Boston Weather Data Analysis by Dylan Gomes

Start Date 3/29/23

Project Goals:

To practice cleaning data with a mix of SQL and Excel and building visualizations in PowerBi, along with helping me answer some questions that have been bothering me as I have noticed a drastic difference in weather based on the last couple of years.

\*\*\*We can answer the questions with a mix of SQL and a mix of visuals with either PowerBi or Tableau\*\*\*

Answer the following questions:

Is there a noticeable trend that the weather is consistently warmer during each season of the year.

\*Answered in the PowerBi Visualization\*

SPRING IS 3 to 5, SUMMER IS 6 to 8, FALL is 9 to 11, WINTER is 12 to 2.

What we need to do with the data,

Data needs to only include winter months, December, January, and February.

Data must include all of the years 2013 to 2023.

\*\*\* THE MAIN QUESTION \*\*\*  
Is global warming effecting the weather in Boston Massachusetts.

We are going to gather and separate the data from winter, summer, fall and overall. Then we will clean and analyze the data allowing us to separate the information we need to answer each of these questions. From there we can then make vizualizations and answer the questions on if it is effecting the weather year over year.

When looking at the excel data compared to the data we had imported into our SQL server, I have found that multiple dates were missing from years by comparing a exported version of what I thought was fully cleaned data to a csv then using excel to compare the data using the exact(function) only to find out the SQL data is missing many dates compared to the original csv. I am now taking a look at the data that was imported to SQL to see where something might have went wrong in the import process.

Steps taken to troubleshoot the problem:

Reinputted data from the original csv to a test table through another method, then I used excel again to compare but same issue we end up starting to miss some of the dates which is a big problem because it will throw of the entire analysis.

(The easiest solution would be just to just preform the data analysis in excel but I want to do it in SQL so we are going to try and make it work).

Looked into the original data csv file, noticed that the data that happened to be missing always seemed to have a null somewhere in the row, might have to replace all the nulls in excel before we input the data going to investigate and see.

While importing data again there has been more errors that were occurring so I believe we are going to need to fill in the nulls before we can import this data.

Once the Nulls were filled, the data imported with 0 rows dropped, while looking through the data I noticed that there was a consistent row that had nulls mostly which was the winds direction which in reality could be useful for our analysis but it wasn’t planned to be used in the analysis so we might be able to get away without it.

After using count() in MYSQL and testing the data in Excel I have confirmed that filling the nulls fixed the data, we now need to clean and figure out what to do with those Nulls which we will do in SQL.

Simply replacing this with nulls was not the simple solution, we need to decide how to handle the nulls currently because SQL is not going to let you query with all of those Nulls in place as it will mess up things like math.