UNSW DataSoc x Atlassian 2021 Datathon Overview

This page outlines the key dates and details of UNSW DataSoc x Atlassian Datathon 2021.

Please note this page is only shared with the Committee of UNSW DataSoc and the participants of UNSW DataSoc x Atlassian 2021 Datathon.

Case Brief

Problem Statement

You are a Data Scientist working for FIFA and you are working on a project to create awareness and engagement for the upcoming 2022 FIFA World Cup in Qatar.

As part of this project, you will have the choice to tackle 1 or both of the following tasks:

1. Identify (predict) the 32 international teams you believe will qualify for the World Cup
2. Prepare a short infographic that uses data insights and visualisations to generate engagement, interest and sponsorship deals. Be creative!

Note that your audience will include technical colleagues and you should ensure that you have the appropriate analytical rigor to your results.

Note:

- Infographic isn't limited to one type of media, can be a video, poster, series of posters etc.
- Think about the project goals, how will the infographic help you generate awareness and engagement?
- Find interesting datasets and most importantly, *tell a story!*

There will be 3 prizes:

- Best Task 1 (prediction of 32 teams - judged by methodology and rigor)
- Best Task 2 (infographic - combination of analytics + creativity)
- Best Overall (combination of Task 1 and Task 2)

You can do one or both of the tasks if you prefer - You will only be eligible for the overall task if you submit both.

Available Data

Below is a list of data resources you can use to come up with a solution to this problem.

Dataset	Description	Link
International match data	• International soccer matches from 1872 to 2021. • 1 match is 1 row in this dataset	https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017?select=results.csv
Player match information	• Match level data at a player level • Large SQLITE dataset • [Guide on how to access and use sqlite]	https://www.kaggle.com/hugomathien/soccer
European match data	• Match level data for European leagues	https://github.com/jalapic/engsoccerdata
Links to many different data sources	• Compilation of football data links	https://github.com/jokecamp/jokecamp.com/blob/master/_posts/2014-03-08-guide-to-football-and-soccer-data-and-apis.markdown
International team information	• Detailed information about each international team including their current full squad • Use in data analysis will involve data scraping / collection which may be time consuming	https://au.soccerway.com/teams/belgium/belgium/281/

Steps you will most likely need to undertake for the 32 team prediction

Other than the core task of predicting the 32 teams that will make it to the 2022 World Cup, there are some other preparation and presentation tasks that you will (probably) need to undertake. Note - not all of these are compulsory!

1. Identify and understand how the rest of the qualification process will work	Qualification for the World cup is a complicated process due to the fact that there are differences between regions. There are 6 different footballing federations that conduct their own qualification process and there are inter-continental play-off matches involved as well. The following Wikipedia page outlines the process https://en.wikipedia.org/wiki/2022_FIFA_World_Cup_qualification We will also include some extra links in the additional resources section to give you a place to start in your research / ideation about how to tackle this problem.
2. Complete the core match dataset	The core dataset of matches played with result has only been updated until 31 Jul 2021 and has an additional 2 months worth of games to be added. Find the list of these games and the additional information as required
3. Find additional datasets that add flavour / value to your analysis	While we have included the core match win/loss data to help you in this task, there are many different approaches you can take that will differentiate you from the competition. One of these ways is by incorporating extra datasets. Some ideas: • Player information (i.e. from the leagues they play in) • Stadium information (i.e. home court advantage / weather etc)
4. Visualisation and presentation	Your ability to communicate your results in a clear and engaging manner will be an important part of success in the judging rounds. Think about how to display your 32 teams, the measure of certainty for each team and all the factors that went into your results! [TIPS FOR PRESENTATION] Remember, there isn't going to be an "answer sheet" - the list of 32 doesn't matter nearly as much as how you got there!

FAQs

Question	Answer
Why are we required to do the research into how the process works as well as data collection and preparation?	One of the primary goals of this hackathon is to provide you with an experience that is both engaging and realistic to working as a Data scientist / analyst. A large portion of a data professionals job is to familiarise themselves with context and prepare / collect data.
Examples of some Infographics	To be added by @ Shahin Elliin

Additional Resources

Resource	Link	Description
ESPN article on how qualification works	https://www.espn.com.au/football/fifa-world-cup/story/3860182/2022-world-cup-how-qualifying-works-around-the-world	Another way the process is explained.
Another way the process is explained.	e.g. https://en.wikipedia.org/wiki/Jordan_national_football_team	Time consuming data collection but each teams wikipedia page does have a list of their current squad

Program Schedule

Key Dates	Time	Program	Zoom Meeting Link / Slack Channels

20 Sep 2021	5:30pm - 7pm AEST	Kick-off Workshop • Kick off by UNSW DataSoc • Introduction to UNSW DataSoc by UNSW DataSoc • Introduction to Atlassian by @ Joyce Lai • The Objective of this Datathon @ Joyce Lai • Announcement of Case Brief & Evaluation Criteria @ Shahin Elliin • Q&A by UNSW DataSoc • Wrap up by UNSW DataSoc	https://www.google.com/url?q=https://atlassian.zoom.us/j/97935399548?pwd%3DQ1V4Z204aDRKamxvckdSVEFXcU1oQT09&sa=D&source=calendar&ust=1631944959331280&usg=AOvVaw36Bmzy9zWP3ZKDJoUI191S
21 Sep 2021	5:30pm - 7pm AEST	Data 101 Workshop • Data Insights Process (CRISP-DM) • Business Context • Demo of Data Exploration (EDA) - Colab, and Tableau (as a BI tool option) • Group Practice • Q&A	https://www.google.com/url?q=https://atlassian.zoom.us/j/97524061075?pwd%3DY2JyL2Y1ZFhRKzJsRFh4bUhqRlJsdz09&sa=D&source=calendar&ust=1631944959331280&usg=AOvVaw18E9_-zinMzkzGvj2I0L-r
21 Sep 2021 - 24 Sep 2021	-	Async Mentoring Sessions	Slack Channel Invites to be shared by UNSW DataSoc
25 Sep 2021	10am - 4pm AEST	Datathon 2021 • 10am | Keynote • 10:30am | Heats Round • 11:45am | Heats Round - Adjudicator Debrief • 12:45pm | Lunch Break • 2pm | Finals Round • 3pm | Finals Round - Adjudicator Debrief • 3:30pm | Announcement of Winner & Thank You	https://www.google.com/url?q=https://atlassian.zoom.us/j/92826558143?pwd%3DY2kvUGhXNlNpNk1LL2tRRHlpd0M5QT09&sa=D&source=calendar&ust=1631944980227710&usg=AOvVaw2RCF1EVLBkJjrS34wS-3uU