

Social Physics On Instagram

Hendriks, Jandie

Witlox, Timothy
Theunissen, Ties

Kooijman, Jeroen
Melskens, Demian

Weijs, Thomas

November 2018

Contents

1	Introduction	1
1.1	Professional Task	1
1.2	Human Behavior	1
2	Components	1
2.1	Category	1
2.2	Location	1
3	Strategies	1
4	Research	2
4.1	What are the most visited locations based on Instagram posts?	2
4.2	What are the reasons for people to go to these places?	2
4.2.1	Most common hash-tags in Belgium	2
4.2.2	Most common hash-tags in France	3
4.2.3	Most common hash-tags in Germany	3
5	Conclusion	3
6	Abbreviations	4

1 Introduction

1.1 Professional Task

This document serves as a research paper for the social physics course. The goal is to take a closer look at human behavior within the scope of our professional task. The aforementioned professional task consists of an analysis of the social media service Instagram. Using a data-set with information about public posts (Global Positioning System (GPS), likes, hashtags, etc.) we try to answer two questions within the professional task. Human behavior plays an important role in answering these questions.

1. *Can we sort the posts into defined categories using machine learning?*
2. *What information can be gathered based on the given GPS data?*

1.2 Human Behavior

This paper will first look at the broad subject that is social physics to define a research question for this paper. Two directions are available since both of our professional task questions contain social physics aspects as seen in section 2. Since the category part of the project has not been developed (as the machine learning is more complex than data visualization) we chose the GPS side of the project to focus on. Our research question based on this is as follows: *"Why do people visit specific locations?"*.

2 Components

2.1 Category

All categories used are based on what people decide to share, which gives human behavior a significant role in our project. The Central Bureau of Statistics (CBS) has provided us with these categories based on a study done aimed at Twitter. These categories came out as the most prominent when the posts gathered from this social media service were categorized. These categories could be changed if the results point to different labels and directions. This means that the behavior of the people that share posts influences our categories. What people choose to share on Instagram makes up our data-set.

Since the Facebook leak last year some people might be more careful with sharing their data. Another side where human behavior is introduced is with advertisements. These are personalized based on the behavior of the user.

2.2 Location

Where people go, take pictures and post these pictures is of course also an important part of human behavior. This GPS data is interesting for the CBS as this gives an insight in the human behavior of the dutch people.

3 Strategies

The sources used are based on the Fontys Hogescholen ICT (FHICT) research framework (Triangulation: FIELD – LIBRARY – WORKSHOP – SHOWROOM – LAB).

Lab Trying out the different machine learning algorithms and visualizing the results.

Showroom We present results to each other within the professional task and give feedback.

Showroom We present our findings to our employer at CBS and use the feedback to further our project.

Library We use the internet to research the different algorithms available for machine learning.

4 Research

4.1 What are the most visited locations based on Instagram posts?

The Instagram data we gather contains latitude and longitude coordinates. With these features we can enrich the data with information about the country and city where the picture is taken. This way we can collect all the Instagram posts that are posted on a single city or country. Then these posts can be compared with each other to create insights about why people would like to go to a specific city or country.

The first step that needs to be taken is to add all the necessary information to the data-set. Because all the data is hosted on a central MongoDB cloud cluster this can be done without cleaning all the data again. First all the data that is not yet updated is retrieved in batches from the database. After that a Python library named "geopy" is used to retrieve the city and country names from the coordinates. The results that are retrieved from the geopy library are then checked for errors and updated in the database.

Once the database is updated with all the location information we can start doing research. This is done by downloading all the data from the database. This data is then analyzed to find out what the most occurring countries are. The most visited countries according to the data that was collected from Instagram are:

- | | | | |
|--------------------|--------------|-----------|------------|
| 1. The Netherlands | 2. Belgium | 3. France | 4. Germany |
| 5. USA | 6. Brazil | 7. Italy | 8. UK |
| 9. Spain | 10.Indonesia | | |

4.2 What are the reasons for people to go to these places?

The most visited countries according to the collected Instagram posts are now known. The next step is to find out why people go to these places. Because most people provide hash-tags that describe what they are posting it could be interesting to analyze those hash-tags for every country. If there is enough data for one country this research could go even further and finding the reasons why people go to specific towns or cities within a country would be possible.

The next step is to go through the previously found countries and analyze which hash-tags are the most common for a given country. This research will be done for the top 3 countries that are not The Netherlands. These countries are; Belgium, France and Germany.

4.2.1 Most common hash-tags in Belgium

- | | | | |
|------------|--------------|------------|---------------|
| 1. Deurne | 2. Antwerpen | 3. Belgium | 4. Maastricht |
| 5. Antwerp | 6. Belgie | 7. Carre | 8. Brussels |
| 9. Limburg | 10.Gent | | |

Looking at the most used hash-tags in Belgium, people are mostly interested in posting about the cities and towns they visit and less about other things.

4.2.2 Most common hash-tags in France

- | | | | |
|--------------|--------------|-------------|------------|
| 1. Carre | 2. Paris | 3. Zadkine | 4. Lisse |
| 5. Brilliant | 6. France | 7. Coiffure | 8. Cheveux |
| 9. Art | 10.Sculpture | | |

These are probably the best results. It's clear that people probably go to Paris because they want to photograph art, to be more specific art at the Zadkine museum which seems to be popular according to the hash-tags.

4.2.3 Most common hash-tags in Germany

- | | | | |
|----------------|--------------|-----------|----------|
| 1. Germany | 2. Amsterdam | 3. Berlin | 4. Heino |
| 5. Instagood | 6. Holland | 7. Travel | 8. Love |
| 9. Photography | 10.Nature | | |

The hash-tags posted in Germany are more random, these are also about cities like Berlin but also about Amsterdam which has little to do with Germany. Other tags that are common like travel, photography and nature do describe a bit better why people would like to post about Germany.

5 Conclusion

Looking at the results it is possible to find out why people go to specific countries or locations. But it is important to have lots of data. This because the results may seem random when looking at the countries themselves. It is possible that the results will be a lot more specific if the possibility exists to look at the cities or even districts within cities. But this was not possible with the currently collected data because of the amount of data. The obtained results are a summary of about 160 thousand cleaned Instagram posts.

6 Abbreviations

CBS Central Bureau of Statistics. 1, 2

FHICT Fontys Hogescholen ICT. 1

GPS Global Positioning System. 1