UNIVERSITY OF TARTU
Institute of Computer Science
Innovation and Technology Management Curriculum

Maria Smirnova

HACKATHONS ARE GOING ONLINE. WHAT'S CHANGED?

Master's Thesis (20 ECTS)

Supervisor: Alexander Nolte,
Associate Professor

Tartu 2022

Title: Hackathons are going online. What's changed?

Abstract: The number of hackathons significantly increased in the past few years. The growth spurt in online hackathons took place in 2020, significantly outweighing the number of offline competitions. Many studies focused on the advantages and disadvantages of online competitions since this has been a topical issue lately, especially in connection with the COVID-19 breakthrough that resulted in restrictions in face-to-face gatherings. While prior work mainly focuses on reports of single events and provides qualitative information on hackathons, it is not clear whether these statements can be generalized for a large number of competitions. We, therefore, add to these studies large-scale quantitative research on different aspects defining the differences between online and offline hackathons. Our findings indicate that hackathons differ in duration of the submission, judging stages and total length. The results also show a significant difference in themes and judging criteria and the absence of distinctions of technologies in use and participants' skills. Moreover, online hackathons' average spatial and time zone distances are higher, indicating a more significant geographic diversity. While we also can see a trend towards an increase in various participating countries, both online and offline.

Keywords: Devpost, hackathon, analysis, online, offline

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

Pealkiri eesti keeles: Häkatonid lähevad *online-i*. Mis muutus?

Lühikokkuvõte: Häkatonide arv kasvas oluliselt viimase mitme aastaga. Online-häkatonide kasvuspurt leidis aset 2020. aastal, ületades oluliselt offline-võistluste arvu. Paljud uuringud fokuseerivad online-häkatonide eelistel ja puudustel, kuna see on viimasel ajal aktuaalne teema, eriti seoses COVID-19 viirusega seotud piirangutega, mille tulemuseks on piiratud näost näkku koosolekud. Kui eelnevad tööd fokuseerivad enamasti üksikute sündmuste raportitele ning annavad kvalitatiivset informatsiooni häkatonidest, ei ole selge kas need väited saab üldistada suurema võistluste arvu jaoks. Seetõttu me lisame nendele uurimustele mastaapsemat kvantitatiivset uuringut erinevate aspektide kohta, defineerides erinevused online- ja offline-häkatonide vahel. Meie tulemused näitavad, et häkatonid erinevad esitamisperioodi kestvuses, hindamise etapides, ja üldises kestvuses. Tulemused näitavad ka olulist erinevust teemades ja hindamise kriteeriumites, ja kasutatud tehnoloogiate ja osalejate oskuste erisuste puudumist. Pealegi online-häkatonide keskmised ruumilised ja ajalised kaugused on kõrgem, näidates olulist geograafilist mitmekesisust. Samal ajal me näeme ka osalevate riikide arvu suurendamise trendi, nii online- kui ka offline-üritustel.

Võtmesõnad: Devpost, Häkaton, analüüs, online, offline

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Table of Contents

1. Introduction

Hackathons are brief, intense competitions where people gather and build teams intending to create an innovative solution and complete tasks of interest to them and provide value for the entire community.(Hackathon Planning Kit, n.d.) Participants might have different backgrounds and knowledge (Hackathon Planning Kit, n.d.; Wang et al., 2021); moreover, they may come from different parts of the world (Bertello et al., n.d.; Braune et al., 2021; Franco et al., 2021; Gupta & Rubalcaba, 2021) and still collaborate on a given task. Hackathons are an essential field of study because of the phenomenon to be beneficial on a wide scale (Cobham et al., 2017; García & Menéndez, 2021) for all: participants (Wang et al., 2021; Yarmohammadian et al., 2021), organizers, industries, and educators taking into consideration all advantages it provides (10 Reasons Why Hackathon Is Important to Make You Industry-Ready | LinkedIn, n.d.). Lately, such competitions have become especially popular, which can be at least partly explained by the ability to address a broad scope of issues (García & Menéndez, 2021) such as environmental, health, social, and others, while at the same time encouraging learning and collaboration between people, fostering the creation of innovative solutions for different spheres and promote education and entrepreneurship (Wang et al., 2021; Yarmohammadian et al., 2021),. With all the restrictions on public gatherings due to the COVID-19 pandemic (COVID-19 Transmission and Protective Measures| WHO Western Pacific, n.d.), we have seen more events online that affected hackathons too. One would expect online events to attract participants from different geographic regions (Braune et al., 2021; Gupta & Rubalcaba, 2021), which will affect collaboration within each team, the tools used, and the length of the hackathon. At the same time, it can be expected that the topic of hackathons will also be biased towards issues related to health and covid (Braune et al., 2021; Gupta & Rubalcaba, 2021).

Lately, a lot of research has been focused on hackathon organizational issues and the advantages and disadvantages of fully online competitions. However, there are only limited scientific studies about quantitative changes brought about by the spread of online hackathons. Therefore, the main objective of this work is to investigate quantitative differences between online and offline hackathons, including event organization, participants distribution, trending topics, and tools in use. This study aims to improve the hackathons development process by gaining numerical insights on online competitions and complement existing theoretical works.

The first research question consequently addresses the stages of hackathons.

RQ1. What are the main steps in online and offline competitions? The main aim is to understand how different the processes are, the main operational steps, the length of each step stated in the schedule, and, therefore, the duration of a hackathon in total.

RQ2. How have themes of the events changed during the transition from offline to online ones? We expect that there might be a slight shift towards social goods and health, which will be mainly due to the pandemic situation and more significant concern about these topics than ever before.

Nowadays, online hackathons spread led to some changes that are actively discussed in scientific works. In principle, they enable participation from anywhere globally (Braune et al., 2021; Franco et al., 2021), but it is not clear whether this is actually happening because there is no sufficient numerical evidence. The second research question, therefore, would be as follows:

RQ3. How do offline and online hackathons vary geographically? The third research question focuses on understanding how different hackathons in terms of distance between participants inside one group and whole hackathon, as well as how different they are in terms of participating countries variety. This leads to also investigating the following and the last research question:

RQ4. What tools are used to create a project solution for hackathons in online and offline events? This research question is closely connected to the previous one since teams formed by participants from different countries have to collaborate using cloud-based technologies more than during offline events to create solutions together.

2. Background

To go deeper into the study's context and define the current state of the art, we analysed related literature. In this section, the work will be considered in the context of previous research on hackathons and the differences between online and offline types, as well as preceding analysis of geographically diversified groups and teams.

Hackathon is a competition where for a short period (in most cases, no more than one week), people with different backgrounds and skills get together in teams and work on the designated problem by developing an innovative solution. Hackathon is given the following definitions in the scientific literature: "…an event in which computer programmers and others involved in software development collaborate intensively over a short period of time on software projects" (Briscoe & Mulligan, n.d.), it "…attracts individuals who collaborate closely during a fast-paced and short time period, where they define a problem and produce a prototype that provides a solution to the problem" (Wang et al., 2021), "type of open innovation"(Soltani et al., 2014) in the form of "the ideation contest" (Soltani et al., 2014) or "…a worldwide phenomenon, with both industry and educators considering the opportunities and benefits that they generate"(Cobham et al., 2017). These definitions, however, focus a lot on software development, which is currently not a rule of thumb for modern hackathons. A more modern definition states that a hackathon is "… short intensive event during which people come together, form teams and attempt to complete a project that is of interest to them. Teams are usually collocated and often composed of people with diverse backgrounds, experience, and expertise" (Hackathon Planning Kit, n.d.). To sum up, a hackathon is mutually beneficial, both for organizers and participants, ideation contest, the main aim of which is to create innovative solutions in the form of a prototype to the stated problem in a short time by a group of individuals.

Hackathons also "reinforce active learning, creative thinking, multidisciplinary team working, innovation, and generation of new ideas" (Yarmohammadian et al., 2021). Participants are usually motivated to take part in such events by gaining knowledge, socializing with other participants, networking, earning a reputation, and what is equally essential affiliating to "cultural «hacker identity» objectives" (Wang et al., 2021). While working on the project, participants also improve collaboration, integration, communication skills and gain presentation and pitching experience in front of a large audience.

Originated from hackers' environment, hackathons now "…have evolved to extend to all professional and academic fields" (García & Menéndez, 2021) and cover diverse themes such as social goods, education, health, technologies, and others. There are also varied areas of implementation – from IT and engineering to public health and social justice, and even various criteria for judging such as creativity, usability, functionality, originality, and others. Therefore, participants "…tend to be diverse in skillsets" (Wang et al., 2021) and backgrounds. Hackathons are popular among companies and enterprises that search for innovative ideas. Moreover, "...more and more universities are launching this type of initiative" (García & Menéndez, 2021)

There are two main types of hackathons based on location: online and offline. Offline is also called physical, and they are usually "…constrained by physical realities where participants have to physically gather in the same place at the same time" (Wang et al., 2021). In contrast, online events are such where physical presence is omitted, and participants can connect to the event from anywhere in the world if there is appropriate equipment and internet connection (as a bare minimum). Here it is also worth mentioning the potential time zone issues as far as the difference in time considered as an “important factor that influences distributed collaborations” (Wei, 2007) and may affect participants in different aspects starting from setting up an adequate team meeting time to selecting collaborations channels, that might be more convenient for the virtual team but not as effective (for example, it is more common for online teams with diverse time zones to use email rather than video calls). (Wei, 2007) Overall, it has commonly been stated that the more significant the difference, the more the hindrance to working together productively (Ford et al., 2017; Olariu & Aldea, 2014)

COVID-19 pandemic clearly “…led to a cultural change in work settings, communication, and collaboration” (Braune et al., 2021) – many events have to be either cancelled or moved online, as well as work and study settings transformed to remote. If before the pandemic, online communication and collaboration seemed to be optional and going beyond what is necessary to get the job done now, it is inevitable, so-called “new normal (COVID-19 Transmission and Protective Measures| WHO Western Pacific, n.d.). The organization of online events differs significantly from face-to-face ones. However, it has been already proved that fully online competitions can bring people together “not only to collaborate in an interdisciplinary manner but also across diverse geographic locations, avoiding the possible inconveniences and costs of travel” (Braune et al., 2021), which in turn can increase innovativeness and creativity of solutions.

2.1 Online and offline hackathons

A lot of studies such as (Bertello et al., n.d.; Braune et al., 2021; Franco et al., 2021; Gupta & Rubalcaba, 2021; Loeffler & Masiga, 2021a; Usher & Barak, 2020; Wang et al., 2021) and others stated the potential as well as limitations of online competitions. Among the main restraints, there is a lack of time, especially in the implementation phase (Loeffler & Masiga, 2021a), the challenge for additional planning and framework, communication between team members is more complicated than during in-person events, it is complex to keep track of extracurricular activities and much easier to lose the attention and interest of the participants (Powell et al., 2021). It is more complicated to team up with other participants due to the lack of each other’s background knowledge (Wang et al., 2021). Moreover, due to the nature, online competitions can be complicated by technical issues arising from both the participants and the organizers (Powell et al., 2021) that can easily distract participants from completing their tasks. While it is relatively easy to “discuss functional things” (Bertello et al., n.d.), deliberation of inventive ideas becomes complicated. Lack of so-called “icebreaker and team building activities” negatively affect the engagement of participants that in turn negatively affect “experience regarding group synergies and team spirit during the event” (Braune et al., 2021). The absence of in-person communication in such events “could make team members losing interest for the development of their projects” (Franco et al., 2021) and make them disengage from the events. Teams are also negatively affected by lack of trust between team members, low level of familiarity between co-workers, misinterpreting peers ideas, and poor communication, including the absence of non-verbal one (facial expressions, gestures, and others) that is a “key source for gathering effective information among engineering team students, such as unclear points, requests for clarification, and acceptance of another team member' side” (Usher & Barak,

2020) and can be mitigated in a face-to-face environment. There is also the problem of time zone difference for virtual competitions' participants that form a team from different parts of the world (Wang et al., 2021). Furthermore, some authors emphasize the difficulty of online mentoring caused by tools and hardware limitations. Doubtless, during online events, it is harder to "…immediately empathize with who you have to mentor" (Franco et al., 2021). There is a lack of in-person familiarity with a team and current team members' state vision (that is helpful to detect confusion, frustration or other emotions) (Nolte et al., 2020). Therefore, the non-availability of real-time support might "impede their [team members] work" (Usher & Barak, 2020). Cloud collaboration tools could be handy in this case. For example, organizers "…may provide an online instant messaging system, such as Slack, to field participant questions and coordinate mentor efforts" (Braune et al., 2021). However, various online platforms may confuse participants due to their interoperability (Braune et al., 2021).

Additionally, there are certain limitations such as data security, especially with used in large numbers video and audio communication tools. Just recently, "Zoom had been slow to address security vulnerabilities" (Paul, 2020). Given the preceding, online hackathons require the most precise preparation, including fundamentally prepared technologies and tools highly in advance before a start date "not only to manage the hackathon but also to favour an effective flow of knowledge and interactions among the different parties" (Franco et al., 2021).

Meanwhile, it became apparent that online hackathons can contribute significantly to solving issues that have arisen due to the COVID-19 breakthrough in the world (Braune et al., 2021). It is stated that online competitions push forward team and mentors diversification, including "cross-region collaboration" (Braune et al., 2021) since both could participate regardless of location and require less time and material costs than offline events, which in turn leads to such advantages for parties as "access to the diverse competencies of a large number of participants and virtual spatial proximity with global attendees" (Gupta & Rubalcaba, 2021). The potential of online hackathons has been already unleashed by an increasing number of competitions and "resulting pandemic solutions" (Gupta & Rubalcaba, 2021). From participants' point of view, remote events give "enormous opportunity…to take advantage of a pool of ideas, knowledge, and relationships that would not have been possible in a physical context" (Bertello et al., n.d.). Accessibility of such events also enlarged due to the availability of meetings records that can be shared among participants who were unable to attend sessions or want to rewatch it at a more convenient time (Loeffler & Masiga, 2021b). Not only country diversification but also skillset of participants noted as clear advantage of online events. (Braune et al., 2021) It is worth mentioning that many competitions' participants highlight that the scope and relevance of online challenges surpass ones of offline hackathons. (Braune et al., 2021) The swiftness of troubleshooting in the mentoring process noted by participants as a positive aspect of online events as far as "mentors can help participants with flexible and short mentorship sessions without altering their working schedules and personal time" (Braune et al., 2021) as well as even located in different regions with participants advisers still could provide needed support which is not the case for in-person events.

Due to their nature, online events also provide more flexibility in monitoring teams' progress (Wang et al., 2021). In addition, in traditional competition, the team is assembled based on familiarity with other members in most cases. In contrast, virtual events are more based on "skills required to solve the pitched problem" (Wang et al., 2021), which can boost problem-solving and lead to more innovative ideas. Table 1 below summarizes the core advantages and disadvantages of online hackathons.

Table 1. **Advantages and disadvantages of online hackathons**

	Advantages	**Disadvantages**
Organization	Lower time costs Lower material costs Wider accessibility	Lack of time in the implementation phase Additional planning and framework needed Complex to keep track of extracurricular activities Require the most precise preparation including fundamentally prepared technologies and tools highly in advance before a start date
Communication	Easy to discuss functional aspects Access to the diverse competencies of many participants Virtual spatial proximity with global attendees	Deliberation of inventive ideas is complicated Lack of team building activities Absence of in-person and non-verbal communication Lack of trust Low level of familiarity between team members Higher risk of misinterpreting peers' ideas
Engagement	Scope and relevance of hackathon	Easier to lose the attention and interest of the participants

Participants	Team and mentors diversification including "cross-region collaboration" The diverse skill set of participants Teams are assembled based on skills required to solve the problem Wider networking	It may be hard to form a team due to lack of each other's background knowledge Time zone difference
Technologies	Variability of cloud collaboration tools Availability of meetings records	Technical issues Tools and hardware limitations Online platforms interoperability Data security issues
Mentoring	Swiftness of troubleshooting Flexible mentorship without altering working process and personal time Flexibility in monitoring teams' progress	Lack of in-person familiarity with a team Non-availability of face-to-face support

Many differences between online and offline hackathons were already discussed in the scientific literature, from organizational to mentoring aspects. However, prior work mainly focuses on reports of single events, and it is not clear whether and how findings are generalizable beyond the studied events. Our work will focus more on quantitative differences between online and offline hackathons in such aspects as an organization by analyzing scheduled stages of hackathons, participants, especially on geographical diversity and skillset, technologies in use, and overall thematic of hackathons. Thus, we will be able to confirm or refute existing assertions about online and offline hackathons based on large scale data gathered from Devpost and add new findings.

2.2 Diversity within a team

It is noted that due to the online nature, there are not only a more significant number of participants, better opportunities for collaboration, less logistical, material, geographical, and time costs and constraints (Braune et al., 2021; Franco et al., 2021) but also more possibilities to create a multinational team. Both positive and negative effects on the team's performance have "variations in group demography" (Phillips & O'Reilly, 1998). On the one hand, it is stated that in teams with higher cultural diversity, there are more "informational advantages" (Bouncken et al., 2016) ; thus, broader information sources are used to do more profound research for complex tasks and issues. Moreover, cultural and geographic diversity leads to creating of a more helpful and dedicated climate inside the team that leads to better teamwork which is an essential factor for creating innovations (Bouncken et al., 2016). It is also stated that "deep-level political diversity predicted greater idea novelty" (Coursey et al., 2020). Therefore it is suggested that teams prosper from the differences of views and attitudes while working on the same project. However, most likely high diversification will violate the functioning of a team and lead to emotional conflicts that will make task implementation more difficult. (Phillips & O'Reilly, 1998). Therefore, actions, such as careful and sustained mentoring, should be taken by hackathons organizers to eliminate the harmful effects of highly diverse teams. There are also side effects from the time difference (if working in online mode from different locations) miscomprehending the context of team members' ideas, and "power of distance in particular" (Bouncken et al., 2016).

Although geographical diversification can affect online and offline teams differently, there are some empirical pieces of evidence about "…significant correlations between the overall team diversity and the innovation level of team projects"(Usher & Barak, 2020) that can be explained by higher creativity and more innovative ideas related to "collaboration with people from diverse backgrounds" (Usher & Barak, 2020).

To summarize what was revealed above, diversity in teams "has both—effects that enhance creativity and innovation as well as effects that are harmful to the quality of teamwork and thus for creativity and innovation" (Bouncken et al., 2016). Therefore, higher attention should be paid to teams with greater geographic diversification. Table 2 below summarizes the core advantages and disadvantages of geographically diverse teams.

Table 2. **Advantages and disadvantages of geographically diverse teams**

Advantages	**Disadvantages**
Informational excellence	Harder to function
The more helpful atmosphere that enhances teamwork	Higher risk of conflicts
Idea novelty	Time and location difference (in online setting)
Greater creativity due to different backgrounds	Miscommunication
	Power of distance

To sum up, prior works mainly were focused on aspects that affect teams due to high spatial and time zone diversity. Knowing that there are both positive and negative effects, we, however, do not know yet how exactly hackathons differ geographically. There are more opportunities to create a multinational team by members from different parts of the world, but how precisely do teams online differ in terms of spatial intergroup or time zone distance from offline ones. Is this difference significant to consider while creating new hackathons or not?

After a more detailed study of the literature, three hypotheses related to the third research question about the geographical diversity of participants were formed:

H1 Diversification: It is expected that there would be more participants from all over the world due to decreased time, monetary and psychological barriers related to relocation.

H2 Expanding spatial boundaries: Projects of online hackathons have a higher chance to be formed by people from different countries. This statement follows the previous hypothesis and the fact that teams must be formed virtually in online events, and there are more possibilities to connect with people from other countries.

H3 Expanding time boundaries: As far as more teams in online hackathons are expected to be formed by people from different countries (Braune et al., 2021; Franco et al., 2021), it is also assumed that while being in different locations, team members are more likely to be in different time zones as well, which in turn will increase time zone differences inside the team

3. Methodology

This chapter will focus on the methodology used in this work. The first part will cover the data collection process. The second part will be devoted to data cleansing. And finally, the last third part is data analysis.

3.1 Data Sources

Devpost (Devpost - The Home for Hackathons, n.d.) was selected as the primary data source because it is one of the most extensive hackathon databases that is used by different actors such as universities, corporations, companies to promote hackathons and engage more people to participate in competitions. It contains general information about hackathons such as schedule (most importantly start and end dates), the total number of participants, number of projects created in the hackathon, location, judging criteria, and prizes available. Moreover, the information about projects created by participants can be retrieved from the site that includes GitHub (or another repository) link, number of participants, technologies used to build the project, likes, and comments. It is also worth mentioning that individual participants' data is also available in Devpost that includes names, skills of a person, number of projects done, number of hackathons in which this person has participated, achievements, followers, and location of the person. To sum up, Devpost contains all the necessary information to answer the research questions stated in this work. However, not all the data is in a perfect state. For example, location data that is needed for research question 3 is presented in a pretty unstandardized format - different languages, different units (country, city, state), and different order of mentioning – which make it difficult to dive deeper into data insights. This exact problem was solved by using geopy, "Python client for several popular geocoding web services"(Welcome to GeoPy's Documentation! — GeoPy 2.2.0 Documentation, n.d.), Nominatim tool that allows searchingOpenStreetMap data by name and address or by latitude and longitude (both methods were used and will are described below).

3.2 Data Collection and Cleaning

This part will be devoted to collecting required data to answer stated research questions as well as the details of the filtering/cleaning. To retrieve the data from Devpost, the author wrote a python script available the GitHub (Smirnova Maria, 2021/2021) repository (see Table 3).

Table 3. **Description of related python notebooks**

Notebook name	**Description**
Data web scrapping final	Obtains information about entities and saves into csv files.
Geopy	Participant location identification code.
Data cleaning and analysis	A notebook for data processing, metrics calculation and hypothesis testing.
Data visualisation	Contains chart visualizations and cloud of words creation.
Map animation	Creates a mp4 video animation of participants location on a worldmap with division of the points by color.

With the web scraping, we managed to obtain the following six information parts that are crucial for our research:

- General information about hackathons
- Project gallery of each hackathon
- Each hackathon individual details
- Each project individual details
- Participants individual details
- Participants’ locations

General hackathons information was obtained using request Python library by sending get request to Devpost hackathons listing page URL with the additional page number parameter set. Here are some of the details that we gathered after this step: id, title, organization name, and most importantly, project gallery URL. The next step is to obtain more informative descriptions of the hackathons by parsing HTML pages with BeautifulSoup Python lib and searching for those details in particular HTML elements. At this point, we have stumbled upon one bug problem: the parsing code was too computational heavy for our machines, and after some number of iterations through the list of pages of hackathons, the machines ran out of memory. It was decided to separate the entities (hackathon, project, participant) into chunks of a specific predefined size, which varies from entity to entity. After the parsing code finishes execution on a particular chunk, the resulting information is saved into separate .csv of length smaller or equal to the chunk size. While operating in the described manner, we gathered detailed information like judging criteria, prizes, online/offline status, etc., from hackathons’ HTML pages. The same approach was used to obtain projects detailed information: GitHub (or other) repository links, participants profile links, technologies used, etc. We increased the chunk size from 1000 to 5000 units for the participants' detailed information extraction. We obtained the following elements from participants’ profiles: name, skills and location of residence.

As was described earlier, location information stated by participants in some cases has a low quality and is not suitable for data aggregation or distance calculation. Locations can be written in various languages or even in a mix of languages. For example, “Ari'el” stands for Israel and “O`zbekiston” for Uzbekistan. To calculate the difference between locations and standardize names of geographical objects geopy library was used. We searched for a stated location in Open Street Map using nom. geocode to get latitude and longitude first for full location (e.g. Yaroslavl, Yaroslavskay oblast, Russia). If the full location was not found, we searched for the first entry, which is usually a city or a state (in the provided example, this entry is Yaroslavl). If the location was still not found, we used the second entry in location (in our example, it’s Yaroslavskays oblast). This solution is not ideal, but it works in most cases. There are some situations when the described algorithm works poorly. For example, if the participant stated that his location is "Tilonia, RJ, India" which is in India, and if Open Street Map didn't find the full location and city, the algorithm will run it on the second entry, which is “RJ”, and it would be latitude and longitude of Rio de Janeiro. Such cases are rare, but they add redundant noise. Finally, the same code was used for the rest of the not found locations, but on the last entry (in our example case, it is Russia) – the data contains only two such cases. In most scenarios, the data was found either containing full location or first entry of location string. Later, we also run the Nominatum to get the English name of a geographical object using the received latitude and longitude from the previous step. This information was needed to check H1 RQ 3. The reverse operation was used to get the full English address and extract city and country from it.

Overall, information about 6150 hackathons, 260656 projects, 280473 unique participants (out of 668660), and 12210 unique participants' locations was collected. While going deeper into the hierarchy levels (from hackathon to project to participant), we had to keep the connection between different datasets. Entity URL and location name (only for location dataset) were selected as a key to establishing relations. An overview of the different datasets that were used to answer research questions is presented in figure 1.

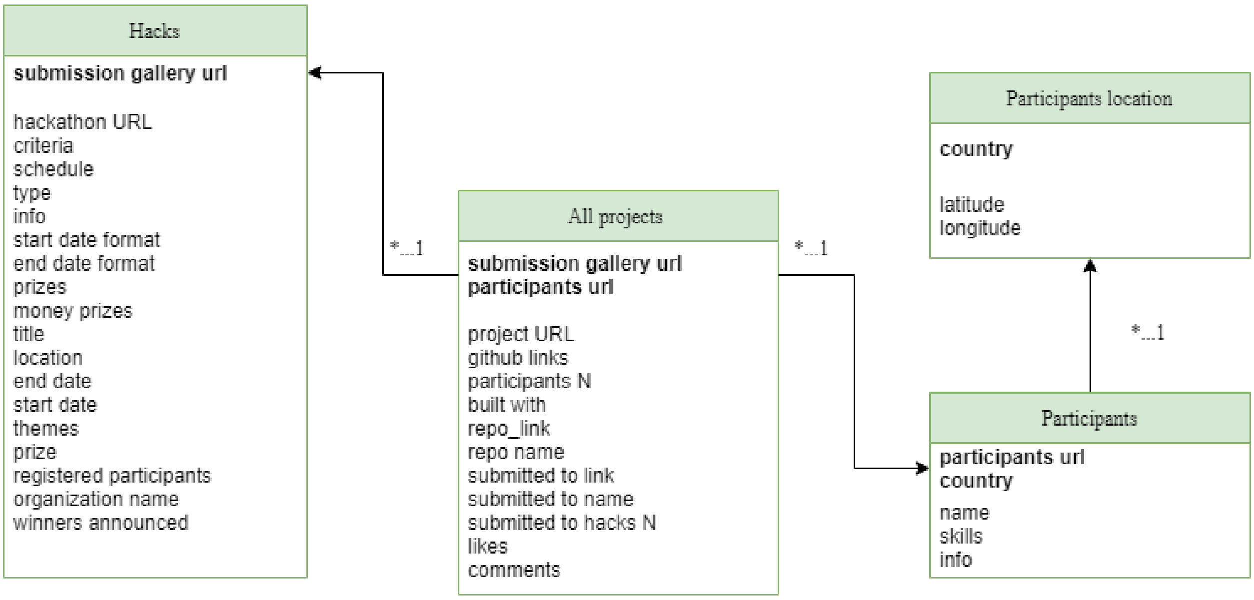

Fig. 1. **Datasets dependence diagram**

We made some conscious decisions about exclusion criteria that we will discuss in the following part (an overview can be found in Table 4). The first step of the data cleaning process was to remove the duplicates. However, it was discovered that a single project could be submitted to multiple hackathons (for example, https://devpost.com/software/break-ly-podic8 was submitted to 2 hackathons: "High Tech Hacks" and "health{hacks}: hackfrom{home} 2021"). We had to consider this situation since all information about projects and a group that created it stays the same except for the hackathon submitted to this project. Therefore, we deleted project duplicates by URL and kept only the first entry (the first hack where this project was submitted). In an example that was presented earlier, only one entry of the project left that was submitted to "health{hacks}: hackfrom{home} 2021x" because it started earlier. After this step, the number of projects in the dataset was reduced to 166895. Hackathons without submission gallery were also removed as we cannot get any information about participants and their projects. Fortunately, from all web-scraped hackathons, only one did not have a submission gallery. To calculate the duration of a hackathon, firstly, start and end dates were transformed to the DateTime format. After an exploration of hackathons duration, it was discovered that some hackathons are longer than or equal to 21 days (fig. 2 left). To be more specific, 639 hacks out of 6149. We decided to remove any competition with a duration longer than three weeks. Additionally, we found that one hackathon had a negative duration because start and end dates were accidentally changed places in the hackathon submission phase. This case was also removed from the dataset. The distribution of hackathons durations after the previous step is shown in fig. 2. right. Projects related to the removed hackathons were also deleted from the dataset.

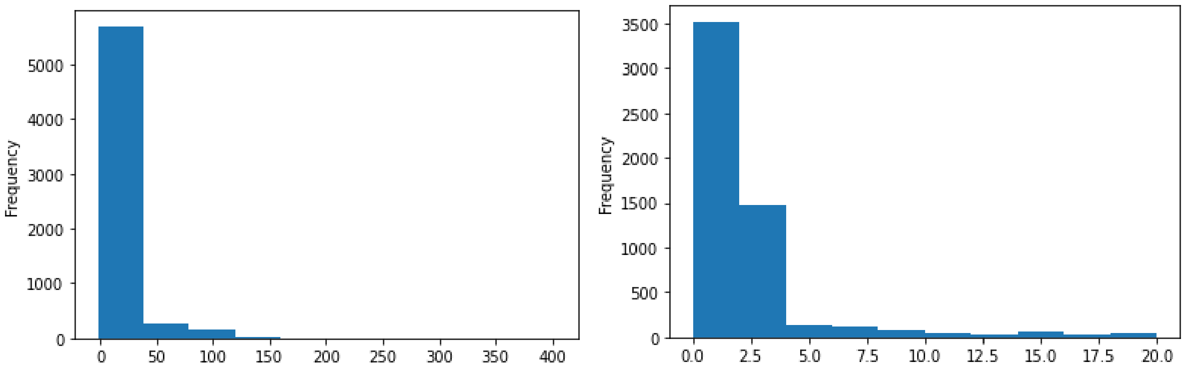

Fig. 2. **Distribution of hackathons' duration (original and cleaned)**

Source: author's calculations

We consider a group to contain at least three persons, so all projects that were done by 1 or 2 participants were removed, which also has led to a decrease in the size of projects down to 74090. Our research questions also contain the calculation of distances (spatial and time zone) between participants. To increase the precision of these calculations, it was decided to remove all participants whose profiles do not contain detailed information about their location (the only country is stated or no location at all). After this step, the number of participants decreased to 199360, and the number of unique locations stated by participants down to 11914. Additionally, we also decided to consider only projects done by groups in which at least 70% of collaborators are present in the filtered participants' dataset.

Table 4. **Exclusion criteria for the dataset cleaning process**

Change source	**Step in the cleaning process**	**Exclusion criteria**	**Number of excluded data points from the related dataset(s)**
Projects	1	Duplicated rows	87615 projects
	2	Duplicated project URLs (1 project submitted to multiple competitions, only first submission was left)	6146 projects
	6	Less than 3 group members	69550 projects
	10	Less than 70% of group members have stated their full location	27478
Hackathons	3	Submission gallery non-existence	1 hackathon
	4	Duration longer than or equal to 3 weeks	639 hackathons & 23050 projects
	5	Negative duration	1 hackathon
Participants	7	No location stated	60918

	8	Only country stated	20195
Locations	9	Only country stated	296

To sum up, the data cleaning step resulted in 5509 hackathons, 46819 projects, and 199360 participants and 11914 participants' locations. Detailed data collection process available in **GitHub** (Smirnova Maria, 2021/2021) repository of the author.

3.3 Data Analysis

To investigate hackathon duration and stages (RQ1), we used the collected data about the start and end date of a hackathon that was extracted from Devpost. The duration was calculated in days by subtracting the end date from the start date and adding 1 to this difference. In some cases, a hackathon might have the same day for start and end date, therefore adding one is crucial to get accurate results (after this step, the duration of such a hackathon would be 1, not 0). We also referred to the schedule of each hackathon that has a form of a string and specified the stages of the event with start and end date-times (example of the string: "Submissions October 22 at 10:30am EAT October 23 at 4:15pm EAT, Public Voting October 22 at 4:15pm EAT October 24 at 4:30pm EAT, Judging October 23 at 4:30pm EAT October 24 at 4:45pm EAT, Winners Announced October 25 at 5:00pm EAT"). The string was split by a delimiter to extract each step information by referring to it by index. The schedule string does not contain the year, so it is taken from the hackathons' start dates. With the stated data, an analysis was conducted to define the average duration of each step and an average distance between stages. During the data analysis phase, it was discovered that in rare cases additional public voting stage is added to the hackathon by the organizers, and it affects the duration of the hackathon. However, the details will be described later in section 4.1. For the last stage, where the winners are announced, there is only the start date, which is logical due to the nature of the announcement process. In some cases, there is no start date or end date for the second stage, but this problem is solved by a mean and std calculating function that ignore all the NaN values. While analyzing the schedule data, it was found out that the stated in Devpost start date and end date are references to the active period of submissions, so it's equal to the first stage of the hackathon. However, in our research question, we are investigating the total duration of the hackathon. To do this, we calculated the total duration of the hackathon as the difference between the date when winners were announced and the start of the submission dates, as the latest and earliest dates available. This approach is suitable because it allows us to investigate the active submission period as well as the total hackathon duration with regards to different stages to be able to conduct analysis in more detail.

To investigate hackathons themes and judging criteria (RQ2), we collected data about hackathons, specifically the type (online/offline), themes, and judging criteria. As far as in most of the cases 1 hackathon can have more than one theme and more than one criteria we split these columns by delimiter, which is, in this case, a comma and stacked back to a column to receive one theme/criteria in 1 row and multiple rows for one hackathon. The hackathons were grouped by themes/criteria, and the percentage rate of each occurrence was calculated by dividing the count of each theme by the sum of all of them.

To study whether there is a significant geographic difference between participants of online and offline hackathons (RQ3), first of all distance matrix between all possible pairs of locations' coordinates was calculated using sklearn.neighbors.DistanceMetric with haversine formula results multiplied by 6371, which is the radius of the Earth in kilometres.

For time zone analysis, it was essential to get the standardized time zones' names by latitude and longitude for which the TimezoneFinder python package was used, the results were added to the participants' dataset.

After this, we created a function that calculates the difference between localized hackathon dates in different time zones for which pytz timezone was used. It's worth pointing out here that we substitute minimum localized date value from a maximum one to avoid any discrepancies, for example, the time zone difference between "America/Los_Angeles" and "America/New_York" on 2021-08-01 can be 3 hours or 21 depending on which of the values is minuend and which is subtrahend (if we subtract date in Los Angeles time zone from date in New York time zone the result would be 3 hours, vice versa it will give 21 hours which is not correct). The defined function also considers DST (daylight saving time) by taking as an input date when the hackathon started, which makes it more accurate.

After this, the function that calculates two types of intergroup distances: in kilometres and in time zone offsets was created. The function receives the following parameters: a data frame of projects, a data frame of participants, a distance matrix (for distance in km), and threshold values (0.7 by default). The first step creates a dictionary where participants represent keys, and their countries or timezones represent values (depending on the type of distance we calculate). It is needed to speed up the following commands by reducing the cost of the reading operation. If some participants' profiles doesn't contain information about the location of a participant, then he or she is not present in the dictionary. Then the function creates a loop iterating through the projects getting their participants.

The next step is to calculate the percentage of the participants in the project whose profiles contain the location information and are present in the dictionary from the previous step. If the percentage of such participants is higher than the threshold, the function will continue to the next step. Otherwise, it is impossible to calculate the distance, and the function returns -1 for this group of participants. If the functions have passed the condition from the previous step, it starts calculating the distances between each pair of the participants and then appends their mean value to the final list of mean group distances, which is returned as the result of the function. The intergroup spatial and time zone distance was grouped by type (offline/online) and averaged (fig. 3).

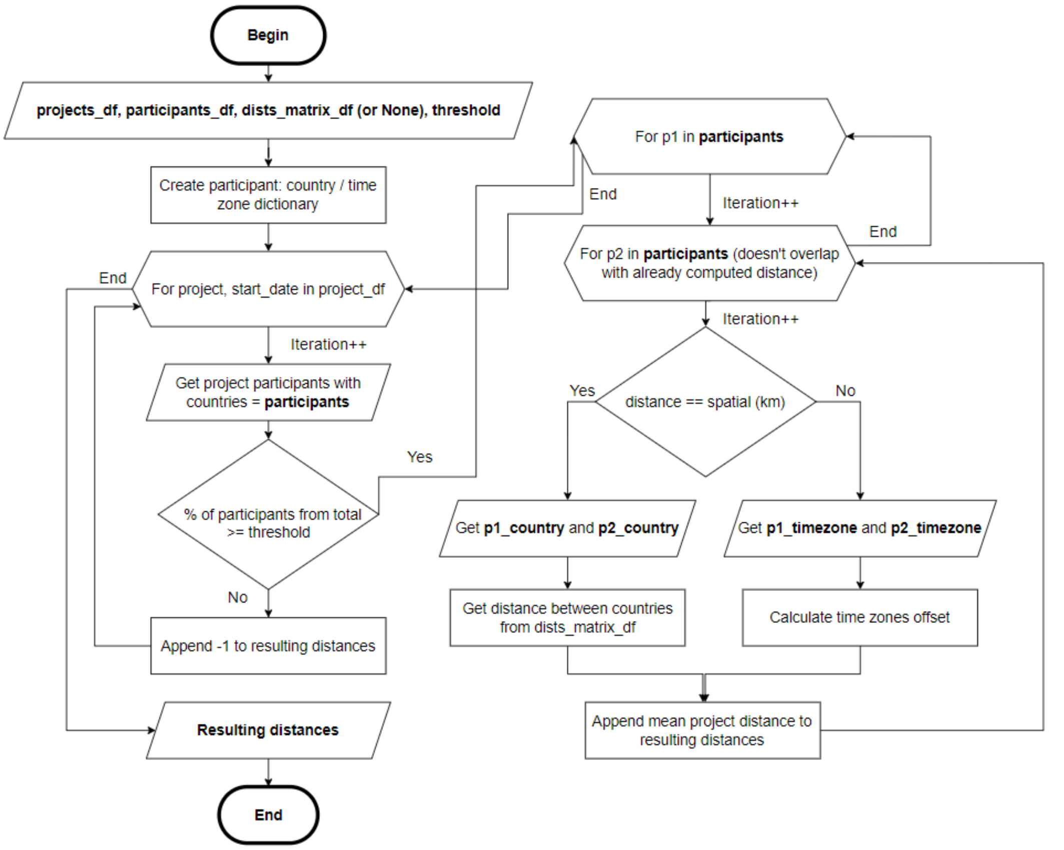

Fig. 3. **Block schema for distance calculation function**

For a better visibility, we also created an animated world map representation for the number of participants in different types of hackathons (fig. 4). It continuously discovers the locations of all participants for the period from 01.01.2013 till 16.10.2021. This visualization helps to get the intuition behind the historical progress of hackathons thought the world, which demographic location contains the most extensive number of participants and the ratios of the hackathon types in those locations.

To create the animation, we had to assemble the dataset from the information about hackathons, projects, and participants that were previously scrapped from the Devpost. It contains the following columns: the date when a participant claimed his participation in a hackathon, the location of a participant, the type of the hackathon, and the coordinates of a participant. The next step was to create a matplotlib figure with a world map background and project the axes into a latitude-longitude coordinate system. The background image was taken from NASA's Blue Marble (NASA Visible Earth - Home, 2022) and was set using the background_img axis function. The axes were created using the Mercator projection with Cartopy. In the second step, we added a function that creates a scatter plot of locations with color division by hackathon type using the historical information from the start date (01.01.2013) till the selected day.

The transformation of latitude and longitude coordinates into appropriate locations on the map was done with ccrs.PlateCarree () projection. In the end, our function saves a png snapshot of the selected day. Then we added description texts like date and participants count using standard functions from matplotlib and ran the function from the second step in the loop from start date till end date to create a dataset of frames for our animation. Finally, we create the video using the ffmpeg program running the following command in the terminal in the folder with frame images: ffmpeg -framerate 30 -pattern_type glob -i '* .png' \ -c: v libx264 -pix_fmt yuv420p out.mp4

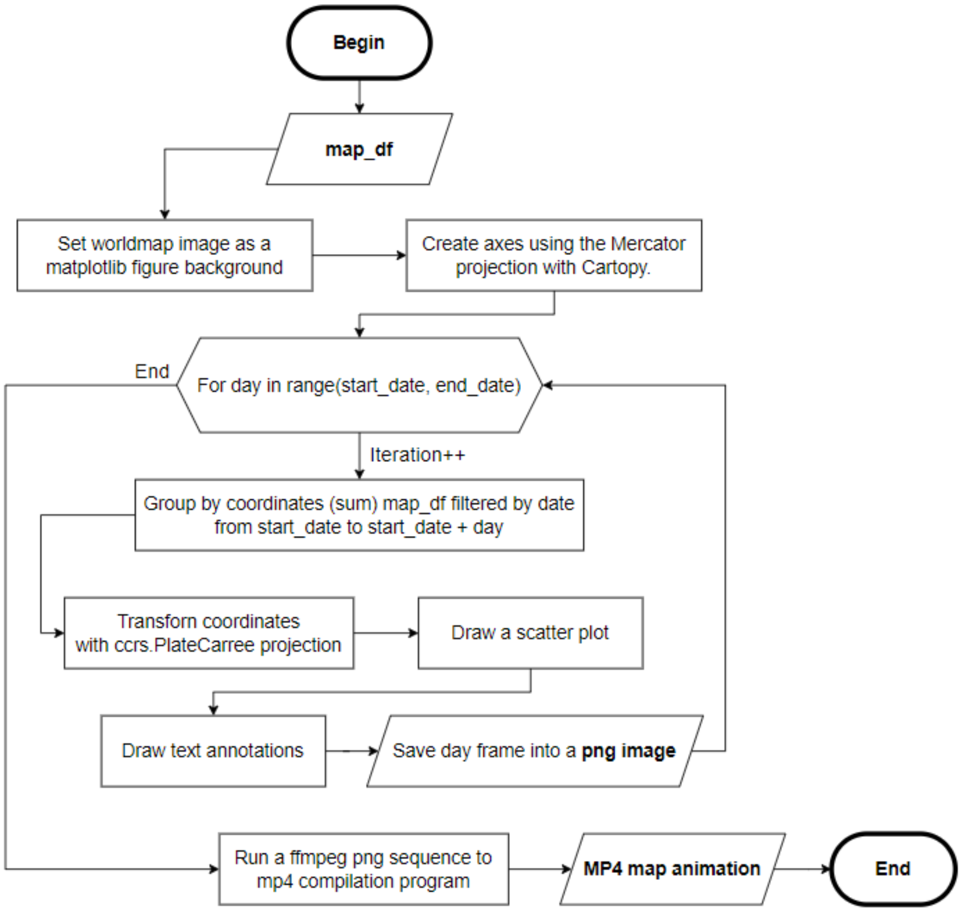

Fig. 4. **Block schema for video map animation function**

To investigate tools and technologies used in hackathons (RQ4), the same approach as for RQ2 was used. However, we used data on projects (tools that were used to build a project) and participants levels (skills that the person uses in hackathons). Furthermore, to dive deeper into GitHub usage, we calculated the ratio of projects where this repository was mentioned. The information about GitHub was collected on the project level by searching for all links and from all links defining only lines that contain GitHub names.

As the final step, we have conducted a two-tailed proportion z-test for themes, judging criteria, tools (built with) and participants skills to validate the significance of the percentage ratios difference between online and offline competitions. Thus, we received not only information about each ratio in both types of hackathons but also the p-value based on which we can conclude the significance of the proportional difference.

For the significance validation of the difference between online and offline hackathons for the mean aggregated metrics such as active stage duration, total duration, spatial inter-group distance, time zone inter-group distance, spatial inter-hackathon distance, time zone inter-hackathon distance it was decided to use Mann-Whitney U test since it was discovered that the majority of the populations of observed variables do not follow the normal distribution.

4. Findings

With the growing number of hackathons, only according to Devpost data, the number of hackathons grew from 56 in 2013 to 831 in 2020, and the number is about to increase in 2021. According to data collected on 16.10.2021, there were already 854 hackathons in 2021. An overview of hackathons amount is shown in fig. 5, highlighting the growing trend of software competitions in 10 years.

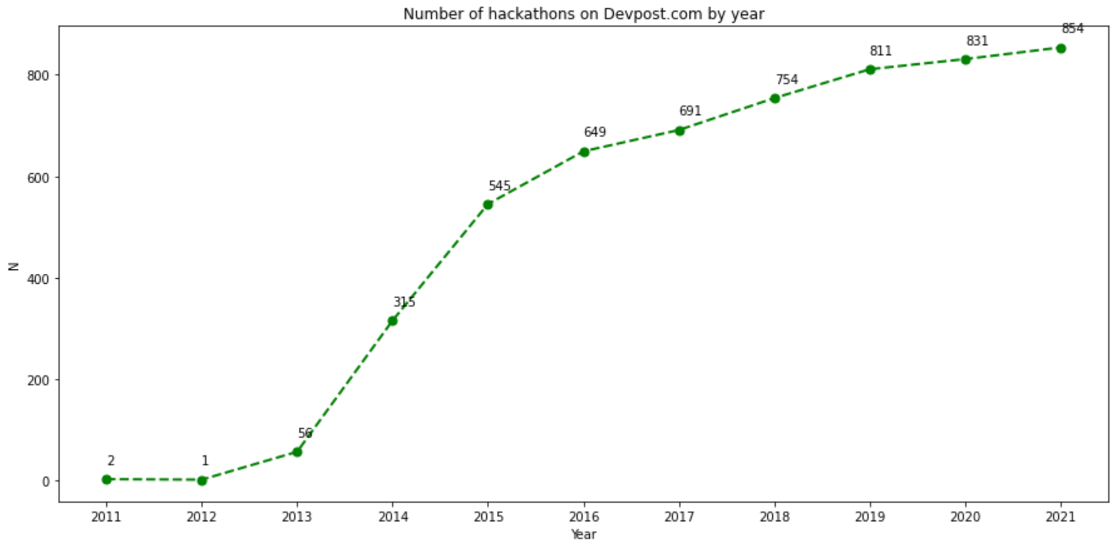

Fig. 5. **Number of hackathons on Devpost in a period from 2009 to 2021**

Source: author’s calculations

The significant growth of online hackathons is also worth mentioning. In general, a significant leap was noticed in 2020 that can be related to the COVID-19 outbreak. Until 2020 the growth was mainly happening due to the increase of offline hackathons amount, but the tables turned, and the outbreak of COVID-19 started the rapid growth of the number of online hackathons. An overview of the separate numbers of online and offline hackathons is shown in fig. 6, which also clearly shows the gap between the two types of competition that occurred in 2020.

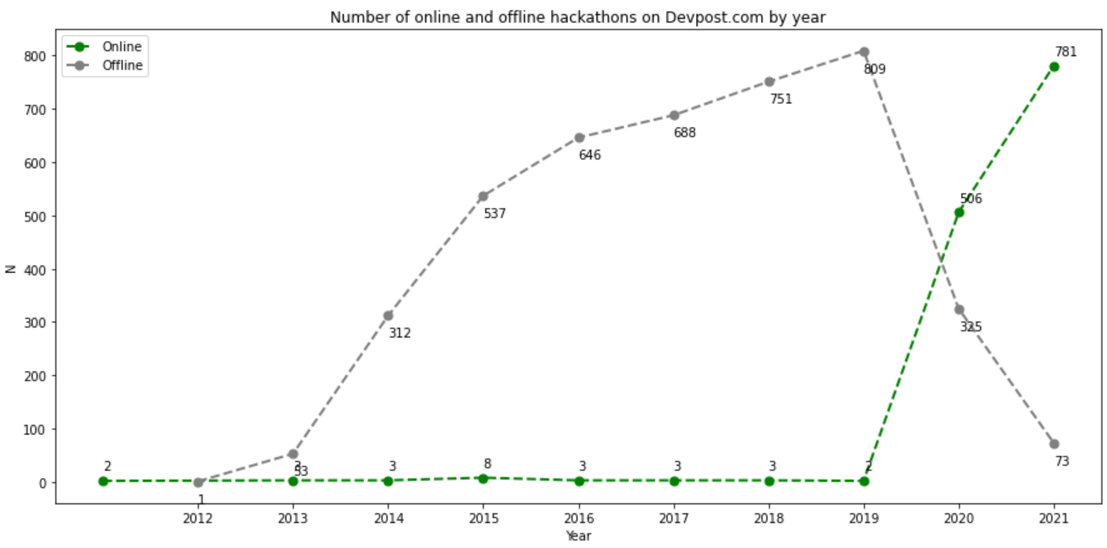

Fig. 6. **Number of online and offline hackathons on Devpost from 2012 to 2021**

Source: author’s calculations

Therefore, it becomes unavoidable to define the difference between online and online hackathons clearly. Here we will discuss our findings in the context of defined research questions and stated hypotheses.

4.1 Online and offline hackathon stages (RQ1)

As mentioned in the introduction section, we will focus on the main stages of online and offline hackathons, the duration of each step, and the total duration of hackathons. First, we adhere to the submission duration of hackathons as it is the most active stage for participants, and the submission period is usually referred to as the start and end date of hackathons. The result of the analysis is presented in fig. 7, showing that the average duration of online hackathons is 1.73 times higher than the average duration of offline hackathons, where 4.26 days is the average duration of online competitions, and 2.45 is the average duration of offline competitions with the standard deviation of 4.07 and 2.19 respectively. Moreover, we can see that the difference is statistically significant [p < .001].

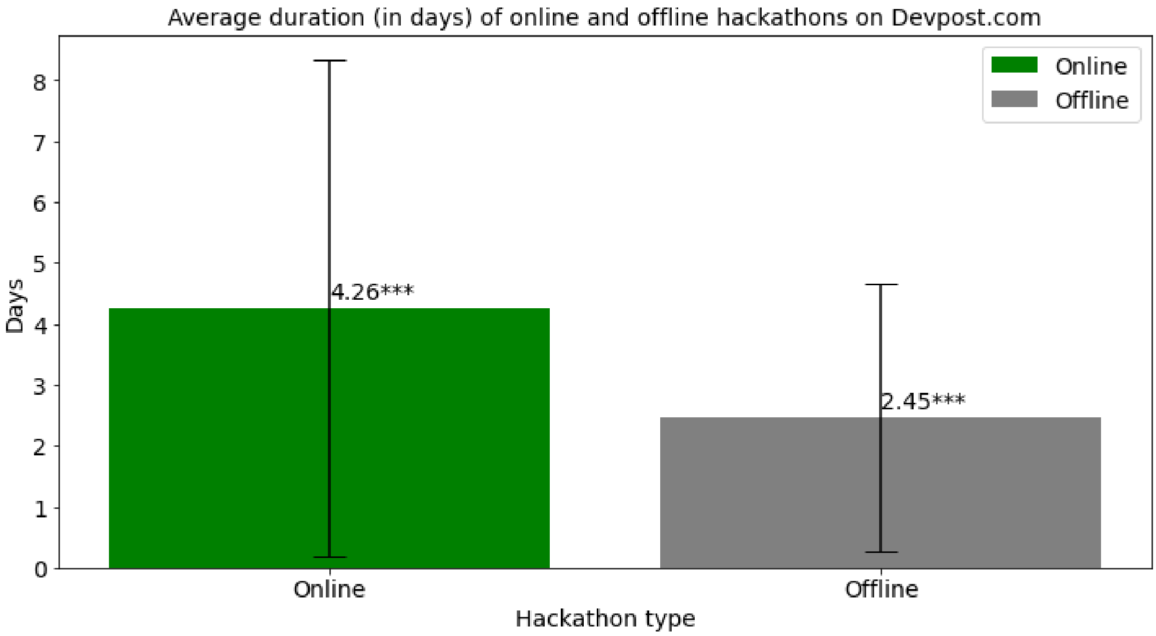

Fig. 7. **The average active stage duration of online and offline hackathons in days**

Source: author's calculations
Note: the bars indicate the mean (m) and standard deviation (SD) for each hackathon type, significance is represented in star notations, where *** p<0.001, ** p<0.01, * p<0.05

During the next step, the schedule of the online and offline events was investigated. According to Devpost hackathons' schedules that we analyzed, there are 3-4 official stages: 1) submission period, 2) public voting (which is present in 11.7% of all hackathons), 3) judging and 4) announcing winners. Winners' announcement is usually not a continuous step and does not have an end time, so we did not calculate the duration for this step. Public voting is an optional step that appears in 17% of online hackathons and only in 11% of offline ones. Therefore, we separated schedules of events for both cases:

1. **Case 1 (the most common):** 1) submission; 2) judging; 3) winners announced
2. **Case 2:** 1) submission; 2) public voting; 3) judging; 4) winners announced

The average duration of each stage and distance between them in case 1 and case 2 are shown in tables 5 and 6, respectively. All metrics are presented in the measure of days.

Table 5. **The average duration (days) of each stage and distance between them in the first type of schedule.**

Hackathon type	Submission duration	Judging duration	Time between submission and judging	Time between judging and winners' announcement
Online	m = 4.15 SD = 3.79	m = 2.31 SD = 2.53	m = 0.31 SD = 1.13	m = 1.47 SD = 3.01
Offline	m = 2.38 SD = 1.98	m = 1.29 SD = 1.28	m = 0.19 SD = 1.05	m = 1.46 SD = 3.09

Source: author's calculations
Note: m represent mean and SD stands for standard deviation

Table 6. **The average duration (days) of each stage and distance between them in the second type of schedule.**

Hackathon type	Submission duration	Public voting	Judging duration	Time between submission and public voting	Time between public voting and judging	Time between judging and winners' announcement
Online	m = 4.64 SD = 4.37	m = 3.07 SD =3.57	m = 2.86 SD = 3.05	m = 0.16 SD = 1.44	m = -0.64 SD =1.32	m = 2.09 SD = 3.53
Offline	m = 2.70 SD = 2.58	m = 1.85 SD = 2.67	m = 1.47 SD = 1.62	m = -0.14 SD =0.62	m = -0.04 SD =0.78	m = 1.46 SD = 2.45

Source: author's calculations
Note: m represent mean and SD stands for standard deviation

Based on the presented information, we can conclude that in both types of schedules, the duration of the submission stage is higher in online hackathons. This behaviour is expected for online competitions because teams need more time to get to know each other, cope with unexpected technical issues, and work on a given task from different places, maintaining communication exclusively virtually. Similarly, judging duration lasts longer in online-based competitions. One possible explanation for this difference is the location of experts who evaluate project solutions in different countries and time zones and the possibility of time extension due to technical problems. Moreover, neither participants nor judges are cut from their daily life and responsibilities, unlike offline hackathons, where their entire time is invested in the several days of intense competition. Hence, the parties need more time to solve the identified task or evaluate that solution. Similarly, public voting lasts longer in online hackathons than in offline ones.

Interestingly, judging in online and offline competitions starts right after submission or public voting ends, in some cases even slightly earlier. The same picture can be observed for 4-stage hackathons where public voting as a next step starts on average a bit before the end of submissions and lasts for a couple more days. In 3-stage hackathons, winners are announced with an average gap of 1.47 and 1.46 days for online and offline hackathons accordingly. Whereas, in 4-stage ones, the gap is a bit bigger for online hackathons and, on average, is equal to two days. The gap between winners' announced stage and judging seems to be unnecessarily big, as far as the practical stage of judging is already over. Schematic interpretations of online and offline hackathons with different amounts stages are represented in figures 8 and 9 accordingly.

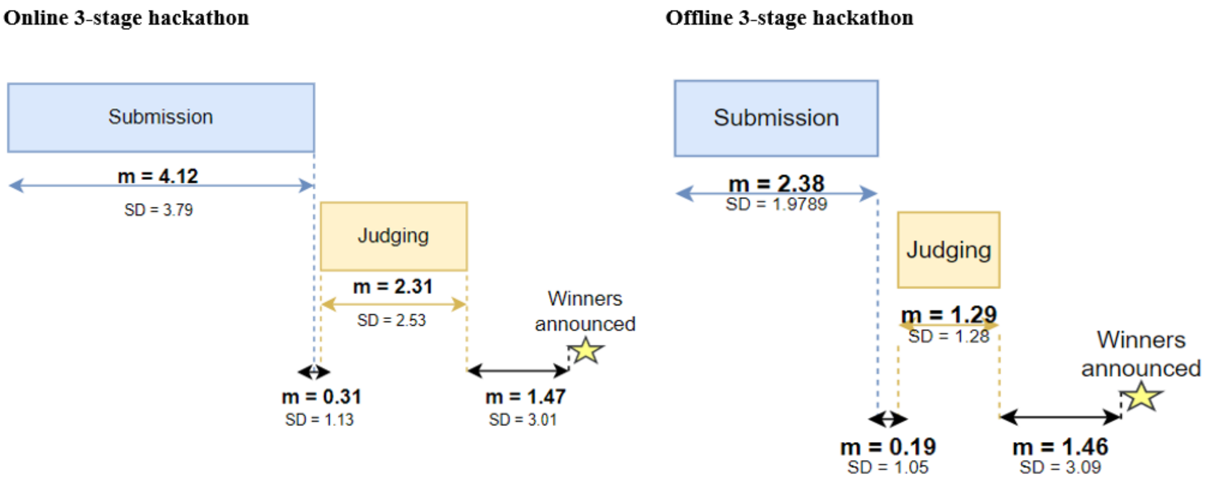

Fig. 8. **Schematic interpretation of 3-stage hackathons.**
Source: author's calculations
Note: mean duration in days (m) and standard deviation (SD) for each type.

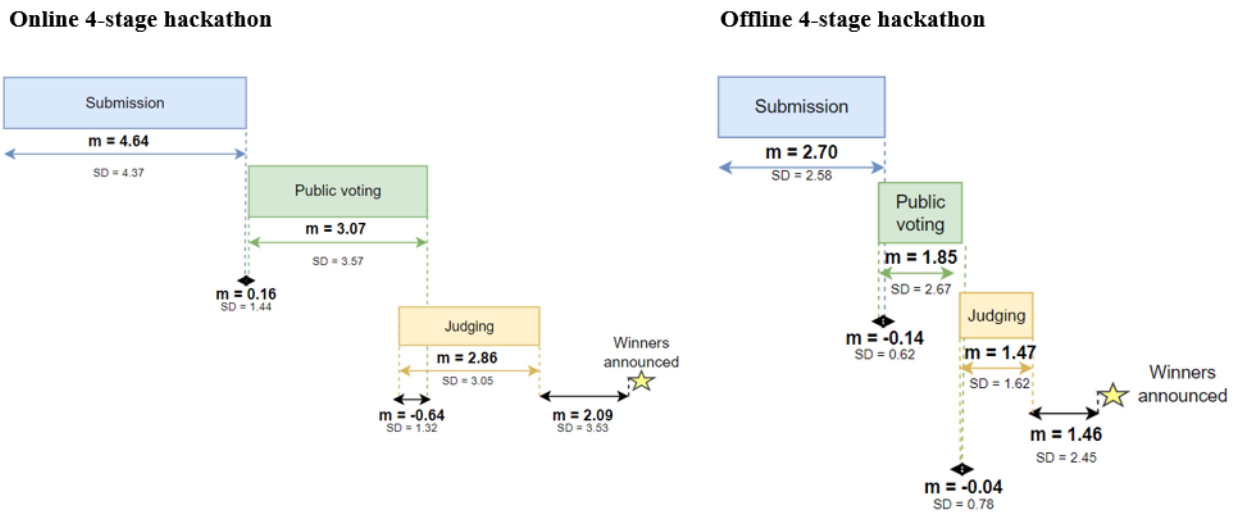

Fig. 9. **Schematic interpretation of 4-stage hackathons.**
Source: author's calculations
Note: mean duration in days (m) and standard deviation (SD) for each type.

As mentioned earlier in this section, the start and end times of hackathons stated in the Devpost hackathon description are the "effective" dates that participants use to create projects. However, if we take a closer look at the actual end date, which is the date when winners are announced, then the average total duration of online competitions (m = 7.63, SD = 7.82) would also be significantly higher than offline ones (m = 4.91, SD = 6.37), which is also represented in figure 10.

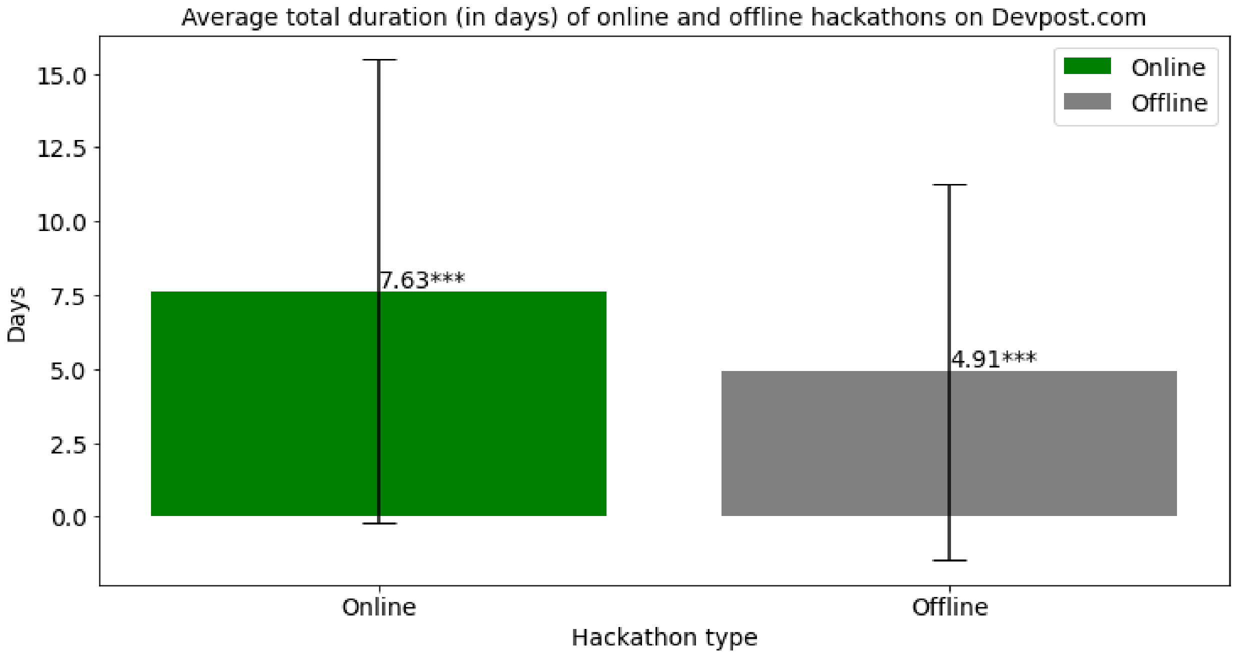

Fig. 10. **The average total duration of online and offline hackathons in days**
Source: author's calculations
Note: the bars indicate the mean (m) and standard deviation (SD) for each hackathon type, significance is represented in star notations, where *** p<0.001, ** p<0.01, * p<0.05

To sum up, online and offline hackathons differ significantly in terms of the duration of stages (both active submission and total time). Based on the overall process analysis, online-based events and offline are similar, with more extended submission, judging and public voting stages in online hackathons.

4.2 Online and offline hackathon topics (RQ2)

This part will present results on competition themes and how they differ in online-based compared to offline events. As stated in the introduction, we assume there will be a shift towards social goods and health-related topics due to the COVID-19 breakthrough. In the majority of the cases, one hackathon can have different themes. For example, in the **#WirVsVirus** hackathon, there are two themes which are "Social Good" and "Covid-19", whereas, in the **#YB Hackathon**, the topics are "Blockchain", "Machine Learning/AI", and "IoT". The difference in online and offline hackathons' themes is presented in a word cloud (fig. 11) and table 7 (the significance here shown using star notation).

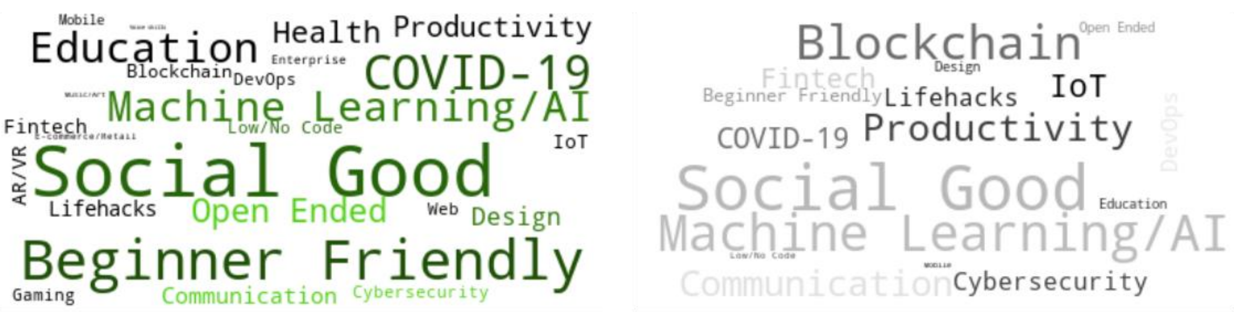

Fig. 11. **Word cloud with online (left) and offline (right) hackathons' themes**

As was expected, the "COVID-19" (9.5% online, 4% offline) and "Health" (5.1% online, 1.5% offline) topics ratio increased significantly in online hackathons compared to offline ones.

Moreover, from table 7, we can see that in online competitions, such topics as "Beginner-friendly" (12.4% online, 1.8% offline), "Education" (7.2% online, 0.9% offline) and "Open-Ended" (6.4% online, 1% offline) took the lead. It is interesting to note that "Education" hackathons, which were supposed to gain more and more popularity among universities (Cobham et al., 2017; García & Menéndez, 2021), are really at the leading positions in online events. "Social Good" (21.7% online, 21.9% offline) remains in the first position, and there was not even a slight significant change there. At the same time, "Machine Learning/AI" (7.5% online, 14.8% offline) and "Productivity" (4.5% online, 7.2% offline) are significantly inferior in online hackathon positions compared to offline ones, although they retain their leading positions in the rating. It is also important to note that such competition themes as "Blockchain" (1.7% online, 13.5% offline), "Communication" (2.7% online, 4.6% offline), "IoT" (2% online, 4.8% offline), "AR/VR" (2% online, 3.8% offline), "Fintech" (2% online, 3.3% offline) and "Gaming" (1.5% online, 3.1% offline) are more prevalent among offline hackathons and less among online ones.

Since the topics of online and offline hackathons differ significantly and they, in turn, can affect assessment criteria, it is also worth analyzing hackathons' benchmarks for grading participants' projects. The results of the analysis can be found in figure 12 and table 7, showing that "Technical Difficulty" (3.2% online, 6% offline) in online competitions is not as crucial as in offline ones, which can be linked to the fact that there are more "Beginner-friendly" online competitions. However, with slight changes in position, such benchmarks as "Originality" (5.6% online, 5.5% offline), "Creativity" (4.5% online, 5.5% offline) and "Design" (5.9% online, 4.1% offline) are taking the lead in both online and offline events. Following "Education" that has become one of the leading themes of online hackathons, "learning" has become one of the leading judging criteria, especially in contrast with its position in offline hackathons (2.9% & 5th in rank online, 1.2% & 15th in rank offline). Similarly, such criteria as "technology" (2.9% online, 1.3% offline) and "completion" (2.4% online, 1.2% offline) have higher rates and ranks in online hackathons than in offline. On the other hand, "usefulness" (1.4% online, 3.5% offline), "polish" (1.2% online, 3.5% offline) have higher positions in offline hackathons' criteria ranking than in online. It should be noted here that all listed above criteria changes are significant [p<0.001].

Fig. 12. **Word cloud with online (left) and offline (right) hackathons' criteria**

Table 7. **Comparison of the most common themes and judging criteria in online and offline hackathons**

Comparison of the most common themes in online and offline hackathons					
	Online		Offline		
Theme	%	Rank	%	Rank	sig
Social Good	21.7	1	21.9	1	
Beginner Friendly	12.4	2	1.8	14	***
COVID-19	9.5	3	4	7	***
Machine Learning/AI	7.5	4	14.8	2	***
Education	7.2	5	0.9	18	***
Open Ended	6.4	6	1	17	***
Health	5.1	7	1.5	15	***
Productivity	4.5	8	7.2	4	**
Design	3.1	9	0.9	19	***
Communication	2.7	10	6.4	5	***
Lifehacks	2.4	11	2.7	12	
Fintech	2	12	3.3	9	*
AR/VR	2	13	3.8	8	**
IoT	2	14	4.8	6	***
Blockchain	1.7	15	13.5	3	***
Web	1.6	16	0.7	20	*
Gaming	1.5	17	3.1	10	**
Low/No Code	1.4	18	0.4	21	**
DevOps	1.4	19	2.5	13	*
Cybersecurity	1.3	20	2.8	11	**
Voice skills	0.2	25	1.2	16	***

Comparison of the most common judging criteria in online and offline hackathons					
	Online		Offline		
Criteria	%	Rank	%	Rank	sig
design	5.9	1	4.1	4	***
originality	5.6	2	5.5	3	
creativity	4.5	3	5.5	2	***
technical difficulty	3.2	4	6	1	***
learning	2.9	5	1.2	15	***
presentation	2.9	6	1.8	7	***
technology	2.9	7	1.3	11	***
completion	2.4	8	1.2	13	***
adherence to theme	2.1	9	0	-	
technical complexity	1.9	10	1.4	10	***
functionality	1.9	11	1.2	12	***
impact	1.8	12	1.6	9	***
innovation	1.5	13	1.7	8	*
usefulness	1.4	14	3.5	5	***
wow factor	1.3	15	1.2	14	
polish	1.2	16	3.5	6	***
execution	1	17	0.8	19	***
feasibility	0.8	18	0.3	32	
practicality	0.6	20	1.1	16	
user experience	0.4	28	0.8	20	
technicality	0.2	56	0.9	17	
technical	0.1	116	0.8	18	

Source: author's calculations
Note: sig represents significance in star notations, where *** $p<0.001$, ** $p<0.01$, * $p<0.05$. Top 20 rank is highlighted with a green color.

Summing up, online and offline hackathons differ significantly in terms of themes. For online-based events, such topics as "Beginner-friendly", "COVID-19", and "Education" are prevailing, whereas, for face-to-face ones ", Blockchain", "Productivity", and "Communication" are rated higher. The similarity between the two types of hackathons lies in the "Social Good" topic, which is the most popular among both, and "Machine Learning/AI" that also belongs to the high rated topics, even though the ratio difference is almost twice higher in offline events. There is also a moderate difference in judging criteria. "Technical difficulty" that is leading in offline competitions is not that important in online ones, similarly to "usefulness" and "polish". For online competitions, on the other hand, "learning" is one of the top-rated criteria as well as "technology" and "completion". Whereas "Design", "Originality", and "Creativity" with some difference in ratios belong to the most popular criteria in both types of hackathons.

4.3 Online and offline hackathon geography (RQ3)

At the beginning of section 4, it was already mentioned that there is a rapid growth in hackathons, and for the last two years, the growing trend began to shift towards online competitions. The geographic diversity of participants also increased significantly from 22 unique countries in 2011 to 74 in 2019 in offline hackathons and from 1 in 2013 to 92 in 2020 for online hackathons. The number is about to increase in 2021. According to data collected on 16.10.2021, there were already 67 unique countries for online hackathons and 26 for offline ones in 2021 (the total number for both types is 74 unique participants' countries). These facts show that not only the number of online hackathons increases but also its popularity in different parts of the world. The overview of the amounts of participants' countries in online and offline hackathons is shown in fig. 13.

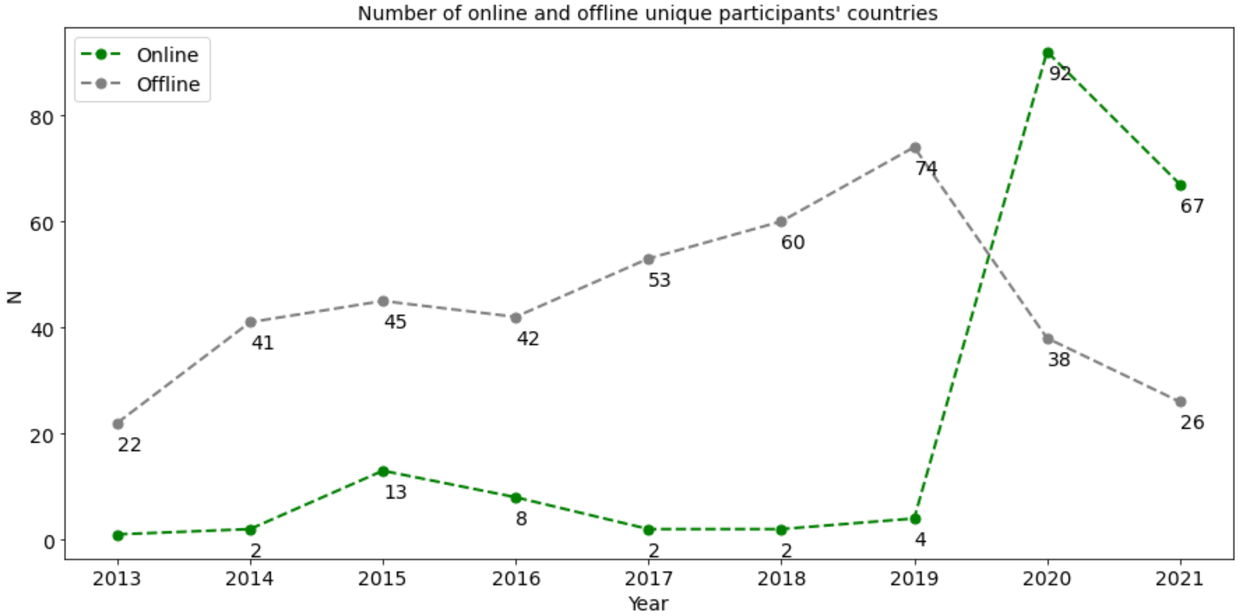

Fig. 13. **Number of online and offline unique participants' countries per year**
Source: author's calculations

We can still state the numerical superiority of the United States of America and Canada participants for online and offline hackathons, as shown in fig. 14 and 15. However, this is most likely since Devpost was started as a mainly US-based database. On the other hand, the fact that in recent years the percentage of participants from India, Australia, Mexico is getting to the top-7 in online hackathons cannot but mentioned. There is also a growing tendency in offline hackathons for participating countries like the United Kingdom, Switzerland, Germany, India, and Pakistan. However, compared to online competitions, the significant growth in offline ones falls on European countries.

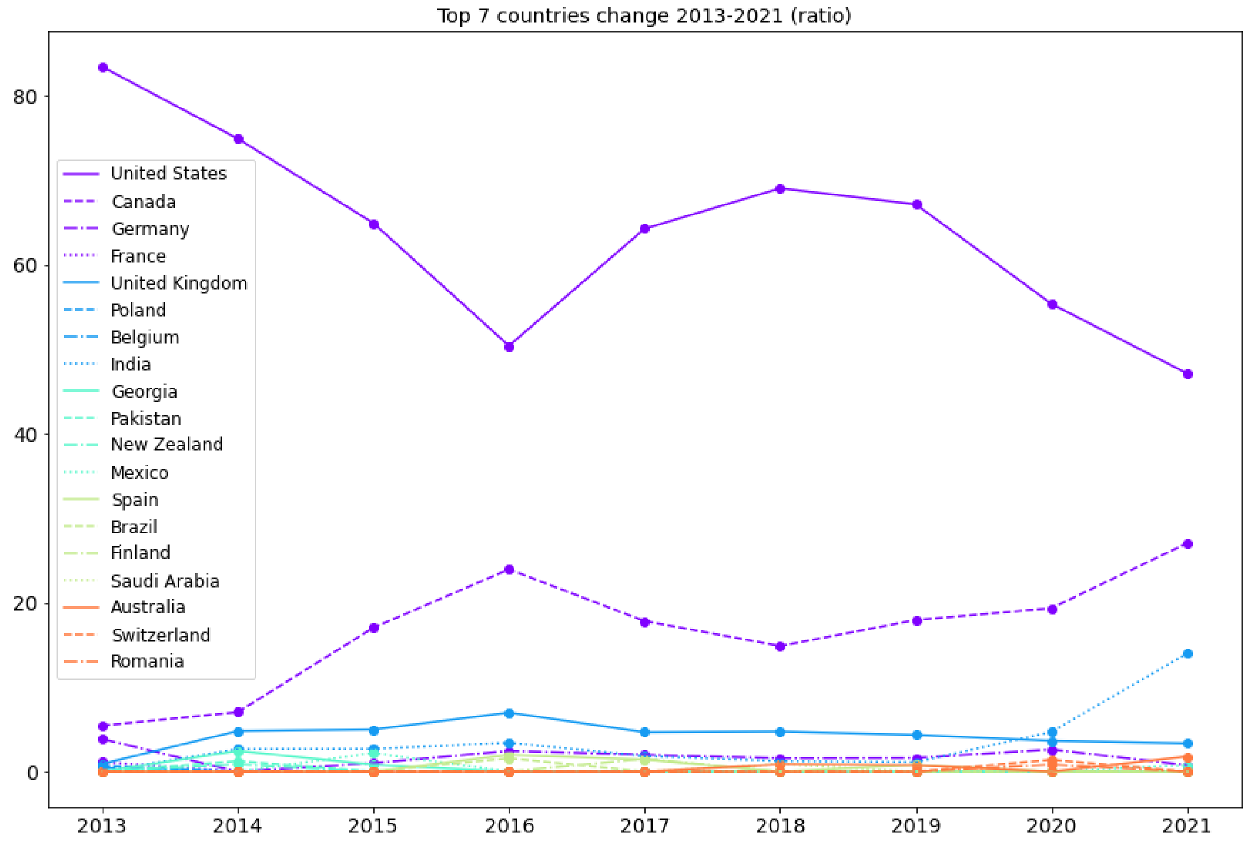

Fig. 14. **Top 7 participating countries in offline hackathons 2013-2021**
Source: author's calculations

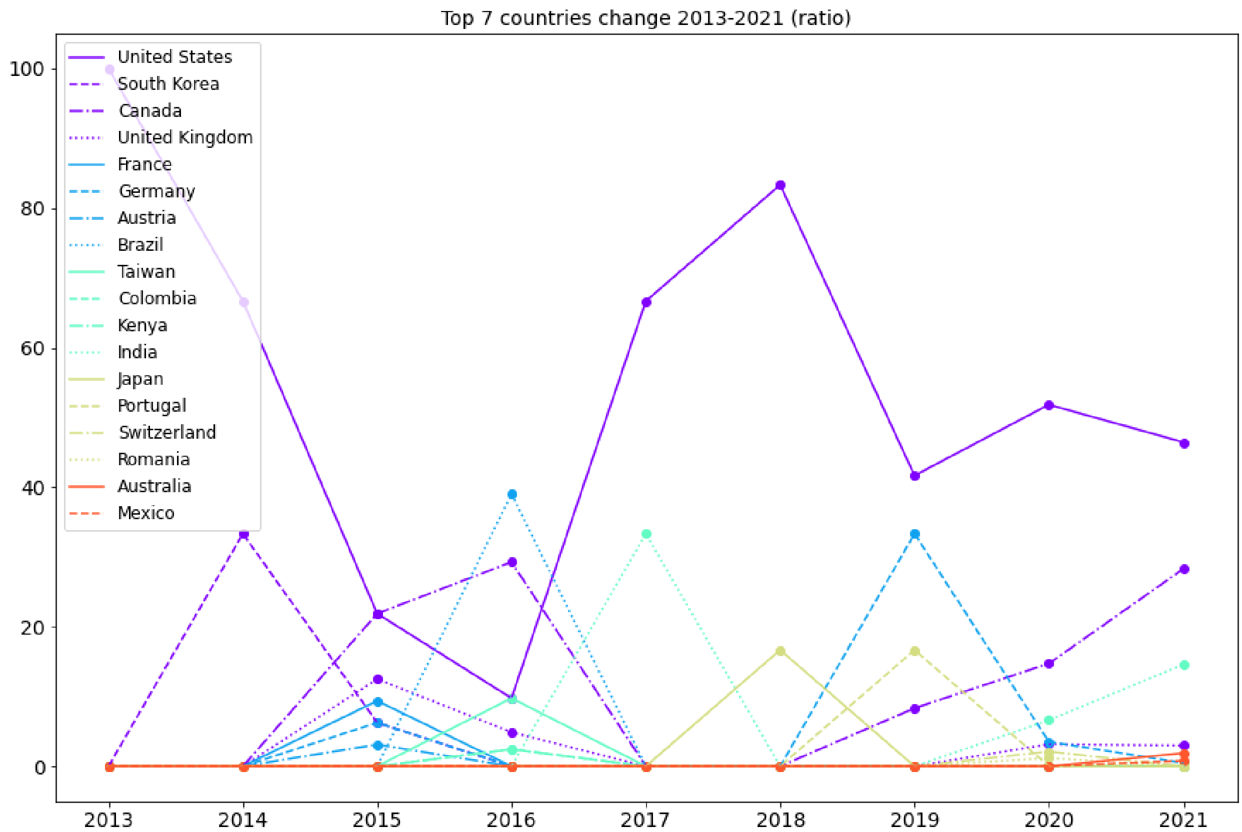

Fig. 15. **Top 7 participating countries in online hackathons 2013-2021**
Source: author's calculations

For a better representation of information, the video map that shows day-by-day changes across the globe in online and offline hackathons participants with the counter of each is presented in the **authors' open google drive file.** Only based on two pictures – the end of 2019 (fig. 16) and the middle of October 2021 (fig. 17), it is already clearly seen that not only the number of online hackathons' participants is growing but also their geographical variety.

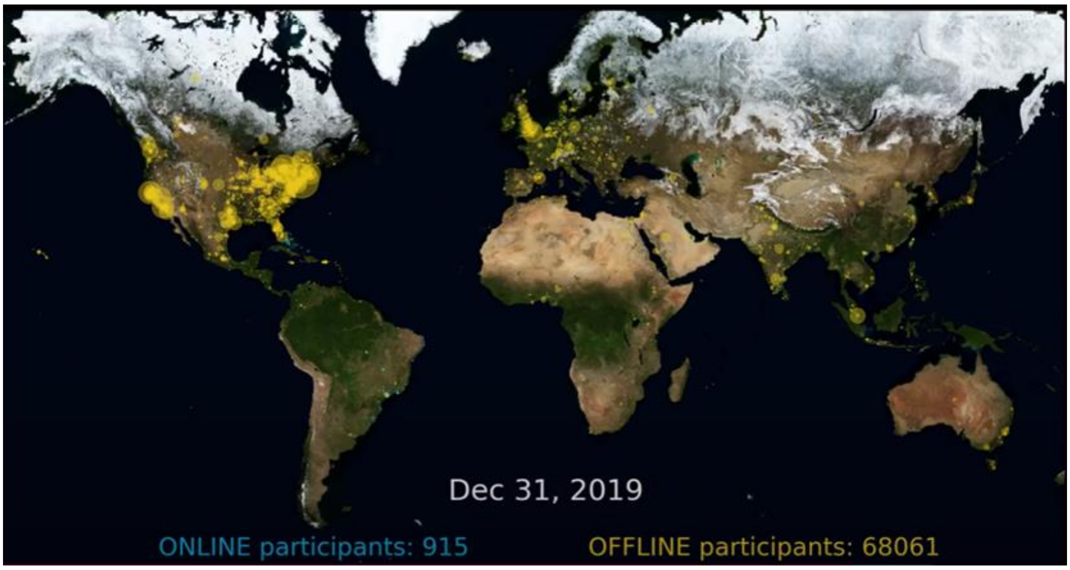

Fig. 16. **Online and offline hackathons participants dispersion on the world map at the end of 2019**
Source: author's calculations

Fig. 17. **Online and offline hackathons participants dispersion on the world map at the middle of October 2021**
Source: author's calculations

As the participating countries in online hackathons become more diverse, we expect the spatial and time zone intergroup and inter-hackathon differences to grow, as stated in hypotheses 2 and 3 RQ3. To deep dive into this question, we calculated spatial and time zone distances inside each group of participants and hackathon as was explained in section 3.3 and compared them between online and offline hackathon types. As expected, the spatial distance between groups that participated in online hackathons was higher than those that participated in offline ones. The difference is statistically significant, as shown in figure 18.

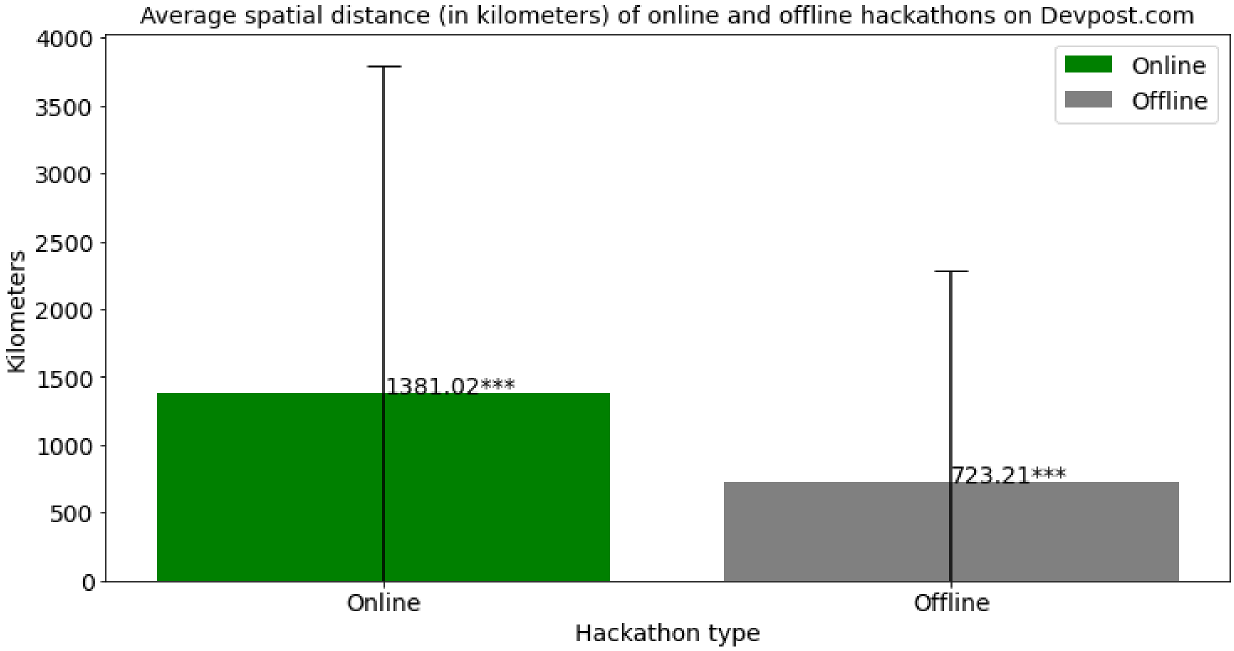

Fig. 18. **Average spatial distance (in kilometers) inside participants groups**
Source: author’s calculations
Note: the bars indicate the mean (m) and standard deviation (SD) for each hackathon type, significance is represented in star notations, where *** p<0.001, ** p<0.01, * p<0.05

The average spatial distance between participants of the online hackathon is 1382 kilometres (SD = 2409), which is twice higher than in offline hackathons (m = 723.21, SD = 1564.79). The biggest spatial zone distance observed is 13341 kilometres between teammates where 1 participant is from South Tangerang (Indonesia), one from Toronto (Canada), one from Santiago (Chile), and one without stated location, who created the **Shopnation** project, which became a winner of OneHacks hackathon. Another good example is a team consisting of 3 people whose intergroup distance is 13336 kilometres (from Paris (France) to Ambato (Ecuador) and Moka District (Mauritius). The average time zone difference inside each group is 8 and 12 hours, respectively, making it harder to coordinate. However, as we can see, the first group was a winner of the hackathon, which means that even with high spatial and time zone differences, teams can create an innovative solution that is worth the first place. The average time zone distance difference between online and offline hackathons teams is not as high as was predicted. However, the same pattern as with spatial distances is observed here: intergroup time zone distance is twice bigger in online hackathons (m = 1.04, SD = 2.20) than in offline (m = 0.54, SD = 1.52). The visual representation can be seen in figure 19.

The most considerable time zone distance observed is between teammates, where 1 participant is from Auckland and two from San Francisco who created the **ArtFinder** project. Further, the following pattern was observed: even if the intergroup spatial is extensive, time zone differences can be 0, making it much easier to collaborate on project creation and communicate in real-time with other team members. The most considerable intergroup spatial distance with a time zone difference of 0 is 3004. Many groups have an intergroup spatial distance between 2000 and 3000 kilometres and a time zone difference of 0.

While we cannot compare the diversification of hackathons origination since there is no location information for the online events, we can look at hackathon participants' spatial and time zone distances. The average distance between participants inside hackathons (fig. 20) have the same consistent pattern as intergroup differences: in online type, it is significantly higher than in offline. Thus, the spatial distance between participants of online hackathons is three times higher than in offline ones (m = 3079.73, SD = 2504.24 for online, m = 806.93, SD = 1160 for offline)

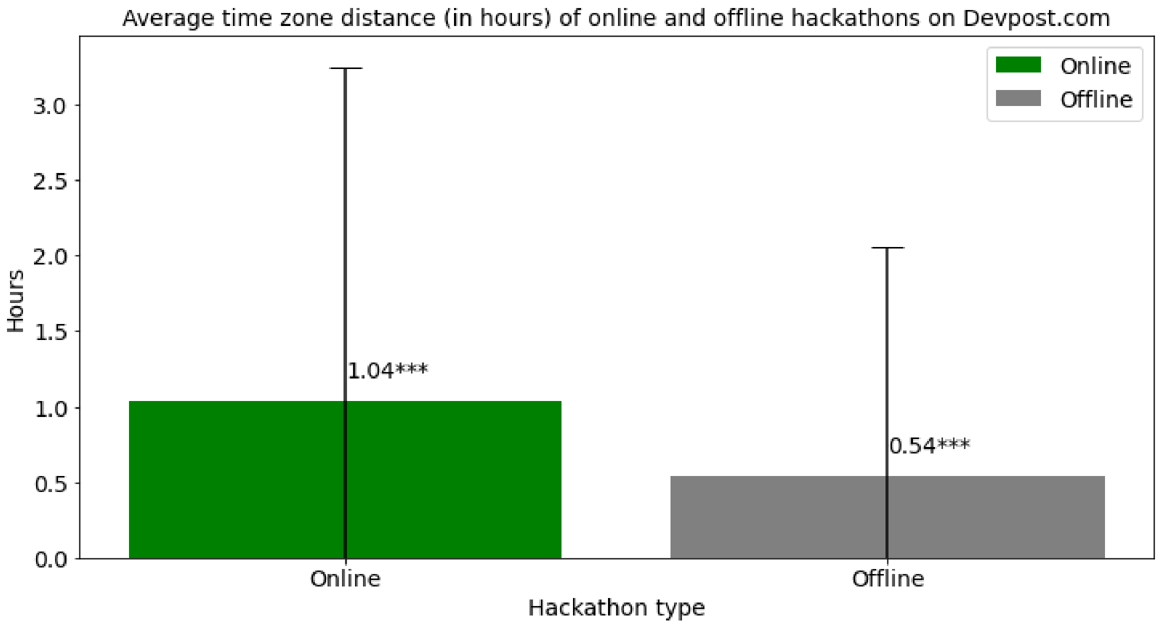

Fig. 19. **Average time zone distance (in hours) inside groups participated in online and offline hackathons**

Source: author's calculations

Note: the bars indicate the mean (m) and standard deviation (SD) for each hackathon type, significance is represented in star notations, where *** p<0.001, ** p<0.01, * p<0.05

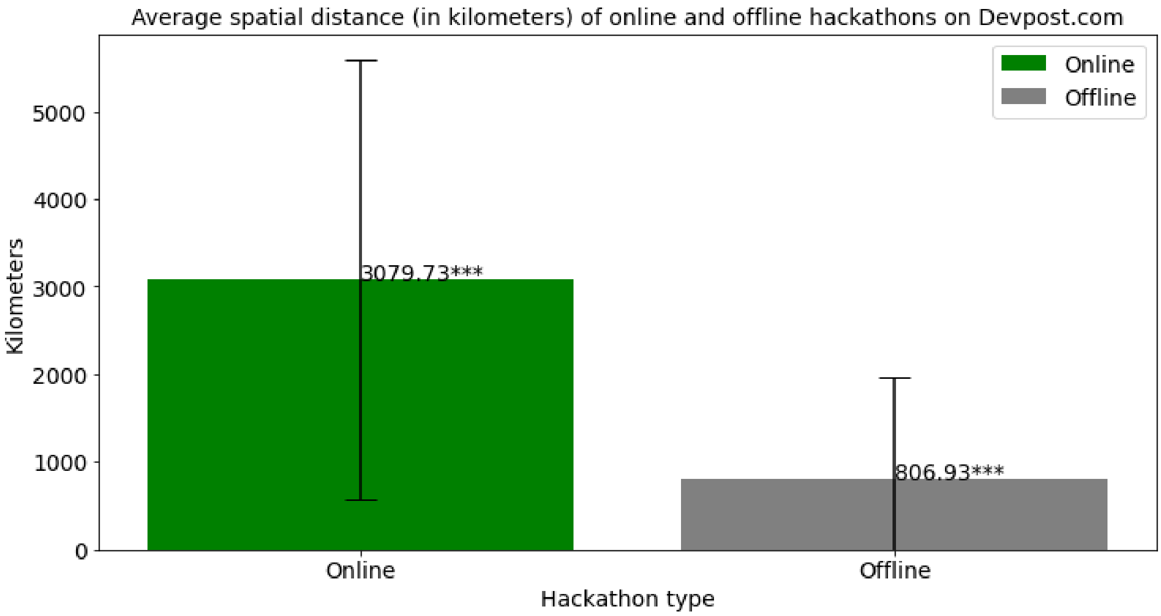

Fig. 20. **Average spatial distance (in kilometers) in online and offline hackathons**

Source: author's calculations

Note: the bars indicate the mean (m) and standard deviation (SD) for each hackathon type, significance is represented in star notations, where *** p<0.001, ** p<0.01, * p<0.05

Reviewing the average time zone distance (fig. 21) between online and offline hackathons participants, we can see that in the first case, the difference (m = 2.43, SD = 2.56) is significantly higher than in the second one (m = 0.51, SD = 1.10). Moreover, this difference is more significant than what we observed earlier for teams. From here, we can conclude that there is a more considerable average distance between participants of hackathon than between individual group team members, that can be explained by the fact that while creating a group, participants might take into consideration the fact of time zone difference and more likely people will collaborate when there are fewer obstacles in communication, including time distinction issue.

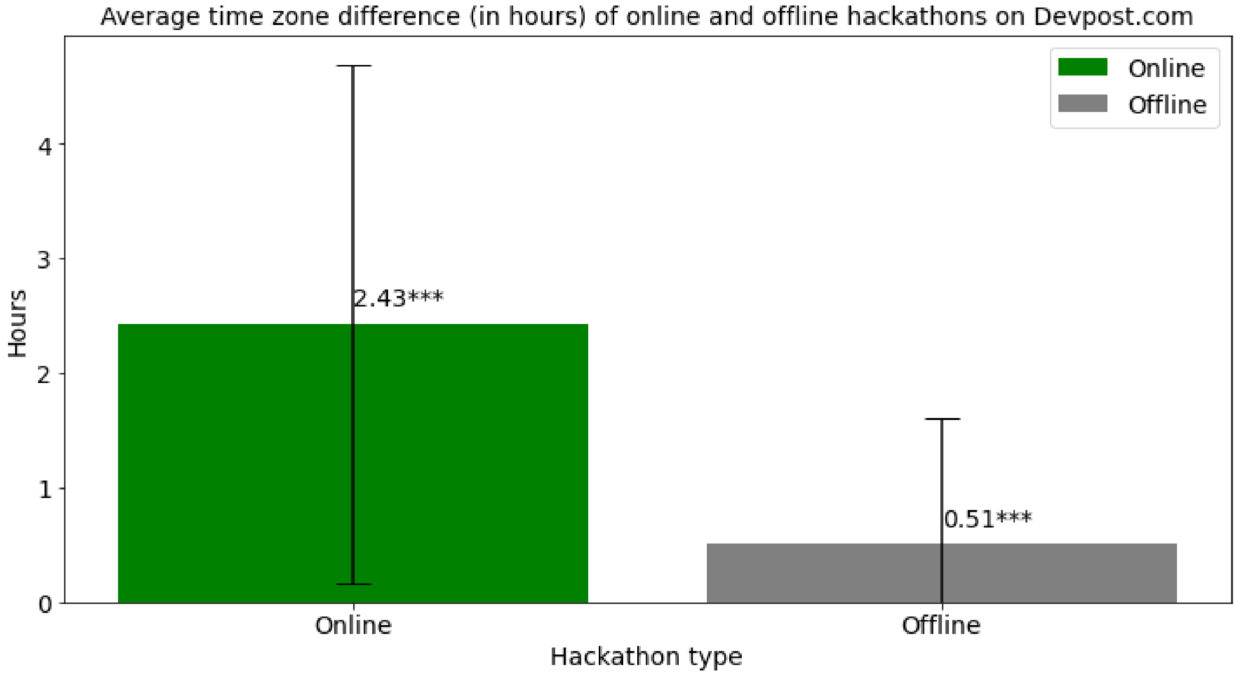

Fig. 21. **Average time zone distance (in hours) in online and offline hackathons**
Source: author's calculations
Note: the bars indicate the mean (m) and standard deviation (SD) for each hackathon type, significance is represented in star notations, where *** $p<0.001$, ** $p<0.01$, * $p<0.05$

To sum up, there is a growing number of participants from all over the world, both for offline and online hackathons, which can be linked to the popularity of such competitions overall. However, the online hackathon's growth in the top-7 countries is more diverse geographically and includes mainly positive changes in India, Australia, and Mexico. In offline hackathons, this growth is caused mainly by European countries. Therefore, intergroup, as well as inter-hackathon spatial distance, is significantly higher for online hackathons. However, surprisingly, with striking spatial distance, intergroup time zone one is not that big, even though the difference between online and offline competitions averages is statistically significant. On the other hand, at the hackathon level, the average time zone difference between participants in online events is significantly more significant than offline events and higher than the intergroup value.

Based on these findings, we also assume that using online or cloud tools in online hackathons would be higher than in offline ones, which is also stated in our fourth research question.

4.4 Online and offline hackathon tools and technologies (RQ4)

In this question, we investigated tools used by participants to create their projects, which are usually stated in the "Built with" section on every project description page. We also paid attention to GitHub repositories created during offline and online events. Finally, participants' skills were investigated. This is done to get a complete picture of technologies used because not all tools that were useful for solution creation were mentioned in the "Built with" section but could be mentioned in the personal skills of team members.

The analysis is represented in figure 22 and table 8, showing that regardless of hackathon type, the main tools are javascript (7% online, 6.64% offline), python (6.43% online, 5.95% offline), HTML (4.29% online, 3.54% offline), and css (4.53% online, 3.48% offline).

Fig. 22. **Word cloud of tools used to create projects in online and offline hackathons**

There is slight growth in usage of such collaboration tools as GitHub (1.03% online, 0.53% offline), google cloud (1.52% online, 1.11% offline), Heroku, and discord (last two tools, however, have still too small a percentage to be even in top-20) in online hackathons. On the other hand, there are higher ratios for azure, Twilio, and amazon-web-services usage in offline hackathons' project creation. However, the percentage is too low for being included in top-20. Therefore, we cannot clearly say that more online or cloud technologies are used in online competitions. Similarly, we cannot state that such technologies are popular or at least stated as "built with" tools neither in online nor in offline events. Moreover, the top-4 "Built with" tools look similar in both hackathon types, with slight (even though significant) changes in ratios.

As far as GitHub is stated as a tool with a higher ratio in projects created during online events, we investigated the mentions of GitHub on the project description page. Surprisingly, the ratio of offline projects containing GitHub links is much higher than online ones, as shown in figure 23.

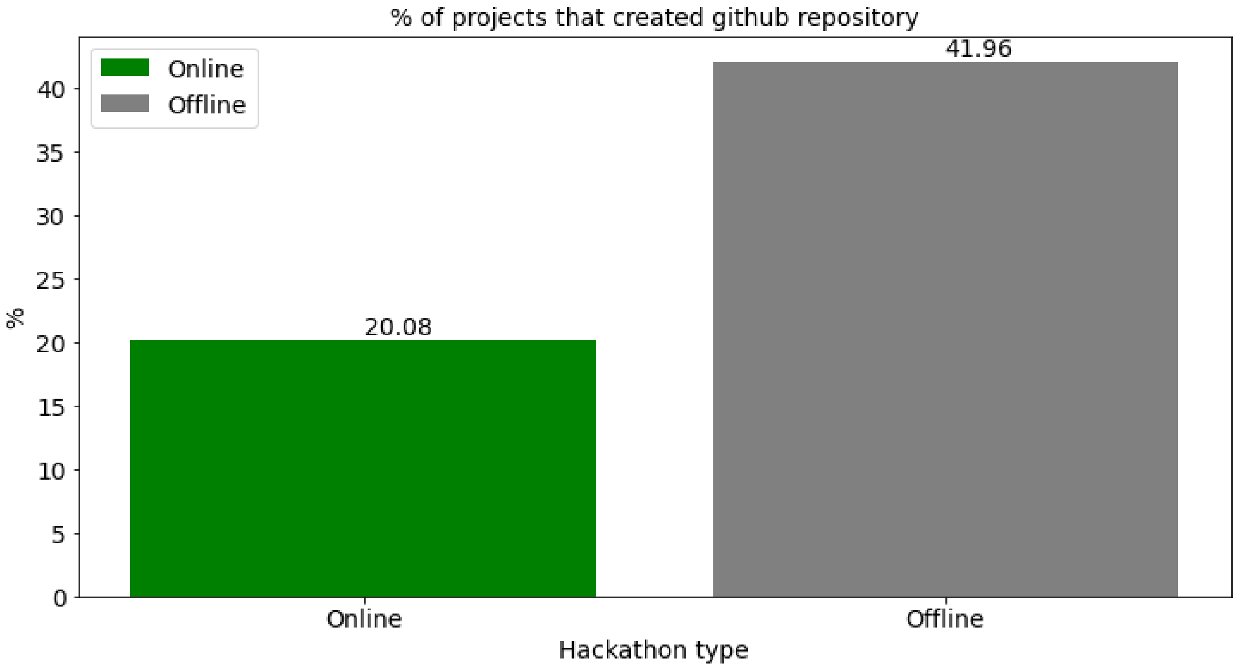

Fig. 23. **The proportion of online and offline projects that contains GitHub link**
Source: author's calculations
Note: the bars indicate the mean (m) and standard deviation (SD) for each hackathon type, significance is represented in star notations, where *** p<0.001, ** p<0.01, * p<0.05

From here, we can state that GitHub is a universal tool that is not only used by online teams collaboration but also during offline hackathons more likely for collaborating on code creation and tracking version history.

Likewise, there are no significant differences in top online hackathons' participants' skills compared to offline ones (fig. 24., table 8). The most popular skills are python (11.4% online, 9.4% offline), java (8.5% online, 9.3% offline), javascript (7.2% online, 7% offline), and c++ (4.8% online, 4.6% offline), which are quite like the tools used to build projects.

Fig. 24. **Online (left) and offline (right) hackathons' participants' skills word cloud**

Table 8. **Online and offline hackathons' top-20 tools and participants' skills comparison**

Comparison of the most popular tools to build projects in online and offline hackathons						**Comparison of the most common participants' skills in online and offline hackathons**					
	Online		Offline				Online		Offline		
tool	%	Rank	%	Rank	sig	skills	%	Rank	%	Rank	sig
javascript	7	1	6.64	1	***	python	11.4	1	9.4	1	***
python	6.43	2	5.95	2	***	java	8.5	2	9.3	2	***
css	4.53	3	3.48	4	***	javascript	7.2	3	7	3	*
html	4.29	4	3.54	3	***	c++	4.8	4	4.6	4	*
react	3.51	5	1.52	11	***	html	3.6	5	2.3	9	***
node.js	2.73	6	3.05	5	***	css	3.6	6	3.4	6	*
firebase	2.42	7	1.55	10	***	html5	3.1	7	3.6	5	***
html5	2.34	8	2.22	7	***	c	3.1	8	3.1	7	
flask	2.23	9	1.72	8	***	react	2.6	9	1.5	12	***
figma	1.81	10	0.21	63	***	node.js	1.9	10	1.9	10	
css3	1.61	11	1.36	14	***	css3	1.9	11	1.4	13	***
bootstrap	1.54	12	1.36	13	***	sql	1.5	12	1.3	17	***
google-cloud	1.52	13	1.11	15	***	c#	1.3	13	1.6	11	***
java	1.45	14	2.57	6	***	photoshop	1.2	14	0.8	24	
express.js	1.34	15	0.9	22	***	design	0.9	15	0.4	41	
mongodb	1.11	16	1.02	18	***	machine-learning	0.9	16	0.8	22	
github	1.03	17	0.53	35	***	flask	0.8	17	0.5	35	
flutter	0.86	18	0.16	81	***	mongodb	0.8	18	0.7	27	
google-maps	0.8	19	0.89	23	***	firebase	0.8	19	0.4	44	
react-native	0.78	20	0.5	36	***	mysql	0.7	20	0.7	26	
android-studio	0.73	21	1.41	12	***	php	0.7	23	1.3	15	
unity	0.52	27	1	19	***	arduino	0.6	31	1	20	
jquery	0.5	32	1.09	16	***	android	0.5	36	2.5	8	
android	0.48	34	1.58	9	***	jquery	0.4	41	1.2	19	
swift	0.47	35	0.94	20	***	ios	0.3	47	1.4	14	
arduino	0.37	43	1.02	17	***	web	0.1	130	1.3	16	

Source: author's calculations

Note: sig represents significance in star notations, where *** $p<0.001$, ** $p<0.01$, * $p<0.05$. Top 20 rank is highlighted with a green color.

To sum up, we did not find any evidence of increased cloud or online tools or technologies usage for this research question. Moreover, the top-4 "built with" tools and participants' skills are the same both for online and offline events, with changes only in ratios (even though these changes are significant). Unexpectedly, the ratio of GitHub repositories mentioned in offline projects appeared to be higher than in online ones.

Summarized findings described in this section are represented in table 9.

Table 9. **Summary of findings**

	Finding
Overall	Significant growth of hackathons (from 56 in 2013 to 831 in 2020) & positive trend in hackathons in 10 past years
	2019-2020 is a turning point in trends for hackathons: rapid growth for online (from 2 in 2019 to 506 in 2020) and sharp decline for offline (from 809 in 2019 to 325 in 2020)
Duration (RQ1)	Active stage of online hackathons submissions is on average (m = 4.26, SD = 4.07) longer than offline ones (m = 2.45, SD = 2.19)
	Judging stage and public voting are on average longer in online hackathons than in offline ones
	Public voting and judging in majority cases start right after submission ends
	In most cases some time is needed after judging to announce winners
	The average total duration of online competitions (m = 7.63, SD = 7.82) is significantly higher than offline ones (m = 4.91, SD = 6.37)
Themes & Judging criteria (RQ2)	“COVID-19” and “Health” topics ratio increased significantly in online hackathons compared to offline ones
	“Social Good” is the most popular topic in both online and offline hackathons, without any significant differences in ratio
	Besides “Social Good”, “COVID-19” and “Health” such topics as “Beginner friendly”, “Machine Learning/AI”, “Education” and “Open Ended” are the most common for online hacks whereas “Machine Learning/AI”, “Blockchain”, “Productivity” and “Communication” for offline ones
	“Technical Difficulty” as a judging criteria in online competitions is not as important as in offline ones as well as “usefulness” and “polish”
	“Originality”, “Creativity” and “Design” are taking the lead in both online and offline events’ judging criteria
	For online hackathons “learning” has become one of the leading judging criteria as well as “technology” and “completion” that have higher rates and ranks in online hackathons than in offline.
Geography (RQ3)	Geographic diversity of participants increased (from 22 unique countries of participants in 2011 to 74 in 2019 in offline and from 1 in 2013 to 92 in 2020 for online type of hackathons)
	Percentage of participants from India, Australia, Mexico is getting to the top-7 in online hackathons

	Percentage of participants from United Kingdom, Switzerland, Germany, India, and Pakistan is getting to the top-7 in offline hackathons
	The spatial distance between groups that participated in online hackathons is significantly higher than in groups that participated in offline ones.
	The time zone distance inside groups that participated in online hackathons is higher than in groups that participated in offline ones, but not as much as expected.
	For teams with the big spatial distance in many cases time zone difference is not bigger than 2 hours, that makes it easier to communicate in real time and collaborate.
	The spatial distance between participants of online hackathon is significantly higher than offline ones.
	The time zone difference inside online hackathons is significantly higher than in offline ones.
	Both average time zone and spatial distance is higher if aggregation is done on hackathon level (and smaller on team level).
Tools & skills (RQ4)	Regardless of hackathon type the main tools are javascript, python, HTML, and css.
	The usage of GitHub and google cloud as tools are higher in online competitions, however, there is no big difference in other online or cloud technologies (the % of use in all cases is small)
	GitHub is mentioned more frequently in offline hackathons' project descriptions than in online ones.
	Similarly, to tools, there are no big differences in online and offline participants' skills, the most common ones are python, java, javascript, and c++.

5. Implication

The findings of this study have several implications both for practice and research. First, the results statistically support some theoretical assumptions highlighted in other articles and refute others that will be discussed in the following section.

Before we did an analysis, it was stated that hackathons could contribute significantly to solving issues that have arisen due to the COVID-19 breakthrough (Braune et al., 2021), leading to the enlarged number of participants, competitions and "resulting pandemic solutions"(Gupta & Rubalcaba, 2021). Based on our findings, we can confirm a given statement as we can see the rapid growth of online hackathons and participants from all over the world. Even though we cannot say much about practical solutions for pandemic issues, but themes analysis (RQ2) partly supports this statement as such topics as "COVID-19" and "Health" ratios increased significantly in online hackathons compared to offline ones. That means that a considerable number of hackathons and respectively projects were created for these themes.

It was also discussed in the scientific literature that one of the biggest restraints of online hackathons is a lack of time, especially in the implementation phase (Loeffler & Masiga, 2021a). However, based on analyzed hackathons' schedule data, we can state that the average active stage of online hackathons submissions (m = 4.26, SD = 4.07) is longer than offline ones (m = 2.45, SD = 2.19). The same goes for the judging stage, which is longer online hackathons than offline. Thus, on average total duration of online competitions (m = 7.63, SD = 7.82) is also significantly higher than offline ones (m = 4.91, SD = 6.37). Therefore, there might be a shortage of time in the implementation phase in isolated cases, but online hackathons generally have more time. A longer duration is possible since online events require more reserve time to solve possible technical issues (Franco et al., 2021; Powell et al., 2021). However, we can neither confirm nor deny this statement based on the conducted analysis.

It is widespread in the scientific literature that hackathons have now expanded to a broad range of areas, cover a large number of topics, and no anymore focus purely on software development (García & Menéndez, 2021; Yarmohammadian et al., 2021). This statement can be proven based on our findings for RQ2. There is a wide variety of top-rated themes from "COVID-19", "Health", "Machine Learning/AI", and "Education" in online hackathons to "Blockchain", "Productivity" and "Communication" in offline ones. Whereas "Social Good" is the most common topic for both types of hackathons.

Nowadays, in scientific articles, there is a notion that hackathons have become more popular among universities that launch such competitions as education initiatives, especially with the transition to a hybrid form of studying due to the pandemic. (García & Menéndez, 2021) However, before we did a study, there was no numerical evidence except for isolated cases. Nevertheless, now we can support this statement because, among all themes of online hackathons, "Education" and "Beginner-friendly" belong to one of the most popular and are rated significantly higher than in offline events. It is worth noticing that "Technical Difficulty" as a judging criterion in online competitions is not as crucial as in offline. On the contrary, "learning" criteria strengthens its position in the rating. Based on these findings, we can prove that educational hackathons with the main aim of learning and suitable for beginners are gaining popularity among online hackathons.

One of the advantages of online hackathons stated in scientific articles is the opportunity to collaborate in the same hackathon from different parts of the world due to reduced costs and other travel restraints. (Braune et al., 2021; Franco et al., 2021; Gupta & Rubalcaba, 2021) However, it was not clear whether this was happening in reality or not. However, now we have the necessary shreds of evidence to prove the essentiality of this point of view. First of all, the geographic diversity of participants increased (from 1 in 2013 to 92 in 2020). Moreover, comparing average spatial and time zone distance both inside teams and inside hackathons shows that it is higher for online competitions than for offline. It points to the fact that both teams and hackathons in the online setting are more geographically diverse than offline.

Consequently, there were doubts in the scientific literature regarding big time zone differences that might negatively affect participants in terms of collaboration and engagement (Bouncken et al., 2016; Wang et al., 2021). However, based on the results of the conducted analysis, we cannot see the drastically bigger intergroup time difference for online hackathons. More precisely, in online hackathons, the average time zone distance in hours is 1.04 (SD = 2.20), and in offline, 0.54 (SD = 1.52), the difference is significant but not as high as expected. On the other hand, overall time zone difference between online hackathons' participants is significantly higher than in offline ones (m = 2.43, SD = 2.56; m = 0.51, SD = 1.10 respectively). Therefore, on average, even when hackathons' participants vary geographically and have quite a high time zone distance, there is a tendency for this difference to be lower within the group.

There was an assumption regarding the increasing usage of cloud and online communication tools in online hackathons (Braune et al., 2021). However, based on our analysis, this statement is refuted because regardless of hackathon type, the main tools are javascript, python, HTML, and css. Mentioning other tools is not significant. Moreover, GitHub, one of the most popular technologies for collaborative development of projects, usage is higher in offline hackathons.

Likewise, based on our findings, we cannot confirm the assumption regarding more diversified participants' skills and knowledge in online hackathons (Braune et al., 2021; Usher & Barak, 2020; Wang et al., 2021). Similarly to tools, there are no noteworthy differences in online and offline participants' skills. The most common ones are python, java, javascript, and c++. However, we checked overall online and offline participants' skills, not the diversification of teammates' knowledge inside one group, which would be more helpful and should be addressed in our future work.

As a practical implication, organizations should pay higher attention to geographically diverse teams as their amount in online hackathons is much bigger than in offline ones that lead to higher spatial and time zone differences that make it harder to collaborate, especially without the growth of cloud and online tools usage. This led to the following suggestion, organizers should not only provide but also promote usage of cloud technologies that can improve intergroup collaboration, as of now we can see that teams becoming more diverse and useful tools (for example GitHub) are still in low usage. The gap between winners' announced stage and judging seems to be unnecessarily big and can be reduced by organizers to decrease tension among teams and keep them as potential participants for upcoming events. Analysis of participants' skills can also give a broader understanding of the most widespread attainments among team members that can be of use for further hackathons organizations.

6. Limitations

The data in Devpost is self-reported, so people can add any information that they deem necessary, and there is no way to check if the provided information is accurate or not. Moreover, our data only consist of a snapshot at a specific point in time when it was collected. It is thus possible that people have moved since participating in a hackathon which means that the location might not be the place they participated from. They might also change the skills in their profile which again affects the reliability of the analysis.

The data about participants' locations in Devpost is highly unstandardized - geographic objects can be written in different languages or even in a mix of languages. To calculate the distance between these locations, OpenStreetMap API was used. However, not all locations were found with the full name, for example, "Tiloniya, RJ, IN". Therefore full names of not found locations were split into three separate entries (e.g. 1: Tiloniya, 2: RJ, 3: IN). If the full location was not found, we searched for the first entry (in most cases city or state), then for the second, and then for the last one. However, this algorithm can give a wrong location latitude and longitude in rare cases. In the given an example, the latitude and longitude refer to Rio de Janeiro instead of the city in India.

The following limitation is closely linked with the previous one. Suppose the location latitude and longitude were not defined correctly. In that case, the time zone will not be correct as far as we are using object latitude and longitude as inputs for the time zone detection function.

As far as one project can be submitted to multiple hackathons to avoid inconsistency, we removed all project duplicates except for the first submission. However, this decision might slightly affect numbers in RQ3 and RQ4.

Overall, online hackathons' growth linked a lot to the spread of COVID-19 and subsequent restrictions on public gatherings. Therefore, it is logical to assume some effect on gathered statistics from coronavirus breakthroughs, for example, in such topics as hackathons themes and judging criteria.

7. Conclusion and future work

In this study, we investigated the differences between online and offline hackathons in terms of schedules and duration, themes, participating teams' and overall hackathons diversity in terms of geography, usage of tools, with a particular focus on online and cloud ones. We found that online and offline hackathons differ significantly in stages' duration. Overall, the schedule of online-based events varies from face-to-face ones only if a public voting stage is present. If this stage is omitted, the schedule has almost equal average duration and the time between steps.

Moreover, the difference between online and offline hackathons' themes was also significant. In online ones, there is a spread of such topics as "Beginner friendly", "COVID-19" and "Education", whereas, for face-to-face ones "Blockchain", "Productivity" and "Communication" rated higher. However, "Social Good" stays as the top-rated theme for both types of hackathons. One other leading topic is "Machine Learning/AI", notwithstanding a much higher ratio in offline events. While themes of hackathons differ a lot, top judging criteria, with slight changes in the ratios, stayed the same: "Design", "Originality", and "Creativity". The main difference is that "Technical difficulty" leads to offline competitions and is not that important in online ones.

There is a significant growth in participants from different countries, both in online and offline hackathons. However, if offline hackathons' significant impact is from European countries, for online ones, it is from such diverse countries as India, Australia, and Mexico. Consequently, intergroup spatial and time zone distance is significantly higher in online hackathons than offline ones, even though time zone distance is not as high as expected. One would expect that this diversity will increase the usage of online and cloud tools for better collaboration. However, we did not find any evidence for this. The ratio of GitHub repositories mentioned in offline projects appeared to be higher than in online ones.

There are several ways to extend this research, e.g. use not only Devpost as the database source, as it is an American platform and became popular in other parts of the world not so long ago. The study can also be extended to exact GitHub repositories of participants to understand if they are active there or not, how the code was changed during the hackathon and after, and most importantly, how online hackathons in these terms differ from offline ones. As for spatial and time zone distance, the isolation index can also be calculated to understand geographical differences better. If in a team of four people, three came from one country and only one from another, it is still much more accessible to collaborate than in a team where four people out of four are coming from all over the world.

Reference List

10 Reasons why Hackathon is Important to Make you Industry-ready | LinkedIn. (n.d.). Retrieved January 6, 2022, from https://www.linkedin.com/pulse/10-reasons-why-hackathon-important-make-you-sathiyanarayanan/?trk=public_profile_article_view

Bertello, A., Bogers, M. L. A. M., & De Bernardi, P. (n.d.). Open innovation in the face of the COVID-19 grand challenge: Insights from the Pan-European hackathon 'EUvsVirus.' *R&D Management*, *n/a*(n/a). https://doi.org/10.1111/radm.12456

Bouncken, R., Brem, A., & Kraus, S. (2016). Multi-cultural teams as sources for creativity and innovation: The role of cultural diversity on team performance. *International Journal of Innovation Management*, *20*(01), 1650012. https://doi.org/10.1142/S1363919616500122

Braune, K., Rojas, P.-D., Hofferbert, J., Sosa, A. V., Lebedev, A., Balzer, F., Thun, S., Lieber, S., Kirchberger, V., & Poncette, A.-S. (2021). Interdisciplinary Online Hackathons as an Approach to Combat the COVID-19 Pandemic: Case Study. *Journal of Medical Internet Research*, *23*(2), e25283. https://doi.org/10.2196/25283

Briscoe, G., & Mulligan, C. (n.d.). *Digital Innovation: The Hackathon Phenomenon*. 13.

Cobham, D., Gowen, C., Hargrave, B., Jacques, K., Laurel, J., & Ringham, S. (2017). *FROM HACKATHON TO STUDENT ENTERPRISE: AN EVALUATION OF CREATING SUCCESSFUL AND SUSTAINABLE STUDENT ENTREPRENEURIAL ACTIVITY INITIATED BY A UNIVERSITY HACKATHON*. 789–796. https://doi.org/10.21125/edulearn.2017.1172

Coursey, L. E., Williams, B. C., Kenworthy, J. B., Paulus, P. B., & Doboli, S. (2020). Divergent and Convergent Group Creativity in an Asynchronous Online Environment. *The Journal of Creative Behavior*, *54*(2), 253–266. https://doi.org/10.1002/jocb.363

COVID-19 transmission and protective measures| WHO Western Pacific. (n.d.). COVID-19 Transmission and Protective Measures| WHO Western Pacific. Retrieved December 21, 2021, from https://www.who.int/westernpacific/emergencies/covid-19/information/covid-19-new-normal

Devpost—The home for hackathons. (n.d.). Devpost - The Home for Hackathons. Retrieved November 30, 2021, from https://devpost.com/

Ford, R. C., Piccolo, R. F., & Ford, L. R. (2017). Strategies for building effective virtual teams: Trust is key. *Business Horizons*, *60*(1), 25–34. https://doi.org/10.1016/j.bushor.2016.08.009

Franco, S., Presenza, A., & Petruzzelli, A. M. (2021). Boosting innovative business ideas through hackathons. The “Hack for Travel” case study. *European Journal of Innovation Management*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/EJIM-06-2021-0300

García, T. M., & Menéndez, F. A. (2021). University Hackathons: New Ways of Educational and Social Innovation. The Experience of the University of Salamanca. *Journal of Higher Education Theory and Practice*, *21*(7), Article 7. https://doi.org/10.33423/jhetp.v21i7.4486

Gupta, V., & Rubalcaba, L. (2021). Creating business value from online hackathons. *IEEE Potentials*, *40*(5), 18–20. https://doi.org/10.1109/MPOT.2021.3088778

Hackathon planning kit. (n.d.). Hackathon Planning Kit. Retrieved January 1, 2022, from https://hackathon-planning-kit.org/

Hackathons are becoming a common event. (n.d.). The Daily Record. Retrieved January 6, 2022, from https://www.the-daily-record.com/story/business/2018/04/01/hackathons-are-becoming-common-event/12820137007/

Imam, A., & Dey, T. (2021). Tracking Hackathon Code Creation and Reuse. *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*, 615–617. https://doi.org/10.1109/MSR52588.2021.00085

Imam, A., Dey, T., Nolte, A., Mockus, A., & Herbsleb, J. D. (2021). The Secret Life of Hackathon Code Where does it come from and where does it go? *2021 Ieee/Acm 18th International Conference on Mining Software Repositories (Msr 2021)*, 68–79. https://doi.org/10.1109/MSR52588.2021.00020

Loeffler, E., & Masiga, C. (2021a). #WirVsVirus: Communities co-producing new solutions to meet COVID-19 challenges through a hackathon in Germany. In O. Williams, D. Tembo, J. Ocloo, M. Kaur, G. Hickey, M. Farr, & P. Beresford (Eds.), *COVID-19 and Co-production in Health and Social Care Research, Policy, and Practice* (1st ed., pp. 57–66). Bristol University Press. http://www.jstor.org/stable/j.ctv1p6hqk9.11

Loeffler, E., & Masiga, C. (2021b). *Communities co-producing new solutions to meet Covid-19 challenges through a hackathon in Germany* (O. Williams, D. Tembo, J. Ocloo, M. Kaur, G. Hickey, M. Farr, & P. Beresford, Eds.; pp. 57–65). Policy Press. https://doi.org/10.47674/9781447361794

NASA Visible Earth—Home. (2022, January 2). [Text.Article]. NASA Visible Earth. https://visibleearth.nasa.gov/collection/1484/blue-marble

Nolte, A., Pe-Than, E. P. P., Affia, A. O., Chaihirunkarn, C., Filippova, A., Kalyanasundaram, A., Angarita, M. A. M., Trainer, E., & Herbsleb, J. D. (2020). How to organize a hackathon—A planning kit. *ArXiv:2008.08025 [Cs]*. http://arxiv.org/abs/2008.08025

Olariu, C., & Aldea, C. C. (2014). Managing Processes for Virtual Teams – A BPM Approach. *Procedia - Social and Behavioral Sciences*, *109*, 380–384. https://doi.org/10.1016/j.sbspro.2013.12.476

O’Leary, M. B., & Cummings, J. N. (2007). The Spatial, Temporal, and Configurational Characteristics of Geographic Dispersion in Teams. *MIS Quarterly*, *31*(3), 433–452. https://doi.org/10.2307/25148802

Paul, K. (2020, April 2). ‘Zoom is malware’: Why experts worry about the video conferencing platform. *The Guardian*. https://www.theguardian.com/technology/2020/apr/02/zoom-technology-security-coronavirus-video-conferencing

Phillips, K., & O’Reilly, C. (1998). Demography and Diversity in Organizations: A Review of 40 Years of Research. In *Research in Organizational Behavior* (Vol. 20, pp. 77–140).

Powell, J., Bailey Hayden, L., Cannon, A., Wilson, B., & Nolte, A. (2021). Organizing online hackathons for newcomers to a scientific community – Lessons learned from two events. *Sixth Annual International Conference on Game Jams, Hackathons, and Game Creation Events*, 78–82. https://doi.org/10.1145/3472688.3472700

Smirnova Maria. (2021). *Devpost-hackathons-analysis* [Jupyter Notebook]. https://github.com/SmirnushaMaria/devpost-hackathons-analysis/blob/199306fc9adb7de0570de075f619923944be0afc/data_webscrapping_final.ipynb (Original work published 2021)

Soltani, P., Pessi, K., Ahlin, K., & Wernered, I. (2014). Hackathon – a method for Digital Innovative Success: A Comparative Descriptive Study. In *Proceedings of the 8th European Conference on Information Management and Evaluation, ECIME 2014*.

Usher, M., & Barak, M. (2020). Team diversity as a predictor of innovation in team projects of face-to-face and online learners. *Computers & Education*, *144*, 103702. https://doi.org/10.1016/j.compedu.2019.103702

Wang, S., Yeoh, W., Ren, J., & Lee, A. (2021). Learnings and Implications of Virtual Hackathon. *Journal of Computer Information Systems*. https://www.tandfonline.com/doi/full/10.1080/08874417.2020.1864679

Wei, K. (2007). Sharing Knowledge in Global Virtual Teams. In K. Crowston, S. Sieber, & E. Wynn (Eds.), *Virtuality and Virtualization* (pp. 251–265). Springer US. https://doi.org/10.1007/978-0-387-73025-7_18

Welcome to GeoPy's documentation! —GeoPy 2.2.0 documentation. (n.d.). Retrieved November 30, 2021, from https://geopy.readthedocs.io/en/stable/

Yarmohammadian, M. H., Monsef, S., Javanmard, S. H., Yazdi, Y., & Amini-Rarani, M. (2021). The role of hackathon in education: Can hackathon improve health and medical education? *Journal of Education and Health Promotion*, *10*, 334. https://doi.org/10.4103/jehp.jehp_1183_20

Appendix I. License

Non-exclusive license to reproduce thesis and make thesis public

I, Maria Smirnova,
(*author's name*)

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,

Hackathons are going online. What's changed?,
(*title of thesis*)

supervised by Alexander Nolte,
(*supervisor's name*)

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.

Maria Smirnova
*Tartu, **06/01/2022***