# WeatherPy

## Introduction

Figure 1. Weather Illustration by Divine Designs

During a casual chat with my group of friends our conversation turned towards the hot weather we were experiencing. One of my friends mentioned - it is not as bad as in equator. This led to a fundamental question: ***"What's the weather like as we approach the equator?".*** "Well, everyone knows that it gets hotter", I exclaimed. They immediately challenged me "But, can you prove it?". "Why don't you find data to support your claim", another friend chimed in.
Pushed into a corner, I accepted the challenge and with my skills in Python requests, APIs and JSON traversals I decided to search the world wide web to find supporting evidence. Here is the problem I decided to investigate:

## Problem Statement:

***What is the weather like as we approach the equator?***

## Target Audience:

Apart from my friends, this report will be interesting to ***(a) weather buffs (b) travel agents*** and ***(c) climatologists*** interested in global temperatures.

## API calls & Libraries used:

1. We created a random list 1500 latitude and longitude coordinates from a uniform distribution spread over the full range of latitude and longitude values.

2. A Python library ***(https://pypi.python.org/pypi/citypy)*** was employed to look for nearest cities with randomly generated geo-coordinates.

3. Weather data for each city was collected by making API calls to **OpenWeatherMap** ***(https://openweathermap.org/api)*** which delivers 2 billion forecasts per day.

## Data collection:

A "try-except" loop was used to retrieve data through successive API calls. Errors detected during execution (exceptions) were handled by the except portion of the loop by printing a running list.

# Data Analysis:

- From a list of random coordinates, ***Citypy*** library found 636 cities
- To study the response of the openweathermap API, we passed the geo-coordinates of Austin. Collected data is shown in **Table 1.**
- Confident of making API calls and handling the response json file, we made successive calls for every city obtained by Citypy.
- Among 636 cities, openweathermap found weather data for ***583 cities*** and skipped 53 cities.
- As shown in Table 2, weather data obtained through API calls was stored in a dataframe ***(df_weather)***
- The millisecond-formatted timestamp in the response file was converted to a readable format in the dataframe
- Temperature in Kelvin format was converted to Fahrenheit in the dataframe.
- The data was exported as a csv file (Cities.csv) in the *Output_Data* folder.

| | Parameter | Austin |
|---|---|---|
| 0 | Country Name | US |
| 1 | Date | 1592328580 |
| 2 | Latitude | 30.27 |
| 3 | Longitude | -97.74 |
| 4 | Current Temp ($^o$F) | 87.836 |
| 5 | Humidity | 52 |
| 6 | Pressure | 1021 |
| 7 | Max Temp ($^o$F) | 89.996 |
| 8 | Min Temp ($^o$F) | 86 |
| 9 | Cloudiness | 20 |
| 10 | Wind speed | 3.1 |

**Table 1. Austin weather collected using openweathermap API**

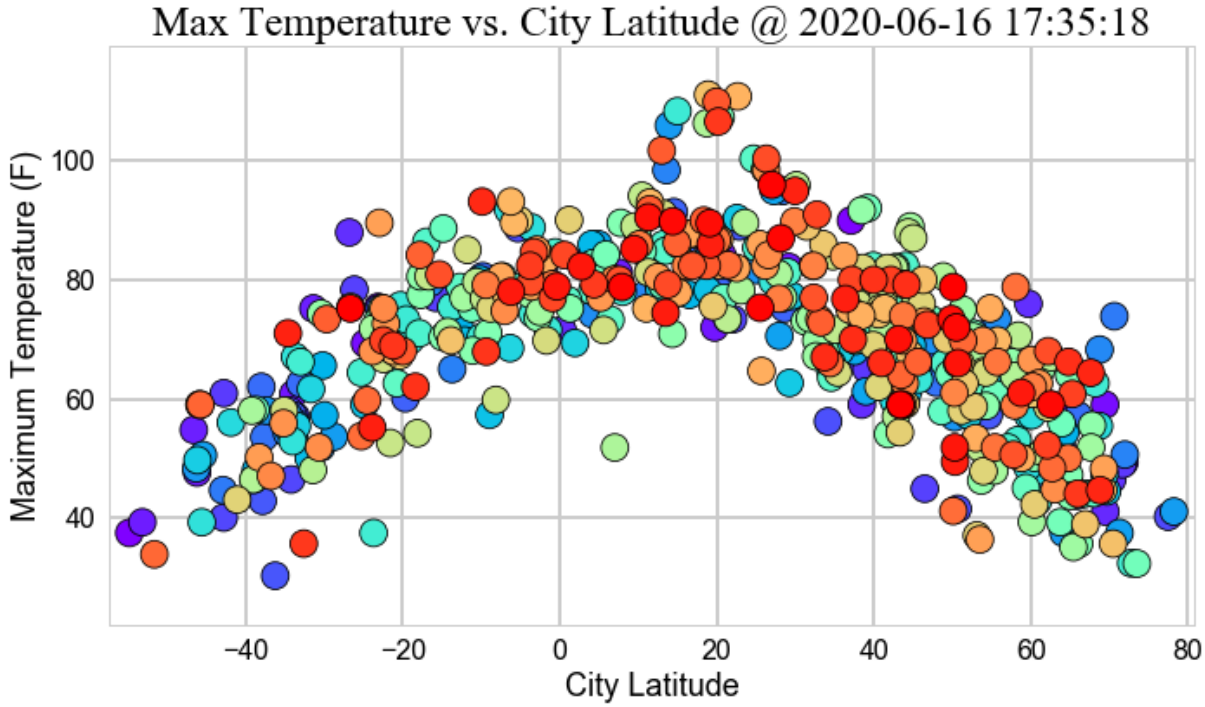
```
The shape of the df_weather dataframe is: (583, 10)
```

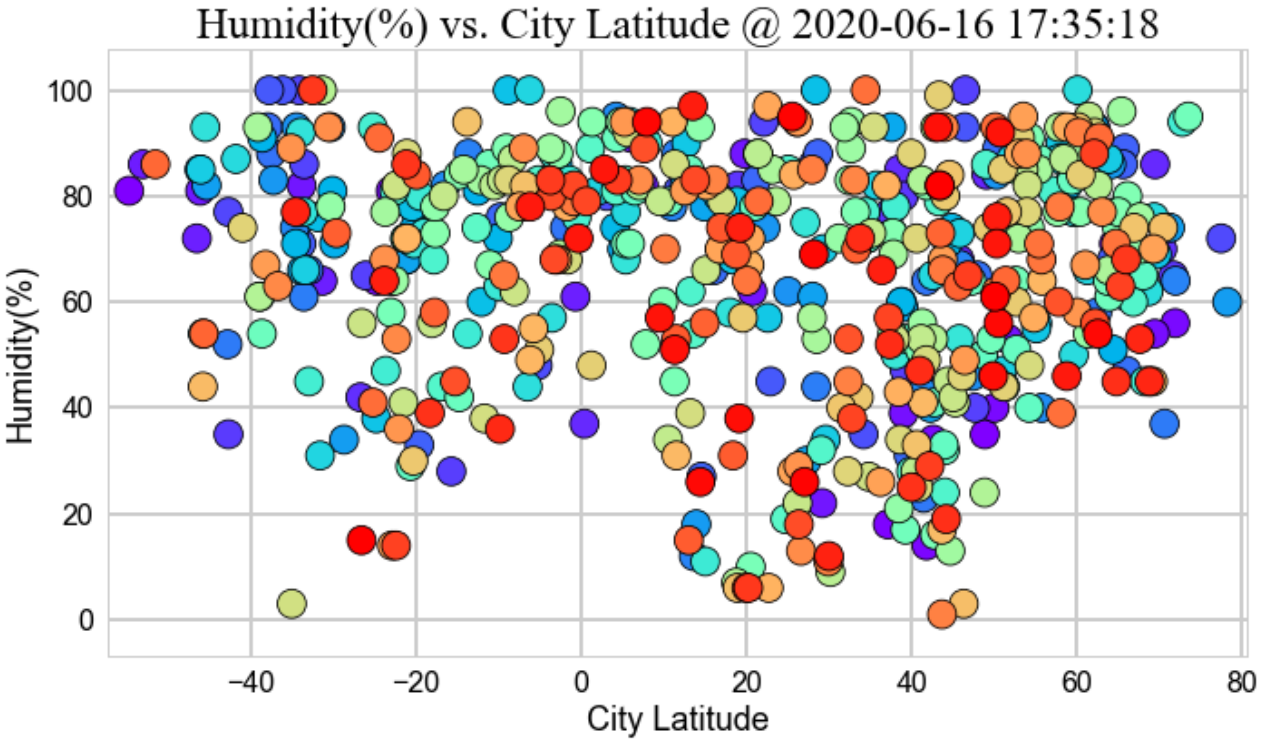
| | City | Country | Date | Latitude | Longitude | Humidity | Pressure | Max_Temp | Cloudiness | Wind_Speed |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Ushuaia | AR | 2020-06-16 17:35:18 | -54.80 | -68.30 | 81 | 1004 | 37.40 | 40 | 2.10 |
| 1 | Hare Bay | CA | 2020-06-16 17:35:18 | 48.85 | -54.01 | 35 | 1021 | 71.60 | 75 | 6.20 |
| 2 | Wolfsburg | DE | 2020-06-16 17:35:07 | 52.43 | 10.80 | 73 | 1013 | 71.01 | 58 | 1.50 |
| 3 | Klaksvík | FO | 2020-06-16 17:35:19 | 62.23 | -6.59 | 93 | 1017 | 51.80 | 100 | 5.10 |
| 4 | Saint George | US | 2020-06-16 17:34:00 | 37.10 | -113.58 | 18 | 1011 | 90.00 | 1 | 12.30 |
| 5 | Rawlins | US | 2020-06-16 17:35:19 | 41.79 | -107.24 | 14 | 1015 | 72.00 | 1 | 11.30 |
| 6 | Rikitea | PF | 2020-06-16 17:35:19 | -23.12 | -134.97 | 81 | 1020 | 75.45 | 22 | 5.74 |
| 7 | Saskylakh | RU | 2020-06-16 17:35:19 | 71.92 | 114.08 | 56 | 999 | 49.30 | 0 | 4.68 |
| 8 | Port-Cartier | CA | 2020-06-16 17:35:19 | 50.03 | -66.87 | 40 | 1020 | 69.80 | 75 | 5.10 |
| 9 | Albany | US | 2020-06-16 17:35:02 | 42.60 | -73.97 | 34 | 1026 | 80.01 | 0 | 0.93 |

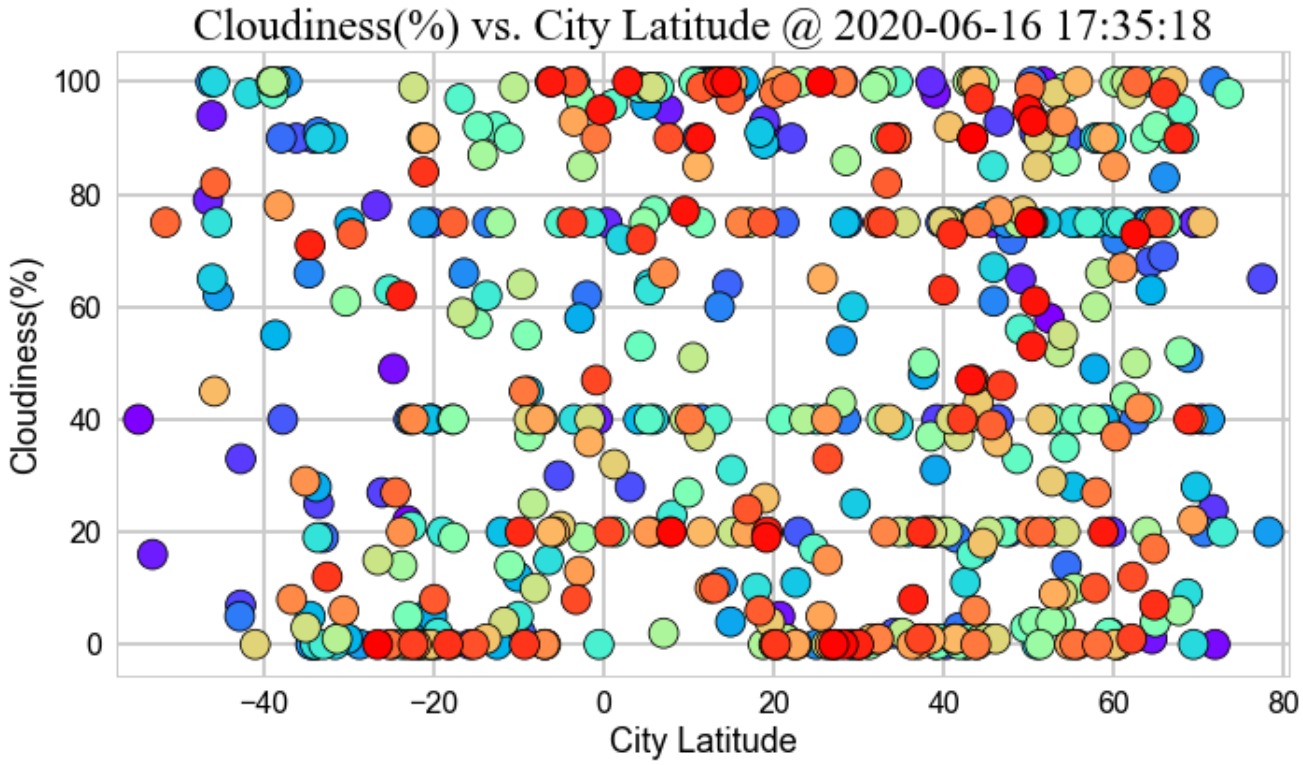
**Table 2. Dataframe of weather data for 583 cities**

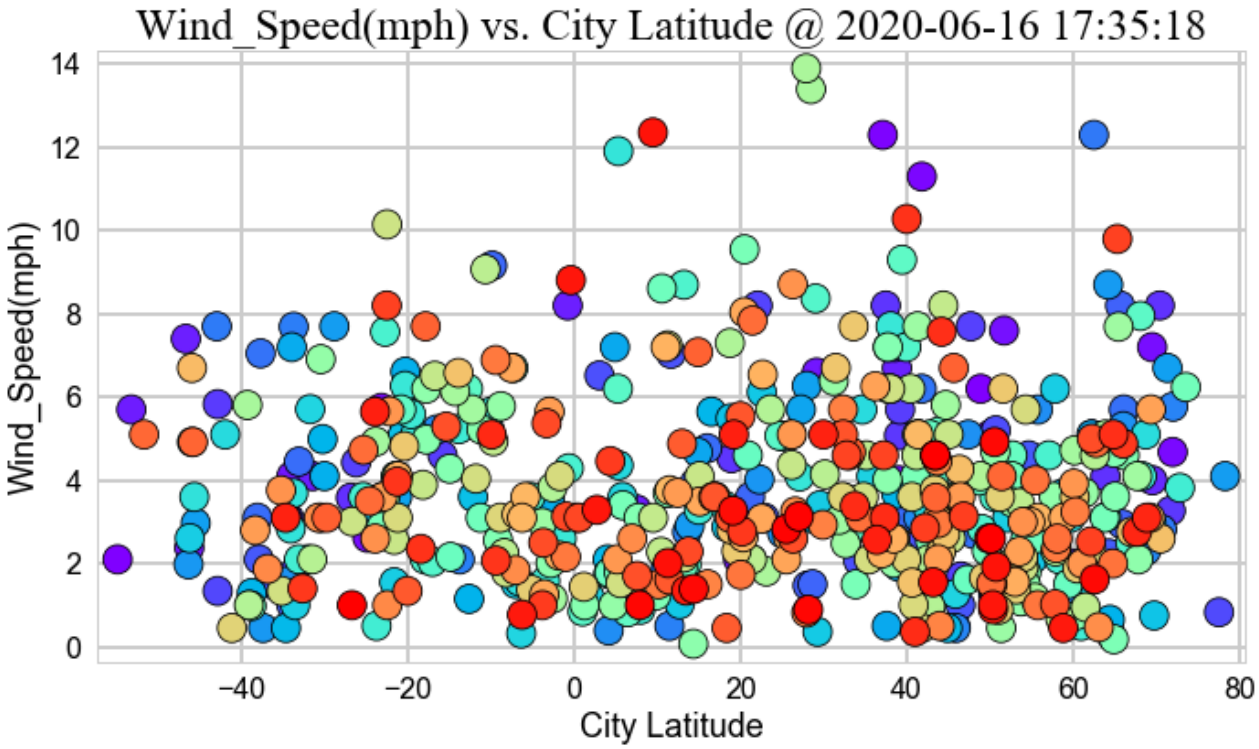
## Scatter Plots:

A function ***(ScatterPlot)*** was created whose inputs are (a) *x & y datasets*, (b) *x & y labels*, (c) *Plot Title* and (d) *Output file name*. ***The function appends the title string with a current date & time stamp for each plot.***




Cloudiness(%) vs. City Latitude @ 2020-06-16 17:35:18
Cloudiness(%)
City Latitude
100
80
60
40
20
0
−40
−20
0
20
40
60
80


Wind_Speed(mph) vs. City Latitude @ 2020-06-16 17:35:18
Wind_Speed(mph)
City Latitude
14
12
10
8
6
4
2
0
−40
−20
0
20
40
60
80

## Conclusions:

### ❖ The maximum temperature is hottest near the Tropic of Cancer:

Tropic of Cancer also known as Northern Tropic is 23.43663° north of equator. This data was taken close to June solstice when the Sun is directly overhead this latitude. Maximum temperature was lowest near the north and south poles.

### ❖ Maximum Humidity (%) is un-related to latitude:

Southern Hemisphere (winter season) has lower average humidity than the Northern Hemisphere (summer season)

### ❖ Maximum Cloudiness (%) is un-related to latitude:

More cloud cover in the Northern Hemisphere (latitudes > 40) as compared to cloud cover in the Southern Hemisphere (latitudes < 40)

### ❖ Average Wind Speed (mph) is un-related to latitude:

Highest wind speed is observed near the Northern Tropic where the maximum temperature is the highest. Summer season is also the onset of hurricanes and atmospheric disturbances which result in wind speeds higher in the Northern Hemisphere as compared to Southern Hemisphere.

## Linear Regression Model:

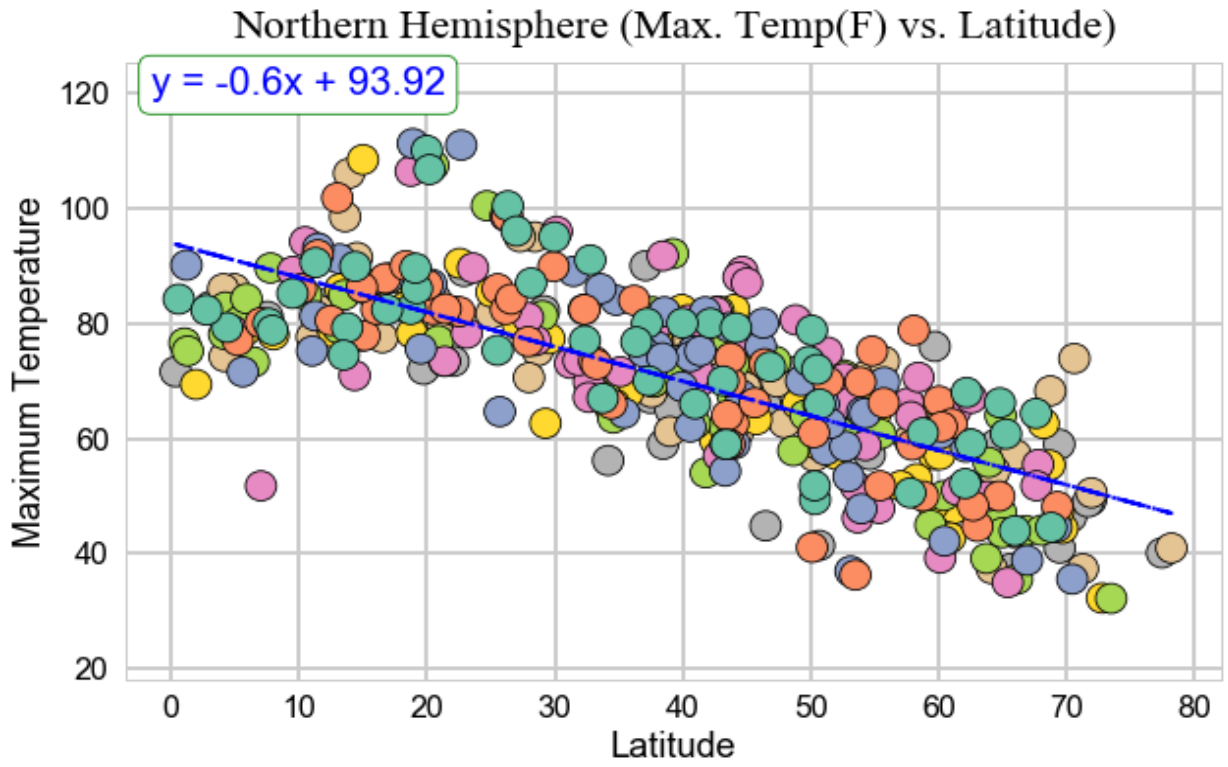
Before creating a linear regression model, the dataframe was divided into two sets: (a) *Northern Hemisphere (Latitude > 0)* and (b) *Southern Hemisphere (Latitude < 0)*. First 5 rows of Northern and Southern Hemisphere dataframe is shown below:

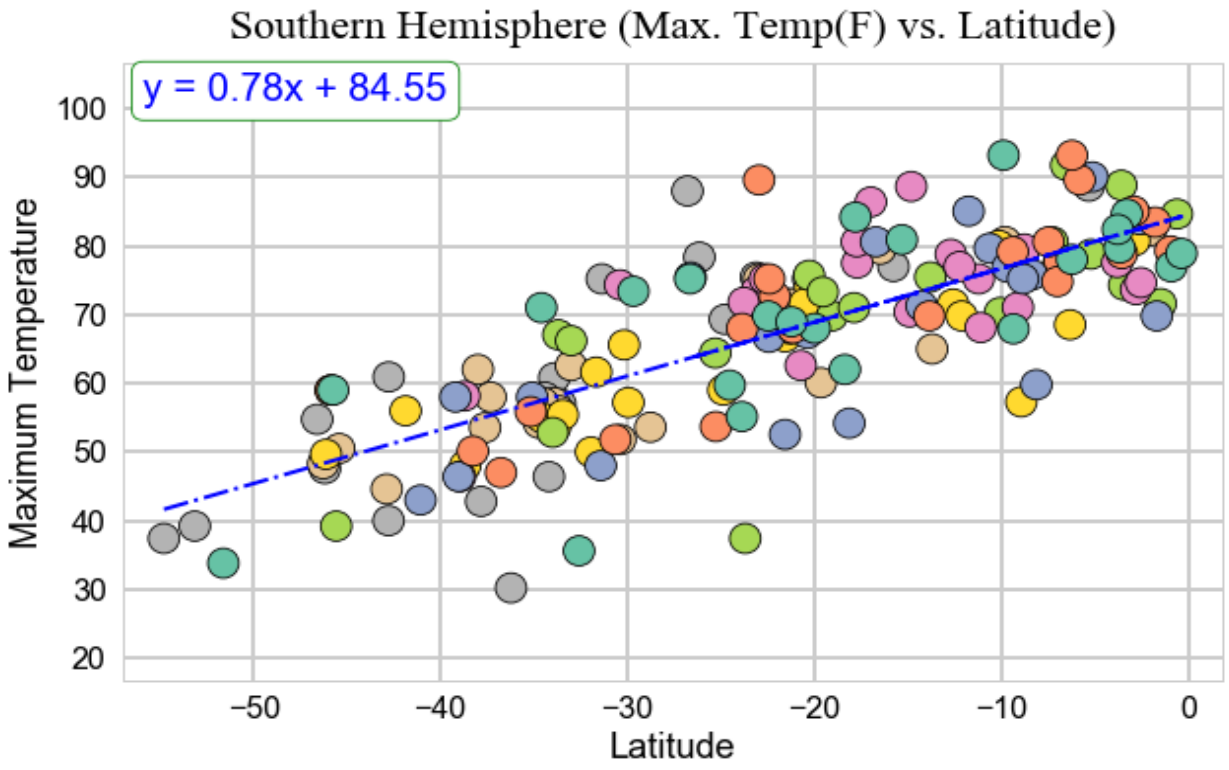
| | City | Country | Date | Latitude | Longitude | Humidity | Pressure | Max_Temp | Cloudiness | Wind_Speed |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Hare Bay | CA | 2020-06-16 17:35:18 | 48.85 | -54.01 | 35 | 1021 | 71.60 | 75 | 6.2 |
| 1 | Wolfsburg | DE | 2020-06-16 17:35:07 | 52.43 | 10.80 | 73 | 1013 | 71.01 | 58 | 1.5 |
| 2 | Klaksvík | FO | 2020-06-16 17:35:19 | 62.23 | -6.59 | 93 | 1017 | 51.80 | 100 | 5.1 |
| 3 | Saint George | US | 2020-06-16 17:34:00 | 37.10 | -113.58 | 18 | 1011 | 90.00 | 1 | 12.3 |
| 4 | Rawlins | US | 2020-06-16 17:35:19 | 41.79 | -107.24 | 14 | 1015 | 72.00 | 1 | 11.3 |

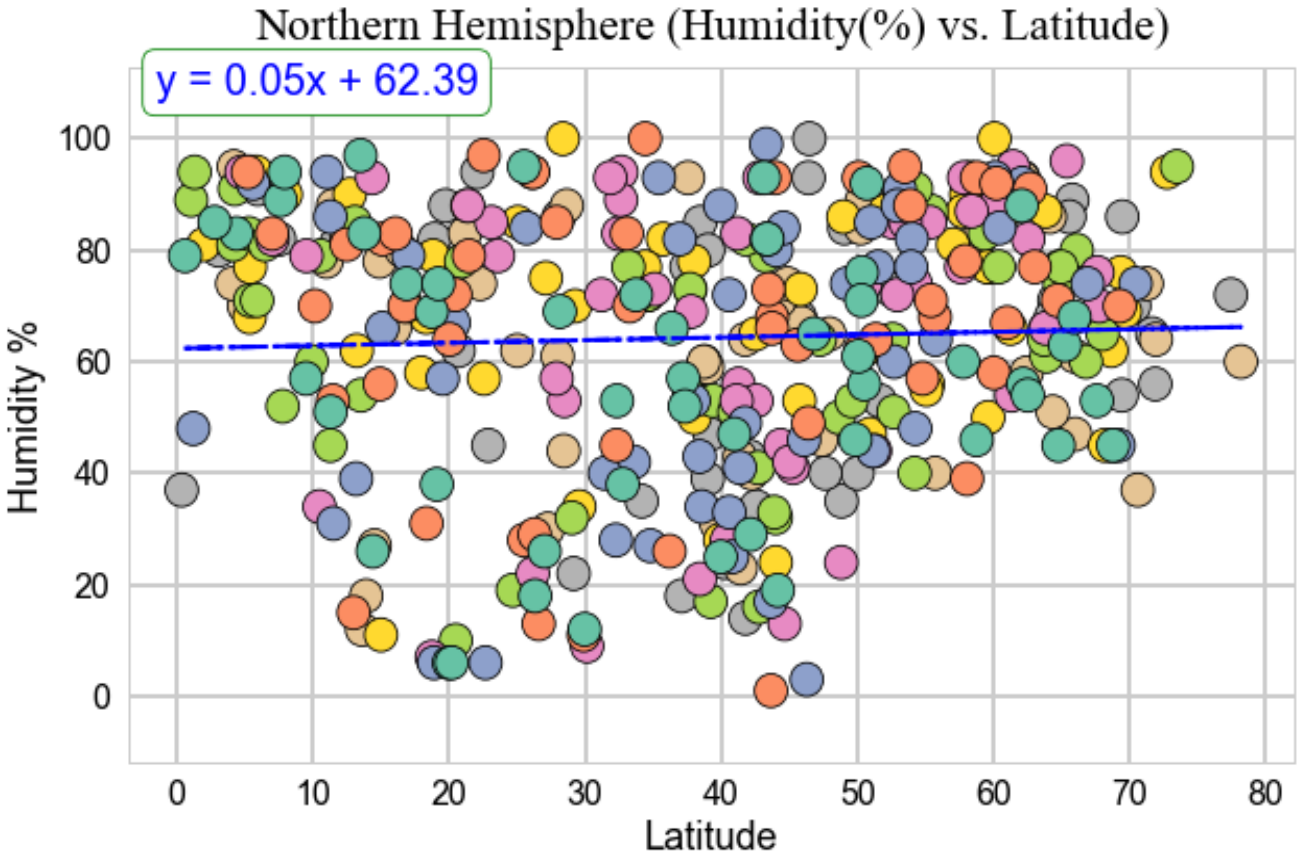
| | City | Country | Date | Latitude | Longitude | Humidity | Pressure | Max_Temp | Cloudiness | Wind_Speed |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Ushuaia | AR | 2020-06-16 17:35:18 | -54.80 | -68.30 | 81 | 1004 | 37.40 | 40 | 2.10 |
| 1 | Rikitea | PF | 2020-06-16 17:35:19 | -23.12 | -134.97 | 81 | 1020 | 75.45 | 22 | 5.74 |
| 2 | Castro | BR | 2020-06-16 17:35:20 | -24.79 | -50.01 | 64 | 1023 | 69.31 | 49 | 2.62 |
| 3 | Port Elizabeth | ZA | 2020-06-16 17:35:21 | -33.92 | 25.57 | 82 | 1033 | 60.80 | 0 | 4.10 |
| 4 | Mataura | NZ | 2020-06-16 17:35:21 | -46.19 | 168.86 | 81 | 1013 | 47.57 | 94 | 2.38 |

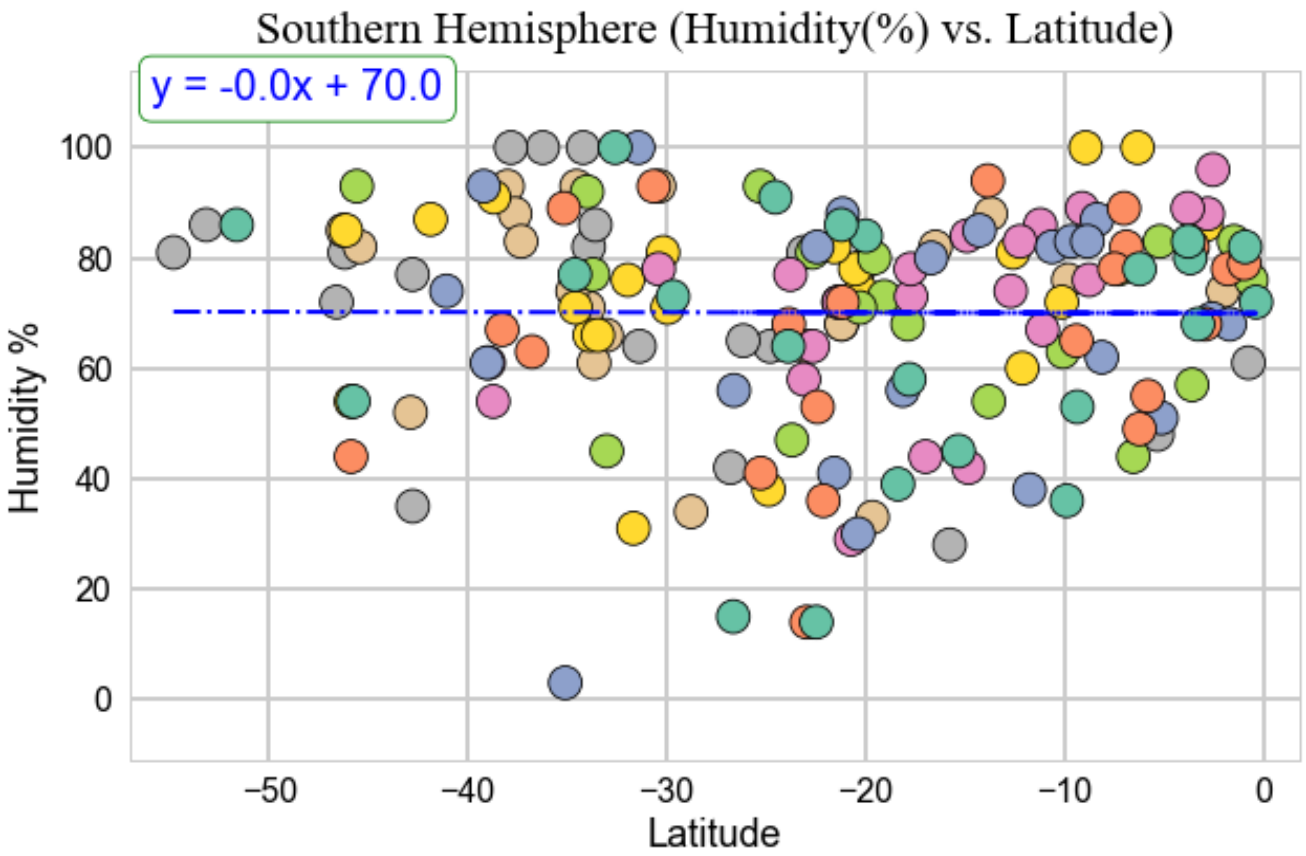
# Scatter Plots with Regression:

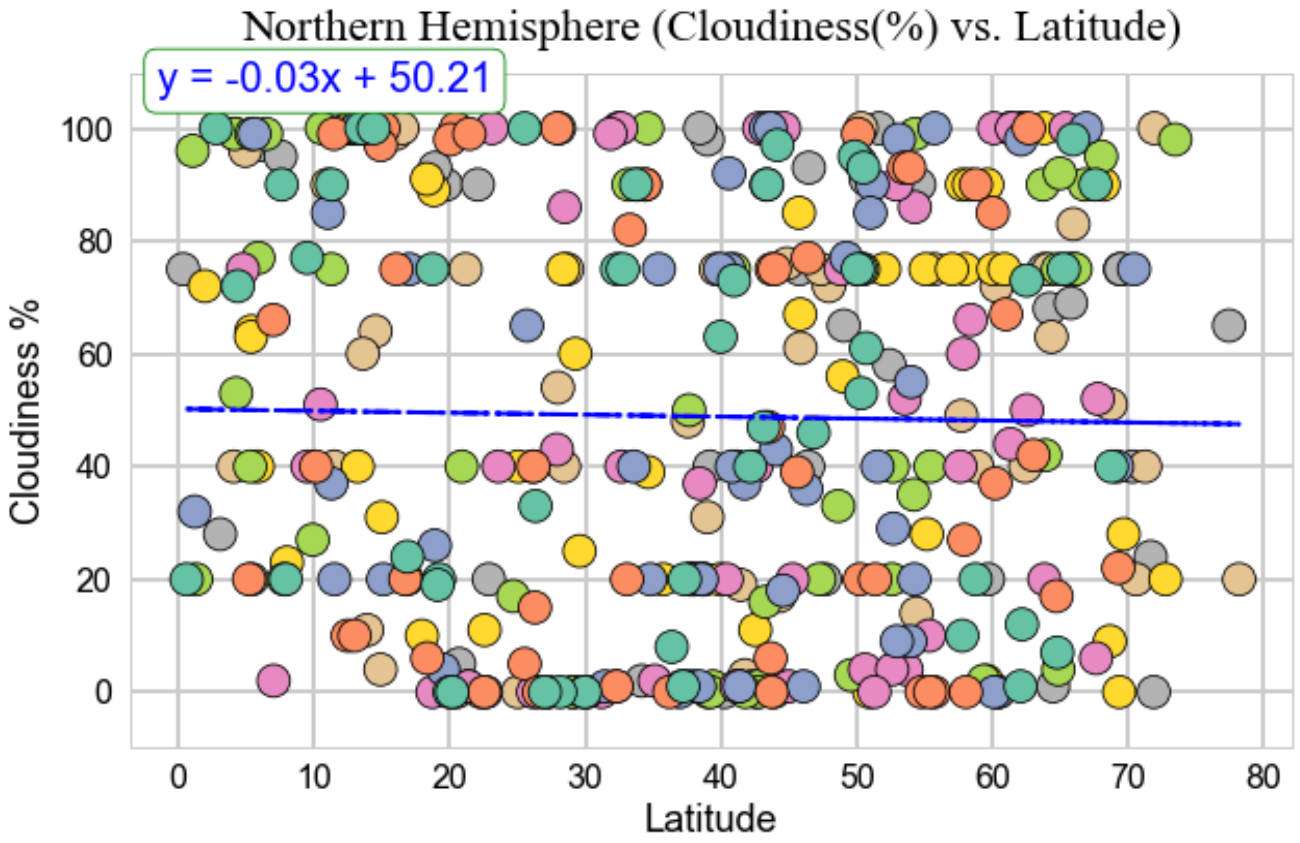
A function ***(LinearRegPlot)*** was created whose inputs are (a) *x & y datasets*, (b) *x & y labels*, (c) *Plot Title* (d) *Title Offset* and (e) *Output file name*. ***The following plots were made by calling the function with different parameters:***

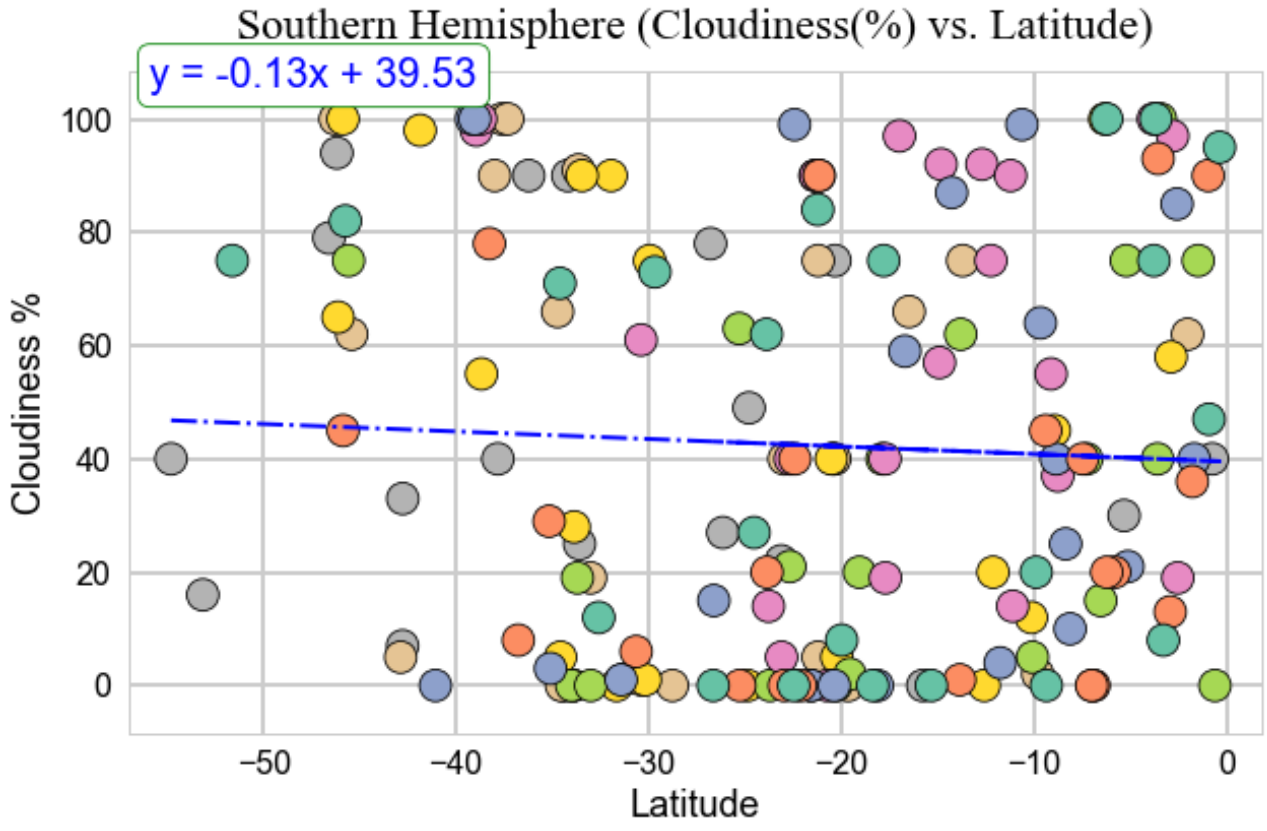


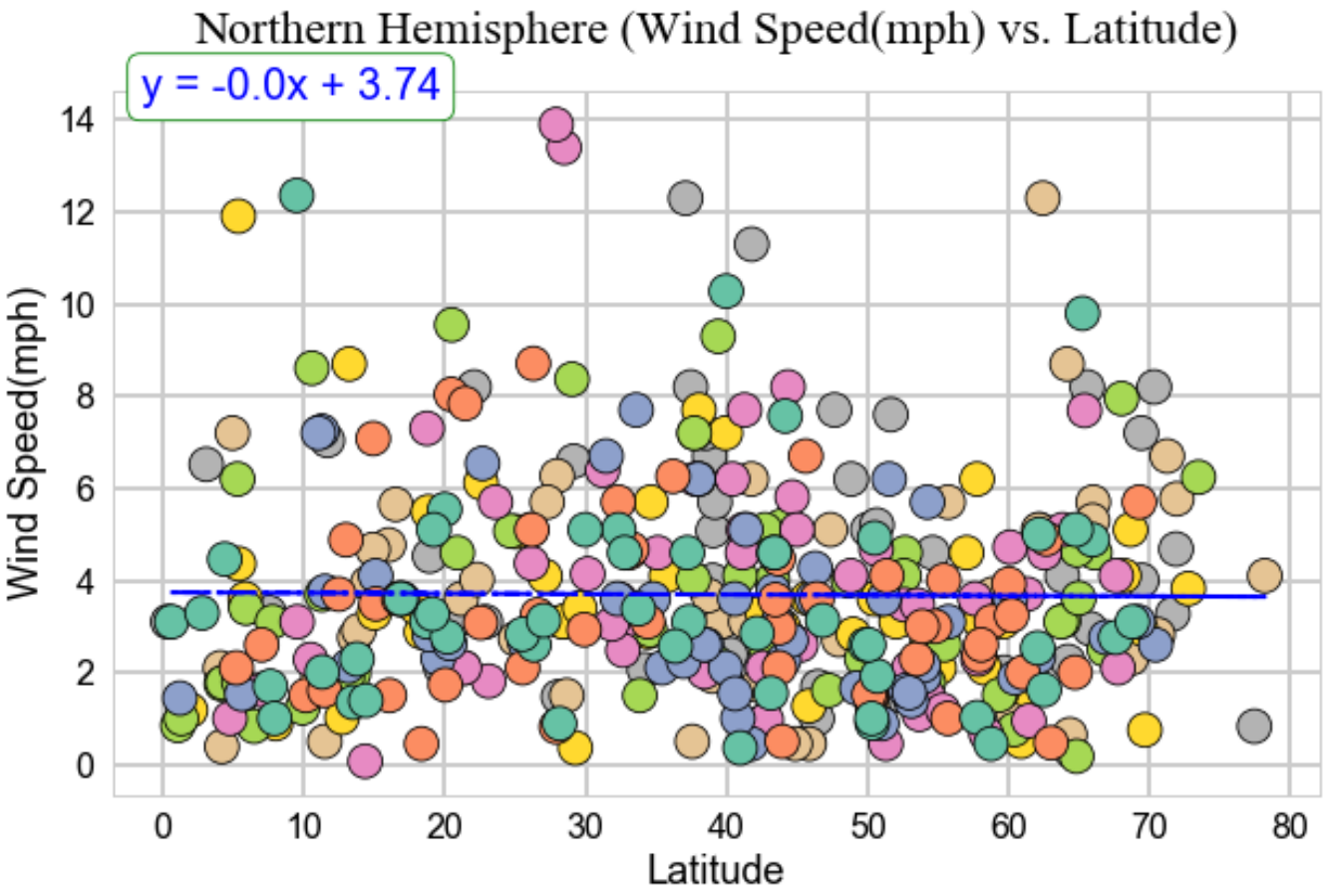



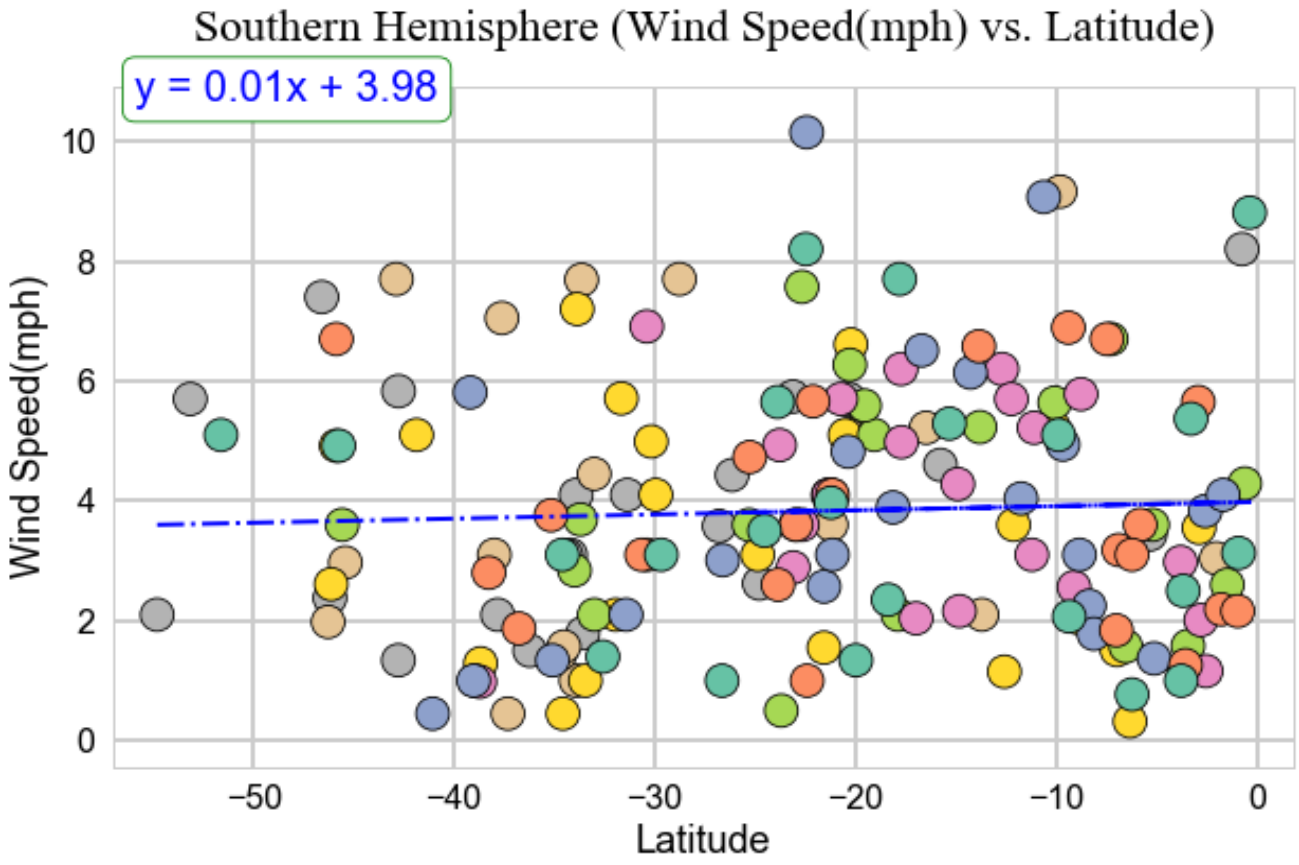

Northern Hemisphere (Humidity(%) vs. Latitude)
y = 0.05x + 62.39
Humidity %
Latitude
0
20
40
60
80
100
0
10
20
30
40
50
60
70
80


Southern Hemisphere (Humidity(%) vs. Latitude)
y = -0.0x + 70.0
Humidity %
Latitude
0
20
40
60
80
100
−50
−40
−30
−20
−10
0


Northern Hemisphere (Cloudiness(%) vs. Latitude)
y = -0.03x + 50.21
Cloudiness %
Latitude
0
20
40
60
80
100
0
10
20
30
40
50
60
70
80


Southern Hemisphere (Cloudiness(%) vs. Latitude)
y = -0.13x + 39.53
Cloudiness %
Latitude
0
20
40
60
80
100
−50
−40
−30
−20
−10
0


Northern Hemisphere (Wind Speed(mph) vs. Latitude)
y = -0.0x + 3.74
Wind Speed(mph)
Latitude
0
2
4
6
8
10
12
14
0
10
20
30
40
50
60
70
80


Southern Hemisphere (Wind Speed(mph) vs. Latitude)
y = 0.01x + 3.98
Wind Speed(mph)
Latitude
0
2
4
6
8
10
−50
−40
−30
−20
−10
0

| Hemisphere | Parameter | r-squared value |
|---|---|---|
| Northern | Maximum Temp (F) | -0.748 |
| | Humidity (%) | 0.05 |
| | Cloudiness (%) | -0.018 |
| | Wind Speed (mph) | -0.01 |
| Southern | Maximum Temp (F) | 0.782 |
| | Humidity (%) | -0.003 |
| | Cloudiness (%) | -0.049 |
| | Wind Speed (mph) | 0.045 |

## Conclusions:

### ❖ Linear Relationship (Maximum temperature vs. Latitude):

Maximum Temperature (F) has a positive linear correlation with latitude in the Southern Hemisphere and a negative linear correlation in the Northern Hemisphere. In both cases, the correlation of determination ("r-squared") is high – equal to or more than 0.75.

### ❖ Other parameters are unrelated with Latitude:

Other parameters like Cloudiness, Wind Speed and Humidity are not correlated with Latitude, with r-squared value less than 0.1

# Summary:

**By analyzing weather data from 583 cities spread over the globe, we conclude:**

a) **Maximum temperature is highest near the Tropic of Cancer (due to summer in Northern Hemisphere)**

b) **The average value of other parameters viz. Humidity, Cloudiness and Wind Speed remain unchanged over the range of latitudes.**