**Final Analysis Report on Childcare Price Dataset**

**Bobga-Herman Gwanvoma**  
Bellevue University  
DSC 640 – Data Presentation & Visualization  
Professor Benjamin Schneider  
February 28, 2025

Summary of Analysis

The National Database of Childcare Prices dataset provides a comprehensive view of childcare costs across the United States at the county level, spanning multiple years. The dataset consists of 227 columns, covering variables such as geographic identifiers, demographic indicators, and childcare cost metrics for different age groups. An exploratory data analysis was conducted to examine trends, gaps, and inconsistencies within the dataset. The findings reveal significant variations in childcare costs by region, with urban areas and care for younger children showing higher costs. Additionally, demographic factors such as unemployment rates were examined to assess their correlation with childcare affordability.

Findings

Key insights from the analysis include variability in childcare costs, which differ significantly by state, county, and year, with urban areas and younger age groups incurring higher expenses. Inflationary trends indicate year-over-year increases in childcare costs, suggesting an inflationary effect impacting affordability. Demographic correlations show that higher unemployment rates often align with lower childcare costs, possibly due to economic constraints or regional subsidies. Policy and economic implications highlight how childcare costs can influence workforce participation, particularly among parents, and reveal disparities in economic accessibility.

Assumptions

Flag columns, such as \_flag, denote data validation or compliance with predefined thresholds. Missing values in demographic indicators are treated as data gaps rather than zero values. Variability in childcare costs is assumed to reflect actual differences unless flagged as anomalies or outliers.

Items That Still Need Clarification

Several aspects require further clarification, including the definitions and significance of flag values in \_flag columns, the methodology behind the childcare cost calculations (such as averages, medians, or weighted values), and the causes and potential impact of missing data on analytical outcomes.

Direction of Story/Plan of Attack/Message

The narrative will emphasize the economic and demographic factors driving childcare cost variability across the United States. The primary focus includes regional disparities in childcare affordability, the impact of childcare costs on family financial stability and workforce participation, and policy recommendations for addressing inequities in childcare pricing.

Target Audience

The primary audience includes policymakers, economists, and advocacy organizations focused on childcare affordability and workforce development. The secondary audience consists of researchers and journalists investigating economic inequalities and demographic trends.

Mediums Included & Why

An interactive dashboard, such as Tableau or Power BI, enables users to explore childcare costs by state, county, and year, with demographic filters allowing for deeper analysis. An infographic provides a visually engaging summary highlighting national and regional disparities using charts, maps, and key insights. A written report delivers a detailed PDF with in-depth analysis, methodology, findings, and policy recommendations for a technical audience.

Design Decisions

Interactivity in the dashboard allows users to drill down into specific regions and demographic groups. Simplicity and clarity in the infographic ensure accessibility for a broad audience. Depth and detail in the written report provide comprehensive insights and citations for rigorous analysis.

Ethical Considerations

Several ethical concerns were addressed. Modifications to the dataset included treating missing values as data gaps rather than zero values to avoid misleading conclusions. Legal and regulatory guidelines were followed, ensuring that the dataset does not contain personally identifiable information (PII) and complies with data privacy laws. Potential risks, such as misrepresentation of trends due to missing or incomplete data, could lead to incorrect policy interpretations. Transparency in data processing was maintained by explicitly documenting all transformations and exclusions to avoid bias.

Assumptions in Data Processing & Visualization

Data filtering was clearly labeled in reports and dashboards to ensure clarity. The dataset was sourced from national databases, and credibility was verified through cross-referencing with official government reports. Data collection followed ethical guidelines, ensuring accuracy and fairness. Ethical concerns, such as potential misinterpretation of childcare cost trends, were mitigated by providing clear documentation and annotations in visualizations.

Lessons Learned

Improvements for future analysis include acquiring more detailed metadata on flag columns and cost calculation methodologies to enhance analysis accuracy. A different approach in the future would involve earlier engagement with domain experts to clarify dataset inconsistencies. The most enjoyable aspect of this project was the process of uncovering trends and patterns in childcare costs and visualizing these insights for diverse audiences.