<u>**Team MARKO - Houston Neighborhood Quality Dashboard**</u>

Houston is the fourth most populated city in the United States with a projected population growth of 1.2 million people between 2020 and 2029 [4]. The 311 Houston Service Helpline is a consolidated call center to provide Houstonians the ability to report non-emergency concerns [1]. The Promises and Pitfalls of 311 Data states that "the collection of millions of geocoded data points corresponding to problems also has the potential to reveal important information about the distribution of physical conditions and government-provided goods and services within cities" [15]. The 311 data is currently made available to the Houston public with the intent to create a transparent connection between government and its citizens by providing continuously measured, geographically fine-grained, and non-self-reported measure of both parties' interactions [11].

Unfortunately, Houston falls short of providing a transparent and communicative relationship between the city's government and citizens from the use of their publicly available service request data. The online portal provided to Houston's citizens is a messy, data-logged interface that uses rudimentary means of visualization. As a result, the information is not easily discernible and it is difficult for users to gather any meaningful insight for an area and the problems afflicting it [1].

With the projected influx of residents to Houston as well as the large number of existing residents, it is critical for the population to understand the neighborhoods in which they currently live or will live to provide the best opportunity to prosper. However, while resident surveys can be conducted to gather data on neighborhood quality as perceived by its residents, results would simply indicate a subjective view that would not serve as a reliable predictor of true neighborhood quality [10, 13].

It is this team's goal to better utilize the 311 call data for visualization and objective indication of neighborhood quality of life. Data related to neighborhood maintenance, which was proven to be the "most important single determinant of how well people liked their neighborhood" [9], could increase the fidelity of an objective measure of neighborhood quality. However, mistrust of authority has shown to have a strong correlation to neighborhood quality [6] and could introduce bias to be cognizant of during analysis.

While city dashboards have been around since the 1990's, the majority of them have been used for city administrative purposes [5]. The architecture of such dashboards generally follows a structure of overview first, zoom and filter, and then details-on-demand [17]. The widespread understanding of neighborhood selection [16] proposes that residents select their dwellings independent of city services. Influencing neighborhood selection by making city services something that can be easily understood and analyzed could lead to a change in the understanding of the selection process.

We have created a user-friendly, interactive dashboard with visual statistics of service requests filterable by Houston neighborhoods. The visualization informs current and prospective residents of the quality of neighborhoods based on measurements of objective service data spanning years 2017-2019. The novelty of this dashboard can be attributed to the intended use for residents instead of city administrators. The dashboard also includes statistical visualizations of our dataset, and the outputs of a PCA analysis designed to measure the neighborhood quality rating.

The dashboard uses single public source data, processed with Python and visualized via a Power BI dashboard. The dashboard approach uses objective measures to create ratings of neighborhood quality instead of allowing the user to base their ratings on subjective measures of quality. By catering the visualization to residents instead of city administrators, the dashboard repurposes an existing visualization technique in a new way to benefit an underserved audience. Enhancements of tooltip functionality and improved user experience to navigate different levels of data detail provide new insights for consumers as they investigate their neighborhood. Unlike other commonplace tools, our dashboard allows current and prospective residents to navigate quantitative and qualitative assessments of Houston neighborhoods in parallel. By combining heat-map and statistical visualization in-situ, the user can develop well informed decisions on the quality of Houston neighborhoods based on their own queries. In addition, our Principal

Component Analysis (PCA) determined neighborhood quality rating offers a novel output that can help current and prospective residents identify their optimal neighborhood.

We implemented a layer of design thinking to establish a team environment with organizational, intellectual, and physical capabilities where we were able develop a process that fits our proposed method's needed workstreams [3]. Design thinking provided us the opportunity to review our progress throughout the development of the product. At the conclusion of our project, all team members had contributed equally to the final dashboard and deliverables.

The plan to develop the dashboard was divided into three activity phases. In the first phase to establish Proof of Concept, the key was to identify the proxies from our dataset - which neighborhoods are more favorable compared to others - and customize map features that explain these proxies in a comprehensive analysis. Next, in the Initial Prototype phase we synced the interplay of data with the map interface and investigated the data upload process. The last phase was to refine the Final Product focuses on utilizing a user-centered approach to test functionality and determine if our dashboard is operating as expected.

Table 1. Final Distribution of Activities (2 weeks per phase)

Team Member	**P1: Proof of Concept**	**P2: Initial Prototype**	**P3: Final Product**
Malachi Duffy	Data exploration	Programming support	Report creation
Ahmed Atry	Neighborhood research	Programming support	Programming support
Ricardo Gallardo	Data cleanup	Statistical analysis	GIS Support
Kayla Dunn	Data exploration	Statistical analysis	Dashboard aesthetics
Oyinda Salako	Power BI setup	Power BI Mock-Up	Dashboard testing

Concept dashboards created using Python, D3, and Power BI were all assessed for feasibility during the Proof of Concept phase of development. Of the three, Power BI was selected due to its expansive visualization capabilities and ease of user navigation. Power BI offers the best platform to blend customized analysis for creating a neighborhood quality rating with commercial software to develop a dashboard visualization. In the event of possible transfer of maintenance duties to another party such as the city administrators, there would be little to no effort required to transfer development knowledge.

To make our data more accessible to the intended audience, we took the free sourced Houston 311 request data and performed several data cleaning exercises. First, we combed through the full data set and eliminated any potential neighborhood name discrepancies via search-and-replace, added neighborhood populations, and eliminated any remaining city of Houston department name discrepancies via search-and-replace. Second, in order to simplify one of the dashboard components, we classified each 311 request into one of four classes: Water/Utility, Roads, Trash, and Safety based on the original service request category. Finally, we focus on organizing the data by counts of each service request class and department for each neighborhood. These counts were used in the statistical visualizations deployed in our dashboard.

To further the utility of our dashboard, we calculated a neighborhood quality rating via PCA. To prepare our data for analysis via PCA, we transformed each class count to a total count to a per-capita count by dividing each count by the neighborhood population as of 2017. We did this to ensure that our data was not presenting population-skewed results to the users and so that our PCA was performed on a consistent basis across all neighborhoods. We then ran the PCA on the transformed data and converted the first principal component value (which explains 66% of the data variability) for each neighborhood into a 0-1 scale. The 0-1 scale was then multiplied by 10, using the ceiling function, to determine an integer value, 1-10 neighborhood quality rating with 1 being the lowest and 10 being the highest quality. The results of this analysis were displayed in the neighborhood description tooltip. Data analysis also demonstrated that the first principal component was highly correlated to the Trash, Water/Utility, and Safety classes which indicates that most of the neighborhood quality is credited to service requests submitted in such categories. Moreover, the second principal component was extremely (92%) correlated to the Roads class which is consistent with the local perception of roadings being a major issue in the metropolitan Houston area.

Our dashboard has two distinct views that current and prospective Houston residents can use: "Overview" and "Stats" views. The Overview view allows the user to see the number of requests by class, department, and division in addition to Key Performance Indicator (KPI) values for the percent of requests closed on time and the average cycle time for a request. The Stats view allows the user to drill down into the number of requests, and their type, by neighborhood, in addition to seeing how different neighborhoods compare to one another. To present the two views, our Power BI dashboard uses four main components to display neighborhood 311 request information: Map Area, Tooltip, Filter Area, and Statistics Area (see Figure 1).

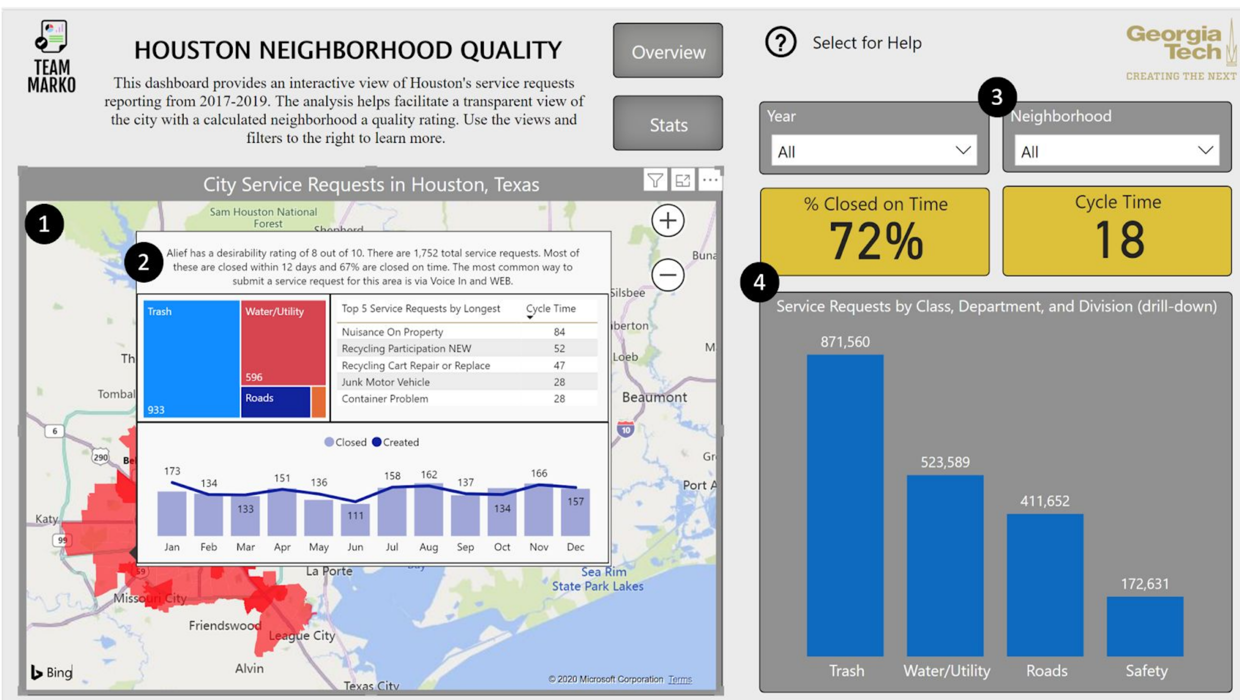

Figure 1. Neighborhood Quality Dashboard

1. Map Area: Houston base map with a color density layer, assigned by total 311 service requests per neighborhood, overlaid on top. This map is interactive and allows the user to drag, zoom in, and zoom out. The color density assigned to each neighborhood changes each time the user changes filter selections (discussed in 3). Additional information for each neighborhood is displayed in the Tooltip that pops up when the user hovers over a neighborhood in the Map Area (discussed in 2). Clicking on a neighborhood in the Map Area will change the statistics area to reflect information for that neighborhood.

2. Tooltip: displayed when the user hovers over a neighborhood to give an in-depth look into that neighborhood. The top of the tooltip provides a synopsis of the neighborhood which gives the user that neighborhood's quality rating, total number of service requests, request cycle time, percent of requests closed on time, and the most common way to submit a service request. Below the neighborhood synopsis is a tree-map, which displays the count of each complaint by category type, and a table of the top 5 service requests by longest cycle time. At the bottom of the Tooltip is a bar-chart and line-graph combination that displays the number of service requests created and closed in a month.
3. Filter Area: allows the user to filter the total data set based by year and by neighborhood. When the user filters the data in the Map Area, Tooltip, and Statistics Area (discussed in 4) will change. In Overview view, the user will also see a tile for the percent of service requests closed on time and a tile for the request cycle time (in days). In Stats view, the user will also see a table of requests, with specifics by neighborhood including: neighborhood quality ranking, total requests, cycle time, and percent of requests closed on time. The user can get more specific information on neighborhood complaints, by call category, by selecting a specific neighborhood in the table.
4. Statistics Area: shows the user specific information about the filtered selections. In Overview view, this area shows a bar graph of the count of calls by class. The user can drill down to get the count of calls by department, and then one step further to the count by division which is the most granular view available to the user. In Stats view, the user will be shown a scatter plot of each neighborhood by percent of request calls completed on time compared to the total count of service calls (see Figure 2). Each data point represents a neighborhood, and selecting any one of the points will adjust the Filter Area and the Map Area.

To guide the user in navigating the four main components of our dashboard and the two views, our dashboard incorporates a help tool. This help tool is designed to answer common questions users might have as they navigate around our dashboard. When selected, two text boxes will appear in key areas that will direct our dashboard users to common tricks, fixes, and tips.

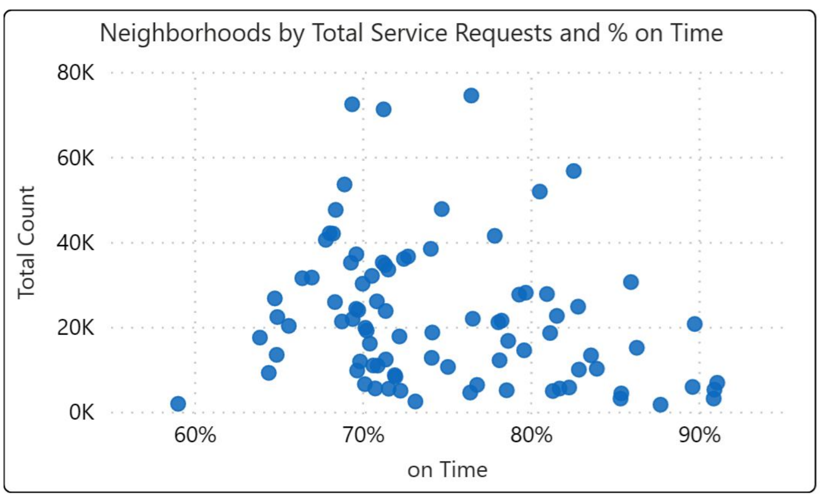

Figure 2. Neighborhood Quality Scatter Plot in the Stats View

We conducted a study of our dashboard with 20 people. Each user was posed a set of data and tool utility related questions. We used these responses to report how accurately each user was able to find the information, and their feedback of the overall tool. Users were presented with one of the below prompts and asked to explore the dashboard with their prompt in mind.

- Prompt 1: You are a new homebuyer, who values clean communities, searching for a suitable neighborhood in Houston to settle down.
- Prompt 2: You and your partner are looking to move from one Houston neighborhood to a more well-kept residential area.

After the user was provided with one of the prompts, the person conducting the experiment provided a basic description of the dashboard including: how to navigate between the Overview and Stats views, what each section of the dashboard was called (Map Area, Tooltip, Filter Area, and Statistics Area), and how to pull up the help tool if they need more assistance during their trial run with the dashboard. Each user was then allotted up to 10 minutes to explore our dashboard. During this time, the experimenter provided no additional help or information to the user.

As the users were performing independent tests on the dashboard, the experimenter was observing: the user's navigation of the dashboard, time spent exploring major components of the dashboard, how much time each user dwelled on a certain issue, and any comments the user had as they navigated the dashboard.

At the end of the allotted 10 minutes, each user was then asked to answer the following questions in the form of a Google Survey. The users were allowed to use our tool while answering this survey. Results from the user surveys are captured in Table 2.

- Question 1: How would you find the percent of service requests completed on time for your neighborhood of interest?
- Question 2: Which of these is the highest quality rated neighborhood in Houston?
- Question 3: What type of issues are most frequently reported in the Midtown neighborhood?
- Question 4: Which neighborhood has the most reported issues related to Trash?
- Question 5: How did you find the navigation of the map?
- Question 6: Do you have any comments for the tool developers?

Table 2. User Survey Results

Metric	**Results**
Question 1 Answered Correctly (%)	80.0
Question 2 Answered Correctly (%)	94.7
Question 3 Answered Correctly (%)	94.7
Question 4 Answered Correctly (%)	78.9
Question 5 Average Rating (1-5)	4.5

Over the course of the user trials, the experimenters were able to notice a number of positive results, in addition some items that could be improved for overall end user benefit. One of the most positive outcomes from our experiments was the ease with which the users were able to interact with our dataset, which was our primary goal at the outset of the product development. Nearly every test user quickly engaged with the filtering capabilities and different views we provided in our dashboard. Another positive outcome from our user tests was the speed with which our help tool was able to answer the common questions that our test users had. After reviewing the help tool, nearly 85.9% of the users who asked questions during their 10 minute trial period were able to get the answers they needed and continue on with their exploration unchallenged.

Many of the users who tested our tool made comments about how they had never thought to use 311 data when looking for new potential neighborhoods to move to but loved the idea of having that information available. Such feedback was reassuring to Team MARKO, since one of our project objectives was to keep our users as informed about Houston neighborhoods as possible. Three of the users also made comments about using this tool, or a similar one, to view complaints specifically related to Hurricane Harvey, which impacted Houston from August to September of 2017.

Of the possible areas for improvement, the most immediately prevalent was the tooltip display when the user filtered down to a highly granular level. When a test user applied multiple filters, the tooltip display would be less meaningful due to granularity. The bar and line chart specifically would show no information for certain months, making the presented information appear sparse. This issue was quickly resolved by Team MARKO prior to project submission. Another area identified for improvement was the overall aesthetic layout of our dashboard. After reviewing some of the early comments from our test users, Team MARKO set out to improve the visual appeal of our final dashboard by modifying the layout and color scheme.

A final comment made by two of the users was related to the visibility of 311 requests. These users said that they would be interested in a map view which included points for every location from which a request originated in lieu of the heat map view that we have included in our final product. In possible future iterations of this product, Team MARKO could consider implementing such a view.

While reducing cost and time to address 311 calls may be a byproduct of the city leveraging our publicly available tool, using finances as a success measure tends to be less significant than the overall innovation of the tool itself [7]. As the novelty of our dashboard is that it is for residents, more emphasis is placed on "implementation, effort and commitment to the analysis, time to market, and user satisfaction" [7] for measuring success. Therefore, we focused on evaluating the effectiveness and value contribution of our dashboard as key measures of impact. It is known that testing products such as ours and collecting feedback serves to evolve and improve the final solution [14]. To improve upon this framework, we developed a small study to understand the user experience. This study ensured the interface matched the expected outcome, helped finalize features, and ultimately tested for overall ease of use and functionality.

The project carries several potential payoffs. Primarily, the ability of users to interact with neighborhood data will prove tremendously beneficial. The perception of a data visualization project is often dependent on its user experience, and the ability of the user to gain an understanding of their desired data [2]. As it exists, this project will help keep current and prospective residents informed of the quality of Houston neighborhoods.

A recognized step in creating a coherent and impactful project is assessing each team members' strengths and weaknesses [12]. Failure to lean in and rely on a team member's strengths could have led to a decline in project quality and perceived output of a team member, both internally and externally. The dashboard was developed by team members with labor free of cost over a six week duration. Each team member dedicated 20% of their time for each week estimating to ~33 hours per week per team member. All members contributed equally to the development of the dashboard and all associated deliverables.

Our dashboard represents a city-wide public good by allowing people to know more about service issues before deciding to move to or work in Houston. It allows for more informed analyses that are crucial to life decisions. While the interactivity of our map may require a bit more technical aptitude to use, the survey showed users not only preferred the ability to customize their own view but also found accurate answers to their questions quicker than if they had searched the web or called the city of Houston themselves [8]. Furthermore, our dashboard presents the opportunity to change how residents interact with the city and vice versa. Our neighborhood quality dashboard empowers Houston residents to make more informed decisions about the places they call home.

Bibliography

[1] Chatfield, Akemi Takeoka, and Christopher G. Reddick. "Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services." Government Information Quarterly 35.2 (2018): 336-347. Web.

[2] Dimara, Evanthia & Perin, Charles. "What is Interaction for Data Visualization?". *IEEE Transactions on Visualization and Computer Graphics* 26 (2020): 119-129. Web.

[3] Dorst, Kees. "The Core of 'Design Thinking' and Its Application." *Design Studies* 32.6 (2011): 521-32. Web.

[4] Egan, John. "Houston Could See Second Biggest Population Surge in U.S. through 2029, Study Says." *CultureMap Houston*, 14 Jan. 2020, houston.culturemap.com/news/city-life/01-14-20-houston-leads-population-growth-2020-2029-report/.

[5] Farmanbar, Mina, and Rong, Chunming. "Triangulum City Dashboard: An Interactive Data Analytic Platform for Visualizing Smart City Performance." Processes 8.2 (2020): 250. Web.

[6] Greenberg, Michael R. "Improving Neighborhood Quality: A Hierarchy of Needs." *Housing Policy Debate* 10.3 (1999): 601-24. Web.

[7] Henttonen, Kaisa, Ojanen, Ville, and Puumalainen, Kaisu. "Searching for Appropriate Performance Measures for Innovation and Development Projects." *R & D Management* 46.5 (2016): 914-27. Web.

[8] Herman, Lukáš, Juřík, Vojtěch, Stachoň, Zdeněk, Vrbík, Daniel, Russnák, Jan, and Řezník, Tomáš. "Evaluation of User Performance in Interactive and Static 3D Maps." *ISPRS International Journal of Geo-information* 7.11 (2018): 415. Web.

[9] Lansing, John B, and Marans, Robert W. "Evaluation of Neighborhood Quality." *Journal of the American Institute of Planners* 35.3 (1969): 195-99. Web.

[10] Mast, Brent D. "Measuring Neighborhood Quality With Survey Data: A Bayesian Approach." *Cityscape (Washington, D.C.)* 12.3 (2010): 123-42. Web.

[11] Minkoff, Scott L. "NYC 311: A tract-level analysis of citizen–government contacting in New York City." Urban Affairs Review 52.2 (2016): 211-246. Web.

[12] Pasha, M., Qaiser, G., & Pasha, U. "A Critical Analysis of Software Risk Management Techniques in Large Scale Systems." *IEEE Access* 6 (2018): 12412-12424. Web.

[13] Pruitt, Sandi L, Jeffe, Donna B, Yan, Yan, and Schootman, Mario. "Reliability of Perceived Neighbourhood Conditions and the Effects of Measurement Error on Self-rated Health across Urban and Rural Neighbourhoods." *Journal of Epidemiology and Community Health* 66.4 (2012): 342-51. Web.

[14] Slavek, N., Blažević, D., & Nenadić, K. "Critical Measures of Success for a Software Project." *Tehnički Vjesnik* (2013): 1119-1127. Web.

[15] White, Ariel, and Kris-Stella Trump. "The Promises and Pitfalls of 311 Data." Urban Affairs

TEAM MARKO

Review, vol. 54, no. 4, July 2018, pp. 794–823. Web.

[16] Wood, Jo, Dykes, Jason, Slingsby, Aidan, and Clarke, Keith. "Interactive Visual Exploration of a Large Spatio-temporal Dataset: Reflections on a Geovisualization Mashup." IEEE Transactions on Visualization and Computer Graphics 13.6 (2007): 1176-183. Web.

[17] Zabel, Jeffrey, and Ioannides, Yannis M. "Interactions, Neighborhood Selection and Housing Demand." Journal of Urban Economics 63.1 (2008): 229-52. Web.

TEAM MARKO