Assessing Supply of Public Tennis Courts with Analysis of Demand Estimation in the Chicago Metropolitan Area

Team 66: Albert Luu, Sang Myun Park, Sang-il Yim, TaeKyu Kang, Xiaoou Zheng
{aluu30, spark930, syim46, shawn52, xzheng325}@gatech.edu

Introduction and Problem Definition

Public tennis courts play a crucial role in promoting public health [1][2], yet there is limited research and analysis on their supply and demand (S/D). This research aims to assess the current availability (supply) of public tennis courts and the level of interest (demand) tennis players have in using these facilities to provide valuable insights for park administrators, local tennis players, and academic researchers. By identifying areas with high demand but low supply of tennis courts, policymakers and stakeholders can prioritize improvements to the right public parks and sports facilities [16] and allocate resources to promote physical activity and health benefits for individuals and communities [19]. To address this issue, we developed a conceptual model to statistically analyze the supply and demand of the public tennis courts. Additionally, we created a comprehensive data visualization tool that helps users determine the supply and demand of public tennis courts in the Chicago metropolitan area.

Literature survey

Literature review suggested that playing tennis has significant health benefits, including improved aerobic fitness, lower body fat percentage, favorable lipid profile, reduced risk of cardiovascular disease, and improved bone health [2]. Additionally, tennis participation has been on the rise in the United States, with a sizable portion of players playing on public parks, school, and community center tennis courts [3][4]. However, resources are limited; decision makers often face the challenge of allocating their limited resources optimally. Identifying areas with high demand but low supply of tennis courts can help policymakers and stakeholders prioritize improvements to the public parks and sports facilities that see higher participation. As a result, this would make physical activity and health benefits assessable to more people. [16][17][19]. Our research aimed to assess the current supply and demand of public tennis courts in the Chicago metropolitan area and provide data driven insights to decision makers in a way that has not been done before; they can then utilize these insights to assist them in allocating their resources to the tennis courts that see more participation [2]. Although the Tennis Participation 2021 Summary Report [10] and other TIA reports exist [9], it may not be applicable to generalize the findings to other cities due to the lack of an algorithmic approach to create a model for generalization. Demographic information can be obtained from data sets provided by the U.S. Census Bureau [8][9][20][21], while official public park entities can provide court data [12]. Tennis player level [11] and age [10] can be used to analyze playing frequency, and an algorithm can distribute matches by day and time to determine minimum and maximum court demand [22]. Popular times for tennis courts can be calculated for each day and hour using Google popular times [23], and visualizations such as interactive maps and reports can provide insights into supply and demand. However, limitations in data accuracy [13][18] and completeness, as well as potential methodological limitations, should be considered in interpreting the results [14][15].

Proposed Method

The previous research in the public park area aimed to provide useful insights for park administrators, local tennis players, and academic researchers by identifying the relationship between park size, park attributes, and physical activity for the public health [5][6][7]. The research was motivated by the growing trend of tennis participation in the United States, with many players using public parks, school, and community center tennis courts. The literature review highlighted the health benefits of playing tennis, including improved fitness levels, reduced body fat, and improved bone health. However, the research recognized limitations in data accuracy and completeness, and potential methodological limitations. It also did not utilize an algorithmic approach to create a model for generalization, which may limit its applicability to other cities. Therefore, our research should be an improvement as it focused on size level and a quantity approach, which can provide more detailed and accurate information about the supply and demand of tennis courts. By using a more comprehensive data set, including data on the size and number of courts, and incorporating an algorithmic approach, our research potentially provides a more nuanced understanding of the supply and demand of tennis courts in the Chicago metropolitan area with the potential to generalize the findings to other cities.

Our approach involved three main steps: data collection and cleaning, analysis, and visualization. We gathered data on population density, demographics, and tennis court locations, cleaned and organized the data, and then used algorithms to calculate demand for different areas and determine the supply-demand ratio. Finally, we created a user-friendly visualization interface that presents the computed parameters to support decision-making. Our approach combined data analysis, computation, and visualization with the goal of providing valuable insights for decision-makers to better promote public health by improving the right public tennis court facilities.

Detailed Description

The algorithm implementation involved analyzing our collected data, a supply analysis, a demand analysis, and a multi-perspective analysis. The supply analysis determined the number of tennis courts in each city within the Chicago metropolitan area. This analysis involved the collection of various data points, including the name of the location, the city, county, and state of the location, latitude and longitude coordinates, and the number of tennis courts at each location. The latitude and longitude data were collected to obtain popular times data from Google Maps. The final dataset contained 3,224 tennis courts at 880 distinct locations across 218 cities within the Chicago metropolitan area.

The demand analysis algorithms comprised of three sub-algorithms. The first algorithm estimated the number of tennis players based on age, ethnicity, gender, education, and income status using defined equations. Given the total population of the city (P), the percentage of tennis players in the city (R) [10], the percentage of the Ethnicity in the city (E), we calculated the number of tennis players in the city by ethnicity.

$$Number\ of\ Ethniciy\ Tennis\ Player\ = round(P \cdot R \cdot E)$$

To accurately estimate the demand for tennis courts in each city, we developed a second algorithm that calculates the weekly number of matches played based on survey data from 10,380 tennis players. The algorithm uses percentages of weekly playing days provided in the survey and symbols including the total population of tennis players (P), the average number of tennis matches per day (M) which is 2.72 [10], the average number of tennis matches per day (T). These calculations take into account the population of players under NTRP 2.5 and above NTRP 5.5, and result in a more precise determination of court demand [9].

$$T = round(P \cdot (0.9411) \cdot M)$$

Finally, the third algorithm estimated the maximum and minimum number of tennis courts demanded by calculating the distribution of tennis matches for each day in a week and each hour in each day according to the average of popular times of tennis courts for each hour from Google Maps. Here are symbols for the formula: demand matches per hour in a day(D), number of tennis matches per day (T), average number of matches per player per day (M).

$$D_{max} = T \cdot Norm_{max} \; D_{min} = T \cdot \left(\frac{1}{7}\right) \cdot \left(\frac{1}{12}\right)$$

The minimum and maximum number of tennis courts demanded were estimated using different equations; the collected data was converted to a Pandas data frame that was then joined according to the geometric values for demand analysis algorithms.

Through the application of supply and demand analyses, we have computed a supply/demand ratio for each city in the Chicago Metropolitan area. Specifically, our analysis produced four distinct supply/demand ratios, which we have labeled as 'S/D Ratio for Minimum Singles', 'S/D Ratio for Minimum Doubles', 'S/D Ratio for Maximum Singles', and 'S/D Ratio for Maximum Doubles'. The minimum and maximum ratios correspond to the minimum and maximum demand levels for tennis courts in each city, while the singles and doubles ratios relate to the type of tennis court, i.e., one for two people and one for four people. To facilitate an interactive user interface, we opted to use the 'S/D Ratio for Minimum Singles' ratio because it had the least amount of missing data; It also captured the highest level of demand for tennis courts. This ratio was further visualized on a map on our webpage.

The last analysis was the multi-perspective analysis which focused on preparing/exploring the data for visual insights on the relationship between demographic factors (Wealth, Education, Age, Ethnicity) and the S/D ratio for each city. This step was also a precursor to developing the user interface that displayed the visualizations. First, we used python to perform a brief exploratory data analysis on the population data that we already had pulled to gain a better understanding of what information we had for wealth, education, age, and ethnicity. Since we also had the S/D ratios for each city, we selected or engineered attributes for each of the four demographic categories to display on scatter plots where the x-axis would be a demographic attribute while the y-axis would be S/D ratio. For wealth, we went with the variables “Average Income” and “Average Home Value.” For age, we pulled the “Percentage Under 18” and “Percentage Over 65” to also create a new field, “Percentage Between 18-65". Then for ethnicity, we included the population percentage attributes for White, Black, Native American, Asian, and Hispanic-Latino. Lastly, for education, we used the percentages in the population data to label each city “Below average high school graduation,” “Above average of high school graduation,” and “Above average bachelor’s degree” with an if-else statement that compared the city’s percentage to the average percentage of all the cities. Now that we had the appropriate data to show the link between the demographic properties and the S/D, we exported and appended the data from a Pandas data frame to the “AllCityData.json” file for the visualization step.

While the analysis work was taking place, the team also scrapped geo-json open-source data and separated it into raw and combined datasets. The raw dataset represented individual city information while the combined and aforementioned “AllCityData.json” incorporated data for all cities. After appending this file with the information from the analysis work, we finally created the web application using the final data from this file. In our web app, the report section included a detailed explanation of assessment and analysis summary, which were major shortcuts for the entire project. We displayed the multi-perspective results in several categories: education, wealth, age, and ethnicity; we utilized the S/D ratios to show its relationship with each demographic category. Lastly, we utilized Mapbox to create the ultimate map-based visualization that included a heatmap to visualize the supply and demand ratio, with distinct colors indicating the level of demand. Overall, this interactive tool is great for the end user to conduct research on the different cities to gain a deeper understanding of the S/D for public tennis courts.

Experiments / Evaluation

To ensure the validity of our results, we reviewed the "plan of activities" outlined in our proposal and assessed/validated whether we accurately satisfied each step. This plan consisted of seven major activities. By carefully following this plan, we were able to systematically check and recheck each step of our process, ensuring that there were no errors or inaccuracies. This rigorous approach allowed us to have confidence in our results and provided a solid foundation for any further analysis or interpretation.

1. Was the tennis court data from each city valid enough?

We had to exclude multi-use areas that are often used for basketball, soccer, tennis, or floor hockey from the data collection. Additionally, private tennis courts from condominium communities in some cities were not counted. We faced some challenges while collecting data. For instance, some cities did not have an official park district website, which required double-checking with Google Maps. Additionally, some tennis court information was only available on Google Maps, while other information was only available on the city park district website and did not appear in Google search results. Despite these challenges, we successfully collected accurate and comprehensive data on tennis courts and related information in 218 cities by manually cross checking the city park district website and Google Maps.

2. Was the population data from the Census valid enough to use?

There were some cities with duplicated names between the city names on the Census website and the targeted city names. By using a data cleaning script in Python, we were able to identify differences of the names between Census dataset and our target city list. Then we reorganized our target city list to the Census dataset. Additionally, we encountered missing data for certain cities in the Chicago Metropolitan area on the Census website. Although we attempted to manually input the population data from other websites, they were not sufficiently credible and did not provide 2021 population data. As a result, we had to exclude invalid cities from our supply-demand ratio analysis.

3. Did the team combine the tennis court data with geometry data?

We combined the tennis court data with geometry data with GeoJson format. Using the open-source repository on GitHub, we were able to identify 218 city boundaries in GeoJson format in the Chicago metropolitan area. However, during our verification process, we discovered that three of these cities (Medinah, Summit Argo, and Wilmot) were invalid GeoJson format. Therefore, we excluded the cities for the analysis and visualization.

4. Did the team design the model to estimate the tennis player for each city?

The number of tennis players is analyzed by the population data according to a USTA report [10]. Because the cities we are evaluating are in the Midwest region of the United States, we used the most recent year's (2020) data point for tennis participation rate of 7.8% in the formula to calculate the number of tennis matches per day (for each city). Our main limitation in validating this process is that we had to rely on the estimated tennis participation rate from the USTA report, not from the survey about the tennis participation data of the Midwest region of the US.

5. Did the team design the model to estimate the demand of the tennis court by tennis player for each city?

We used the second algorithm to calculate the number of tennis matches in a week [11]. The tennis matches are distributed for days and hours by optimal and weighted cases. The optimal cases assume all the matches distributed equally for 7 days and 12 hours per day while the weighted case uses the weekdays and hourly popularity time information. Additionally, the supply-demand ratio is analyzed by the single and double play cases with the tennis court supply data, so the final output is the min/max supply ratio for single/double play cases. However, due to limitations in calculating the exact number of matches that could be played in an hour on any given day, our calculations are approximate. Also, when retrieving popular times data from Google Maps for 880 tennis courts, we found that information for 121 of them was missing. This resulted in some cities having minimum demand data but not maximum demand data. As a result, we used the minimum demand data on the heatmap on our website.

6. Did the team follow implementation of the estimation algorithms?

To validate the proper implementation of the algorithms and data, we compared the resulting scatter plots shown in the "analysis report" section of the web page to the "autoviz" plots generated in the exploratory data analysis (EDA) section in the Jupyter notebook python code. The plots and output from the EDA seemed consistent with what was being displayed on the web page. Educational level and public tennis court supply-demand showed little relationship between each other. Meanwhile, for wealth, there was a positive correlation of 4.78e-3 for Average Income to S/D ratio. There was no noteworthy correlation for the age categories as all the "line of best fits" were all horizontal. Finally, for the ethnicity category, only the white population ethnicity showed a slight positive correlation with the S/D ratio with a slope value of 4.96e-1 while all the other ethnicities showed no correlation.

7. Did the team completely implement the interactive visualization?

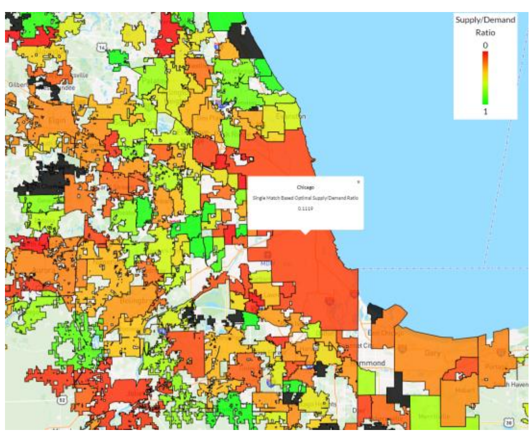

Figure 1. Interactive Map

To visualize the supply/demand ratio of tennis courts in the Chicago metropolitan area, we utilized Mapbox, GeoJson, and D3. Using these technologies, we created interactive-map and report sections on the web page. As the figure 1, the interactive map shows a heatmap that clearly displays areas with high and low supply/demand ratios with color. The greener areas represented a higher supply/demand, while the opposite was colored by red areas. The report section includes two subsections. One is the report summary section that shows the overall summary of the assessment for quick catch-up of supply-demand status in the Chicago metropolitan area. Another is the multi-perspective section that shows scatter plots by incorporating factors such as education, wealth, age, and ethnicity characteristics of each city. The plots provide insight into the relation of the tennis court supply/demand ratio and each factor. With this web page, we provide a comprehensive and nuanced understanding of the state of the art in the public tennis courts in the Chicago metropolitan area. In addition, the webpage is published to the public by the AWS hosting for easy access of the assessment [24].

Conclusion and Discussion

In conclusion, our research project analyzed data and visualized results to assess the supply-demand ratio of public tennis courts in the Chicago metropolitan area. Despite data constraints and time limitations, our web application highlighted the impact of factors such as education, wealth, age, and ethnicity on tennis court demand and provides an interactive interface where the end user can conduct deeper research into the supply-demand of tennis courts in the Chicago metropolitan area. As a result, out of 215 targeted cities, we assessed 185 cities. Among the 185 cities, 172 cities have a supply-demand ratio of less than 1, the average supply-demand ratio is 0.449 and 9 cities do not have any tennis courts, suggesting a potential demand for more public tennis courts to meet the needs of the community. In addition, we could identify several factors from the tennis players in the Chicago metropolitan area. For the foundation for determining the demand for tennis court facilities, we counted 618,902 tennis players. Education, wealth, age, and ethnicity are key factors that affect the number of tennis players and, subsequently, the demand for tennis courts. 60.1% of tennis players are between the ages of 18 to 65. 68.5% of tennis players are identified as white, which is the largest ethnicity group. Chicago metropolitan area tennis players are more likely to have an income of $92,338. 39.9% of tennis players have a college degree or higher.

These insights can inform decision-making for the development of tennis court facilities to meet local community needs and promote tennis participation. As future research, our statistical analysis model could even be applied to other cities or regions of the country; our research can potentially contribute to the development of a conceptual model of the relationship between park design and physical activity [5], as well as introduce a HealthPark concept [6] to investigate park attributes and health benefits [7]. Further data collection and analysis can enhance our understanding of supply-demand dynamics and support evidence-based strategies for promoting physical activity and health in urban areas.

Throughout the course of the project, every individual within the team has wholeheartedly devoted themselves to the task at hand, with each member expending a similar and significant amount of effort. By consistently prioritizing open communication and fostering a collaborative environment, each member was able to contribute, highlight their unique skills, and voice their perspectives towards achieving our shared goal. As a result, the team was able to make noteworthy progress and overcome any obstacles, all while maintaining a sense of fairness and respect for each other's contributions.

References

1. Bedimo-Rung AL, Mowen AJ, Cohen DA. *The significance of parks to physical activity and public health: a conceptual model*. Am J Prev Med. 2005 Feb;28(2 Suppl 2):159-68. doi: 10.1016/j.amepre.2004.10.024. PMID: 15694524.
2. Pluim, B. M., Staal, J. B., Marks, B. L., Miller, S., & Miley, D. (2007, May 15). *Health benefits of tennis*. PubMed Central (PMC). https://doi.org/10.1136/bjsm.2006.034967
3. USTA. (2021, February 11). *U.S. tennis participation surges in 2020, Physical Activity Council (PAC) report finds*. USTA. Retrieved March 2, 2023, <https://www.usta.com/en/home/stay-current/national/u-s--tennis-participation-surges-in-2020--pac-report-finds.html#:~:text=Tennis%20is%20the%20perfect%20social,produced%20by%20Sports%20Marketing%20Surveys>
4. *Parks and Recreation Tennis Programs*. USTA. (n.d.). Retrieved March 2, 2023, <https://www.usta.com/en/home/coach-organize/organization-facilities/parks-recreation.html>
5. Cohen DA, Han B, Nagel CJ, Harnik P, McKenzie TL, Evenson KR, Marsh T, Williamson S, Vaughan C, Katta S. *The First National Study of Neighborhood Parks: Implications for Physical Activity*. Am J Prev Med. 2016 Oct;51(4):419-26. doi: 10.1016/j.amepre.2016.03.021. Epub 2016 May 18. PMID: 27209496; PMCID: PMC5030121.
6. Arena R, Bond S, O'Neill R, Laddu DR, Hills AP, Lavie CJ, McNeil A. *Neighborhood parks have never been assessed nationally to identify their role in physical activity*. Prog Cardiovasc Dis. 2017 Jun-Jul;60(1):152-158. doi: 10.1016/j.pcad.2017.05.010. Epub 2017 May 31. PMID: 28576674.
7. Grilli, Gianluca. Mohan, Gretta. Curtis, John. *Public park attributes, park visits, and associated health status*. Landscape and Urban Planning. 2020, 199. 103814. 10.1016/j.landurbplan.2020.103814.
8. U.S. Census Bureau (2021). *American Community Survey 1-year estimates.* Retrieved from *Census Reporter Profile page for Chicago-Naperville-Elgin, IL-IN-WI Metro Area* <http://censusreporter.org/profiles/31000US16980-chicago-naperville-elgin-il-in-wi-metro-area/>. Retrieved March 2, 2023,
9. Tennis Industry Association (2018). *Getting to know Tennis Players* <https://www.tennisindustry.org/page/research_general>. Retrieved March 2, 2023,
10. Sport Marketing Surveys USA. Tennis Participation Summary Report, USTA Southern, 2021
11. Tishelman, J. Bu, D. Gladstone, J. Colvin A. *Improved General Health Outcomes in US Recreational Tennis Players,* Journal of medicine and science in tennis, 2020 July
12. SPECIAL DISTRICTS (70 ILCS 1205/) Park District Code (2018). <https://www.ilga.gov/legislation/ilcs/ilcs5.asp?ActID=892&ChapterID=15>
13. Wong AY, Ling SK, Louie LH, Law GY, So RC, Lee DC, Yau FC, Yung PS. *Impact of the COVID-19 pandemic on sports and exercise*. Asia Pac J Sports Med Arthrosc Rehabil Technol. 2020 Jul 28;22:39-44. doi: 10.1016/j.asmart.2020.07.006. PMID: 32821648; PMCID: PMC7386263.
14. Office, U.S. Government Accountability. *Assessing Data Reliability (Supersedes Gao-09-680G). U.S. GAO*, Dec. 2019, https://www.gao.gov/products/gao-20-283g.
15. Xie Y. *VALUES AND LIMITATIONS OF STATISTICAL MODELS*. Res Soc Stratif Mobil. 2011 Sep 1;29(3):343-349. doi: 10.1016/j.rssm.2011.04.001. PMID: 22043133; PMCID: PMC3203205.
16. Fatas, Antonio and Singh, Sanjay R., *Supply or Demand? Policy Makers' Confusion in the Presence of Hysteresis*. CEPR. April 2022. Discussion Paper No. DP17232, Available at SSRN: https://ssrn.com/abstract=4121415
17. Thacker SB, Stroup DF, Carande-Kulis V, Marks JS, Roy K, *Gerberding JL. Measuring the public's health*. Public Health Rep. 2006 Jan-Feb;121(1):14-22. doi: 10.1177/003335490612100107. PMID: 16416694; PMCID: PMC1497799.

18. Volenec ZM, Abraham JO, Becker AD, Dobson AP. *Public parks and the pandemic: How park usage has been affected by COVID-19 policies*. PLoS One. 2021 May 19;16(5):e0251799. doi: 10.1371/journal.pone.0251799. PMID: 34010353; PMCID: PMC8133454.
19. Spangler KJ, Caldwell LL. *The implications of public policy related to parks, recreation, and public health: a focus on physical activity*. J Phys Act Health. 2007;4 Suppl 1:S64-71. doi: 10.1123/jpah.4.s1.s64. PMID: 17672224.
20. U.S. Census Bureau (2012). *Chicago-Naperville, IL-IN-WI Combined Statistical Area* < https://www2.census.gov/geo/maps/econ/ec2012/csa/EC2012_330M200US176M.pdf/>
21. U.S. Census Bureau. *QuickFacts* <https://www.census.gov/quickfacts/fact/table/US/IPE120221>
22. Trevor Meier, *The Most Popular Days of the Week to Play Tennis*, Global Tennis Network. 2017 <https://www.globaltennisnetwork.com/tennis-blog/the-most-popular-days-of-the-week-to-play-tennis>
23. Boghe, K. (2021, March 28). The calendar of billions at your fingertips. Medium. Retrieved March 31, 2023, from https://medium.com/analytics-vidhya/the-calendar-of-billions-at-your-fingertips-fa3dc2939fb2
24. Albert Luu, Sang Myun Park, Sang-il Yim, TaeKyu Kang, Xiaoou Zheng. *Public Tennis Court Analysis*. Retrieved April 19, 2023, from http://ec2-3-89-251-158.compute-1.amazonaws.com/