Evaluating the influence of park attributes on user satisfaction from Flickr data for Southeast Michigan

By

Jieping Yang

A practicum submitted

in partial fulfillment of the requirements for the degree of

Master of Landscape Architecture (School for Environment and Sustainability)

at the University of Michigan

August 2022

Abstract

The use of social media data is widely used by academics for research, including landscape architecture. The increased use of social media for expressing opinions about public open spaces presents opportunities to analyze this data from social media to evaluate park user satisfaction. Here we propose the following research question: Is there a correlation between the park attributes and the sentiments of visitors towards the landscape, and if so, can this data be used to inform future designs? To answer this question we use data from Flickr, a popular social media platform, as it offers the most viable data for landscape research since photos with attached descriptive titles or tags contain the geographic information which can be presented in geographic applications like ArcGIS, and it has no limitation of sharing duration.

To conduct the research data from Flickr was accessed via an API. The subsequent data was cleaned, processed, and analyzed via several linear regression models. The results indicate some attributes of the parks have a strong relationship with user satisfaction: Golf course, Hunting Trapping Area, Natural Area, and Monument Historical Features. The landscape designers can focus more on those important attributes as a design guideline for the future, since the result of all the regression models indicates that those attributes influence significantly on user satisfaction.

Acknowledgment

I am grateful to my advisor, Dr. Mark Lindquist, Assistant Professor of Landscape Architecture, who has been providing support during my preparation for this research project, not only useful suggestions, but also mental help. Then I would like to thank Dr. Nathan Fox, Research Fellow at SEAS, who provided indispensable technique support and encouragement for me. In the end, I would like to thank my boyfriend, Xuejian Wang, and my friends, Bowei Li, Yue Zhang, Tianyu Wang, Jianbo Gu, and Changjiu Xia, who are also my family in 1425 McIntyle. Together they accompany me to have a wonderful time at the University of Michigan - Ann Arbor.

Table of Content

List of figures

List of tables

1 Introduction

The use of social media is growing at an alarming rate and people create content, share, bookmark, and network on social media, including Facebook, MySpace, Digg, Twitter, and JISC listservs on the academic side (Asur et al.2010), and the use of social media data is widely used by scientists in the research field.

There are several significant research contributions from social media data, including a better understanding of landscapes, easier access to large amounts of data, and a predictive model which can be used widely. The use of social media data can help us understand human-nature relationships as well. Much research currently using social media is focused on understanding and quantifying social media interactions, while this project will focus on how park attributes influence the spatial experience of a landscape design, and how their presence impacts overall satisfaction based on the Flickr dataset.

The New Zealand Institute of Landscape Architects Best Practice Note identifies three categories of landscape attributes: Biophysical elements, sensory, or perceptual qualities, and associative meanings and values including spiritual, cultural, or social associations (NZILA Education Foundation 2010). For visitors, there will be more experience with sensory or perceptual attributes than the other two categories, and the reflection on a landscape design based on their sense of experience is a good method to improve the further understanding of landscape to inform guidelines for future designs. Based on this reflection, this research will focus on people's attitudes towards different sensory or perceptual qualities based on existing parks in Detroit. This research is a comparison of landscape attributes among a variety of parks on a very large scale.

The research question that guides the project is:
Is there a correlation between the landscape attributes and the sentiments of visitors towards the landscape, and if so, can this data be used to inform guidelines for future designs?

Until recently people lacked the understanding of how to use social media in the landscape research field overall speaking (Shirtcliff 2015), while a variety of research based on social media data has been conducted in different disciplines gradually. For landscape architecture and architecture, there are several relevant studies. Li & Yang (2022) reviewed the academic literature on social media's applications in landscape research. The conclusion includes (1) Social media data can be used validly instead of the traditional methods, and can save time and budget while capturing the intangible and subjective dimension of ecosystem services; (2) Geospatial location, text information and photos are the main parameters in use; and (3) Most research now focuses on the regional scale and nonurban areas, which lacks landscape experience from the visitor's perspective (Li & Yang 2022).

Currently, for site-scale study areas, the spatial information contained in social media data was not used in the majority of research (Li & Yang 2022). The significance of the research presented here is to address this research gap and to provide a guideline for future design toward city open space from a site scale. The public open space plays an important role in improving human well-being, especially for migrant communities, open space brings a visual shared recognition to the residents (Rishbeth 2016). The city of Detroit is a typical migration city since there were large numbers of southern migrants during the American Civil War, and today's version of Detroit is still under migrant conflict (Galster 2012, Eisinger 2013). Given the significance of open space for Detroit and the opportunities which are brought by vacant land in this city, a guideline for open space design is important. The reflection of people's sentiments summarized through social media data could help designers better understand and facilitate the design process when considering the different park attributes.

2 Literature review

In this literature review, three topics will be addressed. The first section provides an overall summary of the history of the city of Detroit and the evolution of its open space. The second provides an overview of the use of fast-developing social media and how people use social media data in the research field. The third section provides an overview of logistic regression and how it is used in the research field.

2.1 Overview of the City of Detroit

The City of Detroit was the fourth largest city in the United States in 1920 (Nolan 1999). Because of the migration of people and industry away from the city, estimates ranging from a low of about 24 square miles to a high of about 40 square miles in Detroit is regarded as vacant land today, considering it is a 139-square-mile city (Gallagher 2019). Detroit is one of the famous examples of shrinking cities in the twenty-first century, along with the significant problems of urban sprawl and inner city decline (Xie et al.2018). The city of Detroit has lost 60% of its population in the past 60 years and lost 25% of its population in the past ten years (Neill et al.2015). According to vacancy data from the census from the Neighborhood Change Database (NCDB 1970–2010) in 2010, the median vacancy rate over the census in the city of Detroit was 26.6%. Detroit lost 58% of its population between 1970 and 2010, while its suburban population increased by 27% (Benfield 2011). The Southeast Michigan Council of Governments reported that between 1965 and 1995, the area of developed land increased by nearly 30%, but its population decreased by 23% (Lui et al.1999).

More recently the city of Detroit filed for Chapter 9 bankruptcy on July 18, 2013. According to the research by McDonald (2014), two major forces impacted the metro area of Detroit: basic market forces and cumulative decline forces. Basic market forces explained the suburbanization from 1950 to 1970, while the cumulative decline forces were dominant in the succeeding decades (McDonald 2014). Ryznar and Wagner (2001) used remotely sensed imagery to detect the changes in Detroit over these years, and they identified the changes in green vegetation for 640 square mile area between two dates, which are 05/10/1975 and 05/16/1992. From the urban planning perspective, the result shows that net vegetation has increased these years (Rynzar et al.2001). This phenomenon indicates that the possible flow of future development could be a population shift from the inner city to the urban area, with the increase of city open space.

According to the book “Driving Detroit” by Galster 2012 and the review of this book by Eisinger 2013, there are several reasons leading to today's situation. Firstly, its excessive dependence on the auto industry caused no fallback while facing a challenging diversity economy. A survey showed that Detroit ranks 95 among the 100 largest cities in the proportion of its population with college graduates degrees, because in the 60 decades last century, the auto industry provided good employment and good living conditions for its residents. Secondly, after World War II, the government encouraged suburban sprawl by leaving the city, which produced the highest level of metro deconcentration problem today. Thirdly, the race problem between white and black residents is also a major reason. Many white Detroiters were Southern migrants to seek job chances in the war in the first half of the last century, which results in today's version of Detroit. Because of the different thoughts between white and black, for instance, white communities without any regional public transit system led to a hodge-podge bus route in Detroit (Galster 2012, Eisinger 2013).

In recent years, researchers focus on how to resolve or release vacant land and shrinking cities' problems. According to Nassauer and Raskin's research, new planning approaches and concepts should be applied to reclaim vacant lands, such as urban agriculture, green infrastructure, and open space planning for vacant urban lands (Nassauer et al.2014). Under Christine and Robert's research, they found that residents prefer to live close to city open space, which can attract migrants to the city through questionnaire surveys (Vogt et al.2004). This conclusion could be used in solving vacant land problems. A major strategy for developing a sustainable city is to expand green infrastructure (Lennon & Scott 2014). Green infrastructure is defined as the development of urban green spaces, such as parks, rain gardens, and greenways, which provide various social and ecological benefits. Recently, the government of Detroit has ambitious policies to develop green infrastructure on a large scale. For instance, expensive stormwater management infrastructure projects are on the agenda in Detroit, which can combine traditional infrastructure elements with various landscape attributes together (Berkooz 2011). According to Meerow et al research conclusion, now the green infrastructure in Detroit is not being used for maximizing the benefits of ecosystem services (Meerow 2017). According to Draus et al research, the residents in Detroit are going to reimagine and remake the city, facing the major challenge of making use of existing open space in ways that contribute to the city's economic redevelopment (Draus et al 2019, Ryan 2012).

Meanwhile, the idea of food cultivation is also being used to address vacant land. According to Colasanti et al (20101), 75% of fruits and vegetables needed by residents in Detroit could be grown locally, even with a limit on the growing season. In addition, Detroit residents indicated the most important thing is the "urban feel" and what constitutes such an aesthetic (Colasanti et al. 2010). The significance of Urban Cultivation is to find a model best fitting a viable economic enterprise and socially green land use in city life. This model also brings big opportunities and challenges to Detroit, especially since cultivation has proven workable in the city, and the next step is to balance the "urban feel", aesthetic, and urban pace with urban cultivation. As for the problem of whether the government will grant permission to city farming, Choo and Kristen discussed its practicality. Urban farming needs time to realize its value since the farmers cannot use heavy amounts of chemical fertilizers and pesticides in the city. The authors give the opinion that without longevity, people will not put time and effort into it, so urban farming is more like a land-based social system. The conclusion is that from a legal point of view, urban cultivation is practical, but ownership is hard to define, which is case by case (Choo et al.2011). Given the success of urban gardening to solve the vacant land problem, the major current issues of urban gardens and the institutions legally have been addressed, and land trusts and other nonprofit organizations are working on these issues (Schukoske 1999).

Over the past few decades Government organizations and environmental groups have made efforts to improve the conditions in Detroit. Detroit Future City (DFC) has conducted research regarding the situation of city space, and it addressed how to best use the city's vacant land. The large-scale plan includes two programs, the Land use and sustainability program and the Community and economic development program (Detroit Future City). The land use and sustainability program include:

1. "Achieving an Integrated Open Space Network in Detroit". This strategy advances the vision of transforming vacant land into an open-space network by outlining a set of consideration factors and giving a framework for the new city style.
2. "Open Space in Detroit". The report helps to transform the vacant land into open space by changing the ownership structure.
3. "Land + Water Works". This campaign is education and installation of green stormwater infrastructure to help residents.
4. "Owner's Guide to Bioretention". This strategy provides details on how to plan for and implement a bioretention basin on non-residential properties.
5. "Working with Lots Program". This is a program that financially supports the installation, activation, and maintenance of the new design for Detroit residents.

The Community and economic development program includes:

1. “Detroit Neighborhood Housing Compact”. DFC will model after the successful initiative in Chicago, and the detailed plan will include bringing together public, private, nonprofit and philanthropic stakeholders to take collective action and regularly collaborate around strategies for building healthier housing markets in Detroit’s neighborhoods.
2. “Single- family Rental Housing Study”. This study which works with the University of Michigan has developed recommendations for incorporating safe, affordable single-family rental housing into a stable and revitalized strategy.
3. “Adaptive Reuse of Industrial Property”. DFC will work with a broad group of public and private-sector partners to reduce blight, mitigate environmental hazards, create neighborhood jobs, and promote healthy, safe, and sustainable neighborhoods.
4. “Technical Assistance and Collaboration”. DFC will lend support and expertise on diverse issues, such as commercial corridor revitalization, economic mobility, entrepreneurialism, equitable development, community land trusts, and neighborhood planning.

This Detroit Future City plan will provide large challenges and also opportunities to this city, which will provide more possibilities for our research topic. Since Detroit has an extensive problem of vacant land, it presents an opportunity for urban development, where green infrastructure is the main strategy (Nassauer 2014, Berkooz 2011). There are plenty of projects in addition to the projects mentioned above that point to a wave of new construction and infrastructure in this hopeful city. The new policies and strategies to address the situation in Detroit could bring variabilities to the development of open space, and also bring possibilities to our research question.

2.2 Overview of the use of social media for perception research

In the past few years, people lack understanding of how to use social media in the landscape research field overall speaking (Shirtcliff 2015), while a variety of research based on social media data has been conducted in different disciplines. For example, in the public health field, there was research based on two geo browsers, Foursquare and google maps, including digital comments and reviews, and the author qualified the public opinion towards

Mecklenburg county using sentiment analysis to create positive community health outcomes (Dony et al.2020). Social media data is also used in demographic and marketing fields. There was a case study based on the application of Foursquare in Kansas City to do the census, and through census tract data there is still ethnic segregation which reflects the historical trends (Fekete 2017).

In the landscape architecture and planning fields, there are several profound pieces of research. A recent paper reviewed the academic literature on social media's applications in landscape research (Li & Yang 2022). For landscape architects, the challenge always is how to put themselves into users' shoes. To address this challenge, public participation has been used extensilty in landscape design tointegrate public perceptions into the design process and aims to elevate the quality of design (Cushing & Renata 2015). Flickr as a social media platform is the best fitness for landscape research (why?), however, Instagram, Facebook, and Twitter are improving their platform features to increase the competitiveness among those social media platforms (Li & Yang 2022). As for the scale and region, most of the research in the landscape field is based in Europe, (20 of 45 papers reviewed), with the remainder in Asia, America, Australia, and South America In addtion the majority of research is based on national parks, regional parks, and conservation lands (Li & Yang 2022).

Next, I will introduce some specific existing research papers in the landscape and urban planning field, to provide more thoughts and find the research gap in this field. There are different topics related to the use of social media in the landscape research field. Li and Yang divided existing research themes into four categories (Li & Yang 2022). The first theme is focusing on figuring out the effectiveness of social media data for landscape research. For instance, Sonter et al. (2016) illustrate the effectiveness of using social media data by the method of finding a significant and positive relationship between the number of visits on Flickr and survey data collected on site . On a continental scale, there was research evaluating the usefulness of different social media platforms, and quantifying landscapes, with the work ultimately introducing a predictive model for quantifying landscape values (Zanten et al. 2016).

The second theme is to develop indicators for landscape assessment from the aspect of people's sense of place, visual preferences, and appreciation for creation, which are all related to cultural ecosystem services. Researchers made assessments such as photo location’s point density, the number of photos, the distance to the site, and so on, to evaluate the function of cultural ecosystem services. This category of research is facing the challenges of bias, because active users may upload more photos than others, and younger groups are always more active than elders. To respond to those biases, there are several researchers, for instance, Gliozzo et al. (2016) developed a new model considering the efforts to travel to the place, the willingness to take a picture, the decision to geolocate the picture, and the action of sharing it through social media, and their work indicated the potential possibilities and contribution of crowdsourcing information to landscape research . For instance, Song & Zhang (2020) evaluated Seattle Freeway Park using social media data to understand such a site-scale landscape design. The methodology they used is getting database construction from Instagram and then doing photo categorization and hashtag categorization for analysis. The results would be how people react to a landmark or a landscape scheme, which could help understand a landscape design on a small scale (Song et al.2020).

The third theme is to discuss the influence of environmental factors on landscape values and seek to find the relationships. Because the landscape types and scales vary a lot, different studies have different landscape values. On a larger scale, the landscape variables include naturalness, landscape diversity, mountainous landscapes, and so on (Li & Yang 2022). According to research by Schreiner et al (2018) they geotagged photo metadata publicly shared on Flickr in Hawaii Volcanoes National Park about infrastructure and natural environment using MaxEnt modeling, and provide assessments that how changes in infrastructure and environmental factors may influence visitor use. For a smaller scale, researchers recently focused more on specific landscape feature variables, for example, distance to water, and proximity to slighting points (Li & Yang 2022).

The fourth theme is to use social media as tool for public engagement while designing. Social media transferred the traditional design process towards two-way communication between designers and users. One example used Khabarovsk, Russia as a case study using Instagram to gather information about people’s sentiment toward the temporary design on the public open space, and according to that information, the designers made modifications to the final design solutions, which is a typical and valued process of two-way communication (Paukaeva et al. 2020).

There are several advantages of researching on social media data. First, the voice on social media is more inclusive and extensive in capturing public opinions, since traditional ways of participation may emphasize experts' opinions, while social media is a more extensive method to gather data (Li & Yang 2022). Also, using social media data is more effective, since the process of data collection can minimize the potential biases compared with traditional methods (Chen et al. 2018), meanwhile, with the lowest investments of time and cost. Third, using social media data to do research can improve the design of landscape and urban planning, because it can record the unconscious landscape experience by visitors (Dunkel 2015) and reflect the relationship between the visitors and designers. Nowadays the expert opinion on landscape assessment dominates the practice and design, and individual landscape preferences recorded by social media could be added into practice to enrich the ways of landscape assessment (Chen et al.2018).

Traditionally how to use social media data effectively and efficiently is still challenging. According to Helles and Jensen's research in 2013, they find that data from social media often are not useful unless the researchers have been manipulated by proper algorithms for data collection and databases into usable records (Helles et al.2013). The data from social media may need to be cleaned which is time-consuming. For instance, Wilkie et al. (2020) found that only around 10% of Tweets are available for their research question during the data preparation process. Also, it is challenging to get detailed information about the user groups, such as ages, family status, and social status. The lack of that information gives the researchers difficulty to balance the bias of social media data. It is obvious that normally older adults and small children have less possibility to upload their opinions via social media platforms, then social equity issues may appear in the research (Langemeyer et al. 2018). Furthermore, 4.9m is the standard spatial accuracy by smartphone users with GPS in open space (Li & Yang 2022), so the efficiency of using spatial information is limited.

With so much information available about sentiments on social media, there are challenges about how to use them effectively. After collecting those sentiments, the researchers might need to absorb the available information and digitize it, which could be sent to different logistic regression models and linear regression models. In previous research, there are lots of methods and models to deal with the calculation. For example, in Kaczynski et al's research, they used logistic regression to examine the relative importance of park size, features, and distance. The method to measure the number of physical activities in their target 33 parks is computing these episodes together by coding. After the coding, they divided these parks into “parks with physical activity” and “parks without physical activity” (Kaczynski et al 2008). Simple summation was very effective in calculating the number of physical activities in each park in this research. In Giles-Corti et al’s research, the authors created a model to calculate a composite score of the parks in three domains: environmental quality factors, amenity factors, and safety factors, and they gave those factors different weights (Giles-Corti et al 2005). In Xiang et al’s research about online review platforms in tourism, they use a Naive Bayesian classifier, which gives each review a sentiment score between 0 and 1 presenting the two extremes of sentiment (Xiang et al 2017).

There are a variety of applied guidelines for social media research. For example, there was a discussion about different methods, such as quantitative methods, qualitative methods, and mixed methods, which extend the research process and link social media resources with other resources (Zeller, 2017). It is expected and practical that future applications, researchers, and design can make full use of the power of social media for public engagement (Li &Yang, 2022). A typical example is a community project, “Face Your World”. In this project, the author Jeanne van Heeswijk and Dennis Kaspori encouraged public engagement while designing (De Lange & De Waal, 2013). On the whole, social media as an effective tool offers a fresh take on landscape design, urban planning, and landscape research.

2.3 Overview of linear regression

Regression analysis is the most widely used model while analyzing multi-factor data. According to Montgomery et al’s book, the definition of regression analysis is “a statistical technique for investigating and modeling the relationship between variables (Montgomery et al 2021)”. Nowadays, the applications of linear regression are very wide, because it is an efficient and high-accurate analysis model for multi-factor data.

The linear regression model is one of the regression models, and its definition from Wikipedia is “In statistics, linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables) (Linear regression Wikipedia)”. There are various models of linear regression. The simplest model is simple linear regression, which only involves one regressor variable. The model could be written as the equation below:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1.1}$$

The equation (1.1) is called a simple linear regression model, where x is the predictor or regressor variable, y is the response variable, ε is a statistical error, β_0 is the intercept and β_1 is the slope (Montgomery et al 2021).

One most widely used model is the general linear model, also called the multiple linear regression model, which involves more than one regressor. The model could be written as the equation below:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta \;\; x \;\; + \varepsilon \tag{1.2}$$

The equation (1.2) is called a multiple linear regression model, where the βs indicate different parameters and all the other alphabets will be the same as the simple linear regression model. The model is employed as linear regression since it is linear in βs, not because y is a linear function of the xs (Montgomery et al 2021).

Along with the developments and advances in computing, the theoretical aspects of the regression model are to solve the math problem such as least squares and maximum likelihood firstly, and those types of problems are almost set in stone (Seber & Lee 2012). Nowadays, more wide uses have been applied in linear regression. An important objective of regression analysis is to estimate the unknown parameters (Montgomery et al 2021). But more importantly, a regression model does not imply a “cause-and-effect” relationship between the variables. Even if there might be a strong relationship between some variables, this cannot be evidence that the regressor variables and the response are related in a “cause-and-effect” manner. It can only be considered as an aid in confirming a “cause-and-effect” relationship but not a basis (Montgomery et al 2021).

Except for the basic theoretical aspects such as least squares and maximum likelihood, there are various practical applications of linear regression models. One of the continued applications is to make inferences about the relative importance of predictor variables (Nimon & Oswald 2013), which will be a good fit for this research. P-values and coefficients in the theoretical model can tell which relationships in the model are statistically significant, where the coefficients indicate the mathematical relationship between each independent and dependent variable, and P-values describe whether these relationships are statistically significant. If P-values are higher than the significant level, it indicates that there is no sufficient evidence to conclude that there is an effect at this level of significance between the variables. Normally, the significance level in statistics is 0.05, and if the P-values are less than this significance level, it indicates the sample data reject the null hypothesis. In other words, the data favor the hypothesis of non-zero correlation, which means that there is evidence showing the relationship between the variables is significant. As for the coefficients, their sign can tell if the correlation between each independent variable and the dependent variable is positive or negative. If the sign of the coefficient is negative, it means that if the value of the independent variable increases, the mean of the dependent variable will also tend to increase, and vice versa. This prediction model is all about determining how the independent variables influence the dependent variables, and it is determined by the P-values and coefficients (Jim Frost, How to Interpret P-values and Coefficients in Regression Analysis).

There are more and more applications of linear regression in the landscape research field, which is mainly used for landscape assessment and digital intelligent design. For landscape assessment, there are various pieces of research right now and this paper will also be an example of landscape assessment. In the very early stages of the model, some researchers used linear regression models to quantify the scenic beauty evaluations and regress the sets of quantitative landscape descriptors against scenic beauty estimates, to predict the scenic beauty of forested environments. As a result, they found that while all prediction models explained substantial portions of perceptual preferences, measures of manageable landscape features tended to have more significant relationships to mensuration parameters than did design features (Arthur 1977). In a later study, in Baek and Park's research in 2014, they employed a linear regression model when examining the associations between characteristics of green spaces, physical activity, perceived health status, and Body Mass Index, and the model indicates that the residents' physical activities are positively and directly influenced by the number of available public parks and green spaces, at the significance level at 10% (Baek & Park 2014). In Roth et al's research in 2018, the authors tend to examine the landscape impacts in the Strategic Environmental Assessment by the nationwide standardized scenic quality data. They employed a multiple linear regression model and recorded all the coefficients of the total 17 regressors. Also, the model showed a representative distribution in comparison with the total population distribution, with a P-value less than 0.001 (Roth et al 2018). The linear regression also could be interpreted in the climate change field. In Sangirova et al's research, the authors used the linear regression model to analyze the past and future climate change over Bostanliq district. The linear model represents that in 10 years there was a rise of 0.3 °C for the period, and it predicts within thirty years temperature would go up by 1.4 °C (Sangirova et al 2019). The linear regression model also could be used in intelligent design, which benefits landscape architects. In Harmon et al's research, they proposed Tangible Landscape —- a technique for rapidly and intuitively designing landscapes by geospatial modeling, linear regression, and simulation. They used a linear regression model when doing correct shifts in scanning and georeferencing (Harmon et al 2018).

Overview, the linear regression model always could be used to find the relationships between the independent variables and the dependent variables and to predict the future unknown parameters.

3 Methods

Social media posts from the website Flickr were downloaded by Flickr API, and these datasets combined with the datasets of the park information from SEMCOG (Southeast Michigan Council of Governments) were used to do linear regression, after a series of processes, such as data cleaning, data analysis, and statistical calculations.

3.1 Study area

The City of Detroit was selected as a case study for several reasons. First, Detroit is famous for its contribution to music, art, architecture, and design, with its profound historical background, which will bring more diversity to the categories of landscape attributes. Second, the Detroit river runs through the city, which offers more diverse landscape designs than cities without Waterfront landscape attributes. Third, due to the existing vacant land, many researchers and government officials make a great effort to improve the city space. Recently, there are plenty of projects for Detroit to be hopeful, since much vacant land has been changed to productive use and landscape space. For example, Riverside Park’s new skatepark, constructed in 2019, was made possible by a 2015 Land Exchange Agreement to expand the waterfront green space (Celebrate the grand opening of Riverside Park’s new skatepark). For another example, the Circle Forest project in 2022 located on Palmer between Elmwood and Moran, intends to bring a native meadow and 200 native trees to 1.3 acres of vacant land to connect the greenspaces in this community (Circle Forest). The new wave of construction in the City of Detroit allows us to do research among those large amounts of city parks since it presents an opportunity for urban development, where green infrastructure is the main strategy in this city (Nassauer 2014, Berkooz 2011).

3.2 Spatial and park amenity data

This research uses the open dataset, including the shapefile and extra information of all the parks which are under active use in the city of Detroit from SEMCOG (Southeast Michigan Council of Governments). This GIS dataset included 2657 parks in the city of Detroit, and the geographic location information is all included in these shapefiles. The extra information includes all these parks’ attributes information, such as whether the parks have these attributes or not, and the categories of attributes are seen in the table below. These attributes are features of parks that potentially influence people’s activities in these parks.

Amphitheater	Ballfields	Basketball courts	Beach	BMX area	Boating	Community recreation center	Concessions
Crosscountry skiing	Golf	Dog park	Equestrian activities	Farm garden	Fishing	Fitness equipment	Geocaching
Golf course	Gymnasium	Hiking nature trails	Hockey	Hunting trapping area	Ice skating	Indoor event facilities	Kayaking canoeing
Monument historic feature	Multiple athletic field	Museum interpretive center	Mountain biking trails	Natural area	Swimming pool	Pickleball court	Picnic shelter
Play area	Restrooms	Shooting range	Shuffleboard	Skate park	Sledding hill	Snowboarding	Soccer field
Tennis court	Track	Volleyball court	Walking biking trails paved	water park/spray park	Wildlife watching		

Table 1. The categories of the park attributes

Parks amenities vs. park size

To determine the relationship with number of park amenities and the park size, the dataset from SEMCOG was used to generate a scatter plot, which can show the distribution of those two variables, the number of park amenities, and park size.

Classification of the parks by size

Another possible factor is the size of the park. According to Kaczynski et al's research, the result indicates the size of the park will influence park-based activity significantly (Kaczynski et al 2008). Increasing the size of a park may increase the activities of human beings, the popularity of the park, and the volume of visitors. Various researches are indicating this conclusion. For example, based on Giles-Corti et al’s research, after considering the distance to public open space, the size was more important than attractiveness by a Logistic regression model. And this research found that a larger attractive park will attract more family activities than a smaller one. Furthermore, the larger park tends to have more attributes providing satisfying experiences, which lead to an experience called “lose oneself” in nature (Giles-Corti et al 2005).

The National Recreation and Park Association’s (NRPA) Park, Recreation, Open Space, and Greenway Guidelines defines park classifications including the size. The APD classifications are as follows: Small Parks, Neighborhood Parks, Community Parks, Special Use Parks, and Open Lands (Mertes & Hall 1996). Since here we only consider the factor of the size, we will focus more on the size of different types of parks. Small parks usually range from only 2500 square feet up to 1 acre. For a neighborhood park, 5 to 10 acres is considered a normal size. For a community park, the size is normally larger than 25 acres. For the special use parks, the size varies a lot because it depends on the purposes. For the open lands, it can be any size, but normally it contains very large sites. We divided our target parks into three categories by their size and these definitions are also referred in our classification. Those categories include small scale, which is up to 5 acres, medium scale from 5 acres to 25 acres, and large scale larger than 25 acres. We decrease the influence of the different sizes by dividing those parks into three different size categories, then comparing them in their category.

We divided those 378 parks which include valid comments and sentiments into three categories, according to the area information of each park from SEMCOG (Southeast Michigan Council of Governments). Because the information of those parks from SEMCOG might miss the important information, which we need to analyze based on, the parks with missing information were removed from our lists. After the cleaning, it only leaves 290 parks with valid information and valid comments from Flickr, which could be sent to do sentiment analysis. The pie chart below shows the percentage of different categories.

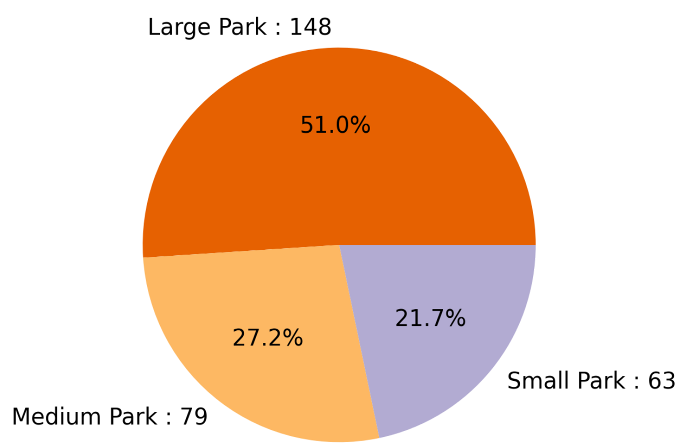

Figure 1. The pie chart showing the percentage and the number of different park categories

3.3 Social media data

The Flickr API was used to get the dataset of published opinions by visitors shared in those parks correspondingly on Flickr in the past fifteen years, from 05/12/2007 to 05/12/2022. This resulted in 23361 photos and comments in the past fifteen years published corresponding to its location. All these Geographic location information, parks' attributes information, and public comments on Flickr constitute the database construction of the research.

According to Li & Yang's review paper, Flickr data are particularly fit for studying people's behaviors and landscape perceptions (Li & Yang 2022), since photos with attached descriptive titles or tags contain the geographic information which can be presented into geographic applications like ArcGIS (Dunkel, 2015), and it has no limitation of sharing duration (Li & Yang 2022). This is the reason why Flickr is chosen as the social media platform for this research.

For these 23361 data sets downloaded from Flickr, there are some invalid data sets that we need to eliminate before analysis. The invalid data includes two types:

1) Blank description. Some users just uploaded the photos without a description, and this type of data should be eliminated. After this step, the data frame has 13862 valid descriptions.
2) Repeated description. Some users uploaded the same descriptions for a series of photos. To avoid repeated calculation of sentiments, we eliminate those data. After this step, the data frame has 3063 valid descriptions.

3.4 Analysis by CLASSECOL

For these data sets downloaded from Flickr, I use Classecol: vignette to analyze the sentiments towards those parks. Classecol is an R package to perform nature-related text classification of public opinion data, which is a text cleaning, processing, and classification tool to support the analysis of nature languages. Classecol can avoid the interpretation issues of sentiment analysis and can also identify relevant texts, describe the stance and declare the type of users that produced the texts. There are 16 different types of classifiers that can be trained on public data, and fall within these three categories: Hunting, Nature, and Biographical (Johnson et al.2021). These 16 classifiers' outputs will be merged by a logistic regression algorithm then the package generates an ensemble text classifier. The 16 different classifiers are shown below.

Negator	Amplifier	Deamplifier	Ad_conjunction
Land_eng(df$text)	Jockers_rinker	Jockers	Huliu
Loughran_mcdonald	Senticnet	Sentiword	Socal_google
Nrc	Afinn	Bing	Meanr

Table 2. The categories of the 16 classifiers under CLASSECOL

According to the analysis in this research, this package will be used to determine whether texts are relevant to nature, and if so, whether the sentiments are positive or negative. In the nature classifiers, overall accuracies ranged from 0.82 - 0.92, with moderate to high accuracy among all the categories except Pro-nature (negative phrasing) in the model (Johnson et al.2021). There are four categories under nature classification, and they are (Johnson et al.2021):

1. Irrelevant—text does not discuss nature or nature related activities.
2. Pro-nature (positive phrasing)—text endorses nature with positive language, for example, interest.
3. Pro-nature (negative phrasing)—text endorses nature with negative language, for example, concern.
4. Against-nature—text indicates opposition or frustration towards nature, for example, fear.

Taking an example from the datasets we have, there is a description, "Autumn colors on the trees in this historic district of Detroit". And the result from Classecol is "Pro-nature (positive phrasing)", and more detailed scores of different classifier models are shown below. Similarly, each description from those 3063 comments in 378 parks will have a result including a conclusion and detailed scores of different classifiers.

negator	amplifier	deamplifier	ad_conjunct	lang_eng(df:	jockers_rinke	jockers	huliu	loughran_m	senticnet	sentiword	socal_googl	nrc	afinn	bing	meanr
0	0	0	0	1	0.12	0.12	0	0	0.01	0.04	0	0	0	0	0

Table 3. An example of analysis result from Classecol

We calculated the number and the percentage of the sentiments, including irrelevant, pro-nature (positive phrasing), pro-nature (negative phrasing), and against-nature. The result is that there are 1808 pro-nature (positive phrasing) valid sentiments, 53 pro-nature (negative phrasing) valid sentiments, 1202 irrelevant comments, and 0 against-nature sentiments. The pie chart below shows the percentage of different sentiments.

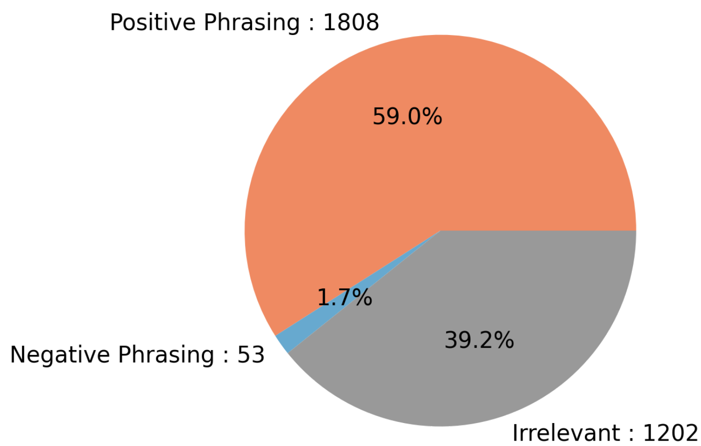

Figure 2. The pie chart showing the percentage and the number of different sentiments by CLASSECOL analysis

According to the categories we mentioned under the classification of the parks by size, we will focus on the analysis of the relationship between sentiments and park attributes under different categories, since the size of parks is a significant factor that influences the satisfaction of the parks. Doing experiments under the three categories, including Large, Medium, and Small will decrease the influence of the park size.

3.5 Calculating the score of each park

In previous research, there are various ways to give a score to a public open space based on the topic of the research. In Xiang et al's research mentioned under the literature review, their method, using the sentiment scores run by the Naive Bayesian classifier, is very referential to our research compared with other accumulative methods, because each comment sent to CLASSECOL will generate 16 scores by 16 different classifiers (seeing Table.2). In Xiang et al's research about online review platforms in tourism, they use Naive Bayesian classifier, which gives each review a sentiment score between 0 and 1 presenting the two extremes of sentiment(Xiang et al 2017). For the 16 classifiers under CLASSECOL, all the extreme scores are different among those classifiers.

Based on the previous research methods and the information about how researchers calculate social media sentiments under the literature review, the method to calculate the ratings of each sentiment will be: Using the Normalization method to make sure that each classifier has the same extreme from 0 to 1, then using the average score as the final score for each sentiment. After calculating each sentiment's scores, a cumulative method will be applied to get the score of each park.

Normalization is a method in statistics and applications of statistics. Normally, normalization of ratings means adjusting values measured on different scales to a notionally common scale, often prior to averaging (Normalization Wikipedia). The intention is that these normalized values allow the comparison of corresponding normalized values for different datasets in a way that eliminates the effects of certain gross influences (Normalization Wikipedia). In this research, since all the 16 classifiers have different scales and extremes, it is hard to compare and do calculations, and Normalization is a method that eliminates the influence of differences among those classifiers. After the normalization, we use the mean value of each classifier as the score for each sentiment. The higher the score is, the more positive the sentiment is. After getting the score of each sentiment, the method to calculate the final score of each park according to the satisfaction from Flickr is to calculate the mean value of each sentiment's score. Using this mean value as the final score has such advantages:

1. Making negative sentiment has a negative effect on ratings. All the scores in our calculation are positive, so even if there is a negative pro-nature sentiment, its score will always be positive. If we choose to add them up, the negative pro-nature sentiment will still bring positive ratings to the park's score.
2. It can decrease the effectiveness of different amounts of comments in each park. Since the amounts of comments we collected vary a lot, if we simply add them together, the parks' scores will have a huge difference.

These are the reasons we used the mean value as the final score of each park.

3.6 Linear Regression

The linear regression model can predict the relationships between the variables. P-values and coefficients in the theoretical linear regression model can tell which relationships in the model are statistically significant, where the coefficients indicate the mathematical relationship between each independent and dependent variable, and P-values describe whether these relationships are statistically significant. If P-values are higher than the significant level, it indicates that there is no sufficient evidence to conclude that there is an effect at this level of significance between the variables, normally the significance level is 0.05. This prediction model is all about determining how the independent variables influence the dependent variables, and it is determined by the P-values and coefficients(Jim Frost, How to Interpret P-values and Coefficients in Regression Analysis). The confidence interval in linear regression is a range of estimates for an unknown parameter. Normally, the 95% confidence interval is most commonly used, and it is used in this research. A 95% confidence interval means there is a 95% probability that the parameter lies within the internal (Wikipedia Confidence interval).

Before the linear regression analysis, since we have 46 attributes, we need to check if any similar attributes provide a similar contribution to the model. Variance inflation factor (VIF) is a measure of multicollinearity among the independent variables in a multiple regression model (Investopedia Variance Inflation Factor (VIF)). Mathematically, the VIF can be written as:

$$VIF = 1/(1 - R^2) \tag{1.3}$$

The equation (1.3) is called VIF ratio, where R^2 is the coefficient of determination of the regression equation (Wikipedia Variance Inflation Factor (VIF)). This ratio is calculated for each independent variable. The higher VIF, the higher the associated independent variables in collinear with the other variables in the model (Investopedia Variance Inflation Factor (VIF)). Normally, if the VIF of one variable is bigger than 10, we regard this variable as collinear with other variables, which could be removed from the regression to reduce ineffective work. See appendix E for more details about the VIF analysis for 46 attributes as independent variables in our model. It is obvious that all the VIFs of variables are under 10, which indicates that all the attributes have their contributions to the regression model, then we could keep all these variables sending to the linear regression.

Using the data after processing, we got three linear regression models for each category and a linear regression model for the whole park.

4 Results and discussion

4.1 Spatial and park amenity data

Park amenities versus park area

Through the observation of park amenities versus park area (Figure 3), we can see that from the overall review, along with the increase of the park area, the park amenities would possibly increase, except for the few outliers such as Sutherland-Wilson Farm and Belle Isle Park. See the image of the scatter plot below for more details. Since the size of a park can contribute a lot to the number of park amenities, then also affect the satisfaction of people, it is not appropriate to compare all the parks together, then we will need to classify those parks by size.

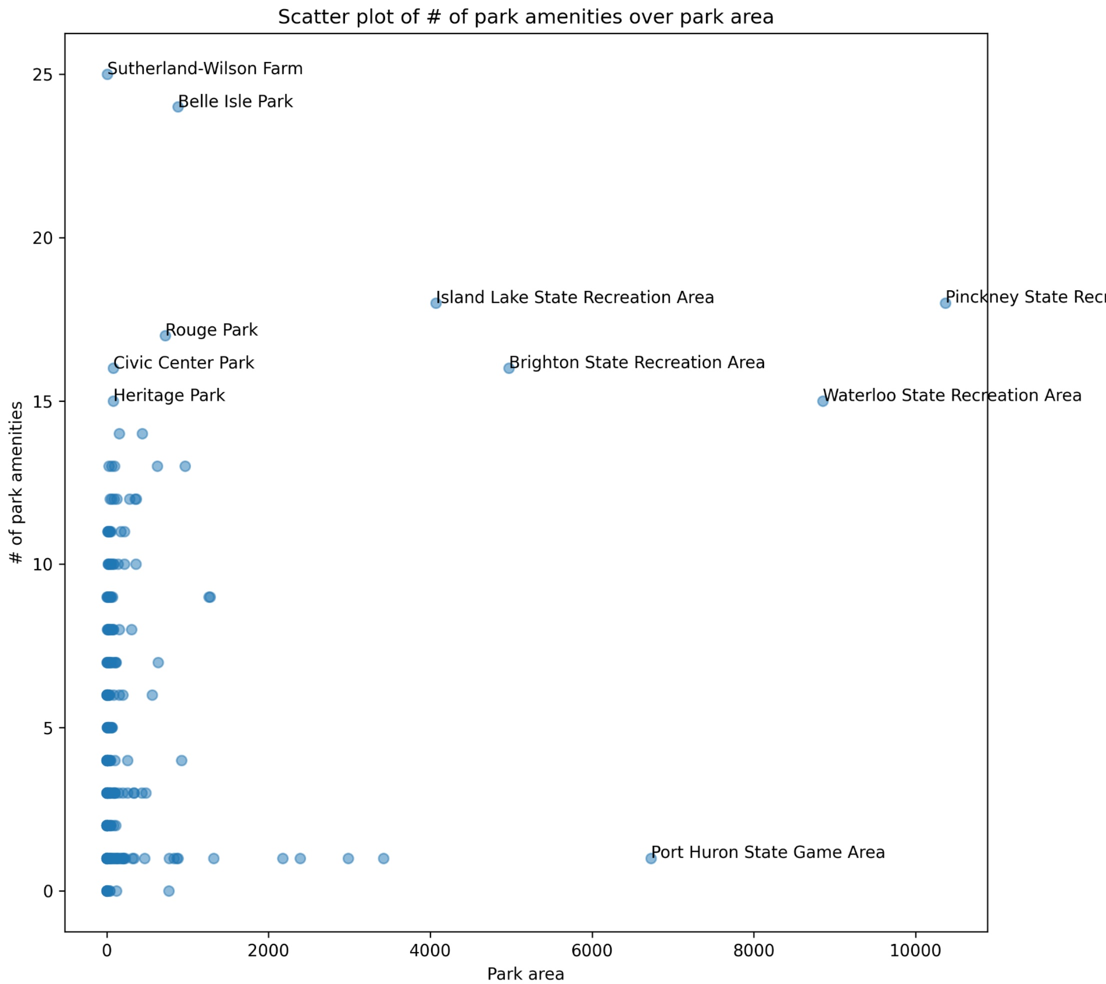

Figure 3. The scatter plot of the park average rating over the number of park comments

4.2 Social media data

In all 2657 parks in Detroit and 23361 comments published in those boundaries in the past fifteen years were collected (Figure 4). Due to the limitation of map size, the legend of red points only shows the point density of those comments, rather than showing the 23361 points on the map.

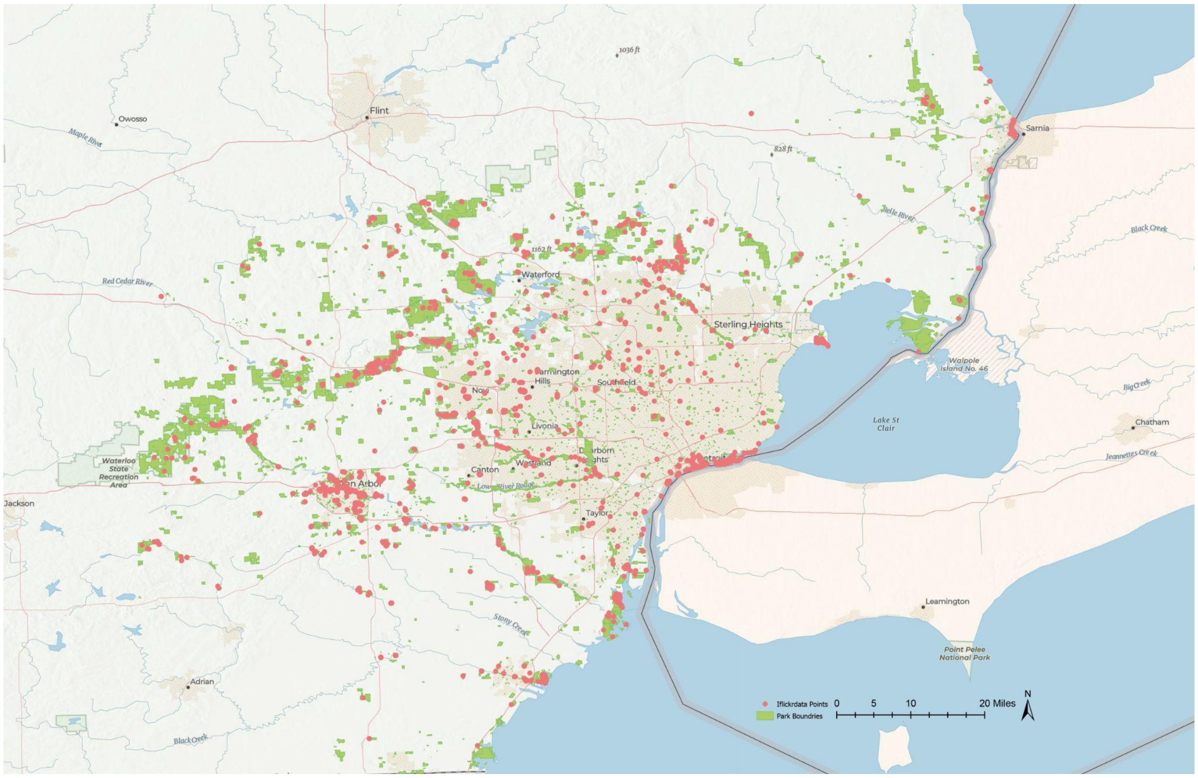

Figure 4. The map showing the geographic information of Flickr data points in 2657 park boundaries by ArcGIS

Data source: The park boundaries Shapefile from Southeast Michigan Council of Governments

The geographic information of Flickr data points from Flickr

Geographic Coordinate System: WGS 84

After data cleaning, our dataset includes 3063 valid descriptions with geographic information in 378 parks out of 2657 parks (Figure 5)., including the boundaries of those parks in Detroit and the geographic information of the valid 3063 comments published in those boundaries in the past fifteen years.

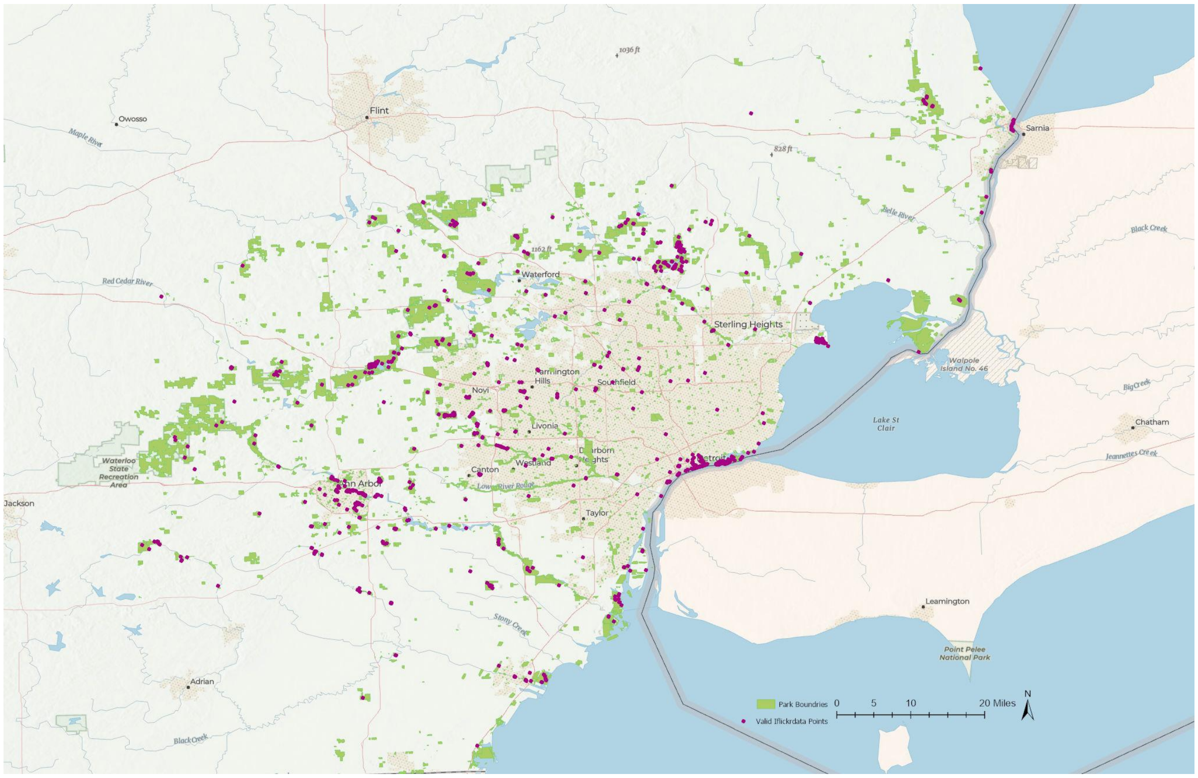

Figure 5. The map showing the geographic information of 3063 valid Flickr data points by ArcGIS

Data source: The park boundaries Shapefile from Southeast Michigan Council of Governments

The geographic information of Flickr data points from Flickr

Geographic Coordinate System: WGS 84

4.3 Sentiment analysis

The table below shows the descriptive statistics of the result scores of each sentiment after normalization. The Figure 6. below shows the distribution of the sentiment scores after normalization.

count	3063
mean	0.401122
std	0.087102
min	0.196672
25%	0.346710
50%	0.377764
75%	0.441876
max	0.909604

table 4. The descriptive statistics of the result scores of each sentiment

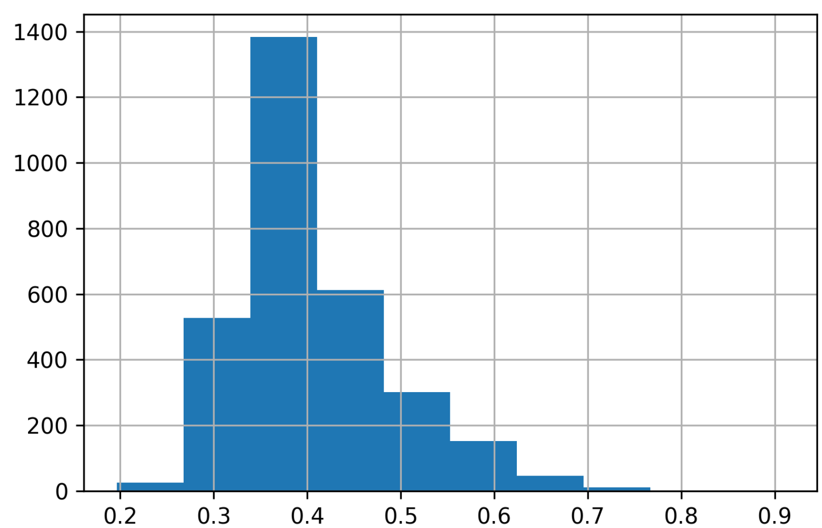

Figure 6. The distribution of the sentiment scores after normalization

Below is the scatter plot of the park average rating over the number of park comments. The parks which have more than 10 comments are selected to be shown, in order to show the data more clearly. We can see that Swift Run and Belle Isle Park are two outliers with a high rating and a high satisfaction.

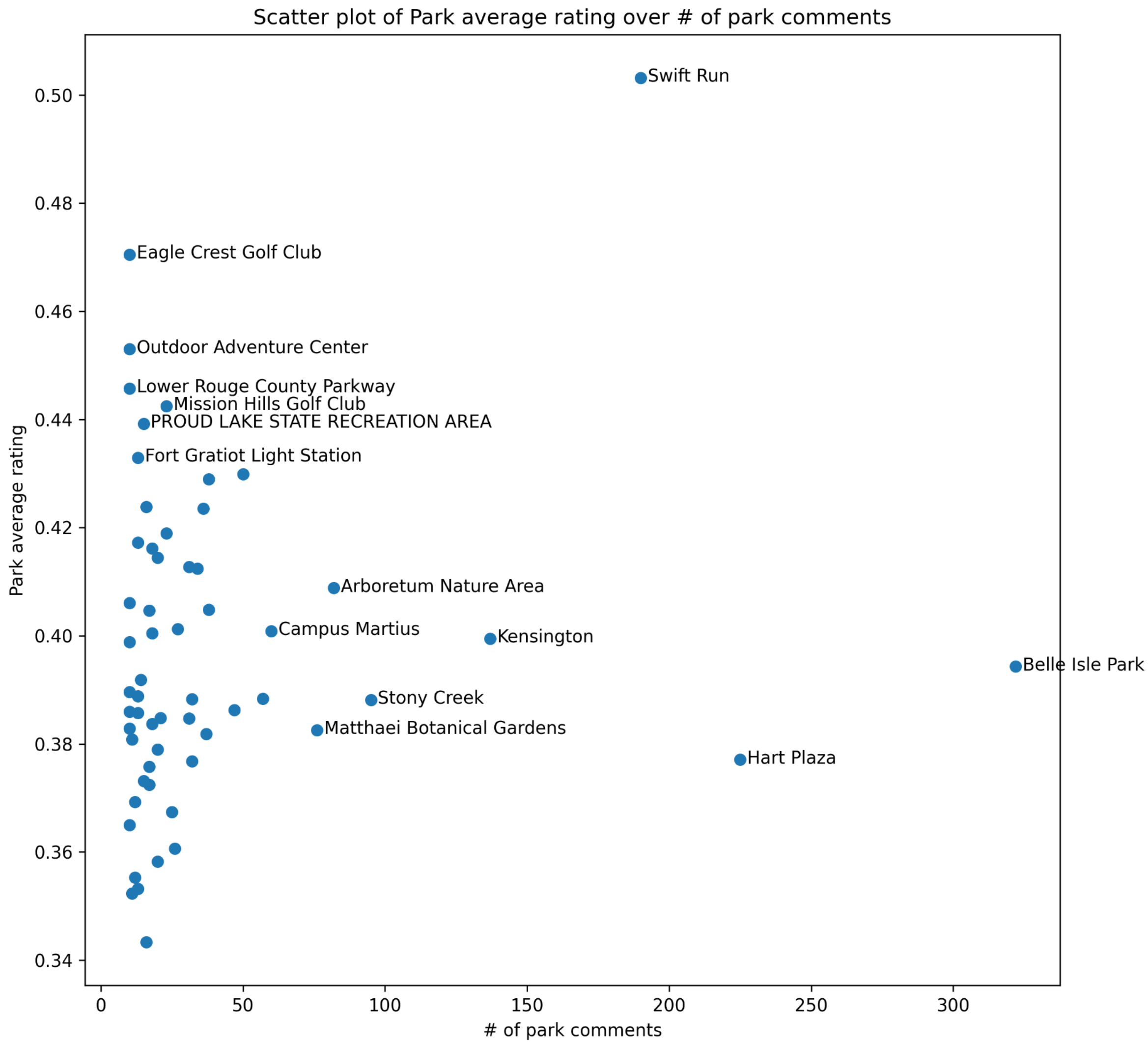

Figure 7. The scatter plot of the park average rating over the number of park comments

4.4 Linear regression for the whole dataset

From the linear regression model for the whole parks dataset, we include the size as a dummy variable and use this dummy variable with other 46 attributes to do linear regression analysis. Under this model for the whole dataset, at the significance level 0.05, the significant attributes are: BMX Area(bicycle motocross or bike motocross area)(P-value = 0.041), Equestrian Activities(P-value = 0.048), Golf Course(P-value = 0), Hunting Trapping Area(P-value = 0), Natural Area(P-value = 0), Monument Historical

Feature(P-value = 0), Walking Biking Trails Paved(P-value = 0.003) and Picnic Shelter(P-value = 0.011). And there are some of them bringing negative effects on user satisfaction, such as BMX Area, Equestrian Activities. See appendix A for more details. Furthermore, the dummy variable size categories of small, medium, and large are all significantly influencing user satisfaction since their P-values are equal to 0, however, the coefficients of those three variables are about the same, which are 0.4055 for the small size, 0.3972 for the medium size and 0.4199 for the large size. Since the influences of those three variables are about the same positive, the dummy variable size of the park will be small on the whole.

From this linear regression summary for the whole parks, we have four conclusions:

1. The significant attributes which have a statistically significant influence on users' satisfaction are BMX Area(bicycle motocross or bike motocross area), Equestrian Activities, Golf Course, Hunting Trapping Area, Natural Area, Monument Historical Feature, Walking Biking Trails Paved and Picnic Shelter.
2. Most attributes have a positive influence on users' satisfaction, however some of them are influencing user satisfaction negatively, like BMX Area and Equestrian Activities.
3. Since the influence of the size is about the same positive according to their coefficients, the dummy variable of size will not influence user satisfaction on the whole.
4. The R-squared value is much lower than the model for the small parks, the medium parks, and large parks, which indicates that the accuracy of the model for the whole parks is lower than that of the other three models.

4.5 Results by park size

The percentage of the three sentiments under small, medium, and large parks are shown in Figure 8, Figure 9, Figure 10.1 and Figure 10.2 below. It is obvious that the amount of positive comments published in larger parks is more than the amount in smaller parks. With the increase of the park area, the positive sentiments of the park increase, while the irrelevant evaluation gradually decreases. To ascertain the influence factors of park attributes, further experiments will be carried out under each category, since under similar conditions, the results are more persuasive and accurate.

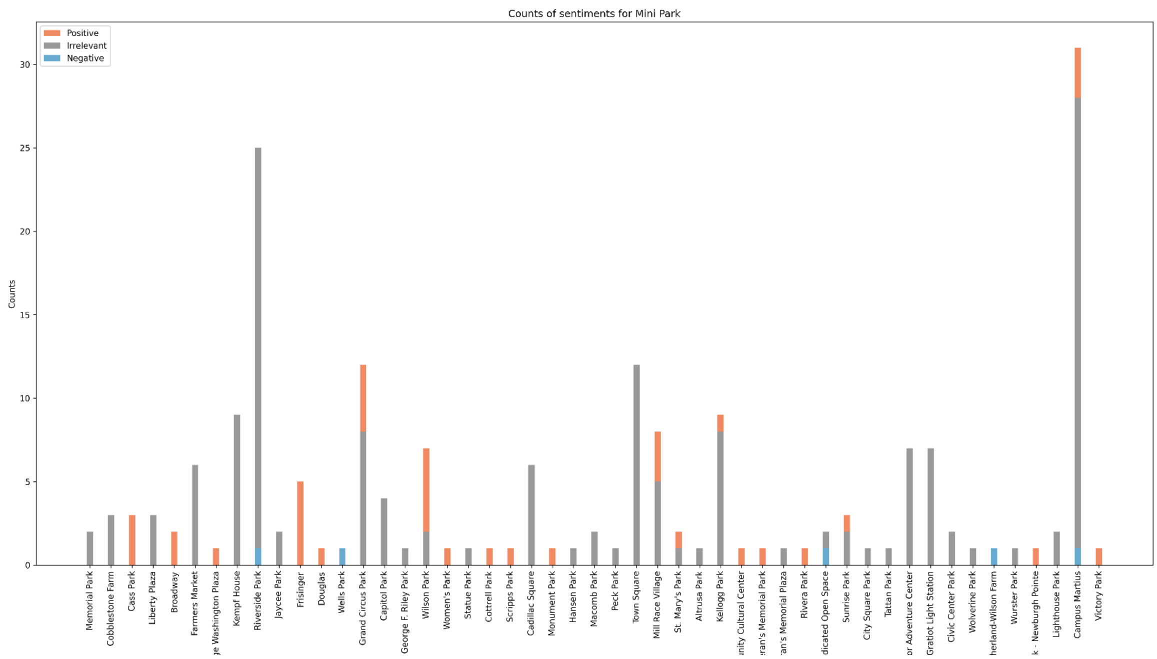

Figure 8. The stacked bar chart showing the percentage of different three sentiments by CLASSECOL analysis under small parks

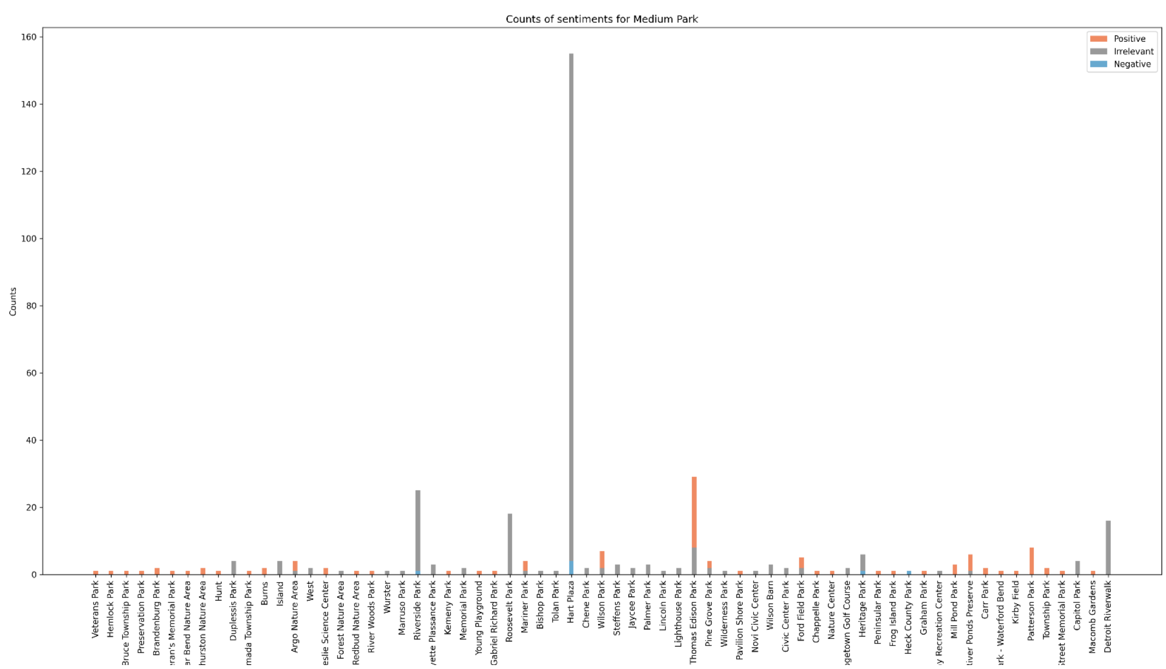

Figure 9. The stacked bar chart showing the percentage of different three sentiments

by CLASSECOL analysis under medium parks

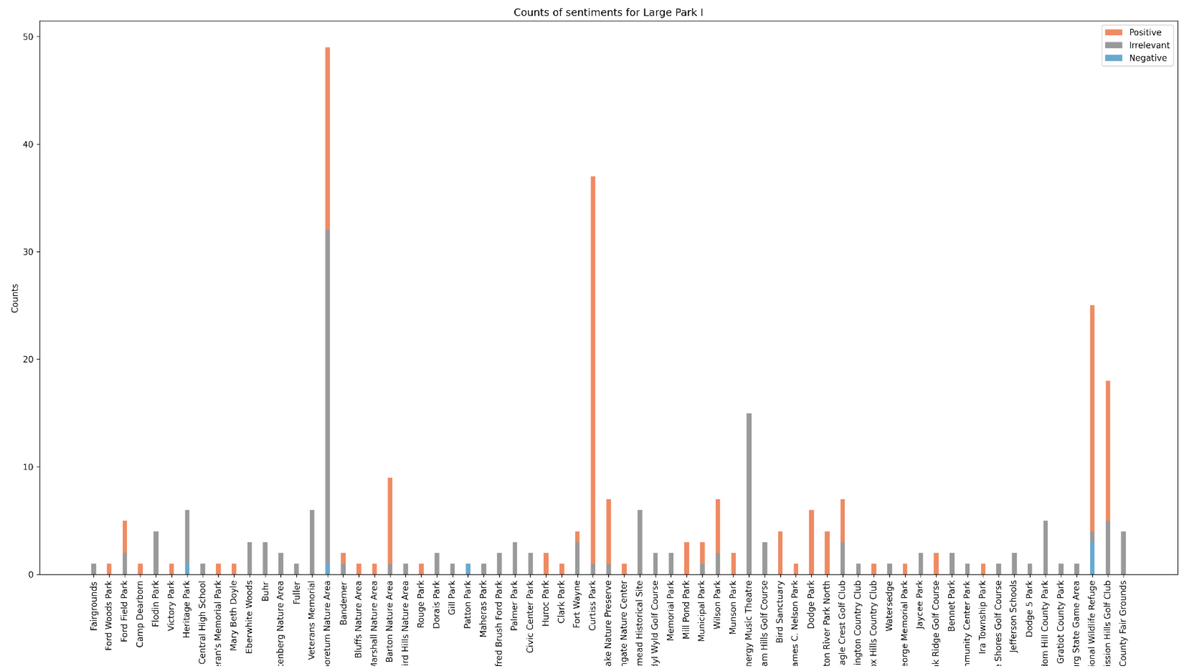

Figure 10.1 The stacked bar chart showing the percentage of different three sentiments by CLASSECOL analysis under large parks I

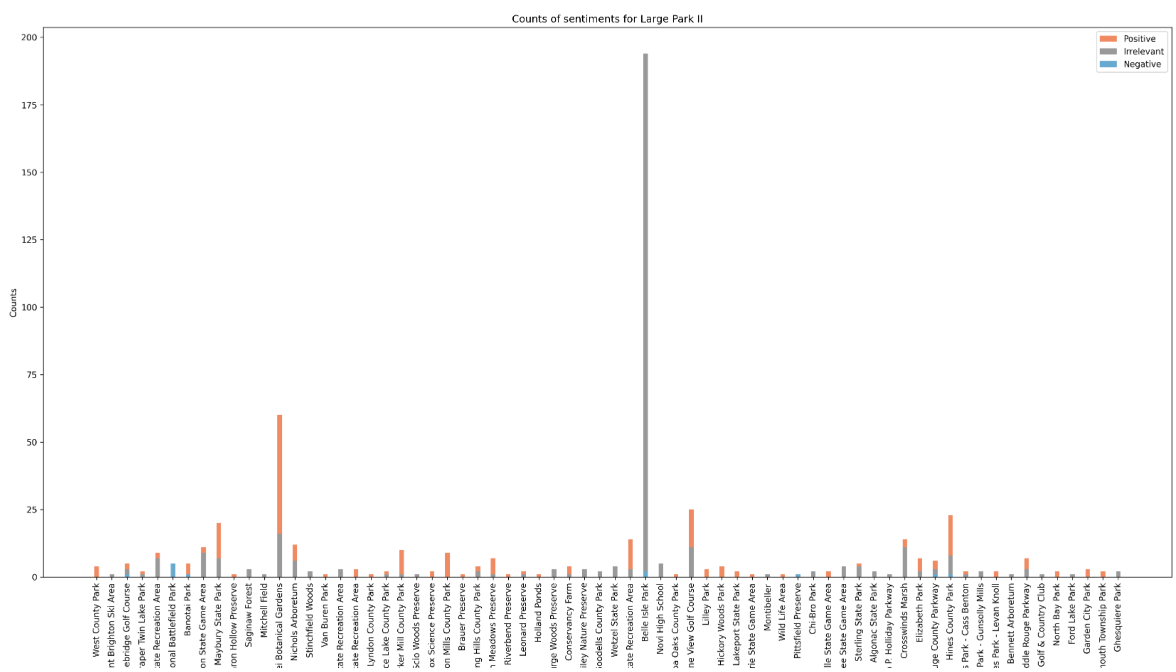

Figure 10.2 The stacked bar chart showing the percentage of different three sentiments by CLASSECOL analysis under large parks II

The result of the linear regression for the small parks

The linear regression model for the small parks is not persuasive, since there are large amounts of P-values equal to Nan, which indicates the result cannot be used to illustrate the significant relationship. The possible reason might be that the number of attributes is higher than the number of predictors, which is the number of parks with effective data in the small park category. The performance of the regression model is that there are no significant attributes that affect the users' satisfaction. See appendix B for more details.

The result of the linear regression for the medium parks

The linear regression model for the medium parks indicates that, at the significance level 0.05, the significant attributes are: Monument Historical Feature(P-value = 0.009) and Amphitheater(P-value = 0.049). The coefficient value of Amphitheater is negative, which indicates that it will have a negative effect on users' satisfaction, while Monument Historical Feature will have a positive effect. See appendix C for more details.

From this linear regression summary for the medium parks, we have two conclusions:

1. The significant attributes which have a statistically significant influence on users satisfaction are Monument Historical Feature and Amphitheater.
2. Amphitheater might have a negative effect on users' satisfaction, according to the linear regression model.

The result of the linear regression for the large parks

The linear regression model for the large parks indicates that, at the significance level 0.05, the significant attributes are: Golf course(P-value = 0), Hunting Trapping Area(P-value = 0), Natural Area(P-value = 0), Monument Historical Feature(P-value = 0.003), Gymnasium(P-value = 0.034) and Snowboarding(P-value = 0.044). All these attributes' coefficients are positive, which indicates that these attributes will have positive effects on the parks' ratings. The confidence interval indicates that the coefficient value has 95% possibilities from the value of 0.025 to the value of 0.975, instead of having an exact effect. See appendix D for more details.

From this linear regression summary for the large parks, we have two conclusions:

1. The significant attributes which have a statistically significant influence on users satisfaction are Golf course, Hunting Trapping Area, Natural Area, Monument Historical Feature, Gymnasium, and Snowboarding.
2. Most attributes have a positive influence on user satisfaction, however, some of them are not significantly influencing satisfaction.

From these three linear regressions, it is obvious that the linear regression model for the large parks contains more information, and there are more significant attributes that affect the users' satisfaction. In the model for the large parks, the differences among each attribute are more pronounced, compared to the other two models. The possible reason is that there are more parks and more effective comments under each park, compared to the other parks under the small and medium parks categories. With more data, the analysis will be more accurate and persuasive.

4.6 Discussion

4.6.1 The influence of park size

After the observation of Figure 3, and the analysis of the size categorization, we found that the size of a park can contribute a lot to the number of park amenities, then also affect the satisfaction of people. These findings are consistent with those of previous studies that the size of the park will influence park-based activity significantly, and a larger attractive park will attract more activities than a smaller one. Furthermore, the larger park tends to have more attributes providing satisfying experiences (Kaczynski et al 2008; Giles-Corti et al 2005).

4.6.2 Overall sentiment analysis

After calculating the number and the percentage of the sentiments, including irrelevant, pro-nature (positive phrasing), pro-nature (negative phrasing), and against-nature, there are 1808 pro-nature (positive phrasing) valid sentiments, 53 pro-nature (negative phrasing) valid sentiments, 1202 irrelevant comments, and 0 against-nature sentiments. This result indicates the majority of the comments from Flickr API is pro-nature (positive phrasing). Throughout the sentiment analysis of each comment and each park, the result indicates that the distribution of the sentiment score of each park after normalization basically shows a normal distribution, and most of comments rate from 0.35 to 0.4.

4.6.3 The analysis of four linear regression models

After comparison of these four linear regression models, the conclusion is that since the influence of the size is about the same positive according to their coefficients, the dummy variable of size will not influence user satisfaction on the whole.

From the analysis of the number of park amenities versus park size, associated with the former research, the results indicate the size of the park will influence park-based activity significantly. However, through the comparison of the three linear regressions under each category and the regression model of the whole dataset with the dummy variable, size, we can see that even though the size influences the result significantly, however, the degree of influence which could be reflected by the value of coefficient, is about the same. This result demonstrates that regardless of the size of the park, it affects user satisfaction in a similar degree but in different attributes. For instance, for a smaller scale, Amphitheater may influence the user satisfaction more significantly while for a larger scale, Natural Area and Golf Course may influence the user satisfaction more significantly.

4.6.4 The discussion of the overall result

Overall, from the comparison of the significant attributes of those different regression models, they have a high degree of similarity with each other in a result. The attributes including Golf course, Hunting Trapping Area, Natural Area, Monument Historical Feature have strong relationships with the users' satisfaction, over the observation towards those four regression models, compared with other attributes. However, there are some perverse results from the regression analysis. For example, the linear regression model for medium parks indicates that the amphitheater will have a strong negative effect on the users' satisfaction. The possible reason behind this might be that certain amphitheater has some negative effects, or there are some certain parks with amphitheaters but they are not satisfied by visitors. So this result cannot be on behalf of all the amphitheater attributes, only if there will be a larger dataset to do analysis.

Since all the regression models indicate the significance of Golf course, Hunting Trapping Area, Natural Area, Monument Historical Feature, the landscape designer can contribute more to these attributes. Also, even if some significant relationships are negative, it still needs to be taken seriously, since it still significantly influences user satisfaction.

5 Conclusions for limitations and future work

There are limitations to this study. First, the valid comments posted on Flickr in the past fifteen years are not enough, and there are only 3063 valid commons that could be sent to the analysis models. Even in those valid commons, there is still a large amount of irrelevant information that has nothing to do with nature. This may lead to low-quality results, low accuracy, and unreliability problems. Second, there is a serious imbalance in the number of data sets among those different parks. This may be due to the following reasons. 1) Different construction times among those parks may lead to a phenomenon that newly constructed parks have less information from the visitors. 2) The parks are different in popularity and geographical location, and more popular parks have more information from the internet. In this research, we have to remove the parks without enough valid comments from our lists. Third, we cannot perfectly control variates in this research. The control variates method is a variance reduction technique used in Monte Carlo methods. It exploits information about the errors in estimates of known quantities to reduce the error of an estimate of an unknown quantity (Lemieux 2017). In our research, the unknown variables are the park attributes shown in table 1, but we cannot control the other variables and make them consistent among all those parks. Those variables include the location of the park, the popularity of the parks, and so on. Another limitation is that we have faced a phenomenon that lots of parks share the same name, and there is no better way to match them by coding since they don't have the same information to match. The only way to match them is using the geometry location downloaded by API and using GIS to relocate them into certain parks. But this method can only be used toward a small amount of data since this method is complicated.

For future research, the thought of processing and analysis could be used in a wider field towards larger datasets. The next step for this research is to find more suitable and larger datasets to do deeper research.

Appendices

Appendix A

The summary of linear regression for the whole parks including size as dummy variables

Appendix B

The summary of linear regression for the small parks

Appendix C

The summary of linear regression for the medium parks

Appendix D

The summary of linear regression for the large parks

Appendix E

The summary of the VIF regression for the park attributes

Appendix A

The summary of linear regression for the whole parks including size as dummy variables

OLS Regression Results						
Dep. Variable:	y	R-squared:	0.297			
Model:	OLS	Adj. R-squared:	0.088			
Method:	Least Squares	F-statistic:	1.422			
Date:	Mon	15 Aug 2022	Prob (F-statistic):	0.0547		
Time:	15:29:52	Log-Likelihood:	299.47			
No. Observations:	211	AIC:	-500.9			
Df Residuals:	162	BIC:	-336.7			
Df Model:	48					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
AMPHITHEATER	-0.0298	0.022	-1.349	0.179	-0.073	0.014
BALLFIELDS	0.0514	0.019	2.703	0.008	0.014	0.089
BASKETBALL_COURTS	-0.005	0.018	-0.277	0.782	-0.04	0.03
BEACH	0.0031	0.032	0.095	0.924	-0.061	0.067
BMX_AREA	0.0172	0.059	0.293	0.77	-0.099	0.133
BOATING	-0.0123	0.036	-0.337	0.737	-0.084	0.06
CAMPING	0.0114	0.046	0.246	0.806	-0.08	0.103
COMMUNITY_RECREATION_CENTER	0.0655	0.032	2.059	0.041	0.003	0.128
CONCESSIONS	0.0439	0.026	1.672	0.097	-0.008	0.096
CROSSCOUNTRY_SKIING	-0.0113	0.029	-0.388	0.699	-0.069	0.046
DISC_GOLF_COURSE	0.0295	0.031	0.966	0.335	-0.031	0.09
DOG_PARK	-0.0012	0.042	-0.03	0.976	-0.083	0.081

EQUESTRIAN_ACTIVITIES	-0.0524	0.042	-1.254	0.212	-0.135	0.03
FARM_GARDEN_ACTIVITIES	2.47E-05	0.031	0.001	0.999	-0.061	0.061
FISHING	-0.0051	0.019	-0.269	0.788	-0.043	0.033
FITNESS_EQUIPMENT	0.038	0.034	1.107	0.27	-0.03	0.106
GEOCACHING	0.0406	0.05	0.811	0.418	-0.058	0.139
GOLF_COURSE	-0.002	0.021	-0.096	0.923	-0.043	0.039
GYMNASIUM	-0.0515	0.048	-1.063	0.289	-0.147	0.044
HIKING_NATURE_TRAILS	0.0064	0.016	0.399	0.69	-0.025	0.038
HOCKEY	0.0111	0.036	0.306	0.76	-0.061	0.083
HUNTING_TRAPPING_AREA	-0.0051	0.028	-0.182	0.856	-0.061	0.051
ICE_SKATING	-0.0634	0.029	-2.189	0.03	-0.121	-0.006
INDOOR_EVENT_FACILITIES	-0.0273	0.021	-1.315	0.19	-0.068	0.014
KAYAKING_CANOEING	0.0167	0.024	0.687	0.493	-0.031	0.065
MONUMENT_HISTORIC_FEATURE	0.0143	0.017	0.825	0.41	-0.02	0.048
MULTIPURPOSE_ATHLETIC_FIELD	0.0457	0.021	2.192	0.03	0.005	0.087
MUSEUM_INTERPRETIVE_CENTER	0.0304	0.033	0.926	0.356	-0.034	0.095
MOUNTAIN_BIKING_TRAILS	0.0119	0.035	0.339	0.735	-0.057	0.081
NATURAL_AREA	-0.0226	0.017	-1.324	0.188	-0.056	0.011
SWIMMING_POOL	0.0575	0.026	2.175	0.031	0.005	0.11
PICKLEBALL_COURT	5.22E-17	3.06E-17	1.704	0.09	-8.29E-18	1.13E-16
PICNIC_SHELTER	-0.033	0.014	-2.298	0.023	-0.061	-0.005

PLAY_AREA	0.008	0.018	0.454	0.65	-0.027	0.043
RESTROOMS	-0.0282	0.014	-1.945	0.053	-0.057	0
SHOOTING_RANGE	-0.0833	0.07	-1.193	0.234	-0.221	0.055
SHUFFLEBOARD	-0.0672	0.052	-1.289	0.199	-0.17	0.036
SKATE_PARK	-0.0502	0.034	-1.476	0.142	-0.117	0.017
SLEDDING_HILL	0.0075	0.025	0.301	0.764	-0.042	0.056
SNOWBOARDING	-0.0381	0.068	-0.558	0.577	-0.173	0.097
SOCCER_FIELD	-0.0589	0.019	-3.087	0.002	-0.097	-0.021
TENNIS_COURT	-0.0249	0.019	-1.288	0.2	-0.063	0.013
TRACK	-0.0843	0.042	-2.006	0.046	-0.167	-0.001
VOLLEYBALL_COURT	0.0667	0.023	2.917	0.004	0.022	0.112
WALKING_BIKING_TRAILS_PAVED	0.0143	0.012	1.158	0.249	-0.01	0.039
WATER_PARK_ SPRAY_PARK	0.0426	0.031	1.381	0.169	-0.018	0.104
WILDLIFE_WATCHING	-0.0057	0.016	-0.354	0.724	-0.038	0.026
Small	0.4055	0.014	29.455	0	0.378	0.433
Medium	0.3972	0.015	26.259	0	0.367	0.427
Large	0.4119	0.014	30.301	0	0.385	0.439
Omnibus:	96.488	Durbin-Watson:	2.2			
Prob(Omnibus):	0	Jarque-Bera (JB):	656.354			
Skew:	1.613	Prob(JB):	2.98E-143			
Kurtosis:	11.016	Cond. No.	1.19E+16			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The smallest eigenvalue is 3.24e-30. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

Appendix B

The summary of linear regression for the small parks

OLS Regression Results						
Dep. Variable:	y	R-squared (uncentered):	0.776			
Model:	OLS	Adj. R-squared (uncentered):	0.515			
Method:	Least Squares	F-statistic:	2.97			
Date:	Thu	11 Aug 2022	Prob (F-statistic):	0.0116		
Time:	14:49:15	Log-Likelihood:	9.9296			
No. Observations:	39	AIC:	22.14			
Df Residuals:	18	BIC:	57.08			
Df Model:	21					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
AMPHITHEATER	0.0896	0.233	0.385	0.705	-0.399	0.578
BALLFIELDS	0.0979	0.284	0.345	0.734	-0.499	0.695
BASKETBALL_COURTS	0.3533	0.318	1.11	0.282	-0.315	1.022
BEACH	-0.0254	0.064	-0.398	0.696	-0.159	0.109
BMX_AREA	-0.0254	0.064	-0.398	0.696	-0.159	0.109
BOATING	-0.0254	0.064	-0.398	0.696	-0.159	0.109
CAMPING	-0.0254	0.064	-0.398	0.696	-0.159	0.109
COMMUNITY_RECREATION_CENTER	0.0619	0.253	0.245	0.809	-0.47	0.594
CONCESSIONS	0.0052	0.265	0.02	0.984	-0.553	0.563
CROSSCOUNTRY_SKIING	-0.0254	0.064	-0.398	0.696	-0.159	0.109
DISC_GOLF_COURSE	-0.0254	0.064	-0.398	0.696	-0.159	0.109
DOG_PARK	0.1554	0.214	0.725	0.478	-0.295	0.606
EQUESTRIAN_ACTIVITIES	-0.0254	0.064	-0.398	0.696	-0.159	0.109
FARM_GARDEN_ACTIVITIES	-0.0189	0.24	-0.079	0.938	-0.524	0.486
FISHING	0.1865	0.304	0.614	0.547	-0.451	0.824

FITNESS_EQUIPMENT	-0.0254	0.064	-0.398	0.696	-0.159	0.109
GEOCACHING	-0.0254	0.064	-0.398	0.696	-0.159	0.109
GOLF_COURSE	-0.0254	0.064	-0.398	0.696	-0.159	0.109
GYMNASIUM	-0.0254	0.064	-0.398	0.696	-0.159	0.109
HIKING_NATURE_TRAILS	-0.0603	0.278	-0.217	0.831	-0.645	0.524
HOCKEY	-0.0083	0.35	-0.024	0.981	-0.743	0.726
HUNTING_TRAPPING_AREA	-0.0254	0.064	-0.398	0.696	-0.159	0.109
ICE_SKATING	-0.11	0.316	-0.348	0.732	-0.774	0.554
INDOOR_EVENT_FACILITIES	0.0148	0.365	0.041	0.968	-0.752	0.782
KAYAKING_CANOEING	-0.0603	0.278	-0.217	0.831	-0.645	0.524
MONUMENT_HISTORIC_FEATURE	0.1062	0.16	0.662	0.516	-0.231	0.443
MULTIPURPOSE_ATHLETIC_FIELD	0.0485	0.35	0.138	0.891	-0.688	0.785
MUSEUM_INTERPRETIVE_CENTER	0.1869	0.239	0.782	0.444	-0.315	0.689
MOUNTAIN_BIKING_TRAILS	0	0	nan	nan	0	0
NATURAL_AREA	0	0	nan	nan	0	0
SWIMMING_POOL	0	0	nan	nan	0	0
PICKLEBALL_COURT	0	0	nan	nan	0	0
PICNIC_SHELTER	0.2465	0.126	1.951	0.067	-0.019	0.512
PLAY_AREA	0.022	0.216	0.102	0.92	-0.433	0.477
RESTROOMS	0.1893	0.162	1.167	0.258	-0.151	0.53
SHOOTING_RANGE	0	0	nan	nan	0	0
SHUFFLEBOARD	0	0	nan	nan	0	0
SKATE_PARK	0	0	nan	nan	0	0
SLEDDING_HILL	0	0	nan	nan	0	0
SNOWBOARDING	0	0	nan	nan	0	0
SOCCER_FIELD	0	0	nan	nan	0	0
TENNIS_COURT	-0.5401	0.482	-1.121	0.277	-1.552	0.472
TRACK	0	0	nan	nan	0	0

VOLLEYBALL_COURT	-0.5733	0.422	-1.36	0.191	-1.459	0.312
WALKING_BIKING_TRAILS_PAVED	0.1477	0.144	1.027	0.318	-0.154	0.45
WATER_PARK_SPRAY_PARK	0.0064	0.262	0.025	0.981	-0.544	0.556
WILDLIFE_WATCHING	0.125	0.254	0.493	0.628	-0.408	0.658
Omnibus:	1.564	Durbin-Watson:	1.117			
Prob(Omnibus):	0.457	Jarque-Bera (JB):	1.394			
Skew:	0.443	Prob(JB):	0.498			
Kurtosis:	2.727	Cond. No.	1.52E+16			
Notes:						
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[3] The input rank is higher than the number of observations.						
[4] The smallest eigenvalue is 1.76e-31. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

Appendix C

The summary of linear regression for the medium parks

OLS Regression Results						
Dep. Variable:	y	R-squared (uncentered):	0.923			
Model:	OLS	Adj. R-squared (uncentered):	0.749			
Method:	Least Squares	F-statistic:	5.315			
Date:	Thu	11 Aug 2022	Prob (F-statistic):	0.00041		
Time:	14:47:28	Log-Likelihood:	40.737			
No. Observations:	52	AIC:	-9.474			
Df Residuals:	16	BIC:	60.77			
Df Model:	36					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
AMPHITHEATER	-0.3566	0.167	-2.132	0.049	-0.711	-0.002
BALLFIELDS	0.3887	0.21	1.85	0.083	-0.057	0.834
BASKETBALL_COURTS	-0.1036	0.168	-0.615	0.547	-0.461	0.254
BEACH	0.601	0.498	1.206	0.245	-0.455	1.657
BMX_AREA	4.84E-15	9.92E-15	0.488	0.632	-1.62E-14	2.59E-14
BOATING	-0.5031	0.309	-1.627	0.123	-1.158	0.152
CAMPING	-1.11E-15	1.46E-15	-0.761	0.458	-4.21E-15	1.98E-15
COMMUNITY_RECREATION_CENTER	0.4648	0.584	0.796	0.438	-0.773	1.702
CONCESSIONS	0.1585	0.6	0.264	0.795	-1.113	1.43
CROSSCOUNTRY_SKIING	-0.3849	0.758	-0.508	0.618	-1.991	1.221

DISC_GOLF_COURSE	-0.6629	1.496	-0.443	0.664	-3.834	2.508
DOG_PARK	0.1201	0.258	0.465	0.648	-0.428	0.668
EQUESTRIAN_ACTIVITIES	-5.60E-16	3.41E-16	-1.641	0.12	-1.28E-15	1.63E-16
FARM_GARDEN_ACTIVITIES	-0.4382	0.334	-1.314	0.208	-1.145	0.269
FISHING	0.2967	0.186	1.599	0.129	-0.097	0.69
FITNESS_EQUIPMENT	0.979	1.445	0.678	0.508	-2.083	4.041
GEOCACHING	-4.98E-15	4.75E-15	-1.049	0.31	-1.50E-14	5.08E-15
GOLF_COURSE	0.3378	0.199	1.695	0.109	-0.085	0.76
GYMNASIUM	-1.3425	1.285	-1.045	0.312	-4.066	1.381
HIKING_NATURE_TRAILS	0.1073	0.175	0.612	0.549	-0.265	0.479
HOCKEY	0.1201	0.258	0.465	0.648	-0.428	0.668
HUNTING_TRAPPING_AREA	-1.78E-15	1.63E-15	-1.091	0.292	-5.22E-15	1.68E-15
ICE_SKATING	-0.2648	0.526	-0.503	0.622	-1.381	0.851
INDOOR_EVENT_FACILITIES	0.3058	0.185	1.656	0.117	-0.086	0.697
KAYAKING_CANOEING	0.2186	0.266	0.821	0.424	-0.346	0.783
MONUMENT_HISTORIC_FEATURE	0.6316	0.211	2.991	0.009	0.184	1.079
MULTIPURPOSE_ATHLETIC_FIELD	0.0593	0.141	0.421	0.68	-0.24	0.358
MUSEUM_INTERPRETIVE_CENTER	-1.761	1.484	-1.187	0.253	-4.907	1.385
MOUNTAIN_BIKING_TRAILS	-4.15E-16	2.16E-16	-1.924	0.072	-8.72E-16	4.23E-17
NATURAL_AREA	0.1797	0.156	1.152	0.266	-0.151	0.51
SWIMMING_POOL	-1.5593	0.883	-1.767	0.096	-3.43	0.312

PICKLEBALL_COURT	-2.22E-15	1.98E-15	-1.124	0.278	-6.41E-15	1.97E-15
PICNIC_SHELTER	-0.1098	0.163	-0.675	0.509	-0.455	0.235
PLAY_AREA	0.1819	0.237	0.767	0.454	-0.321	0.685
RESTROOMS	-0.0496	0.105	-0.472	0.643	-0.272	0.173
SHOOTING_RANGE	-7.50E-17	4.43E-17	-1.695	0.109	-1.69E-16	1.88E-17
SHUFFLEBOARD	-0.9423	0.627	-1.502	0.152	-2.272	0.387
SKATE_PARK	-0.3669	0.363	-1.01	0.327	-1.137	0.403
SLEDDING_HILL	0.4351	0.245	1.779	0.094	-0.083	0.953
SNOWBOARDING	0	0	nan	nan	0	0
SOCCER_FIELD	-0.2287	0.203	-1.126	0.277	-0.659	0.202
TENNIS_COURT	-0.0386	0.174	-0.222	0.827	-0.407	0.33
TRACK	0.4795	0.284	1.687	0.111	-0.123	1.082
VOLLEYBALL_COURT	1.1879	0.722	1.645	0.12	-0.343	2.719
WALKING_BIKING_TRAILS_PAVED	0.1616	0.098	1.657	0.117	-0.045	0.368
WATER_PARK_SPRAY_PARK	-0.6363	0.428	-1.487	0.156	-1.543	0.271
WILDLIFE_WATCHING	0.049	0.152	0.323	0.751	-0.273	0.371
Omnibus:	40.956	Durbin-Watson:	2.403			
Prob(Omnibus):	0	Jarque-Bera (JB):	129.103			
Skew:	2.198	Prob(JB):	9.24E-29			
Kurtosis:	9.345	Cond. No.	1.65E+16			
Notes:						
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[3] The smallest eigenvalue is 4.01e-31. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

Appendix D

The summary of linear regression for the large parks

OLS Regression Results						
Dep. Variable:	y	R-squared (uncentered):	0.871			
Model:	OLS	Adj. R-squared (uncentered):	0.794			
Method:	Least Squares	F-statistic:	11.28			
Date:	Thu	11 Aug 2022	Prob (F-statistic):	1.58E-19		
Time:	14:47:49	Log-Likelihood :	61.89			
No. Observations:	120	AIC:	-33.78			
Df Residuals:	75	BIC:	91.66			
Df Model:	45					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
AMPHITHEATER	0.1486	0.118	1.259	0.212	-0.087	0.384
BALLFIELDS	-0.0797	0.095	-0.836	0.406	-0.27	0.11
BASKETBALL_COURTS	-0.0515	0.085	-0.609	0.544	-0.22	0.117
BEACH	-0.1686	0.117	-1.437	0.155	-0.402	0.065
BMX_AREA	-0.2538	0.228	-1.113	0.269	-0.708	0.201
BOATING	-0.0803	0.148	-0.542	0.59	-0.376	0.215
CAMPING	0.2073	0.158	1.313	0.193	-0.107	0.522
COMMUNITY_RECREATION_CENTER	0.1089	0.198	0.55	0.584	-0.286	0.503
CONCESSIONS	-0.0491	0.128	-0.383	0.703	-0.304	0.206
CROSSCOUNTRY_SKIING	0.0167	0.092	0.183	0.856	-0.166	0.199
DISC_GOLF_COURSE	-0.0591	0.121	-0.489	0.626	-0.3	0.182
DOG_PARK	-1.82E-16	2.43E-16	-0.749	0.456	-6.66E-16	3.02E-16
EQUESTRIAN_ACTIVITIES	-0.1947	0.136	-1.432	0.156	-0.465	0.076

FARM_GARDEN_ACTIVITIES	-0.1856	0.146	-1.273	0.207	-0.476	0.105
FISHING	0.027	0.074	0.363	0.718	-0.121	0.175
FITNESS_EQUIPMENT	0.0649	0.185	0.351	0.726	-0.303	0.433
GEOCACHING	0.1324	0.16	0.828	0.41	-0.186	0.451
GOLF_COURSE	0.3604	0.053	6.815	0	0.255	0.466
GYMNASIUM	0.6683	0.309	2.165	0.034	0.053	1.283
HIKING_NATURE_TRAILS	0.0317	0.058	0.549	0.585	-0.083	0.147
HOCKEY	0.1493	0.156	0.96	0.34	-0.161	0.459
HUNTING_TRAPPING_AREA	0.3381	0.073	4.603	0	0.192	0.484
ICE_SKATING	0.0458	0.125	0.366	0.715	-0.203	0.295
INDOOR_EVENT_FACILITIES	0.0956	0.11	0.869	0.388	-0.124	0.315
KAYAKING_CANOEING	0.0665	0.089	0.75	0.455	-0.11	0.243
MONUMENT_HISTORIC_FEATURE	0.2318	0.076	3.046	0.003	0.08	0.383
MULTIPURPOSE_ATHLETIC_FIELD	0.177	0.095	1.859	0.067	-0.013	0.367
MUSEUM_INTERPRETIVE_CENTER	-0.0385	0.166	-0.231	0.818	-0.37	0.293
MOUNTAIN_BIKING_TRAILS	-0.078	0.121	-0.643	0.522	-0.32	0.164
NATURAL_AREA	0.2887	0.048	6.037	0	0.193	0.384
SWIMMING_POOL	-0.1336	0.112	-1.196	0.235	-0.356	0.089
PICKLEBALL_COURT	-2.81E-17	6.25E-17	-0.449	0.654	-1.53E-16	9.64E-17
PICNIC_SHELTER	0.0882	0.095	0.931	0.355	-0.1	0.277
PLAY_AREA	0.078	0.075	1.036	0.304	-0.072	0.228
RESTROOMS	0.0021	0.087	0.024	0.981	-0.172	0.176
SHOOTING_RANGE	-0.3974	0.206	-1.926	0.058	-0.808	0.014
SHUFFLEBOARD	-0.1343	0.24	-0.56	0.577	-0.612	0.344
SKATE_PARK	-0.0387	0.135	-0.286	0.775	-0.308	0.231
SLEDDING_HILL	0.1554	0.101	1.543	0.127	-0.045	0.356
SNOWBOARDING	0.3739	0.183	2.046	0.044	0.01	0.738
SOCCER_FIELD	-0.033	0.073	-0.453	0.652	-0.178	0.112

TENNIS_COURT	0.1253	0.101	1.24	0.219	-0.076	0.327
TRACK	-0.3982	0.297	-1.342	0.184	-0.989	0.193
VOLLEYBALL_COURT	0.0774	0.095	0.817	0.417	-0.111	0.266
WALKING_BIKING_TRAILS_PAVED	0.1097	0.062	1.777	0.08	-0.013	0.233
WATER_PARK_SPRAY_PARK	-0.0989	0.133	-0.741	0.461	-0.365	0.167
WILDLIFE_WATCHING	0.0413	0.06	0.686	0.495	-0.079	0.161
Omnibus:	1.362	Durbin-Watson:	1.756			
Prob(Omnibus):	0.506	Jarque-Bera (JB):	1.014			
Skew:	0.216	Prob(JB):	0.602			
Kurtosis:	3.127	Cond. No.	3.08E+16			
Notes:						
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[3] The smallest eigenvalue is 2.8e-31. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

Appendix E

The summary of the VIF regression for the park attributes

	Attribute	VIF
0	AMPHITHEATER	1.39608316
1	BALLFIELDS	3.688793527
2	BASKETBALL_COURTS	2.606213309
3	BEACH	3.220807869
4	BMX_AREA	2.222657189
5	BOATING	3.491346096
6	CAMPING	4.760816891
7	COMMUNITY_RECREATION_CENTER	2.037203465
8	CONCESSIONS	2.666596222
9	CROSSCOUNTRY_SKIING	3.018168191
10	DISC_GOLF_COURSE	2.287556133
11	DOG_PARK	1.47776159
12	EQUESTRIAN_ACTIVITIES	2.663710439
13	FARM_GARDEN_ACTIVITIES	1.878561777
14	FISHING	2.787131553
15	FITNESS_EQUIPMENT	2.053311255
16	GEOCACHING	2.22899679
17	GOLF_COURSE	1.273103157
18	GYMNASIUM	4.148000307
19	HIKING_NATURE_TRAILS	3.229891708
20	HOCKEY	2.056867482
21	HUNTING_TRAPPING_AREA	1.812119866
22	ICE_SKATING	2.764401459
23	INDOOR_EVENT_FACILITIES	2.417615918

24	KAYAKING_CANOEING	2.485618011
25	MONUMENT_HISTORIC_FEATURE	1.78343664
26	MULTIPURPOSE_ATHLETIC_FIELD	2.02562444
27	MUSEUM_INTERPRETIVE_CENTER	1.874113879
28	MOUNTAIN_BIKING_TRAILS	2.467131307
29	NATURAL_AREA	2.792000659
30	SWIMMING_POOL	1.998608219
31	PICKLEBALL_COURT	NaN
32	PICNIC_SHELTER	4.357372157
33	PLAY_AREA	5.507152713
34	RESTROOMS	3.284527498
35	SHOOTING_RANGE	2.113518088
36	SHUFFLEBOARD	1.212816025
37	SKATE_PARK	1.286021798
38	SLEDDING_HILL	1.471081767
39	SNOWBOARDING	1
40	SOCCER_FIELD	2.05028377
41	TENNIS_COURT	2.330757855
42	TRACK	3.132586958
43	VOLLEYBALL_COURT	2.078311705
44	WALKING_BIKING_TRAILS_PAVED	2.689711312
45	WATER_PARK_SPRAY_PARK	1.898231291
46	WILDLIFE_WATCHING	3.699606863

Citation

Arthur, L. M. (1977). Predicting scenic beauty of forest environments: some empirical tests. Forest science, 23(2), 151-160.

Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology* (Vol. 1, pp. 492-499). IEEE.

Baek, S. K., & Park, K. H. (2014). Associations between characteristics of green spaces, physical activity and health-Focusing on the case study of Changwon city. Journal of the Korean Institute of Landscape Architecture, 42(3), 1-12.

Benfield, K. (2011). Detroit: The "Shrinking City" That Isn't Actually Shrinking. The Atlantic, 6.

Vogt, C. A., & Marans, R. W. (2004). Natural resources and open space in the residential decision process: a study of recent movers to fringe counties in southeast Michigan. Landscape and Urban Planning, 69(2-3), 255-269.

Celebrate the grand opening of Riverside Park's new skatepark. Retrieved from https://detroit.curbed.com/2019/6/21/18701437/riverside-park-skatepark-grand-opening-event on 06/03/2022

Chen, Y., Parkins, J. R., & Sherren, K. (2018). Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. Landscape and Urban Planning, 170, 283-292.

Choo, K. (2011). Plowing Over: Can Urban Farming Save Detroit and Other Declining Cities-Will the Law Allow It. ABAJ, 97, 43.

Circle Forest. Retrieved from https://www.arbdetroit.org/circle-forest on 06/08/2022

Colasanti, K., Litjens, C., & Hamm, M. (2010). Growing food in the city: The production potential of Detroit's vacant land. The CS Mott Group for Sustainable Food Systems at Michigan State University, East Lansing, MI, 1-13.

Cushing, D. F., & Renata, A. (2015). Themes in landscape architecture publishing: Past trends, future needs. Landscape Journal, 34(1), 15-36.

De Lange, M., & De Waal, M. (2017). Owning the city: New media and citizen engagement in urban design. In Urban land use (pp. 109-130). Apple Academic Press.

Detroit: South Eastern Michigan Council of Governments.

Detroit Future City. Retrieved from https://detroitfuturecity.com/our-programs/ on 05/10/2022

Dony C C, Fekete E. Leveraging Social Media to Track Urban Park Quality for Improved Citizen Health[M]//Geospatial Technologies for Urban Health. Springer, Cham, 2020: 157-177.

Draus, P., Haase, D., Napieralski, J., Roddy, J., & Qureshi, S. (2019). Wounds, ghosts and gardens: Historical trauma and green reparations in Berlin and Detroit. Cities, 93(June), 153-163.

Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. Landscape and Urban Planning, 142, 173–186.
Eisinger, P. (2013). Book Review: Driving Detroit: The Quest for Respect in the Motor City, by George Galster.
Fekete, E. (2017). Foursquare in the City of Fountains: Using Kansas City as a case study for combining demographic and social media data. Thinking big data in geography: New regimes, new research, 165-88.
Galster, G. (2012). Driving Detroit. In Driving Detroit. University of Pennsylvania Press.
Gliozzo, G., Pettorelli, N., & Haklay, M. (2016). Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales, UK. Ecology and Society, 21(3).
Harmon, B. A., Petrasova, A., Petras, V., Mitasova, H., & Meentemeyer, R. (2018). Tangible topographic modeling for landscape architects. International Journal of Architectural Computing, 16(1), 4-21.
Helles, R., & Jensen, K. B. (2013). Introduction to the special issue' Making data-Big data and beyond'. First Monday.
Jim Frost. How to Interpret P-values and Coefficients in Regression Analysis. Retrieved from https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/ on 07/10/2022
John Gallagher.(2019). Social media is arguing about how much vacant land is in Detroit — and the number matters. Retrieved from https://www.freep.com/story/money/business/john-gallagher/2019/10/26/detroit-vacant-land/4056467002/on 06/04/2022
Johnson, T. F., Kent, H., Hill, B. M., Dunn, G., Dommett, L., Penwill, N., ... & González‐Suárez, M. (2021). classecol: Classifiers to understand public opinions of nature. Methods in Ecology and Evolution, 12(7), 1329-1334.
Kaczynski, A. T., Potwarka, L. R., & Saelens, B. E. (2008). Association of park size, distance, and features with physical activity in neighborhood parks. American journal of public health, 98(8), 1451-1456.
Langemeyer, J., Calcagni, F., & Baro, F. (2018). Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. Land use policy, 77, 542-552.
Lemieux, C. (2017). "Control Variates". Wiley StatsRef: Statistics Reference Online: 1--8.
Mertes, J. D., & Hall, J. R. (1996). Park, recreation, open space and greenway guidelines. Urban Land Inst.
Lennon, M., & Scott, M. (2014). Delivering ecosystems services via spatial planning: reviewing the possibilities and implications of a green infrastructure approach. Town Planning Review, 563-587.
Berkooz, C. B. (2011). Green infrastructure storms ahead. Planning, 77(3).
Li, S., & Yang, B. (2022). Social media for landscape planning and design: a review and discussion. Landscape Research, 1-16.

Linear regression Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Linear_regression on 07/10/2022

Liu, X., & Rogers, J. (Eds.). (1999). Land use and land development in southeastern Michigan.

McDonald, J. F. (2014). What happened to and in Detroit?. Urban Studies, 51(16), 3309-3329.

Galster, G. (2012). Driving Detroit. In Driving Detroit. University of Pennsylvania Press.

Giles-Corti, B., Broomhall, M. H., Knuiman, M., Collins, C., Douglas, K., Ng, K., ... & Donovan, R. J. (2005). Increasing walking: how important is distance to, attractiveness, and size of public open space?. American journal of preventive medicine, 28(2), 169-176.

Meerow, S., & Newell, J. P. (2017). Spatial planning for multifunctional green infrastructure: Growing resilience in Detroit. Landscape and Urban Planning, 159, 62-75.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.

Nassauer, J. I., & Raskin, J. (2014). Urban vacancy and land use legacies: A frontier for urban ecological research, design, and planning. Landscape and urban planning, 125, 245-253

NCDB (2010). Neighborhood change databases – Tract data for 1970–80-90-00-10, Version 2.1, Geolytics Inc.

Neill, W. J. (2015). Carry on shrinking?: The bankruptcy of urban policy in Detroit. Planning Practice and Research, 30(1), 1-14.

Nimon, K. F., & Oswald, F. L. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. Organizational Research Methods, 16(4), 650-674.

Nolan, J. (1999). How Prohibition made Detroit a bootlegger's dream town. Detroit News, 15.

Normalization(statistics) Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Normalization_(statistics) on 08/01/2022

Confidence interval Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Confidence_interval on 08/11/2022

NZILA Education Foundation. Landscape Assessment and Sustainable management, date of edit current version: 02.11.10

Paukaeva, A. A., Setoguchi, T., Luchkova, V. I., Watanabe, N., & Sato, H. (2021). Impacts of the temporary urban design on the people's behavior-The case study on the winter city Khabarovsk, Russia. Cities, 117, 103303.

Rishbeth, C. (2016). Landscape experience and migration: Superdiversity and the significance of urban public open space (Doctoral dissertation, University of Sheffield).

Roth, M., Hildebrandt, S., Röhner, S., Tilk, C., Schwarz-von Raumer, H. G., Roser, F., & Borsdorff, M. (2018). Landscape as an area as perceived by people: Empirically-based nationwide modeling of scenic landscape quality in Germany. J. Digit. Landsc. Archit, 3, 129-137.

Ryan BD. Design after decline. InDesign After Decline 2012 May 22. University of Pennsylvania Press.

Ryznar, R. M., & Wagner, T. W. (2001). Using remotely sensed imagery to detect urban change: Viewing Detroit from space. Journal of the American Planning Association, 67(3), 327-336.
Sangirova, U., Isakuzieva, R., Zakirova, G. S., Sattorov, O. B., Tulaboev, A., Ismailova, Z. K., ... & Mahsudov, P. M(2019). ARCHITECTURE. LANDSCAPE ARCHITECTURE.
Schukoske, J. E. (1999). Community development through gardening: State and local policies transforming urban open space. NYUJ Legis. & Pub. Pol'y, 3, 351.
Seber, G. A., & Lee, A. J. (2012). Linear regression analysis. John Wiley & Sons.
Shirtcliff, B. (2015). Big Data in the Big Easy: How social networks can improve the place for young people in cities. Landscape Journal, 34(2), 161–176.
Sonter, L. J., Watson, K. B., Wood, S. A., & Ricketts, T. H. (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. PLoS one, 11(9), e0162372.
Van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016). Continental-scale quantification of landscape values using social media data.
Variance Inflation Factor (VIF) Investopedia. Retrieved from https://www.investopedia.com/terms/v/variance-inflation-factor.asp#:~:text=Variance%20inflation%20factor%20(VIF)%20is,only%20that%20 single%20 independent%20variable. on 08/13/2022
Variance Inflation Factor (VIF) Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Variance_inflation_factor on 08/13/2022
Walden-Schreiner, C., Leung, Y. F., & Tateosian, L. (2018). Digital footprints: Incorporating crowdsourced geographic information for protected area management. Applied Geography, 90, 44-54.
Weisberg, S. (2005). Applied linear regression (Vol. 528). John Wiley & Sons.
Wilkie, S., Thompson, E., Cranner, P., & Ginty, K. (2020). Attention restoration theory as a framework for analysis of Tweets about urban green space: a case study. Landscape Research, 45(6), 777-788. Proceedings of the National Academy of Sciences, 113(46), 12974-12979.
Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65.
Xie, Y., Gong, H., Lan, H., & Zeng, S. (2018). Examining shrinking city of Detroit in the context of socio-spatial inequalities. Landscape and Urban Planning, 177, 350-361.
Yang Song & Bo Zhang (2020) Using social media data in understanding site-scale landscape architecture design: taking Seattle Freeway Park as an example, Landscape Research, 45:5, 627-648, DOI: 10.1080/01426397.2020.1736994
Zeller, F. (2017). Analyzing social media data and other data sources: A methodological overview. The Sage handbook of social media research methods, 386-404.