OwusuSefah640JupyterMedium

November 10, 2024

Term Project

Bernard Owusu Sefah

National Database of Childcare Prices

DSC 640

0.1 Final Medium: Jupter Presententaion For Data Analysts and Technical Professionals

Introduction

Childcare Cost Analysis: A Data-Driven Approach   This notebook provides an analysis of the **National Database of Childcare Prices** dataset. The goal is to explore trends in childcare costs, identify regional disparities, and analyze cost differences across age groups such as infants, toddlers, and preschoolers.

Feel free to interact with the notebook by running the code cells and modifying the parameters to suit your exploration needs.

Data Import and Cleaning

```
[1]: # Importing necessary libraries
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns

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

     # Clean the data: remove duplicates and handle missing values
     childcare_data = childcare_data.drop_duplicates().dropna()

     # Display the first few rows of the dataset
     childcare_data.head()
```

```
[1]:   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]
```

Exploratory Data Analysis

Regional Disparities in Childcare Costs (Bar Chart)

```python
[2]: # Average childcare cost by state for preschoolers
     avg_cost_by_state = childcare_data.groupby('State_Name')['MFCCPreschool'].
      ↪mean().sort_values(ascending=False)

     # Plot: Average childcare cost by state (preschool)
     plt.figure(figsize=(12, 6))
     avg_cost_by_state.plot(kind='bar', color='skyblue')
     plt.title('Average Childcare Cost by State (Preschool)')
     plt.xlabel('State')
     plt.ylabel('Average Cost')
     plt.xticks(rotation=90)
```

```
plt.tight_layout()
plt.show()
```

Average Childcare Cost by State (Preschool)
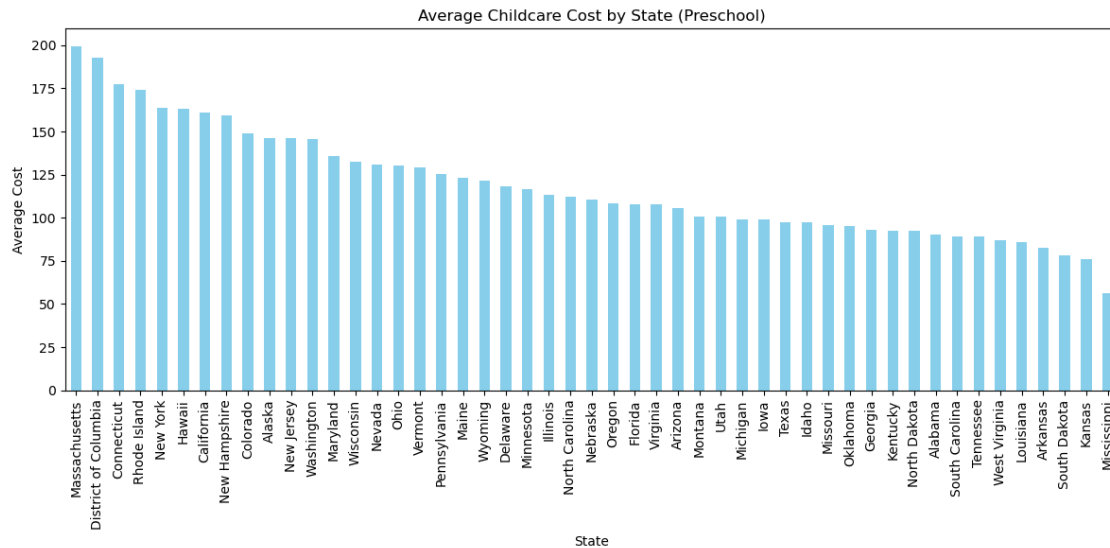
Description: This bar chart represents the average childcare cost for preschoolers across different states. The chart is sorted in descending order, showing the states with the highest costs on the left and the states with the lowest costs on the right. Insights: * Massachusetts and the District of Columbia have the highest average childcare costs for preschoolers, both nearing $200.

- Mississippi, Kansas, and South Dakota have the lowest costs, falling below $75.

- This visualization highlights the stark regional disparities in childcare costs, with a clear gap between the most and least expensive states.

Trend Analysis Over Time (Line Chart)

```
[3]: # Trend analysis: Childcare costs over time for preschoolers
     cost_trends = childcare_data.groupby(['StudyYear',␣
     ↪'State_Name'])['MFCCPreschool'].mean().unstack()

     # Plot: Childcare cost trends over time
     cost_trends.plot(figsize=(10, 6))
     plt.title('Childcare Cost Trends Over Time (Preschool)')
     plt.xlabel('Year')
     plt.ylabel('Average Cost')
     plt.grid(True)
     plt.tight_layout()
     plt.show()
```

C:\Users\berna\AppData\Local\Temp\ipykernel_30348\753182436.py:10: UserWarning:
Tight layout not applied. The bottom and top margins cannot be made large enough

```
to accommodate all axes decorations.
  plt.tight_layout()
```


Childcare Cost Trends Over Time (Preschool)
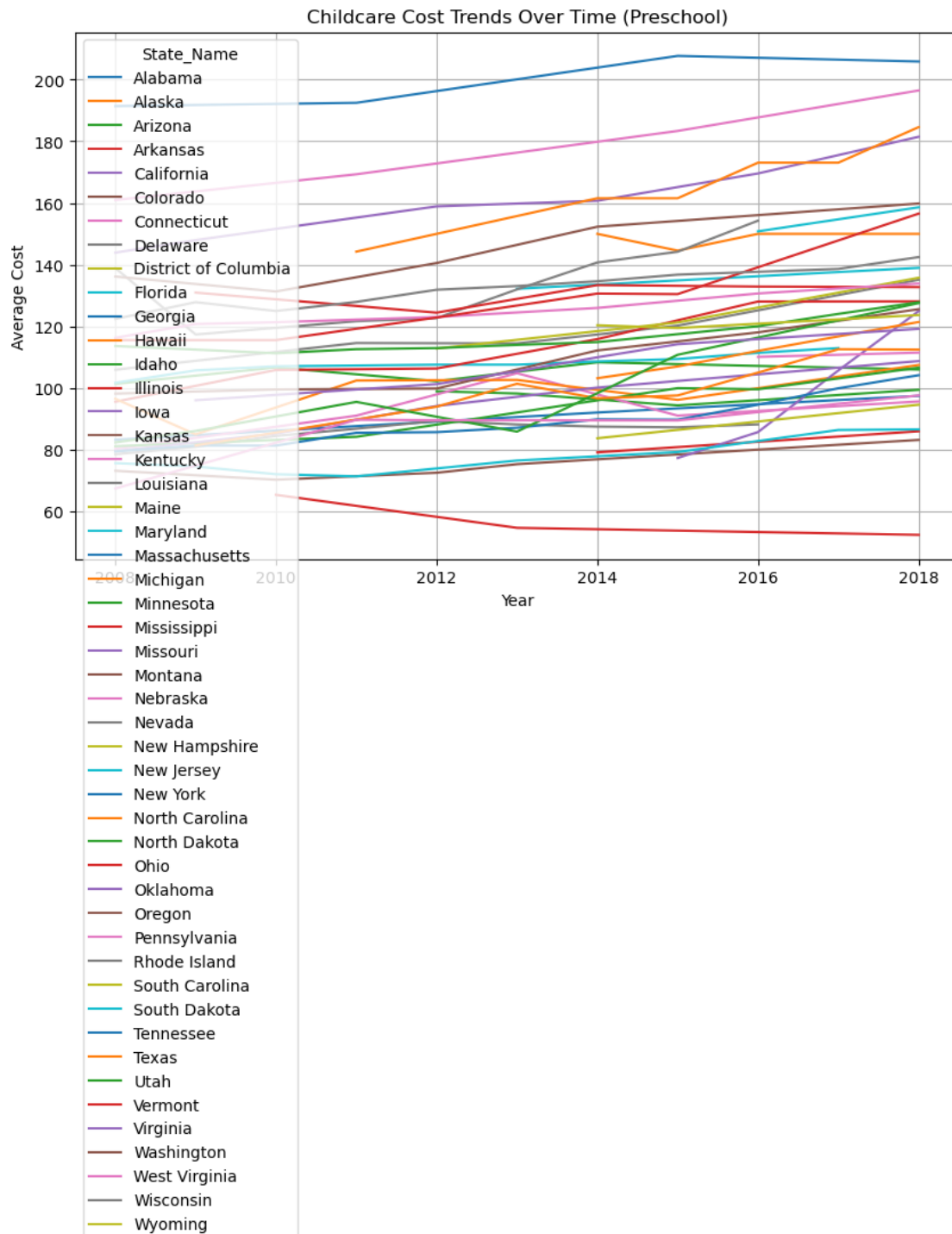
Description: This line chart illustrates how childcare costs for preschoolers have evolved over time

(from around 2008 to 2018) across various states. Each colored line represents a different state, allowing for a comparison of trends. Insights: * Most states show a gradual increase in childcare costs over time.

- Some states, such as Alaska and Massachusetts, exhibit sharper upward trends, indicating a more rapid rise in childcare costs.

- States like Kansas and Kentucky remain relatively flat, showing minimal cost increases during the same period.

- This visualization is useful for identifying states where childcare costs have increased significantly over time, providing context for regional policy discussions or economic analysis.

Cost Breakdown by Age Group (Pie Chart)

[4]:
```python
# Average costs for infants, toddlers, and preschoolers
age_group_columns = ['MFCCInfant', 'MFCCToddler', 'MFCCPreschool']
cost_by_age_group = childcare_data[age_group_columns].mean()

# Plot: Distribution of childcare costs by age group
plt.figure(figsize=(6, 6))
cost_by_age_group.plot(kind='pie', autopct='%1.1f%%', colors=['lightblue',
 ↪'lightgreen', 'coral'])
plt.title('Childcare Costs Distribution by Age Group')
plt.ylabel('')
plt.tight_layout()
plt.show()
```

Childcare Costs Distribution by Age Group
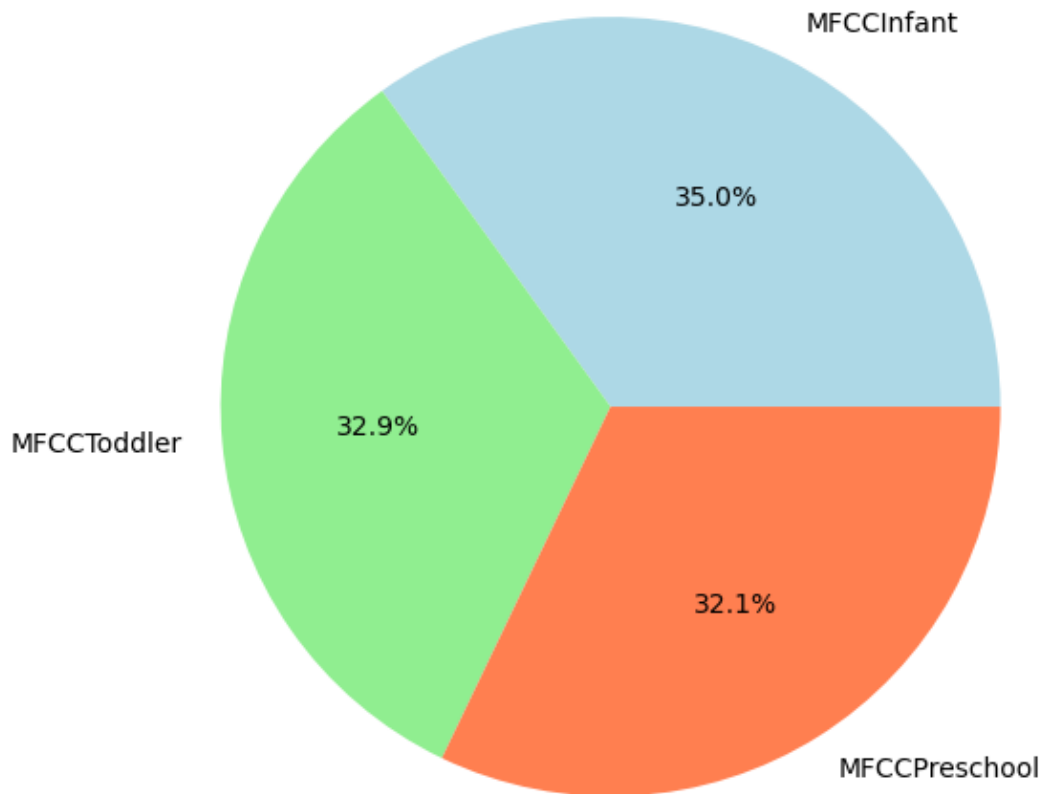
Description: This pie chart shows the distribution of childcare costs across different age groups: infants, toddlers, and preschoolers. Each segment of the pie represents the proportion of total childcare costs attributed to each age group. Insights: * Infant care represents the largest portion of childcare costs at 35%, followed by toddler care at 32.9%, and preschool care at 32.1%.

- This indicates that, on average, infant care tends to be more expensive than toddler and preschool care, reflecting the higher demands of caring for infants.

- Providers and policymakers may focus on reducing infant care costs as they constitute the largest financial burden on families.

0.2 Interactive Exploration

This section allows the user to interactively explore childcare costs for a specific state or multiple states

```python
[15]: # Import necessary libraries
      import matplotlib.pyplot as plt

      # Interactive exploration: Filter by multiple states and visualize cost trends
      states = input("Enter states separated by commas (e.g., 'California, Texas,␣
       ↪Florida'): ")
      states_list = [state.strip() for state in states.split(',')]  # Convert input␣
       ↪to a list of states

      # Filter data based on the input states
      filtered_data = childcare_data[childcare_data['State_Name'].isin(states_list)]

      # Plot childcare costs over time for the selected states
      filtered_cost_trends = filtered_data.groupby(['StudyYear',␣
       ↪'State_Name'])['MFCCPreschool'].mean().unstack()

      # Plot
      filtered_cost_trends.plot(figsize=(10, 6))
      plt.title(f'Childcare Costs Over Time for Selected States (Preschool)')
      plt.xlabel('Year')
      plt.ylabel('Average Cost')
      plt.grid(True)
      plt.tight_layout()
      plt.show()
```
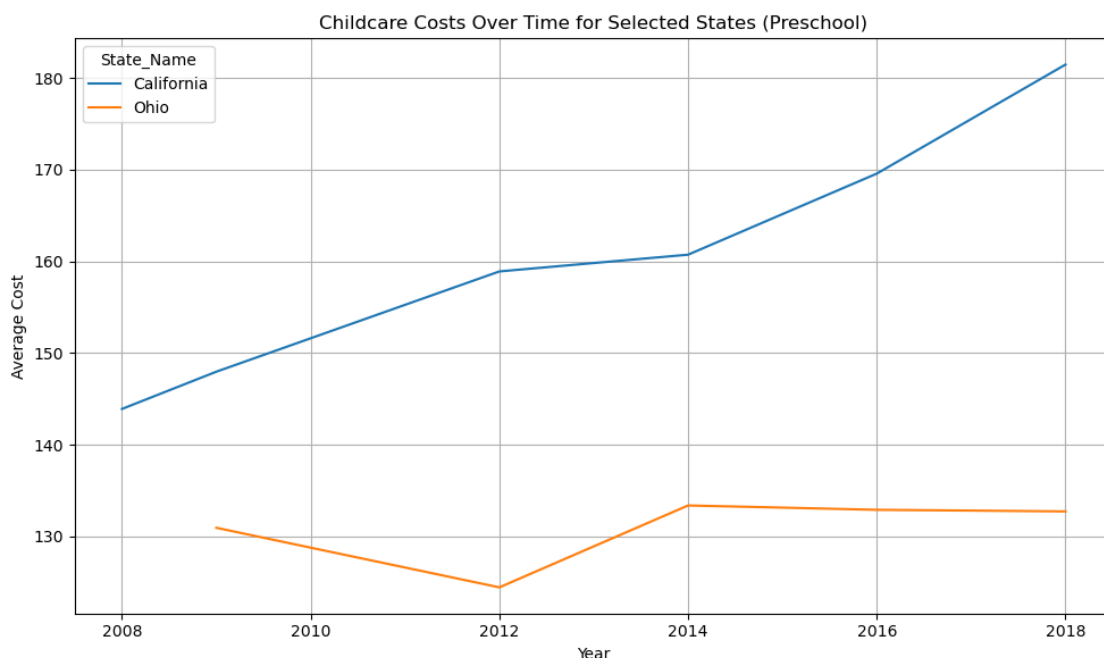
Enter states separated by commas (e.g., 'California, Texas, Florida'):
California, Ohio

Childcare Costs Over Time for Selected States (Preschool)

California (Blue Line):

- California shows a consistent upward trend in childcare costs over the 10-year period.

- In 2008, the average childcare cost for preschoolers in California was around $140.

- By 2018, this cost increased significantly to around $180, indicating a rise of approximately 28% over the decade.

Ohio (Orange Line):

- Ohio displays a more stable trend, with a slight decline in childcare costs from 2008 to 2012, followed by a small increase around 2014.

- After 2014, the costs in Ohio seem to level off, remaining just below $130 in 2018.

- This suggests that while California has experienced a substantial increase in childcare costs, Ohio's costs have remained relatively flat.

Key Takeaways: * California has seen a much sharper increase in childcare costs compared to Ohio, making it significantly more expensive for preschool childcare in 2018.

- Ohio's relatively stable trend suggests less financial strain from rising childcare costs compared to states like California, where the continuous rise could pose challenges for families.

Conclusion   This analysis provides insights into regional disparities and trends in childcare costs across the U.S., focusing on differences across age groups. We observed significant variations between states, with certain regions having notably higher costs for preschoolers. The data also reveals that infant care tends to be the most expensive across the board.

The notebook encourages further exploration, allowing users to filter by state and year to delve deeper into the data.

Next steps for exploration: - Investigate how specific policy changes or regional programs have impacted childcare costs over time. - Explore correlations between childcare costs and other demographic or economic factors.

Narrative for Design Decisions   In designing this Jupyter Notebook, the primary goal was to create an interactive, accessible analysis tool tailored to data analysts and technical professionals. The notebook uses simple but clear color schemes (e.g., sky blue, light green) to ensure that the visual elements are easy to interpret. Sizing and spacing are optimized for readability, especially in the bar and pie charts. The interactive components (e.g., state-specific cost exploration) allow users to engage with the data directly, providing flexibility in the analysis.

The notebook is designed to allow users to explore various aspects of childcare costs, from regional disparities to age group breakdowns and trends over time. By focusing on the National Database of Childcare Prices, the analysis stays relevant to current trends in the childcare industry

[ ]: