

JSC370 - Final Project Report

Search Trends in 2022 on Sports Event

Wise Chua

Introduction

Every year, sports events are one of the most important activities that people from all around the world are interested in. Major sport events are extremely competitive as this is the time for athletes to demonstrate their skills, determination, and performance in the sports. Moreover, people usually interact with some kinds of sport throughout their childhood life and during school activities. This will develop their interest and hobbies in different kinds of sports. It is also important for a person to do some form of sports to maintain health and wellness such as running, playing basketball and soccer.

Due to the impact in COVID-19, most events that were originally held in 2020 and 2021 were postponed as travel restrictions were in place and athletes' health safety. Some of the several major athletic events postponed in these two years were Tokyo 2020 Summer Olympic and Paralympic Games, UEFA Euro 2020, 2021 Southeast Asian Games, and COPA America 2020. With social distancing measures lifted and travel restrictions lifted, most sports events had resumed in 2022.

In recent years, the popularity surge in esports had become a different kind of sport due to the highly competitive events and increased popularity of video gaming due to advancement in technology, it had become a new kind of sport. In 2018, esports had been recognized as a sports event in the 2018 Indonesia Asian Games. Moreover, the Olympic Committee is going to hold an Olympic Esports Series in 2023 for supporting the development of esports in the past few years and possibly becoming an official sport event in the Olympic Games.

With the increase usage in technologies due to COVID-19 as people are required to stay at home during the lockdown period, people started to gain real time information through search engines, online newspapers and social media platforms. In particular, sport and esports events have become more popular and well known over these years due to technological advancement such as live time information on these activities can be delivered worldwide efficiently.

Google search engine is one of the most widely used search engines for academically, general purpose and any information that people want to know. One of the reasons that Google has become so powerful and highly popular is due to their strong search algorithms which can provide information that users would need to find immediately. Moreover, Google also provides different services such as image search, news information, and live time map information which are extremely useful designs for people to use in daily life.

For this research, we would want to find the different searching trends for sports and esports events that happened in 2022. The dataset will be generated by using the Google Trends R package which can provide a scaled range from 0 to 100 based on a topic's proportion to all searches on all topics (By Google Documentation) that people had searched on Google Search Engine. In each generated trends queries, it also provides the proportion searches in each region, countries and related queries to the given keywords.

Methods

Data Collection

The data for this research project is acquired through the gtrends function from the gtrendsR package in R. This function can find the number of a keyword proportion to all searches on all topics at the same time. It can also provide the proportion of the searches on each keyword with each country and cities over the specified time period. It will also provide similar keywords to the searched keywords for number of hits. Due to the limitation of the function, it can at most take in 5 keywords each time which means it will generate a different data set. Hence, several keywords will need to be formed with different function calls and we will need to store them as different variables. For each event, we will use 5 keywords that can represent the sport event and use the other top 10 related queries from the results of the keywords.

Since we are comparing over 5 different major sports events, the data set will be combined together after extracting the information from the results of the function call. As our main research question is comparing these sports events within 2022, we will limit our searches to only search queries data that occurred in 2022. We will generate with date restrictions to our search trends for the whole 2022 year (2022-01-01 to 2022-12-31). Note that in each result of the function call, it will have 4 data frames in total where each of them are search hit over time, country, city and major designated market area (DMA) worldwide.

Note that since Google trends does not store information for locations that have extremely low search hits, the results from each enquiry of Gtrends package may not always contain the same countries, cities, and major designated market areas in each query to Google trends.

The unit for the search hits are calculated by a sample of Google searches. Google uses re-sampling data on that sample and by having the resampling data, it become the representative of all Google searches. Thereafter, it normalized the search data by location, time, and between different terms. A more detail on the search results can be found at Google Trends Help.

The table below shows the keywords that we used for each sport events. It has the initial 5 keywords that we started with and the remaining 10 keywords based on the related queries.

Table 1: Keywords to Search on Google Trends for Each Sports Events

2022 Qatar World Cup 5 Search Keywords	world cup 2022	world cup	2022 world cup	qatar world cup	fifa world cup
2022 Qatar World Cup 10 Related Search Words	fifa	fifa world cup 2022	fifa 2022	2022 world cup schedule	qatar 2022 world cup
	world cup 2022 fixtures	world cup football 2022	world cup 2022 live	world cup live	world cup table 2022
2022 Beijing Olympic Games 5 Search Keywords	2022 winter olympic	olympic games beijing 2022	olympic games 2022	beijing olympics 2022	winter olympic games
2022 Beijing Olympic Games 10 Related Search Words	olympics 2022	olympics	winter games 2022	winter olympics	olympic games
	2022 winter olympic games	winter olympic games 2022	olympic medals 2022	winter olympic medal 2022	beijing 2022

2022 Birmingham Commonwealth Games 5 Search Keywords	2022 birmingham commonwealth games	2022 birmingham commonwealth	commonwealth games	commonwealth games 2022	2022 commonwealth games
2022 Birmingham Commonwealth Games 10 Related Search Words	birmingham commonwealth games 2022	birmingham games	birmingham games 2022	commonwealth games 2022 india	commonwealth games 2022 table
	birmingham 2022 commonwealth games india	commonwealth games medal tally 2022	commonwealth games medal tally	commonwealth games medal table	commonwealth games 2022 medals
2022 League of Legend World Championship 5 Search Keywords	2022 league of legends world championship	2022 league of legends	2022 lol world championship	2022 lol worlds	lol worlds championship
2022 League of Legend World Championship 10 Related Search Words	league of legends worlds	league of legends worlds 2022	worlds 2022	lol world championship 2022	league of legends world championship 2022 tickets
	league of legends champions	league of legends world championship prize pool	league of legends world championship 2022 prize pool	league of legends world championship 2022 schedule	worlds league of legends
2022 The International Dota 2 5 Search Keywords	the international dota 2 championship	2022 the international dota 2 championship	dota 2 championship	the international dota 2	dota the international 2022
2022 The International Dota 2 10 Related Search Words	dota 2 world championship	dota 2 championship 2022	dota 2 world championship 2022	dota 2 championship prize	the international 2022 dota 2
	dota 2 international 2022	the international 2022	international dota 2 2022	dota 2 liquipedia	the international dota 2 championship

From table 1 above, we can see that the related words are often different keywords that people will search for that specific event. For example, people will search “fifa 2022” instead of “world cup” as some people used the organization that is holding this event. Similarly, people will search for “winter games 2022” instead of olympics as these are words that leads people to think of that event. For sport events, people are curious on the rank of the leading in each sport, but in the esports activity, other noticing keywords that appears in each event was the prize pool for the esports championship games.

Data Wrangling

The data cleaning for each sport event will be relatively similar because the generated results have fixed columns and information inside are the same. We will first extract the **interest over time** dataframe from each queries and combine them into a single dataframe. Afterwards, we will be changing the date columns from character values into Date by using `as.Date` in R. Next, we will place all values in the hits column (which indicate the proportion of search on the topic compared to all samples of all searches) that has <1 into 0. This is to have all values that has less than 1% of search in the week to be considered as 0 as these are too small. Finally, we will group the data by their dates and use the summarise function to compute the mean of hits in each week with missing values removed.

For the remaining 3 data frames where there are categories by country, city, and major DMA worldwide, the numbers of hits column in these data frames is not separated by weeks, as they are the highest score that occurred during the search period (2022-01-01 to 2022-12-31). We will group them by either their country, city and major DMA depending on the data frame and compute the average of hits. We will handle the same as previous for hits that have <1 and remove NA values.

Notice that when using the `gtrends` function call, sometimes the type for the number of hits are not in integer and stored as character type. Hence, we will need to convert it to integer before merging the data from each function call.

Finally, we will sort the remaining 3 data frames by the highest search queries in their region for better demonstration of the data frame.

Below are the first six rows after data cleaning for each sports of the remaining 3 data frames:

Table 2: First six rows of the World Cup Data by City, Country and Designated Market Area (DMA)

City	Hits	Countries	Hits	DMA	Hits
Chingjui Khullen	100	Solomon Islands	100.00000	Miami-Ft. Lauderdale FL	88.53333
Kalimpong	100	Qatar	81.50000	Washington DC (Hagerstown MD)	75.66667
Lusail	100	Papua New Guinea	74.00000	New York NY	71.66667
Shani Arjun	100	Bangladesh	68.78571	Dallas-Ft. Worth TX	68.33333
Tirur	98	Maldives	64.77778	West Palm Beach-Ft. Pierce FL	66.76923
Bandar	96	Belize	55.00000	Houston TX	66.75000

From table 2, it is showing the top 6 location of search hits in different location categories on world cup. We can see that cities that are appearing in the top 6 are locations that has extremely low population except Lusail which is one of the city that is in the hosting country. In the countries categories, most of them are relatively small countries and the hosting country does have an extremely high search hits. For designated market area, we can see the high search hits are location that had high population rate in the city and it an important city for the United States.

Table 3: First six rows of the Olympic Games Data by City, Country and Designated Market Area

City	Hits	Countries	Hits	DMA	Hits
Beijing	100.0	Mongolia	76.57143	Marquette MI	94.00000
North Bay	100.0	China	62.75000	Burlington VT-Plattsburgh NY	93.14286
Riga	100.0	Canada	62.69231	Zanesville OH	85.50000
Ljubljana	97.0	North Macedonia	56.14286	Alpena MI	78.00000
Cordova Bay	90.5	Slovenia	53.66667	Duluth MN-Superior WI	77.20000
Canberra	90.0	Armenia	52.33333	Ottumwa IA-Kirksville MO	70.00000

From table 3 above, we can see that it follows a similar trend that the location where the event was held has a 100 search hits, however, due to google access limitations in China, the countries search hit is to be

lower than Mongolia. Notice that Canada has a similar search hits result as China which is reasonable as Winter Olympics is one of the international events that Canada usually does well. From the DMA location, it seems that the top scores are locations in United states that have low population and it does not follows the pattern with previous table.

Table 4: First six rows of the Commonwealth Games Data by City, Country and Designated Market Area.

City	Hits	Countries	Hits	DMA	Hits
Marston Green	100	New Zealand	87.90909	Roanoke-Lynchburg VA	100.0
Nelson	100	India	85.00000	Charlottesville VA	85.0
Uttarpara Kotrung	100	Papua New Guinea	56.00000	Peoria-Bloomington IL	69.0
Napier	98	United Kingdom	40.66667	Washington DC (Hagerstown MD)	68.5
Upper Hutt	88	Australia	36.72727	New York NY	63.4
Invercargill	82	Kenya	35.44444	Dallas-Ft. Worth TX	61.0

From table 4, we can see that the cities that have the top six are locations that are small town all around the world. The countries that have the top search hits produce a different results comparing to previous as most of the countries here are big countries that are well known. Although the event was held in Brimingham, United Kingdom, the search hits was only 40.6. For the DMA locations, we can see that some location are big cities in United States which follows a similar trend to previous result.

Table 5: First six rows of the League of Legends Games Data by City, Country and Designated Market Area.

City	Hits	Countries	Hits	DMA	Hits
Bay Lake	100.00000	Iceland	100.00000	Lafayette IN	100.00000
Davis	100.00000	North Macedonia	72.00000	Madison WI	81.75000
Kuala Lumpur	100.00000	Denmark	67.83333	New York NY	79.66667
San Mateo	82.00000	Singapore	65.50000	Ft. Smith-Fayetteville-Springdale-Rogers AR	73.00000
Berkeley	70.00000	Czechia	59.42857	Huntsville-Decatur (Florence) AL	69.00000
Irvine	67.33333	Sweden	57.28571	Minneapolis-St. Paul MN	66.80000

From table 5, cities that are having high search hits are mostly United states cities where there is a famous college. This could due to League of Legends is highly popular between students as this is a video game. From the countries, it is relatively different as most countries are in Europe and the championship events were actually hold in United States. From the DMA section, the locations seems to varies between big cities and small cities. It also shows a similar result with League of Legend games but not with other sports event.

Table 6: First six rows of the Dota Games Data by City, Country and Designated Market Area.

City	Hits	Countries	Hits	DMA	Hits
Huancan	98.00000	Sierra Leone	100.00000	Miami-Ft. Lauderdale FL	94.00000
Juliaca	97.00000	Mongolia	89.60000	Washington DC (Hagerstown MD)	90.33333
Huancayo	94.83333	Peru	75.85714	Orlando-Daytona Beach-Melbourne FL	90.00000
Puno	92.00000	Myanmar (Burma)	56.40000	New York NY	82.40000
Miraflores District	88.00000	Singapore	50.71429	Detroit MI	82.00000
Huamancaca	86.00000	North Macedonia	43.00000	Austin TX	78.00000

From table 6, the top cities that has the highest search hits are all small cities, which comparing to previous table, it shows differently. For the countries, it is countries that has a relatively low population and the location where Dota 2 championship were held was not in it. For DMA, most area were big cities in United States with a very high population comparing to other cities in

Data Merging

We will merge all cities data with different sport events into one data by their city with using full join since the number of cities in each data could be different. We will also do this to the countries data and designated market area data. With merging all 5 sports into 1 data frame, we will need to rename the columns to the games name.

Data Exploration

For data exploration, we will use barchart graph to see how the general trends goes in each sport events and compare between different sports events by different city, country and designated market area. We will only choose locations that have all 5 data.

City

Kindly refer to the website on Figure 1.

Figure 1 is the barplot shows cities that have at least 2 search hits in the 5 sport events. I limited to at least 2 because most of the cities have either only have 1 search hits. We can see that most of the cities from the bar plot are considered as one of the major city in the province. We can see that World Cup and Olympic Games has very high search trends as it appears in all cities but commonwealth games does not. A very surprising result was League of Legend where it appears in most of the cities. By filtering the events, we can compares the search hits for an isolated event. As the default is showing cities that has two results in the sport hits, there are other cities that are not shown.

Country

Kindly refer to the website on Figure 2.

Figure 2 is barplot that only shows countries that have all 5 sport events and search queries. From the country search trend, we can see most of the countries have at least some search trend in the World Cup event. Some countries have relatively low search in Olympic games. For League of Legends, we can see most of the countries have very high search hits compared to Dota 2. We can filter out some events and only comparing between 1 to 4 events instead of 5. It shows the sums of search hits in total for the sport events which also can determine whether there are some countries that have extremely high search hits in all different sport games. For the Commonwealth Games, it seems that some countries have more searches in the Commonwealth compared to the Olympics but would need further investigation.

Designated Market Area

Kindly refer to the website on Figure 3.

Figure 3 is a barplot that shows the search hits for designated market area with all 5 sports events. It can produce a comparison between area whether the search trends are similar and whether there is an event is more outstanding in some areas. Also, it can tell which area have the highest search trends among all events. However, from the search trend plot above, we can see most events have a relatively consistent search hits among all different games. One of the most surprising results was that Dota 2 hits were relatively higher than League of Legends but this trend was not found in previous barplot.

Results

Preliminary Results

Below are the summary results for the search trends for each sports event from 2022/01/01 to 2022/12/31 worldwide.

Table 7: Summary Results for All Sport Events Search Trend with Time

Summary Results for World Cup Search Trend

	date	hits
	Min. :2022-01-02	Min. : 0.3333
	1st Qu.:2022-04-01	1st Qu.: 0.4000
	Median :2022-06-29	Median : 0.7333
	Mean :2022-06-29	Mean : 4.5449
	3rd Qu.:2022-09-26	3rd Qu.: 1.4667
	Max. :2022-12-25	Max. :49.8000

Summary Results for Olympic Games Search Trend

	date	hits
	Min. :2022-01-02	Min. : 0.4000
	1st Qu.:2022-04-01	1st Qu.: 0.4667
	Median :2022-06-29	Median : 0.6000
	Mean :2022-06-29	Mean : 2.5782
	3rd Qu.:2022-09-26	3rd Qu.: 0.9500
	Max. :2022-12-25	Max. :42.8000

Summary Results for Commonwealth Games Search Trend

	date	hits
	Min. :2022-01-02	Min. : 0.0000
	1st Qu.:2022-04-01	1st Qu.: 0.0000
	Median :2022-06-29	Median : 0.0000
	Mean :2022-06-29	Mean : 1.7705
	3rd Qu.:2022-09-26	3rd Qu.: 0.1333
	Max. :2022-12-25	Max. :47.3333

Summary Results for League of Legends Search Trend with Time

	date	hits
	Min. :2022-01-02	Min. : 2.667
	1st Qu.:2022-04-01	1st Qu.: 3.267
	Median :2022-06-29	Median : 3.833
	Mean :2022-06-29	Mean : 6.500
	3rd Qu.:2022-09-26	3rd Qu.: 4.783
	Max. :2022-12-25	Max. :25.533

Summary Results for League of Legends Search Trend with Time

	date	hits
	Min. :2022-01-02	Min. : 1.000
	1st Qu.:2022-04-01	1st Qu.: 1.830
	Median :2022-06-29	Median : 2.267
	Mean :2022-06-29	Mean : 3.892
	3rd Qu.:2022-09-26	3rd Qu.: 3.818
	Max. :2022-12-25	Max. :31.800

Table 7 show the results that for most of the mean are relatively low because most of the high search queries occurred during the event. We can also see that the maximum search queries for each event is approximately 40% to 50% which is not really that high. Although League of Legends has the lowest highest hits, their mean is the highest compared to others. This could represent that the search trend hits numbers were more spread out throughout 2022.

Research Question

Our main question was to find whether there is any trend in the search hit of sports events during 2022.

Analysis with Search Hits over time

Transition Over Time Graph - Animation

Kindly refer to the website on Figure 4.

Figure 4 above shows how the search hits transition for each sport events in 2022.

Transition Over Time Graph - Interactive

Kindly refer to the website on Figure 5.

Figure 5 above shows the different search hits in weekly with the date as the start of the week and the number represents the search hits for that event. Each line represents different sports event and we can choose only certain events if we would want to compare a particular set of sport events.

We can see that most events have a higher peak when the event is on-going. However, for the World Cup, we can see the search trend did not go down significantly compared to others and there are higher search trends before the events start compared to other events. For both Commonwealth Games and Olympic games, the search trend went down extremely fast after the event had concluded. For both the esports (Dota 2 and League of Legends), we can see there are still some slightly search queries being requested throughout the year.

Conclusions

In this project, we were to find the relationship of search trends in people for different sport events that occurred in 2022. Based on the line graphs and summary statistics table, we can see that the search queries increase significantly during the event and the search queries will decrease significantly. During other periods of time than the event time, there are relatively low search queries and most events will have no search queries. For the World Cup, there are more search queries even after the event compared to others. This could be due to the World Cup is an event that everyone around the world will be following compared to other events

that have lower popularity. For esports events, we can see that even it is not during the championship game period, the search trend will stay fluctuating throughout the year, this could be due to event structure are based on the teams performance throughout the year.

An interesting result from the peak of each events was the sports events will have a higher peak than esports events. We can see that from the summary statistics table that their maximum value was higher than esports while their median are lower than esports events. Notice that for all sports events, the third quartile have a lower values than esports event which also represents most of the data are very small.

Limitations

There are several limitations in this research. First of all, there are several empty values in the city, country and designated market area data frame as Google trends does not capture all search queries in those regions. Google only keep partial of these data and will if the search hits in the region are too low or none, it will not appeared in the generated results. These empty values result in the data in different cities, countries, and designated market area being very limited as we can't compare different regions very well. We limited the queries from January 2022 to December 2022, but some of the events happened at the beginning of 2022 and some events occurred during the end of 2022. This could highly affect the comparison of the queries between before and after the events. Finally, the units in the search hits are relatively wierd scale, based on Google documentation, the search hits are based on topic's proportion to all searches on all topics. The sample searching that Google used might not be representative. Some other limitations could be the keywords that we chose are not representative to the event.

Future Steps

My plan for the future steps is to incorporate different small events that occurred in the Games and we look deeper into those trends. Also, I would want to change the time trends and look at different timing of each Sports event based on one week before the event starts to one week after the event. Furthermore, look at other sport events that also occurred in 2022 and consider whether events that had been postponed by COVID-19 have different trends with events that were originally planned for 2022. If it is possible, I would want to try comparing with other search engines such as Yahoo or Bing on the same events to see whether there is any difference.