

```
In [ ]: Course: DSC630 - Predictive Analytics<br>
Instructor: Fadi Alsaleem<br>
```

```
In [47]: '''Karthikeyan Chellamuthu '''

# Assignment Week1 -Exercise1. Week 1 Assignment Python refresher

''' Karthikeyan Chellamuthu - Date :06-12-2022'''
```

```
Out[47]: ' Karthikeyan Chellamuthu - Date :06-12-2022'
```

```
In [ ]: This assignment is a refresher of data analysis and visualization using Python and/or
Find a data set that interests you and has appropriate data to create some interesti
A few good sources for finding datasets include Kaggle, UCI ML Repository, and the U

With the dataset that you choose, perform the following steps using Python and/or R:

About Data set

Considering the past few month situation between Russia and Ukraine there are variet
across the globe and some small countries like srilanka is suffering due to inflatio
very simple data set to understand about the features available in the Matplotlib.
library for creating static, animated, and interactive visualizations in Python. Mat
easy and hard things possible.
```

```
In [35]: # 1. Write a summary of your data and identify the questions to explore visually wit
```

```
In [ ]: ...

The given data set is containing the decade of data for each country, it's a very si
we can compare the visualization to create a line plot and bar charts to visualize t
In addition, we can add another country to display the legend. here are my question
'''
```

```
In [ ]: # Here are the two questions that i am planning to apply the summary of my results s
...

1. Perform the analysis for the given data gas price to uk using bar chart.
2. Plot a linear graph compare USA vs UK the gas price and look for the price hike d
3. Perform a yearly percentage for a county and create a pie chart. '''
```

```
In [5]: #Import required libraries for matplotlib
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
%matplotlib inline
%matplotlib inline
import matplotlib.path_effects as path_effects
import os
os.chdir('/Users/LENOVO/Desktop/BU/DSC 630')
os.getcwd()
```

```
Out[5]: 'C:\\Users\\LENOVO\\Desktop\\BU\\DSC 630'
```

```
In [6]: #Create a dataframe
gasPrice = pd.read_csv('gas_prices.csv')
```

```
In [7]: #Display the data frame
gasPrice.head()
```

```
Out[7]:
```

	Year	Australia	Canada	France	Germany	Italy	Japan	Mexico	South Korea	UK	USA
0	2010	1.63	1.38	3.87	3.34	3.84	2.82	1.49	3.04	4.06	1.06
1	2011	1.72	1.52	3.85	3.42	3.87	3.27	1.79	3.80	4.29	1.17
2	2012	1.94	1.86	3.80	3.45	3.77	3.65	2.01	4.18	4.58	1.51
3	2013	1.71	1.72	3.51	3.40	3.57	3.27	2.20	3.76	4.13	1.46
4	2014	1.76	1.69	3.62	3.67	3.74	3.15	2.24	3.84	4.16	1.36

```
In [8]: # display the number of rows and columns
gasPrice.shape
```

```
Out[8]: (13, 11)
```

```
In [9]: # display the dtype for each of the column
gasPrice.dtypes
```

```
Out[9]: Year          int64
Australia    float64
Canada       float64
France       float64
Germany      float64
Italy        float64
Japan        float64
Mexico       float64
South Korea  float64
UK           float64
USA          float64
dtype: object
```

```
In [32]: #Simple Histogram chart for the gas price in UK.
```

```
# Read CSV into pandas
gasPrice.head()
df = pd.DataFrame(gasPrice)

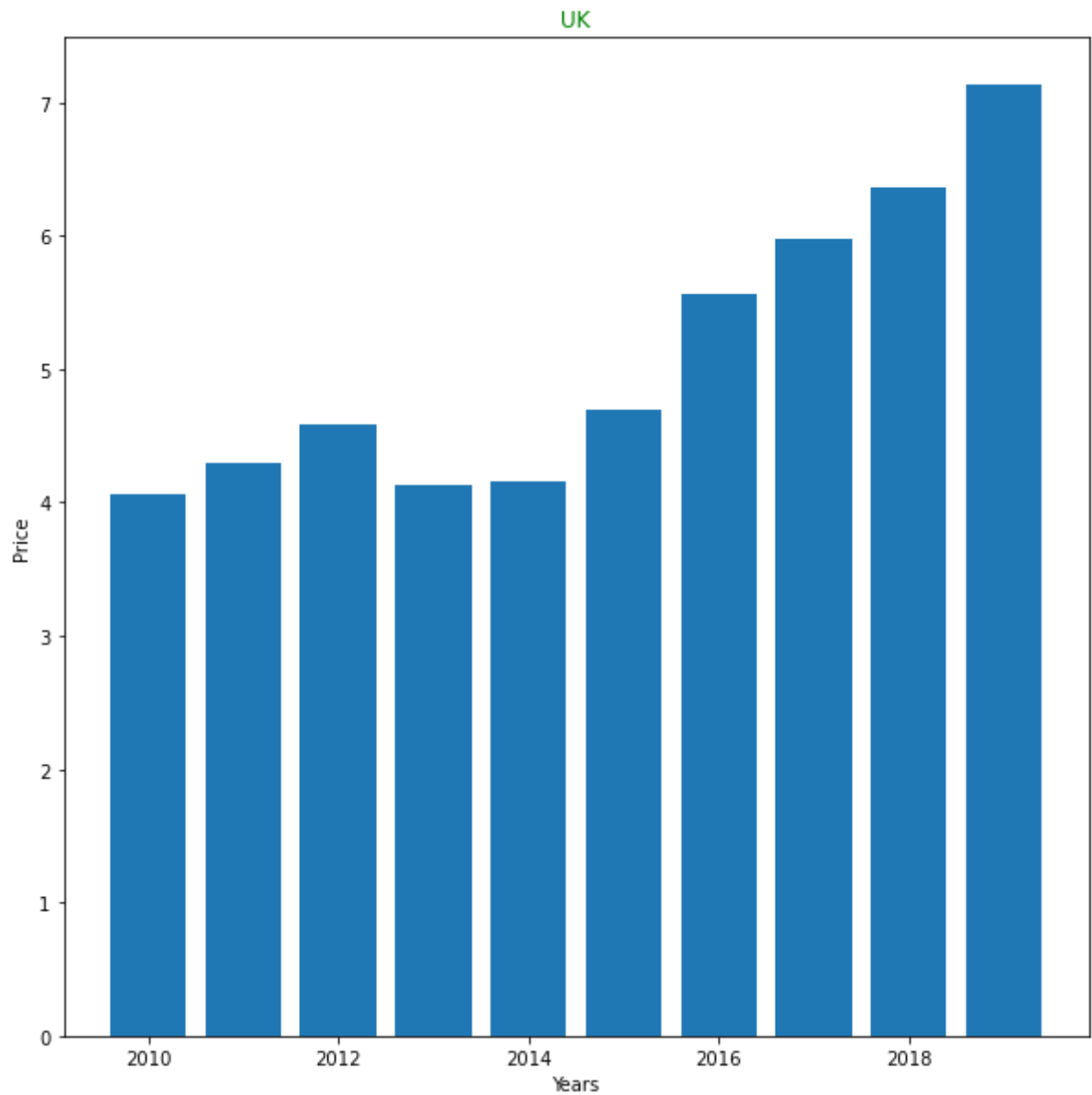
name = df['Year'].head(12)
price = df['UK'].head(12)

# Figure Size
fig = plt.figure(figsize =(10, 10))

# Horizontal Bar Plot
plt.bar(name[0:10], price[0:10])

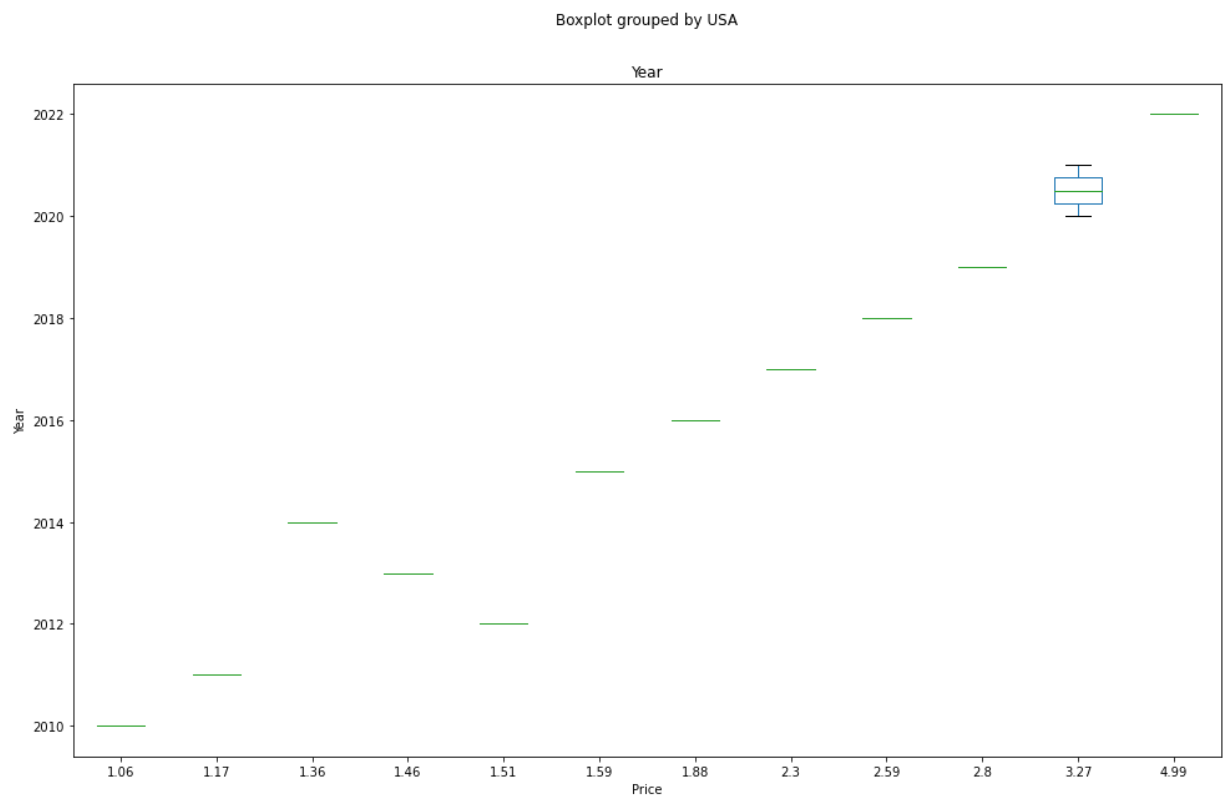
# Show Plot
plt.xlabel('Years',)
plt.ylabel('Price in Euro')
```

```
plt.title('UK', color='GREEN')  
plt.show()
```



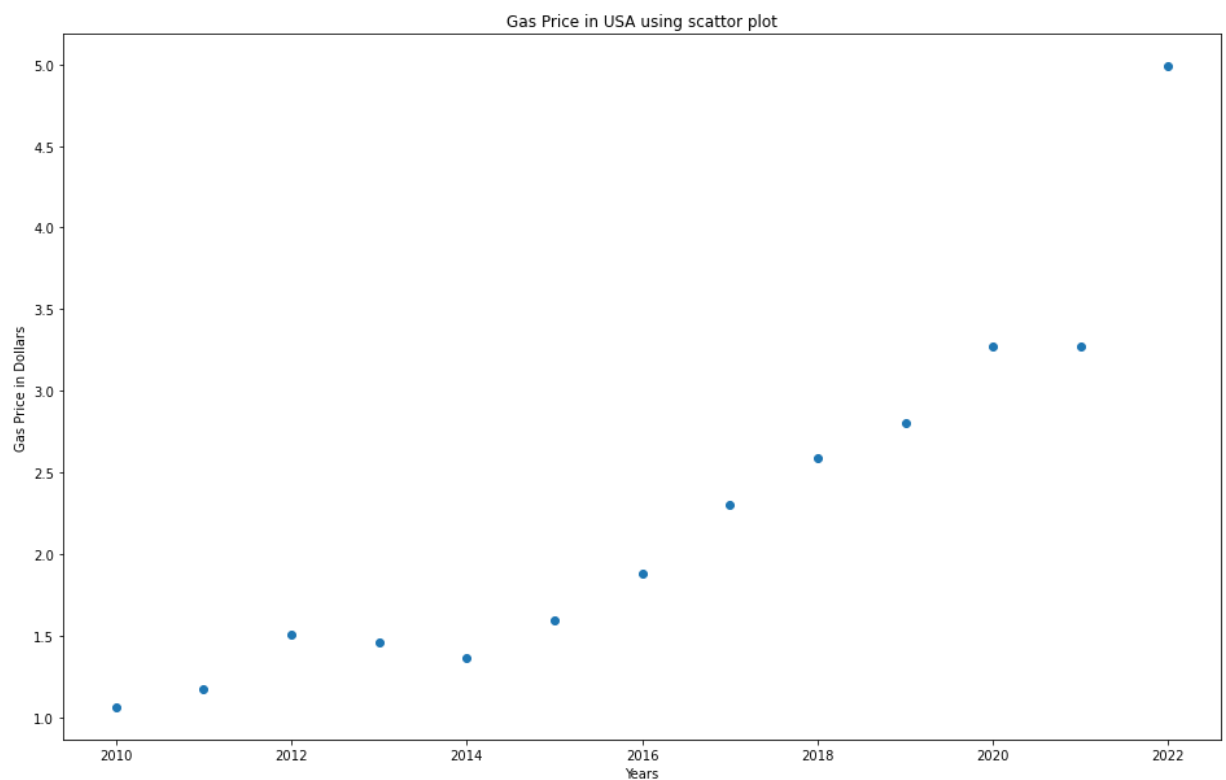
```
In [38]: # boxplot of gas oruce (scale of 0-10) with respect to gas price  
bplot = gasPrice.boxplot(by='USA', column=['Year'], grid=False, figsize=(16,10))  
bplot.set_ylabel('Year')  
bplot.set_xlabel('Price')
```

```
Out[38]: Text(0.5, 0, 'Price')
```



```
In [39]: # Scatterplot of Gas Price in Dollars by Latest Price
plt.figure(figsize=(16, 10))
plt.scatter( gasPrice.Year,gasPrice.USA)
plt.title('Gas Price in USA using scattor plot')
plt.xlabel('Years')
plt.ylabel('Gas Price in Dollars')
```

```
Out[39]: Text(0, 0.5, 'Gas Price in Dollars')
```



```
In [40]: #Simple bar chart for the price hike in USA.
```

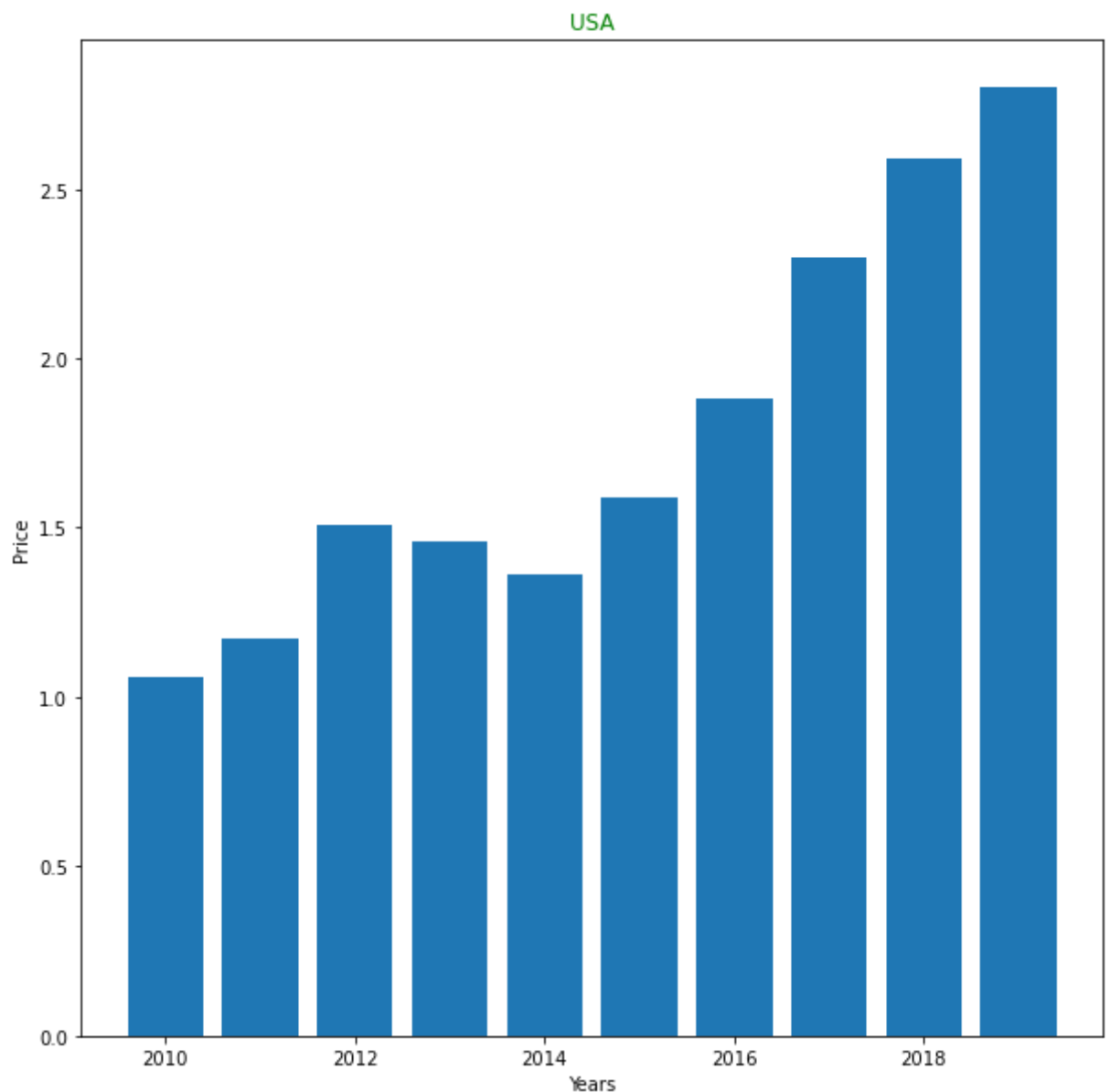
```
# Read CSV into pandas
gasPrice.head()
df = pd.DataFrame(gasPrice)

name = df['Year'].head(12)
price = df['USA'].head(12)

# Figure Size
fig = plt.figure(figsize=(10, 10))

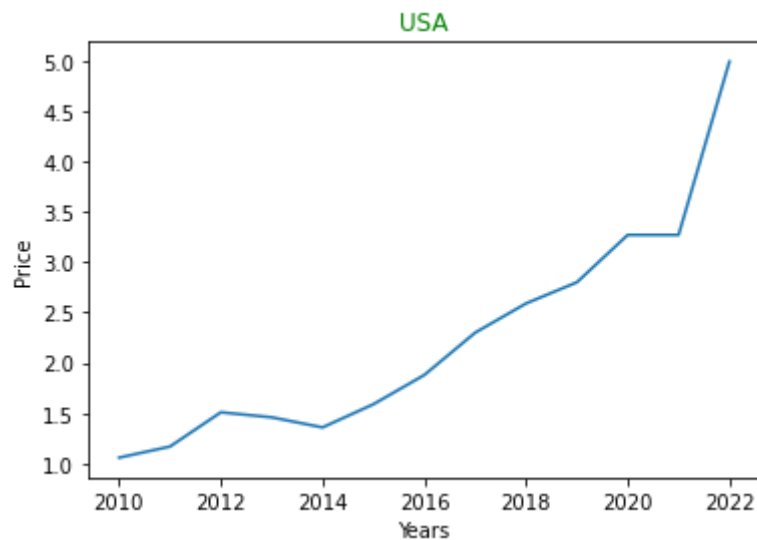
# Horizontal Bar Plot
plt.bar(name[0:10], price[0:10])

# Show Plot
plt.xlabel('Years',)
plt.ylabel('Price')
plt.title('USA', color='GREEN')
plt.show()
```



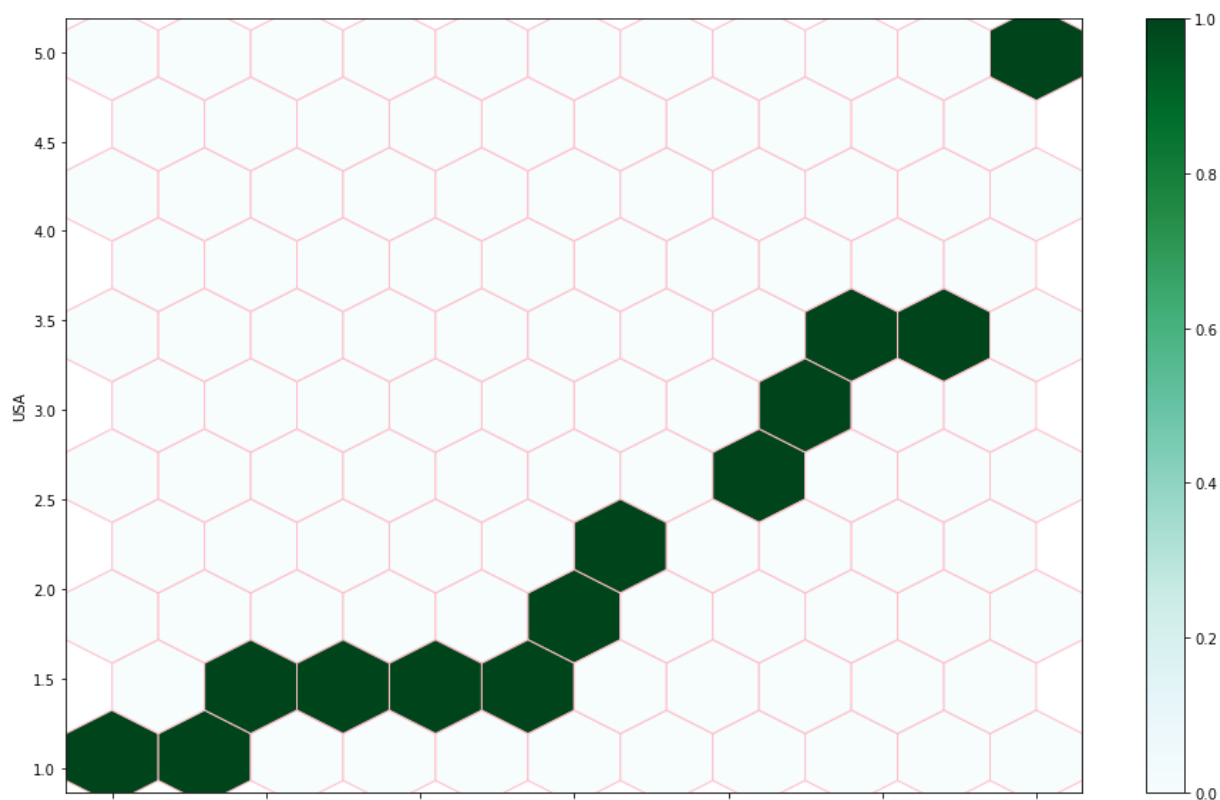
```
In [41]: #Sample line plot for the gas price
plt.plot(gasPrice.Year, gasPrice.USA)
plt.xlabel('Years',)
plt.ylabel('Price')
plt.title('USA', color='GREEN')
```

Out[41]: Text(0.5, 1.0, 'USA')



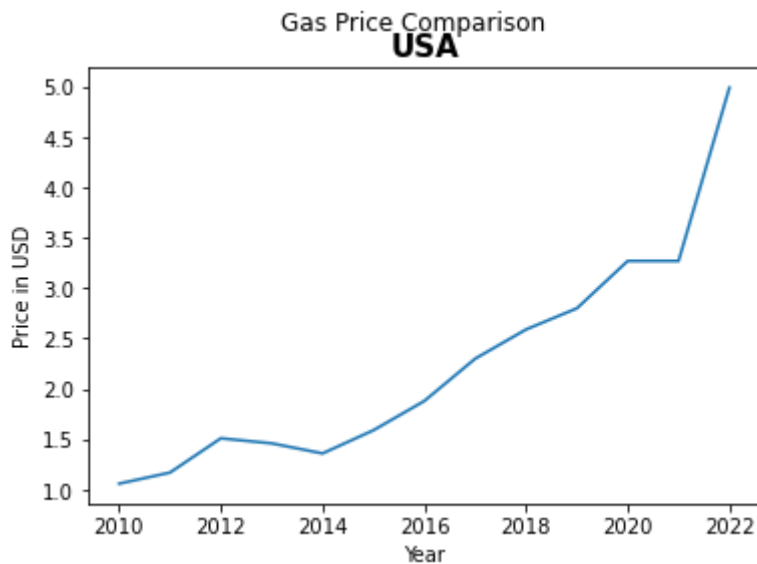
In [17]: `## Hex plot for Gas price and the hex plot for the same, color='GREEN'`  
`gasPrice.plot.hexbin(x='Year', y='USA', gridsize=10, figsize = (16,10), color='PINK')`

Out[17]: <AxesSubplot:xlabel='Year', ylabel='USA'>



In [23]: `#Gas price comparison over the years`  
`gasPrice = pd.read_csv('gas_prices.csv')`  
`plt.plot(gasPrice.Year, gasPrice.USA)`  
`plt.suptitle('Gas Price Comparison')`  
`plt.title('USA', fontdict={'fontsize':15,'fontweight':'bold'})`  
`plt.xlabel('Year')`  
`plt.ylabel('Price in USD')`

Out[23]: Text(0, 0.5, 'Price in USD')



```
In [24]: # what is the data type of this data
gasPrice.dtypes
```

```
Out[24]: Year          int64
Australia    float64
Canada       float64
France       float64
Germany      float64
Italy        float64
Japan        float64
Mexico       float64
South Korea  float64
UK           float64
USA          float64
dtype: object
```

```
In [25]: # removing null values to avoid errors
gasPrice.dropna(inplace = True)

# percentile list
perc = [.20, .40, .60, .80]

# list of dtypes to include
include = ['object', 'float', 'int']

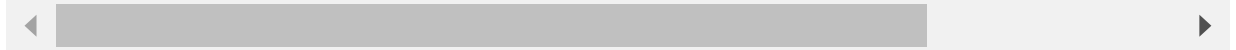
# calling describe method
desc = gasPrice.describe(percentiles = perc, include = include)

# display
desc
```

```
Out[25]:
```

	Year	Australia	Canada	France	Germany	Italy	Japan	Mexico	South Korea	UK
<b>count</b>	13.00000	13.000000	13.000000	13.000000	13.000000	13.000000	13.000000	13.000000	13.000000	13.000000
<b>mean</b>	2016.00000	2.893846	2.654615	5.266154	5.302308	5.390000	4.155385	2.160000	4.155385	4.155385
<b>std</b>	3.89444	1.145481	1.051274	1.584085	1.796743	1.622673	1.036714	0.286298	1.036714	1.036714
<b>min</b>	2010.00000	1.630000	1.380000	3.510000	3.340000	3.570000	2.820000	1.490000	3.570000	3.570000
<b>20%</b>	2012.40000	1.736000	1.702000	3.820000	3.432000	3.798000	3.270000	2.018000	3.798000	3.798000

	Year	Australia	Canada	France	Germany	Italy	Japan	Mexico	
<b>40%</b>	2014.80000	2.140000	1.964000	4.254000	4.406000	4.398000	3.614000	2.168000	4.
<b>50%</b>	2016.00000	2.720000	2.370000	4.990000	5.240000	5.290000	3.930000	2.220000	4.
<b>60%</b>	2017.20000	3.292000	2.964000	5.544000	5.734000	5.812000	4.318000	2.254000	5.
<b>80%</b>	2019.60000	4.198000	3.884000	7.146000	7.402000	7.270000	5.240000	2.430000	5.
<b>max</b>	2022.00000	4.450000	4.080000	7.510000	7.750000	7.630000	5.740000	2.450000	6.



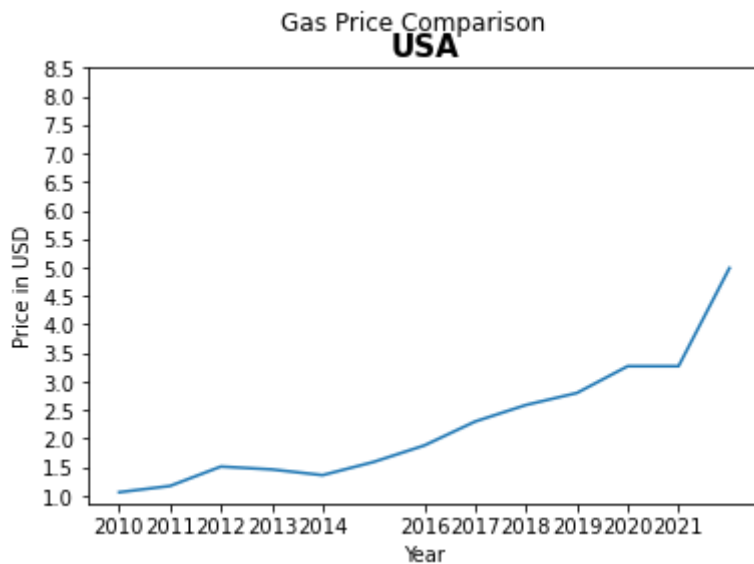
In [26]:

```
#Price vs year chart to USA with Tick set to the graph
gasPrice = pd.read_csv('gas_prices.csv')
plt.plot(gasPrice.Year, gasPrice.USA)
plt.suptitle('Gas Price Comparison')
plt.title('USA', fontdict={'fontsize':15,'fontweight':'bold'})
plt.xlabel('Year')
plt.ylabel('Price in USD')
plt.xticks([2010,2011,2012,2013,2014,2016,2017,2018,2019,2020,2021])
plt.yticks([1,1.5,2,2.5,3,3.5,4,4.5,5,5.5,6,6.5,7,7.5,8,8.5])
```

Out[26]:

```
([<matplotlib.axis.YTick at 0x1ff4471bfa0>,
<matplotlib.axis.YTick at 0x1ff4471b8b0>,
<matplotlib.axis.YTick at 0x1ff446e2c70>,
<matplotlib.axis.YTick at 0x1ff4475f790>,
<matplotlib.axis.YTick at 0x1ff4475a580>,
<matplotlib.axis.YTick at 0x1ff44742ee0>,
<matplotlib.axis.YTick at 0x1ff4475fb80>,
<matplotlib.axis.YTick at 0x1ff44767310>,
<matplotlib.axis.YTick at 0x1ff44767a60>,
<matplotlib.axis.YTick at 0x1ff4476e1f0>,
<matplotlib.axis.YTick at 0x1ff44767d90>,
<matplotlib.axis.YTick at 0x1ff4475fbe0>,
<matplotlib.axis.YTick at 0x1ff4476e9d0>,
<matplotlib.axis.YTick at 0x1ff4476ee50>,
<matplotlib.axis.YTick at 0x1ff447755e0>,
<matplotlib.axis.YTick at 0x1ff44775d30>],
[Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, ''),
Text(0, 0, '')[15]])
```





In [42]:

```
#Gas price comparision USA Vs UK
gasPrice = pd.read_csv('gas_prices.csv')
plt.plot(gasPrice.Year, gasPrice.Canada,'b.-',label = 'UK',color='green')
plt.plot(gasPrice.Year, gasPrice.USA,'b.-',label = 'USA',color='red')
plt.suptitle('Gas Price Comparison')
plt.title('UK Vs USA', fontdict={'fontsize':15,'fontweight':'bold'})
plt.xlabel('Year')
plt.ylabel('Price in USD')
plt.xticks([2010,2011,2012,2013,2014,2016,2017,2018,2019,2020,2021])
plt.yticks([1,1.5,2,2.5,3,3.5,4,4.5,5,5.5,6,6.5,7,7.5,8,8.5])
plt.legend()
```

C:\Users\LENOVO\AppData\Local\Temp\ipykernel\_20564\3916492106.py:3: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "b.-" (-> color='b'). The keyword argument will take precedence.

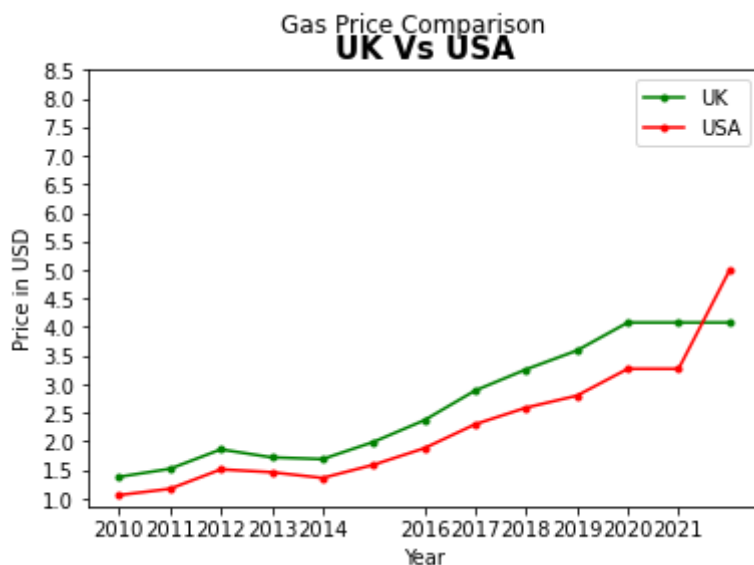
```
plt.plot(gasPrice.Year, gasPrice.Canada,'b.-',label = 'UK',color='green')
```

C:\Users\LENOVO\AppData\Local\Temp\ipykernel\_20564\3916492106.py:4: UserWarning: color is redundantly defined by the 'color' keyword argument and the fmt string "b.-" (-> color='b'). The keyword argument will take precedence.

```
plt.plot(gasPrice.Year, gasPrice.USA,'b.-',label = 'USA',color='red')
```

Out[42]:

```
<matplotlib.legend.Legend at 0x22966d91f10>
```



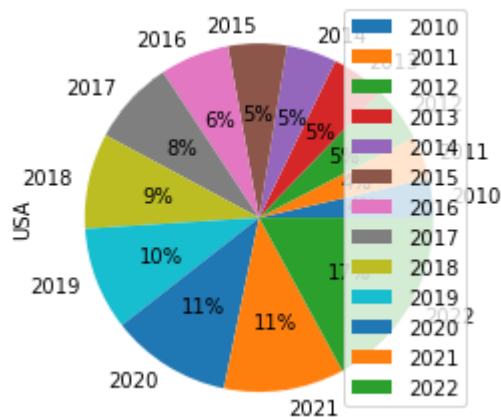
In [84]:

```
#Pie chart
gasPrice.groupby(['Year']).sum().plot(kind='pie', y='USA', autopct='%1.0f%%')
```

```
plt.title('Price per Year in percentage USA', fontdict={'fontsize':15,'fontweight':'
```

Out[84]: Text(0.5, 1.0, 'Price per Year in percentage USA')

### Price per Year in percentage USA



In [33]: *#Simple bar chart for the price hike in USA.*

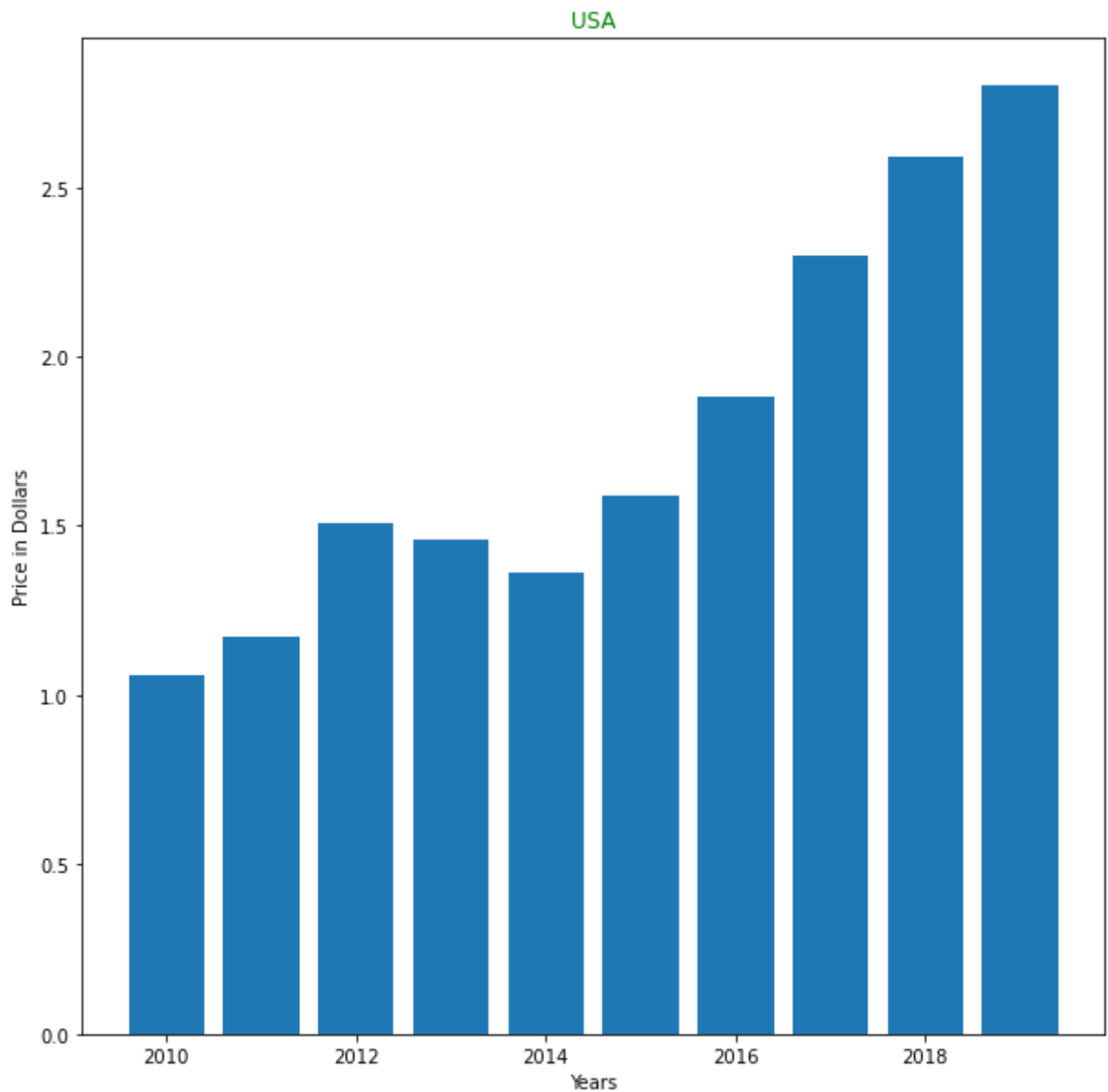
```
# Read CSV into pandas
gasPrice.head()
df = pd.DataFrame(gasPrice)

name = df['Year'].head(12)
price = df['USA'].head(12)

# Figure Size
fig = plt.figure(figsize =(10, 10))

# Horizontal Bar Plot
plt.bar(name[0:10], price[0:10])

# Show Plot
plt.xlabel('Years',)
plt.ylabel('Price in Dollars')
plt.title('USA', color='GREEN')
plt.show()
```



```
In [45]: #Final summary and conclusion

...
Addressed the following questions and

* Python is a great language for doing data analysis, primarily because of the fant
* Gas price comparison over the years in a country to understand the trend and how
* Understanding easily the given data type of any data.
* Learnt how to perform the plots using matplotlib,
* Learned about describe () method is used for calculating some statistical data li
* It analyzes both numeric and object series and also the DataFrame column sets of
* Cleansed the data to produce clean visualization.

1. Perform the analysis for the given data gas price to usa using bar chart.
2. Plot a linear graph compare USA vs UK the gas price and look for the price hike d
3. Perform a yearly percentage for a county and create a pie chart.
...
```

```
Out[45]: '\nAddressed the following questions and \n\nPython is a great language for doing da
ta analysis, primarily because of the fantastic ecosystem of data-centric Python pac
kages. Pandas is one of those packages and makes importing and analysing data much e
asier.\nGas price comparison over the years in a country to understand the trend and
how inflation raised over the years. \nUnderstanding easily the given data type of a
ny data. \nLearnt how to perform the plots using matplotlib, \nLearned about describ
```

e () method is used for calculating some statistical data like percentile, mean and standard deviation of the numerical values of the Series or DataFrame. \nIt analyzes both numeric and object series and also the DataFrame column sets of mixed data types.\nCleansed the data to produce clean visualization. \n\n1. Perform the analysis for the given data gas price to usa using bar chart. \n2. Plot a linear graph compare USA vs UK the gas price and look for the price hike due to war in 2022. \n3. Perform a yearly percentage for a county and create a pie chart. \n'

In [ ]: