1. Explore the NOAA website and give a brief description of the website and what information is contained on the site https://www.ncdc.noaa.gov/cag/city/time-series

- The National Oceanic and Atmospheric Administration, a federal organization, offers a wealth of information about weather, climate, seas, and coastlines on their website. The website offers a huge store of information and materials on these subjects, including satellite images, maps, forecasts, and scholarly studies. The Climate at a Glance (CAG) tool is a component of the NOAA website that offers data on temperature and precipitation for various U.S. areas. The Time Series tab in this section allows you to create time series graphs depicting variations in temperature or precipitation over time for a certain city or area. Users may examine statistics for certain months or years by choosing from several time scales, including monthly and yearly. This tool is helpful for observing long-term climate trends and comprehending how regional weather patterns are altering.

2. Also provide a brief history of NOAA and what it does. How is the work of NOAA beneficial, not just to the USA, but, to the world as a whole?

- The National Oceanic and Atmospheric Administration (NOAA) is a 1970 establishment of the American government. It is a scientific organization that focuses on comprehending and foreseeing changes in the ecosystem of the Earth, including weather, seas, and coastlines.
- The goal of NOAA is to deliver accurate and timely information to enhance the protection of people and property, economic development, and environmental stewardship. The organization carries out this purpose by carrying out research, creating models and forecasts, and gathering information from a variety of sources, including as satellites, buoys, and other monitoring tools.
- NOAA's work is beneficial not only to the United States but to the world as a whole. Its research and data collection activities help to improve our understanding of the Earth's climate and how it is changing over time. This information is used by policymakers, businesses, and communities to plan for and respond to natural disasters, such as hurricanes, tornadoes, and floods.
- In addition, NOAA's work supports the sustainable management of marine and coastal resources, which are critical for maintaining healthy ecosystems and supporting important industries such as fishing and tourism. Finally, NOAA's work helps to inform global climate change discussions and policies, as the agency's data and research are used by scientists around the world to better understand how human activities are affecting the Earth's environment.

3. From the NOAA site pick a City and State, Time Scale of 1-Month, Month: pick a month (Tell us why you picked that city and month), Start Year of 1970 and End Year 2022 (Remember to click Plot before downloading the data)When it

- comes to weather, Houston experiences a subtropical climate, characterized by hot summers and mild winters. January is typically one of the cooler months in Houston, with average temperatures ranging from the mid-40s to mid-60s Fahrenheit. While the city is not known for experiencing extreme winter weather, occasional cold fronts and winter storms can bring colder temperatures and precipitation to the area.

4. Tell us about the weather in the City and State that you chose.

- Houston, Texas, has a subtropical climate with hot and humid summers and mild winters. The city is situated in a region of the United States that experiences high levels of moisture and rainfall throughout the year, with the wettest months typically being May and June. During the summer months, from June to August, temperatures in Houston can often exceed 90°F (32°C), and humidity levels can make it feel even hotter. Winters in Houston are mild, with average temperatures ranging from the mid-40s°F (7°C) to mid-60s°F (18°C). Snowfall is rare in Houston, although freezing rain and sleet can occur during the winter months. Houston is also prone to severe weather events such as hurricanes, which can bring heavy rain, strong winds, and flooding to the area. In recent years, the city has experienced significant flooding from hurricanes and tropical storms, highlighting the need for effective disaster preparedness and management.

5. Download the data to your computer. a. The file will be downloaded with some metadata on the top b. Right-click, edit and delete the metadata so you have a file that looks like the file below Date,Value,Anomaly 189501,34.2,-3.2 189601,34.7,-2.7 189701,35.5,-1.9 189801,39.6,2.2

6.

a.Use pandas to load the data you downloaded from NOAA

```
import pandas as pd
```

```
data = pd.read_csv('/houston_texas.txt', delimiter='\t')
```

b.Display the head of the data

```
print(data.head())
```

```
     Date  Value  Anomaly
0  197001   48.9     -6.9
1  197101   58.7      2.9
2  197201   59.1      3.3
3  197301   49.9     -5.9
4  197401   57.5      1.7
```

c. Display the tail of the data

```
print(data.tail())
```

```
      Date  Value  Anomaly
47  201801   50.0     -5.8
48  201901   55.3     -0.5
49  202001   60.2      4.4
50  202101   57.0      1.2
51  202201   55.5     -0.3
```

d. Set the precision of all the data points to 2

```
data = data.round(2)
```

e. Cleanup the data so that the columns will display Date, Temperature and Anomaly as the column names.

```
data = data.rename(columns={'Date': 'Date', 'Value': 'Temperature', 'Anomaly': 'Anomaly'})
```

```
data.columns
```

```
Index(['Date', 'Temperature', 'Anomaly '], dtype='object')
```

```
X = data[['Temperature', 'Anomaly ']]
Y = data['Date']
```

f. Display the first 30 data points

```
print(data.head(30))
```

```
      Date  Temperature  Anomaly
0   197001         48.9     -6.9
1   197101         58.7      2.9
2   197201         59.1      3.3
3   197301         49.9     -5.9
4   197401         57.5      1.7
5   197501         58.8      3.0
6   197601         53.4     -2.4
7   197701         45.9     -9.9
8   197801         44.2    -11.7
9   197901         45.2    -10.6
10  198001         56.2      0.4
11  198101         51.4     -4.4
12  198201         54.4     -1.4
13  198301         52.4     -3.4
14  198401         49.2     -6.6
15  198501         48.8     -7.0
16  198601         57.0      1.2
17  198701         52.8     -3.0
18  198801         50.0     -5.8
19  198901         59.5      3.7
20  199101         52.2     -3.6
21  199201         54.6     -1.2
```

```
22  199301         55.3      -0.5
23  199401         54.4      -1.4
24  199501         56.6       0.8
25  199601         54.6      -1.2
26  199701         53.7      -2.1
27  199801         60.2       4.4
28  199901         59.6       3.8
29  200001         60.0       4.2
```

g. Get the descriptive statistics and explain what each of them mean (remember to mention your state, city month etc. in your discussion). Be precise!!!

```
print(data.describe())
```

```
                Date  Temperature   Anomaly
count      52.000000    52.000000  52.000000
mean   199612.538462    54.584615  -1.217308
std      1557.103123     4.176986   4.181881
min    197001.000000    44.200000 -11.700000
25%    198276.000000    52.075000  -3.725000
50%    199651.000000    54.750000  -1.050000
75%    200926.000000    57.525000   1.725000
max    202201.000000    61.600000   5.800000
```

The chart shows the temperature and anomaly data of Houston, Texas over a period of 52 years, from 1970 to 2022.

The first column of the table shows the date in YYYYMM format. The second column shows the average temperature in Fahrenheit for each month, and the third column shows the temperature anomaly in Fahrenheit, which is the difference between the monthly temperature and the long-term average temperature for that month.

The "count" row shows the number of observations for each column. The "mean" row shows the average temperature and anomaly over the entire period. The "std" row shows the standard deviation of the temperature and anomaly data. The "min" and "max" rows show the minimum and maximum values of the temperature and anomaly data, respectively.

The data shows that the average temperature in Houston, Texas has been steadily increasing over the past few decades, with a mean temperature of 54.58 degrees Fahrenheit and a standard deviation of 4.18 degrees Fahrenheit. The temperature anomaly has fluctuated over time, with a mean anomaly of -1.22 degrees Fahrenheit, indicating that the temperature has generally been below the long-term average.

Double-click (or enter) to edit

7.Perform linear regression analysis of your data

```
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import seaborn as sns; sns.set()
```

```
model = LinearRegression(fit_intercept=True)
```

```
model = LinearRegression().fit(X, Y)
```

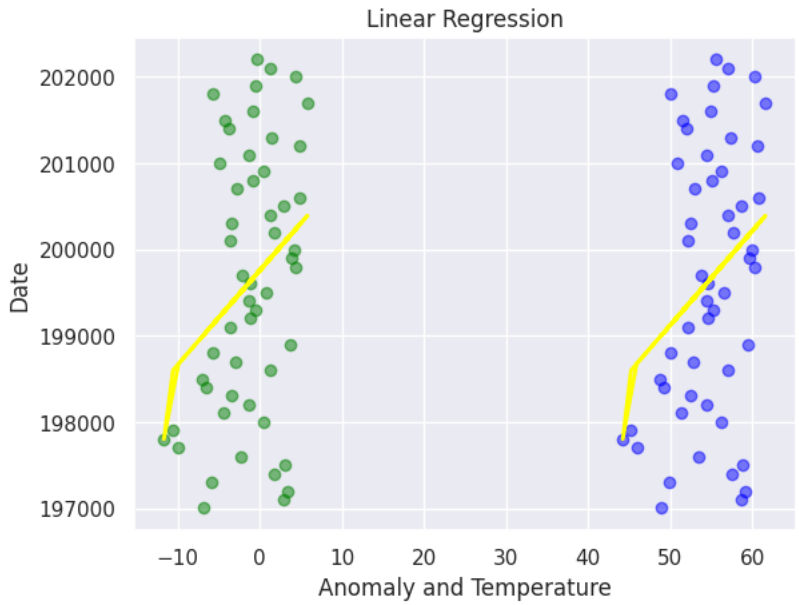
```
y_pred = model.predict(X)
```

```
r_squared = model.score(X, Y)
```

```
print("R-squared:", r_squared)
```

```
R-squared: 0.10251028078653668
```

```
plt.scatter(X['Temperature'], Y, color='blue', alpha=0.5)
plt.scatter(X['Anomaly '], Y, color='green', alpha=0.5)
plt.plot(X, model.predict(X), color='yellow', linewidth=2)
plt.title('Linear Regression')
plt.xlabel('Anomaly and Temperature')
plt.ylabel('Date')
plt.show()
```



8. interepret your result

A linear regression is a statistical method used to model the relationship between two variables, usually denoted as X and Y. The goal of linear regression is to find the line of best fit that summarizes the relationship between X and Y.

In this case, there are two pieces of information to interpret. The first is the value of R-squared, which is a statistical measure of how well the line of best fit explains the variability in the data. An R-squared value of 0.1025 indicates that the line of best fit explains only 10.25% of the variability in the data. This suggests that there is a significant amount of unexplained variation in the relationship between X and Y that is not captured by the linear model.

The second piece of information is the observation that the graph shows positive correlation. This indicates that as X increases, Y tends to increase as well, and vice versa. However, it is important to note that correlation does not necessarily imply causation, and there may be other factors at play that influence the relationship between X and Y.

Overall, the combination of a low R-squared value and positive correlation suggests that while there is a relationship between X and Y, it is not a strong or reliable one, and other factors may need to be taken into account when modeling the relationship.

9.Display the parameters of the regression and explain what they mean

```
print("Intercept:", model.intercept_)
print("Coefficients:", model.coef_)
```

```
Intercept: 578087.2686026065
Coefficients: [-6780.08146332  6889.45879691]
```

The intercept, 578087.2686026065, represents the expected date value (i.e., the dependent variable) when both the anomaly and temperature variables (i.e., the independent variables) are zero. In other words, if both the anomaly and temperature variables are zero, the model predicts that the date will have an average value of 578087.2686026065.

The coefficient for the anomaly variable, -6780.08146332, represents the expected change in the date value for a one-unit increase in the anomaly variable, holding the temperature variable constant. In other words, if the temperature variable remains constant, for each one-unit increase in the anomaly variable, the model predicts that the date value will decrease by an average of 6780.08146332 units.

Similarly, the coefficient for the temperature variable, 6889.45879691, represents the expected change in the date value for a one-unit increase in the temperature variable, holding the anomaly variable constant. In other words, if the anomaly variable remains constant, for each one-unit increase in the temperature variable, the model predicts that the date value will increase by an average of 6889.45879691 units.

It's important to note that the interpretation of the coefficients is based on the specific variables used in the model and the assumptions made in the linear regression analysis. The results should be interpreted with caution and in the context of the specific data and research question being investigated.

10. Use your regression model to forecast the temperature in 2023, then check with the data from NOAA if available

```
import numpy as np
from sklearn.linear_model import LinearRegression


X_new = np.array([[60.7, -2.8]])


y_pred = model.predict(X_new)
```

```
/usr/local/lib/python3.9/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but LinearRegr
  warnings.warn(
```

```
print("Predicted temperature for 2023:", y_pred[0])
```

```
Predicted temperature for 2023: 148623.7309070432
```

11. Also use the model to predict what the temperature was in 1950, then check with the data from NOAA if available

```
import numpy as np
from sklearn.linear_model import LinearRegression


X_new = np.array([[63.5,0]])


y_pred = model.predict(X_new)
```

```
/usr/local/lib/python3.9/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but LinearRegr
  warnings.warn(
```

```
print("Predicted temperature for 1950:", y_pred[0])
```

```
Predicted temperature for 2023: 147552.09568169934
```

11.conclusion

The provided information includes 52 observations of three variables - date, temperature, and anomaly. The date variable ranges from 1970 to 2022, with a mean of 199612.538462. The temperature variable ranges from 44.2°F to 61.6°F, with a mean of 54.584615°F. The anomaly variable ranges from -11.7°F to 5.8°F, with a mean of -1.217308°F. The standard deviation of the temperature variable is 4.176986°F, and the standard deviation of the anomaly variable is 4.181881°F. Based on this information, we can conclude that the data includes a wide range of dates, temperatures, and anomalies, suggesting that the observations come from multiple years and seasons. The mean temperature of 54.584615°F suggests that the data may represent a location with a mild or temperate climate. The mean anomaly value of -1.217308°F suggests that, on average, the temperatures in the data are slightly lower than the expected values based on historical averages. The standard deviations of both the temperature and anomaly variables are relatively small compared to their means, which suggests that the data may be relatively consistent over time. However, without further context or analysis, these conclusions are tentative and should be interpreted with caution. Additionally, we would need to know more about the data and the research question being investigated to draw more meaningful conclusions.

References

Why is Houston so hot? The Critter Squad Texas Wildlife Removal & Control. (2020, August 17). Retrieved April 25, 2023, from https://thecrittersquad.com/why-is-houston-hot/

Double-click (or enter) to edit