**Milestone 4: Childcare Cost Analysis for Policymakers Report**

Data Science: Bellevue University

Milestone 1 Report Modified

**1. Summary of Analysis**

This analysis focused on understanding childcare costs across the United States, drawing from the **National Database of Childcare Prices** dataset. The primary goals were to identify regional disparities in childcare costs, analyze trends over time, and break down costs by age group (infants, toddlers, preschoolers). Visualizations, including bar charts, pie charts, and line graphs, were used to present the key findings in an accessible format, targeting both policymakers and families.

The dataset was explored for trends that can inform decision-making, particularly for regions where childcare costs are prohibitively high, and for understanding how costs differ depending on the age of the child. This analysis aimed to provide actionable insights for stakeholders, including families, childcare providers, and policymakers.

**2. Changes from Milestone 2**

Since **Milestone 2**, there have been strategic changes in both the **target audiences** and the **mediums** used:

* **Target Audience Changes:**
  + Initially, the plan was to address **policymakers**, **parents/guardians**, and **childcare providers** as the main audiences. Based on feedback and further refinement, the audience was restructured to focus on:
    1. **Policymakers** as the primary audience, who need detailed data-driven insights to guide decisions on childcare policies.
    2. **Data Analysts & Technical Professionals** as a new audience, with an interactive approach to understanding the data analysis through code.
* **Mediums Changes:**
  + The **Interactive Dashboard** initially planned for data analysts was replaced with a **written report** tailored to policymakers, providing a comprehensive overview of the findings and recommendations.
  + A **Jupyter Exploratory Notebook** was introduced for the technical audience, allowing a more structured and engaging walkthrough of the data analysis. The interactive dashboard was combined with the introduction of the Jupyter Notebook for more of the technical audience.
  + The **Infographic** remains unchanged but was refined to focus on clear, digestible visual elements that highlight key insights for parents.

These changes were made to ensure that the analysis is communicated effectively, with each audience receiving information in the most suitable format.

**3. Findings**

* **Regional Disparities:** There are significant regional disparities in childcare costs, with the northeastern states, particularly **Massachusetts**, the **District of Columbia**, and **Connecticut**, being the most expensive. In contrast, southern and midwestern states like **Mississippi** and **Kansas** have the lowest costs.
* **Age Group Breakdown:** Infant care is consistently the most expensive, accounting for **35%** of overall childcare costs. This is followed by toddlers at **32.9%** and preschoolers at **32.1%**.
* **Trends Over Time:** Childcare costs have generally risen over the last decade, especially in high-demand states like California, where costs increased by **28%** over the ten-year period analyzed (2008–2018).

**4. Assumptions**

* **Cost Proportions by Age Group:** It is assumed that the higher costs for infant care are largely due to the specialized care and lower child-to-caregiver ratios required.
* **Trend Extrapolation:** The assumption was made that the trends observed from 2008 to 2018 would likely continue unless significant policy or economic changes occur.

**5. Items that Still Need Clarification**

* **Regional Economic Factors:** A deeper analysis into the specific economic drivers behind the cost disparities in each region is necessary to understand the root causes of high childcare costs in the Northeast.
* **Policy Impact:** There is limited data on how state-level childcare subsidies and programs influence the cost trends observed, and further analysis is required to measure their impact.

**6. Direction of Story/Plan of Attack/Message**

The message of this analysis is that **childcare costs** are rising significantly, and the burden on families is unevenly distributed across regions and age groups. The analysis provides a clear call to action for **policymakers** to address these disparities by implementing subsidies, standardizing childcare pricing, and expanding access to affordable childcare services. The **plan** is to use the data to push for regional solutions, targeted assistance for families, and policies that ensure equitable access to childcare.

**7. Target Audience**

The primary audience is **policymakers** who need data-driven insights to formulate childcare policies, particularly in states where costs are unsustainable for families. Secondary audiences include **data analysts** seeking technical insights from the analysis. **Parents/guardians** were also considered, but now the focus is on presenting information that policymakers can use to drive decisions benefiting families nationwide.

**8. Mediums Included & Why**

* **Written Report:** Provides in-depth analysis tailored for policymakers, focusing on detailed data and policy recommendations. This replaced the interactive dashboard, which was determined to be less effective for a non-technical policy audience.
* **Jupyter Notebook:** Introduced for technical professionals, this medium includes detailed code and visualizations and allows for further exploration of trends and patterns in the data, presenting the findings in a structured manner.
* **Infographic (Canva):** Designed for parents/guardians, providing a simple, visual summary of key findings (most and least expensive states, cost breakdown by age group).

**9. Design Decisions**

The design choices across all mediums were made to balance clarity and accessibility. The **color choices** (sky blue, light green, red) ensure clear differentiation between data points. **Simple bar charts and pie charts** were selected for parents and policymakers, while more complex visualizations (e.g., line charts for trends over time) were used for the presentation aimed at technical professionals. The **interactive elements** in the Jupyter Notebook allow data analysts to explore the data further, providing flexibility in analysis.

**10. Ethical Considerations**

* **Data Transparency:** No data was filtered or excluded without clearly labeling or explaining it. All steps in cleaning and analyzing the data were documented.
* **Data Privacy:** The dataset used did not include any personally identifiable information, thus ensuring compliance with privacy regulations.
* **Assumptions and Fairness:** Assumptions regarding costs by age group and regional trends were clearly identified, and no conclusions were drawn without sufficient data support.

**What Changes Were Made to the Data?**

* The data was cleaned by removing duplicates and handling missing values. No significant transformations were made that would alter the nature of the dataset.

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

* The project adhered to ethical guidelines regarding data use. The dataset, being publicly available, had no restrictions related to privacy or data protection laws.

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

* If regional disparities are exaggerated or misunderstood, it could lead to misinformed policy decisions. However, care was taken to ensure an accurate representation of the data.

**Did you make any assumptions in cleaning/transforming or when presenting the data?**

* The only assumptions made were related to cost proportions across age groups. No data was filtered without explanation.

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

* The **National Database of Childcare Prices** is a credible source, widely recognized and used for policy and research purposes. It was cross-verified with secondary data sources to ensure consistency.

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

* Yes, the data was publicly available and ethically sourced without violating any privacy or regulatory concerns.

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

* By maintaining transparency in how the data is cleaned and presented and by clearly documenting any assumptions or gaps in the analysis, we mitigate the risk of misrepresentation or bias.

**11. Lessons Learned**

* **What would you do differently next time?**
  + Next time, I would explore additional datasets (e.g., demographic or economic indicators) to analyze the factors driving childcare costs comprehensively.
  + I would also consider using more dynamic visualizations, such as interactive dashboards, for a more engaging presentation of the data.
* **What did you enjoy the most?**
  + The most enjoyable aspect was creating visualizations that clearly communicated complex data in an accessible way. Seeing how bar charts and pie charts made the insights more understandable was a highlight of the project.