Seattle weather project

Introduction

We need to understand the climate, particularly the amount of rainfall in a given place, as excessive rainfall can lead to problems, while a lack of rainfall can also cause issues. Which is why the project mainly focuses on identifying the amount of rainfall in Seattle and a city I am completely unaware of but much more curious about, which is Tampa. This project aims to compare the rainfall in Seattle, Washington, and Tampa, Florida, based on two main attributes from the dataframe, date and precipitation, for the time period 2018 to 2022. This is what we will be mainly focusing on in this project, to identify which city has more rainfalls and which city receives the most rainfall and the precipitation levels in both the cities.

Data Description

The data for this project is taken from two sources. The Seattle rainfall data was extracted from an open-source GitHub dataset, and the Tampa rainfall data was extracted from the National Oceanic and Atmospheric Administration (NOAA). Both of the datasets contained daily precipitation readings with the date and the amount of rainfall, the precipitation in inches. After combining the two datasets, the new dataset contained 3,624 observations. I added month and day of year columns, which made it easier to examine the seasonal trends. Missing precipitation values were imputed using the average rainfall for that month. I added a new column called any_precipitation to indicate whether it rained on that day (1 for rain, 0 for no rain).

Methods and Analysis

The methods and the data analysis were divided into several phases:

1. Data Collection: I saved the data and placed them in my weather data folder. Then I loaded the libraries needed to analyze the dataset and imported the files with Pandas Read_CSV.

2. Data Understanding: I explored the data with the functions such as head(), info(), columns() to understand the important columns that i have to focus on.

3. Data preparation/ Cleaning: The Seattle and Tampa datasets were not in a particular datetime format, so I converted the data columns to datetime format using the Panda's for the_Datetime function. Since we noticed that the Seattle dataset gave us the wrong results, I explicitly changed the data column to Datetime format. Then I merged the data sets in Seattle and Tampa together. Then I joined the columns we had to work with individually and removed the zero values from the dataset.

4. Data Exploration: I began to look at rainfall on a daily basis, and compared the two cities over time to see patterns. I created plots to confirm monthly and daily rainfall for Seattle and Tampa. Then I planned seasonal and monthly patterns in histograms and box plots to compare the rainfall distributions to Seattle and Tampa. This helped us identify which month got more or less rainfall.

5. Statistical testing is done to test whether the differences between the cities were meaningful, I used A two-sample t-test to compare average rainfall amounts per month between Seattle and Tampa.A two-proportion z-test to compare the proportion of rainy days each month in the two cities seattle and tampa.

6. To make the results easy to read, I used clear and simple graphs that indicate and show the differences. Barplots were annotated with asterisks (*) to mark statistically significant differences between cities. And the contingency table did give the Boolean values of precipitation.

Results

There were various findings from the project, all the findings with visualizations are as follows:
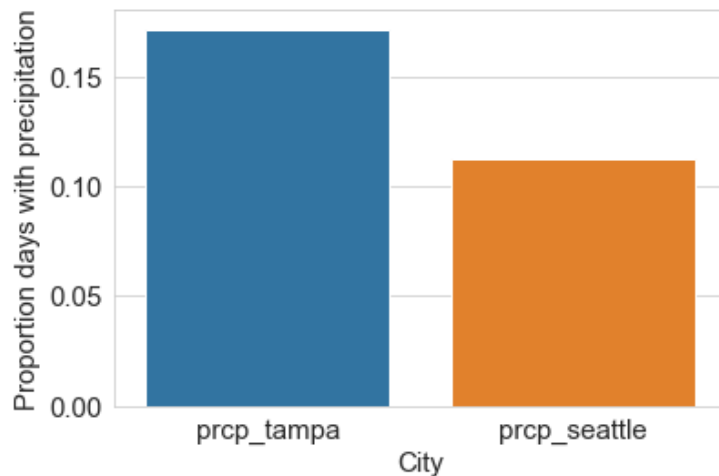

The fact that Tampa's bar is higher than Seattle's suggests that precipitation fell on a higher percentage of Tampa's observed days than Seattle's. Given that Seattle is stereotyped as having a lot of rain, this outcome may come as a surprise. However, Seattle experiences more frequent, lighter precipitation days, whereas Tampa experiences heavy bursts of rainfall more often, particularly during specific months. These statistics may also be impacted by variations in data collection times, missing data, or station locations.
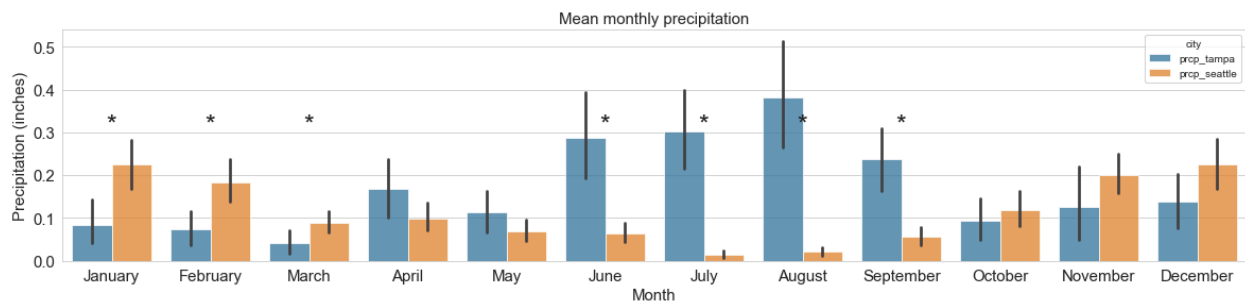
Tampa experiences a sharp increase in rainfall from June to September, which also happens to be the hurricane season this year. During these months, Tampa consistently receives more rainfall on average

than Seattle. In the fall, winter, and spring, particularly in January, February, and November, December, Seattle experiences more even and moderate rainfall. Tampa receives more rainfall from June to September comparatively than other months than Seattle, which receives less overall. The diagram makes it clear that rainfall patterns and seasons are very different, like Tampa's summers are much wetter, while Seattle's winters are wetter.
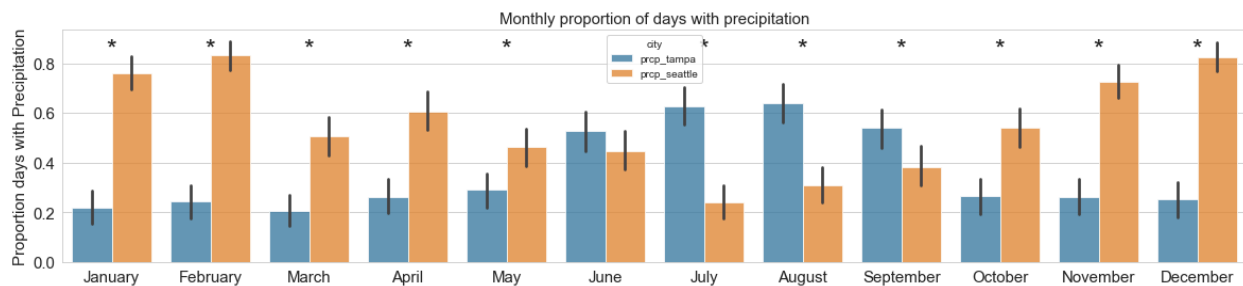
The t-test showed that Tampa has a significantly higher average rainfall in the summer months compared to Seattle. The z-test showed Seattle has significantly more rainy days in most months, even though individual storms are lighter. These differences were marked with significance stars on the barplots. Seattle experiences rain more frequently throughout the year, but in smaller amounts on average. Tampa, on the other hand, tends to have fewer rainy days, but when it rains, the amount of precipitation is higher.

Conclusion

In the conclusion, this analysis reveals the answer to the original question, **"Does it rain more in Seattle or Tampa?"** Seattle has a high number of rainy days, which means that rainfall is frequent and even throughout most of the year. Tampa receives heavy rainfall over the course of short days, concentrated primarily during the summer season. Practically, this means that Seattle has a drizzly, frequently rainy climate, while Tampa has intense, seasonal rainfall. These findings match climate patterns typical of the northwest Pacific and the subtropical climate in Florida.

References

The Seattle rain is downloaded from the github.

https://github.com/brian-fischer/DATA-5100/blob/main/weather/seattle_rain.csv

The source for the city of Tampa data is taken from this page.

https://www.ncei.noaa.gov/cdo-web/orders?email=smurala@seattleu.edu&id=4131968