

# Predicting Park Access Rankings Across U.S. Cities

DSCI 310 Group 08

Shivani Aggarwal, Luna Gulec, Jingyuan Liu, Sarena Ulman

## Table of contents

Summary . . . . .	1
Introduction . . . . .	2
Background Information . . . . .	2
Research Question . . . . .	3
About the Dataset . . . . .	4
Data Dictionary . . . . .	5
Methods & Results . . . . .	6
Data Wrangling and Cleaning . . . . .	6
Exploratory Data Analysis & Visualization . . . . .	7
Regression Analysis . . . . .	12
Results & Visualizations . . . . .	12
Discussion & Summary of Our Findings . . . . .	14
Interpretation and Expectations . . . . .	16
Impact of Our Findings and Practical Implications . . . . .	16
Limitations and Improvements . . . . .	17
Future Work and Further Questions . . . . .	17
References . . . . .	18

## Summary

Urban parks play a crucial role in promoting health, well-being, and social cohesion within cities. Recognizing the importance of equitable access to high-quality parks, this project aims to analyze the characteristics of urban park systems in the United States, identify the factors most strongly associated with top-performing park systems, and predict a city’s park access ranking, using a dataset from the **Trust for Public Land’s** ParkScore index up to 2021.

Our exploratory analysis revealed that while some metrics, like park size and land percentage, have consistent distributions across cities, others, like the percentage of the population living near a park, vary significantly. We developed a Ridge Regression model to predict a city's rank. After identifying and removing the `rank_last_time` feature to prevent data leakage, our final model achieved a strong performance with an  $R^2$  value of 80.3% on the test set.

The results indicate that while temporal and geographical factors have the highest impact on rankings, specific actionable features like spending per resident and park access are the primary park-based predictors of a higher and better ranks. These findings suggest that urban planners can improve their city's standing by prioritizing investment and optimizing park locations to increase the number of residents within a 10-minute walk.

## Introduction

### Background Information

In today's world, cities are expanding rapidly, while urban green space is declining. According to the Husqvarna Group Urban Green Space Insights (HUGSI) Report (Husqvarna Group 2025), which uses AI and satellite data to measure green spaces in cities worldwide, between 2023 and 2024, the 516 cities analyzed recorded a loss of 95 million m<sup>2</sup> of green area – a loss equivalent to the size of Paris – primarily due to human activities such as construction and urban expansion. Among the analyzed cities, 73% show a negative development. North America stands at a 37% average urban green space, among the middle of all regions. In contrast, the Nordic region stands out as a green hub, hosting some of the greenest cities globally, with an impressive 49% share of urban green space and an average urban tree canopy cover of 35% (Husqvarna Group 2025).

Researchers have found that people with greater access to greenspaces are 44% less likely to be diagnosed with an anxiety disorder, and that accessible greenspaces are known to improve cardiovascular health, reduce obesity rates, prevent chronic diseases, reduce stress, and improve mood. Reports from the National Recreation and Park Association also suggest that physical activity in greenspaces has stronger mental health benefits than physical activity in non-greenspaces (“The Fundamentals of Urban Greenspaces” 2025). These facts highlight how essential greenspaces are for both physical and mental well-being, but beyond these benefits are long-term sustainable benefits to the community. Parks also serve as communal hubs where people from all walks of life can meet, socialize, and build relationships. With increased social cohesion comes more engaged and committed communities more likely to take responsibility for maintaining these spaces, making them safer, more welcoming, and better cared for, ensuring the space remains a valuable resource for future generations.

In addition, modern urban lifestyle is associated with chronic stress, insufficient physical activity and exposure to anthropogenic environmental hazards. Urban green spaces, such as parks, playgrounds, and residential greenery, can promote mental and physical health, and reduce

morbidity and mortality in urban residents by providing psychological relaxation and stress alleviation, stimulating social cohesion, supporting physical activity, and reducing exposure to air pollutants, noise and excessive heat (World Health Organization Regional Office for Europe 2016).

However, it's also important to note that not all communities have equal access, and that it's crucial to engage with the community in a way that is both meaningful and equitable. If the park doesn't reflect the cultural identity of a community, it can become a space that feels alienating, rather than inclusive, and it can actually cause more harm than good to individuals of the surrounding community. Capital Trees' "The Fundamentals of Urban Greenspaces" (2025) references an article from the American Planning Association, which asks three fundamental questions when designing parks and public spaces:

- *Who is Helped:* engage communities to help address systemic inequities, provide equal access to recreational resources, serve as spaces for cultural expression, healing, and collective action, and prioritize the needs of all groups, so that parks can become tools for social justice, equity, and well-being.
- *Who is Harmed:* how a lack of inclusive engagement perpetuates existing inequalities or historical injustices by neglecting the voices of those who are already marginalized or underrepresented financially, physically, culturally, and psychologically.
- *Who is Missing:* actively seek typically underrepresented voices for whom parks might be designed in ways that do not consider safety concerns, such as vulnerable populations, including people with disabilities, children, and seniors.

Access to greenspaces shouldn't be a privilege; it should be a right. Because not all communities have equal access, we believe that there is a need for scores, indexes, and metrics like the ParkScore index that allow parks to be ranked on many different metrics, such as such as park amenities, to help improve community experiences. Designing parks and green spaces that are accessible, equitable, meet the environmental needs of the area, and consider community and cultural needs, can bring mental and physical benefits to individuals as well as help reduce health disparities. When everyone, especially vulnerable populations, has access to high-quality, inclusive greenspaces, the community as a whole thrives. Investing in community engagement through parks and greenspaces is about building healthier, happier, and more equitable communities for the future, and we intend to explore this in our project.

## **Research Question**

The objective of this project is to investigate whether characteristics of a city's park system can be used to predict its park access ranking and determine the ranks of cities among the top-performing park systems in the United States. Specifically, we ask:

*Can features describing park availability, amenities, and investment be used to predict a city's ranking in the Trust for Public Land's ParkScore index?*

To answer this question, we developed a regression model that predicts a city’s park system rank based on several explanatory variables, including park size, percentage of city land dedicated to parks, access to parks, spending per resident, and availability of amenities such as playgrounds, dog parks, and recreation facilities. By examining how these features relate to park rankings, we aim to identify which characteristics are most strongly associated with higher-performing park systems.

Understanding these relationships and identifying the features that distinguish top-ranked cities may help urban planners and policymakers better understand the factors that contribute to successful park systems and guide decisions about future investments in public green spaces.

**Proposed Research Question:** *Can we predict a city’s park access ranking based on park characteristics and population features such as total park acreage, number of parks, population size, and population density?*

## About the Dataset

The dataset used in this analysis (jonthegeek 2021) contains information about park systems across major cities in the United States up to 2021 and is sourced from the **ParkScore index**, developed by the **Trust for Public Land**, whose mission is to ensure that every American has access to a good quality park within a 10-minute walking distance. The ParkScore index evaluates park systems in the largest U.S. cities based on several metrics, including park access, acreage, investment, and amenities. The data dictionary is in Table 1. The first 5 rows of the raw data is presented in Table 2.

For this project, we use a version of the dataset provided through the **TidyTuesday** repository (Data Science Learning Community 2024), which compiles publicly available datasets for data analysis and visualization practice, and can be accessed [here](#).

Each observation in the dataset represents a U.S. city in a given year and contains variables describing characteristics of the city’s park system. These include measures such as median park size, percentage of city area dedicated to parks, percentage of residents living near a park, park spending per resident, and availability of park amenities such as playgrounds, dog parks, basketball courts, and recreation or senior centers (Poon and Patino 2021).

The main goal of collecting the data was to assess if the residents of major US cities had sufficient access to parks, and to measure and rank the quality of each city’s parks. The dataset is publicly hosted on GitHub, ensuring maximum reproducibility. You can read more about Trust for Public Land’s initiative [here](#) (Chapman et al. 2021).

## Data Dictionary

Note that “points” are essentially their yearly normalized values (higher points = better).

Table 1: Parks Ranking Data Dictionary

Table 1

variable	class	description
year	double	Year of measurement
rank	double	Yearly rank
city	character	City Name
med_park_size_data	double	Median park size acres
med_park_size_points	double	Median park size in points
park_pct_city_data	character	Parkland as percentage of city area
park_pct_city_points	double	Parkland as % of city area points
pct_near_park_data	character	Percent of residents within a 10 minute walk to park
pct_near_park_points	double	Percent of residents within a 10 minute walk to park points
spend_per_resident_data	character	Spending per resident in USD
spend_per_resident_points	double	Spending per resident in points
basketball_data	double	Basketball hoops per 10,000 residents
basketball_points	double	Basketball hoops per 10,000 residents points
dogpark_data	double	Dog parks per 100,000 residents
dogpark_points	double	Dog parks per 100,000 residents points
playground_data	double	Playgrounds per 10,000 residents
playground_points	double	Playgrounds per 10,000 residents points
rec_sr_data	double	Recreation and senior centers per 20,000 residents
rec_sr_points	double	Recreation and senior centers per 20,000 residents points
restroom_data	double	Restrooms per 10,000 residents
restroom_points	double	Restrooms per 10,000 residents points
splashground_data	double	Splashgrounds and splashpads per 100,000 residents
splashground_points	double	Splashgrounds and splashpads per 100,000 residents points
amenities_points	double	Amenities points total (ie play areas)
total_points	double	Total points (varies in denominator per/year)
total_pct	double	Total points as a percentage
city_dup	character	City duplicated name
park_benches	double	Number of park benches

This can also be found at the data website [here](#).

## Methods & Results

### Data Wrangling and Cleaning

We load data from the original source on the web. Here, we present the first 5 rows of the raw data in Table 2.

*Note: Table 2, Table 3, and Table 4 only display a subset of key features to make sure that the tables do not run off the page in the report for clarity. The excluded columns are consistent with the characteristics of the displayed data.*

Table 2: Raw Data Sample

Table 2

year	rank	city	med_park_size_data	med_park_size_points	park_pct_city_data
2020	1	Minneapolis	5.700000	26.000000	15%
2020	2	Washington, D.C.	1.400000	5.000000	24%
2020	3	St. Paul	3.200000	14.000000	15%
2020	4	Arlington, Virginia	2.400000	10.000000	11%
2020	5	Cincinnati	4.400000	20.000000	14%

The original raw dataset has 713 rows and 28 columns. Then, we perform data wrangling to clean the data from its original format to the format necessary for the purpose of this analysis.

Since this dataset records the yearly rankings of park access across U.S. cities, it is inherently time-series data. However, advanced time-series methods are beyond the scope of this course. Instead, we extract each city's rank from the last time, denoted as `rank_last_time`, to capture basic year-to-year trends. For the first observation year of a city, we decided to impute its `rank_last_time` with its ranking in the current year (i.e., `rank`). We also then remove data from year 2012, 2013, and 2014 because 21.0% of values are missing in those rows, which would not help with building a predictive model.

We noticed that more than 13.9% of the data in columns `restroom_data`, `restroom_points`, `splashground_data`, `splashground_points`, `total_points`, `total_pct`, `city_dup`, and `park_benches` is missing, with some columns approaching 59.2%. Therefore, we decided to drop these columns as advanced data imputation methods are also beyond the scope of this course.

There are a total of 102 cities in this dataset, but each city only has fewer than 10 observations. This makes `city` not an ideal categorical variable for regression due to the potential overfitting issue. Therefore, we perform feature aggregation and add a new column called `state` so that each category has more valid observations. We manually map each city into a state.

Finally, we convert columns `year`, `city`, and `state` to type `category`, and columns `rank` and `rank_last_time` to type `int`. In addition, we know that variables ending with “points” are essentially their yearly normalized values (higher points = better), so we decided to remove raw numerical variables (i.e., variables ending with “\_data”) and just keep the normalized ones with potential categorical variables in the model. In particular, our response variable is **rank!**

Here, we present the first 5 rows of the processed data in Table 3.

Table 3: Processed Data Sample

Table 3

year	rank	city	med_park_size_points	park_pct_city_points	pct_near_park_points
2015	13	Albuquerque	8.000000	20.000000	32.000000
2016	20	Albuquerque	8.000000	20.000000	31.000000
2017	17	Albuquerque	8.000000	20.000000	31.000000
2018	40	Albuquerque	8.000000	20.000000	30.000000
2019	34	Albuquerque	20.000000	50.000000	82.500000

Now we are ready to proceed! Our analysis is based on the processed data as shown in Table 3.

## Exploratory Data Analysis & Visualization

Before conducting exploratory analysis, we split the data into training and test sets, so as to avoid touching the test data and accidentally learn its characteristics during analysis. We save 20.0% of the data as the test set and set it aside.

Here is the distribution of ranks (our target variable) in the training set:

- Minimum rank: 1
- Maximum rank: 98
- Average rank: 47.0
- Median rank: 46.0
- Most common rank: 24

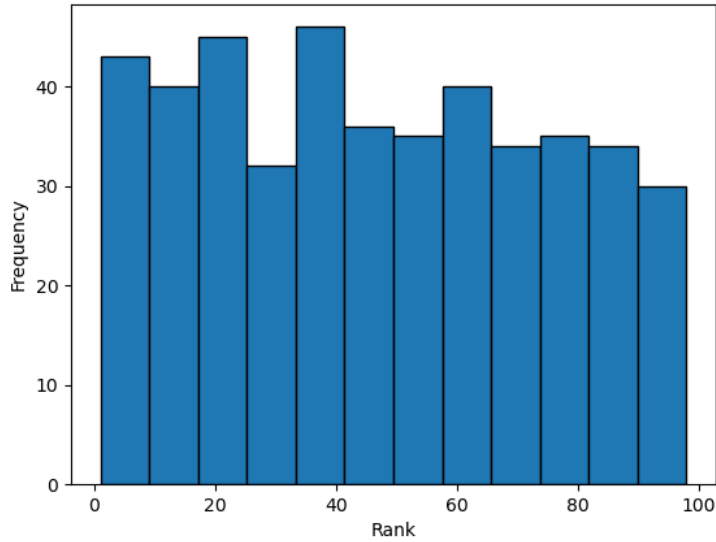


Figure 1: Ordinal Rank vs Value Frequency in the Training Set

From Figure 1, we can see that the ranks are somewhat evenly distributed, with a slight majority of parks receiving a lower rank.

Here, we have the summary of the features in the training data in Table 4.

Table 4: Summary of the Training Features

	year	med_park_size_points	park_pct_city_points	pct_near_park_points
count	450.000000	450.000000	450.000000	450.000000
mean	2017.566667	15.497778	15.811111	35.292222
std	1.664847	11.909364	12.548032	24.822211
min	2015.000000	1.000000	1.000000	1.000000
25%	2016.000000	7.000000	7.000000	18.000000
50%	2018.000000	13.000000	12.500000	29.000000
75%	2019.000000	20.000000	20.000000	43.750000
max	2020.000000	50.000000	50.000000	100.000000

The first thing to note is that every column has 450 values, meaning all of the NA values were successfully removed. Most of the point scales are out of 50 or 100, and the average number of points hovers between 15.5-44.7 points in all categories.

Since we expect previous ranks to be quite important in predicting the current rank, we look at the distribution of the feature `rank_last_time`.

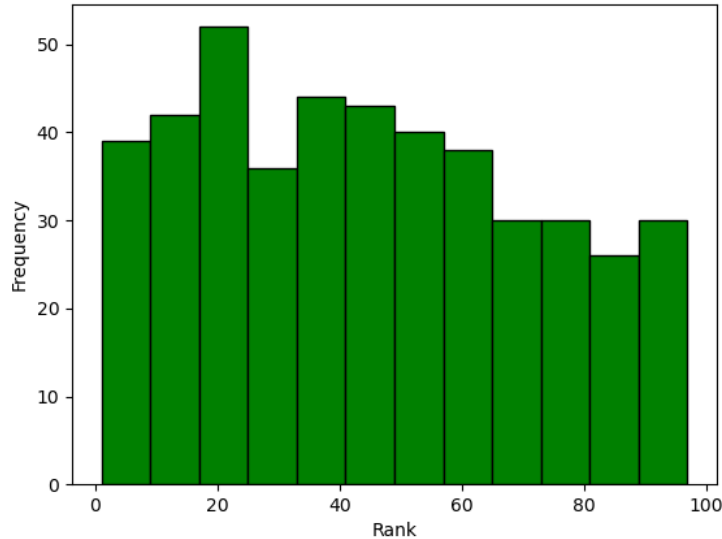


Figure 2: Previous Rank Value vs Frequency it Appears in the Training Set

In Figure 2, we see a similar distribution as the target variable, with a decent majority of the ranks being on the lower half of the spectrum. Due to its similarity to the target variable, it is important to check the correlation between `rank_last_time` and `rank`.

We can see it in the scatterplot in Figure 3 that the two variables are almost perfectly correlated with a correlation of 97.0%. This either means that `rank_last_time` is a strong predictor for `rank` or that it presents data leakage by acting as a proxy for `rank`. To prevent the potential data leakage problem, we remove `rank_last_time` as a feature before setting up the predictive model.

Lastly, let's inspect the distributions of the point-based features in a more visually interpretable boxplot. We are also able to check for outliers this way.

Figure 4 shows that the medians are relatively similar, except for `pct_near_park_points`, and that most columns have similar standard deviations. Something to note is every feature, with the exception for `amenities_points`, has several outliers. Additionally, the features `med_park_size_points` and `park_pct_city_points` have nearly identical distributions.

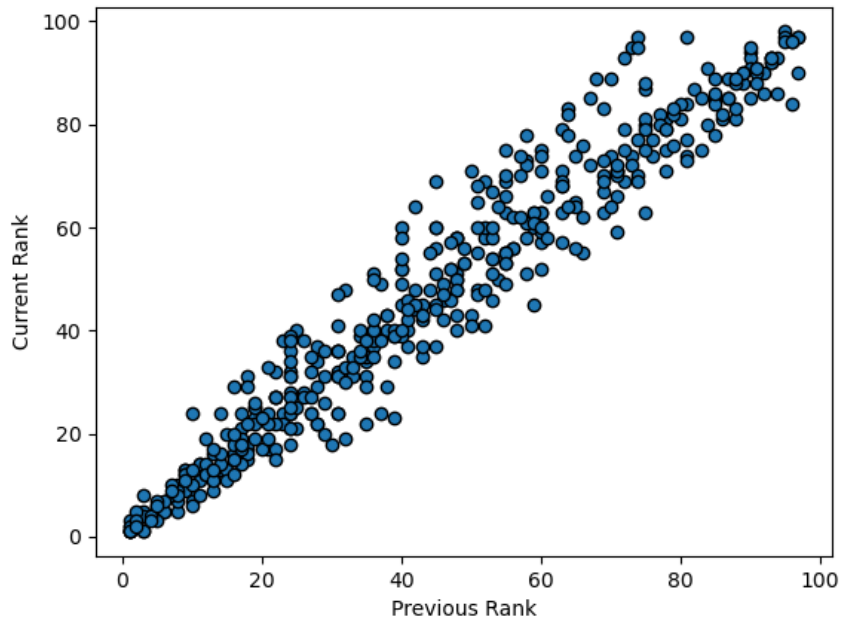


Figure 3: Previous Rank vs Current Rank

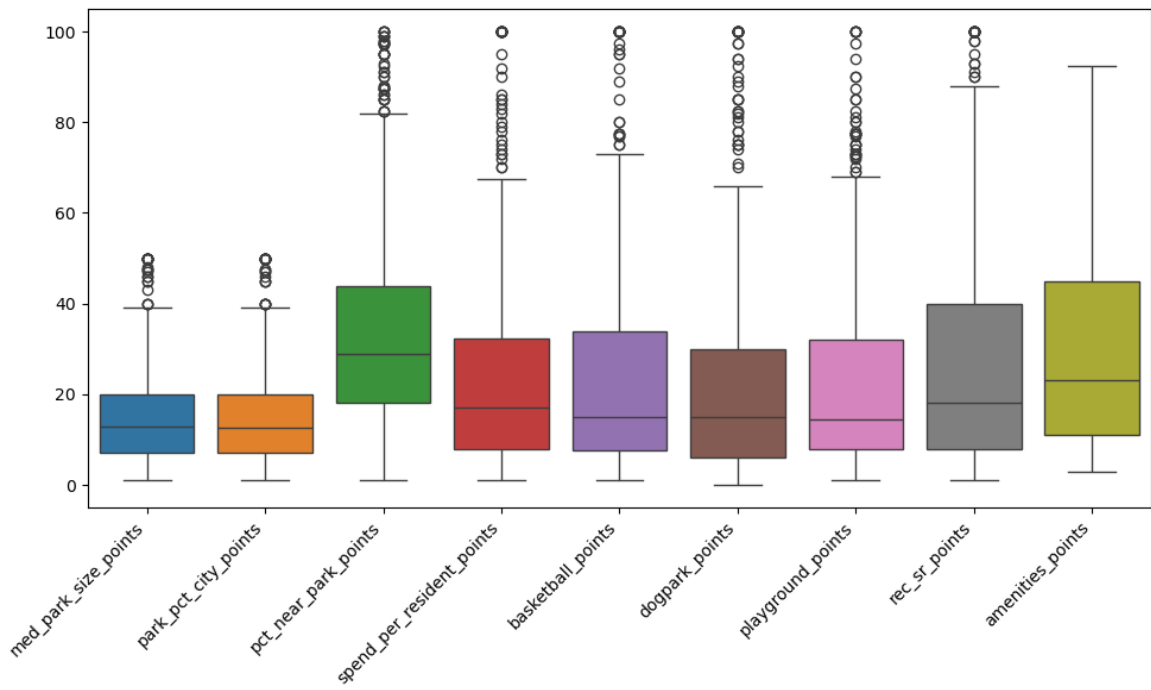


Figure 4: Boxplots of the Point-based Features

## Regression Analysis

Since our target variable **rank** is a discrete numerical variable ranging from 1-98, it would be best to use **ridge regression** to build a predictive model. We drop **city** as a feature as it is represented by **state**. We drop **rank\_last\_time** as it may act as a proxy to the target and leak information into the predictive model. Since all the remaining numerical features are standardized point-based scores, we use them as is.

Before implementing the model on the data, we optimize the regularization hyperparameter *alpha* through randomized grid search. This tests the model performance using different values of *alpha* and picks the one that results in the best performance. The grid search results in an *alpha* of 0.3994 that produces a mean cross-validated  $R^2$  of 74.6%.

Now, we fit the model on the training data using  $alpha = 0.3994$  and evaluate its performance on the test set.

## Results & Visualizations

The model performed exceptionally well on the testing data. It explains 80.3% of the variance in park rankings on the test set. The model has a **RMSE** (Root Mean-Squared Error) score of 12.46. This means that on average, predictions are off by about 12.0 ranks. This is reasonable given the ranks range from 1 to 98.

Since our target is rank, an ordinal variable, we also calculate the **Spearman's Rank Correlation**. We use this measure to evaluate the correlation between the predicted and actual ranks. The Spearman correlation on the test data is 0.913%. This means that the predicted rankings are very strongly correlated with the actual rankings in terms of order. Even if the exact rank is off by ~12.0 positions, the model is still very good at getting the relative ordering of cities correct (i.e. high-ranked cities are predicted high, low-ranked cities predicted low). The Spearman p-value is essentially 0.0, meaning the correlation is highly statistically significant.

Let's take a look at the model coefficients and understand the relationships between the predictor variables and the target.

Table 5: Ridge Model Coefficients (Top 15)

Table 5

feature	coefficient
year_2020	44.872492
year_2019	43.312074
year_2016	-25.556359
year_2017	-24.981293

feature	coefficient
state_IN	24.878675
year_2015	-21.892940
state_MI	16.754294
year_2018	-15.753973
state_IL	-14.618183
state_KS	13.918173
state_WI	-13.119980
state_TN	12.653788
state_NC	12.369204
state_NE	-11.883857
state_MA	-11.851997

Since a lower rank is better, a positive coefficient predicts a worse rank whereas as a negative coefficient predicts a better rank.

- From Table 5, we can see that cities in 2020 and 2019 tend to receive much worse predicted ranks compared to the baseline year, 2021, whereas from 2015-2018, cities tend to rank better. These shifts in ranking across years possibly reflects changes in the ranking methodology or shifts in overall park system scores across years rather than true differences in park quality.
- Cities in Indiana and Michigan tend to rank worse versus cities in Illinois and Wisconsin tend to rank better than the baseline.

Since the largest coefficients correspond primarily to year and state indicators, we can say that temporal and geographic variation plays a significant role in predicting rank

Let's look at coefficients after removing features `state` and `year`.

Table 6: Ridge Model Coefficients after Removing State and Year

Table 6

feature	coefficient
park_pct_city_points	-0.742548
med_park_size_points	-0.479936
spend_per_resident_points	-0.449792
pct_near_park_points	-0.439352
amenities_points	-0.413770
playground_points	0.080544
dogpark_points	-0.016685
rec_sr_points	-0.014377

feature	coefficient
basketball_points	0.013002

When examining only the point-based park features in Table 6, several variables showed meaningful relationships with park rankings. The features that predict a higher rank with the largest magnitude are `park_pct_city_points` (percentage of city land dedicated to parks), `med_park_size_points` (median park size), `spend_per_resident_points` (spending per resident), `pct_near_park_points` (percentage of residents living near a park), and `amenities_points` (overall amenity score).

These findings align with the idea that park accessibility, funding, land allocation, and park size are the key components of high-quality park systems. Amenity-specific features such as `playground_points`, `dogpark_points`, `rec_sr_points`, and `basketball_points` had near-zero coefficients, suggesting that individual amenities contribute less to overall ranking compared to broader measures of park access and investment.

We can better visualize the model performance using the scatterplot of Actual vs Predicted Ranks.

## Discussion & Summary of Our Findings

Our ridge regression model was able to predict city park system rankings with relatively strong performance, fitting quite well. The model achieved an RMSE score of approximately 12.0 and a Spearman’s rank correlation of 0.913%, meaning that the predictions are roughly off by only ~12.0 ranks while still capturing the relative order of cities very accurately. Overall, the model is explaining 80.3% of the variation in rankings on the test set. Looking at Figure 5, we can see that the predicted ranks follow the actual ranks closely, with the regression line passing through the bulk of the data points, indicating that the model captures the overall trend in rankings well. Therefore, this suggests that measurable park attributes, such as park size, accessibility, and amenities, as well as year of ranking and the state in which the park is located do contribute meaningfully to how cities are ranked in the ParkScore system.

Among the predictor variables as listed in Table 5 and Table 6, it is interesting to see that the `year` and `state` features had such a large impact on the prediction, with the highest magnitude coefficients. Intuitively, the rankings should not depend that much simply on the year in which the measurements were taken. However, this is somewhat to be expected, and it suggests that park rankings vary significantly across time and location.

Overall, the model captured meaningful relationships between park characteristics and ranking outcomes, even without relying on previous rankings (`rank_last_time`).

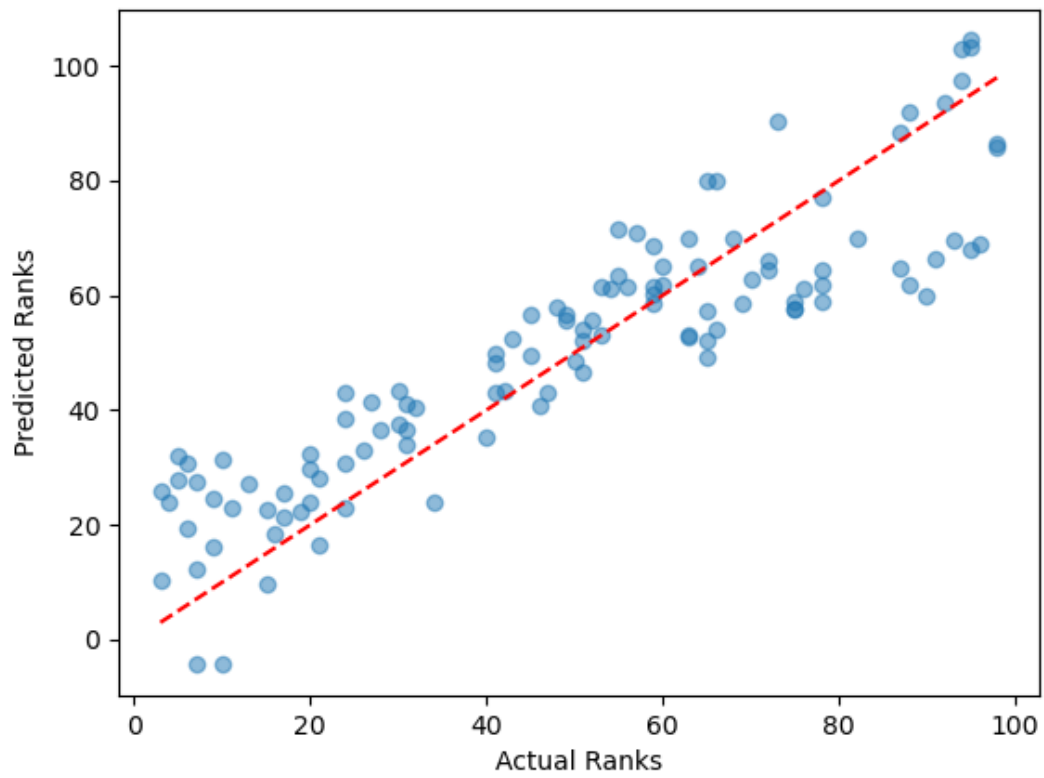


Figure 5: Actual vs Predicted Ranks

## Interpretation and Expectations

Many of the observed relationships are consistent with our expectations. For example, features related to park accessibility (`pct_near_park_points`) and park coverage (`park_pct_city_points`) were associated with improved rankings. This is intuitive because cities where more residents live near parks and where parks occupy a greater portion of urban land are likely to provide better recreational opportunities and public access to green space.

Similarly, spending per resident on parks (`spend_per_resident_points`) also showed a meaningful relationship with improved rankings, which makes sense because higher funding may enable cities to maintain parks more effectively, build new facilities, and improve park amenities, all of which can contribute to stronger performance in the ParkScore evaluation.

One unexpected result was the strong influence of the `year` feature; while we were expecting an association, we didn't expect one of this magnitude. Ideally, park rankings should primarily reflect park quality rather than the year in which the data was collected. However, this effect may be explained by changes in ParkScore evaluation methodology or shifts in national benchmarks over time. If scoring criteria evolve, the same park characteristics could produce different rankings across different years.

Additionally, several `state` indicators had relatively large coefficients, which may reflect regional differences in urban planning priorities, climate, population density, or historical investment in public green spaces.

## Impact of Our Findings and Practical Implications

Since the most important features corresponded to immutable attributes, our results don't suggest many actionable changes. However, if we only look at the point-based features, we were able to uncover some meaningful relationships between point-based park characteristics and rankings. The results particularly suggest that cities seeking to improve their park rankings may benefit from focusing on:

- Increasing the total size of your park within city boundaries
- Increasing the percentage of residents within a 10 minute walking distance
- Increasing public spending on park systems
- Improving park amenities and recreational facilities.

These factors are directly tied to park accessibility and quality, which are key components of urban livability. Urban planners and policymakers could use insights from models like this to better understand which aspects of park infrastructure contribute most strongly to overall system performance. Improving park accessibility and investment may also support public health, environmental sustainability, and social equity, since access to green space has been associated with improved physical activity, mental health, and community well-being.

## Limitations and Improvements

Despite its usefulness, our analysis has several limitations.

First, ridge regression produces continuous predictions, which means that predicted rankings can fall outside of the valid range of rankings. In our final scatterplot visualization of *Actual vs. Predicted Ranks* in Figure 5, we can see that some predicted values were negative, which is not possible for real park rankings. This occurs because linear models like ridge regression can theoretically produce prediction values ranging from  $-\infty$  to  $+\infty$ . To address this, we could explore alternative modelling approaches that may be better suited for ranking outcomes, such as gradient boosting ranking models (example, LightGBM Ranker), which are commonly used to predict ordinal data, and which may produce predictions that better reflect the discrete nature of rankings. However, this is beyond the scope of this course, so we state it as a limitation of our selected model.

Second, the dataset contains relative competitive rankings rather than absolute scores. For instance, a city's rank may change even if its park system remains unchanged, simply because other cities improved or worsened, which introduces additional variability that is not able to be captured by the available features. One possible way around this could be using absolute scoring rather than competitive rankings, which might better reflect improvements in park systems over time and allow models to focus directly on park characteristics rather than relative comparisons between cities.

Third, some variables that we believe are still important in this context and that may influence park rankings are not included in this dataset, such as park maintenance quality, safety, user satisfaction, or environmental conditions. Including such variables might improve predictive accuracy and provide deeper insights into the factors that determine park system performance.

Finally, we decided to remove the `rank_last_year` feature, because the previous ranking may not always be available when we try to predict the new ranking, and because of potential data leakage. We wanted our model to rely more on temporal and geographical features and be able to produce an accurate prediction without the previous ranking, which may act as a proxy for the target rank. So, while removing `rank_last_year` improved the fairness of the model and allowed us to focus more on the other features, it also removed a feature that potentially contained strong predictive information. As a result, the model may not capture some missing context that naturally exists in park rankings, which may have been an interesting addition to have.

## Future Work and Further Questions

This project gave us lots of insights regarding this topic, but it also raised several interesting directions for future research.

We previously mentioned that our model could potentially be improved by exploring alternative approaches, such as LightGBM Ranker, to better reflect the discrete ranking, and that evaluating absolute scoring rather than relative competitive rankings could allow models to focus more on park characteristics.

Another interesting direction would be to examine which park improvements are most cost-effective. For example, policymakers may want to know whether expanding park acreage, increasing spending, or adding certain amenities provides the greatest improvement in ranking per dollar invested, rather than just seeing which features provide the overall greatest improvement in ranking.

Investigating equity in park access would also be interesting. While this analysis focuses on overall city rankings, an important question to ask is whether all residents within a city have equal access to parks. Understanding such disparities in park accessibility across neighbourhoods could help cities create more equitable urban environments so that the benefits can be experienced among all citizens, rather than only the select ones who are not as disadvantaged.

Finally, sustainability would also be an important feature to incorporate, as sustainability for future generations is important in this context.

Some follow up questions we have included:

- How do we ensure that every person has access to a quality park?
- Out of the features that are both important in predicting rank and can be changed, which ones are the easiest and cheapest to improve in a park?
- What would the coefficients of the features look like if every park was ranked on a scale of 1 to 10 instead of being ranked competitively? Would the coefficients change significantly?
- How would the results change if we added a sustainability ranking?

## References

- Chapman, Ronda, Lisa Foderaro, Linda Hwang, Bill Lee, Sadiya Muqueeth, Jessica Sargent, and Brendan Shane. 2021. “Parks and an Equitable Recovery.” The Trust for Public Land. <https://www.tpl.org/parks-and-an-equitable-recovery-parkscore-report>.
- Data Science Learning Community. 2024. “Tidy Tuesday: A Weekly Social Data Project.” GitHub. <https://github.com/rfordatascience/tidytuesday/blob/main/README.md>.
- Husqvarna Group. 2025. “Urban Green Space Continues to Decline Globally- Nordic Region Emerges as a Green Hub.” Husqvarna Group. <https://www.husqvarnagroup.com/en/press/urban-green-space-continues-decline-globally-nordic-region-emerges-green-hub-2406952>.
- jonthegeek. 2021. “Park Access.” GitHub; Data Science Learning Community. <https://github.com/rfordatascience/tidytuesday/blob/main/data/2021/2021-06-22/readme.md>.

Poon, Linda, and Marie Patino. 2021. "The Inequality of American Parks." *Bloomberg*. <https://www.bloomberg.com/news/articles/2021-05-27/the-cities-where-people-of-color-can-walk-to-a-park>.

"The Fundamentals of Urban Greenspaces." 2025. Capital Trees. <https://capitaltrees.org/march-2025-the-fundamentals-of-urban-greenspaces/>.

World Health Organization Regional Office for Europe. 2016. "Urban Green Spaces and Health." World Health Organization. <https://iris.who.int/handle/10665/345751>.