US Unemployment Dataset (2010 - 2020)

From January 2010 until 2020, this dataset offers time series data on the unemployment rate in the United States. It offers data on unemployment rates for adults according on their educational attainment, race, and gender.
This dataset also includes data on unemployment rates by state for the year 2020. The data set is tidy, thus there isn't much pre-processing required.
This dataset was obtained from the Kaggle repository and has been identified as coming from the "US. BUREAU OF LABOR STATISTICS" https://data.bls.gov/cgi-bin/surveymost?

The image below depicts the importing of the dataset and the verification of the dataset's metadata to give us a better idea of what our dataset contains. The dataset comprises 13 columns and 132 entries, according to the information gathered.

	Year	Month	Primary_School	Date	High_School	Associates_Degree	Professional_Degree	White	Black	Asian	Hispanic	Men	Women
0	2010	Jan	15.3	Jan-2010	10.2	8.6	4.9	8.8	16.5	8.3	12.9	10.2	7.9
1	2011	Jan	14.3	Jan-2011	9.5	8.1	4.3	8.1	15.8	6.8	12.3	9.0	7.9
2	2012	Jan	13.0	Jan-2012	8.5	7.1	4.3	7.4	13.6	6.7	10.7	7.7	7.6
3	2013	Jan	12.0	Jan-2013	8.1	6.9	3.8	7.1	13.7	6.4	9.7	7.5	7.2
4	2014	Jan	9.4	Jan-2014	6.5	5.9	3.3	5.7	12.1	4.7	8.3	6.2	5.8

```
# Checking the information of the dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 132 entries, 0 to 131
Data columns (total 13 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Year                 132 non-null    int64
 1   Month                132 non-null    object
 2   Primary_School       123 non-null    float64
 3   Date                 132 non-null    object
 4   High_School          123 non-null    float64
 5   Associates_Degree    123 non-null    float64
 6   Professional_Degree  123 non-null    float64
 7   White                123 non-null    float64
 8   Black                123 non-null    float64
 9   Asian                123 non-null    float64
 10  Hispanic             123 non-null    float64
 11  Men                  123 non-null    float64
 12  Women                123 non-null    float64
dtypes: float64(10), int64(1), object(2)
memory usage: 13.5+ KB
```

On the figure below, we are checking for the missing values in our dataset and its shows that almost all of our numerical column has missing data. The function fillna is then used to fill the missing data with the mean of each of the columns.

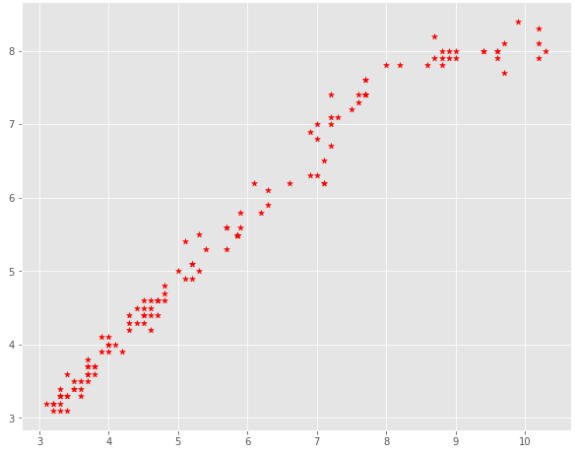
```
# Checking the number of missing data in each columns
data.isnull().sum()
```

```
Year                 0
Month                0
Primary_School       9
Date                 0
High_School          9
Associates_Degree    9
Professional_Degree  9
White                9
Black                9
Asian                9
Hispanic             9
Men                  9
Women                9
dtype: int64
```

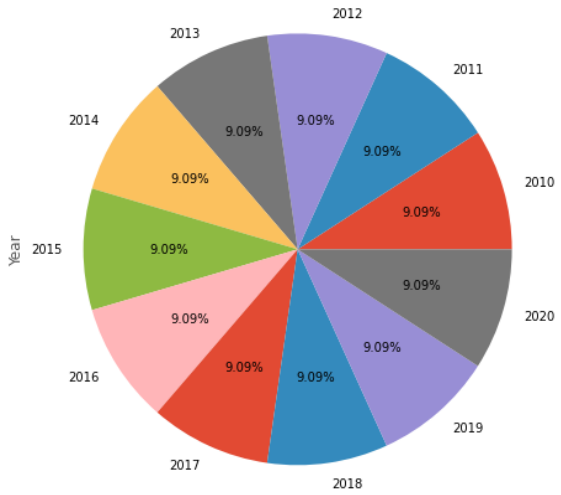
```
# Checking the summary of the dataset
data.describe()
```

	Year	Primary_School	High_School	Associates_Degree	Professional_Degree	White	Black	Asian	Hispanic	Men	Wom
count	132.000000	132.000000	132.000000	132.000000	132.000000	132.000000	132.000000	132.000000	132.000000	132.000000	132.000
mean	2015.000000	9.329268	6.391057	5.338211	3.127642	5.455285	10.692683	4.679675	7.660163	5.848780	5.485
std	3.174324	3.226293	2.145864	1.829050	0.904805	1.815674	3.489486	1.610033	2.738076	2.085038	1.689
min	2010.000000	4.800000	3.400000	2.700000	1.900000	3.100000	5.400000	2.100000	3.900000	3.100000	3.100
25%	2012.000000	6.475000	4.600000	3.700000	2.400000	3.800000	7.500000	3.375000	5.100000	4.000000	4.000
50%	2015.000000	8.650000	5.750000	5.200000	3.050000	5.050000	10.692683	4.400000	6.850000	5.350000	5.350
75%	2018.000000	12.000000	8.100000	6.900000	3.925000	7.025000	13.725000	5.900000	9.925000	7.350000	7.225
max	2020.000000	15.800000	11.000000	8.900000	5.000000	9.000000	16.800000	8.300000	12.900000	10.300000	8.400

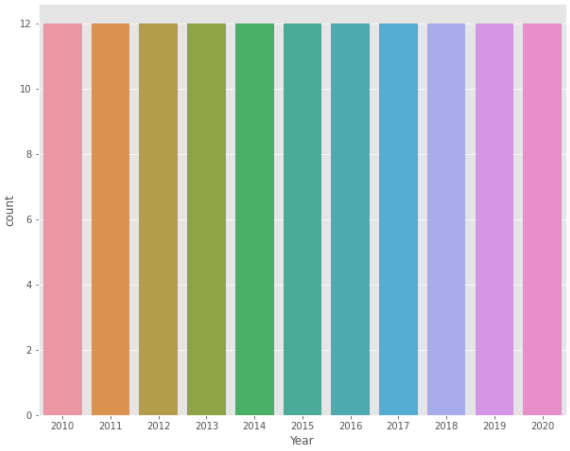
On the figure above we show the statistical summary of the dataset showing the standard deviation, mean, min, max and count of the dataset



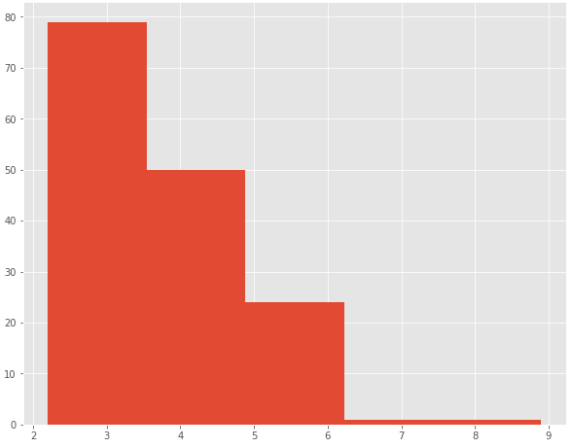
The figure above shows the scattered plot distribution between the Men and Women column on the dataset. The graph shows that there is a correlation between the two columns



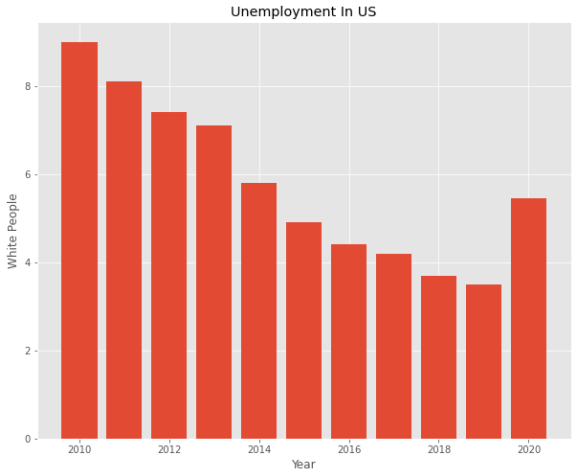
The figure above shows the pie chart percentage count on the year column shown the percentage distribution of each of the categories of the year present in the column. The bar chart representation below shows that the years have almost the same range of counts which are not much differing.



The figure below shows the histogram plot distribution of the unemployment rate; this presents the unemployment rate which reduces drastically down along the line.



The figure below shows the Unemployment in US of the White people, which shows that from year 2010 it decreases down to the year 2019 and the unemployment rate jumped up in year 2020 which shows that the Covid 19 period actually affected the unemployment rate of the white negatively.



In conclusion, we can see the rise in unemployment rate since the start of 2020. This may be the impact of COVID-19 and needs further analysis.

Unemployment rate in the United States has affected everybody irrespective of education qualification, race, and gender. The US government needs to create a better system that could help to prepare for an unexpected pandemic in order not to affect the rate of loss of job. This improve system will also help the country economic to grow.