Licensed Child Care Providers and Facilities

- **INTRODUCTION:**

In Data Science Environment, Time is of the essence. We need Good result in a shorter span of time. I made Data collection, Data Exploration, Feature Engineering, Data Visualization Here.

As I started Off with the dataset, I looked first and foremost number of instances. Having a dataset Of 1005 Rows and 22 Columns in the Dataset. This Dataset gave me an idea on how to proceed. I made a point to look for. This Dataset Contain Family Child Care Providers, Large Family Child Care Providers and Child care centers that are licenses By the state Of Delaware Concerns regarding a child care provide and facilities .

1. Null Values
2. Look at categorical Columns
3. Look at Numerical columns
4. Explore the data set
5. Feature Engineering
6. Data Visualization

- **Data Dictionary for Dataset: -**

1. License Number: - The license Number issued by the state of Delaware
2. County: - The county where the childcare is located

3. Name: - The name of the childcare provider or facility.
4. Name Reserved: - Last Name, First Name when it is an individual provider.
5. Type of Child Care: - Family Child Care, Large Family Child Care, Child Care Center.
6. Enforcement Action: - Current Enforcement Status
7. Intent to revoke: - The office of child care licensing has begun the process to revoke the license of the child care provider.
8. Street Address: - Street Address of the Child Care.
9. City: - City Of the childcare
10. State: - State of the childcare
11. Zip Code: - Zip Code of the Child Care
12. Phone Number: - phone number of the childcare
13. Age Range: - Ages of children served by this childcare.
14. Age Group: - Age Group of Children served by this child care facility.
15. Capacity: - The maximum number of children that can be served by this childcare at one time.
16. Opening Time: - opening time of the childcare.
17. Closing Time: - Closing Time Of the childcare.
18. Special Conditions: - Special Conditions for this childcare
19. Government Programs: - Government Programs that the childcare provider participates in childcare programs.
20. Delaware Stars Level: - State of Delaware ‘s Voluntary Quality rating and improvement System.
21. Geocoded Location: - Street Address, City, State and Zip Code Combined For geolocating Records on a map.

22. Count: - This is Used Strictly For providing Counts and Summaries Of the data and is Otherwise not relevant to this specific row in this dataset. The value will always be “1”.

- **Import the Required Libraries For the Project**

1.Import the Required libraries

In [3]:
```
# Python ≥3.5 is required
import sys
assert sys.version_info >= (3, 5)

# Scikit-Learn ≥0.20 is required
import sklearn
assert sklearn.__version__ >= "0.20"

# To plot pretty figures
%matplotlib inline
import matplotlib as mpl
import matplotlib.pyplot as plt
mpl.rc('axes', labelsize=14)
mpl.rc('xtick', labelsize=12)
mpl.rc('ytick', labelsize=12)

# Where to save the figures
import os
PROJECT_ROOT_DIR = "."
CHAPTER_ID = "Childcare_project"
IMAGES_PATH = os.path.join(PROJECT_ROOT_DIR, "images", CHAPTER_ID)
os.makedirs(IMAGES_PATH, exist_ok=True)

def save_fig(fig_id, tight_layout=True, fig_extension="png", resolution=300):
    path = os.path.join(IMAGES_PATH, fig_id + "." + fig_extension)
    print("Saving figure", fig_id)
    if tight_layout:
        plt.tight_layout()
    plt.savefig(path, format=fig_extension, dpi=resolution)

# Ignore useless warnings (see SciPy issue #5998)
import warnings
warnings.filterwarnings(action="ignore", message="^internal gelsd")
```

- **Import CSV DATA SET USING PANDAS**

Import CSV Data set Using Pandas

In [4]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

In [5]:
```
import numpy as np
```

In [7]:
```
import pandas as pd
url = '/content/Licensed_Child_Care_Providers_and_Facilities.csv'
Data1 = pd.read_csv(url)
Data1.head(5)
```

Out[7]:

	License Number	County	Name	Name Reversed	Type of Child Care	Enforcement Action	Intent to Revoke	Street Address	City	State	Zip Code	Phone Number	Age Range	Age Group	Ca
0	27390	Sussex	LISA STOECKEL	STOECKEL, LISA	Licensed Family Child Care	NaN	NaN	22448 PETERKINS ROAD	GEORGETOWN	DE	19947	(302) 856-2651	6 months through 5 years.	Infant through Pre-School	9
1	27399	New Castle	YMCA OF DELAWARE / CENTRAL BRANCH YMCA CHILDRE...	YMCA OF DELAWARE / CENTRAL BRANCH YMCA CHILDRE...	Licensed Child Care Center	NaN	NaN	501 WEST 11TH STREET	WILMINGTON	DE	19801	(302) 254-9622	6 weeks through 5 years.	Infant through Pre-School	9
2	27407	Kent	DOVER EDUCATIONAL & COMMUNITY CENTER	DOVER EDUCATIONAL & COMMUNITY CENTER	Licensed Child Care Center	Warning of Probation	Intent to Place on Warning of Probation	744 RIVER ROAD	DOVER	DE	19901	(302) 883-3092	1 year through 12 years.	Toddler through School-Age	5
3	27410	New Castle	ELIZABETH JOHNSON	JOHNSON, ELIZABETH	Licensed Family Child Care	NaN	NaN	2011 WEST STREET	WILMINGTON	DE	19802	(302) 287-5733	6 weeks through 12 years.	Infant through School-Age	9
4	27411	New Castle	JANEL DEMONIA	DEMONIA, JANEL	Licensed Family Child Care	NaN	NaN	1113 CADE STREET	WILMINGTON	DE	19802	(302) 656-1344	6 weeks through 12 years.	Infant through School-Age	9

- **Let's explore the data a little bit by checking the number of rows and columns in our datasets.**

Let's explore the data a little bit by checking the number of rows and columns in our datasets.

In [8]:
```
Data1.shape
```

Out[8]: (1005, 22)

- **Let's Check here Information of the dataset explaining which datatype.**

Let's Check here Information of the dataset explaining which datatype.

In [9]:
```
Data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1005 entries, 0 to 1004
Data columns (total 22 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   License Number        1005 non-null   int64 
 1   County                1005 non-null   object
 2   Name                  1005 non-null   object
 3   Name Reversed         1005 non-null   object
 4   Type of Child Care    1005 non-null   object
 5   Enforcement Action    10 non-null     object
 6   Intent to Revoke      24 non-null     object
 7   Street Address        1005 non-null   object
 8   City                  1005 non-null   object
 9   State                 1005 non-null   object
 10  Zip Code              1005 non-null   int64 
 11  Phone Number          1005 non-null   object
 12  Age Range             1005 non-null   object
 13  Age Group             1005 non-null   object
 14  Capacity              1005 non-null   int64 
 15  Opening Time          1005 non-null   object
 16  Closing Time          1004 non-null   object
 17  Special Conditions    28 non-null     object
 18  Government Programs   923 non-null    object
 19  Delaware STARS Level  1005 non-null   int64 
 20  Geocoded Location     1003 non-null   object
 21  Count                 1005 non-null   int64 
dtypes: int64(5), object(17)
memory usage: 172.9+ KB
```

- **To see the statistical details of the dataset.**

To see the statistical details of the dataset

In [10]:
```
Data1.describe()
```

Out[10]:

	License Number	Zip Code	Capacity	Delaware STARS Level	Count
count	1.005000e+03	1005.000000	1005.000000	1005.000000	1005.0
mean	5.968706e+05	19824.040796	50.530348	1.537313	1.0
std	4.793050e+05	103.157730	63.682795	2.005434	0.0
min	2.739000e+04	19701.000000	5.000000	0.000000	1.0
25%	1.145190e+05	19713.000000	9.000000	0.000000	1.0
50%	5.285800e+05	19805.000000	12.000000	0.000000	1.0
75%	9.962730e+05	19941.000000	76.000000	3.000000	1.0
max	1.481094e+06	19977.000000	585.000000	5.000000	1.0

- **Let's Check How many Unique Object in the Dataset**

Let's Check How many Object Unique In County

```
In [11]: Data1['County'].unique()
```

```
Out[11]: array(['Sussex', 'New Castle', 'Kent', 'NEW CASTLE', 'SUSSEX', 'KENT'],
               dtype=object)
```

Let's Check How many Object Unique In Type Of Child Care

```
In [12]: Data1['Type of Child Care'].unique()
```

```
Out[12]: array(['Licensed Family Child Care', 'Licensed Child Care Center',
                'Licensed Large Family Child Care'], dtype=object)
```

Let's Check How many Object Unique In City

```
In [13]: Data1['City'].unique()
```

```
Out[13]: array(['GEORGETOWN', 'WILMINGTON', 'DOVER', 'NEW CASTLE', 'MIDDLETOWN',
                'MILTON', 'BEAR', 'MILLSBORO', 'NEWARK', 'HOCKESSIN', 'HARTLY',
                'BETHEL', 'FELTON', 'CAMDEN WYOMING', 'MILFORD', 'FRANKFORD',
                'LINCOLN', 'SEAFORD', 'FREDERICA', 'DELMAR', 'LEWES', 'CLAYMONT',
                'LAUREL', 'SELBYVILLE', 'DELAWARE CITY', 'MAGNOLIA', 'HARBESON',
                'DAGSBORO', 'SMYRNA', 'CLAYTON', 'GREENWOOD', 'BRIDGEVILLE',
                'TOWNSEND', 'HARRINGTON', 'BLADES', 'OCEAN VIEW', 'CHRISTIANA',
                'ELLENDALE', 'REHOBOTH BEACH', 'CAMDEN', 'NEWPORT', 'FARMINGTON',
                'ODESSA', 'LONG NECK'], dtype=object)
```

Let's Check How many Object Unique in Goverment Programs.

```
In [14]: Data1['Government Programs'].unique()
```

```
Out[14]: array(['Child Care Food Program',
                'Child Care Food Program;Nonprofit;Purchase of Care',
                'Purchase of Care', 'Child Care Food Program;Purchase of Care',
                'Private', 'Child Care Food Program;Profit;Purchase of Care', nan,
                'Nonprofit', 'Child Care Food Program;Private;Purchase of Care',
                'Profit;Purchase of Care', 'Child Care Food Program;Private',
                'Nonprofit;Private', 'Nonprofit;Purchase of Care',
                'Private;Purchase of Care',
                'Child Care Food Program;Nonprofit;Private;Purchase of Care',
                'Child Care Food Program;Nonprofit',
                'Child Care Food Program;Private;Profit;Purchase of Care',
                'Private;Profit', 'Publicly Operated;Purchase of Care',
                'Nonprofit;Private;Purchase of Care'], dtype=object)
```

Let's Check How Many Object Unique in Age Group

```
In [15]: Data1['Age Group'].unique()
```

```
Out[15]: array(['Infant through Pre-School', 'Toddler through School-Age',
                'Infant through School-Age', 'Pre-School through School-Age',
                'Pre-School', 'Toddler through Pre-School', 'School-Age',
                'Infant through Other', 'Infant through Toddler',
                'Toddler through Other'], dtype=object)
```

Let's Check How many Object Unique In State.

```
In [16]: Data1['State'].unique()
```

```
Out[16]: array(['DE'], dtype=object)
```

Let's Check How Many Object Unique In Age Range

```
In [17]: Data1['Age Range'].unique()
```

```
Out[17]: array(['6 months through 5 years.', '6 weeks through 5 years.',
                '1 year through 12 years.', '6 weeks through 12 years.',
                '5 years through 12 years.', '3 months through 10 years.',
                '6 weeks through 14 years.', '3 months through 12 years.',
                '3 months through 14 years.', '3 months through Kindergarten.',
                '6 weeks through 10 years.', '3 years through 5 years.',
                '2 years through 4 years.', '6 years through 12 years.',
                '6 weeks through 3 years.', '2 years through 12 years.',
                '6 months through 12 years.', '2 years through 5 years.',
                '6 weeks through 8 years.', '3 years through 10 years.',
                '2 years through 13 years.', '6 weeks through 6 years.',
                '8 weeks through 5 years.', '6 weeks through Kindergarten.',
                '8 weeks through 12 years.', '8 weeks through Kindergarten.',
                '6 weeks through 9 years.', '18 months through 5 years.',
                '1 year through 5 years.', '3 months through 5 years.',
                '6 years through 17 years.', '3 months through .',
                '2 years through 8 years.', 'Kindergarten through 12 years.',
                '1 year through 9 years.', '6 weeks through 4 years.',
                '18 months through 12 years.', 'Kindergarten through 15 years.',
                '1 year through 10 years.', '6 weeks through 13 years.',
                '6 weeks through 7 years.', '3 months through 7 years.',
                '2 years through 6 years.', '3 months through 11 years.',
                '8 weeks through 11 years.', '2 years through 9 years.',
                'Kindergarten through 17 years.', '3 years through 12 years.',
                '18 months through 13 years.', '1 year through 6 years.',
                '1 year through 13 years.', '6 weeks through 11 years.',
                '5 years through 13 years.', '2 years through Kindergarten.',
```

- ## Check The Null Value is in the Dataset

```
In [18]: Data1.isnull()
```

Out[18]:

	License Number	County	Name	Name Reversed	Type of Child Care	Enforcement Action	Intent to Revoke	Street Address	City	State	Zip Code	Phone Number	Age Range	Age Group	Capacity	Opening Time	Closing Time	Spec Cond
0	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
1	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True
3	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
4	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1000	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
1001	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
1002	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
1003	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True
1004	False	False	False	False	False	True	True	False	False	False	False	False	False	False	False	False	False	True

- **Handle Categorical Features Using One Hot Encoding:**

Handle Categorical Features Using One Hot Encoding:

In [19]:
```
import pandas as pd
Data1 = pd.read_csv('/content/Licensed_Child_Care_Providers_and_Facilities.csv',usecols=['County'])
Data1.head(15)
```

Out[19]:

	County
0	Sussex
1	New Castle
2	Kent
3	New Castle
4	New Castle
5	New Castle
6	NEW CASTLE
7	New Castle
8	NEW CASTLE
9	Sussex
10	New Castle
11	Sussex
12	New Castle
13	Kent
14	New Castle

In [20]:
```
pd.get_dummies(Data1).head()
```

Out[20]:

	County_KENT	County_Kent	County_NEW CASTLE	County_New Castle	County_SUSSEX	County_Sussex
0	0	0	0	0	0	1
1	0	0	0	1	0	0
2	0	1	0	0	0	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0

In [21]:
```
pd.get_dummies(Data1,drop_first=True).head()
```

Out[21]:

	County_Kent	County_NEW CASTLE	County_New Castle	County_SUSSEX	County_Sussex
0	0	0	0	0	1
1	0	0	1	0	0
2	1	0	0	0	0
3	0	0	1	0	0
4	0	0	1	0	0

In [22]:
```
Data1['County'].unique()
```

Out[22]:
```
array(['Sussex', 'New Castle', 'Kent', 'NEW CASTLE', 'SUSSEX', 'KENT'],
      dtype=object)
```

```
In [22]: Data1['County'].unique()
Out[22]: array(['Sussex', 'New Castle', 'Kent', 'NEW CASTLE', 'SUSSEX', 'KENT'],
               dtype=object)

In [23]: Data1.dropna(inplace=True)

In [24]: pd.get_dummies(Data1,drop_first=True).head()
```

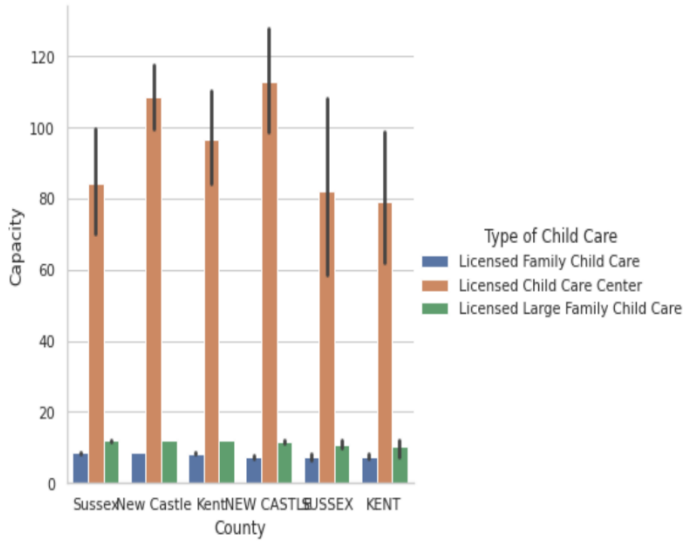
Out[24]:

	County_Kent	County_NEW CASTLE	County_New Castle	County_SUSSEX	County_Sussex
0	0	0	0	0	1
1	0	0	1	0	0
2	1	0	0	0	0
3	0	0	1	0	0
4	0	0	1	0	0

- **<u>Let's Visualize The dataset. Use of Visualization Packages.</u>**

➢ Statistical analysis is a process of understanding how variables in a dataset relate to each other and how those relationships depend on other variables. Visualization can be a core component of this process because, when data are visualized properly, the human visual system can see trends and patterns that indicate a relationship. Visualize the Type of Child Care. County and Capacity depend on each other in this visualization.

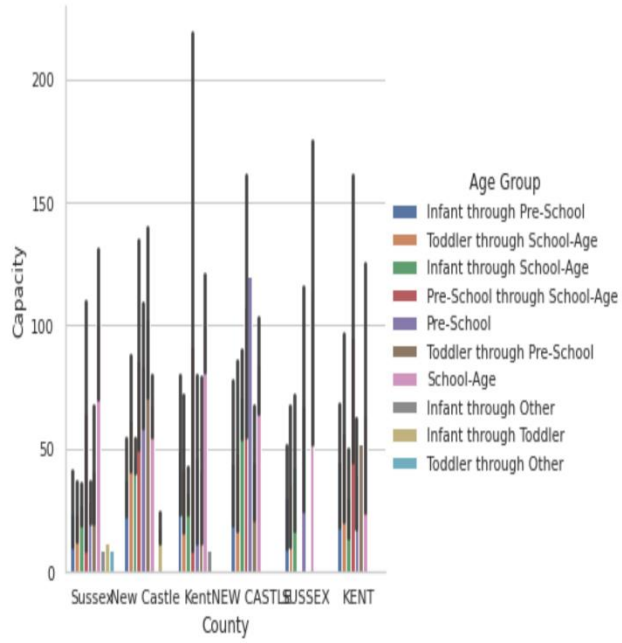
```
In [35]: import seaborn as sns
         import matplotlib.pyplot as plt

In [ ]: sns.catplot(x="County", y="Capacity", hue="Type of Child Care", kind="bar", data=Data1)

Out[ ]: <seaborn.axisgrid.FacetGrid at 0x7f5fef922e50>
```



➢ Visualize the Age Group.

```
In [ ]: sns.catplot(x="County", y="Capacity", hue="Age Group", kind="bar", data=Data1)
```

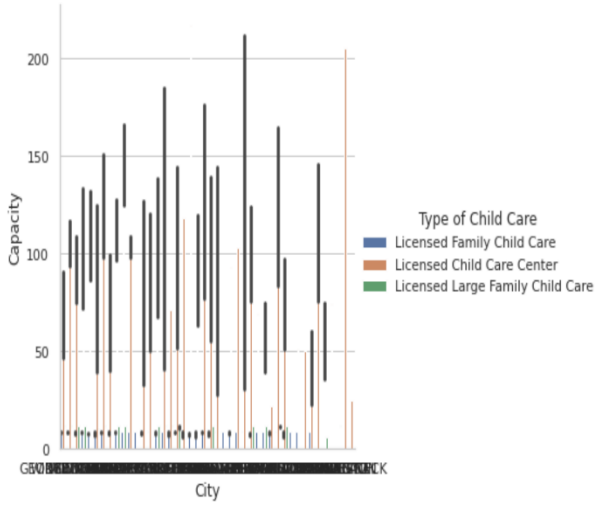
```
Out[ ]: <seaborn.axisgrid.FacetGrid at 0x7f5fef927a90>
```



➢ Visualize the Type Of Child Care.

```
In [ ]: sns.catplot(x="City", y="Capacity", hue="Type of Child Care", kind="bar", data=Data1)
        sns.catplot(x="City", y="Capacity", hue="Age Range", kind="bar", data=Data1)
        sns.catplot(x="City", y="Capacity", hue="Age Group", kind="bar", data=Data1)
        sns.catplot(x="City", y="Capacity", hue="Government Programs", kind="bar", data=Data1)
```

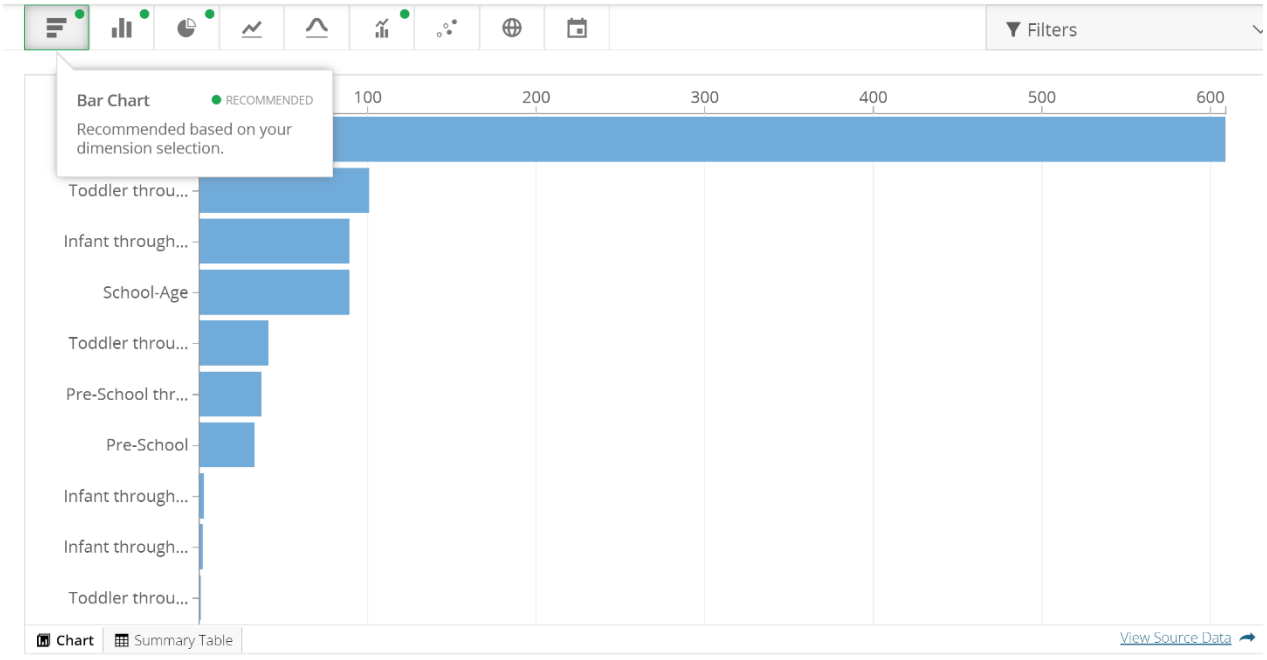
```
Out[ ]: <seaborn.axisgrid.FacetGrid at 0x7f5fee88f090>
```



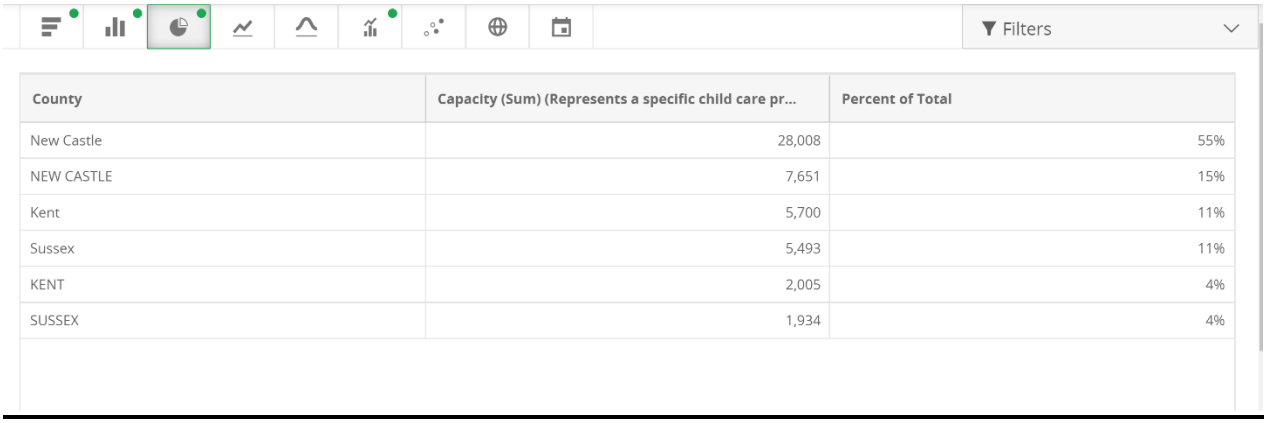
➢ Age Group and Count is relationship with each other in this dataset

Age Group	Count (Sum) (Represents a specific child care provider or facility licensed by the...
Infant through School-Age	609
Toddler through School-Age	101
Infant through Pre-School	89
School-Age	89
Toddler through Pre-School	41
Pre-School through School-Age	37
Pre-School	33
Infant through Toddler	3
Infant through Other	2
Toddler through Other	1

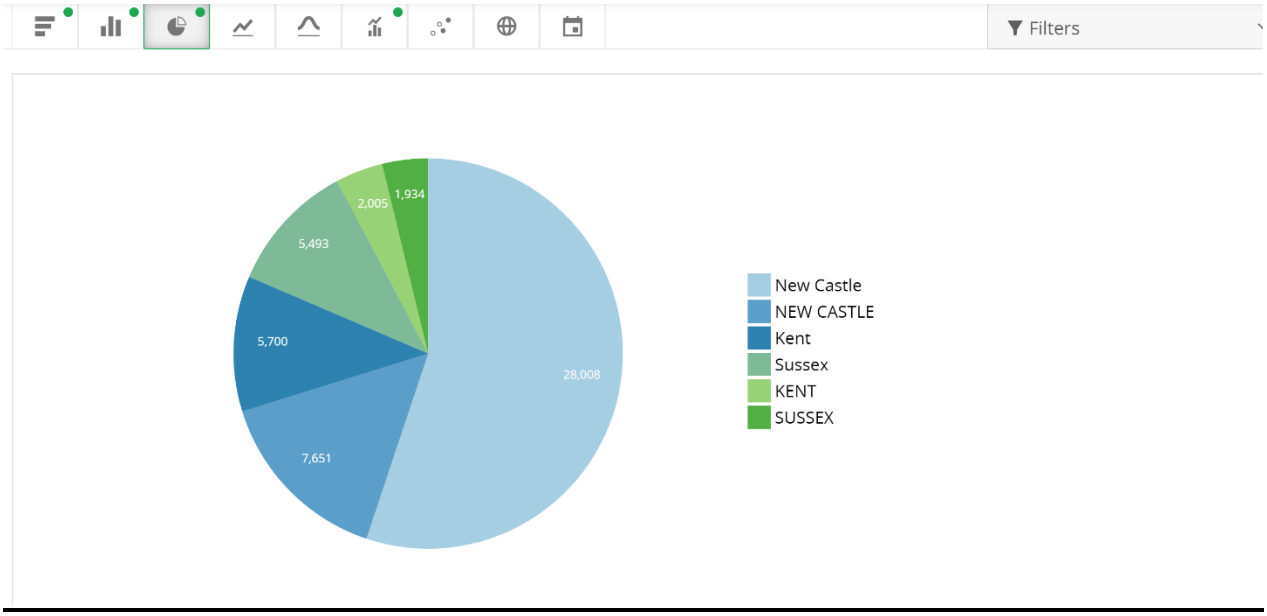
Toddler through Pre-School



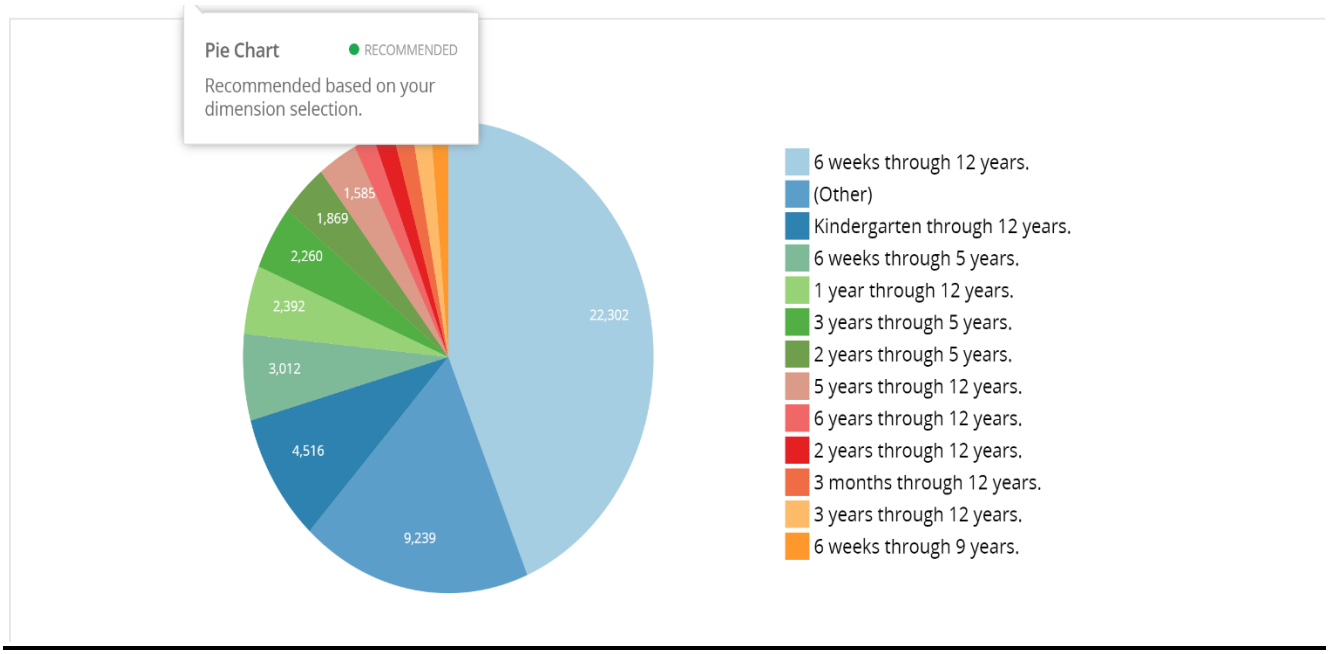
- **County and Capacity is relationship with each other (represent a specific Child Care Provider)**



County	Capacity (Sum) (Represents a specific child care pr...	Percent of Total
New Castle	28,008	55%
NEW CASTLE	7,651	15%
Kent	5,700	11%
Sussex	5,493	11%
KENT	2,005	4%
SUSSEX	1,934	4%



- **Age range and Capacity is relationship with each other in this dataset**



Age Range	Capacity (Sum) (Represents a specific child care pr...	Percent of Total
6 weeks through 12 years.	22,302	44%
(Other)	9,239	18%
Kindergarten through 12 years.	4,516	9%
6 weeks through 5 years.	3,012	6%
1 year through 12 years.	2,392	5%
3 years through 5 years.	2,260	4%
2 years through 5 years.	1,869	4%
5 years through 12 years.	1,585	3%
6 years through 12 years.	811	2%
2 years through 12 years.	793	2%
3 months through 12 years.	767	2%
3 years through 12 years.	766	2%

- **Closing Thoughts And Future Scope: -**

- In this project data is repeated and redundant. I Need to use the Exploratory data analysis Feature engineering techniques like One hot Encoding, Encoding with the multiple Categorical Variable. Feature selection Techniques.
- This dataset is most of the features are categorical. For the visualization I need numerical dataset as well. For that. I will do categorical variable to numerical variable Convert using Feature Engineering Method.
- After the Data Cleaning Done the Main Purpose is the data analysis using Different kinds Of Visualization Methods. After we get the clear view about the data.
- We choose the different Algorithm methods like Linear regression, Logistic regression and Decision tree, Random Forest. And train and test the dataset here. For the Dataset model accuracy.
- Once we get the dataset accuracy using different Algorithm. and model creation. We will start working on model deployment using the docker and Kubernetes.