Milestone 5 - Childcare Costs in America: A Growing Burden

***Name**: Akshay Sharma*

***Course**: DSC640 – Data Presentation and Visualization*

***Term**: Summer, 2025*

***Project Title:** Childcare Costs in America: A Growing Burden*

Summary of Analysis

This project rigorously analyzed the National Database of Childcare Prices (NDCP), specifically focusing on Family Child Care (FCC) weekly costs at the 75th percentile for Infants, Toddlers, and Preschoolers spanning the years 2008 to 2018. Utilizing Python for data processing and visualization, the analysis aimed to uncover and present national and regional cost trends over this decade. The primary objective was to highlight the escalating financial burden of childcare on American families.

Findings:

The analysis yielded several critical findings:

- ***Infant Care is the Most Expensive:*** *Consistently, infant care has emerged as the most expensive form of childcare. Its weekly cost rose significantly from approximately $97 in 2008 to about $114 in 2018, representing a 17% increase over the decade.*
- ***Consistent Increases Across Age Groups:*** *Toddler and preschool care also experienced steady increases, with weekly costs rising by $17-$18 during the same period.*
- ***Linear Upward Trajectory:*** *The upward trend in childcare costs appears almost linear, suggesting systematic increases rather than fluctuations tied to specific economic anomalies.*
- ***Weak Correlation with Unemployment:*** *A heatmap analysis revealed a weak positive correlation between the unemployment rate (age 20-64) and infant care costs. This suggests that higher unemployment rates do not consistently lead to lower childcare prices, due to fixed operational costs within childcare facilities.*
- ***Significant State-Level Variability:*** *Box plots and bar charts demonstrated substantial cost variability across states and even within counties. States like Massachusetts, California, and New York consistently ranked among the costliest for infant care, indicating pronounced regional disparities. The annual cost of infant care in high-cost states can even exceed public university tuition.*

Assumptions:

Throughout this analysis, the following assumptions were made:

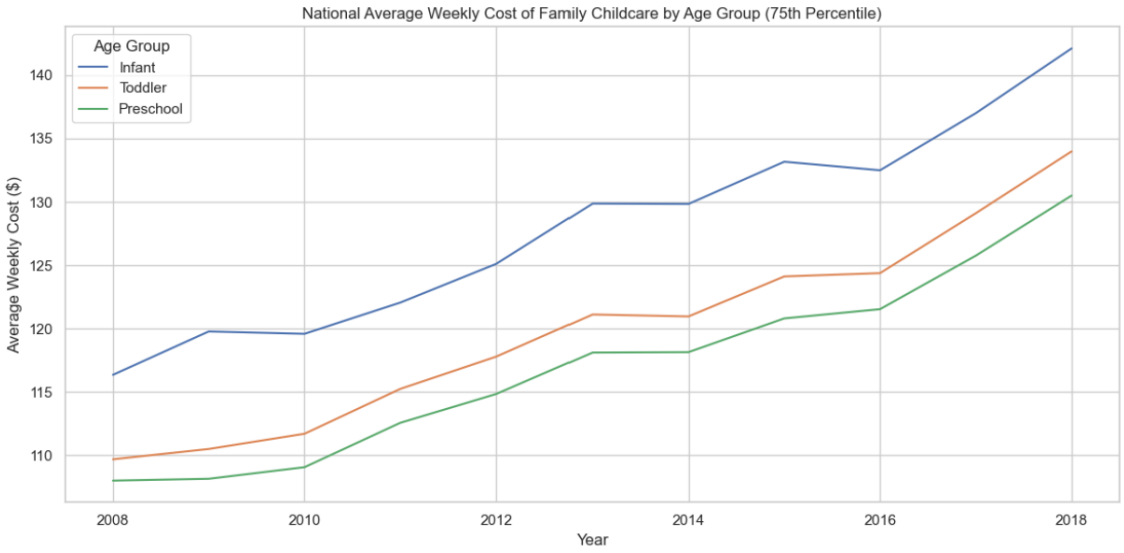
- ***Data Accuracy:*** *It was assumed that the data provided in the "_75FCCInfant", "_75FCCToddler", and "_75FCCPreschool" columns accurately reflect market trends for family childcare at the 75th percentile.*
- ***Data Representativeness:*** *The dataset was assumed to include complete and representative data for all U.S. counties, although the primary focus of this analysis was on national and state-level averages.*
- ***Cost Drivers:*** *While the analysis highlights cost increases, it acknowledges that these may also be influenced by local laws, provider availability, and income inequality, which were not explicitly modeled.*

Items That Still Needing Clarification:

To further enrich the understanding of childcare costs, the following aspects warrant additional investigation:

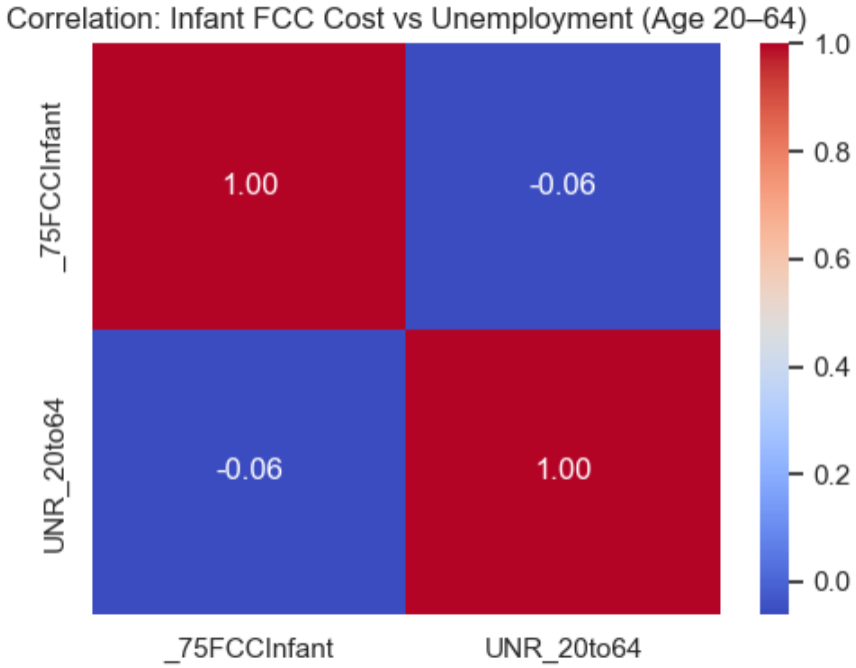
- **Specific Policy and Economic Drivers:** A deeper dive into the specific policy changes, supply chain dynamics, or labor market shifts that contributed most significantly to the observed cost increases would provide valuable context.
- **Inflation and Income Adjustment:** A comparative analysis of childcare costs adjusted for inflation and median income per region would offer a more nuanced understanding of the true financial burden on families.

Visualization:

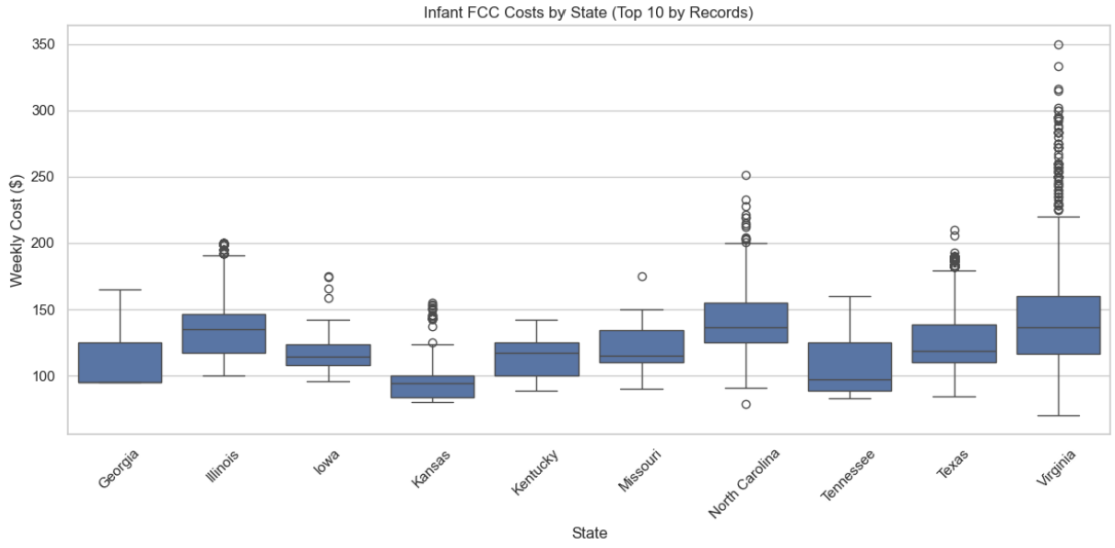


Correlation: Infant Care Costs vs. Unemployment Rate (Age 20–64)

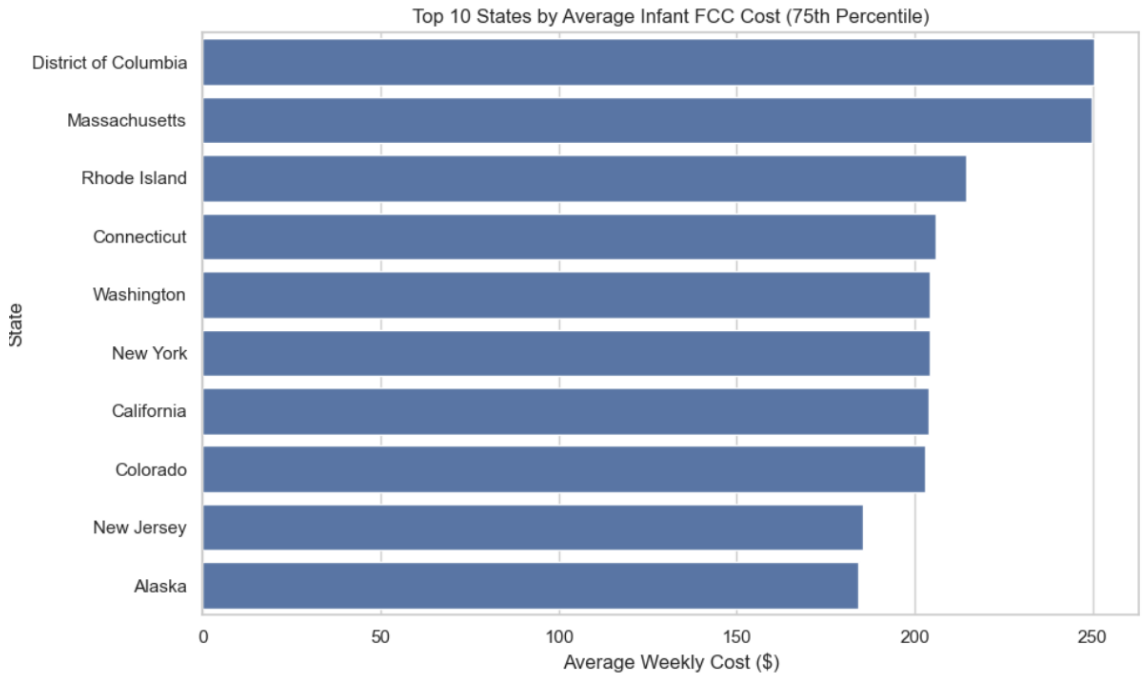
This heatmap shows a weak positive correlation between unemployment rate and infant care costs, suggesting that higher unemployment may not consistently lower prices, due to fixed operational costs in childcare.



Box Plot: Infant FCC Costs by State (Top 10 by Record Count) This plot reveals significant cost variability within states like California and Texas, highlighting regional or county-level differences in care pricing.



Bar Chart: Top 10 States by Average Infant Care Cost (75th Percentile) The top states by average infant care cost include Massachusetts, California, and New York, pointing to urban and coastal states as costliest for families.



Direction of Story / Plan of Attack/Message You Want to Get Across:

The overarching story this data tells is one of ***the rising childcare costs that are increasingly pricing out working families, especially those with infants.*** *The core message is clear:* ***affordable childcare is becoming progressively inaccessible****, with significant implications for workforce participation, particularly for women.*

Our plan for attack is to highlight the urgent need for:

- ***Public policy awareness and intervention*** *to address the escalating costs.*
- ***Regional cost comparisons*** *to pinpoint areas facing the highest financial burden.*
- ***Advocacy for subsidized childcare support*** *to alleviate the strain on household budgets.*

Target Audience:

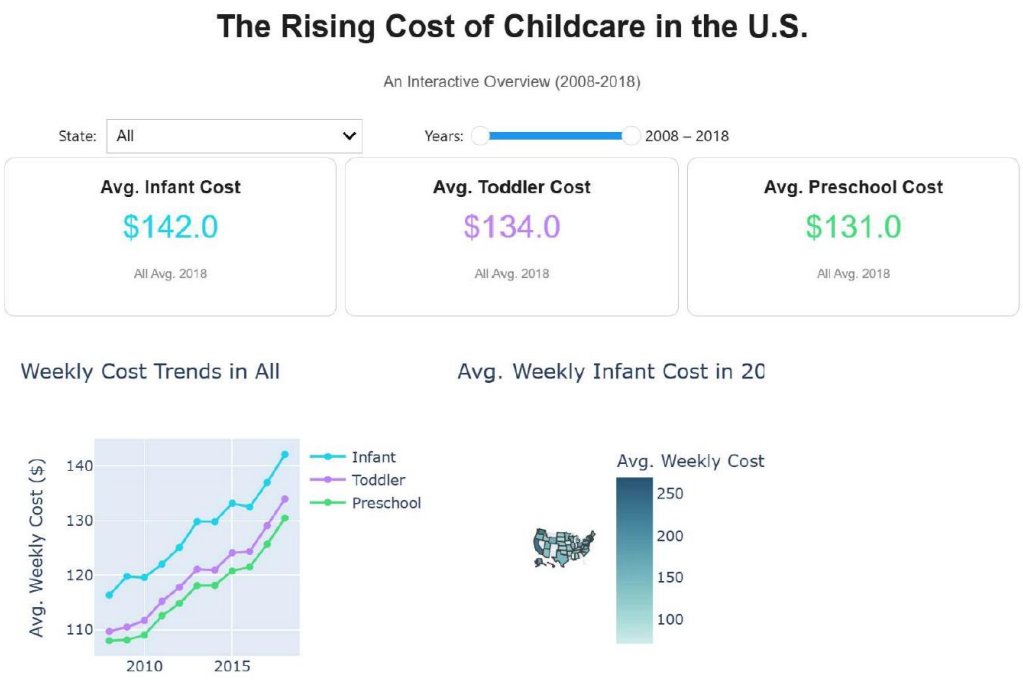
The communication strategy is tailored to reach several key audiences:

- ***Policymakers & Government Officials:*** *To drive data-informed decisions regarding childcare subsidies, tax credits, and workforce policies.*
- ***Nonprofits and Advocacy Groups:*** *To empower their campaigns for childcare equity and support for struggling families.*
- ***Researchers & Economists:*** *To provide a robust data foundation for deeper analysis on the labor market, education, and family planning.*
- ***Parents & Caregivers:*** *To increase transparency around cost expectations and assist in personal financial planning.*

Mediums for Communication:

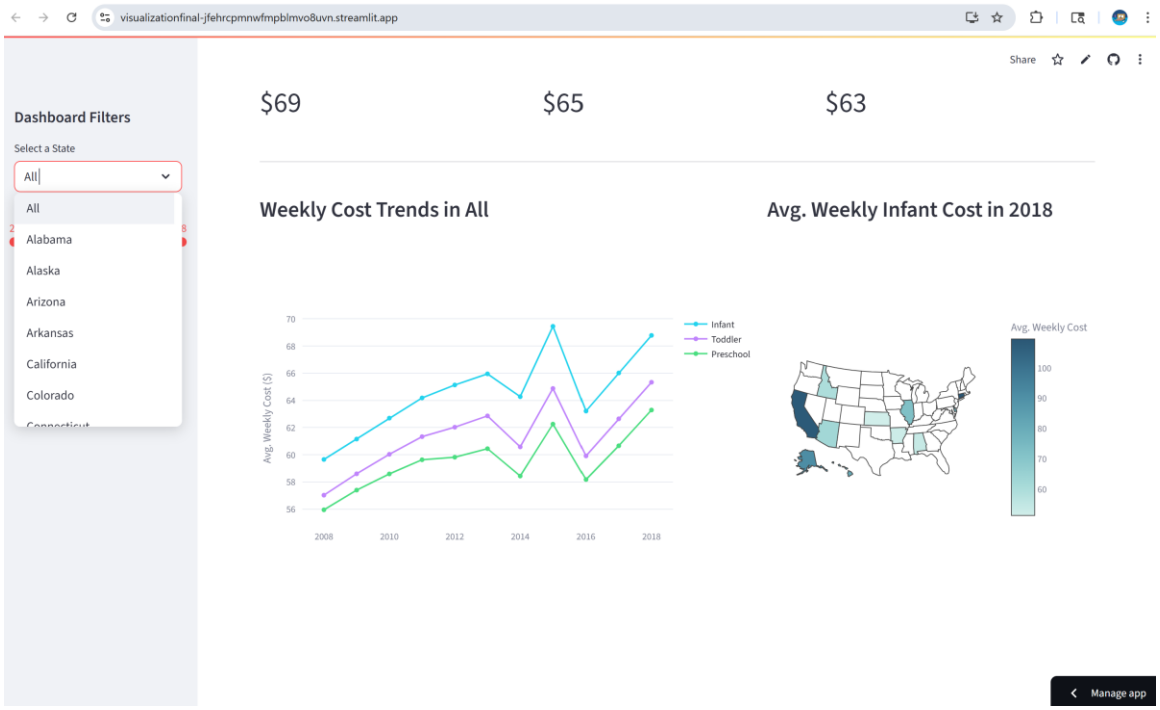
Three distinct mediums were chosen to effectively communicate the findings to diverse audiences:

1. ***Interactive Web Dashboard (Python/Jupyter with Plotly Dash):*** *This medium is designed for data-savvy audiences (policymakers, researchers, journalists) who need to engage deeply with the data. It allows for self-directed exploration, filtering by state, year, and age group, providing nuanced insights beyond national averages. Its interactivity transforms the dataset into a dynamic resource for evidence-based decision-making.*

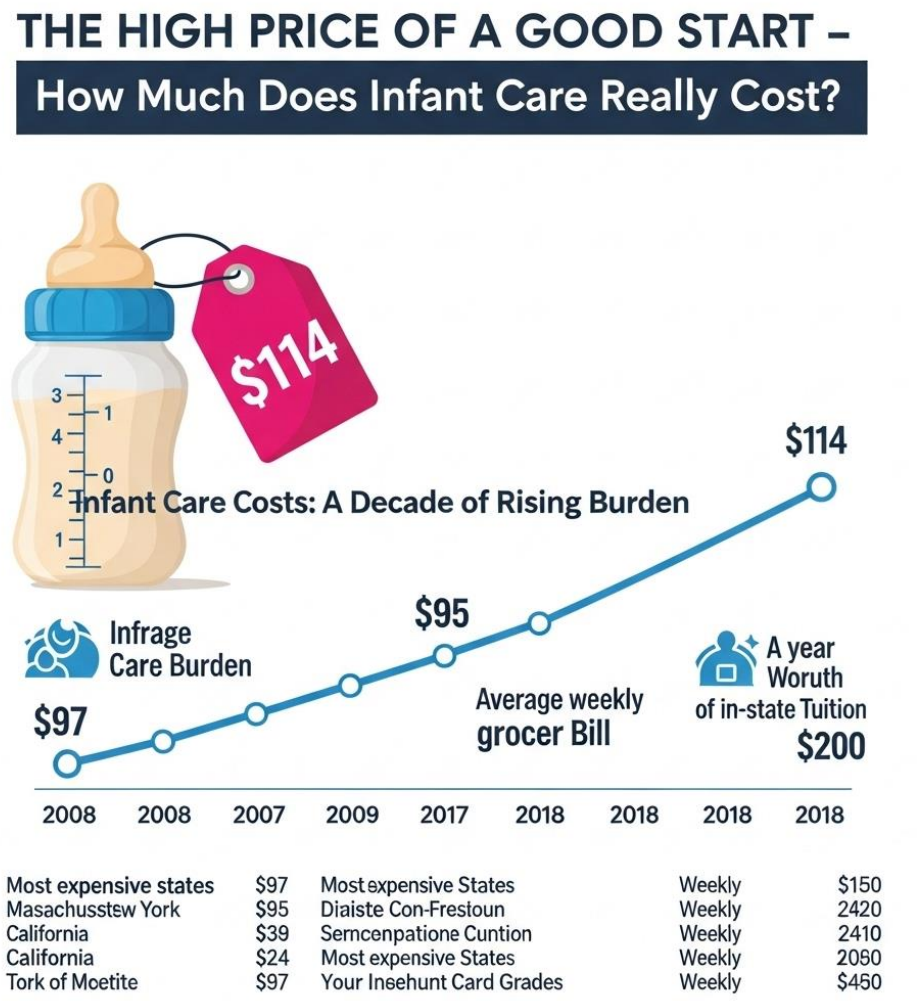


This medium is hosted on streamlit applications as well. Please find the link along with the screenshot below :

https://visualizationfinal-jfehrcpmnwfmpblmvo8uvn.streamlit.app/



2. ***Social Media Infographic:*** *Tailored for the public, parents, and social advocacy groups on fast-paced platforms like Instagram and LinkedIn. The goal is instant understandability and emotional resonance. It distills complex data into a single, highly shareable image, focusing on the most impactful numbers (e.g., infant care costs) to raise widespread awareness and prompt conversation.*



3. ***Executive Presentation Slide Deck (PDF):*** *Crafted for a formal, time-constrained audience of executives and policymakers. Its purpose is persuasion through concise, high-impact data storytelling. Each slide delivers one key message supported by clean, unambiguous visuals, framing childcare as an economic issue warranting intervention. It concludes with actionable, high-level recommendations.*

Childcare_Costs_Pres
entation_Sharm640.pı

Design Decisions

Across all mediums, design decisions were made to enhance clarity, credibility, and persuasion:

- ***Color & Theme:*** *Professional and conservative palettes (deep blue, gray, white) were used for formal presentations, with a single accent color to highlight key data. For the dashboard, a dark theme with distinct, consistent colors for age groups (cyan for infants, purple for toddler, green for preschool) was chosen to reduce eye strain and make visualizations pop. The infographic used high-contrast, simple colors to ensure readability on small screens, with stark red for inflated costs to evoke urgency.*
- ***Visual Choices:*** *The principle of "one key idea per slide" guided the executive presentation, using tailored visuals (line charts for trends, bar charts for comparisons, icons for recommendations). The dashboard featured a line chart for trends and a choropleth map for geographic context, complemented by prominent KPI cards. The infographic relied on a single, central icon (baby bottle with price tag) as a universal symbol for immediate topic communication, with typography as the main visual element.*
- ***Sizing & Spacing:*** *Large, clear headlines communicated main takeaways. Visuals occupied most of the space with minimal text. Abundant white space was used to create a polished, professional feel and draw attention to key messages. The infographic vertical layout was optimized for mobile scrolling with generous white space to prevent clutter.*

Ethical Considerations:

Addressing ethical considerations was paramount throughout the project:

- ***What changes were made to the data?***

  The data from the National Database of Childcare Prices was loaded from an Excel file. Key columns related to state name, abbreviation, county name, study year, and the 75th percentile weekly costs for infant, toddler, and preschool care were selected. The data was then melted to transform the cost columns into a single 'weekly_cost' column with an associated 'metric' (later extracted as 'age_group'). Missing values in 'weekly_cost' or 'age_group' were dropped. The 'StudyYear' was converted to an

integer 'year', and 'age_group' was capitalized for consistency. Column names were standardized (e.g., 'State_Name' to 'state_name'). No other significant transformations or aggregations that would alter the fundamental meaning of the data were performed beyond what was necessary for visualization and analysis.

- ***Are there any legal or regulatory guidelines for your data or project?***

  The dataset used, the National Database of Childcare Prices, is public and ethically sourced. As the data reflects aggregated county/state-level metrics, there are no direct violations of privacy regulations such as HIPAA (Health Insurance Portability and Accountability Act) or FERPA (Family Educational Rights and Privacy Act), which pertain to individual-level health and educational records, respectively. The project adheres to general data ethics principles for public data use.

- ***What risks could be created based on the transformations or how the visualizations are presented?***

  - ***Misinterpretation of Correlation:*** *While a weak positive correlation between unemployment and infant care costs was noted, presenting this without careful explanation could lead to misinterpretation, implying a causal link where none is strongly established. The risk is that audiences might oversimplify complex economic relationships.*

  - ***Overgeneralization of Averages:*** *Focusing on national or state averages, while useful, risks obscuring significant variations at the county or local level. This could lead to a perception that the problem is uniform, when in reality, some communities face much higher burdens.*

  - ***Emotional Manipulation (Infographic):*** *While the infographic aims for emotional resonance, using stark colors (e.g., red for high costs) could be perceived as manipulative if not balanced with factual presentation. The risk is that the emotional appeal overshadows the data's objectivity.*

  - ***Cherry-picking Data:*** *By focusing specifically on the 75th percentile of FCC costs and infant care, there is a risk of being accused of cherry-picking data to support a particular narrative, even if these choices were justified by the project's scope.*

- ***Did you make any assumptions in cleaning/transforming or when presenting the data? Did you filter any data without labeling or clearly identifying that the data was not included?***

  Yes, assumptions were made as outlined in the "Assumptions" section regarding data accuracy and representativeness. In cleaning, rows with missing 'weekly_cost' or

'age_group' were dropped. When presenting, particularly in the executive slide deck, the line chart for national trends focused only on infant data to make a focused point, explicitly stating that "All other age groups are removed to focus the message." Similarly, the bar chart compared infant care costs to tuition in a "high-cost state" (Massachusetts) as an example, rather than an exhaustive list, which was clearly indicated. The infographic also focused solely on infant care costs for simplicity. These filters were intentional for clarity and narrative focus, and efforts were made to contextualize them.

- ***How was your data sourced/verified for credibility?***

  The data was sourced from the National Database of Childcare Prices (NDCP). This database is a credible and publicly available resource, often compiled by government agencies or reputable research institutions. Its public nature and comprehensive scope lend it significant credibility for this type of analysis.

- ***Was your data acquired in an ethical way?***

  Yes, the data was acquired ethically. As it is a publicly available dataset, there were no privacy concerns or issues related to unauthorized access or collection.

- ***How would you mitigate any of the ethical implications you have identified?***

  - ***For Misinterpretation of Correlation:*** *Provide clear disclaimers and explanations that correlation does not imply causation. In the dashboard, tooltips or accompanying text could elaborate on the complexities.*

  - ***For Overgeneralization of Averages:*** *Emphasize the interactive dashboard's ability to filter by state and county, encouraging users to explore local data. In presentations, explicitly mention the variability and use examples of high-cost states to illustrate the extreme burden.*

  - ***For Emotional Manipulation:*** *Ensure that emotional appeals in the infographic are always grounded in accurate data and are used to draw attention to information, not to distort it. Maintain a professional tone in formal mediums.*

  - ***For Cherry-picking Data:*** *Clearly articulate the scope of the analysis (e.g., "focusing on Family Child Care (FCC) weekly costs at the 75th percentile for Infants, Toddlers, and Preschoolers"). Justify the focus on infant care due to its status as the most expensive category.*

Lessons Learned

- ***What would you do differently next time?***

  Next time, I would aim to integrate external economic data (e.g., inflation rates, median household income) directly into the analysis from the outset. This would allow for a more robust "real cost" analysis and provide deeper insights into the affordability aspect, rather than just nominal cost increases. Additionally, exploring more advanced statistical modeling to identify specific drivers of cost increases (beyond simple correlations) would be beneficial. I would also consider incorporating qualitative data, such as anecdotal evidence from parents or childcare providers, to add a human element to the quantitative findings.

- ***What did you enjoy the most?***

  The most enjoyable aspect of this project was the process of transforming raw data into compelling visual narratives across different mediums. Seeing the interactive dashboard come to life and designing the infographic and executive slides to convey a powerful message, was incredibly rewarding. It was particularly satisfying to realize how data visualization can effectively communicate complex societal issues and potentially inform policy discussions. The iterative process of refining the visualizations and the story to best suit each target audience was also a highly engaging and educational experience.

Appendix – Code for Hosting Dashboard on Streamlit