Katherine Garcia Grijalva
INST414
Sprint 3
April 27, 2025

Sprint 3

Executive Summary For Non-Technical Stakeholders

Healthcare accessibility is based on several societal factors, more so when advocating for digital healthcare. These societal factors include those classified as economic and digital infrastructure influences. Adoption of telecommunication platforms in the healthcare industry has become increasingly more prominent, i.e. video call doctor's appointments, pharmacy prescription delivery services, patient laboratory portals, etc. As these services become more widespread, successful consumer and community engagement relies heavily on adequate internet access, providing this research's reasoning for determining accessibility to telehealth services based on internet access levels.[1] Throughout the model organized in this predictive analysis research, unemployment and employment data were gathered for the population of the DMV-Metropolitan area in an effort to examine how this socioeconomic factor could form a relationship with telehealth/healthcare accessibility. Using employment status to base the research upon was determined through employment status's connection to income levels, employment healthcare benefits, digital literacy probabilities, etc.[2] Viewing employment status

[1] Phuong, Jimmy, et al. "Telehealth and Digital Health Innovations: A Mixed Landscape of Access." *PLOS Digital Health*, Public Library of Science, doi.org/10.1371/journal.pdig.0000401. Accessed 16 May 2025.
[2] Silver, S. R., Li, J., & Quay, B. (2022, January). *Employment status, unemployment duration, and health-related metrics among us adults of prime working age: Behavioral risk factor surveillance system, 2018-2019*. American journal of industrial medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC8678322/
[3] Zuckerman, S., Gonzalez, D., & Karpman, M. (2020, February). *On Eve of 2020 Census, Many People in Hard-to-Count Groups Remain Concerned about Participating*. https://www.urban.org/sites/default/files/publication/101732/on20eve20of20202020census2c20many20people20in20hard-to-count20groups20remain20concerned20about20participating.pdf.

as an umbrella term in this research for these specific factors allows for further investigation, if necessary, into which of the mentioned areas affected by employment status should be prioritized to address by stakeholders.

Problem Statement & Business/Social Context

As mentioned previously, in today's world, access to healthcare services is increasingly dependent on both digital connectivity and economic stability. The expansion of telehealth platforms were first heavily implemented during eras like the COVID-19 pandemic where in-person meeting was a difficult normality to navigate and emphasized the popular guideline that quality healthcare did not necessarily need to be carried out in person.[1] This switch to relying on telehealth services made broadband internet access a necessary requirement for receiving quality and safe healthcare. The presence of socioeconomic barriers, such as employment status, also limit individuals' ability to afford healthcare services, insurance coverage, or even devices necessary to access this specific kind of care.

This shift toward digital healthcare presents the growing challenge that those without adequate internet access or employment resources are at risk of being left out of this expansion of telehealth-focused services. In regions like the DMV-Metropolitan area, where socioeconomic demographics are very diverse, understanding the relationship between internet and economic infrastructures with healthcare access is essential to successfully advance these kinds of services to similar areas to the DMV-Metropolitan throughout the country, and not solely in the context of social and economic demographics.

Delving into this social issue, the goal of this research is to help address the most influential factors enlarging the gap between communities with disadvantaged standings and

access to healthcare services. By analyzing U.S. Census public datasets, this research seeks to answer the research question: “Is there a clear correlation between socioeconomic factors and healthcare access in the DMV-Metropolitan area, using broadband availability and employment status as high-stakes determining factors in access to telehealth services?”

This question opens up a plethora of accessibility research points where the variables being analyzed could be further applied to more complex issues, such as healthcare coverage and digital literacy levels among the same groups focused on in this research, which would be valuable to many business structures and ventures. Non-governmental organizations (NGO) working in the healthcare realm and advocating for the improvement of community healthcare, such as healthcare clinics, insurance companies, university-level healthcare centers and/or pharmacies, benefit from this kind of research through the identification of target populations most prone to exclusion from their services, and in this case not just from physical healthcare services, but from digital healthcare channels as well. For these NGOs and other business models, the findings from this analysis can help support and develop adequately targeted outreach efforts and refine healthcare services for more equitable service delivery. Understanding how deep digital and economic factors influence the accessibility of these services will empower any possible stakeholders to invest limited resources into the most influential areas to have the greatest positive impact.

Methodology

My research data is collected through one primary data source: the U.S. Census Bureau’s American Community Survey Data (ACS), Selected Economic Characteristics. The ACS databases include several demographic and socioeconomic filtering characteristics that are

relevant to this research, making for tailored curations of datasets to study the key factors of healthcare access on the geographic scale of the DMV-Metropolitan area. Data initially projected to be used in my research, which included the Prosperity Index dataset, as well as several ACS databases, such as the "Selected Social Characteristics Data Set," and the, "Demographic and Housing Estimates Data Set," did not end up being used for several reasons. As I progressed in finalizing my models, I discovered the lack of relevance that several datasets proposed in the areas I wanted to focus on in my research. Because of the large amount of data in the ACS database, I initially planned on utilizing multiple influencing factors, not solely employment status and broadband access, but quickly realized the immense amount of feature engineering I would have to incorporate in my research in order for accurate conclusions to be made. Ultimately, the more factors used, the more overlap I had with how closely related each possible influencing factor was to one another, making for inconclusive and/or possibly indirect correlations to healthcare access. Further optimization of the Prosperity Index dataset, this dataset was filtered to retain only U.S. data, and in doing so only displayed a single relevant score for healthcare prosperity for the entire U.S. Because of this, the data was irrelevant to the minimized scope I set forth for this research, and therefore was dropped from being included in the refinement of my data sources.

Because the datasets provided by the ACS databases incorporated many variables in their extensive data survey results, a lot of the data cleaning involved in my data preparation process revolved around removing irrelevant variables that would not be used for models and data visualizations; the variables filtered allowed for datasets with only isolated variables that had the strongest possible ties to healthcare access in the scope of my research. In this case, some of the excluded variables that were removed included: income levels, poverty status, home computer

access, etc. Further data manipulation done with these datasets involved removing incomplete entries, particularly rows with missing values in critical variable fields. Through this data preparation process, I ensured that my modeling choices would be concise and based on clean and relevant data. The final datasets with the finalized variables and fields used are provided below:

Broadband Access Dataset (ACS: S2801)

Broadband of any type	Number of Households

Unemployment (ACS: DP03)

Unemployed Households	Number of Households

Employment (ACS: DP03)

Employed Households	Number of Households

Model Evaluation and Selection

When refining my modeling approach, I made sure to take into account the more-than-one influential factor element that exists in the data for my research. The primary goal set for my models was to assess how the relationship between the population considered to be employed/unemployed and the population with broadband access is representative of access to healthcare services, with an emphasis on telehealth services. I decided on constructing a combined scatter plot, with the x-axis representing population and the y-axis representing counties within the DMV-Metropolitan area. Multiple data points were plotted per county to show employed and unemployed households, as well as the number of households with

broadband access. This kind of scatter plot allowed for easier interpretation of any patterns that could be found between the three data points within each county's x-value. With essentially 3 different values in a single X-plot, per county, the most influential and/or correlated socioeconomic factor would essentially be blatantly obvious through viewing which of the two plots in the X-plot, 'Unemployed' or 'Employed', is closer in proximity to the plot point representing broadband access. Nonetheless, there will be more advanced models used to determine exact correlation measurements. This model also perfectly aligned with the mentioned goal of keeping the research data concise, clean, and relevant.

To apply a model that would determine the relationship between employment levels and broadband access, a linear regression model was used. The linear regression model would be able to demonstrate both the relationships, "Unemployed vs. Broadband," as well as, "Employed vs. Broadband." The employment data, defined through the number of households defined through 'Unemployed' and 'Employed', served as the independent variable, while the dependent variable was defined through the number of households with broadband access. Through these visualization models, I was able to then utilize model evaluation metrics, specifically R^2 score and correlation coefficient calculations, which would in turn provide me with strong statistical conclusions around the accuracy of the models created. The evaluation metrics serve as assessments for the accuracy of the prediction capabilities of the linear regression models, with the R^2 score and correlation coefficients measuring variance and strength of positive or negative impact, respectively. The significance of evaluating these values lies in the fact that the study of a diverse area like the DMV-Metropolitan area could be applied to different metropolitan areas like it. Generalization and the extent of possible error of the data findings would sway any conclusion's credibility for possible further application in future stakeholder needs.

Blatant limitations to the selection of models and content of models was mentioned previously; the complex overlapping factors affecting healthcare accessibility was difficult to display without largely extensive analysis and cross-validation efforts. Again, because of the large list of possible affecting factors, a lot of the initial variables and datasets thought to be relevant and useful were ultimately discarded from the scope of my research. Through this decision, the limitations to the model's correlation and explanatory power and relevance increased as well. Focusing solely on a specific sector of the influential factors to healthcare accessibility limits any models from considering other possible strongly-suited factors, and could possibly completely ignore higher correlated conclusions, as well as highlight uncorrelated variables.

Implementation Strategy and Business Value

The overall point of this research project was to evaluate how effective predictive models could be to provide compelling evidence of correlations between low-resource environments with access to a basic human right. The findings from this project would serve to support businesses and/or organization initiatives that work to improve healthcare access. In the geographic scope of this project, areas in the country that resemble the DMV-Metropolitan Area would find this project the most useful and compelling to implement in their business practices and strategies. Addressing factors affecting low-resource groups, especially in areas that lean more towards having technologically advanced infrastructures like the DMV-Metropolitan, provides these businesses in these areas with insights into any flaws in their implementation strategies of their telehealth healthcare access services, or even their broadband internet systems depending on the stakeholder in question. Using broadband access and socioeconomic standings,

this research was able to conduct correlation analysis between these factors and set forth a basis for equipping stakeholders with information to implement initiatives to expand their client reach and exposure of their services.

The key point to this research's findings is to determine if in a technologically advanced area, like the DMV-Metropolitan, socioeconomic infrastructure factors, like employment status and broadband access, can be indicators for identifying telehealth healthcare access barriers. Implementing the findings from this research in a real-world setting could involve taking into account the possible change of demographics found in an area over time, differences in low-resource levels, as well as curating geographic mapping tools and clinic business partnerships:

1. **Changing Demographics**: Having access to information archives that provide insights on the possible changing demographics within an area, i.e. broadband access levels by county and employment status by county, would aid in keeping findings updated and accurately representative of the area in which business initiatives are targeting. Being introduced to the important access to databases like the ACS through this research can be used to monitor these socioeconomic factor levels and enable these models to be updated regularly.
2. **High Risk Scoring**: A model that scores geographic areas on their likelihood of following the same trends that lead to poor healthcare access could be an additional vital predictive model to accompany this research. A system like this would essentially be able to flag any areas that follow a pattern of low employment and low broadband access and have a 'low score' indicating a high risk of conflict between business consumers and the business's services, if that were to be the correlation findings of this research.

3. **Mapping Tools**: Using mapping tools that visualize geographic data will be essential in pinpointing which regions the ideal levels of broadband access and employment levels are low. It could serve as the visualization model to the 'High Risk Scoring' system. This visual mapping aids in the strategic communication within business teams of where to target their telehealth services and programs to expand awareness.
4. **Clinic Business Partnerships**: With overlap in the factors that correlate to healthcare access barriers, initiatives to partner between stakeholders to co-target areas that fall within their industry could be put into place. For example, clinics that are looking to push their use of their telehealth services could partner with local offices of internet providers that are highly used in the area, providing these clinics with internet access and an analysis and optimization of their telehealth services systems as a whole. A possible marketing promotion that spreads awareness to their local consumer base of their new healthcare partnership would serve as a marketing pull for people to seek out these healthcare options. This could look like a co-sponsored clinic event targeted towards the local community that promotes community health while also serving as a digital literacy learning experience.

From a business and social impact perspective, the results offer clear business value in the form of positive community growth, as well as great business expansion opportunities. These implementation examples touch on improving business infrastructure to better communicate with communities in areas like the DMV-Metropolitan. These strategies provide businesses with optimized investments to reduce healthcare inequalities between different employment groups. It also serves as a catalyst for other future models that look to address correlations between other possible factors and healthcare access. Furthermore, the research done could also help expand

business' research methods by adopting the same structure of data analysis and applying it to aspects of healthcare accessibility outside of the scope of this investigation.

Ethical Considerations and Limitations

The ethical considerations and limitations touched upon in this research project centers around underrepresentation and missrepresentation in data collection, as well as marginalized sensitivities.

With large databases, such as the American Community Surveys (ACS) used, the data collected may underrepresent communities that form part of the DMV-Metropolitan area. While the analysis done in this research was not explicitly aimed towards analyzing the gap between healthcare access and multiple marginalized populations, it is vital to work with representative and ethically sourced data for conclusive results to be accurately data-driven and supported. Common underrepresented communities in Census collected research could include: undocumented immigrants, unwilling survey recipients, and individuals with low English proficiency, which could make up an important population for further healthcare accessibility research.[3]

Moving onto missrepresentation in the data used, using broadband access as an accurate representation for digital inclusion into telehealth services assumes populations with internet access automatically participating in telehealth services and being able to afford access to telehealth services, which is not always the case. For example, some households may have broadband availability, but are not able to afford subscriptions and/or lack the digital literacy to

seek out and make use of telehealth platforms, as well as simply choosing to not partake in using telehealth services, all factors that are harder to singularly quantify.

To reach the project's goal of effectively utilizing predictive analysis models to identify areas that are more prone to the lack of telehealth resources, communicating the results found in a non-stigmatizing manner, making sure to not promote negative labels put onto marginalized populations, ensures ethical implementations for all stakeholders involved. While discussing implementation strategies, the "high risk scoring" portion, without sufficient context, could enforce a negative connotation with the community it is set out to identify.

Research Conclusions

Overall, this analysis relies on a small portion of a plethora of factors linked to the social issue of healthcare access. Healthcare accessibility is formed by intersecting factors of social, economic, demographic, and overall complex infrastructures that also differ widely depending on geographic location. While the current model does not reflect the complete nuances when addressing healthcare accessibility, it is a stepping stone in the correct direction of advancing predictive model analysis applications and power. My analysis used statistical modeling techniques, including linear regression and correlation analysis, to determine whether a relationship exists between employment status and broadband access, showing that certain communities experience both high unemployment and low internet access, conditions that can severely limit their ability to access healthcare services, especially virtual care.

GitHub Repo: https://github.com/kp-gg/INST414final.git