Instagram Activity Analysis: GDPR, JSON & Me: Final Report

Ryan Traviss

TBC 2020/2021

1 Abstract

This gets done last...

Contents

List of Tables

List of Figures

2 Introduction

2.1 Motivation

I want to discover what Instagram could learn about me if they analysed my personal data in depth. I am specifically interested in what my usage trends reveal from the data I know they hold. For example, could they work out when I catch the train to and from school purely from the timestamps on when I "like" photos? Or can they determine when I have just got back from a Scout camp based on a spike in the number of "followers" I have? Or are my usage habits just too random?

I am quite comfortable with my usage and feel like Instagram adds value to my life by allowing me to keep in touch with friends both new and old. Society often has privacy concerns about the data social media sites such as Instagram collect and a by-product of carrying out the above is seeing all the data they hold on me. Is this what I expected to find?

I propose to focus this project on Instagram as I know I use it regularly and have done since mid-2017 when I created my account. I considered Facebook but I mostly only browse less frequently and don't interact with posts much. I use Snapchat more to communicate with individuals directly similar to WhatsApp or text messages instead of sharing posts and seeing what others share like I do on Instagram.

2.2 Background

2.2.1 General Data Protection Regulation (GDPR)

The General Data Protection Regulation [1] was written to give clear legal rights to people whose data is being used by other people or organisations[2]. It is a European Union(EU) Regulation that came into force on the 25th May 2018 and is complemented by the Data Protection Act 2018[3] in the UK. EU regulation automatically comes into legal force in all EU member states[4] which the UK was at the time. After the UK left the EU on the 31 January 2020 at 11:00 pm (often colloquially referred to as "Brexit"), GDPR was transferred to UK law under the European Union (Withdrawal) Act 2018 and it is now called UK GDPR to differentiate it from its EU counterpart. The UK GDPR was also amended for clarity by removing mentions of EU institutions and replacing "Union" with "United Kingdom"[5][6]. Before GDPR, across the EU, data protection laws varied and the penalty here in the UK was capped at £500,000[7] which was quite low. This research is only possible through the use of Subject Access Requests (SAR), commonly referred to as the "Right of Access". Basically, anyone can request an organisation that holds data on them to give them a copy of it all which they must usually do within a month. There is no specific format the request must be made under [8]. I made a request each to Facebook, Instagram and Google on the 27th April 2020 which they complied with the same day. I further explain why I am only using my data due to GDPR in section 2.4.

The Information Commissioner's Office (ICO) enforces GDPR and the Data Protection Act 2018 within the UK. They describe their themselves as [9]:

> The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

The ICO is an executive non-departmental public body [10] run by the Information Commissioner [11]. Article 4 of GDPR defines many relevant key terms including:

> 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

Therefore, anonymised data, where the data subject cannot be identified, is not personal data [12] and so not covered by GDPR. This may allow further research using other peoples Instagram data.

Instagram has both Terms of Use [13] and a Data Policy [14] which together gives information on the personal data they process.

2.2.2 Instagram

Instagram[15] is a social media platform, owned by Facebook[16], where you follow other people and share pictures with your followers to keep in touch. Social media sites often use lots of terminology but

sometimes in conflicting ways and therefore follows a list of how I would use these key words to describe Instagram in particular.

- A “user” is simply a person who has an Instagram account. Eg. Ryan Traviss
- A “username” identifies an account to other users and always begins with an @ symbol. Eg. @exetermathssch Note: ”usernames” can be only be changed twice within a 14 day period.
- A user can create a “post” which must include between 1 & 10 pictures or videos. I made 50 posts during the period under analysis.
- A post has a “caption” which is added by the user making the post, usually a poor attempt at a pun, and can be edited after posting. Eg. “We’re all one big IKEA family!”
- Based on privacy settings, other users can then add “comments” to a post which consist of text or “like” the post by clicking a heart which shows they enjoyed it. “comments” can also be “liked”. I have liked 8242 posts and 223 comments up until I downloaded my data on 2020-04-27.
- A user can “follow” another user meaning they see their posts if their privacy settings allow or they accept a “follow request” and these are referred to as “connections” in the data. I have 361 followers and I follow 537 accounts at time of writing.
- A user can “message” another user which is similar to sending a text message but can include pictures and other media.

2.2.3 Java Script Object Notation (JSON)

In response to my requests, I received a nested file structure consisting of JavaScript Object Notation files [17][18]. A JSON file consists of objects which are name/value pairs inside curly brackets such as ”language_code”: ”en” where the name is on the left and the value is on the right. There are also arrays which are an ordered list of values such as [”2020-03-24T10:31:20+00:00”, ”exetermathssch”] which is an array, as it is enclosed by square brackets, with 2 values inside. I further explain how I have manipulated the data in this format in section 3.3.1.

2.2.4 Dates & Times

A key part of this project is manipulating dates and times of which there are many formats used to express them[19]. All dates and times I will be working with are in the format ISO 8601 [20] outlines. “2020-03-24T10:31:20+00:00” is an example which contains the date in the format year followed by month followed by day. This is followed by a T to represent the time portion and then hh:mm:ss for hours (in 24 hour format), minutes and seconds and then the time zone relative to Universal Coordinated Time (UCT) is the last part. This helpfully allows dates and times to be sorted alphabetically and be in order unlike the date format we normally use in the UK which is DD/MM/YYYY. Under this format, days of the week are numbered 1 through 7 where 1 is Monday and 7 is Sunday. Therefore the days of the week 1-5 can be referred to as weekdays and 6&7 collectively as the weekend.

There are no formal agreed definitions of many time terms so for ease of understanding below is a list of terms and how I will use them in this document.

- AM - The hours between 00 and 11 inclusive.
- Morning - See AM above.
- Noon/Midday - hour 12 under ISO 8601.
- PM - The hours between 12 and 23 inclusive.
- Afternoon - The hours between 12 and 17 inclusive, ie. 12:30pm (half-past twelve) is in the afternoon.
- Evening - The hours between 18 and 23 inclusive, ie. 23:59 is in the evening but 00:01 is in the morning.

I have not defined any terms related to the night as these are almost exclusively set by sunrise and sunset.

2.2.5 Metadata

Metadata can be colloquially defined as "data about data". When using Instagram, the main data may be a picture I have posted and the corresponding metadata includes when I uploaded that picture alongside its size and colour depth. This research relies on two types of metadata: datetimes and usernames.

Almost all actions on Instagram which are available in the data download include the datetime of when the action was carried out.

Many actions involve another user and their username is also stored such as in the likes data, where the username is of whose post was liked.

2.2.6 Friendship

explain friendship here

2.2.7 Sets

A set is an unordered collection of unique elements[21]. Curly brackets are used to show the members of a set and capital letters typically are used to identify sets.

For example, A is a set containing 1, 2 and 3 is shown by:

$$A = \{1, 2, 3\}$$

The cardinality (size) of a set can be shown with vertical bars:

$$|A| = 3$$

The union of two sets is simply all the members of both sets:

$$B = \{2, 4, 6\}$$

$$A \cup B = \{1, 2, 3, 4, 6\}$$

The intersection of two sets is the set containing the members of both sets:

$$A \cap B = \{2\}$$

The difference of two sets is the elements in the 1st set which are not in the 2nd:

$$A - B = \{1, 3\}$$

$$B - A = \{4, 6\}$$

A subset is a set which contains only elements of another set:

$$C = \{1, 2\}$$

$$C \subseteq A$$

In fact C is a proper subset of A as there is at least one element in A which is not in C:

$$C \subset A$$

These can be shown using venn diagrams...

2.3 Tasks

2.3.1 Original Tasks

Extracted from the Research Proposal[22]

1. To investigate my usage trends of Instagram using "likes"
   (a) By time
      i. On an hourly basis to identify peak times
      ii. On a daily basis to identify differences in weekend/weekday usage and on events such as Scout camps
      iii. On a monthly basis to identify wider trends such as between term-time and holiday usage
   (b) By user's post I've liked: Is it possible to work out who my best friends are by this method? Or does this just reflect users who post more?
2. Aggregate Measure: To create an aggregate measure of usage based on "posts", "likes", "messages" and "comments"
   (a) Investigate trends over time and contrast to using only "likes" as above
3. Identify Best Friends: Investigate if you can work out who my best friends are through whose posts I've "commented" on
4. Compare Followers: To compare "followers" with events such as new schools and scout camps to determine the strength of this relation and thus if you could infer such events with only the "followers" data
5. Extension Graphs: Alternative methods of presenting the above information such as climate stripes for usage

2.3.2 New Tasks

The following tasks have been added over time after the research proposal was finalised so they can all be considered extension objectives.

6. Contact Instagram
   (a) Request any documentation explaining the data
   (b) Request the dates of when different features were first released
   (c) Request information on how long different types of data is retained for
   (d) Request data not included in the Data Download eg. Ad categories
   (e) Include Instagram's responses as an appendix to this report
7. Publish the code used for this project
   (a) Set up a GitHub repository
   (b) Research Code Maintainability
   (c) Declare a public API
   (d) Release versions according to Semantic Versioning 2.0.0 [23]
   (e) To add a page to my website (ryantraviss.github.io) explaining the project
   (f) Select an open source code license [24]
   (g) Potentially distribute this as a package
8. Automate the analysis as much as possible
   (a) Create a single subroutine that carries out all the all time analysis
   (b) Create a single subroutine that carries out all the yearly analysis
   (c) Include an option allowing all the analysis to be written to a file as well as/instead of being printed.

9. Create tools to assist future development
    (a) Make a function that returns all different time zones used in the data
    (b) Make a function that returns the oldest and newest piece of data to assist manual analysis
    (c) Make a function that returns the date and time the account was created on
    (d) Create a subroutine that maps the contents of the JSON files in a human readable format
10. Investigate using other people's Instagram data via the Data Download tool
    (a) Research carefully the definition of personal data
    (b) Consider if the downloaded data could be cleaned (eg. by removing usernames)
    (c) Alternatively, consider if people could use the automated tools to conduct the analysis themselves and send me the results (tables/graphs)
    (d) If possible, review this with the Data Protection Officer (DPO) of my school to gain a second opinion
    (e) Gain approval of my supervisor before commencing any research using other peoples data
11. Investigate using the Instagram API to conduct further analysis
    (a) Look into the instagram-scraper Python package [25]
    (b) Look into the Instaloader Python package [26]

2.4 Ethics & Bias

I plan on only using my own data in part to avoid ethical issues. I will not publish any messages I have sent/received and will be mindful of other people's reasonable expectation of privacy throughout. My Instagram username will not be published to preserve my privacy.

To prevent me altering my behaviour due to carrying out this research, I will work only using the data I downloaded on the 2020-04-27 prior to commencement of this project. At the moment, due to current restrictions in place to prevent the spread of COVID-19, I am also not meeting new people or going on Scout Camps[27] so my behaviour is clearly going to be different than it would usually be. I also recorded my best friends as set out in Section 3.5 before conducting the relevant data analysis.

I investigated using other people's data to allow comparisons to be made and to see if correlations that are true for me are also true for them. However under GDPR, my lawful basis for processing[28] would be consent meaning they could, at any point, ask me to delete their data under the "Right of Erasure"[29] which I would then be obligated to do or potentially face legal penalties which are quite high[30]. Furthermore, if I used data from my friends at school, they would be children whose rights are specifically highlighted in GDPR[31].

Notwithstanding, if the data is anonymised and so GDPR does not apply it would be a possibility to conduct parts of this research on other users data. Great care would be needed to ensure that the data does not identify them and that they have no influence on my findings especially if my friends contribute their Instagram data.

3 Methodology

The intention of the following section is that you can both understand the decisions I have made and could carry out similar research into your own data using the code in Appendix C.

3.1 Time Period of Analysis

My data ranges between when I created my account on 2017-07-19 & 2020-04-27 which is when I downloaded my data and started this project. The data for 2020-04-27 may be incomplete as I downloaded my data in the morning and I may have used Instagram later in the day. The effect of this should be negligible due to it making up such a small part of the 1014 days in total under analysis. 2020 was a leap year so comparing the month 2020-02 (February 2020) to any other February may lead to minor issues over the difference in the number of days. Overall, the only full years for which I have data are 2018 and 2019.

3.2 Time Zone

I checked the time zones using the subroutine below and ['+00:00'] was the only result. I am unsure of how Instagram handles time zones as it doesn't appear to change for British Summer Time (BST)[32]. I believe Instagram itself handles this and stores the datetime in terms of your current local time. I have not left the UK since the creation of my Instagram account so the effect of any timezone changes would be minimal ie. if I had been to Australia this could completely change the hours I like posts.

Figure 1: Timezone Test subroutine

```
def timezone_test(self):
    """
    Prints all of the timezones when I liked a post/comment.

    Returns
    -------
    None.

    """
    timezones = self.__data_time__(slice(10,11), "T", slice(19,26), return_ints=False)
    print(np.unique(timezones, return_counts=False))
```

3.3 Program Development

The program that provides the tools to conduct this analysis was written in *Python* [33] version 3.7 . I have used the Object Orientated Programming paradigm[34] to reduce code duplication and ensure future maintainability. I wrote the source code in the Spyder IDE[35] due to the wide variety of libraries included in the distribution.

3.3.1 Code Analysis

3.4 Events

To assist in the completion of task 1(a)ii, I have compiled a table of all the events I have been on where I have spent a night away from home. This list is not definitive as I have been unable to find information about some events especially those in 2017/2018. As seen in Appendix A Table 2, I have been very busy with a total of 75 nights spent other than at my home.

Whilst conducting the analysis I have checked the data for the period of the event with a day either side to look for a peak after I have returned especially if there was not WiFi wherever I was.

3.5 Best Friends

For tasks 1(b) & 3, I needed to determine a proper subset[1] of my friends to refer to as my best friends. Friendship is a two-way relationship[REF] however there is no reliable method to find out how they feel about me (and if there was would I want to know or would it simply damage my mental health) and thus all measures are based only on how I feel. Trust is the sole criterion I have used to classify my friends meaning that some friends I enjoy spending time with are not included. I imagined a variety of situations and thought who I would go to for advice or reassurance. Obviously, over the past 3 years my best friends will have changed and they will continue to change going forward so the list represents a snapshot in time from when it was written (2020-05-26).

After getting a list of 19 people, I then attempted to separate these into three broad groups labeled A, B and C[2] where A is those who I trust most and I would go to them with issues I wouldn't go to people in B about and so on. I have made no attempt to rank people within each group. I have put 3 people into the A category, 7 into the B category and 7 into the C category. I also decided I trusted two people less than those in category C as I used to know them better and they would have been in a category in the past. Therefore they were included erroneously in the original list and were removed leaving me with 17 best friends.[3]

[1] A set such that my best friends is not simply just all my friends so there are people who are my friends but not my best friends. $BestFriends \subset Friends \Rightarrow |BestFriends| < |Friends|$

[2] Expressed formally as: $A \subset BestFriends$, $B \subset BestFriends$ and $C \subset BestFriends$.

[3] Expressed formally as: $|A| = 3$, $|B| = 7$, $|C| = 7$, $|BestFriends| = |A \cup B \cup C| = 17$

This was all carried out before I saw any Instagram data to prevent me labelling people as best friends to make the data match my predictions(intentionally or otherwise[REF]).

As set out in Section 2.4, my own privacy is a key issue, therefore my best friend list was compiled on paper and will never be released. This brings up reproducibility issues however if you follow the method set out above to create your own best friend categories and then use your own Instagram data you can test my conclusions. This research is not going to be submitted to a peer-reviewed journal so any findings will likely just suggest further research that should be carried out. If asked by anyone, I will say you are on the list regardless of if this is true or not. The reason for this is to prevent the list growing longer solely due to peer pressure[REF].

talk about Dunbars number[REF] here. Include diagram if possible

For the purpose of this research, I have manually combined the like counts of users who have multiple accounts with different usernames. It is not uncommon to have a "private" account (also called "finsta" as in fake Instagram) of which you are more selective of your followers than your "main" account. I have not combined multiple accounts if one is clearly a business account or for a person's pet. If a username is a shared account belonging to multiple users I have taken no action but this will have an inconsequential effect on my results due to the low frequency of such accounts. I will identify using footnotes where this has been done on any tables/graphs.

For the purpose of referring to different members of each category they are labelled by the category letter and a number so A_1-A_3[4], B_1-B_7 & C_1-C_7 are my best friends. The number is simply in the order I wrote them down on paper, it is not a ranking within the category nor is it alphabetical or follows any other pattern.

3.6 Statistics

For all following equations: $x = x_1, x_2, \ldots, x_n$

The mean of a variable is equal to the sum of the elements divided by the number of elements as shown by:

$$\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{1}$$

The acronym SD refers to population standard deviation which is calculated using the following formula:

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2)}{n}} \tag{2}$$

show median represented by a tilde

3.7 Data Cleaning

The data x1 and x2 appears to be the same but this demonstrates the importance of cleaning data. In some contexts, such as how many posts have I liked per hour, the absence of any data should be interpreted as a zero. Whilst the sum of the x values is obviously unchanged, both the mean, median and measures of variation are altered sometimes significantly.

Table 1: Data Cleaning Demonstration

						$\sum x$	n	$\bar{x}$	$\tilde{x}$
x1	2	3	4			9	3	3	3
x2	2	3	4	0	0	9	5	$\frac{9}{5}$	2

3.8 Data Sources

3.8.1 Time

[4]Expressed formally as: $A = \{A_1, A_2, A_3\}$

Table 2: Time Data Sources Comparison

Data Source	Mode	Size
media_likes		8242
comment_likes		223
comments		285
stories		69
posts	single	50
	multiple	320
direct		174
messages		4120
message_likes		234
chaining_seen		1269
following		536
followers		362
	Default	14666

3.8.2 User

following table excludes my username so numbers may differ from above

Table 3: User Data Sources Comparison

Data Source	Mode	Size
media_likes		8208
comment_likes		223
comments	smart	284
	post	187
	reply	285[5]
chaining_seen		1269
messages	sender_individual	3613
	sender_group	14942
	sender_both	18555
	participants_individual	165
	participants_group	239
	participants_both	404
message_likes	my_messages_individual	172
	my_messages_group	130
	my_messages_both	302
	my_likes_individual	102
	my_messages_group	126
	my_messages_both	228
	my_messages_likes_both	530
	all_likes_individual	173
	all_likes_group	5960
	all_likes_both	6133
following		536
followers		362
	Default	14127

4 Results

4.1 Usage Trends using Likes by Time

8242 posts and 223 comments from 556 different users were liked during the period under analysis.

4.1.1 Hourly

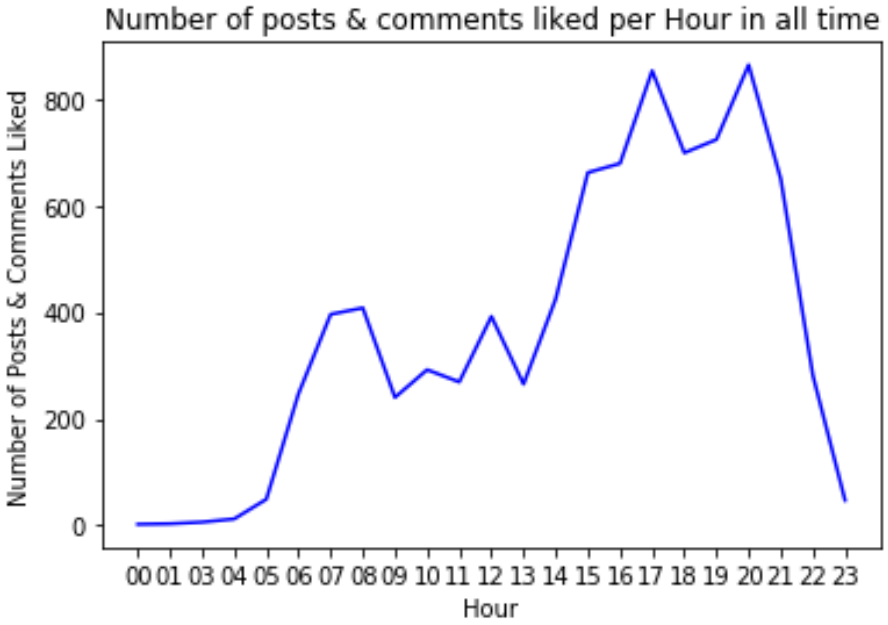
The average number of posts liked at each hour of the day was 352 but with such a large standard deviation of 282 indicates a high level of variability in the times when I use Instagram. A bot could continue posting at a consistent level[REF] including over the night which would lead to a low Standard deviation and a line graph appearing almost flat. Appendix A includes the full data as tables including a variety of summary statistics and Appendix B is for extra graphs that could not fit within the body of the report.

Figure 2 shows a graph of the total number of posts & comments liked at each hour between the creation of my account until I downloaded my data. Overall, the graph shows clear negative skew with bimodal peaks at 5pm and 8pm with above 850 likes. Very few posts were liked in the morning with none at 2am and the number of likes grows slowly until 5am when it increases to over 400 likes at both 7am and 8am. The number of posts liked then fluctuates until the afternoon where the hourly rate of likes is at roughly twice the morning level. Figure 3 shows the median time to like a post was 4pm with 50% of posts being liked between noon & 7pm. So few posts are liked between midnight and 2am they are considered outliers as they are more than 1.5x the interquartile range below the lower quartile at 12.

Now I will go through each year and attempt to identify trends in what time posts & comments have been liked and explain the causes.

In 2017, I liked 471 posts & comments which makes up a small proportion of the 8465 posts & comments liked over the period under analysis. My like distribution by hour is bimodal with the highest peak at noon with a slightly lower peak at 5pm. My usage is in an almost "M" shape with lows in the morning, 2pm and 10pm. From the boxplot, the range in hours that I liked posts & comments is less than in any other year.

Figure 2: Likes Hourly Line Graph All Time



In 2018, I liked 2027 posts & comments. The like distribution by hour has negative skew with the median hour being the latest of any year at 5pm. The overall shape is similar to the all time distribution. The interquartile range is the lowest out of any year due to the concentration of likes in the afternoon/evening.

In 2019, I liked 4579 posts & comments which is over half of my likes during the period of analysis. This year is the only where I have liked any posts at midnight or 3am

In 2020, I liked 1388 posts & comments.

days, months ect

4.1.2 Events

See Table 2

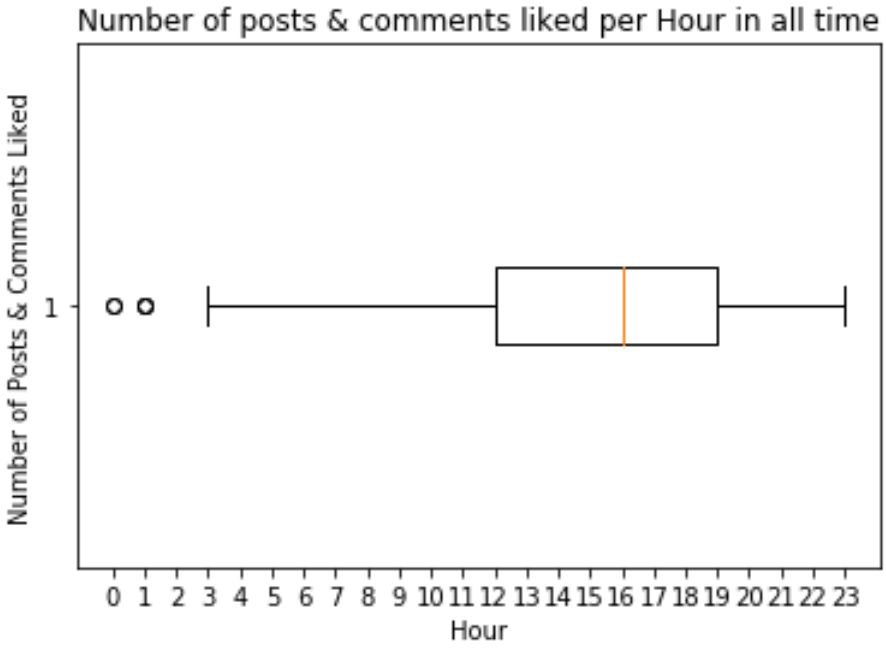
4.2 Aggregate Measure

4.3 Best Friends

As I have liked 8465 posts & comments from 556 different usernames the mean number of posts & comments liked per user is 15. Table 3 in Appendix A shows my the mean likes out of my seventeen "best friends" is 27 and thus I like my best friends posts more that the average user. This is also true for the median with 5 posts & comments liked for all users and 18 for my best friends. However, there is not a threshold above which all or even most of the users are my best friends. The top users two are actually not individuals' personal accounts eg. I have liked 528 posts from "@visitfaroeislands" during the period of analysis placing it in the top spot. This user is actually the national tourism agency[36] of the Faroe Islands and therefore post with a very high frequency. Also only 5.5% of the posts & comments I have liked were made by one of my best friends. This measure appears to be ineffective at determining who my best friends are.

Table 4 shows comments I have liked is a much better measure as 85/223 of the comments I have liked are by my best friends. The mean number of comments liked is 2.5x more for my best friends and the median is 5x greater. By ranking, my best friends take up first place and joint second, third and fourth places with non-best friends. However this method still isn't totally accurate as four of my best friends: A2, B3, C3 and C7 have no comments of theirs I have liked.

Figure 3: Likes Hourly Boxplot All Time



5 Conclusion

5.1 Review of my Data

5.2 Timeline Review

Figure 4: Original Timeline

Task	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
1(a) Likes by Time									
1(b) Likes by user's post									
2 Aggregate Measure									
3 Identify Best Friends									
4 Compare Followers									
5 Ext: Graphs									
Report									
Poster									
Presentation									

6 Evaluation

It would have benefited this research if I were to record hypothesis before commencing the data analysis. In my head, I expected certain patterns to appear such as my followers over time spiking when I joined Exeter Mathematics School (2019-09) however including this in the report after I had looked at such trends is of no benefit. I was too excited to get into writing code and getting results without the necessary preparation and background research.

7 Acknowledgements

I would like to thank...

References

[1] *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL.* URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN (visited on 04/27/2020).

[2] *Protection of personal data (from 2018).* URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum:310401_2 (visited on 08/03/2020).

[3] *Data Protection Act 2018 - PART 4 - CHAPTER 3 - Rights - Section 94.* URL: http://www.legislation.gov.uk/ukpga/2018/12/section/94 (visited on 04/27/2020).

[4] *Regulations, Directives and other acts.* URL: https://europa.eu/european-union/eu-law/legal-acts_en (visited on 08/03/2020).

[5] *The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019.* URL: https://www.legislation.gov.uk/uksi/2019/419/schedule/1/made (visited on 08/03/2020).

[6] *Keeling Schedules showing changes which would be affected by the Data Protection, Privacy and Electronic Communications (Amendments etc)(EU Exit) Regulations 2019 laid on 14 January 2019.* URL: https://www.gov.uk/government/publications/data-protection-law-eu-exit (visited on 08/03/2020).

[7] *Information Commissioner's guidance about the issue of monetary penalties prepared and issued under section 55C (1) of the Data Protection Act 1998.* URL: https://ico.org.uk/media/for-organisations/documents/1043720/ico-guidance-on-monetary-penalties.pdf (visited on 05/26/2020).

[8] *(Guide to the General Data Protection Regulation).* URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/ (visited on 04/27/2020).

[9] *Information Commissioner's Office.* URL: https://ico.org.uk/ (visited on 08/03/2020).

[10] *GOV.UK Information Commissioner's Office.* URL: https://www.gov.uk/government/organisations/information-commissioner-s-office (visited on 08/03/2020).

[11] *Elizabeth Denham CBE, Information Commissioner.* URL: https://ico.org.uk/about-the-ico/who-we-are/information-commissioner/ (visited on 08/03/2020).

[12] *Anonymisation:managing data protection risk code of practice.* URL: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf (visited on 08/08/2020).

[13] *Terms of Use.* URL: https://help.instagram.com/581066165581870 (visited on 07/27/2020).

[14] *Data Policy.* URL: https://help.instagram.com/155833707900388 (visited on 07/27/2020).

[15] *Instagram.* URL: https://www.instagram.com (visited on 07/15/2020).

[16] *Facebook.* URL: https://about.fb.com/company-info/ (visited on 07/15/2020).

[17] *Introducing JSON.* URL: https://www.json.org/json-en.html (visited on 05/02/2020).

[18] *The JavaScript Object Notation (JSON) Data Interchange Format.* URL: https://tools.ietf.org/html/rfc8259 (visited on 08/03/2020).

[19] *xkcd: ISO 8601.* URL: https://xkcd.com/1179/ (visited on 05/02/2020).

[20] *ISO 8601-1:2019 Date and time — Representations for information interchange — Part 1: Basic rules.* URL: https://www.iso.org/standard/70907.html (visited on 05/02/2020).

[21] Bob Reeves. *AQA A-level Computer Science.* Hodder Education, 2019, pp. 164–168.

[22] *Instagram Activity Analysis: GDPR, JSON & Me: Research Proposal.* May 27, 2020. URL: https://github.com/ryantraviss/instagram_analysis/blob/master/ResearchProposalFinal.pdf.

[23] *Semantic Versioning 2.0.0.* URL: https://semver.org/ (visited on 07/23/2020).

[24] *Choose an open source license.* URL: https://choosealicense.com/ (visited on 07/23/2020).

[25] *Instagram Scraper.* URL: https://github.com/arc298/instagram-scraper (visited on 07/27/2020).

[26] *Instaloader.* URL: https://instaloader.github.io/ (visited on 07/27/2020).

[27] *Coronavirus update*. URL: https://www.scouts.org.uk/news/2020/march/coronavirus-update/ (visited on 07/23/2020).

[28] *Lawful basis for processing*. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/ (visited on 05/17/2020).

[29] *Right to erasure*. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/ (visited on 05/17/2020).

[30] *Penalties*. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-law-enforcement-processing/penalties/ (visited on 05/17/2020).

[31] *Children and the GDPR*. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/children-and-the-gdpr/ (visited on 05/06/2020).

[32] *When do the clocks change?* URL: https://www.gov.uk/when-do-the-clocks-change (visited on 07/14/2020).

[33] *Python*. URL: https://www.python.org/ (visited on 05/02/2020).

[34] Peter Wegner. "Concepts and Paradigms of Object-Oriented Programming". In: *SIGPLAN OOPS Mess.* 1.1 (Aug. 1990), pp. 7–87. ISSN: 1055-6400. DOI: 10.1145/382192.383004. URL: https://doi.org/10.1145/382192.383004.

[35] *Spyder: The Scientific Python Development Environment — Documentation*. URL: https://docs.spyder-ide.org/ (visited on 05/05/2020).

[36] *Visit Faroe Islands*. URL: https://www.visitfaroeislands.com/ (visited on 07/15/2020).

A Tables

Table 4: Number of posts & comments liked by hour

Hour	2017[6]	2018	2019	2020[7]	all time
00	0	0	2	0	2
01	0	2	1	0	3
02	0	0	0	0	0
03	0	0	6	0	6
04	0	2	10	0	12
05	1	33	13	2	49
06	10	51	173	13	247
07	14	75	245	62	396
08	19	92	210	87	408
09	4	36	149	51	240
10	52	47	151	42	292
11	48	58	121	42	269
12	71	73	189	59	392
13	11	55	153	46	265
14	8	107	240	70	425
15	17	137	361	147	662
16	30	173	375	101	679
17	58	210	443	142	853
18	22	172	399	106	699
19	32	210	368	114	724
20	30	213	510	111	864
21	34	189	304	123	650
22	10	78	138	55	281
23	0	14	18	15	47
Total	471	2027	4579	1388	8465
Mean	19	84	190	57	352
Median	12.5	65.5	163	53	286.5
SD(3sf)[8]	20.3	72.8	154	47.8	282
Range	71	213	510	147	864

[6]This covers the period from when I created my account on 2017-07-19.
[7]This covers the period until I downloaded my data on 2020-04-27.
[8]This is the population standard deviation to three significant figures.

Table 5: Events

Event	Start Date	Finish Date	Total nights
Frosties	2018-01-27	2018-01-28	1
IMC 2018	2018-04-06	2018-04-14	8
Ten Tors 2018	2018-05-11	2018-05-13	2
Sidvale Groupcamp	2018-06-08	2018-06-10	2
Bronze DofE Qualifying	2018-07-12	2018-07-13	1
Chill Camp	2018-07-20	2018-07-22	2
RTTF4	2018-07-29	2018-08-05	7
Cornwall Family Holiday	2018-08-13	2018-08-16	3
EMS Summer School	2018-08-20	2018-08-22	2
Dartmoor Cycle ride with South Brent	2018-09-15	2018-09-16	1
Explorer Belt Workshop 1	2018-09-21	2018-09-22	1
Cremyll Challenge 2018	2018-10-05	2018-10-07	2
Explorer Belt Workshop 2	2018-11-02	2018-11-03	1
AMC 2018	2018-12-28	2019-01-05	8
Bristol Mini Explorer Belt	2019-02-22	2019-02-24	2
March March 2019	2019-03-01	2019-03-03	2
Silver DofE Training Weekend	2019-03-30	2019-03-31	1
Afterprom	2019-06-19	2019-06-20	1
Sidvale Summer Camp	2019-06-21	2019-06-23	2
Gilwell 24 2019	2019-07-12	2019-07-14	2
Silver DofE Practice	2019-08-03	2019-08-05	2
South Brent Summer Expedition	2019-08-09	2019-08-17	8
Silver DofE Qualifying	2019-09-21	2019-09-23	2
Cremyll Challenge 2019	2019-10-04	2019-10-06	2
St Andrews Uni visit	2019-10-21	2019-10-24	3
AMC 2019	2019-12-28	2020-01-04	7
Total			75

Table 6: Comparison of Best Friends by posts & comments liked

Best Friends Category	# of posts & comments liked	posts & comments liked rank
A_1[9]	13	66
A_2	3	76
A_3[4]	53	28
B_1[4]	41	40
B_2	20	59
B_3	0	NA
B_4	54	27
B_5[4]	17	62
B_6	18	61
B_7	97	13
C_1	8	71
C_2[4]	55	26
C_3	11	68
C_4	36	44
C_5	9	70
C_6	30	50
C_7	0	NA
Best Friends Total	465	
Mean	27	
Median	18	
SD(3sf)	25.2	
Range	97	
Overall Total	8465	
Mean	15	
Median	5	
SD(3sf)	36.6	
Range	527	

test

[9]Two usernames have been combined for this user.

Table 7: Comparison of Best Friends by comments liked

Best Friends Category	# of comments liked	comments liked rank
A_1	1	10
A_2	0	NA
A_3	7	4
B_1	7	4
B_2	10	2
B_3	0	NA
B_4	7	4
B_5	8	3
B_6	1	10
B_7	3	8
C_1	3	8
C_2	5	6
C_3	0	NA
C_4	18	1
C_5	5	6
C_6	10	2
C_7	0	NA
Best Friends Total	85	
Mean	5	
Median	5	
SD(3sf)	4.73	
Range	18	
Overall Total	223	
Mean	2	
Median	1	
SD(3sf)	2.65	
Range	18	

B Extra Graphs

C Code

The most up-to-date version of this code can be found at github.com/ryantraviss/instagram_analysis

Figure 5: Likes Hourly by year