Adelaide Weather Change

The following data is about the temperature changes in Adelaide. We are provided with two datasets from the Australian Bureau of Meteorology about climate changes in Adelaide from the year 1955 to the year 2022. The datasets have information about the maximum and minimum temperatures observed every day at Adelaide airport. We will clean the data first and process the data by merging the two datasets provided to us to extract the information about minimum temperature and maximum temperature observed in the day and the average temperature range for each year. We will support the information with relevant graphs which will help us to understand the changes that occured over time.

In [244…
```
#Importing libraries
import pandas as pd import
numpy as np
import matplotlib.pyplot as plt import
seaborn as sns
```

In [245…
```
min_temp=pd.read_csv(r'IDCJAC0011_023034_1800_Data.csv')
```

In [246…

In [247…
```
max_temp=pd.read_csv(r'IDCJAC0010_023034_1800_Data.csv')
```

In [248…
```
#Cleaning the csv files by re                oving the null v alues
min_temp=min_temp.dropna()
max_temp=max_temp.dropna()
```
Out[248]:

```
min_temp
```

	Product code	Bureau of Meteorology station number	Year	Month	Minimum Daytemperature (Degree C)	Days of accumulation of minimum temperature	Quality
47	IDCJAC0011	23034	1955	2	17	17.1	1.0	Y
48	IDCJAC0011	23034	1955	2	18	17.2	1.0	Y
49	IDCJAC0011	23034	1955	2	19	15.1	1.0	Y
50	IDCJAC0011	23034	1955	2	20	13.4	1.0	Y
51	IDCJAC0011	23034	1955	2	21	11.4	1.0	Y
...	...	...	...	...	...	...	...	...
24695	IDCJAC0011	23034	2022	8	12	11.1	1.0	N
24696	IDCJAC0011	23034	2022	8	13	9.8	1.0	N
24697	IDCJAC0011	23034	2022	8	14	7.8	1.0	N

24698	IDCJAC0011	23034	2022	8	15	10.1	1.0	N
24699	IDCJAC0011	23034	2022	8	16	8.1	1.0	N

24638 rows × 8 columns

In this datset,we are provided with project code, bureau of meteorology station number, year, month, day, maximum temperature(Degree C), days of accumulation of maximum temperature, quality.

In [249…
```
max_temp
```

Out[249]:

	Product code	Bureau of Meteorology station number	Year	Month	Day	Maximum temperature (Degree C)	Days of accumulation of maximum temperature	Quality
47	IDCJAC0010	23034	1955	2	17	29.9	1.0	Y
48	IDCJAC0010	23034	1955	2	18	27.2	1.0	Y
49	IDCJAC0010	23034	1955	2	19	22.6	1.0	Y
50	IDCJAC0010	23034	1955	2	20	25.6	1.0	Y
51	IDCJAC0010	23034	1955	2	21	26.6	1.0	Y
…	…	…	…	…	…	…	…	…
24694	IDCJAC0010	23034	2022	8	11	15.8	1.0	N
24695	IDCJAC0010	23034	2022	8	12	14.6	1.0	N
24696	IDCJAC0010	23034	2022	8	13	14.5	1.0	N
24697	IDCJAC0010	23034	2022	8	14	15.2	1.0	N
24698	IDCJAC0010	23034	2022	8	15	15.0	1.0	N

24651 rows × 8 columns

In this datset,we are provided with project code, bureau of meteorology station number, year, month, day, minimum temperature(Degree C), days of accumulation of minimum temperature, quality.

In [250…
```
#Minimum temperature obeserved
min_temp.loc[min_temp['Minimum temperature (Degree C)'].idxmin()]
```

Out[250]:
```
Product code                                    IDCJAC0011
Bureau of Meteorology station number                 23034
Year                                                 1982
Month                                                   6
Day                                                     8
Minimum temperature (Degree C)                       -2.6
Days of accumulation of minimum temperature           1.0
Quality                                                 Y
Name: 10020, dtype: object
```

The minimum temperature was observed during the year 1982 with temperature of -2.6C. The min_temp dataset has 24638 rows with 8 columns after cleaning the data.

In [251…
```
#Maximum temperature observed
max_temp.loc[max_temp['Maximum temperature (Degree C)'].idxmax()]
```

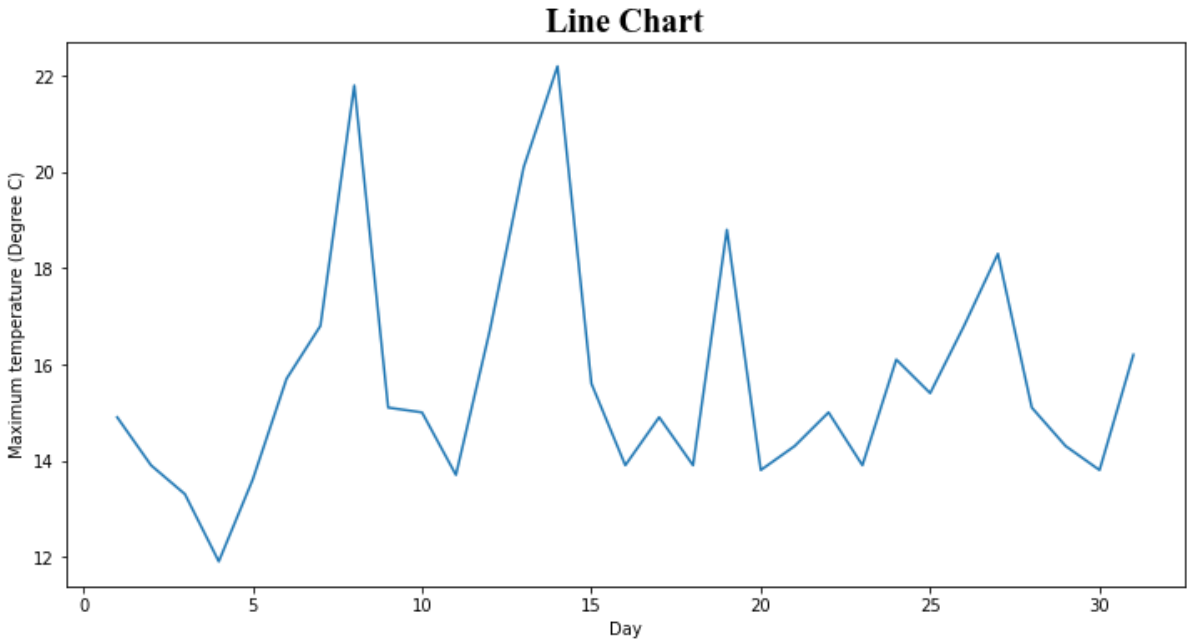
Out[251]:
```
Product code                                        IDCJAC0010
Bureau of Meteorology station number                 23034 Year
2019
Month                                                    1
Day                                                     24
Maximum temperature (Degree C)                        45.8
Days of accumulation of maximum temperature            1.0
Quality                                                  N
Name: 23399, dtype: object
```

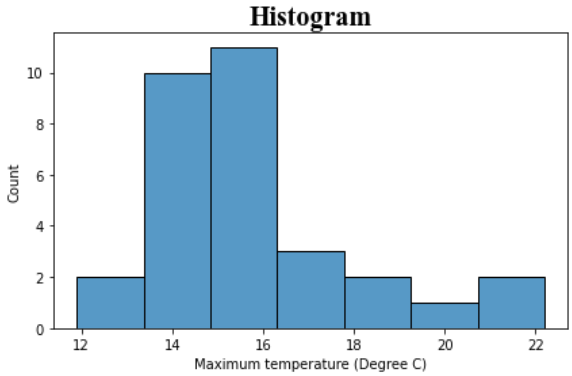
The maximum temperature was observed during the year 2019 with temperature of 45.8C.The max_temp dataset has 24651 rows with 8 columns after cleaning the data.

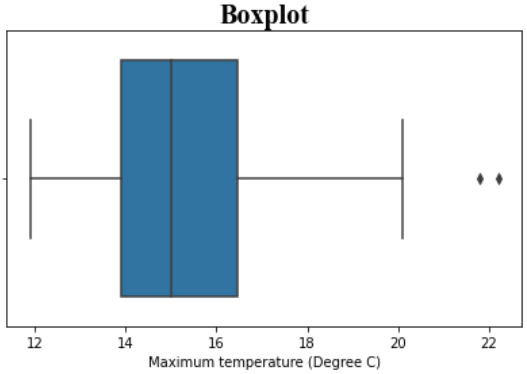
In [252…
```
#Function to calculate the descriptive statistics with year and month passed
def descriptiveStats(year,month):
    descriptiveStats= max_temp.loc[(max_temp['Year'] == year) & (max_temp['M
print("Mean: ",round(descriptiveStats['Maximum temperature (Degree C)'].
print("Median: ",round(descriptiveStats['Maximum temperature (Degree C)'
print("Standard Deviation: ",round(descriptiveStats['Maximum temperature
print("Skewness: ",round(descriptiveStats['Maximum temperature (Degree C
print("Kurtosis: ",round(descriptiveStats['Maximum temperature (Degree C
quartile75, quartile25 = np.percentile(ds['Maximum temperature (Degree C
iqr = quartile75 - quartile25     print("IQR: ",round(iqr,2))
    # Plotting the line chart for Day vs Maximum temperature (Degree C)
fig, ax = plt.subplots(figsize=(12, 6))     ax.plot(descriptiveStats["Day"],
descriptiveStats['Maximum temperature (     ax.set_title("Line
Chart",fontname="Times New Roman",fontweight="bold",s     plt.xlabel("Day")
    plt.ylabel("Maximum temperature (Degree C)")
    #Plotting the box plot and histogram     fig1,
ax = plt.subplots(1, 2, figsize=(15,4))
    sns.histplot(data=descriptiveStats, x='Maximum temperature (Degree C)',
ax[0].set_title("Histogram",fontname="Times New Roman",fontweight="bold"
sns.boxplot(data=descriptiveStats, x='Maximum temperature (Degree C)', a
ax[1].set_title("Boxplot",fontname="Times New Roman",fontweight="bold",s
```

In [253…
```
descriptiveStats(1961,8)
```

```
Mean:  15.64
Median:  15.0
Standard Deviation:  2.41
Skewness:  1.37
Kurtosis:  1.68
IQR:  2.55
```

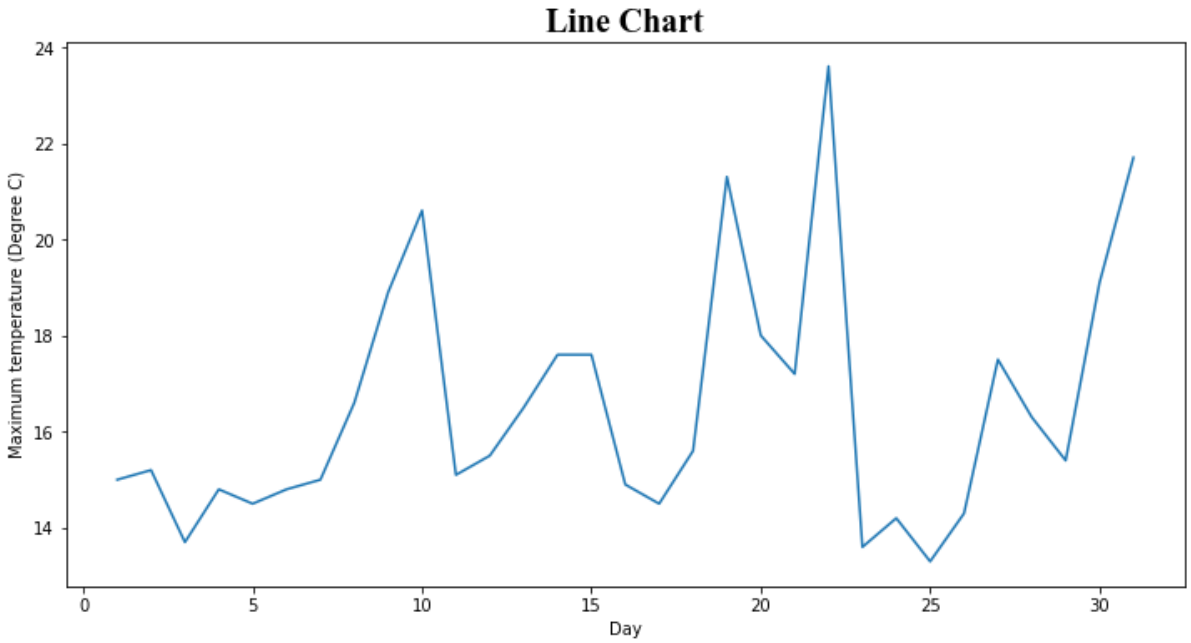


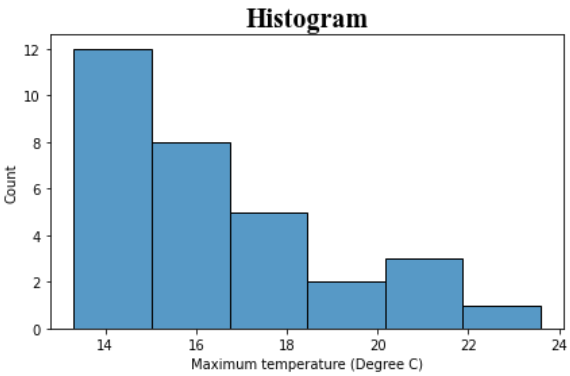


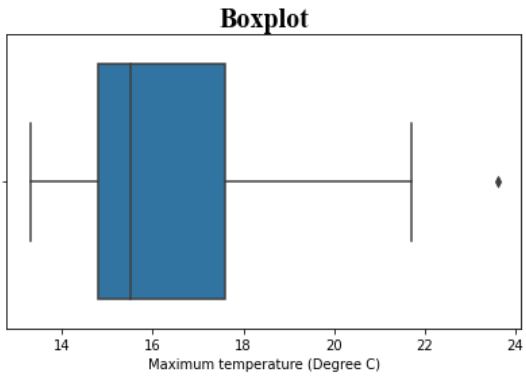


In [254… descriptiveStats(2021,8)

```
Mean:  16.51
Median:  15.5
Standard Deviation:  2.58
Skewness:  1.16
Kurtosis:  0.8
IQR:  2.55
```







The descriptive stats function provides us the information about the mean, median, skewness, kurtosis, standard deviation, and IQR.

The mean and median maximum temperature has increased from August 1961(15.64,15.0) compared to August 2021(16.51,15.5).

The standard deviation value of 2.41 in August 1961 compared to the standard deviation value of 2.58 in August 2021 means that the observations are almost equally spread out.

The IQR value of August 1961 and IQR value of August 2021 is 2.55 which indicates the distribution of 1961 and 2021 is similar.

The skewness value of 1.37 in August 1961 and skewness value of 1.16 in August 2021 indicate both have high positive skewness.

The kurtosis value of 1.68 in August 1961 means it is thick-tailed compared to the kurtosis value of 0.8 in August 2021 means it is thin-tailed.

The line chart of August 1961 shows the minimum value and maximum value of maximum recorded temperature are 12 and 22 degrees C respectively. Most of the observations lie between 14 degrees C and 20 degrees C with the possibility of two outliers.

The line chart of August 2021 shows the minimum value and maximum value of the maximum recorded temperature is 13.5 and 23.5 degrees C. Most of the observations lie between 14 degrees C and 21 degrees C with the possibility of one outlier.

The histogram of both August 1961 and August 2021 follows right-skewed distribution.

The boxplot of August 1961 is nearly symmetrical and has high positive skewness(1.37) possibly due to the presence of two outliers.

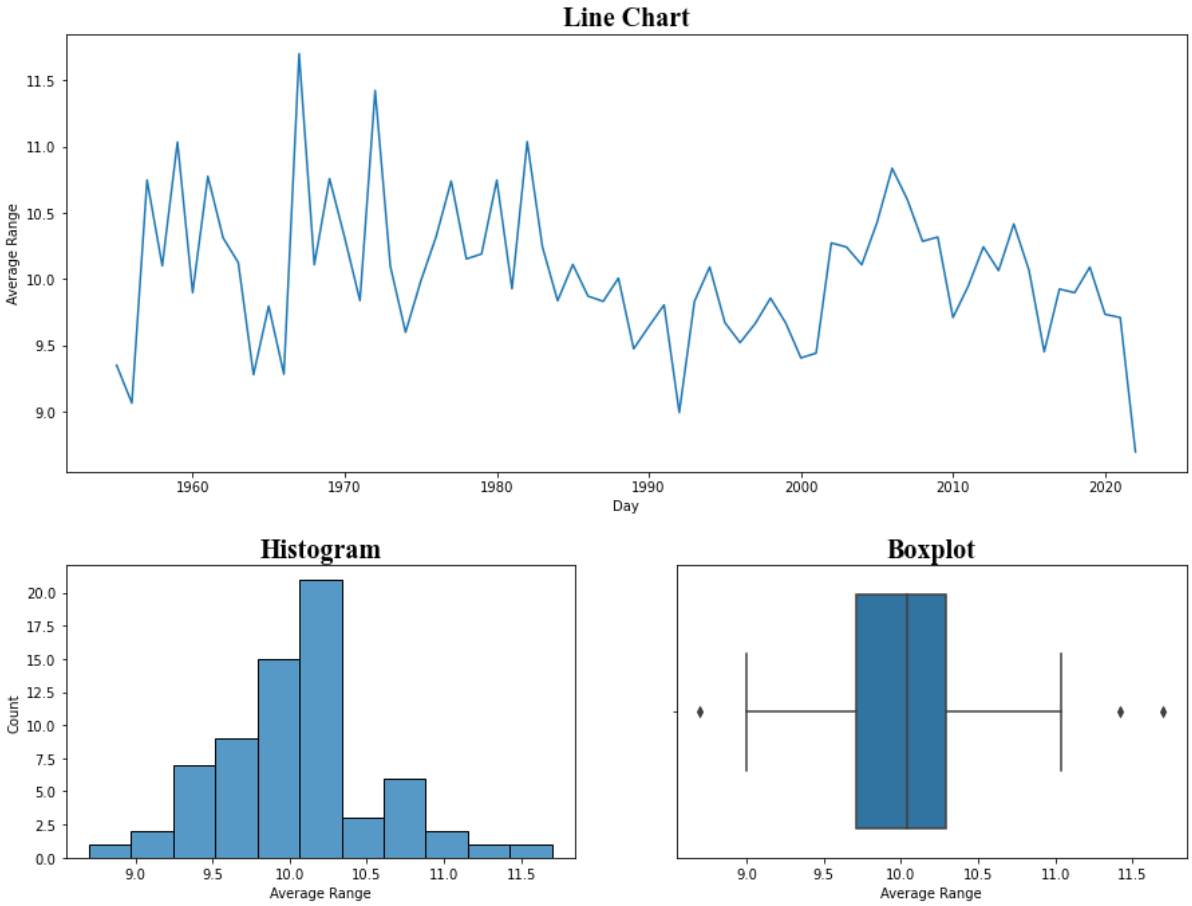
The boxplot of August 2021 is right skewed and has high positive skewness(1.16) possibly due to the presence of one outlier.

In [255…

```
#Creating daily range column to get the average range for each year
merged_data=pd.merge(min_temp,max_temp,on=['Year','Month','Day'],how='inner'
merged_data['Daily range']=merged_data['Maximum temperature (Degree C)']-mer
average_range=merged_data.groupby('Year')['Daily range'].mean().reset_index(
print("Mean: ",round(average_range['Average Range'].mean(),2)) print("Median:
",round(average_range['Average Range'].median(),2)) print("Standard
Deviation: ",round(average_range['Average Range'].std(),2)) print("Skewness:
",round(average_range['Average Range'].skew(),2)) print("Kurtosis:
",round(average_range['Average Range'].kurt(),2)) quartile75, quartile25 =
np.percentile(average_range['Average Range'], [75 , iqr = quartile75 -
quartile25 print("IQR: ",round(iqr,2)) #Plotting the line chart
```

```
fig, ax = plt.subplots(figsize=(15, 6))
ax.plot(average_range["Year"], average_range['Average Range'])
ax.set_title("Line Chart",fontname="Times New Roman",fontweight="bold",size=
plt.xlabel("Day") plt.ylabel("Average Range") plt.show()
#Plotting the histogram and box plot fig1, ax
= plt.subplots(1, 2, figsize=(15,4))
sns.histplot(data=average_range, x='Average Range', ax=ax[0])
ax[0].set_title("Histogram",fontname="Times New Roman",fontweight="bold",siz
sns.boxplot(data=average_range, x='Average Range', ax=ax[1])
ax[1].set_title("Boxplot",fontname="Times New Roman",fontweight="bold",size=
```

```
Mean:  10.04
Median:  10.04
Standard Deviation:  0.54
Skewness:  0.46
Kurtosis:  1.03 IQR:
0.58
```



The mean is 10.04 and the median is 10.02.

The standard deviation is 0.54 and the IQR value is 0.58 which shows that there is a dense distribution present.

The skewness value is 0.46 which indicates it has low positive skewness.

The kurtosis value is 1.03 which shows the distribution is heavy-tailed. From the line chart, we can observe that the highest temperature is 11.8 and the lowest temperature is 8.8 and there is no sufficiently great variation in the temperature over the years.

The histogram of the average range for each year shows that the distribution is right skewed.

The boxplot of the average range indicates that the distribution is right skewed and there is the

```
average_range
```

presence of three outliers.

In [256…

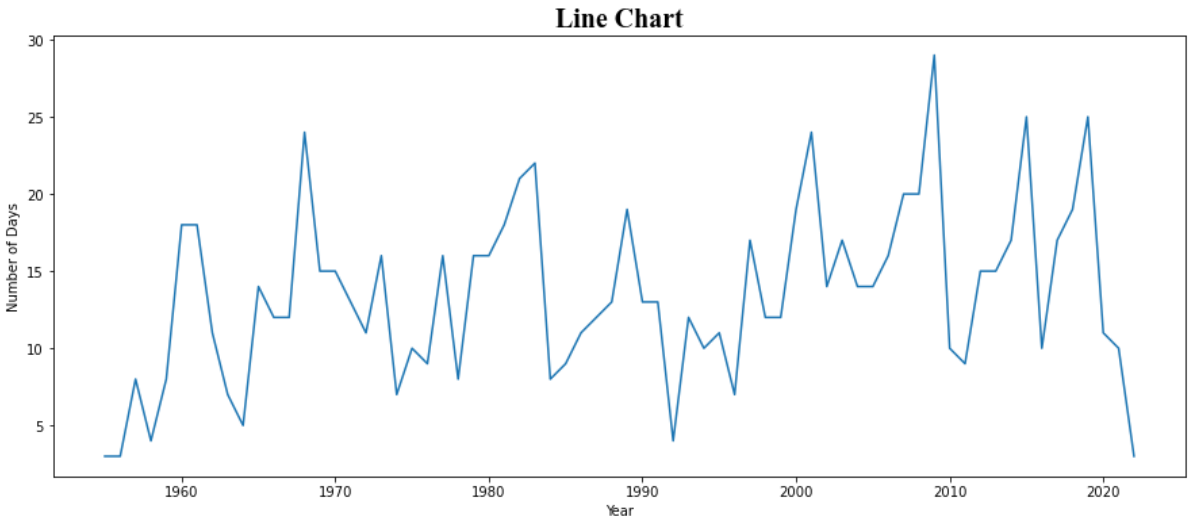
Out[256]:

	Year	Average Range
0	1955	9.350000
1	1956	9.065027
2	1957	10.746575
3	1958	10.100274
4	1959	11.033973
...	...	...
63	2018	9.897808
64	2019	10.090959
65	2020	9.734699
66	2021	9.710411
67	2022	8.697357

68 rows × 2 columns

In [257…

```
#Calculating number of days with temperature greater than 35 degree C
temp_35=merged_data.loc[merged_data['Maximum temperature (Degree C)']>35]
days_with_tempmore_35=temp_35.groupby('Year')['Day'].count().reset_index(nam
fig, ax = plt.subplots(figsize=(15, 6))
ax.plot(days_with_tempmore_35["Year"], days_with_tempmore_35['days_with_temp
ax.set_title("Line Chart",fontname="Times New Roman",fontweight="bold",size=
plt.xlabel("Year") plt.ylabel("Number of Days") plt.show()
print("Mean: ",round(days_with_tempmore_35['days_with_tempmore_35'].mean(),2
print("Median: ",round(days_with_tempmore_35['days_with_tempmore_35'].median
print("Standard Deviation: ",round(days_with_tempmore_35['days_with_tempmore
print("Skewness: ",round(days_with_tempmore_35['days_with_tempmore_35'].skew
print("Kurtosis: ",round(days_with_tempmore_35['days_with_tempmore_35'].kurt
quartile75, quartile25 = np.percentile(days_with_tempmore_35['days_with_temp
iqr = quartile75 - quartile25 print("IQR: ",round(iqr,2))
```



```
Mean:  13.47
Median:  13.0
Standard Deviation:  5.7
Skewness:  0.33
Kurtosis:  0.0
IQR:  7.0
```

In [258…
```
days_with_tempmore_35
```

Out[258]:

	Year	days_with_tempmore_35
0	1955	3
1	1956	3
2	1957	8
3	1958	4
4	1959	8
…	…	…
63	2018	19
64	2019	25
65	2020	11
66	2021	10
67	2022	3

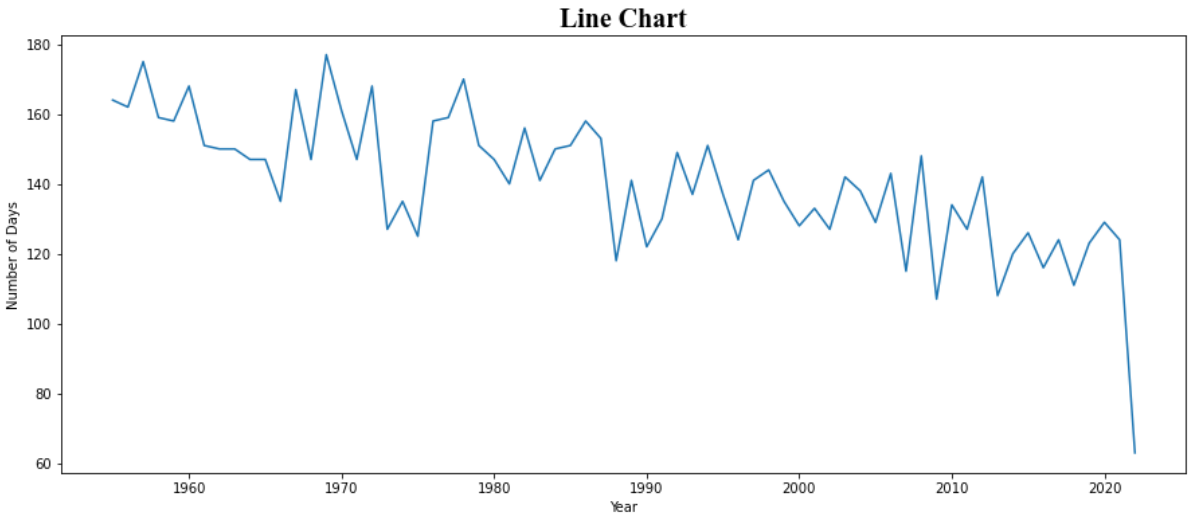
68 rows × 2 columns

The mean is 13.47 and median is 13.0.
The standard deviation is 5.7 and the IQR value is 7 which indicates that observations are spread out.
The skewness value is 0.33 which shows that it has low positive skewness. The kurtosis value is zero which indicates that the distribution is platykurtic which means it is flat tailed.
The minimum number of days with temperature greater than 35 was observed in the year 1955,1956 and 2022 which is 3 days.
The maximum number of days with temperature greater than 35 was observed in the year 2009 which is 29 days.

In [259…

```
# Calculating number of days with temperature less than 10 degree C
temp_10=merged_data.loc[merged_data['Minimum temperature (Degree C)']<10]
days_with_templess_10=temp_10.groupby('Year')['Day'].count().reset_index(nam
fig, ax = plt.subplots(figsize=(15, 6))
ax.plot(days_with_templess_10["Year"], days_with_templess_10['days_with_temp
ax.set_title("Line Chart",fontname="Times New Roman",fontweight="bold",size=
plt.xlabel("Year") plt.ylabel("Number of Days") plt.show()
print("Mean:   ",round(days_with_templess_10['days_with_templess_10'].mean(),2
print("Median:   ",round(days_with_templess_10['days_with_templess_10'].median
print("Standard  Deviation:  ",round(days_with_templess_10['days_with_templess
print("Skewness:   ",round(days_with_templess_10['days_with_templess_10'].skew
print("Kurtosis:   ",round(days_with_templess_10['days_with_templess_10'].kurt
quartile75, quartile25 = np.percentile(days_with_templess_10['days_with_temp
```

```
iqr = quartile75 - quartile25 print("IQR:
",round(iqr,2))
```




```
Mean:  140.29
Median:  141.5
Standard Deviation:  19.15
Skewness:  -0.86
Kurtosis:  2.66 IQR:
24.0
```

The mean is 140.35 and the median is 142.0.

The standard deviation is 19.1 and the IQR value is 24 which indicates that observations are spread out.

The skewness value is -0.87 which shows that it has low negative skewness. The kurtosis value is 2.72 which indicates that it is heavy-tailed.

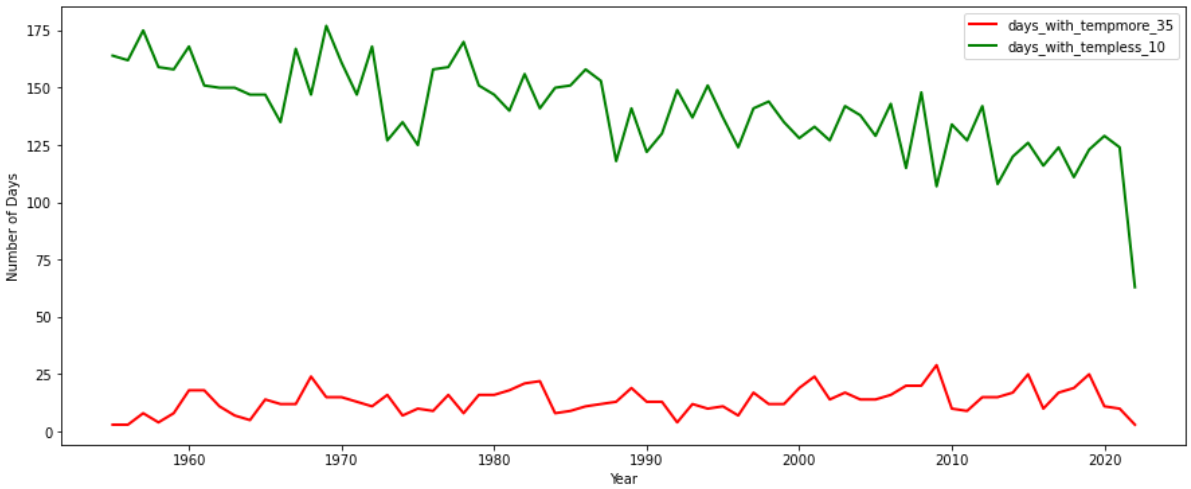
We have taken a threshold of temperature less than 10 degrees C as minimal temperature because World Health Organisation(WHO) considers the temperature as cold.

The minimum number of days with a temperature less than 10 was observed in the year 2022 which is 63 days.

The maximum number of days with a temperature less than 10 was observed in the year 1969 which is 177 days.

In [260…

```
plt.figure(figsize=(15,6))
plt.plot('Year','days_with_tempmore_35', data=days_with_tempmore_35, color='
plt.plot('Year','days_with_templess_10', data=days_with_templess_10, color='
plt.legend() plt.xlabel("Year")
plt.ylabel("Number of Days")
# show graph plt.show()
total_days=pd.merge(days_with_tempmore_35,days_with_templess_10,on='Year',su
total_days['Total']=total_days['days_with_tempmore_35']+ total_days['days_wi
print("Mean: ",round(total_days['Total'].mean(),2)) print("Median:
",round(total_days['Total'].median(),2)) print("Standard Deviation:
",round(total_days['Total'].std(),2)) print("Skewness:
",round(total_days['Total'].skew(),2)) print("Kurtosis:
",round(total_days['Total'].kurt(),2))
quartile75, quartile25 = np.percentile(total_days['Total'], [75 ,25])
iqr = quartile75 - quartile25 print("IQR: ",round(iqr,2))
```




```
Mean:  153.76
Median:  157.0
Standard Deviation:  18.92
Skewness:  -1.35
Kurtosis:  5.64
IQR:  22.5
```

In [261…

```
total_days
```

Out[261]:

	Year	days_with_tempmore_35	days_with_templess_10	Total
0	1955	3	164	167
1	1956	3	162	165
2	1957	8	175	183
3	1958	4	159	163
4	1959	8	158	166
…	…	…	…	…
63	2018	19	111	130
64	2019	25	123	148

65	2020	11	129	140
66	2021	10	124	134
67	2022	3	63	66

68 rows × 4 columns

The mean is 153.82 and median is 157.0.
The standard deviation is 18.88 and the IQR value is 22.5 which indicates that observations are spread out.
The skewness value is -1.36 which shows that it has high negative skewness.
The kurtosis value is 5.72 which indicates that it is heavy tailed.
The minimum number of days with extreme temperature was observed in the year 2022 which is 66 days.
The maximum number of days with extreme temperature less than 10 was observed in the year 1969 which is 192 days From the line chart, we can see that number of days with temperature less than 10 degree C are gradually decreasing over the years and number of days with temperature more than 35 degree C over the years does not see significant variation.
The number of days with temperature less than 10 degree C over the years are decreasing indicating climate change as the temperature is rising with increase of minimal temperature of the days.

In this report, we have done a climate analysis of the changes in temperature over the Adelaide airport from the year January 1955 to August 2022. The datasets are provided with information on maximum and minimum temperature on each day over the years. From the datasets, we found out about the days when the maximum and minimum temperatures were recorded.

We made a function to calculate descriptive statistics for a given month and year and we run it for August 1961 and August 2021 and found out that the mean(15.64) and median(15.0) have increased from August 1961 to mean(16.51) and median(15.5) of August 2021. The standard deviation and IQR of August 1961 are 2.41 and 2.55 respectively and for August 2021 the standard deviation is 2.58 and IQR is 2.55 respectively which indicates the observations are almost equally spread out.

We merged the data sets and calculated the range between maximum temperature and minimum temperature for each day and calculated the average range for each year and found out that the most observations lies between 9 degrees C and 11 degrees C with few outliers and is densely distributed.

We have calculated the number of hot days(Temperature greater than 35 degrees C) and cold days(Temperature less than 10 degrees C) observed for each year and after merging both the results we found that the number of cold days(Temperature less than 10 degrees C) is gradually decreasing over the years. We can conclude that the climate is becoming hotter and there is an indication of global warming happening.