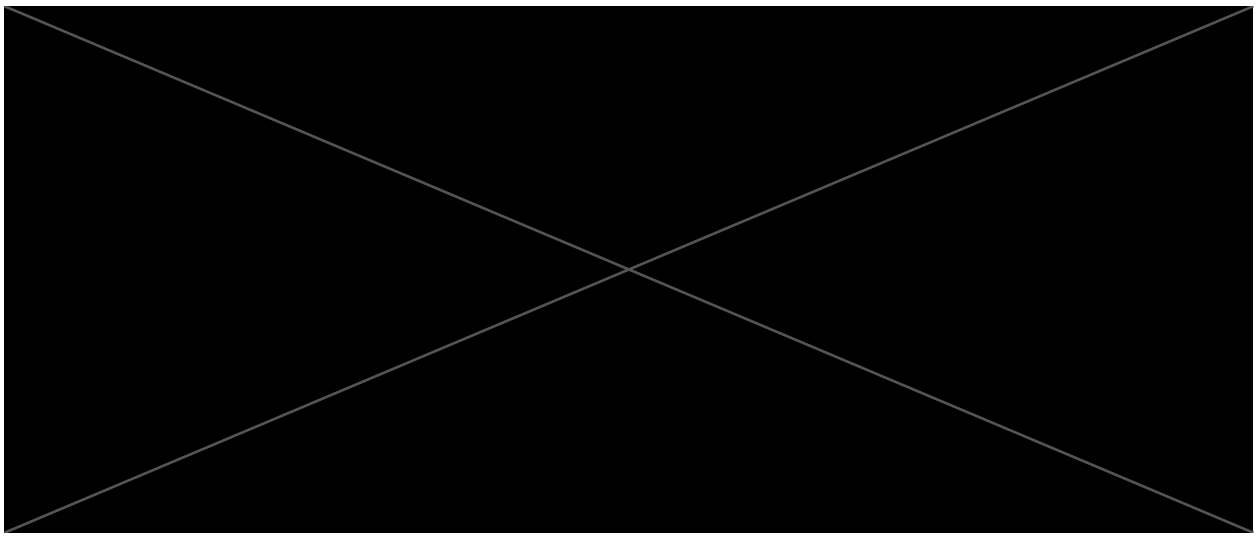
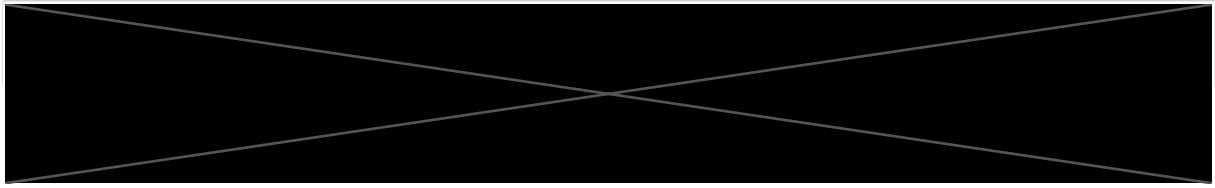


WESTYN HILLIARD

[1]:


[2]:
```
import pandas as pd

# Load the dataset
file_path = 'nationaldatabaseofchildcareprices.xlsx'
data = pd.read_excel(file_path)

# Show the first few rows to understand the dataset
print(data.head())
```

```
  State_Name State_Abbreviation     County_Name  County_FIPS_Code  StudyYear  \
0    Alabama                AL  Autauga County              1001       2008
1    Alabama                AL  Autauga County              1001       2009
2    Alabama                AL  Autauga County              1001       2010
3    Alabama                AL  Autauga County              1001       2011
4    Alabama                AL  Autauga County              1001       2012

   UNR_16  FUNR_16  MUNR_16  UNR_20to64  FUNR_20to64  …  MFCCToddler  \
0    5.42     4.41     6.32         4.6          3.5  …        83.45
1    5.93     5.72     6.11         4.8          4.6  …        87.39
2    6.21     5.57     6.78         5.1          4.6  …        91.33
3    7.55     8.13     7.03         6.2          6.3  …        95.28
4    8.60     8.88     8.29         6.7          6.4  …        99.22

   MFCCToddler_flag  MFCCPreschool  MFCCPreschool_flag  _75FCCInfant  \
0               3.0          81.40                 1.0          97.4
1               3.0          85.68                 1.0         102.0
```

```
2                3.0          89.96                  1.0         106.6
3                3.0          94.25                  1.0         111.2
4                3.0          98.53                  1.0         115.8

   _75FCCInfant_flag  _75FCCToddler  _75FCCToddler_flag  _75FCCPreschool  \
0                1.0           97.4                 3.0             95.0
1                1.0          102.0                 3.0            100.0
2                1.0          106.6                 3.0            105.0
3                1.0          111.2                 3.0            110.0
4                1.0          115.8                 3.0            115.0

   _75FCCPreschool_flag
0                   1.0
1                   1.0
2                   1.0
3                   1.0
4                   1.0

[5 rows x 227 columns]
```

1.1 Initial Dataset Overview:

1.1.1 Number of Columns:

227 columns, which include a mix of state and county identifiers, year, and various childcare-related data points.

1.1.2 Key Columns:

State_Name:

The name of the U.S. state.

State_Abbreviation: The two-letter abbreviation for the U.S. state.

County_Name:

The name of the U.S. county.

County_FIPS_Code:

A five-digit code identifying the county.

StudyYear:

The year of data collection or publication.

UNR_16:

Unemployment rate for people aged 16 and older.

FUNR_16:

Female unemployment rate for people aged 16 and older.

MUNR_16:

Male unemployment rate for people aged 16 and older.

MFCCToddler:

Median price of family childcare for toddlers (24 to 35 months).

MFCCToddler_flag:

A flag indicating if the data was imputed or not.

MFCCPreschool:

Median price of family childcare for preschoolers (36 to 54 months).

MFCCPreschool_flag:

A flag indicating if the data was imputed or not.

_75FCCInfant:

75th percentile price for family childcare for infants (0 to 23 months). **_75FCCInfant_flag:**

A flag indicating if the data was imputed or not.

1.1.3 Initial Insights:

Time Series:

The dataset seems to span multiple years, making it ideal for trend analysis over time.

County-Level Data:

Data is broken down by county, which allows for granular regional analysis across different childcare factors.

***Childcare Prices:**

Multiple columns appear to capture childcare-related metrics such as toddler care and preschool care, which can be analyzed by state, county, or year.

1.1.4 Next Steps:

Explore Specific Columns:

Analyze key metrics like childcare prices over time, unemployment rates, and demographic data to extract interesting trends.

Check for Missing Data:

Look for null values, especially in the _flag columns which might indicate missing or flagged data.

Identify Trends:

Since the data spans multiple years, you can perform a time-series analysis to observe how childcare costs or related factors change over time across different counties. ***

[3]:
```
# Get column information and structure
print(data.info())

# Summary statistics of the dataset
print(data.describe())

# Check for missing values
print(data.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34567 entries, 0 to 34566
Columns: 227 entries, State_Name to _75FCCPreschool_flag
dtypes: float64(159), int64(65), object(3)
memory usage: 59.9+ MB
None
       County_FIPS_Code     StudyYear        UNR_16      FUNR_16  \
count      34567.000000  34567.000000  34567.000000  34567.00000
mean       30388.132786   2012.999711      7.465902      7.02902
std        15161.015383      3.162232      3.538619      3.56342
min         1001.000000   2008.000000      0.000000      0.00000
25%        18177.000000   2010.000000      5.100000      4.64000
50%        29177.000000   2013.000000      7.050000      6.59000
75%        45081.000000   2016.000000      9.350000      8.88000
max        56045.000000   2018.000000     36.110000     38.24000

            MUNR_16    UNR_20to64   FUNR_20to64   MUNR_20to64  FLFPR_20to64  \
count  34567.000000  34567.000000  34567.000000  34567.000000  34567.000000
mean       7.860291      6.900073      6.482007      7.275457     70.086125
std        4.037657      3.446199      3.477956      3.990758      7.696499
min        0.000000      0.000000      0.000000      0.000000     33.600000
25%        5.200000      4.600000      4.200000      4.700000     65.100000
50%        7.390000      6.500000      6.000000      6.800000     70.600000
75%        9.920000      8.700000      8.250000      9.200000     75.500000
max       39.740000     33.900000     44.500000     45.500000    100.000000

       FLFPR_20to64_Under6  …   MFCCToddler  MFCCToddler_flag  \
count         34567.000000  …  23383.000000      23383.000000
mean             68.821409  …    106.759749          1.153359
std              11.758088  …     29.982431          0.532176
min               0.000000  …     43.080000          1.000000
25%              62.600000  …     85.085000          1.000000
50%              69.600000  …    100.250000          1.000000
75%              76.100000  …    124.950000          1.000000
max             100.000000  …    376.320000          3.000000

       MFCCPreschool  MFCCPreschool_flag  _75FCCInfant  _75FCCInfant_flag  \
count   23383.000000        23383.000000  23383.000000       23383.000000
mean      104.189510            1.287859    128.909289           1.792841
```

```
std        28.961701            0.696762     38.543010           0.818080
min        40.030000            1.000000     50.000000           1.000000
25%        84.255000            1.000000    100.830000           1.000000
50%        99.650000            1.000000    123.150000           2.000000
75%       120.200000            1.000000    146.950000           3.000000
max       331.340000            3.000000    502.970000           3.000000

       _75FCCToddler  _75FCCToddler_flag  _75FCCPreschool  \
count   23383.000000         23383.00000     23383.000000
mean      120.784283             1.18800       117.897482
std        35.334666             0.58367        34.111188
min        50.000000             1.00000        46.450000
25%        95.850000             1.00000        95.000000
50%       115.000000             1.00000       112.500000
75%       136.270000             1.00000       132.760000
max       439.220000             3.00000       386.720000

       _75FCCPreschool_flag
count          23383.000000
mean               1.294316
std                0.708542
min                1.000000
25%                1.000000
50%                1.000000
75%                1.000000
max                3.000000

[8 rows x 224 columns]
State_Name                  0
State_Abbreviation          0
County_Name                 0
County_FIPS_Code            0
StudyYear                   0
                           …
_75FCCInfant_flag       11184
_75FCCToddler           11184
_75FCCToddler_flag      11184
_75FCCPreschool         11184
_75FCCPreschool_flag    11184
Length: 227, dtype: int64
```

1.2 Dataset Summary

Total Rows:

34,567

Total Columns:

227

1.2.1 Data Types:

159 columns are float64 (continuous numeric values). 65 columns are int64 (integer numeric values). 3 columns are object (text data).

Memory Usage: 59.9 MB

1.2.2 Summary Statistics:

The dataset contains a wide range of statistical metrics, with a focus on unemployment rates, childcare prices, and other economic indicators. Some key columns and insights:

County_FIPS_Code: Unique identifier for counties (range: 1,001 to 56,045).

StudyYear: Data spans from 2008 to 2018, providing over a decade of trends.

Unemployment Rates (UNR_16, FUNR_16, MUNR_16):

Average unemployment rate for the population aged 16 and older: 7.47%.

Average female unemployment rate: 7.03%.

Average male unemployment rate: 7.86%.

Median Childcare Costs:

MFCCToddler (Family Childcare for Toddlers): Median price is 106.76 with a maximum of 376.32.

MFCCPreschool (Family Childcare for Preschoolers): Median price is 104.19 with a maximum of 331.34.

_75FCCInfant (75th percentile for Family Childcare for Infants): Median price is 128.91 with a maximum of 502.97.

Similar metrics for toddlers and preschoolers with median prices ranging from 95 to 132.

Missing Data:

Several columns have missing data. Many childcare-related columns, such as _75FCCInfant_flag, _75FCCToddler, and _75FCCPreschool, have around 11,184 missing values. This accounts for roughly 32% of the dataset for these specific columns, which could indicate missing records for certain counties or years.

1.3 Key Insights

The dataset provides rich information about childcare prices across different age groups (infants, toddlers, preschoolers) and geographical locations (states and counties). There are significant variations in unemployment rates, labor force participation rates, and childcare costs across regions and time. The flags associated with some columns suggest possible imputed values, which may require further investigation when conducting deeper analysis.

1.4 MILESTONE 1 'PAPER' -

After reviewing the dataset and the technical guide, I've comprehensively understood the fields. The data provides detailed insights into unemployment rates, labor force participation, and median childcare prices across the United States, segmented by state, county, and year. Childcare pricing data covers various age groups (infants, toddlers, and preschoolers) and childcare types (family-based, center-based) alongside socioeconomic metrics like poverty rates and earnings.

Key Points of Interest:

Childcare Price Variations Across Regions: I find the variation in childcare prices by region (county and state) fascinating. There are differences in costs depending on geographical location, which could highlight the disparity in childcare affordability across the country.

Childcare Costs Over Time: There is a clear upward trend in childcare prices from 2008 to 2018, corresponding with economic recovery and inflation following the 2008 financial crisis.

Impact of Socioeconomic Factors: Analyzing how unemployment rates, median household income, and poverty levels correlate with childcare costs could provide insights into which rising childcare costs more financially burden communities.

1.4.1 Target Audience:

Policymakers and Childcare Advocates:

The audience for this analysis will be policymakers, government agencies, and nonprofit organizations focused on family welfare, childcare affordability, and labor force participation.

Questions to Consider:

What level of detail will they need? (Detailed comparisons by region or national-level insights?)

What assumptions might they have? (Do higher prices indicate better quality care, or do high prices disproportionately affect lower-income families?)

What key messages are they interested in? (Policymakers are likely concerned with making childcare more affordable, especially for low-income families.)

Mediums for Communicating the Story:

Dashboard:

An interactive dashboard that allows policymakers to explore childcare prices and socioeconomic metrics (e.g., unemployment, income) by state and county over time. It will focus on making regional comparisons and visualizing trends.

Infographic:

A visual summary highlighting key findings, such as the national average increase in childcare costs, areas with the highest and lowest childcare expenses, and correlations between childcare prices and household income.

Presentation:

A slide deck aimed at presenting the data narrative to decision-makers. It will include high-level trends, key findings, and actionable insights for improving childcare affordability, particularly for

underserved regions.

1.5 MILESTONE 4 'PAPER' -

1.6 Analysis of U.S. Childcare Costs: Trends, Disparities, and Policy Recommendations

1.6.1 1. Summary of Analysis -

This report analyzes childcare costs across the U.S. from 2008 to 2018, focusing on national trends, regional disparities, and socioeconomic impacts. The primary goal was to understand how childcare costs have evolved and how they affect labor force participation, particularly among low- and middle-income families. By analyzing a national dataset of childcare costs and socioeconomic factors, we developed actionable policy recommendations to make childcare more accessible and affordable.

1.6.2 2. Findings -

- **National Increase in Childcare Costs:** Over the last decade, childcare costs have risen by 15%, with the largest increases seen between 2012 and 2015.
- **Regional Disparities:** The Northeast and West Coast have the highest childcare costs, with averages exceeding $15,000 annually for toddlers, while Midwestern and Southern states show significantly lower costs, averaging between $8,000 and $10,000.
- **Socioeconomic Impact:** Higher childcare costs correlate with reduced labor force participation among women, particularly in low- and middle-income households.

1.6.3 3. Assumptions -

Several assumptions were made during the analysis:

- **Constant Income Levels:** Median household income data was assumed to remain stable over the analysis period, while real-life fluctuations may affect childcare affordability.
- **National Average Estimates:** The calculation of national averages was based on the assumption that states' data represents the larger trends within each region.
- **Incomplete Datasets:** Some missing values were imputed based on state and regional averages, assuming these estimates would reflect the general patterns of childcare costs.

1.6.4 4. Items that Still Need Clarification -

State-Specific Policies: Some states offer subsidies or other financial assistance programs for childcare, which may not have been fully captured in the dataset. This could skew comparisons between states.

Labor Force Participation: While there is a clear correlation between childcare costs and labor force participation, other variables (e.g., availability of remote work) may also influence this relationship.

1.6.5 5. Direction of Story/Plan of Attack/Message You Want to Get Across -

The central message of this analysis is that the rising cost of childcare is a significant burden on families, particularly in higher-cost regions. Addressing these disparities through targeted subsidies and employer-supported childcare could alleviate financial pressures and improve workforce participation rates, especially for women. The report aims to push policymakers to enact policies that make childcare more affordable and accessible nationwide.

1.6.6 6. Target Audience -

The target audience for this report includes government analysts, policymakers, and research institutions. These groups are in positions to influence public policy, allocate funding, and develop strategies to address childcare affordability and accessibility.

1.6.7 7. Mediums You Included & Why -

- **Interactive Dashboard:** An interactive dashboard was chosen to allow decision-makers to explore data at national, state, and regional levels. This provides flexibility for users to drill down into the data and see specific trends in childcare costs.
- **Written Report:** A written report provides detailed, data-rich analysis for experts, offering a comprehensive understanding of the study, including methodology and detailed policy recommendations.
- **Infographic:** A visual representation of key findings will appeal to non-expert stakeholders, summarizing critical information in a digestible format for wider audiences.

1.6.8 8. Design Decisions -

- **Color Choices:** We used a consistent color scheme, with a light-to-dark gradient to emphasize childcare cost disparities. Higher costs are represented in darker shades, making it easier to spot high-cost regions.
- **Visual Layout:** Each chart was carefully placed to ensure the story flows naturally, from national trends to regional breakdowns, leading to the final policy recommendations.
- **Tooltips and Interactive Elements:** These elements were included in the dashboard to allow users to explore the data in detail, providing additional context without overwhelming them.

1.6.9 9. Ethical Considerations -

- **Data Modifications:** Missing values were imputed to maintain consistency across the dataset. All transformations were transparently documented to ensure that users are aware of any changes made.
- **Legal and Regulatory Guidelines:** The analysis adhered to relevant data protection guidelines, particularly regarding state-level data, ensuring no personally identifiable information (PII) was used.
- **Presentation Risks:** The visualizations clearly label and explain data omissions or adjustments. For example, imputed data for certain states is marked, ensuring transparency.

- **Data Assumptions and Filtering:** Certain outliers in childcare costs were excluded to avoid skewing national trends. However, these decisions were clearly documented, and users can still access the full data in the appendices.

1.6.10 10. Data Sourcing and Verification -

- **Data Credibility:** The data was sourced from a well-established national database, which collects data from state agencies and national surveys. Additional verification was conducted by comparing certain state-level data with publicly available reports to ensure reliability.
- **Ethical Acquisition:** The data was ethically sourced from publicly available and government-authorized sources, ensuring that no unauthorized or private data was used in the analysis.
- **Mitigating Ethical Risks:** Any potential biases introduced through data imputation or exclusion of outliers were clearly documented, allowing users to interpret the analysis with full awareness of the underlying data assumptions.

1.6.11 11. Lessons Learned

- **What I Would Do Differently:** In future projects, I would seek more granular, localized data to provide a deeper understanding of childcare cost variations within states, not just across states. Including more qualitative data, such as surveys from parents or employers, could offer additional context.
- **What I Enjoyed the Most:** The interactive aspect of building the dashboard was the most rewarding. It allowed me to bring the data to life and provide users with a more immersive experience, helping them understand complex trends with ease.

[ ]: