

Child Care

December 8, 2024

```
[12]: import pandas as pd

# Define the correct file path
file_path = '/Users/cheribeda/Desktop/nationaldatabaseofchildcareprices (1).
↪xlsx'

# Load the Excel file into a Pandas DataFrame
try:
    data = pd.read_excel(file_path)
    # Display the first few rows of the dataset
    print("File loaded successfully!")
    print(data.head())
except FileNotFoundError:
    print("File not found. Please check the file path.")
except Exception as e:
    print(f"An error occurred: {e}")
```

File loaded successfully!

	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]

```
[14]: # Get the shape of the dataset
print("Dataset shape:", data.shape)

# Display all column names
print("Column names:")
print(data.columns)

# Check for missing values
print("Missing values per column:")
print(data.isnull().sum())
```

Dataset shape: (34567, 227)

Column names:

```
Index(['State_Name', 'State_Abbreviation', 'County_Name', 'County_FIPS_Code',
      'StudyYear', 'UNR_16', 'FUNR_16', 'MUNR_16', 'UNR_20to64',
      'FUNR_20to64',
      ...,
      'MFCCToddler', 'MFCCToddler_flag', 'MFCCPreschool',
      'MFCCPreschool_flag', '_75FCCInfant', '_75FCCInfant_flag',
      '_75FCCToddler', '_75FCCToddler_flag', '_75FCCPreschool',
      '_75FCCPreschool_flag'],
      dtype='object', length=227)
```

Missing values per column:

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
```

```
[16]: # Define relevant columns
columns_of_interest = [
    'State_Name', 'County_Name', 'StudyYear',
    '_75FCCInfant', '_75FCCToddler', '_75FCCPreschool',
    '_75FCCInfant_flag', '_75FCCToddler_flag', '_75FCCPreschool_flag'
]

# Create a subset
subset_data = data[columns_of_interest]

# Preview the subset
print(subset_data.head())
```

	State_Name	County_Name	StudyYear	_75FCCInfant	_75FCCToddler	\
0	Alabama	Autauga County	2008	97.4	97.4	
1	Alabama	Autauga County	2009	102.0	102.0	
2	Alabama	Autauga County	2010	106.6	106.6	
3	Alabama	Autauga County	2011	111.2	111.2	
4	Alabama	Autauga County	2012	115.8	115.8	

	_75FCCPreschool	_75FCCInfant_flag	_75FCCToddler_flag	\
0	95.0	1.0	3.0	
1	100.0	1.0	3.0	
2	105.0	1.0	3.0	
3	110.0	1.0	3.0	
4	115.0	1.0	3.0	

	_75FCCPreschool_flag
0	1.0
1	1.0
2	1.0
3	1.0
4	1.0

```
[18]: subset_data = subset_data.dropna()
```

```
[21]: # Find the highest and lowest infant care costs
highest_infant_cost = subset_data.loc[subset_data['_75FCCInfant'].idxmax()]
lowest_infant_cost = subset_data.loc[subset_data['_75FCCInfant'].idxmin()]

# Find the highest and lowest toddler care costs
highest_toddler_cost = subset_data.loc[subset_data['_75FCCToddler'].idxmax()]
lowest_toddler_cost = subset_data.loc[subset_data['_75FCCToddler'].idxmin()]
```

```

# Find the highest and lowest preschool care costs
highest_preschool_cost = subset_data.loc[subset_data['_75FCCPreschool']].
    idxmax()]
lowest_preschool_cost = subset_data.loc[subset_data['_75FCCPreschool'].idxmin()]

# Display the results
print("Highest Infant Care Cost:")
print(highest_infant_cost[['State_Name', 'County_Name', 'StudyYear',
    ↪'_75FCCInfant']])
print("\nLowest Infant Care Cost:")
print(lowest_infant_cost[['State_Name', 'County_Name', 'StudyYear',
    ↪'_75FCCInfant']])

print("\nHighest Toddler Care Cost:")
print(highest_toddler_cost[['State_Name', 'County_Name', 'StudyYear',
    ↪'_75FCCToddler']])
print("\nLowest Toddler Care Cost:")
print(lowest_toddler_cost[['State_Name', 'County_Name', 'StudyYear',
    ↪'_75FCCToddler']])

print("\nHighest Preschool Care Cost:")
print(highest_preschool_cost[['State_Name', 'County_Name', 'StudyYear',
    ↪'_75FCCPreschool']])
print("\nLowest Preschool Care Cost:")
print(lowest_preschool_cost[['State_Name', 'County_Name', 'StudyYear',
    ↪'_75FCCPreschool']])

```

Highest Infant Care Cost:

State_Name	Wyoming
County_Name	Teton County
StudyYear	2018
_75FCCInfant	502.97

Name: 34533, dtype: object

Lowest Infant Care Cost:

State_Name	South Dakota
County_Name	Bennett County
StudyYear	2008
_75FCCInfant	50.0

Name: 26003, dtype: object

Highest Toddler Care Cost:

State_Name	Wyoming
County_Name	Teton County
StudyYear	2018
_75FCCToddler	439.22

Name: 34533, dtype: object

Lowest Toddler Care Cost:

State_Name	South Dakota
County_Name	Bennett County
StudyYear	2008
_75FCCToddler	50.0

Name: 26003, dtype: object

Highest Preschool Care Cost:

State_Name	Wyoming
County_Name	Teton County
StudyYear	2018
_75FCCPreschool	386.72

Name: 34533, dtype: object

Lowest Preschool Care Cost:

State_Name	Mississippi
County_Name	Carroll County
StudyYear	2018
_75FCCPreschool	46.45

Name: 15497, dtype: object

[]: