

University of Tartu
Faculty of Science and Technology
Institute of Computer Science

Jan Moppel

Welcome to my couch: Why some people attract more guests than others on Couchsurfing?

Master's Thesis (30 ECTS)
Software Engineering Curriculum

Supervisor:

Rajesh Sharma, PhD

Tartu 2020

Welcome to my couch: Why some people attract more guests than others on Couchsurfing?

Abstract: Couchsurfing is a social networking platform that helps travelers in finding a free couch (or place to stay). However, not everyone is lucky to find guests on this platform. It has been observed that some hosts have to put more effort to get a couch request from surfers, because of their low popularity. In this thesis, we collected public data of 47 564 hosts which span across 65 cities and 6 continents to understand the characteristics of popular hosts on Couchsurfing. This is the first quantitative research on this topic that uses a dataset of that size. It is important to note that we focused on the popularity with respect to hosts and not to guests. Our findings reveal that i) popular hosts have fewer friends than somewhat popular or unpopular ones; ii) hosts from Europe hosted surfers from more countries than hosts from other continents, and iii) popular hosts have a bigger percentage of positive personal references than somewhat popular or unpopular ones.

CERCS: P170 Computer science, numerical analysis, systems, control

Keywords: understanding popularity, Couchsurfing, social media analysis, exploratory analysis

Tere tulemast minu diivanile: Miks meelitavad mõned inimesed Couchsurfingul rohkem külalisi kui teised?

Lühikokkuvõte: Couchsurfing on suhtlusvõrgustik, mis aitab reisijatel leida tasuta ööbimiskoha. Paraku pole kõigil siiski õnne sellel platvormil külalisi leida. On täheldatud, et mõned võõrustajad peavad madala populaarsuse tõttu rohkem pingutama, et surfajatelt diivanipäring saada. Käesolevas töös koguti avalikke andmeid 47 564 võõrustaja kohta, mis hõlmavad 65 linna ja 6 maailmajagu, et mõista populaarsete Couchsurfingu võõrustajate omadusi. Tegemist on esimese selleteemalise kvantitatiivse uurimistööga, milles kasutatakse sellise suurusega andmestikku. Oluline on märkida, et keskendusime populaarsusele võõrustajate, mitte külaliste osas. Meie tulemused näitavad, et i) populaarsetel võõrustajatel on vähem sõpru kui mõnevõrra populaarsetel või ebapopulaarsetel võõrustajatel; ii) võõrustajad Euroopast võõrustasid surfajaid suuremast arvust riikidest, võrreldes võõrustajatega teistest maailmajagudest; iii) populaarsetel võõrustajatel on positiivsete isiklike viidete protsent suurem võrreldes mõnevõrra populaarsete või ebapopulaarsete võõrustajatega.

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Märksõnad: populaarsuse mõistmine, Couchsurfing, sotsiaalmeedia analüüs, uuriv analüüs

Contents

1	Introduction	4
2	Background	7
3	Related work	12
3.1	Popularity in real life	12
3.2	Popularity on the internet	13
3.3	Couchsurfing	15
4	Data description	17
4.1	Description	17
4.2	Data preparation	20
4.3	Identifying popular users	23
5	Methodology	27
6	Results	29
6.1	Macro analysis	29
6.2	Meso analysis	30
7	Conclusion	36
7.1	Limitations of this study	37
7.2	Future work	37
	References	38
	Licence	43

1 Introduction

Travel has always been a significant part of human nature. From the beginning of time people moved around to pursue different objectives. At first, the goals of these trips were very practical and straightforward: get food, escape the danger, find a place to sleep, etc. Slowly, with the development of humanity and technology, traveling evolved from just an activity to stay alive into something different. Religious travels, trades, explorations [1] were no longer necessities, but an idea and desire-driven actions that required relocation. And in this paper, the discussion will revolve around one of these forms - tourism.

Nowadays, it is one of the biggest industries in the world. In 2019, travel and tourism contributed approximately 2.893 trillion U.S. Dollars to the global economy, which is 31% of the total contribution to Gross Domestic Product (GDP) [2] [3] [4]. This includes different sectors of the industry: transportation, food & beverages, entertainment etc. [5] One of the most significant parts of it is accommodation. It varies from bed & breakfast to luxury hotels and resorts [6]. All of them are not free and require payments to stay in. But there is one particular alternative that stands out of the traditional hospitality options and its name is Couchsurfing [7].

Couchsurfing is a hospitality exchange web-platform, where users (guests) can travel (surf) to different destinations and stay at the other user's (host) place. The main difference between the usual accommodation options and this alternative, is that it is free. As a part of the gift economy [8], Couchsurfing doesn't allow hosts to charge guests for lodging [9]. On this platform travelers can discover new places with the locals and get free accommodation, while hosts can meet a lot of new people with different backgrounds and actively participate in the cultural exchange.

As a unique hospitality option, Couchsurfing remains a very popular topic in the scientific world. From the beginning of 2019, more than 1000 articles were written that explored the different aspects of the platform [10]. Considering the popularity of the service, it is no surprise that this subject interests the researchers from all around the world. By December 2019, Couchsurfing had a community of over 14 million people in more than 200,000 cities [11].

Throughout the years, researchers explored the user trust-building phenomenon on Couchsurfing through the perspective of local communities [12], differences in user behaviors on Couchsurfing and Airbnb [13] [14], the challenging nature of managing this non-profit hospitality exchange platform [15] and other topics that are interesting for people from different scientific fields.

Some works, were very close, to the subject that we try to explore in this thesis. In particular, works by Maura Rae Cherney from the Illinois State University [16] and Alexander Ronzhyn from the Universität Koblenz-Landau [17] [18]. Maura Rae Cherney's "Surf's Up: Communicative Aspects of Online Trust-Building Via Reducing Uncertainty Online in Couchsurfing" [16], gave insights into how hosts make a decision of whether to accept or reject surfers and which attributes they consider as relevant in making such a decision. But that study was about surfer's popularity and characteristics that contribute to it. Alexander Ronzhyn's articles "Online identity: constructing interpersonal trust and openness through participating in hospitality social networks" [17] and "Conveying the Message of Trust through Written Texts in CouchSurfing.org" [18], focused on exploring, which characteristics of both hosts' and surfers' personalities, made them gain more positive references and therefore more popularity. Although, these studies have been done through the text analysis of references that were left on hosts' and surfers' accounts. They didn't consider other factors that can also influence the user's popularity on the platform.

The goal of this thesis, however, is to explore the phenomenon of host's popularity on the platform. With the community so big, the guest on Couchsurfing is served with a lot of options where to stay. This inevitably created a situation, where some users are more popular than others, and thus, people who are looking for an accommodation often pick a popular one. Therefore, it became much harder for non-popular hosts to receive any couch request, and take part in this unique experience. Specifically, we investigated the following questions in this thesis:

1. **RQ1:** Is the popularity on the platform completely randomized or there is a correlation between certain attributes of the host's profile and popularity?
2. **RQ2:** What are the differences between popular and non-popular hosts?
3. **RQ3:** Are there any geographical differences in the hosts' popularity?

Despite all the previous research, this thesis still brings some new perspective on the platform, as all the questions listed above are still relevant. The main contribution of this thesis is large scale data analysis for understanding popularity on Couchsurfing. To the best of our knowledge, this is the first study that exclusively explores the hosts' popularity and considers all attributes that users publicly share on their accounts. This is a quantitative research that uses the anonymized public data of 47 564 hosts from 65 cities from six different regions/continents: Asia, Europe, North America, South America, Australia, and Africa. A study performed on a scale like this provides some insights into the subject of hosts' popularity that, potentially, can be interesting not only for the scientific community, but also for the Couchsurfing users.

Based on our study we found out that there are specific correlation patterns between popularity and hosts' profile features. Also, some of these characteristics demonstrate different behavior for hosts from distinct continents.

The rest of the thesis is organized as follows:

1. In Chapter 2, we describe the ecosystem of Couchsurfing, what this service offers, and how users interact on the platform.
2. In Chapter 3, we discuss related works that are relevant for this study: popularity, Couchsurfing, and social networks. We discuss articles that were previously written on these subjects and try to understand the current state of the art on them. By doing so, we narrow down the spectrum of things that should be researched in this work.
3. In Chapter 4, we describe the dataset collection step and available features of the data.
4. In Chapter 5, we describe the data preparation steps and approaches to data analysis.
5. In Chapter 6, we give an overview of the results by displaying different findings that were made throughout the research, with examples and plots to justify them.
6. In Chapter 7, we conclude the thesis with a discussion of results, limitations of this study and future directions.

2 Background

As was stated in the Chapter 1, Couchsurfing is a social networking platform that has the goal of providing hospitality exchange. It has all the usual features of social networking platforms:

1. People can register and customize their profiles: add personal information, pictures, interests;
2. Write private messages to others;
3. Send friend requests;
4. Find groups and events by interests etc.

However, Couchsurfing has specific functionalities that other social networks do not have, such as:

1. Users are divided by their status: hosts and guests. Guests are couch surfers, who are looking for a place where to stay. Hosts are users who provide that place;
2. Guests can send "Couch requests" to hosts;
3. Guests and hosts can leave feedback on their interaction in real life on their profiles. This is called a reference.

Usually, if a person wants to find a place where to stay via Couchsurfing, he searches for hosts in the needed location. The platform has over 14 million people in more than 200,000 cities by 2019 [11], so there is a high chance of having at least one host in the required location. Hosts are usually sorted by either "Best match" or hosting experience, which means the amount of hosted guests throughout the years. The example of searching view interface is on the Figure 1.

When guests spot a host that they would like to pick, they open the host's profile and send "Couch requests". The way user profiles look like on the web-site is displayed on the Figure 2. The request is a message that briefly introduces the guest and the dates on which he wants to stay. It can be seen on the Figure 3. If the host accepts the request, they can proceed with their virtual interaction, which eventually leads to the interaction in real-life.

After finishing couch surfing, guests and hosts leave references on the page for host or guest respectively [19]. References can be positive or negative/neutral, and usually describe the experience of the host/surfer, including their interaction, activities etc. Members can also leave their comments on them. There are 3 types of references:

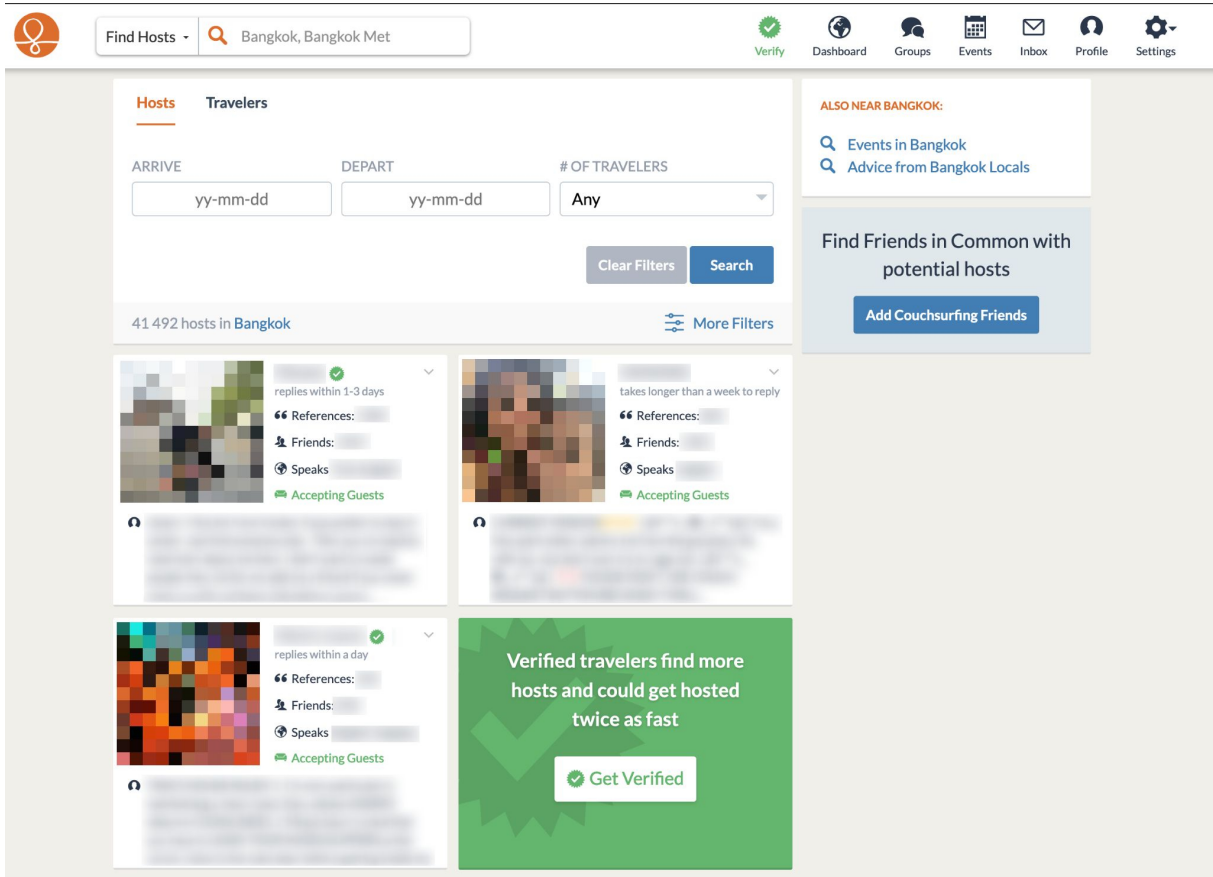


Figure 1: Example of host search

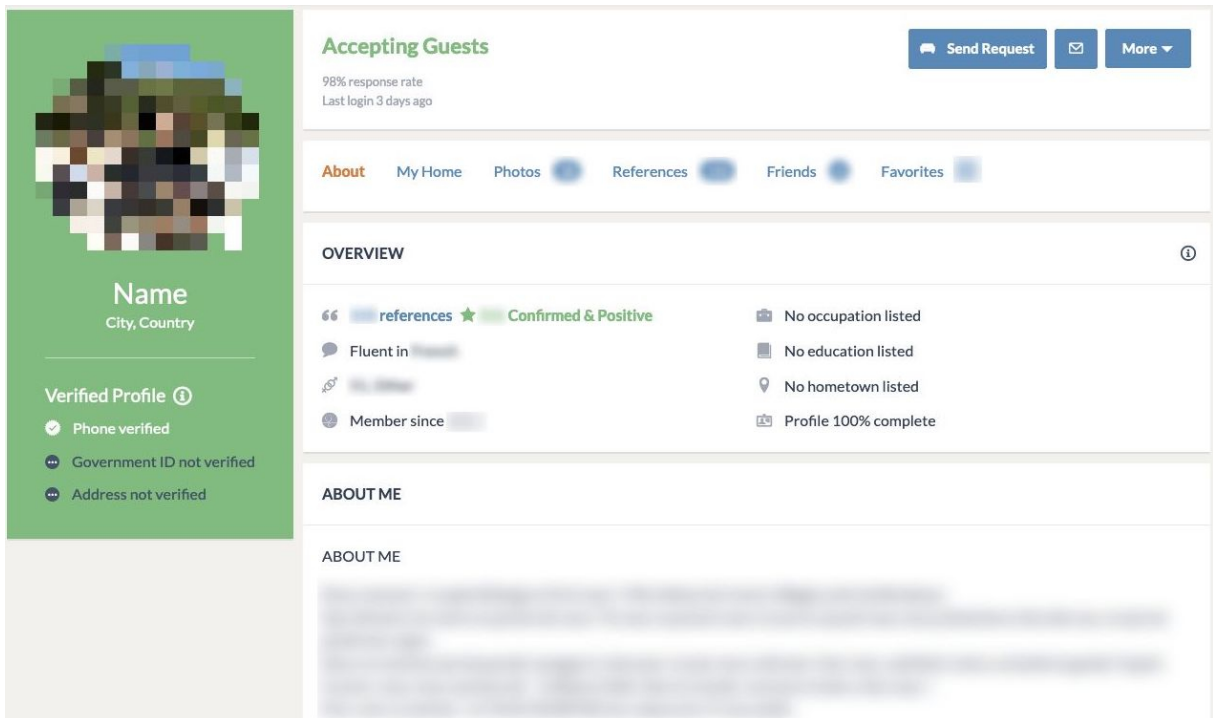


Figure 2: Example of the user profile

Figure 3: Example of the couch request

1. **Hosting experience references.** These are comments left by surfers after their stay at the host's place. Guests write them on the host's page. The example is shown in Figure 4.
2. **Surfing experience references.** These are comments left by hosts about their guests. They are left on the surfer's page. The example is shown in Figure 4.
3. **Personal experience references.** These are comments that can be left by anyone. They usually include all types of messages, from a usual reference to just saying "Hi!". The example is shown on Figure 5

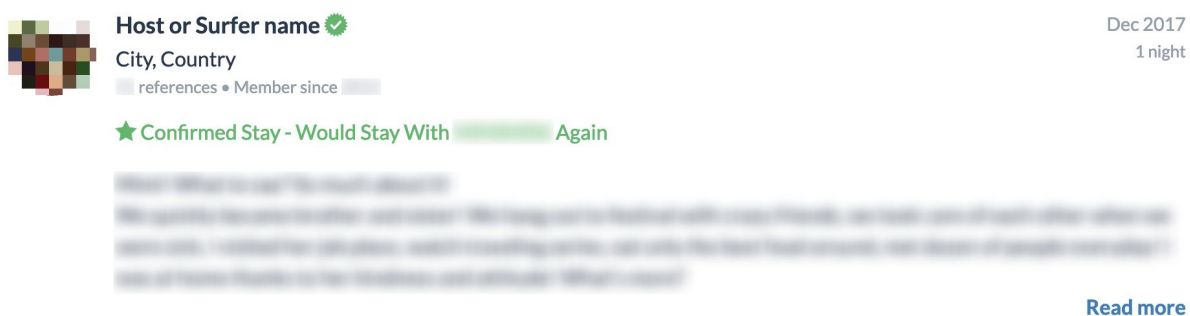


Figure 4: Example of the hosting and surfing reference

In addition to references, Couchsurfing members can also leave anonymous feedback after their



Author of Personal reference

City, Country

references • Member since [redacted]

Feb 2016

Positive

[Redacted text]


 [redacted]'s response
Comment on the reference

Figure 5: Example of the personal reference with a comment from host

stays [20]. Feedback is implemented through the tags system. Members can select positive or negative tags based on their experience. If selected tags are positive, then they become public, and other users can see them. However, if they are negative, then they remain private and are reviewed by the Safety Team. The examples of positive and negative tags are shown in Figures 6 and 7.

ENDORSEMENTS

It's great to hear that you'd recommend Sonja. Please select reasons why you'd recommend.
This feedback will be grouped anonymously on Sonja's profile.

thoughtful **friendly** fun clean
wanted to hang out okay with not hanging out
punctual followed house rules flexible knowledgeable

Figure 6: Example of the positive feedback tags [20]

PRIVATE FEEDBACK

We're sorry to hear that you wouldn't recommend Sonja. Please tell us your reasons why.
This feedback is private.

messy **unfriendly** didn't follow house rules not punctual
not flexible didn't want to hang out wanted to hang out too much
I felt unsafe Other

Figure 7: Example of the negative feedback tags [20]

Now, when we described how Couchsurfing works, we can move on to the next chapter that discusses the platform and topics to it from the scientific perspective.

3 Related work

In this chapter of the thesis, we examine works that revolve around the main topics discussed in this work - popularity, Couchsurfing, and methods that help to analyze user behavior on social networks in general. First, we explore the theme popularity in real life and on the internet - what does the term "popularity" mean, what are the differences between this concept for offline and online lives, how social media popularity differs from platform to platform, and how it can be measured. Second, we discuss papers that have provided insights into how Couchsurfing works from different perspectives - what are the benefits that make this service popular, what's the difference between couch surfing and usual travel styles and how the offline/online human interaction is organized on the platform.

3.1 Popularity in real life

According to the Cambridge Dictionary, the word "popularity" means "the fact that something or someone is liked, enjoyed, or supported by many people" [21]. This concept of one thing or person being more preferred than the other one, has always been a part of our lives. From an early age, we experience and learn about it [22] [23] [24], and even if we don't completely understand it, "popularity" already affects the way we see and think about the world. It greatly influences our behavior, relationships with others and overall, plays a huge role in our development as individuals.

When talking about person's popularity, according to W.M. Bukowski [25], it really depends on the 2 contradictory processes: *differentiation* and *assimilation*. Differentiation in this context, means the ability of the person to outstand the others and be different. Assimilation, on the other hand, means an individual's skill to immerse himself in the group by endorsing the group's values and traditions and keeping its dynamics. In other words, in order to become popular in a group of people, the person has to be a unique member of it (differentiation), but at the same time keep a strong connection with the goals and values of the group (assimilation).

However, when we discuss the subject of popularity, we often tend to mix 2 different terms. From the perspective of the layman, if someone is recognized and supported by many people on social media or in real life, he is a popular person, but there is a clear distinction between a *popular person* and an *elitist* that we need to establish [25]. This difference lies in the way a person becomes popular. The elitist is someone, who tends to concentrate only on differentiation from the group and ends up being too far away from the group's values and dynamics. For example, being an extremely wealthy individual or elite athlete, very often results in forgetting about the group's goals and, instead, focusing solely on individual goals [25] [26].

The popular person, on the other hand, differs himself from the rest of the group, but at the same time stays in it. By doing that, the group itself recognizes him as a "key member" and makes him popular. The popular person will strive to achieve the goals of the group before achieving his personal ones. In other words, elitist makes himself popular, but popular person is made popular by his group.

Popularity often results in power and dictates the way people make decisions that can affect not only an individual's life, but also the whole nation [27] [28]. The importance of this social construct in real-life is undeniable. However, with the development of the internet, the concept of popularity was given a completely new dimension, that needs to be discussed in particular.

3.2 Popularity on the internet

Even in 2020, when we talk about popularity and its impact, we need to separate offline and online presence. And there are multiple reasons for that.

First and the obvious one, not all people know, what "online" is. Despite that World Wide Web (www) is available to the general public from the 1990s [29], not all people have the luxury to use it. According to the Internet World Stats [30], by December 2019 41.3 % of the world population still doesn't have access to the internet. For these people, such concepts as "likes", "reposts", "followers" don't mean anything. They still assess popularity solely based on factors, that were used before the social networks and media became what they are today. For example, what television or newspapers present to the public.

Second, online and offline presence usually is not equal and have a different impact. Even if business or person is presented in the real world and on the internet, the weight and influence of popularity may be different for online and offline audiences.

For businesses, online popularity can dramatically influence its offline situation in a very short period of time. For example, the quality and quantity of hotel reviews posted on the TripAdvisor platform [31] can significantly impact the offline popularity of the hotel [32]. In the article by Karen Xie, Chihchien Chen & Shinyi Wu - "Online Consumer Review Factors Affecting Offline Hotel Popularity: Evidence from Tripadvisor" [32], authors tried to explore this phenomenon and used reviews to estimate the effect it would cause on the offline hotel occupancy. After analyzing 56 284 reviews of over 1000 hotels, they not only found the correlation between the percentage of positive reviews and hotel booking, but also that reviews which were posted in the recent 2 quarters have the most influence.

From the individual's perspective, we also need to separate online and offline popularity, as it is influenced by different personality traits [33] [34]. In the online world, we tend to behave differently than in the real one. According to the psychologist Leora Trub [35], depending on the individual's nature, the online persona can be more or less self-revealing and include more or fewer negative traits than the same individual's offline persona. As a result, the same person can be more appealing to the public and thus popular from their online or offline presence based on traits that the audience wants to see. This all means, that online and offline presence are separated. However, at the same time, both of them have a strong impact on each other.

Finally, even when talking solely about the subject of online popularity, we need to be aware of the platform, we are discussing. Each social networking and media platform has its own popularity measurements, metrics and audience. This means that having an impact on one service doesn't automatically guarantee the same level of impact on the other one.

For example, differences between Instagram [36] and YouTube [37]. The main visible metrics of popularity are similar: followers/subscribers, number of comments and number of likes. However, the user behavior on two platforms is different. According to the article by Emilio Ferrara, Roberto Interdonato & Andrea Tagarelli [38], Instagram users seem to have a relatively short attention span and a limited vocabulary, which is why they usually don't get involved in long conversations in the comment section. Also, popular users on the platform produce content in a specific way, where they either show a very narrow and definitive content or a very broad one. At the same time people with broader interests tend to be more popular than others. On the other hand, the article by Dustin J. Welbourne and Will J. Grant [39] shows that users on YouTube pay more attention to other things. After analyzing 390 science communication videos, that were produced by 21 professional channels and 18 amateur content creators, authors demonstrated that despite the number of produced videos by professionals was much higher, content done by user-generated channels was much more popular. Also for YouTube, the presentation of content plays a big role in user popularity. As a result, shorter videos with the presenter, who appears from video to video were more successful on this platform.

Differences between popularity influences on social media and networking platforms lie not in the metrics (likes, comments, etc.), but the type of content and its target audience. Cases like Instagram and YouTube demonstrate distinctions in the main types of content: Instagram - pictures, Youtube - videos. But, for example, to understand popularity on YouTube and TikTok [40], we have to look at the target audience of the service as they are both video platforms. And the audience, will dictate, what type of content should be produced by creators to become popular [41] [42].

All written above, demonstrates that popularity depends both on the offline and online worlds,

as both of them have an impact on each other. However, while analyzing social media and networks, the platform's main idea, user base and interaction needs to be understood thoroughly, to explore its popularity aspect.

3.3 Couchsurfing

The way Couchsurfing operates, its ecosystem and user base has been already described above in the Chapter 2. From there, it is clear that this service provides a very unique experience, where user interaction is meant to go from virtual to real world. So there is no surprise, that it has always been an interesting topic for research.

The most interesting aspect of the platform that was explored in many papers is trust. The level of trust that has to be built between users on Couchsurfing is much higher than on other social networking platforms, as the whole concept of accepting a stranger into a personal space (house) would not work without a high level of trust. One of the main ways of building confidence in a host or a surfer on Couchsurfing is through analyzing references on the user's account. According to the works by Alexander Ronzhyn [17] and Eugenia Kuznetsova [18], the trust can be conveyed through the reference in different ways. After analyzing multiple datasets of more than 450 references using the corpus-based linguistic analysis [43], authors came to the conclusion that references with a high number of adjectives are considered as a sign of good personal communication between users. Besides that, there were very specific words found in the positive references, including "interesting", "generous", "positive", "trustful", "easy-going" etc. All of them seem to represent the personal traits that Couchsurfing members find very important. Also a lot of references had a mention of a specific activity, like "town tour" or "excursion", which also shows that members of this community value the time that hosts and guests spend together.

The confirmation for that can be also found in works done by De-Jung Chen [44] and Xiao Liu [45]. According to them, the unique experience that Couchsurfing provides to its members made it more than just a hospitality exchange platform. It has transformed it into a completely separate traveling style, where the traveler is not a tourist anymore, but a person who tries to discover new places from the perspective of the local. Instead of visiting as many sightseeing attractions as possible, surfers prefer to put emphasis on communication with the host and learn more about the place from it. And this was possible, because of all the mechanisms that Couchsurfing provides. For example, their user reputation system helps to maintain the trust, which is very important for its members. Having a lot of neutral or negative references, can affect the host's or surfer's experience, so they try their best to maintain a positive image, by being friendly, generous, and spending quality time together.

A good example, of how Couchsurfing mechanics help to provide a positive experience to its members and maintain this new style of traveling is described in the article by Maura Rae Cherney [16]. Based on the survey responses of 231 hosts, the author identified which features and attributes influenced the host's decision to accept or reject a guest the most. In other words, the information that most likely is used to reduce uncertainty in the decision-making process is (in importance order):

1. Couch request message. Is the message appropriate, does the arrival date suits the host, etc.?
2. References. How many negative references are on the surfer's page, why they are negative, etc.?
3. Other profile info. What are the surfer's interests, description, etc.?
4. Photos. Does a surfer has some photos on their page, what are on these pictures, etc.?

Some of these features are very popular on the social networking platforms, like photos or profile information. However, the couch request messages and references are very uncommon attributes of Couchsurfing. All of them in combination, help to either build or destroy the trust in a member.

Taking into account what discussed above, it is clear that Couchsurfing is a unique hospitality exchange platform, which is of interest not only for users, but also for researchers. To contribute to the current state of the art, in this thesis, we performed an exploratory research of the host's popularity. In the next chapter, we describe the data that we used in our analysis.

4 Data description

In this chapter, we describe our dataset that was used for analysis. First of all, we explain the origins of the data: what type of information we decided to use, how we extracted, anonymized, and cleaned it. Then we talk about evaluation methods: why we divided hosts into different categories based on their popularity, what these categories are, and how we evaluated which user belongs to which group.

4.1 Description

This study intends to explore the subject of hosts popularity on Couchsurfing. To make it happen, we needed data that would fit our criteria:

1. **It should be publicly available to all surfers.** Surfers send couch requests based on things that they see on hosts' profiles. Therefore, information that we use in our research had to be identical to what couch surfers can see on the platform.
2. **It should be anonymous.** The results of this study should be done in the form of aggregated reports that don't need any personal information. E.g. name, Facebook profile link. Therefore, our dataset should not contain any information that can be used to identify a host.

To understand what type of data we have to deal with, we had to create an account on the platform and look for some random host profiles. Based on what we saw on the website, we identified 2 types of information that suit our needs and can be extracted:

1. **Fixed type data.** Everything that involves general information. E.g. gender, age, number of friends, etc.
2. **Free text.** Text that can include everything that person wants to write there. E.g. interests, couch descriptions, hobbies etc.

To perform the analysis, we had to decide how large the dataset should be and in which locations our gathered users should host. As this is a quantitative research, our dataset should be as big as possible to give more objective results. Therefore, in terms of the amount of collecting data, our decision was to not put any limit on the number of collected hosts.

For the location, we also concluded that it is better to have as many different cities as possible. The idea behind that was to take cultural diversity into account, because hosts in one part of the world can be popular for one set of reasons, but in the other one, the logic can be completely

different. Having multiple contrasting locations would allow us to analyze them separately and then compare the results between them, making our analysis more comprehensive. In order to get closer to the real-life scenario in our research, we decided to take the most popular cities to visit according to the previous works. This is made based on the assumption that the more popular city is, the more surfers would like to visit it and thus generate more hosts of different popularity levels.

We chose to use the list of 20 most visited cities in the World by 2018 published in the Mastercard Global Destination Cities Index 2019 [46]. These cities are (Figure 8): Bangkok, Singapore, Tokyo, Seoul, Osaka, Phuket, Pattaya, Bali, Hong Kong, Dubai, Istanbul, Antalya, Mecca, Kuala Lumpur, Paris, London, Milan, Barcelona, Palma de Mallorca, New York City. It is clear from the list, that the choice was diverse enough as there are cities from different parts of the world and thus different cultures.



Figure 8: 20 most visited cities by 2018 (first location pool)

From each city, we selected all hosts that were ready to accept guests and had at least 1 reference from surfers. This was done to exclude hosts that didn't host anyone before and thus don't have any experience with guests.

However, during the data collection, we found out that some of the cities didn't have enough hosts matching our criteria. In particular, Milan, Phuket, Pattaya, and Palma de Mallorca had less than 100 users of that kind. Also, looking at the mapped cities (Figure 8), it became clear that many more cities were presented from Asia than from any other region. Consequently, we

decided to significantly increase the location pool to include cities from six continents: Asia, Europe, North America, South America, Australia, Africa. The result can be seen in Figure 9.



Figure 9: 65 cities with at least 100 hosts (second location pool)

In the end, after collecting all necessary data from the second location pool, we end up with a total of 47 564 hosts from 65 cities. The Figure 10 shows detailed statistics of our dataset.

Hosts info

Characteristics	Value
# of hosts	47 564
# of hosts in Asia	13 123
# of hosts in Europe	17 780
# of hosts in North America	4 598
# of hosts in South America	9 551
# of hosts in Australia	1 605
# of hosts in Africa	907

Cities info

Characteristics	Value
# of cities	65
# of cities in Asia	17
# of cities in Europe	16
# of cities in North America	10
# of cities in South America	14
# of cities in Australia	4
# of cities in Africa	4

Figure 10: In total, data of 47 564 hosts from 6 regions was collected for research

4.2 Data preparation

The process of data collection for this research had to be done in a specific way as Couchsurfing doesn't provide any public API. That meant that custom solution had to be used. Usually in these cases web-scrapers [47] or web-crawlers [48] are used to get data, but we were able to find the custom-built API (Application Programming Interface) solution [49], that had all necessary endpoints to retrieve the information. It worked as a usual API [50]:

1. To get access to the endpoints, a user provides his own Couchsurfing credentials (username, password);
2. Sends the desired request to the server;
3. Gets a response in the JSON format.

This helped us in the process of data collection. However, to suit our needs, we had to significantly modify the original script, as unchanged it was just a pure API. Therefore, to collect large amounts of data, we needed to write an automation tool that would not only retrieve the necessary **publicly available** data, but also anonymize and save it in a suitable format. As a result, finished automation script's algorithm works as follows (Figure 11):

1. Pick a city from the array of all 65 cities;
2. Request a list of users sorted by hosting experience (100 users per page);
3. Loop through these users and request full profile information for each one of them;
4. Anonymize & clean retrieved data;
5. Request references info separately;
6. Format and merge references data with the host data;
7. Repeat the cycle until encountering the first host with 0 hosting references for given city;
8. Write retrieved information into a separate .csv file.

The process of getting host data was slow. On average, it took about an hour to get the data for each city. This was caused by the requirement of retrieving each host data separately. As a result, each page with 100 users, needed a total of 201 GET requests: 1 to get the hosts listed on the page + 2 per host to get profile information and references.

The data anonymizing and cleaning was also done in the script. The main purpose of this besides

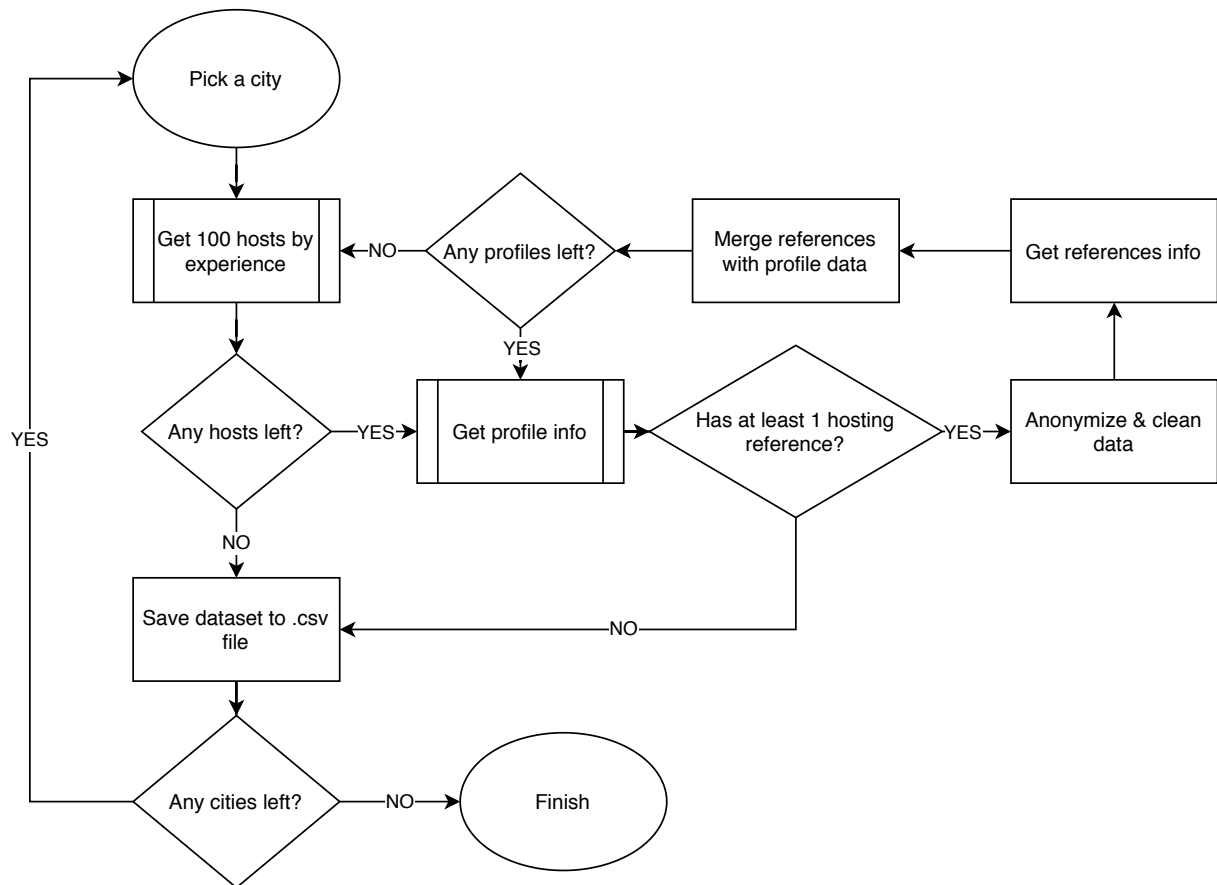


Figure 11: Data collection algorithm

anonymization was to eliminate attributes that were unnecessary for the research (e.g. any data related to the profile which credentials were used for the API calls) and structure everything appropriately (e.g. reformatting nested fields). As a result, part of the host's information was removed. The deleted data can be divided into 6 types:

1. **Unnecessary personal information.** Fields like host name, id etc.
2. **Empty fields.** Fields that didn't contain any value for all hosts. E.g. firstName, email, address etc.
3. **Identical fields.** Fields, which values were the same for all hosts. E.g. status, isDeleted etc.
4. **Redundant fields.** Cases where other fields reflected the same information better. E.g. isVerified, photoCount etc.
5. **Hidden fields.** Fields containing personal information that cannot be seen in the host's profile. E.g. isFacebookLinked, facebookFriendCount, friendStatus etc.
6. **Restructured fields.** Fields that were nested first and were restructured to fit the .csv format, but information in which hasn't been changed. E.g. verification, couch, oldSchool etc.

7. **Reformatted fields.** Fields that had some redundant information, and thus were reformatted to have only relevant data. E.g. `guestPreferences`, `albumList` etc.

Most of the information in the received data was saved, as we had to leave as many potential popularity influences as possible. Although, the examples of the restructured and reformatted fields can be seen in the Figure 12.



Figure 12: Restructured and Reformatted fields examples

In this thesis, we didn't plan to perform any type of linguistic analysis [43], therefore to make use of text variables they should've been converted to the numeric or boolean types. Textual characteristics then were transformed according to the types of information they contained. In the end, there were 3 types of text variables:

1. **Variables that can be encoded.** E.g. gender (male = 0, female = 1, other = 2), type of sleeping arrangements (shared room = 0, public room = 1, private room = 2, shared surface = 3) etc.
2. **Variables that can be counted.** E.g. number of fluent languages, visited countries etc.
3. **Variables that can only indicate their presence.** E.g. description of the host's interests, public transit etc.

As a result of anonymization and cleaning the formatted data was represented as follows:

1. **Personal characteristics.** This is the information representing the host's personal information, such as age, gender, interests, etc.
2. **Account activity.** In other words, everything related to the user's activity on the platform: last hosting date, response rate, days from the last activity etc.
3. **Social information.** Things such as the amount of friends, communities that host is a part of, badges, all fall into that category.
4. **Hosting, surfing & personal experience.** These are all characteristics related to the previous experience of the host, including the number of hosted guests, number of positive & negative references, previous surfing & personal experience etc.
5. **Hosting characteristics.** All the information about guest preferences, couch availability, and its description.

For each city we had a separate data file with different numbers of hosts in it. Depending on the number of experienced users the dataset contained from 100 to 1000+ entries each.

4.3 Identifying popular users

After the data collection process was over, we could start to identify different popularity categories. To do so, we had to choose a certain characteristic that can be the definition of the host's popularity on the platform. This will serve as a threshold that separates hosts into different groups by popularity.

While looking at the data from the platform, it became clear that "host popularity" in the context of Couchsurfing is measured by the amount of positive references from guests or, in other words, positive hosting experience. For the thesis, we decided to go with the Couchsurfing's approach. As a result, our popularity metric shows what is the total number of guests who have been staying at the host's place and left a positive reference.

As mentioned above, to analyze our data, we agreed to divide it into 3 groups: **popular**, **somewhat popular** and **unpopular hosts**. In the beginning, we tried to consider hosts, who had at least 80% of the maximum number of hosting experience for the city as popular, from 50% to 80% (not included) as somewhat popular and less than 50% as unpopular hosts. However, in the process of division, turned out that we cannot simply use a universal percentage margin for each city as the gap between the number of positive hosting experience can be too big. For

example, with the use of this threshold, out of 491 hosts in Bangkok, only 2 could be considered "popular", 2 - "somewhat popular" and others were "unpopular". For the same reason, dividing the dataset into quartiles [51] wasn't an option as well. Every city, had the same picture, where 75% of all hosts fell into the "unpopular" group. Objectively, both of these approaches weren't suitable for our case, as both of them factually mix "somewhat popular" and "popular" categories together.

In order to make the evaluation more realistic with very distinct "somewhat popular" and "popular" groups, we agreed to go with the alternative solution. To divide hosts into categories based on popularity, we decided to use histograms [52]. A histogram is a bar graph that is used to visualize quantitative data. When given a variable, the histogram divides it into predefined intervals that are called bins, and displays them on the X-axis. On the Y-axis, it shows the number of instances, which variables fall into that interval. Histograms are usually used to explore the data distribution based on a particular variable.

In our case, we used histograms to see how hosts are distributed based on their popularity (positive hosting experience). To do so, first, we had to find an optimal bin width (interval), and then visually divide hosts into popularity groups based on the final histogram. For each city, we used intervals of 1, 5, 10, and 15 positive hosting references as bin sizes. As a result, we had 4 histograms for each city. Figure 13 displays these plots for Rome.

For each city, judging by its histograms, we chose the most suitable bin sizes based on clear visual distinction between groups. If it turned out that even bin size of 15 was not big enough to recognize popularity categories without any doubt, we experimented further, until finding a suitable bin width. As a result, for each city we drew the final histogram with the appropriate positive hosting experience interval and then visually divided them into groups by popularity. E.g. Mumbai hosts were divided into popularity groups based on the histogram with bin width equal to 15, and it is depicted in the Figure 14.

In the end, results for cities varied. For example, for Bangkok 50+ positive hosting references can be considered as popular, 25-50 as somewhat popular, less than 25 as unpopular. But for Antalya this number is lower: 22+ is popular, 8-22 is somewhat popular, less than 8 is unpopular.

After popularity thresholds were clear, we additionally separated all hosts from each city based on their group. As a result, each city had 3 separate groups that had to be explored individually and compared with each other. The next chapter describes methods that were used to analyze all these categories.

Rome histogram

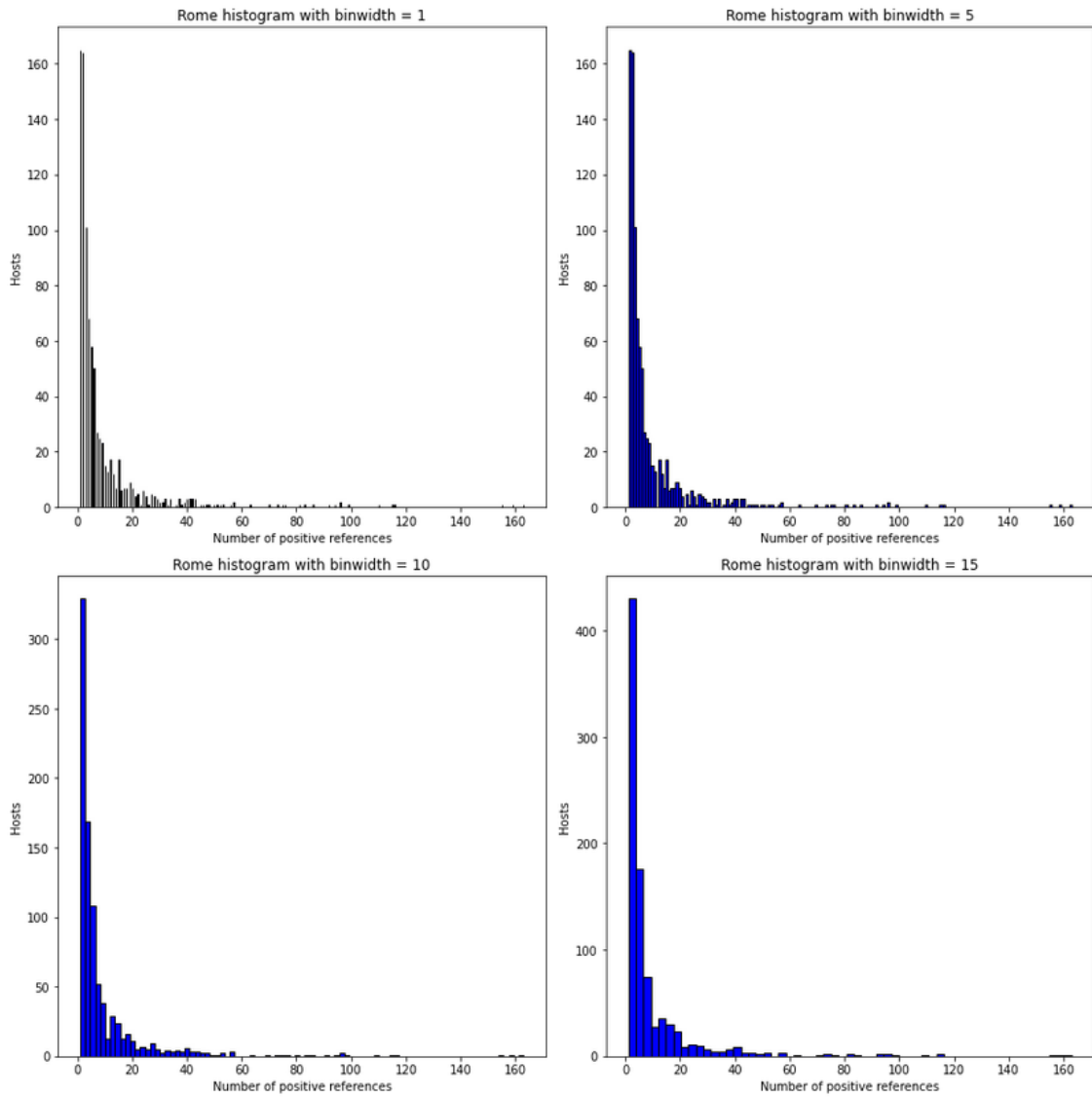


Figure 13: Rome hosts histograms with bin sizes of 1, 5, 10, and 15. On the X-axis is the number of positive hosting references and on the Y-axis - number of hosts

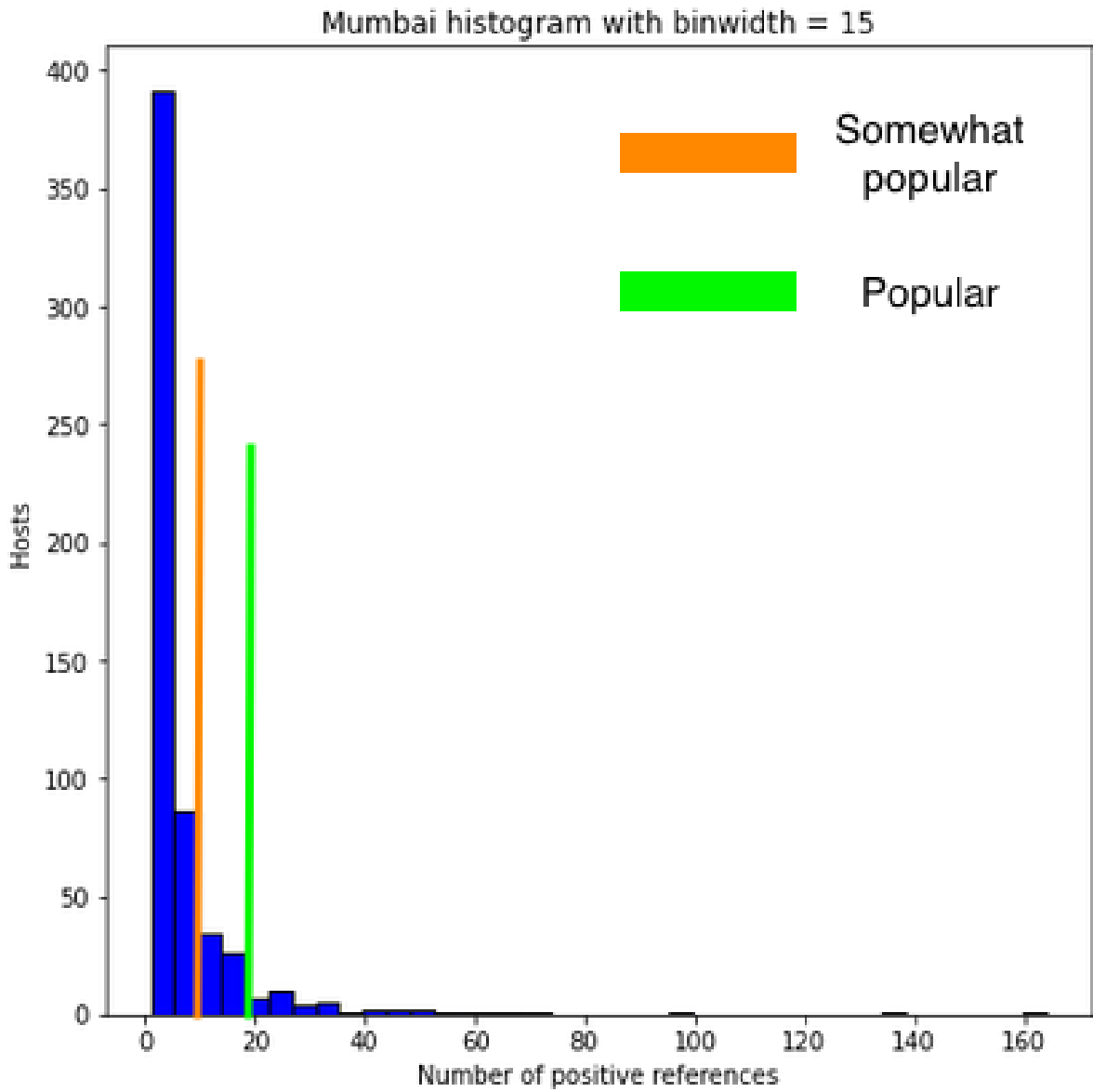


Figure 14: Popular hosts from Mumbai have at least 20 positive hosting references, somewhat popular - at least 10 and unpopular - less than 10.

5 Methodology

In this chapter, we describe our research methodology approach. As described in the previous section, our popularity metric was the same as on the Couchsurfing - amount of positive hosting references. Now, the question was, how to find out, whether there is any dependence between hosts' popularity and other characteristics publicly available on their profiles. And if there are, then how big and different it is for all 6 geographical regions. To answer these questions, we had to find appropriate methods to analyze our dataset and draw logical conclusions.

In statistics, to calculate a correlation between 2 quantitative variables the correlation coefficient is used. It is a number that lays in between -1 and +1, where the mark $-/+$ shows the direction of the linear relationship between 2 variables, and numeric value - the strength of the correlation [53].

There are 2 types of correlation coefficients that are mostly used [54]: Pearson product-moment correlation coefficient and Spearman's rank correlation coefficient. Differences between them lie in the:

1. **Type of data relationship.** Pearson coefficient is used for linear relationship, Spearman - for monotonic relationship;
2. **Type of variables.** If variables are of ordinal type (e.g. satisfaction score), then the Spearman correlation coefficient is more appropriate than Pearson's one [55].

For our research, we chose to use the Pearson correlation coefficient, as we didn't have any ordinal variables in our dataset. In the beginning, the plan was as follows:

1. Calculate the correlation score for popular, somewhat popular and unpopular hosts of each city;
2. Merge results of each city together into a correlation matrix by their geographical region and popularity group;
3. Compare popularity matrices between each other within a region and between all regions.

This plan is also displayed in the Figure 15. After we agreed on the approach, we were going to use for analysis, we could start to explore the data.

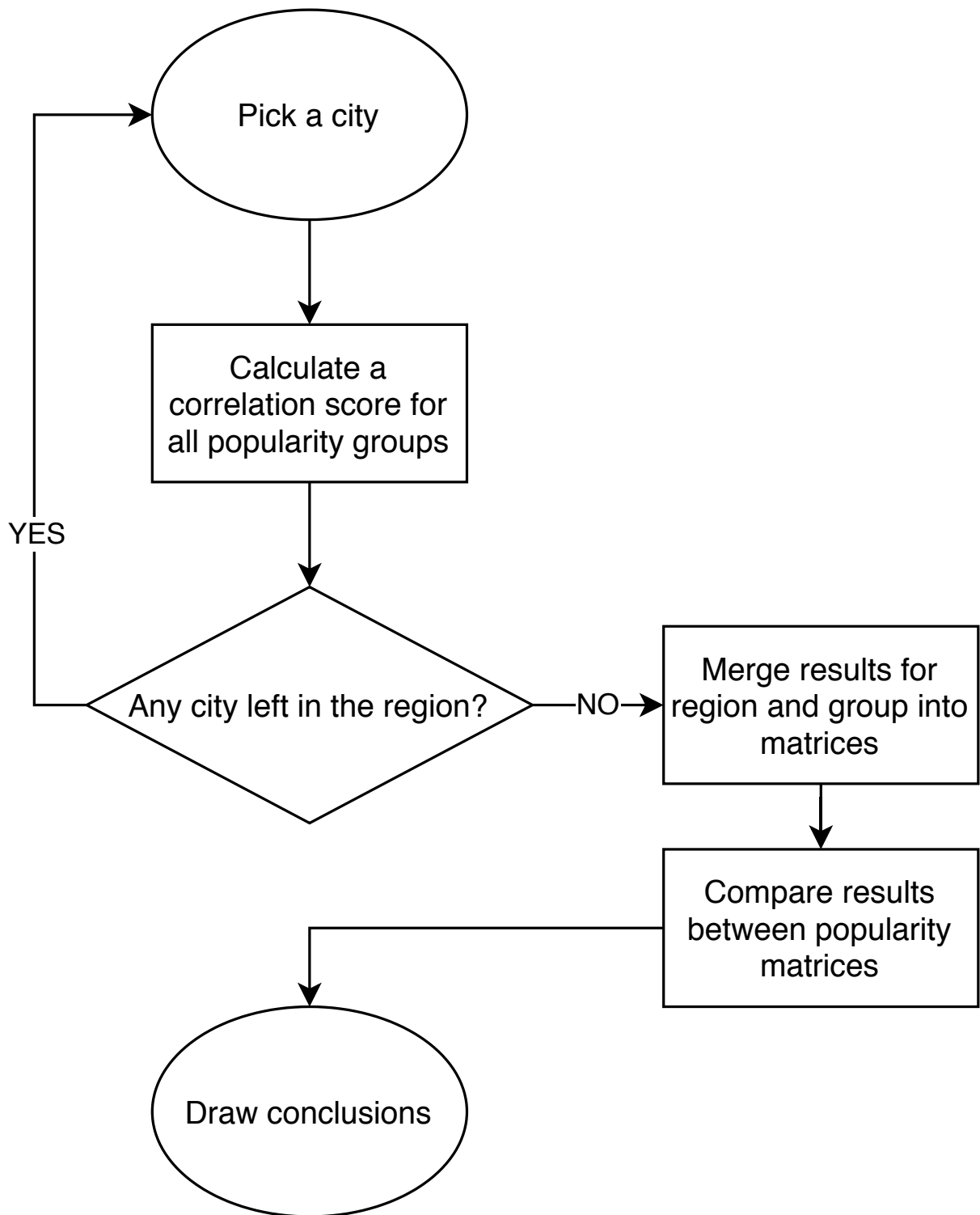


Figure 15: The initial plan of analysis

6 Results

In this chapter, we first describe our results for macro analysis, where we created a correlation matrix for each region and found out the strong characteristics that influence hosts' popularity (Section 6.1). Next, we describe our meso analysis results, where we analyzed each of the strong characteristics separately from the perspective of the popularity group and region (Section 6.2).

6.1 Macro analysis

When we started to perform our analysis, we quickly realized that the initial plan does not work as we thought. It turned out that each region has 3 matrices with too many characteristics to compare and interpret. For example, in Figure 16 is pictured correlation matrix for popular hosts in Asia. The more saturated square is, the stronger correlation is. Cells with a red tint have a positive correlation, with a blue tint - negative. Empty cells mean that all hosts from the city have the same variable value and, therefore, there is no correlation.

The better approach was to filter out characteristics that didn't have a strong enough correlation with popularity. We defined a heuristic, that characteristics with a big influence on the host's popularity should reflect it on the popular hosts matrix. This would allow us to then look at how this value changes, comparing to somewhat popular and unpopular hosts' scores. So, we introduced changes to our initial plan that included 3 additional steps (Figure 17):

1. Find characteristics with a minimum correlation score of +/- 0.5 (moderate positive or negative correlation) in at least half of the cities from region's popular hosts' matrix;
2. Extract these characteristics from each popularity group matrix;
3. Compare matrices between each other and between regions.

As a result, we removed characteristics that didn't have sufficient correlation scores for the region and thus made it possible to analyze each matrix separately. The filtered out correlation matrices for Asia are demonstrated on Figures 18, 19 and 20.

In total, we had 3 strong characteristics that needed to be looked into further:

1. **Number of friends.** How many friends does the host have?
2. **Number of countries host's guests are from.** Surfers from how many different countries visited the host's place?
3. **Number of positive personal references.** How many positive personal references have people left on the host's account?

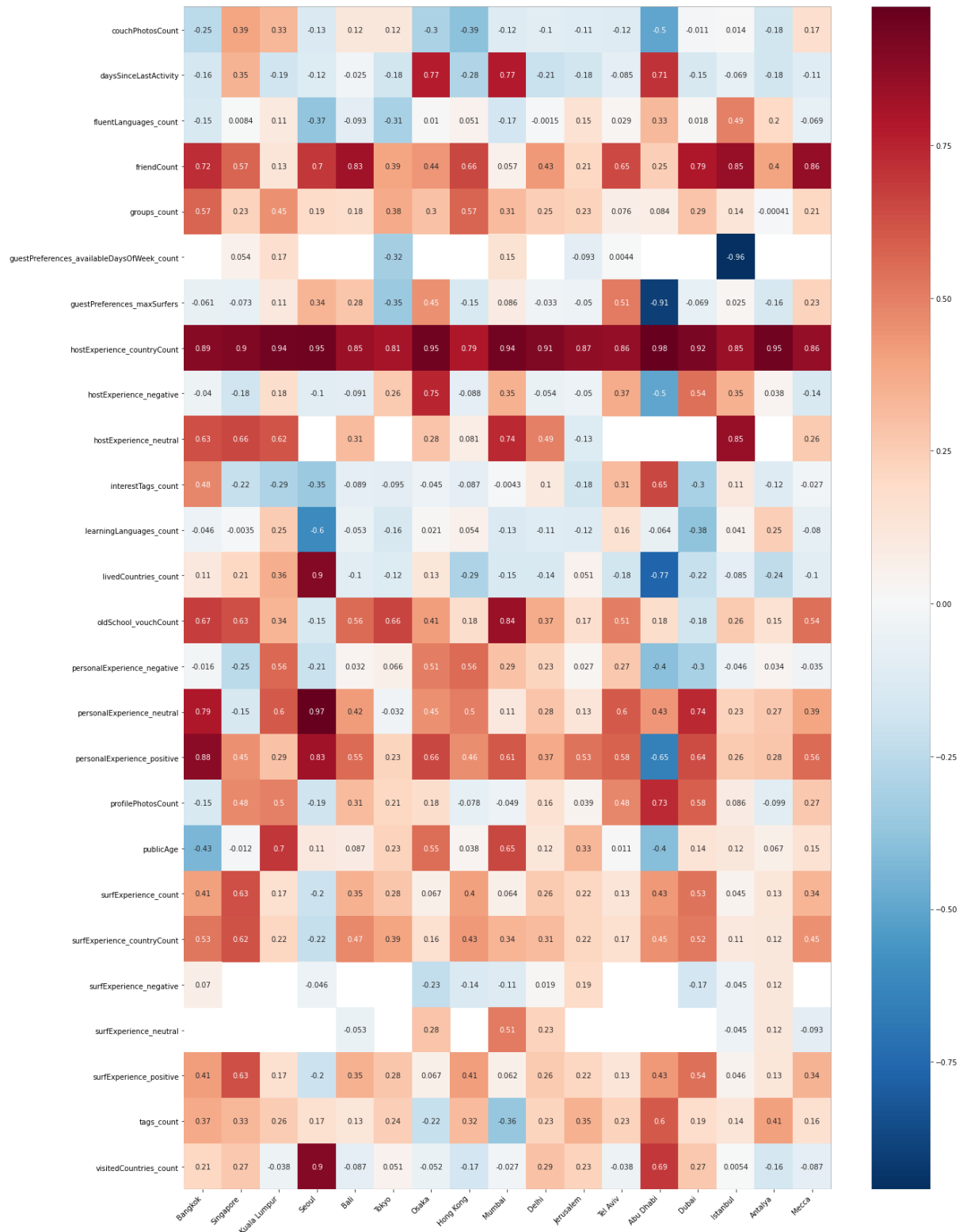


Figure 16: Heatmap of the Pearson correlation matrix for popular hosts in Asia.

6.2 Meso analysis

To conclude the analysis, we needed to use correlation matrices that were built before to demonstrate dependencies between popularity and these characteristics.

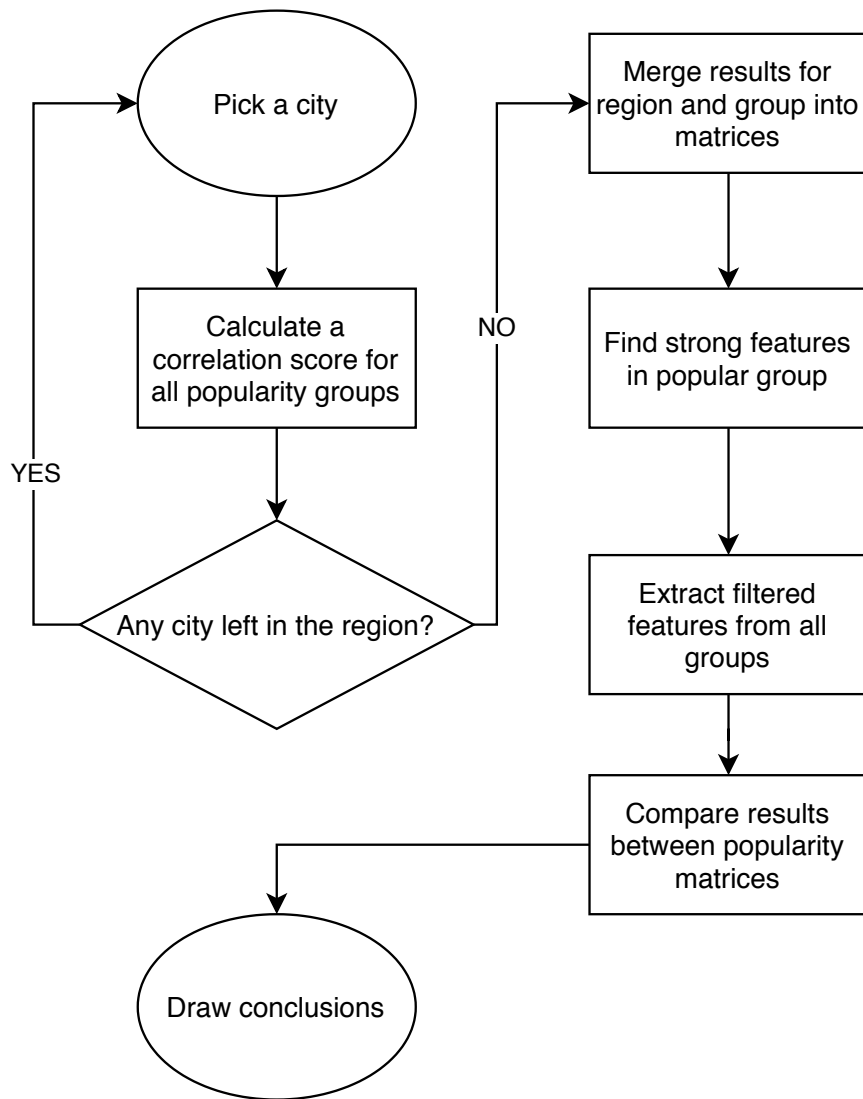


Figure 17: The updated plan of analysis

First, the number of friends. How does the number of friends affect the host’s popularity? To answer this question, for each region, we calculated the percentage of people with more friends than the number of positive hosting references.

It turned out that having more friends doesn’t make a host more popular. This result is consistent throughout all regions. The only slight difference between somewhat popular and unpopular hosts is demonstrated in Australia and Africa. However, it doesn’t contradict our conclusion, as the percentage of popular people with more friends than the number of positive hosting references is still lower than for somewhat popular and unpopular people. The barplot, demonstrating our conclusion can be seen on the Figure 21.

Second, the number of countries surfers came from. For this characteristic, we wanted to see results, not from the popularity groups’, but the regions’ perspective. In other words, not how

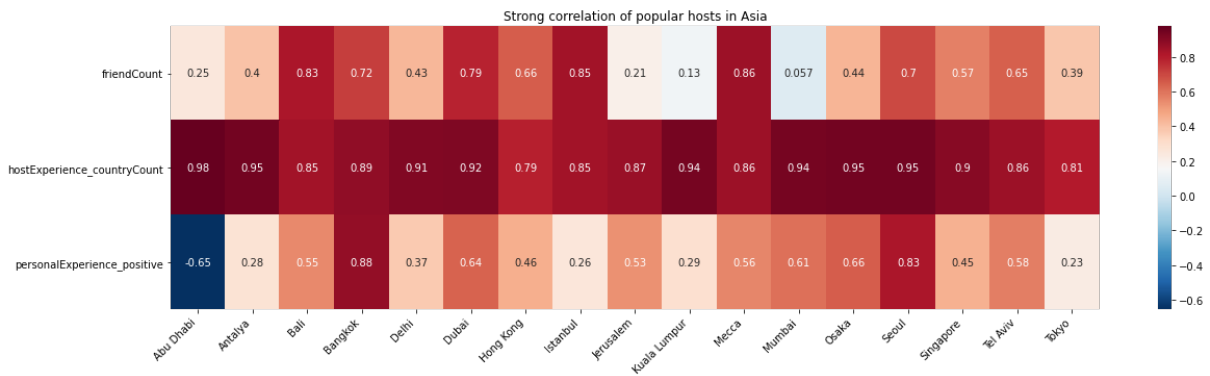


Figure 18: Strong characteristics for Popular Asian hosts

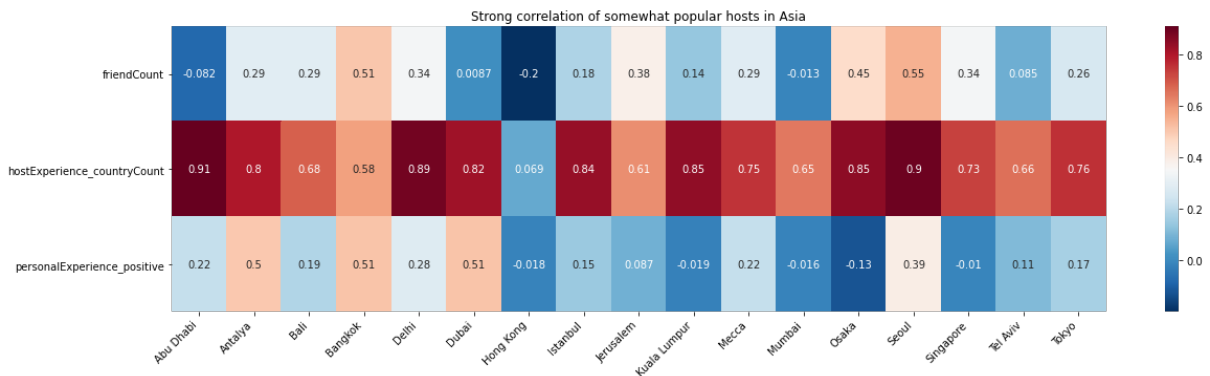


Figure 19: Strong characteristics for Somewhat Popular Asian hosts

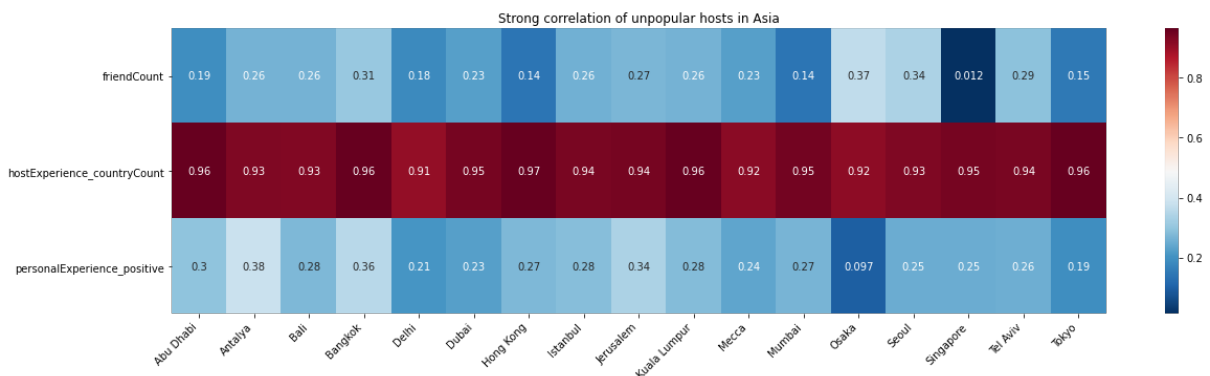


Figure 20: Strong characteristics for Unpopular Asian hosts

different numbers are within a region between its popularity categories, but what are the differences in a particular group between regions.

To find it out, we counted the average number of countries that surfers are from for each region and separated them by popularity groups. On the Figure 22 we display the resulted plot, which shows how this characteristic differs from region to region.

From the plot, we can draw a conclusion that an average host from Europe has hosted surfers

More friends doesn't make a host more popular

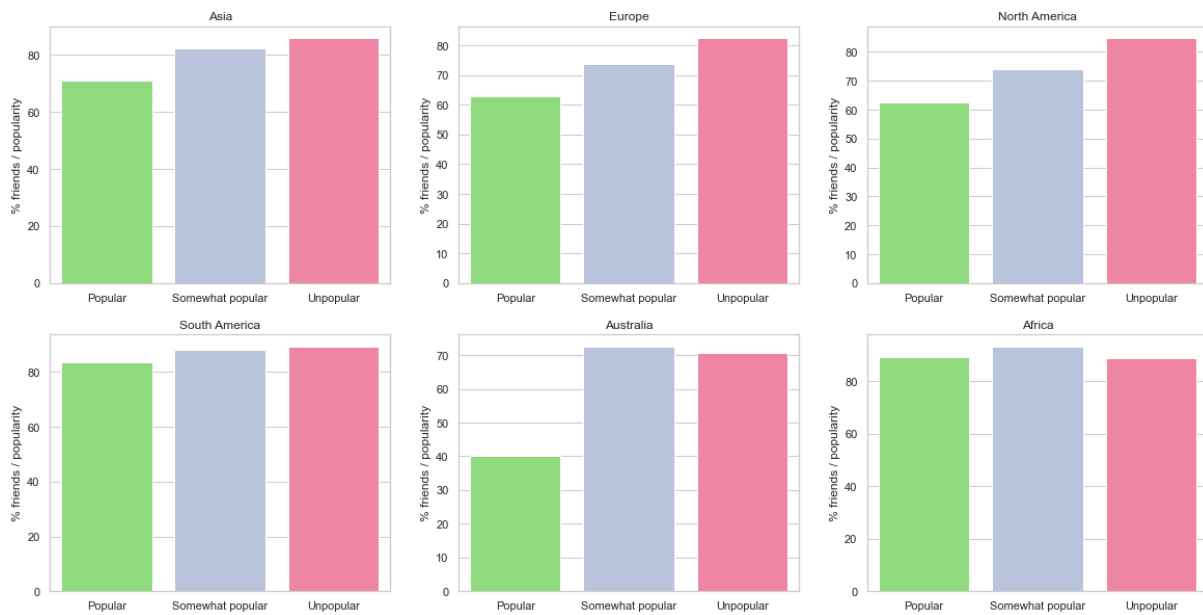


Figure 21: Percentage of hosts, who have more friends than amount of positive hosting references is lower for popular than for somewhat popular or unpopular hosts

from more countries than hosts from any other region. This can be due to several factors that can cause bigger traveler flow from different countries:

1. **An average person from Europe has higher income** [56]. Despite that couch surfer doesn't have to pay for lodging, there are still other travel costs, like food or plane tickets. Higher-income can make traveling much more affordable for an average person.
2. **Within Europe, there are a lot of small, but popular countries for tourists** [57]. Visiting neighboring countries can be done very quickly if they are small. E.g. a trip from Estonia to Latvia and back can be done in a matter of one day. This time aspect makes it easy to couch surf.
3. **European Union citizens don't need a visa to visit European countries** [58]. Citizens from the European Union don't need to apply for a visa to travel to EU countries, and can simply use a passport of their country. This simplifies visiting other countries and makes traveling much more convenient.

The last characteristic that is strongly correlated with the popularity is the number of positive personal references. To explore, what this variable can say about popularity, we first, decided to compare values between regions to see whether there are any significant distinctions inside the popularity group caused by geographical differences. Similarly to the number of friends

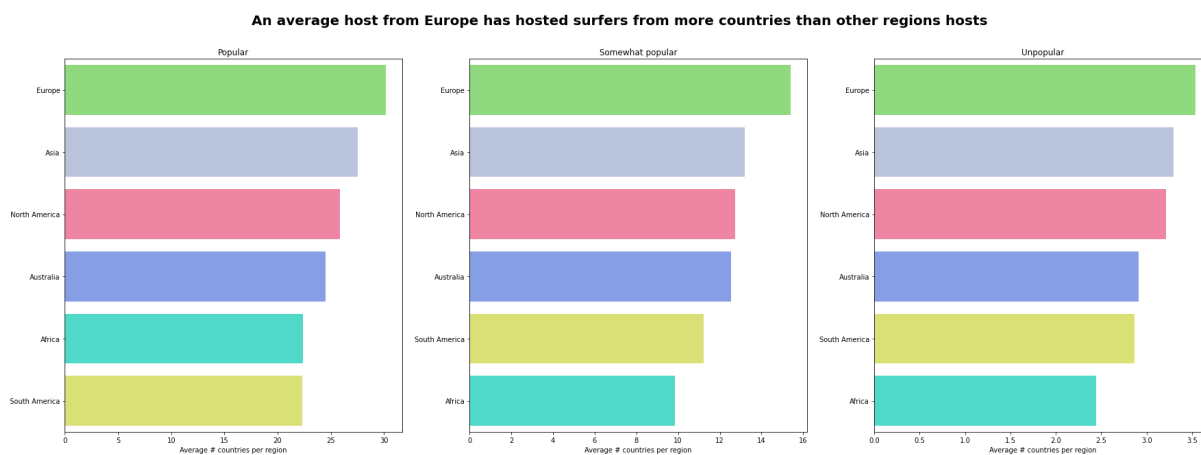


Figure 22: Average number of countries surfers were from is higher for Europe than for any other region

that hosts have, we calculated the average percentage of positive personal references out of all personal references per region. This resulted in a plot shown on the Figure 23.

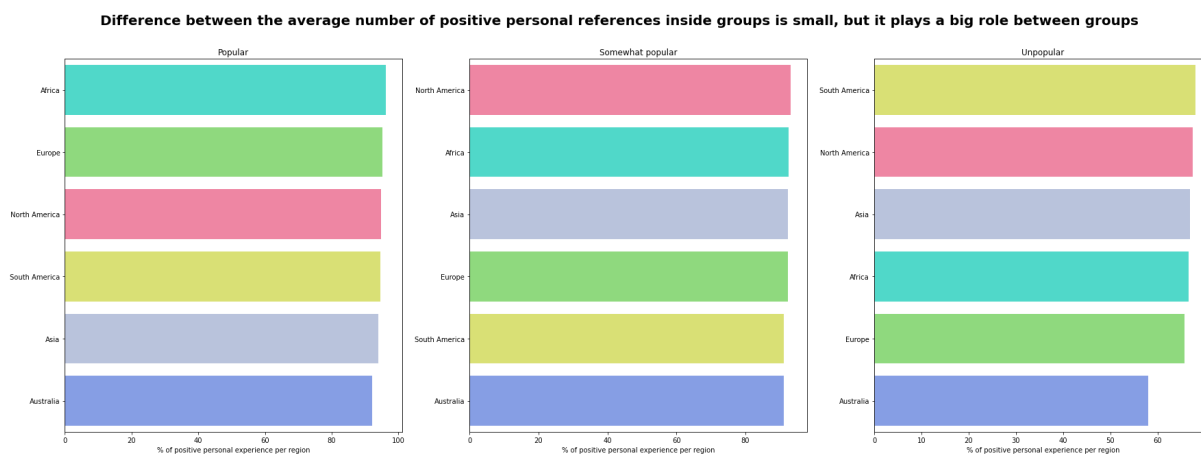


Figure 23: Average percentage of positive personal references doesn't significantly change between regions

The results of this experiment showed that there is no significant difference between countries within a group. However, there was a clearly visible contrast between groups. That is why we decided to explore the average percentage of positive personal references inside each region separately to see how it changes depending on hosts' popularity. The results can be seen on the Figure 24.

Judging by the final plot, we saw that the more popular host is, the bigger number of positive personal references he has. We assume that popular hosts not only have more guests, but also are more social and outgoing. This means that potentially they meet more people than an average host, and thus, more people can leave a positive reference. With personal references, there is

The less popular host is, the less is the percentage of positive personal references

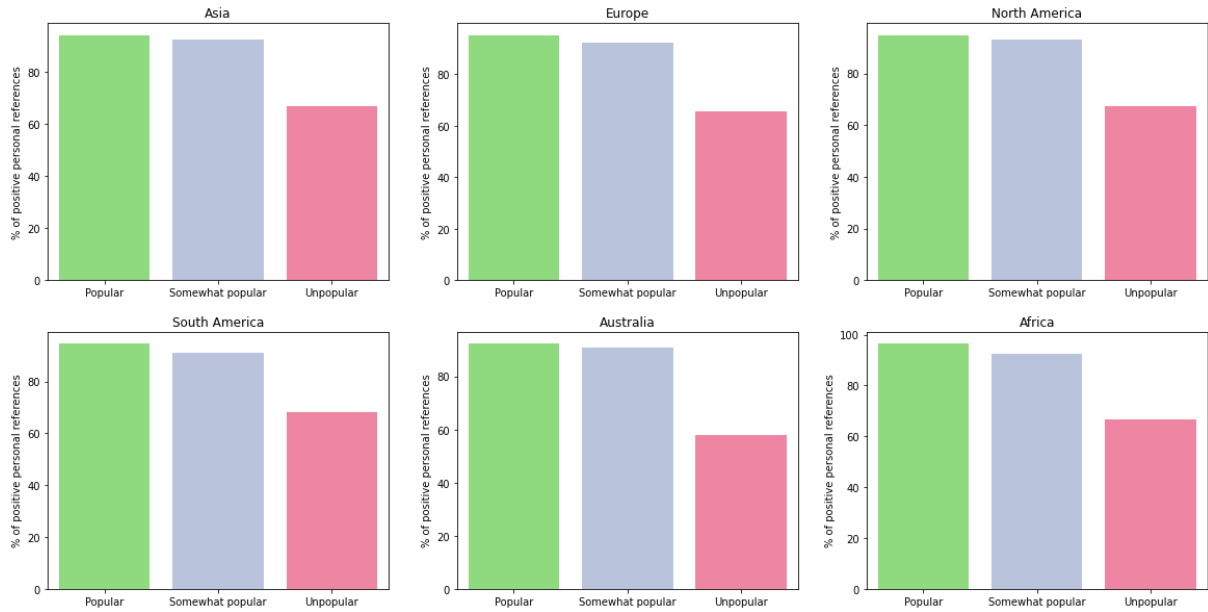


Figure 24: Average percentage of positive personal references per region. The less popular host is - less positive personal references he gets

always a chance that a comment can be not only from host or surfer, but also from a random person. That is why a person's communication skills can play a much bigger role here than his hospitality

7 Conclusion

Couchsurfing is a hospitality exchange platform that provides its users with a unique traveling experience. Surfers - people, who look for lodging during their trips, not only get it for free, but also have an opportunity to explore the new place through the eyes of the locals. Hosts - people, who provide a place to stay, have a chance to meet new interesting individuals and fully participate in the cultural exchange.

As a popular internet platform, Couchsurfing still grows and is used by many people all around the world. However, for new members it becomes harder and harder to experience, what makes this platform special, as they lack the key aspect that influences users' interaction - experience. The number of times a person has hosted others or couch surfed, for the Couchsurfing ecosystem means a person's popularity. The more popular user is - the more chance he gets to participate in the cultural exchange.

To analyze the subject of hosts' popularity, we collected the data of 47 564 hosts from 65 cities from all around the world, to perform quantitative exploratory research. By using histograms to see the distribution of the positive hosting references (host's popularity), we were able to divide all hosts into three popularity groups: popular, somewhat popular, and unpopular. We then used the Pearson correlation coefficient to analyze our dataset.

As a result of this thesis, we were able to find answers to the research questions established:

1. **RQ1:** *Is the popularity on the platform completely randomized or there is a correlation between certain attributes of the host's profile and popularity?*

The hosts' popularity on the platform is not randomized and has a clear correlation pattern with profile features. As a result of macro analysis, we found out that characteristics with the strongest correlation are: the number of friends, the number of countries that surfers originated from, and the amount of positive personal references.

2. **RQ2:** *What are the differences between popular and non-popular hosts?*

While performing the meso-level analysis and looking into each characteristic separately, we found out that popular hosts usually have much less friends than unpopular or somewhat popular ones. Also users from the popular category have a much higher percentage of the positive personal references than others.

3. **RQ3:** *Are there any geographical differences in the hosts' popularity?*

As for the geographical differences, the biggest difference was found in the number of countries that surfers originated from. On average, hosts from Europe have guests from

more different countries than hosts from any other continent.

As a result, this research concluded that the popularity of the host on the Couchsurfing platform, is not a randomly accomplished characteristic. There are certain factors in each part of the world that influences it. Despite being similar, these features can behave differently depending on the region or a characteristic itself.

7.1 Limitations of this study

In this thesis, we performed mainly macro- and meso-level analysis and not micro-level analysis. Popularity is a really sensitive topic that can be affected by many things. Micro-level analysis might help to identify more features that influence hosts' popularity on the platform, but it hasn't been done explicitly in this study.

Also, despite that we were careful in setting three categories of popularity, they were still arbitrary. This could affect our results. Therefore, having a more precise framework of grouping can help to eliminate this concern.

7.2 Future work

Taking into account all said above, for the future work, we would like to:

1. **Investigate qualitative characteristics in more details.** The analysis of nominal and ordinal variables of the dataset can help to draw more conclusions. In the future work, we would like to look into that domain more.
2. **Collect data from more cities.** Throughout this study, we noted that the population of the city in the real world is not always a guarantee of its popularity on Couchsurfing. Cities with fewer people, can still have a very high number of active hosts. In the future work, we would like to have a bigger dataset that would help to include more of these cases.
3. **Combine this work with survey-based analysis.** Combining a quantitative with qualitative data from Couchsurfing community members would help to make the analysis in the future work more complete and can also lead to some interesting findings.

This research discussed the topic of hosts' popularity on Couchsurfing. While making this thesis, it became clear that this is a very complicated subject that needs to be researched thoroughly. We hope that taking into account the results of this thesis, the future work will be able to lead to even more interesting findings on this topic.

References

- [1] A history of why people travel. <https://matadornetwork.com/bnt/a-history-of-why-people-travel/>.
- [2] Direct and total contribution of travel and tourism to the global economy from 2006 to 2019. <https://www.statista.com/statistics/233223/travel-and-tourism--total-economic-contribution-worldwide/>.
- [3] Direct contribution of travel and tourism to GDP in leading countries worldwide in 2019. <https://www.statista.com/statistics/292461/contribution-of-travel-and-tourism-to-gdp-in-select-countries/>.
- [4] Gross Domestic Product: An Economy's All. <https://www.imf.org/external/pubs/ft/fandd/basics/gdp.htm>.
- [5] Tourism Industry; Everything You Need to Know About Tourism. <https://www.revfine.com/tourism-industry/>.
- [6] Hospitality: A Complete Overview of the Hospitality Industry. <https://www.revfine.com/hospitality/>.
- [7] Couchsurfing. <http://couchsurfing.com/>.
- [8] What is gift economy? <https://matadornetwork.com/bnt/a-history-of-why-people-travel/>.
- [9] I heard of someone charging for a couch. Is that OK? <https://support.couchsurfing.org/hc/en-us/articles/200639420-I-heard-of-someone-charging-for-a-couch-Is-that-OK->.
- [10] Couchsurfing articles from since 2019. <https://bit.ly/2W6Op6F>.
- [11] About Couchsurfing. <https://www.couchsurfing.com/about/about-us/>.
- [12] Qiuju Luo and Hui Zhang. Building interpersonal trust in a travel-related virtual community: A case study on a guangzhou couchsurfing community. *Tourism Management*, 54:107–121, 2016.
- [13] Jiwon Jung, Susik Yoon, Seunghyun Kim, Sangkeun Park, Kun-Pyo Lee, and Uichin Lee. Social or financial goals?: Comparative analysis of user behaviors in couchsurfing and airbnb. pages 2857–2863, 05 2016.
- [14] Airbnb. <https://www.airbnb.com/>.

- [15] Michael O'Regan and Jaeyeon Choe. Managing a non-profit hospitality platform conversion: The case of couchsurfing. com. *Tourism Management Perspectives*, 30:138–146, 2019.
- [16] Maura Rae Cherney. Surf's up: Communicative aspects of online trust-building via reducing uncertainty online in couchsurfing. Theses and Dissertations, 2014. <https://ir.library.illinoisstate.edu/etd/162/>.
- [17] Alexander Ronzhyn. Online identity: constructing interpersonal trust and openness through participating in hospitality social networks. *The Journal of Education Culture and Society*, 1:47–56, 06 2013. https://www.researchgate.net/publication/268979069_Online_identity_constructing_interpersonal_trust_and_openness_through_participating_in_hospitality_social_networks.
- [18] Alexander Ronzhyn and Eugenia Kuznetsova. Conveying the message of trust through written texts in couchsurfing.org. *Synthesis: an Anglophone Journal of Comparative Literary Studies*, page 92, 05 2014. https://www.researchgate.net/publication/328883913_Conveying_the_Message_of_Trust_through_Written_Texts_in_CouchSurfingorg.
- [19] How can I submit references and feedback for other members? <https://support.couchsurfing.org/hc/en-us/articles/211602518-How-can-I-submit-references-and-feedback-for-other-members->
- [20] What is the difference between Surf, Host and Personal references and Feedback? <https://support.couchsurfing.org/hc/en-us/articles/212281057-What-is-the-difference-between-Surf-Host-and-Personal-referen>
- [21] The definition of popularity. <https://dictionary.cambridge.org/dictionary/english/popularity>.
- [22] Douglas W Nangle, Cynthia A Erdley, and Joel A Gold. A reflection on the popularity construct: The importance of who likes or dislikes a child. *Behavior Therapy*, 27(3):337–352, 1996.
- [23] John C Masters and Wyndol Furman. Popularity, individual friendship selection, and specific peer interaction among children. *Developmental psychology*, 17(3):344, 1981.
- [24] Helen Lois Koch. Popularity in preschool children: Some related factors and a technique for its measurement. *Child Development*, 4(2):164–175, 1933.
- [25] William M Bukowski. Popularity as a social concept. *Popularity in the peer system*, pages 3–24, 2011.

- [26] Bernard C Reimann and Yoash Wiener. Corporate culture: Avoiding the elitist trap. *Business Horizons*, 31(2):36–44, 1988.
- [27] Helmut Norpoth. Economics, politics, and the cycle of presidential popularity. *Political Behavior*, 6(3):253–273, 1984.
- [28] John E Mueller. Presidential popularity from truman to johnson 1. *American Political Science Review*, 64(1):18–34, 1970.
- [29] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. The world-wide web. *Communications of the ACM*, 37(8):76–82, 1994.
- [30] Internet Usage Statistics . <https://internetworldstats.com/stats.htm>.
- [31] TripAdvisor . <https://www.tripadvisor.com/>.
- [32] Karen L Xie, Chihchien Chen, and Shinyi Wu. Online consumer review factors affecting offline hotel popularity: evidence from tripadvisor. *Journal of Travel & Tourism Marketing*, 33(2):211–223, 2016.
- [33] Tamsin C Marriott and Tom Buchanan. The true self online: Personality correlates of preference for self-expression online, and observer ratings of personality online and offline. *Computers in Human Behavior*, 32:171–177, 2014.
- [34] Tim Blumer and Nicola Döring. Are we the same online? the expression of the five factor personality traits on the computer and the internet. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 6(3), 2012.
- [35] Leora Trub. A portrait of the self in the digital age: Attachment, splitting, and self-concealment in online and offline self-presentation. *Psychoanalytic Psychology*, 34(1):78, 2017.
- [36] Instagram. <https://www.instagram.com/>.
- [37] Youtube. <https://www.youtube.com/>.
- [38] Emilio Ferrara, Roberto Interdonato, and Andrea Tagarelli. Online popularity and topical interests through the lens of instagram. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 24–34, 2014.
- [39] Dustin J Welbourne and Will J Grant. Science communication on youtube: Factors that affect channel and video popularity. *Public Understanding of Science*, 25(6):706–718, 2016.
- [40] TikTok. <https://www.tiktok.com/en/>.

- [41] Qiyang Zhou. *Understanding User Behaviors of Creative Practice on Short Video Sharing Platforms—A Case Study of TikTok and Bilibili*. PhD thesis, University of Cincinnati, 2019.
- [42] Andrew N Smith, Eileen Fischer, and Chen Yongjian. How does brand-related user-generated content differ across youtube, facebook, and twitter? *Journal of interactive marketing*, 26(2):102–113, 2012.
- [43] Linguistic Analysis Explained. <https://goascribe.com/linguistic-analysis-explained/>.
- [44] De-Jung Chen. Couchsurfing: Performing the travel style through hospitality exchange. *Tourist Studies*, 18(1):105–122, 2018. <https://journals.sagepub.com/doi/full/10.1177/1468797617710597>.
- [45] Xiao Liu. Social networking sites’ influence on travelers’ authentic experience: a case study of couchsurfing, 2013. <https://jewlscholar.mtsu.edu/handle/mtsu/3682>.
- [46] Global Destination Cities Index 2019. <https://newsroom.mastercard.com/wp-content/uploads/2019/09/GDCI-Global-Report-FINAL1.pdf>.
- [47] What is Web Scraping? <https://www.techopedia.com/definition/5212/web-scraping>.
- [48] What is Web Crawler? <https://www.techopedia.com/definition/10008/web-crawler>.
- [49] Nikolay Derkach. Couchsurfing Python API repository. <https://github.com/nderkach/couchsurfing-python>.
- [50] What is an API? <https://www.howtogeek.com/343877/what-is-an-api/>.
- [51] What is Quartile in Statistics? <https://explorable.com/quartile/>.
- [52] Statistics and Histograms. <https://www.dummies.com/education/math/statistics/statistics-and-histograms/>.
- [53] Correlation Coefficient. <https://www.investopedia.com/terms/c/correlationcoefficient.asp>.
- [54] A comparison of the Pearson and Spearman correlation methods. <https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/a-comparison-of-the-pearson-and-spearman-correlation-methods/>.

- [55] Spearman's correlation. <http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf>.
- [56] Average income around the world. <https://www.worlddata.info/average-income.php>.
- [57] 10 Small Countries That Score Big With Travelers. <https://www.neverstoptraveling.com/small-countries>.
- [58] Visa policy. https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/visa-policy_en.

Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Jan Moppel**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Welcome to my couch: Why some people attract more guests than others on Couch-surfing?,

supervised by Rajesh Sharma.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Jan Moppel

15/05/2020