Leveraging Publicly Available Data to Enhance the Feature Space of the Effect Score

Group D1

Dashboard Link

Bugra Yilmaz, Dragos Pop, Fabian Frank and Pepijn Groenen
Data Systems Project 2022
University of Amsterdam

ABSTRACT

The municipality of Amsterdam (Gemeente Amsterdam) has the responsibility to send inspectors to check buildings on fire risk. Capacity on inspections is limited so some sort of ranking on fire risk has to be done to plan the order of inspections. Fire risk is defined as chance times effect, while the biggest gap of information in the current model of the municipality exists on the cultural effect. Cultural effect is very broadly defined in literature and the multidimensional nature makes it hard to measure. This paper focuses on how to make use of publicly available data to gain insight in the cultural effect on fire risk. Data sources used are Wikipedia, Tripadvisor, Google, and Flickr. The combined data forms the source of a cultural effect score which is then calculated per address in Amsterdam (BAG), and visualized in a Tableau dashboard. The stakeholders deem the result a valuable tool for planners to help them in deciding which buildings to have inspected first.

1 INTRODUCTION

The municipality of Amsterdam collaborates with the department of permits, surveillance and enforcement (Vergunningen, Toezicht en Handhaving, VTH) in conducting fire safety inspections for the buildings of the city. One can think of an inspection as a list of checks that the object has to meet in order to be compliant with the highest degree of security in case of a fire incident. For instance, one check ensures that the fire exit signs always have the lights on, while another one concerns the materials of building's critical walls are inflammable. Given that there are approximately 500.000 assets in Amsterdam, the municipality implemented a risk-based scoring system that evaluates the buildings in a data-driven way. This system produces a ranking that prioritizes the buildings with a larger chance of an incident and effect of a potential incident to be visited first by inspectors. In other words, for every building within the city, the municipality is generating a score through the formula $risk = chance * effect$, appointing inspectors to first visit the objects that scored the highest.

When it comes to the second term of the formula, "effect" is composed of four elements, namely social, economic, infrastructural, and cultural effect. Concerning the last one, it only takes into account if a building is a monument (Dutch or UNESCO) and its construction year, which was regarded as a research gap since the cultural value of a building should depend on more than two features to allow a thorough comparison of two assets from a cultural standpoint. Moreover, in approaching quantifying the cultural value of an object, it was discovered the high degree of subjectivity associated with computing such a measure.

Preliminary research has been done to understand how to possibly measure cultural value. According to Wijesinghe [22], cultural value often is intangible and influenced by its historical significance and context, age, design, visual impact and size, architectural and engineering sophistication. Nijkamp [19] looks at the effects cultural heritage has to assess value, but this approach would involve a lot of observation or working with questionnaires. He notices that to assess the value of cultural assets most information on the attributes is qualitative, soft, or fuzzy in nature, which makes it more or less 'unmeasurable'. Mason [17] states that heritage values are by nature varied and often in conflict, and that no single method or discipline gives a sufficient assessment of heritage values. A combination of methods of various disciplines is required for any comprehensive assessment of the value of heritage, but even then the subjectivity and contingency of heritage value make it difficult to establish a value. The views of experts, citizens, communities, governments and other stakeholders often are not aligned, while another difficulty is the changing nature of value. Mason concludes that heritage values cannot be objectively measured and broken down like a chemist does to substance. Choi [7] has a similar view, namely that effective measurement of cultural value often is very difficult because of its multidimensional nature. Throsby [21] goes even further by saying that models are unable to appropriately and adequately estimate cultural value, because next to its multidimensionality it is unstable, contested, lacks a common unit, and often contains elements that cannot easily be expressed according to any quantitative or qualitative scale.

We discussed the difficulties of cultural value measurement and the multi-interpretability of it with the stakeholders, and they agreed we could shift from a strict view of cultural value to some sort of popularity measurement. The following study will show how publicly available information is used to enlarge the feature space in such a way that a better effect score is generated for each building, ultimately contributing to more accurate ranking of objects that should be inspected first.

2 METHODOLOGY

The proposed solution for the previously explained gap is based on the idea that there is unstructured information publicly available online that can positively contribute to a more accurate effect score. The sources of information picked for the tasks are among the most popular websites on the Internet, namely Google, Wikipedia, Tripadvisor, and Flickr.

The infrastructure of the proposed system can be seen in Figure 1. The system is composed on two main parts, namely the scraping and visualizing. The first one starts by extracting names of public assets within Amsterdam from Wikipedia that will play as the rows the first dataset. Since the assets are public, it was assumed that the popularity of a public building reflects its cultural value. Thus, extra features from Wikipedia, Tripadvisor, Google and Flickr are added to represent interest in the respective objects. The described process is depicted in the upper branch of the diagram, following the black arrows into the POI dataset.

However, since other buildings are also important, their addresses were taken from the Raw BAG dataset and the value of each feature was computed considering the public assets in the vicinity and aggregating the respective variables together. One can see this in the diagram following the blue arrows.

Finally, the two derived datasets are visualized in a Tableau dashboard that allows the user to apply various filter and attribute more importance to certain features if needed.

Wikipedia

Wikipedia is an online encyclopedia that can be edited by its users. It makes a useful data source for this project because of its extensive content on culture, history and society [3].

Scraping a list of culturally relevant POI names. Wikipedia contains lists of pages that are part of specifically defined categories. We scraped the following lists that contain cultural relevant buildings in Amsterdam:

- List of monuments in Amsterdam(“rijksmonumenten” and “gemeentelijke monumenten”)
- List of sights in Amsterdam
- List of subjects in Amsterdam (“monuments and buildings”, “memorial monuments”, “museums”, “concert halls and theaters”, “cinemas”, “parks” and “markets”)

Some of these lists contained some pages that are not buildings in Amsterdam. These were filtered out by checking if the word “Amsterdam” or “Amsterdamse” occured in the first 300 characters of the page summary. We noticed that all pages on Wikipedia start with a one sentence summary that almost always contains the name of the city where the building is located. A small test has been conducted: of the 100+ pages in the List of sights in Amsterdam, 6 pages had been filtered out. All of them were not buildings in Amsterdam.

Scraping features from Wikipedia.
The following features were scraped from either the Wikipedia API or the Wikimedia API. Wikimedia is a site that gives access to metadata from Wikipedia.

- Wikipedia Views (NL and EN). This is the amount of views that the Wikipedia page received between 01-01-2002 and 01-01-2020. These have been extracted separately for the English and the Dutch version of the page. There are many factors that influence the amount of views, among them: general popularity, events and incoming links. In this project specifically, it is interpreted as an indication of online interest in a culturally relevant POI from the Dutch and English speaking world respectively.
- Wikipedia Edits. This is the amount of times that the dutch version of the page is edited. Increased readership potentially increases the number of edits, but the main reason we incorporated this feature in our model is that editing a page requires some form of commitment to its subject. This commitment then comes from people with affinity with the subject. Cultural value is not only a measure of popularity, it also requires specialists and enthusiasts to have some interaction with a subject. However, some extra noise was introduced to this feature as a small proportion of pages are protected [4]. This protection implies that less people are permitted to edit and thus the number of edits will decrease.
- Wikipedia Languages. This is the amount of languages in which a page is available. This feature can be used as an indication of the international online interest in its subject.

Tripadvisor

Cultural buildings attract many visitors, not only tourists but also the locals. Marinakou and Giousmpasoglou [16] recommend social networks and Tripadvisor for the promotion

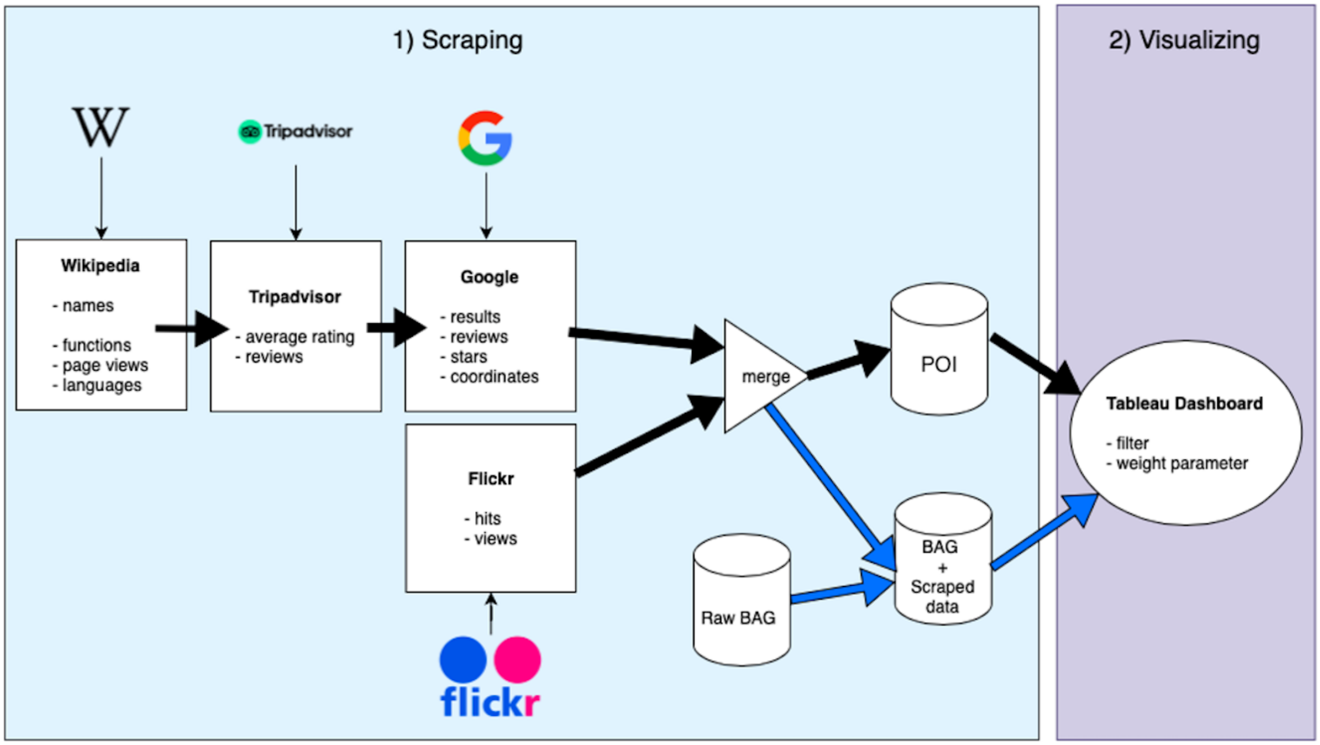

Figure 1: System Infrastructure

of cultural tourism in Bahrain. The findings of the paper show that travellers and visitors collect information from Tripadvisor about key cultural attractions. It suggests the local authorities use Tripadvisor to attract more visitors. For this project, Tripadvisor can help us to reflect the cultural value of the buildings.

Tripadvisor is a platform where users can leave a rating and review about a place. The main three categories are Hotels, Restaurants and Attractions. Each category has its own ranking. The ranking is based on quality, recency and quantity (). The quality has to do with the average rating of a building, recency is about the recent reviews and the quantity covers the volume of consistent positive feedback. The ranking of a place could be an interesting feature to scrape, but this becomes ambiguous as multiple points of interest could share the same rank since they are in different categories. To mirror the ranking feature, the two features that will be extracted are the average rating and the number of reviews per point of interest. The first feature is an indication of how positive a POI is perceived. The second feature gives us an indication of the number of visitors. The computed Tripadvisor score from these two features could be an approximation of the rankings for our points of interest.

Web-scraping techniques are used to extract the features from this platform. To explain the procedure in short, we begin with the points of interest list created by the previous section. The Python packages used are BeautifulSoup, requests and HTMLSession. The web-scraping will be iterated over the POI. The method of scraping consists of two parts, one about finding the right Tripadvisor page and the other for scraping that specific page. The starting point for the first part is Google Search. The reason for this starting point is the fact that the Tripadvisor Search page generates different URLs for different searches. Since these URLs are obnoxious and cannot be easily interpreted, we opted to use Google Search. By feeding Google Search the point of interest with 'tripadvisor' added, we obtain a page of related websites. There are several things to look into. First of all, the Tripadvisor pages need to refer to Amsterdam. Additionally, even if the website is about Amsterdam, we need to be sure that the selected page is not a recommendation page. The latter is a page where 'nearby locations' or '10 other places' are recommended. In most cases where a place has no Tripadvisor page, a restaurant or cafe is caught as the destination place. Unfortunately, it is advised to manually screen the obtained Tripadvisor URLs. For example, since Muntplein does not have its own Tripadvisor page, a restaurant or cafe that's nearby the location is chosen as the destination page. Some ways to deal with this problem is to use the string editing distance. The package 'NLTK' contains the module 'edit_distance'. We need to distinguish whether the current POI is a restaurant or not. Some cafes or restaurants include 'Amsterdam' in the title, therefore we used a threshold of 9. If the editing distance is 10 or more, one can say that the current POI is not a restaurant but the obtained URL is. Another example why there were no additional restrictions imposed is for example the 'He Hua Temple'. The Tripadvisor title for this place is 'Ho Guang Shan Temple'. As the names may differ, the places are still the same. As there is a bag of stopwords in the NLTK package, we manually set up a list of 'stop places'. These places were manually screened and added to the list, so that they can be skipped when running the code multiple times. Once the right URLs have been obtained, we can continue with the second part of the method, the actual scraping of the features. As we are scraping HTML websites, we need to know how the 'HTML bodies' of these pages are created. There are four types of categories: Attractions, Restaurants, Hotels and 'ShowUser'. Knowing that the right Tripadvisor URL has been obtained and to which category a POI belongs to, the scraping process can begin. The only difference between the categories is in which 'divs and classes' the two features are stored. On the Tripadvisor page, the name, average rating and number of reviews are scraped and stored into the dataframe.

Google

Google Search is the most popular search engine, with a global market share of 92% [1]. For this reason, the platform can be seen as a reliable source of information, hence, it was used to further scrape four additional features for the assets derived in the [Wikipedia section], namely the number of reviews, their average rating, the coordinates of a building, and the number of results retrieved after searching the name.

The intuition behind selecting the respective features is as follows: the number of reviews received should be a good indication of the number of visitors of the respective building, based on the assumption that generally someone would not review a place he/she has not visited. One corner case one can think of is when the respective place is the office of a company that is drop-shipping a certain product, in which case the respective product is the one reviewed rather than the place itself. However, this is rarely the case given the names of the assets were extracted from Wikipedia to represent public buildings and monuments.

Secondly, the average rating is seen as a reflection of the reviewer's perception of the particular point of interest. In comparison with the previous variable, the average rating is not as reliable since there are many factors that can influence the rating, as one may rate the overall experience while someone can evaluate a place focusing on a specific element. Nevertheless, the statistical power of the true average rating increases with the size of the sample, hence, for the assets with many reviews, the appraisals and criticisms should balance each other.

Next, the coordinates of a building were extracted for visualization and identification purposes, while the number of results retrieved by Google were regarded as a measure of interest because a building with cultural value would have many articles referencing it which would be returned by the engine. When it comes to the retrieval of coordinates, it was discovered that the button reading "maps" in the bar below the search box contains a link towards the Google Maps page which contains the coordinates of the respective place.

In order to scrape the previously described features, the requests library was used to get the HTML of the webpage returned by Google after searching the name of one place. To avoid as much as possible fetching results of a different object with the same name, the word "Amsterdam" was added at the end of each query. In other words, adding "Amsterdam" after "Nieuwe Kerk" ensures the results point to the church in Amsterdam and not the one in the Hague or Delft. Unfortunately, this issue cannot be fully resolved because Google will also retrieve documents about the Nieuwe Kerk from the Hague that contain the word "Amsterdam" within the article. After the request generates a response, this is parsed using Beautiful Soup package.

One important design choice that happens at this point is extracting all the features in one request, rather than one feature per request. This is done to reduce the number of requests as much as possible because Google is using a complex system of request rate limitation, making scraping a challenging task after the limit is passed [2]. On the same note, splitting the data in batches of ten records and adding a time buffer of ten seconds between each request in a batch and ten minutes in-between batches were implemented to mimic the behavior of a real user and avoid passing Google's scraping limits.

A strategy employed by Google Search to restrict scraping their results is tampering the format of the HTML response page. For this reason, error handling played a significant role in the performance of the scraper. After running the scraping first time, 65% of the reviews, 16% of the average ratings, and 19% of the coordinates for the 202 assets were missing. In addressing this, the scraping script was adjusted for the pages with missing reviews, approximately halving the number. Consequently, 32% of the reviews, 16% of the average ratings, and 19% of the coordinates were missing afterwards. These were then filled in manually since they represent a manageable amount of approximately 130 entries in total.

Lastly, it was found that Google does not return the same number of results for the same query run at different points in time. In order to deal with this, the number of results were scraped three times for each building and averaged with the hopes of making the feature more reliable. However, more iterations would definitely contribute to this scope.

Flickr

The Flickr API is used to scrape the desired data, every result consists of 'id', 'owner', 'date taken', 'date uploaded', 'number of views', 'title', 'tags', 'geometry', and a link to the picture itself in 500px resolution. Data scraping is set to 'per day' (e.g. November 20, 2014), because the API has a limit on the amount of results that are given back. This way we avoid missing certain results, since the daily upload of pictures from within Amsterdam is low enough to never exceed the API limit. The Flickr API gives back the geolocation in latitude and longitude, which is converted to geometry points to be able to plot the data. The borders of Amsterdam are imported from data provided by the government [5]. The geometry points of the Flickr dataset are then projected onto the borders of Amsterdam, and any entry with a point outside of Amsterdam borders is dropped.

If a location is duplicated by the same user, the picture with the most views gets selected while the duplicates are dropped. By taking a picture on a certain point the location is deemed important by the Flickr user already, any extra pictures taken on exactly the same location are considered noise. For example, one user uploaded pictures of rugby games for years, accounting for thousands of pictures on that location. Only his most interesting picture on that location (the one with the highest number of views) got selected into our cleaned up dataset. We want to measure how interesting a location is to different Flickr users, not how interesting it is for a single user. By using that method the amount of pictures is reduced from 561770 to 297167.

It is considered very likely that users uploading photos that are taken within the same city or country over a long period of time are locals, while users that upload photos that are taken only during a short period of time are tourists [9]. Generally it is accepted that a tourist is residing in the same place for a maximum of 30 days [9, 11–13, 15]. If a single user has taken all of his pictures of Amsterdam within 30 days that user is considered a tourist, else he is considered a local. One could argue that tourists take pictures on generally higher rated locations than locals, since tourists have to be selective on which locations they visit because of time limitations. Research done by García-Palomerez et al. (2015) supports this theory by identifying tourist hot spots based on social networks. They came to the conclusion that tourist photos are taken mainly around the main monuments and other tourist attractions representative for the city, while photos taken by locals extend to parks and recreational areas [9]. This is further supported by the study of Ginzarlya et al. (2018) in which they separate tourists and locals based on the country filled in at the user's profile [10]. It could be said that mainstream touristic locations are found attractive by tourists, and less-known hidden places are discovered by locals. Therefore data of both tourists and locals could be useful, while data separation is desirable.

DBSCAN, which stands for Density-Based Spatial Clustering of Applications with Noise, is used to cluster the data and detect outliers. It works on the assumption that clusters are dense regions in space separated by regions of lower density [8]. DBSCAN thus has the ability to group pictures together without being limited to a certain shape or pattern, by looking at the local density of the data points. A cluster has to consist of at least 'X' amount of pictures within maximum 'Y' meters distance. If that requirement is met any additional data point within range gets included in the cluster. We found that setting 'X=10' and 'Y=100' gives us the best balance between calculation time and precision, we don't want too much data filtered out but also not too little (see Figure 2). Karayazi et al. (2021) uses the same 100m radius for clustering within Amsterdam [14]. If a pseudo-cluster doesn't get the required 10 pictures those individual data points will be considered outliers and eventually dropped from the dataset. The number of local pictures gets reduced from 209203 to 14977 while the number of tourist pictures gets reduced from 87964 to 3709.

The clustered data shows patterns expected for tourists and locals. To further support our clustering method we analyzed the amount of tourist pictures taken per year, since there should be a huge dip visible in tourist numbers for 2020 and 2021 because of corona [6, 20]. The share of tourist pictures for 2020 and 2021 indeed dropped immensely as can be seen in Figure 3.

Data Integration and Visualization

Once all the features have been scraped, we end up with the 'fgtw-data.csv' which contains the features for 202 points of interest. This dataset will be used to calculate the feature scores for the addresses in the BAG dataset, which contains over half a million addresses. The line of thought can be explained as the effect of the neighborhood on the points of interest. If one of the addresses that is close to a POI catches fire, this will have an immediate impact on the environment and thereby the POI. The computed dataset is 'bag-fgtw.csv'. We will create two dashboards based on each dataset, one for the points of interest and the other for the BAG. The visualization tool that was used for this purpose is Tableau. The idea is to create an interactive dashboard that reacts to the user-input. The landing page of the dashboard includes a short description of all (eleven) scraped features. Two buttons

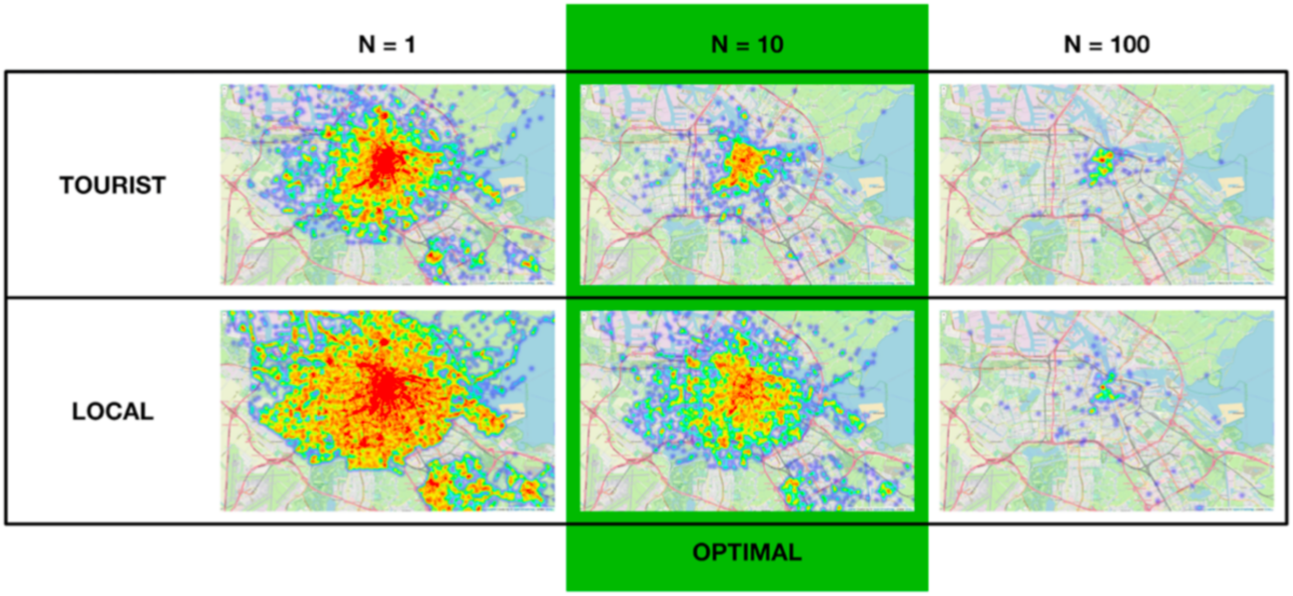

Figure 2: DBSCAN clustering

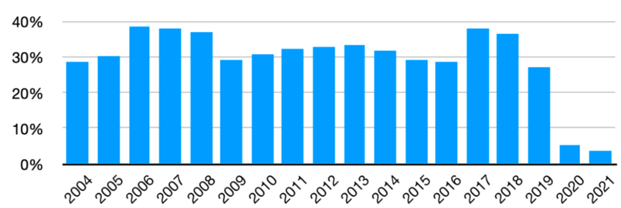

Figure 3: Percentage of pictures taken by tourists in Amsterdam starting from 2004

are provided, one for the POI dashboard and the other for the BAG dashboard.

The places are ranked based on a normalized score. Define formula $x_{i,j}$ and $w_{i,j}$ as respectively the feature score and weight for $i \in \{F, G, T, W\}$ and for $j = (1, ..., n_i)$. Since Flickr and Tripadvisor have two features, $n_F = n_T = 2$. Likewise, for Google and Wikipedia we have $n_G = 3$ and $n_W = 4$, respectively. We do not only put weights on all eleven features, but also on each source i. Thereby, we define $w_{i,0}$ as the weight of source i. Let's define z_i and y_i as respectively the normalized and overall scores. The overall scores are calculated in the following way: $y_i = w_{i,0} * \sum_{j=1}^{n_i} w_{i,j} * x_{i,j}$. Once we have obtained an overall score for every place, we apply normalization: $z_i = \frac{y_i - min(y)}{max(y) - min(y)}$. Hereby, the scores have an upper and lower bound of 1.0 and 0.0, respectively. Both dashboards show a heatmap based on the final normalized scores.

As shown in Figure 4, the POI dashboard contains in total fifteen parameters that impact the final normalized scores and three filters. These three filters are Name, Label and Top N. Four out of fifteen parameters are the weights of each source $w_{i,0}$, namely Wikipedia (in gray), Flickr (in blue), Tripadvisor (in green) and Google (in red). These four parameters can be found in the top right of the POI Dashboard. By changing these values to 0, the total effect of the source on the final score becomes zero. As long as the value of these parameters are greater than zero, the partial weights can be similarly adjusted to impact the final score. The other eleven partial weights can be found in the top left.

Let's take the following scenario into consideration. We are interested in the people that have written a review about a place or have taken a picture of a place. Since Wikipedia does not have any of both the options, the weight of Wikipedia can be set to 0. The partial weights for the features Flickr Views, Tripadvisor Rating, Google Rating and Google Results can be set to 0 as well. The obtained normalized scores are now based on the features Flickr Count, Tripadvisor Reviews and Google Reviews.

Finally, the BAG dashboard (see Figure 5) is similar to the POI Dashboard. The main difference is that the BAG dashboard does not have the weights of all eleven features. These weights are set to 1. The user can still change the four weights of the sources, namely wi,0. An additional parameter is the Flickr Visitor Type, namely Local, Tourist or All. The addresses can be filtered by Purpose, Type and Top N.

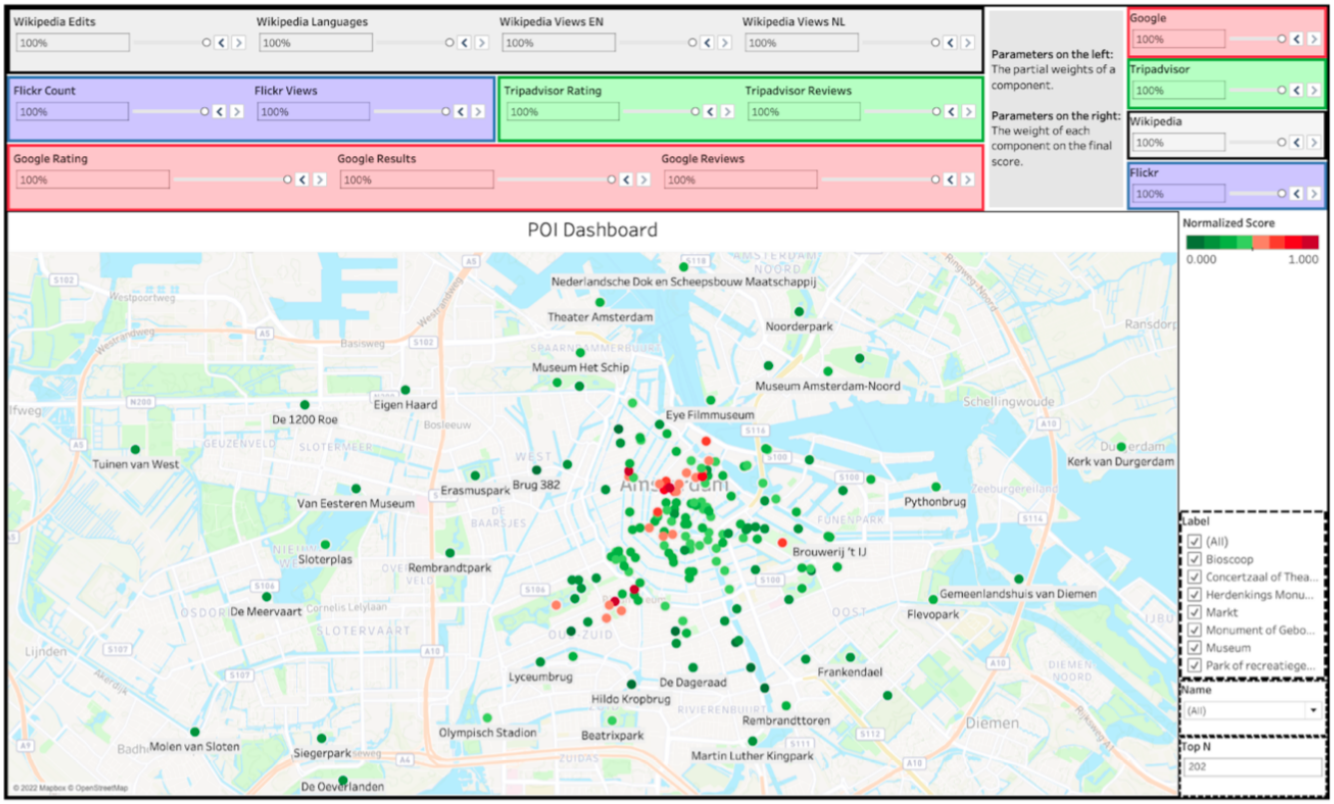

Figure 4: POI dashboard

User Validation

Our product will be used by the people at the fire department of Amsterdam that plan the fire inspections. However, because of the subjectivity surrounding the "cultural value score" as discussed in the introduction, control of the model's parameters have been put in the hands of the user. Thus, our system is designed for users that either have some knowledge on Amsterdam's culture or can cooperate with domain experts.

A summative user validation experiment has been conducted in order to gain inside a user's experience of the feature space and its visualization. Thereby validating how well, given the user's knowledge on Amsterdam's culture, the model estimates this perceived popularity value. And, validating the UX of the visualization's interface. We managed to recruit 2 participants that fulfill the following criteria: 1: They have a clear idea of what the planners of the fire department want out of the tool. And 2: They have an affinity with Amsterdam's culture. These participants were asked by the facilitator to execute a series of tasks on both the BAG and POI Dashboard. As they executed these tasks they were asked to think out loud while their comments were recorded. These tasks included using parameters and filtering options, reflecting on the system output as well as answering open questions on the interaction with the dashboards. A complete list of the tasks and details on the participants has been added as an appendix to this report. The result section consists of observations where participants walked into issues while executing their task. Also, we have taken into account all the feedback they gave while executing tasks or answering questions.

3 RESULTS

In general questions about the interaction with the dashboards, participants reacted positively. They are easily accessed online and participants find they contain decent indications of cultural popularity. The BAG dashboard was viewed as a useful tool for planners of the fire department to select buildings that are in need of extra caution. The problem with this dashboard is it's lack of explainability. That's where participants saw the added value of the POI dashboard. In combination these two tools were viewed as a useful tool for the planners. However, the dashboards still have some issues and improvements that will be described in this section.

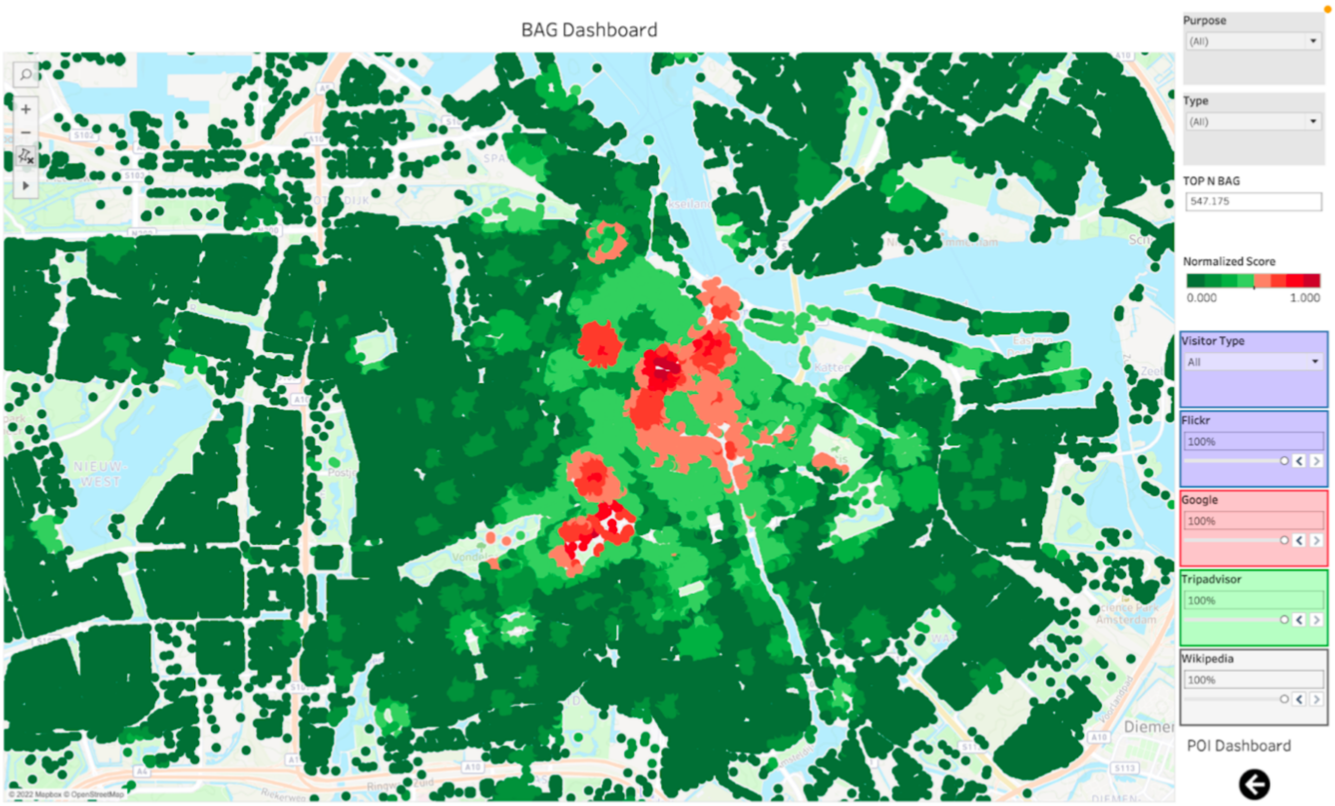

Figure 5: BAG dashboard

Both of the test participants mentioned a problem with the distribution of the normalized feature values. When singling out some of the features it became clear that some features values only use a small portion of their full range. We soon realized that the normalization method that we use is causing this issue, in combination with outliers. These outliers set the min or max value to an extreme, making the rest of the data fall within a limited range.

At first one of the dashboards was called 'monument' dashboard, but there are also non-monuments on it. One of the test participants proposed we should rename it 'POI' (Point Of Interest) dashboard.

Participants also mentioned we should have a list of top-10 results with building details visible on the dashboards, and not hide it. To make room for that they proposed to have less options for parameter adjustment visible, for planners the subparameter adjustment is not desirable anyway. Making the sub parameters an option on both dashboards would give some needed consistency between them as well. Users should be able to easily see what the grouped parameters consist of, with for example an explanation that the Wikipedia weight is consisting of pageviews, edits, etc when you hover over or click on it.

When participants were asked how the feature values match with the participants' expectations, they generally respond positively with the exception of some outliers. Google Results seem unrepresentative of real popularity because of high and unrealistic scores on for example 'Amsterdam Museum' and 'The Movies'.

The main takeaway on letting the participants reflect on single feature values is that rating is not useful as a standalone parameter. These values have a very limited range and are rounded to a single decimal. The results of this is that both the google and the TripAdvisor rating contain very little information.

Both test participants saw value in making a difference between tourists and locals, because it scored as expected with locals more divided over the city. Depending on the target audience it can be picked accordingly.

A participant mentioned that a particularly useful feature for planners is the top-N results in the BAG dashboard, but it should be expanded to a two-option-slider (so you could set both a maximum and a minimum). Planners would use this feature to gradually finish off sets of buildings that have high cultural relevance. They would do top-1000 first, then pick 1000-2000, etc. Our current implementation of the top-N is not suitable for that. We precomputed the range which took two weeks to calculate in a Jupyter notebook, Gemeente Amsterdam proposes to have it running live from a database so that the range could be adjusted in real time.

It is also suggested by the participants that we propose some basic values for the parameters that can be reset, instead of simply putting a 100% weight on each parameter in the initial configuration. We initially made the choice to leave this to the domain experts, but given this feedback we decided to provide it in the discussion.

4 CONCLUSION

Did we solve the stakeholder's needs?

This project succeeded in filling up a gap in available data with regards to the cultural effect score that estimated the cultural damage when certain buildings in Amsterdam fly on fire. We did so by scraping features from the web that could indicate online popularity area's and culturally relevant POI's.As confirmed by one of the participants in the user validation: "the score the model produces is a cultural popularity score, not a cultural value". We also created a tool that let's the planners of the fire department gain inside in this data, enabling these planners to make a more accurate ranking of objects that should be inspected first. The BAG dashboard can be used as a tool to select buildings that are in need of extra caution. The POI dashboard can be used in order to gain inside in the features that make up the effect score.

Conclusions of the User validation

Next to the perception of the general effectiveness and use cases of our model and it's visualization, the user validation also facilitated insight into the strengths and flaws of the current implementation. Notable flaws being the current normalization method and the rounded rating features. Notable strengths were the top-N feature in the BAG dashboard and the way in which the system output matches the user's expectations.

5 DISCUSSION

For this project, the platforms Flickr, Google, Tripadvisor and Wikipedia were scraped to resemble the cultural value of places. The main discussion point could be the credibility of these platforms, as the features may not be representative of the entire population. One has to bear in mind that as much as the cultural value of a building is subjective, so are the scraped features. Another critique would be the credibility of Flickr, as it shows a declining user base since 2017. Nevertheless, as described in the previous section, the users found the system usable and representative.

On the normalization problem we encountered in the user validation

One of the problems that participants of the user validation encountered was the unbalanced distribution of the normalized feature values. Because of the min-max normalization method we used, outliers had a major impact on the range of the other values. For some features, this meant that almost all data entries had a similar color coding in the visualization. To fix this issue, a possible solution would be to rank the data entries with regards to each feature and then use the normalized ranks as the new value.

Distribution of the weights and adjustments in the parameters

The end-users liked the idea to influence the rankings by changing the weights. However, in the case of the POI Dashboard, there were many weights that could be adjusted. This is quite some information to digest. The end-users had some of their own intuitions to change the weights. Our recommendation would be to keep the weight of Flickr at 50% and the other three sources combined at 50%. The reason for this would be the fact that Flickr has significantly more spatial data. Some features could have been combined, such as the Dutch and English Wikipedia views. By providing the option to put weights on these two features, the granularity of the partial scores becomes too detailed. Furthermore, the number of Wikipedia edits also give a good indication of how popular a certain page is, as is the amount of pageviews. Because the amount of possible languages is limited and also subject to diligence of users that create pages, we prefer to give more weight to edits. Since the scraping of Google Results provided different values for every query run and thereby being less reliable, we advise to give less weight to the Google Results than the number of Google Maps reviews. For Flickr, the number of pictures and number of views of those pictures give roughly the same results, but it's nevertheless good to have some combination of it so possible outliers get averaged out a bit more. We propose to take both evenly distributed into a Flickr parameter.

Quality of rating features

The user validation pointed out the google and Tripadvisor ratings contain very little information. They have a very limited range and are also rounded to one decimal. This can be dealt with by either not including these features in the score calculations, or coming up with a customized normalization method for these features. The problem that was discussed before, concerning the combination of the current min-max normalization and outliers, particularly applies here as well.

6 FUTURE WORK

As pointed out in the Results section, the web-scraping is not live and needs to be executed manually over time, which will lead to inefficiencies. A 'live' database needs to be connected to the dashboard in order to feed these features. What makes this process more complex is the fact that the features are scraped from the internet. A database is required that can be updated daily, if possible by the minute.

According to [18], making use of 2 participants in the user validation is not enough to catch all usability problems. According to his estimation, 2 participants together find 50% of usability problems. Therefore, It would be wise to execute the validation 2 more times.

The POI dataset can be scaled up if needed. There are multiple ways of doing so. One option would be to scrape additional culturally relevant POI names from Wikipedia. The lists of museums can be expanded or the list of churches to name a few. The other option would be to make use of existing data entries that have missing values. The full list of Wikipedia pages that have been scraped consists of 427 data entries instead of the 202 in the POI dashboard. The reason for not including these entries is because they have missing values. In a large majority of cases, this is because the Wikipedia page does not have a corresponding Tripadvisor page. You can choose to set these missing values to zero because not having a Tripadvisor page indicates a lack of popularity. This will most likely double the amount of data entries.

REFERENCES

[1] [n.d.]. *Google Search*. https://en.wikipedia.org/wiki/Google_Search

[2] [n.d.]. *Search engine scraping*. https://en.wikipedia.org/wiki/Search_engine_scraping

[3] [n.d.]. *Wikipedia:Contents*. https://en.wikipedia.org/wiki/Wikipedia:Contents

[4] [n.d.]. *Wikipedia:Protection policy*. https://en.wikipedia.org/wiki/Wikipedia:Protection_policy

[5] Gemeente Amsterdam. [n.d.]. . https://maps.amsterdam.nl/open_geodata/?k=202

[6] Gemeente Amsterdam. [n.d.]. *Derde raming bezoekersaantallen Amsterdam 2018-2021*. https://onderzoek.amsterdam.nl/publicatie/derde-raming-bezoekersaantallen-amsterdam-2018-2021

[7] Andy S Choi, Franco Papandrea, and Jeff Bennett. 2007. Assessing cultural values: developing an attitudinal scale. *Journal of Cultural Economics* 31, 4 (2007), 311–335.

[8] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *kdd*, Vol. 96. 226–231.

[9] Juan Carlos García-Palomares, Javier Gutiérrez, and Carmen Mínguez. 2015. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography* 63 (2015), 408–417.

[10] Manal Ginzarly, Ana Pereira Roders, and Jacques Teller. 2019. Mapping historic urban landscape values through social media. *Journal of Cultural Heritage* 36 (2019), 1–11.

[11] Fabien Girardin, Filippo Dal Fiore, Josep Blat, and Carlo Ratti. 2007. Understanding of tourist dynamics from explicitly disclosed location information. In *Symposium on LBS and Telecartography*, Vol. 58. Citeseer.

[12] Fabien Girardin, Andrea Vaccari, Alexander Gerber, Assaf Biderman, and Carlo Ratti. 2009. Quantifying urban attractiveness from the distribution and density of digital footprints. *International Journal of Spatial Data Infrastructures Research. 2009; 4: 175-200* (2009).

[13] Bálint Kádár. 2014. Measuring tourist activities in cities using geotagged photography. *Tourism Geographies* 16, 1 (2014), 88–104.

[14] Sevim Sezi Karayazi, Gamze Dane, and Bauke de Vries. 2021. Utilizing urban geospatial data to Understand heritage attractiveness in Amsterdam. *ISPRS International Journal of Geo-Information* 10, 4 (2021), 198.

[15] Athanasios Koutras, Ioannis A Nikas, and Alkiviadis Panagopoulos. 2019. Towards developing smart cities: Evidence from GIS analysis on tourists' behavior using social network data in the city of Athens. In *Smart tourism as a driver for culture and sustainability*. Springer, 407–418.

[16] Evangelia Marinakou and Charalampos Giousmpasoglou. 2016. Using Tripadvisor© for exploring cultural tourism development in Bahrain. (2016).

[17] Randall Mason et al. 2002. Assessing values in conservation planning: methodological issues and choices. *Assessing the values of cultural heritage* (2002), 5–30.

[18] Jakob Nielsen. 1994. *Usability engineering*. Morgan Kaufmann.

[19] Peter Nijkamp. 1988. Culture and region: a multidimensional evaluation of monuments. *Environment and Planning B: Planning and Design* 15, 1 (1988), 5–14.

[20] Parool. [n.d.]. *Coronaklap voor toerisme kwam het hardst aan in Amsterdam*. https://www.parool.nl/amsterdam/coronaklap-voor-toerisme-kwam-het-hardst-aan-in-amsterdam~b1cd14f2/

[21] David Throsby. 2003. Determining the value of cultural goods: How much (or how little) does contingent valuation tell us? *Journal of cultural economics* 27, 3 (2003), 275–285.

[22] Thilan Wijesinghe. 1993. A proposed Methodology for Measuring the Economic Value of Cultural Monuments. In *Economics of conservation*. p–149.

A DETAILS ON THE USER VALIDATION

- Test participants: Martijn de Jong (Gemeente Amsterdam), Vasilis Deligiannis (Universiteit van Amsterdam)
- Observers:Fabian Frank, Pepijn Groenen

Both user validations are recorded, raw files are available on request.

List of tasks

Dashboard 1 (POI)

1. Go to the Tableau home screen and open the 'Monument' dashboard
2. Filter on 'label' and select only 'Museum' and 'Bioscoop'
3. What is currently the building with the highest score?
4. Reset the view with all buildings visible again
5. What is the Google Rating of Vondelpark?
6. Change the parameter 'Google Results' to 0
7. Adjust the weighting of the overall parameters so that Google and Flickr are 2x more important than TripAdvisor and Wikipedia
8. Check if the ranking of the buildings has changed by displaying a list of the buildings and sort them by score from highest to lowest
9. Display the top 100 buildings on the map itself
10. Set all subfeatures to 0 and set only 1 subfeature to maximum at a time and repeat for each subfeature. What strikes you and do you think these features are a good indication of cultural value?
11. Take a good look at the buildings we used in our Monument dashboard. Do you have examples of certain buildings or types of buildings that you are still missing but could be culturally relevant?

Dashboard 2 (BAG)

1. Go to the Tableau home screen and open the 'BAG' dashboard
2. Filter on 'purpose' and select only 'Sportfunctie' and 'Sportfunctie|Winkelfunctie'
3. Filter by 'type' and deselect everything with 'Sportzaal' in it
4. Select a number of buildings that score remarkably high and try to explain those scores
5. Look at the buildings that score remarkably low and try to explain those scores
6. Reset the data by selecting all buildings again
7. Display only the data related to tourists on the map and compare it with the data from locals. What strikes you most about this?
8. Display all data again, then set the dashboard to display the top-10.000 buildings on the map
9. Would you also like to be able to adjust the subfeatures (number of page views, etc) in this dashboard and why/why not?

General questions

1. What is your initial response to the dashboards?
2. What is possibly misleading about our ranking of buildings relative to their actual cultural value?
3. Do you have any other recommendations and/or comments?