Karthick Pandi

Nano Degree in Data Analyst

In this project, I am going to explore the local and global temperature data and compare the temperature trends of the nearest city with the overall global temperature trends.Creating a visualization and prepare a write up describing the similarities and differences between global temperature trends and temperature trends in the closest big city. ¶

SQL Queries used for downloading the data from the tables

1) Select * from city_list - For Selecting my nearest city

2) select * from city_data where (country = 'Kenya' and city = 'Nairobi') -for getting the data for the city Nairobi in Kenya

3) Select * from global_data - For getting the global data

In [ ]:

In [ ]: # Importing the required packages

In [6]:
```
import pandas as pd
```

In [7]:
```
import numpy as np
```

In [ ]:
```
# Importing the required packages for data visualization
```

In [8]:
```
import seaborn as sns
```

In [9]:
```
import matplotlib.pyplot as plt
```

In [ ]:
```
# For seeing the graphs in jupyter page
```

In [10]:
```
%matplotlib inline
```

In [ ]:
```
# Importing the temperature data set with the data relevent to the city which I choose .Its done with the help of s
ql queries.

# Select * from city_data where (city = 'Nairobi' and country = 'Kenya');
# Select * from global_data;
```

In [11]:
```
city_data = pd.read_csv("citydata.csv")
```

In [12]:
```
city_data.head()
```

Out[12]:

	year	city	country	avg_temp
0	1850	Nairobi	Kenya	15.33
1	1851	Nairobi	Kenya	NaN
2	1852	Nairobi	Kenya	NaN
3	1853	Nairobi	Kenya	NaN
4	1854	Nairobi	Kenya	NaN

In [24]:
```
city_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 164 entries, 0 to 163
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   year      164 non-null    int64  
 1   city      164 non-null    object 
 2   country   164 non-null    object 
 3   avg_temp  141 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 5.2+ KB
```

In [13]:
```
city_data.describe()
```

Out[13]:

	year	avg_temp
count	164.00000	141.000000
mean	1931.50000	16.079716
std	47.48684	0.440104
min	1850.00000	15.110000
25%	1890.75000	15.810000
50%	1931.50000	16.050000
75%	1972.25000	16.330000
max	2013.00000	17.300000

In [15]:
```
city_data.columns
```

Out[15]:
```
Index(['year', 'city', 'country', 'avg_temp'], dtype='object')
```

In [18]:
```
# Importing the global temperature dataset
```

In [16]:
```
global_data = pd.read_csv("global_data.csv")
```

In [17]:
```
global_data.head()
```

Out[17]:

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47

In [18]:
```
global_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 266 entries, 0 to 265
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   year      266 non-null    int64  
 1   avg_temp  266 non-null    float64
dtypes: float64(1), int64(1)
memory usage: 4.3 KB
```

In [19]:
```
global_data.describe()
```

Out[19]:

	year	avg_temp
count	266.000000	266.000000
mean	1882.500000	8.369474
std	76.931788	0.584747
min	1750.000000	5.780000
25%	1816.250000	8.082500
50%	1882.500000	8.375000
75%	1948.750000	8.707500
max	2015.000000	9.830000

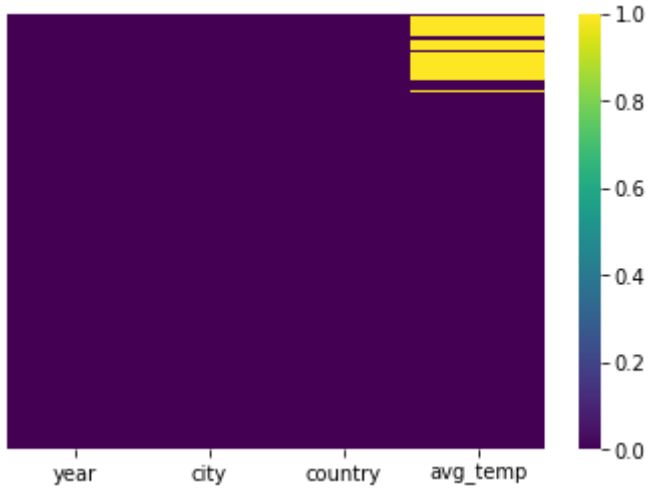
In [20]:
```
global_data.columns
```

Out[20]:
```
Index(['year', 'avg_temp'], dtype='object')
```

In [32]:
```
# Finding NA values in city_data
```

In [21]:
```
sns.heatmap(city_data.isnull(),yticklabels=False,xticklabels=True,cmap='viridis')
```

Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x256698ae488>



In [44]:
```
# It is conclusive from the above heat map that we are having some missing values in avg_temp column..
```

In [22]:
```
city_data.sort_values('year')
```

Out[22]:

	year	city	country	avg_temp
0	1850	Nairobi	Kenya	15.33
1	1851	Nairobi	Kenya	NaN
2	1852	Nairobi	Kenya	NaN
3	1853	Nairobi	Kenya	NaN
4	1854	Nairobi	Kenya	NaN
...	...	...	...	...
159	2009	Nairobi	Kenya	17.30
160	2010	Nairobi	Kenya	16.82
161	2011	Nairobi	Kenya	16.90
162	2012	Nairobi	Kenya	16.84
163	2013	Nairobi	Kenya	16.91

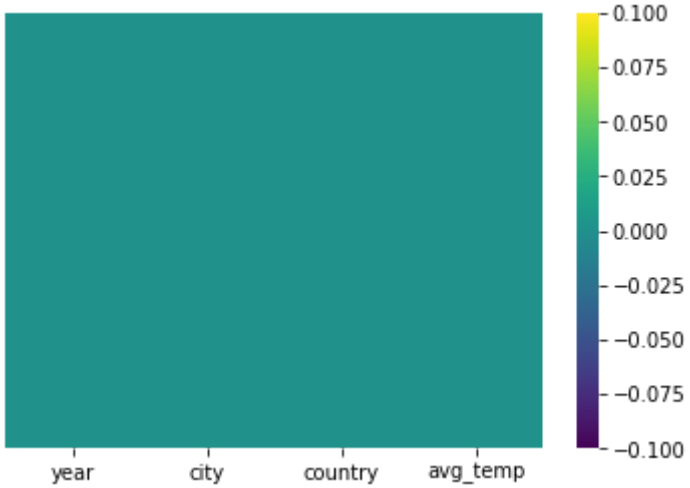
164 rows × 4 columns

In [23]:
```
# Function for imputing the missing average temperature by using the average temperature.
def imputeval(avg_temp):
    if pd.isnull(avg_temp):
        return city_data[city_data['year'] <= 1879]['avg_temp'].mean()
    else:
        return avg_temp;
```

In [24]:
```
# Imputing the values
city_data['avg_temp'] = city_data['avg_temp'].apply(imputeval)
```

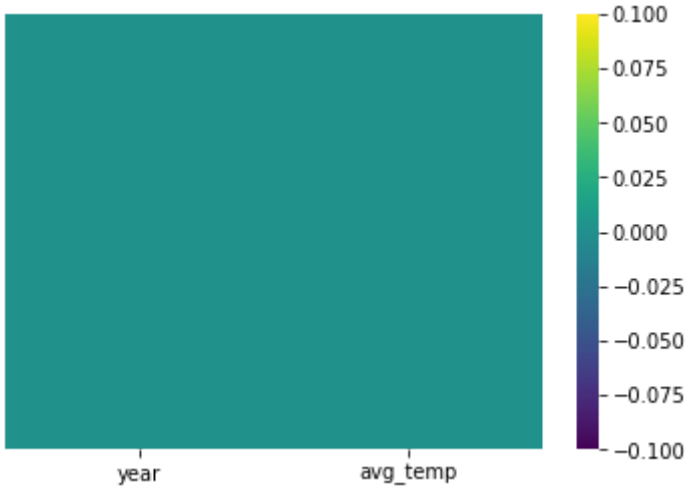
In [25]:
```
# checking wheather it has missing values furthermore..
sns.heatmap(city_data.isnull(),yticklabels=False,xticklabels=True,cmap='viridis')
```

Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x2566a07d888>



In [26]:
```
sns.heatmap(global_data.isnull(),yticklabels=False,xticklabels=True,cmap='viridis')
```

Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0x2566a117cc8>



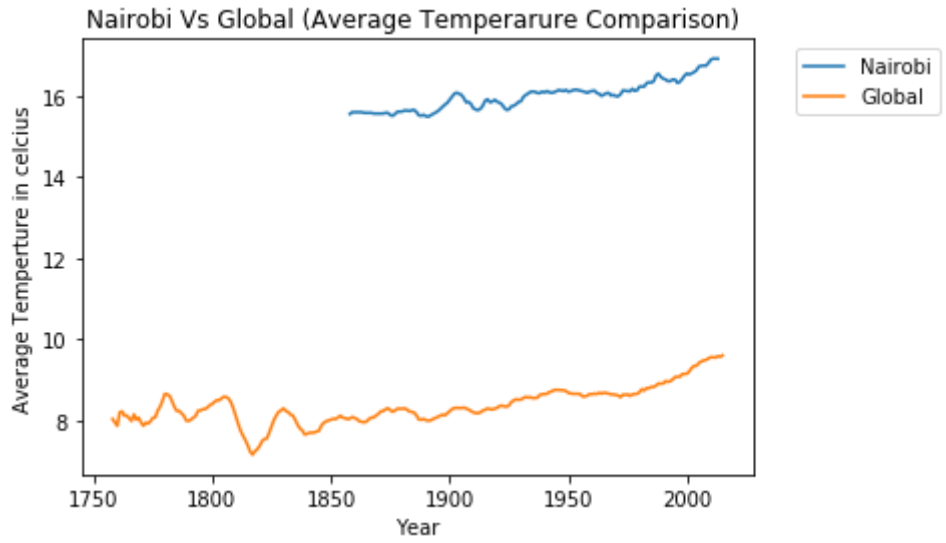
In [28]:
```
plt.figure(figsize=(10,10))
```

Out[28]: <Figure size 720x720 with 0 Axes>

<Figure size 720x720 with 0 Axes>

In [38]:
```
plt.plot(city_data['year'],city_data['avg_temp'].rolling(9).mean(),label = 'Nairobi')
plt.plot(global_data['year'],global_data['avg_temp'].rolling(9).mean(),label = 'Global')
plt.xlabel('Year')
plt.ylabel('Average Temperture in celcius')
plt.title("Nairobi Vs Global (Average Temperarure Comparison) ")
plt.legend(bbox_to_anchor=(1.05, 1))
```

Out[38]: <matplotlib.legend.Legend at 0x2566aa2ba48>



1) My city named Nairobi is the hottest one when compared with the global average temperature. Yes the differnece is consistent when the global average temperature gets increase the city nairobi also faces some increase in its average temperature.

2) In Nairobi, it is conclusive from the graph that years around 1810 - 1880 and the years around 1935 - 1985 the average temperature has maintains its steadiness in average values. For the years around 1960 - 1980 in global temperature graph we can see some evidence for steadiness in the avg values.

3) Yes the overall trend looks like it has constant increase in its average temperature.Yes the world is getting hotter than year by year.Yes, the temperature is consistent for last few hundred years.

4) City Nairobi has an average temperature values as double when compared with global average temperature.