**\**

**Term Project – Milestone 5**

Christina Vosnak

Data Science, Bellevue University

DSC 640

Professor Schneider

March 2nd, 2025

**Summary of Analysis**

For the term project in this course, I will be analyzing trends using data from the national database of childcare prices which includes data from the years of 2008 to 2018 across each state and county in the US. The data within this document includes information regarding Median Household Income, various levels of childcare prices, and demographic information with each metric broken down into various measures. For example, income metrics can be broken down by household demographics, genders, and employment status while childcare prices can be broken down by the age of the child, the percentile of the childcare price average, and whether the childcare is located at a center or at home. Additionally, there is general information regarding specific demographic information such as race, ethnicity, and employment/unemployment rates.

**Findings**

At first glance, the data provides valuable insights into the database of childcare prices. I first looked at childcare prices over time to see how they increased in comparison to median household income. I found that MHI has increased by over 15% in the last 10 years compared to childcare prices, which based on the level of care have increased about 18%-20% meaning that families on average are spending more a greater percentage of their income each year on childcare.

I then was curious about childcare prices by state to see which states had the highest rates for childcare. I found that Washington DC, Massachusetts, and Hawaii rank highest in childcare prices but also rank high in terms of Median Household income. I found that New Jersey has one of the highest Median Household incomes yet the average cost for childcare ranks much lower than states with similar MHIs. I did find it interesting that as median household income increased across states, so did childcare, showing that there is a correlation between the two variables. I then dove into workforce information for males and females to see what percentage of each gender was participating and how it is changing over time. I found that between both genders, labor force participation is decreasing over time, however, particularly with females, those with children are participating in the workforce more in order to support their families. This statistic can be beneficial for employers to understand labor force participation rates and demographics to potentially include childcare benefits in their employment offers for a competitive edge.

**Assumptions**

For this data, I assume that the information provided is correct and that the analysis within PowerBI is accurate, especially for covering inconsistencies with missing data. Across the board, I assume the information provided can be comparable across states and counties in terms of definitions of segmentation between types of childcare costs based on age. There are sections of the data were information was “imputed” however, I did not focus on that information or flag information for this project as I did not see the benefit of including it in this overall analysis.

One thing to note is that I focused on overall state information. If an individual looks at this analysis, information may vary from a state’s average compared to their particular area. An area for future opportunity is to break down the information by County instead of by state to have more relative information for an individual community.

**Items that still need clarification**

One aspect of the information that I didn’t focus on was demographic information, specifically racial disparities. For future expansion of this project, I’d like to further understand how race plays into MHI, childcare costs, and labor force participation rates. Additionally, I focused on the average household and didn’t separate the information by single income household, or differences in family dynamics. While this information is available, it is still another area to dive deeper into in the future of this project when developing a more depth analysis. Ultimately, the purpose of this project is to cover a general understanding of the information to reach general audiences for decision-making.

**Direction of Story/Plan of Attack/Message you want to get across**

Overall, the primary objective of this project is to highlight the financial burden that childcare costs place on families, particularly in areas that high childcare cost to Income ratios. By emphasizing the average cost of childcare costs in terms of a percent of annual income, I hope to address affordability disparities amongst states to ignite policy discussions within local governments to advocate for subsidies, tax benefits or employer-supported childcare options to ease the burden of costs on families. The key message that this analysis will show is that childcare is a critical, yet expensive necessity and it requires assistance from policymakers to help make it affordable. Additionally, I hope to highlight overall labor force participation rates and how the trends are similar except amongst working women with children, to provide an opportunity for employers and policy makers to assist families with childcare costs.

**Target Audience**

There are three target audiences for this project. The PowerPoint will target policy makers to highlight overall trends in the United States surrounding childcare costs, median household incomes, and labor force participation rates. The Dashboard will target the general public, specifically families, and individuals with children to highlight the same information from the PowerPoint but in a different light to urge these individuals to call upon their local governments to enact change for childcare. Finally, the infographic will target employers by highlighting how childcare costs can affect the workforce especially women with children and by providing childcare benefits, companies can benefit economically.

**Mediums you intend to include. You must have 3 and they must all be different.**

I plan to include a PowerPoint presentation as my focus for analysis to present to policymakers that includes recommendations to ease the financial burden amongst families. This PowerPoint will walk policy makers through the statistics on how median household incomes, childcare costs, and labor force participation rates are changing over time and will provide recommendations at the end on how to help working individuals combat the rising costs of childcare so they can effectively support their families. I also plan to create an interactive dashboard within PowerBI to explore childcare costs by state, year, and affordability metrics that targets families across the United States. The PowerPoint and Dashboard will serve as complementary tools, targeting different audiences to shed light on childcare costs. This will allow all audiences to educate themselves about childcare affordability throughout the United States as well as their specific county. Finally, I plan to create an infographic with key points that will serve as a call to action for employers to include childcare benefits as part of their employment offers with families. Overall, labor force participation rates are decreasing, by providing childcare benefits, employers can remain fully staffed to operate efficiently and effectively.

**Design Decisions**

For the most part, I believe that I selected mediums that will resonate well with each target audience. For each medium, I included enough information to inform the audience, but not to overwhelm them. Each visual is simple and easy to understand with clear data and axis labels so that the audience can quickly interpret the information. I kept colors simple to not draw the audiences away from the actual data, but I made sure to present different segments of information in a different color to differentiate each statistic.

**Ethical Considerations**

Overall, the data for this project came from the National Database of Childcare Prices. The information presented is accurate and from a reliable source. Each variable is clearly explained within the technical guide. For my analysis, I removed unnecessary columns to focus on the data for my specific direction. I also created new columns that equated childcare prices in terms of a yearly metric rather than weekly to compare the information to median household income. I also created a new column that compares annual childcare rates to the percentage of household income to easily create a percentage bar chart that shows by state the average percent income spent on childcare each year. Otherwise, I did not manipulate the data in any additional way that would be a concern for ethical considerations.

One critical ethical consideration for future analysis is the effect of regional, racial, and family dynamics on the results. When conducting a deeper analysis, it is essential to approach the data with caution to avoid bias and stereotyping, ensuring that insights are presented fairly and accurately. Misrepresentation or oversimplification of trends could reinforce harmful stereotypes or lead to misleading conclusions about different communities. If visualization reveals a significant trend in the data, it is crucial not to imply causation without proper statistical support. Correlation does not inherently mean one factor directly influences another, and drawing such conclusions without rigorous analysis could mislead audiences or contribute to misinformation.

**Lessons Learned**

From this data analysis, I have learned that it is challenging to incorporate every single piece of data available while still maintaining a clear and impactful narrative. With such a vast amount of information, there are numerous possible directions for exploration. However, regardless of the analytical approach, it is crucial to focus on key findings that can effectively inform and inspire action. Prioritizing the most relevant insights ensures that the analysis remains both meaningful and actionable rather than overwhelming or diluted by excessive details.

Moving forward, I would like to delve deeper into demographic factors, racial disparities, and family dynamics to better understand how these elements intersect with childcare costs. By analyzing these relationships more closely, I hope to uncover patterns that could help inform and shape policies aimed at improving affordability and accessibility for families from diverse backgrounds.

Additionally, this analysis has underscored the importance of high-quality data visualization in effectively communicating complex findings. Well-structured visual representations not only enhance comprehension but also make it easier to highlight disparities and trends that may otherwise be overlooked.