Gainesville Neighborhood Satisfaction Report 2021

Orli Forster, Alok Kothari, Carston Koziol and Clare McAuliffe

City of Gainesville, Strategic Initiatives Branch

BSI IA Capstone Project

SI 405/485

November 10, 2021

Client Background

The City of Gainesville, located in the north central Florida region, has a current population of 143,835 residents and approximate areas of 63.75 square miles. Gainesville is also home to the University of Florida, one of the largest public universities in the nation. The City of Gainesville's Strategic Initiatives Branch is responsible for overseeing the touchpoints between the residents and the city (i.e policies, programs, services, technology, and facilities) and implements projects aimed at improving those interactions. We will be working with Jacqueline Stetson, the Neighborhood-Centered Design Director, and John John, the Service Design Strategist. Since 2019, the Strategic Initiatives Branch has conducted an annual Neighborhood Satisfaction Survey that collects data on its residents -- more specifically, demographic information, needs and priorities.

Project Context

In the past, Gainesville has struggled with inequities in allocation of city resources which can be harmful for community health. For example, East Gainesville received fewer resources compared to the rest of the city. East Gainesville also happens to have a higher African American population in relation to the rest of the city. With data from the 2020 and 2021 Neighborhood Satisfaction Survey, the Strategic Initiatives Branch hopes to address these inequities and understand the lived experiences, needs and priorities of its residents, referred to as "neighbors". The survey collected data from 1,003 households, distributed across 4 common districts. On the City of Gainesville website, there is a dashboard with preliminary analysis from both years (2020 and 2021) of survey data, created by EPC Institute, a market research firm. While this dashboard summarizes the survey results, it lacks a focus on equity and does not provide clear, directed insights. Thus, the City of Gainesville has not been able to act on this data, and therefore, cannot address the historical inequities throughout its community.

Stakeholders

The team's analysis and visualizations of the Neighborhood Satisfaction Surveys and 2020 U.S. Census will be used to inform allocation and prioritization decisions across the City of Gainesville government. The City of Gainesville's Strategic Initiatives branch, who runs the survey, along with the Equity and Inclusion branch, are committed to providing residents with an equitable allocation of resources and will use our analysis to direct this initiative.

One way the Strategic Initiatives branch plans to do this is by distributing the team's report to the City of Gainesville commissioners who convene each year to determine how city resources should be allocated. With this more in-depth analysis, focusing on inequity, the commissioners can make more informed resource allocation decisions.

In the past, individual residents have received an inequitable distribution of resources based on demographics and geographics. As a result of our analysis, the residents can expect their most pressing priorities to be addressed by the city.

Problem Statement

Without equity directed analysis of resident data regarding demographics and priorities, the City of Gainesville can not effectively address historical inequities in the community and work to improve the overall satisfaction of its residents. As a result, the City of Gainesville is seeking insights into where and how to allocate resources most equitably by developing a deeper understanding of demographic trends and how residents' priorities have shifted over time. These insights will enable the City of Gainesville's commissioners to make more informed decisions in determining the city's prioritization agenda and allow the Strategic Initiatives Branch to create a survey that is more efficient for analysis.

Project Objectives

- Understanding of how satisfaction is mapped to geography across Gainesville, FL.
- Generate analysis to give insight into inequalities based on demographics.
- Translate insights into visualizations that are easily digestible by core stakeholders.

Project Scope

The current time frame of four months significantly constricts the breadth of our analysis. Due to this short time frame and the team's expertise in Information Analysis, the team will focus on conducting an in-depth analysis and informative visualizations. The Strategic Initiatives Branch is also looking to improve their survey to better reach resident concerns. Although we will not provide survey design recommendations, our analysis will provide a foundation for future improvements. With only two years of data from the Neighborhood Satisfaction Survey, the team will conduct comparative analysis between the two years of data instead of trend analysis. We will be using 2 Census periods (2010 and 2020) to allow for comparison.

Our minimum viable product is a product report that will contain insights and visualizations to effectively address resident priorities as well as current demographic and geographic inequities.

Impact & Benefits

The data analysis from the Census and Neighborhood Survey has given Gainesville the ability to get views on; perception of life, public safety, mobility, maintenance, utilities, and parks/recreational centers. The main issue is these findings are summed together, and fail to address smaller sub communities within Gainesville. Our exploratory analysis on the Census and Neighborhood Satisfaction Survey will enable the City of Gainesville to make data-driven decisions on how to best serve each community and create an equitable and inclusive environment.

The initial impact will be advice to the City of Gainesville's Equity and Inclusion branch and Communications and Engagement branch, in order to make future decisions regarding

allocation of resources and priorities. Additionally, commissioners will be given the data analysis and visualizations to create more informed community decisions when they convene annually to determine how resources should be distributed.

Advice and findings will stem from demographic trends and an understanding how priorities differ across communities to determine how to allocate resources in the most effective, equitable manner. This will allow for a larger impact on the community, granting commissioners utilize the information and make equitable changes. This analysis will provide the potential opportunity for changes, allowing for residents to benefit from a more equal community Therefore, the overall community of Gainesville will benefit from actionable analysis regarding their satisfaction with the city and in the long term, benefit from more direction in community priorities.

Project Resources & Needs

Research Goal and Justification of Methods

Our main goal is to utilize exploratory data analysis of the recently conducted Neighborhood Satisfaction Survey and Census 2020 to identify correlations between demographics and perspectives of city residents with the proper allocation of financial resources. In addition, this will provide a basepoint to identify potential outliers, and anomalies that may distinguish the Gainesville community.

Neighborhood Satisfaction Survey

Two years of data from the Strategic Initiatives branch's Neighborhood Satisfaction Survey will be one of the main data sources used for analysis. The 2020 and 2021 datasets are publicly available on the City of Gainesville website in CSV format where the rows represent the responses for each Gainesville "neighbor" surveyed, while the columns represent each survey question (and sub-question). The 2020 dataset contains 1042 rows and 223 columns and the 2021 dataset contains 1003 rows and 342 columns.

The survey questions address:

- Gainesville overall as a place to live
- Safety
- Sense of community and belonging
- Quality of life
- Future priorities
- Preferred communication methods
- Mobility
- Past experiences with city leaders
- Logistical information regarding resident's geographic location (zip code, district, and longitude and latitude of street address)

In addition to the survey results datasets, the City of Gainesville has provided additional resources relating to its Neighborhood Satisfaction Surveys:

- Dashboard (2021)
    - Using Microsoft Power BI, the dashboard contains 5 tiles for city performance, who responded, mapping, investment priorities, and benchmarks.
- Key Findings Reports for 2020 and 2021 (PDF format)
    - Finding Report segmented by Age & gender, ethnicity, years lived in the city & expected continued residency in the city, employment status, total household income for 2021 and 2020.
    - Finding Report segmented by Census block groups, City Commission districts for 2021 and 2020.
- Survey questionnaire templates
    - Provided the 8 page survey questionnaire template that was sent out to all residents. Questionnaire available in both English and Spanish.

We will need to rename and condense the survey questions that are set as columns, either by the question number or by an abbreviated synopsis/tagline representing each survey question so that the data is more digestible. Additionally, N/A responses are represented as “9” in the dataframe so we will have to account for this, during the data cleaning phase. Furthermore, each question has multiple subcomponents that will need to be broken down into their own. These components will need to be broken down into smaller data frames to make it easier to navigate for efficient analysis.

U.S. 2020 Census Summary Results for Gainesville

The primary external data source that will be used is the U.S. Government’s 2020 Census results, which is a public dataset in a CSV format. The data contains 83 rows and 3 columns -- each row represents a Census question and the columns represent each “fact” (question), fact note and the corresponding statistic for the City of Gainesville. The Census data is publicly available, and updated twice for redistricting in August and September 2021. For the 2020 Census Data: Since the rows represent each Census question, we plan to match the nomenclature with the Neighborhood Satisfaction Survey columns. Additionally, we will replace abbreviations in the data with full written explanations from the data dictionary.

Since this data reflects a summary overview of Gainesville, FL, it may not provide us with sufficient information to extrapolate how resources should be allocated based on factors such as demographics and geography. If this dataset does not provide the level of granularity we are looking for, we can reach out to Catherine Morse, a Census Data expert at the University of Michigan, for guidance.

Systems & Tools

Folium

Python’s free ecosystem add-on Folium will allow us to plot out the trends over geographical locations. Having a map that organizes trends through scattered plots will help us uncover clustering patterns of demographics based on questions from the survey. Geographical representations of the location corresponding to response allows for a visual aid when working

within the team as well as the client when determining what data is of value and what can be discarded. Carston will be the main owner of this task and research techniques for incorporating the visual tool into the project.

Github

Clare will be in charge of maintaining the GitHub repository, solving any merge conflicts, and ensuring that the data is kept up to date. This will ensure that every stakeholder and UMSI team member has an up to date copy of the data we plan on analyzing.

Tableau

For our main visualizations we will be using Tableau as a BI tool. All University of Michigan students receive a free year for Tableau Desktop, therefore this will be our main product. Clare will be the main lead for this tool given she has exposure, and plans of taking several learning courses prior to our analysis. Tableau will help all of us create visualizations once the data is cleaned and formatted. This paired with some multilevel data analysis in Seaborn will help drive our overall conclusions.

Seaborn

Since we are working with other external data mixed with internal data, Seaborn and Python will allow us to perform significance tests on varieties of data and compare two different sets of data on the same visualizations. Some models we plan to create are chi2 tests and heat maps showing correlations, in addition to utilizing a folium library.

Skills

ETL (Extraction, Transform, Load)

ETL enables the analysis team to monitor specific factors and trends over time. The ETL capabilities will allow us to analyze the categorical and quantitative data on a timeline in parallel to the world pandemic. The SI 330 Data Manipulation course covered aspects of ETL through the Python Data Library, Pandas. Internally, we plan to review this material as well as our own external research, and client conversations on similar tests. Alok will be responsible for reviewing the fundamental components of ETL by the start of the Winter 2022 semester. Supporting team members will be Clare and Carston.

Database Exploration (using Python and Seaborn)

Database exploration methods through python and seaborn will allow our team to find if there are significant correlations between different variables in the datasets, mainly the census. The SI 370 Data Exploration class covered many different ways to visualize data through Matplotlib, Seaborn, and Numpy, therefore reviewing this material will be helpful. This combined with outside documentation regarding folium will allow us to expand our knowledge. Carston will be responsible for learning the folium library by the end of November. Clare and Orli will lead the data exploration aspect by refreshing their understanding of the fundamental components and

procedures involved with conducting a successful exploration by the start of the Winter 2021 semester.

Wider Context

Indirect Comparator - Coral Springs, FL

We have identified our indirect comparators as other Florida city governments with design branches, looking to address equity and inclusion. Research shows that Coral Springs, FL is attempting to approach resident satisfaction and equity very similar to Gainesville. Coral Springs conducted a 2021 Residential survey with 43 questions regarding topics such as demographics, public safety, city services, amenities (i.e available parks/recreation, shopping), and economic development. Evaluation of the publicly available survey shows that Coral Springs attempted similar surveys in 2009, 2011, and 2019. All data collected has been summarized by ETC Institute, the same partner the City of Gainesville has worked with. Announced April of 2021, Coral Springs updated its plans for equity with the creation of the Manager of Diversity, Equity, and Inclusion position under their Human Resources department. The city is now attempting to facilitate this movement forward by reviewing organizational policies and updating policies/practices that may have caused unintentional disparities amongst populations. Educational information will also be disseminated amongst departments to increase awareness and provide support to employees when crafting future decisions and policies. Coral Springs' attention to the issue of equity gives the team the ability to compare and contrast how the surveys are being utilized between the two cities and what insights are being analyzed.

	Gainesville, FL	**Coral Springs, FL**
Population ***per Census Data**	Population = 133,997	Population = 133,759
Ethnicity ***per Census Data**	White alone, not Hispanic/Latino = 56% African American = 21.8% Asian = 7.2%	White alone, not Hispanic/Latino = 41% African American = 22% Asian = 5.4%
Hispanic Origins ***per Census Data**	Hispanic/Latino = 11.9%	Hispanic/Latino = 29.4%
Department Addressing Satisfaction of the Community	Strategic Initiatives Branch	Budget & Strategy

Department Addressing Diversity, Equity, and Inclusion	Office of Equity & Inclusion	Human Resources

Coral Springs Synthesis

With similar population sizes and approaches to identifying and analyzing city internal issues, Coral Springs can be viewed as a viable asset to identify how their available survey was constructed and what issues they were attempting to hone in on. We recommend that Gainesville develop a symbiotic relationship with Coral Springs to dissect what approaches do and don't work in tackling issues related to diversity, equity, and inclusion.

Analogous Research - CMS Baltimore

Centers for Medicare & Medicaid Services (CMS) in Baltimore, MD is a part of the Department of Health and Human Services (HHS). In 2003, CMS took action to make comparative information on hospital performance publicly available for 2 reasons: 1) to create an incentive for hospitals to provide the utmost care to patients and 2) to help individuals make more informed decisions about the hospital they choose. Existing satisfaction surveys do not allow for direct comparison across hospitals. In collaboration with the Agency for Healthcare Research and Quality (AHRQ), development and testing was conducted on the existing surveys with methods including cognitive interviews, consumer focus groups, extensive psychometric analyses, consumer testing, numerous small-scale field tests, and pilot testing. From their analysis, CMS and AHRQ were able to condense the pilot survey from 66 to 29 concise questions to allow for a more accurate comparison across hospitals. Although these methods have proven successful for hospitals across the nation, the team does not have the expertise or training to conduct such development and testing methods.

CMS Baltimore Synthesis

After looking at the methods used by CMS in Baltimore to improve their hospital satisfaction surveys, the team has decided to eliminate this aspect from the project scope. After the team completes analysis and creates visualizations focusing on survey results and equity directed insights, we would advise the City of Gainesville to then look into improving the survey design and we recommend CMS and our analogous research as a helpful resource.

Project Plan

This table demonstrates how each milestone will be reached and who will take the lead. The yellow lines categorize the data into groups. Some milestones within those groups will be completed simultaneously, however larger groupings will be done chronologically. This entails that each section is dependent on the previous.

Task/Activity/Milestone	Duration	Start date	End date	Team Member	Tools / Resources
Data Cleaning					

Filling in NA values	1 hr - 3 hr	1/7/22	1/14/22	Carston	Python/Pandas
Renaming columns for 2020, 2021 Survey	1 hr - 3 hr	1/7/21	1/14/21	Orli	Python/Pandas
Defines groups in data	1 hr - 2 hr	1/14/22	1/21/22	Carston	Python/Pandas
Removing Unnecessary Columns	1 hr - 2hr	1/14/22	1/21/22	Orli	Python/Pandas
Merge Datasets					
Ensure that all types are correct (ie timestamps, ints) and prepare Census data	1 hr - 2 hr	1/21/22	1/28/22	Orli	Python/Pandas
Merge Neighborhood Surveys	1 hr - 3 hr	1/21/22	1/28/22	Carston	Neighborhood Survey Dataset
Merge 2010 and 2020 Census Data	1 hr - 3 hr	1/21/22	1/28/22	Orli	Gov. Census Dataset
Summarizing Data					
Compare / contrasts results between 2020 and 2021 Neighborhood Survey	1 hr - 3 hr	1/28/22	2/4/22	Carston	Python/Sklearn
Identify factors to test statistical significance	1 hr - 2 hr	1/28/22	2/4/22	Orli	Python/Sklearn
Identify visualizations to represent correlations of interest	1 hr	1/28/22	2/4/22	Carston	Python/Seaborn/ Matplotlib
Set Hypothesis	1 hr	1/28/22	2/4/22	Alok	-
Analyze Data					
Identify changes in demographics	1 hr - 3 hr	2/4/22	2/11/22	Alok	Python/Sklearn
Determine if and how prioritization preferences differ across region and demographics	1 hr - 3 hr	2/11/22	2/18/22	Alok	Python/Sklearn
Funding trends per geographic area	2 hr - 4 hr	2/14/22	2/18/22	Alok	Python/Sklearn
Longitudinal Analysis	2 hr - 4 hr	2/18/22	2/25/22	Alok	Folium

BREAK 2/25/22 - 3/7/22					
Midterm Report					
Discover statistically relevant patterns and trends of equity imbalance in the city	3 hr - 9 hr	3/7/22	3/14/22	Orli	Python/Sklearn
Resident Priorities based on Geography	1 hr - 3 hr	3/14/22	3/21/22	Clare	Python/Sklearn
Resident Priorities based on Demographics	1 hr - 3 hr	3/14/22	3/21/22	Alok	Python/Sklearn
Demographics vs Geography	1 hr - 3 hr	3/14/22	3/21/22	Carston	Python/Sklearn
Comparison to Indirect Comparator (Similarities & Differences)	1 hr - 2 hr	3/21/22	3/25/22	Carston	Python/Sklearn
Identify connections to Analogous Research	1 hr - 2 hr	3/21/22	3/25/22	Orli	Data Axel
Data Visualizations					
Use visualization tools to map out certain trends	1 hr - 3 hr	3/25/22	4/1/22	Clare	Matplotlib/Seaborn/Tableau
-- 2D Representations done through Tableau	1 hr - 3 hr	3/25/22	4/1/22	Clare	Tableau
-- 3D Representations done through Seaborne	1 hr - 3 hr	3/25/22	4/1/22	Clare	Seaborne
Final Report					
Draw conclusions of Visualizations and Analysis	2 hr - 4 hr	4/8/22	4/12/22	Carston	-
Completion of final write up	2 hr - 4 hr		4/15/22	Alok	Google Docs
Prep slide deck for presentation	2 hr - 4 hr	4/12/22	4/19/22	Orli	Google Slides
Final Draft	2 hr - 4 hr		4/19/22	Clare	Google Docs/Slides

Client Communication

For all client communication, all team members will attend the following meetings. 3 days prior to each, the team will meet to set the agenda and send it out to the client.

	Date	Topic
Client Meeting	14-Jan	Report Review
Meeting with Equity and Inclusion team	TBD January 2022	Stakeholder Meeting
Client Meeting	21-Jan	Data Cleaning
Client Meeting	14-Feb	Merge and Summary Review
Client Meeting	25-Feb	Analysis Preview
Deliverable Check in	14-Mar	Mid Report Check
Client Meeting	28-Mar	Analysis Review
Client Meeting	4-Apr	Visualization Review
Client Meeting	11-Apr	Final Questions and Concerns
Final Deliverable Due	19-Apr	Review of Final Draft

Role Assignment

Although everyone will communicate and ensure each individual is up to date on data analysis and progress, we will split up main roles to have a lead and supporting lead.

Role Name	Description & Responsibilities	Lead	Supporting Lead
Project Manager/ Communications	Responsible for overseeing all aspects of the project including to update the tracker	**Orli Forster**	**Alok Kothari**

Lead	regularly, delegate the workload appropriately, hold team members accountable, ensure all deadlines are met, facilitate communication, and act as a liaison between the client and the UMSI team.		
Editor	Responsible for reviewing written products and communications to ensure context is appropriate and formatted correctly.	**Carston Koziol**	**Orli Forster**
Data Cleaning	Responsible for filling the n/a values, dropping columns, and merging the data sets to an identified group format, fit for analysis.	**Carston Koziol**	**Orli Forster**
Data Analyst	Responsible for interpreting correlations between variables in datasets and cross relational trends as well. In addition, they will be in charge of testing to decipher if there is statistical significance between variables.	**Alok Kothari**	**Clare McAuliffe**
Data Visualizer	Responsible for constructing visualizations from analyzed data to more easily convey results within the team and to the client.	**Clare McAuliffe**	**Alok Kothari**