

NYC Analysis

Introduction

When is the best time to travel to New York City? While “best” can be subjective from person to person, people's main concerns are price and weather. In this report, we look at data from the past two to three years showing New York City's recent air travel, weather, and housing prices history and make an inference on when the best time to travel to New York would be. We considered flying as the method of transportation to NYC. We looked at data from John F. Kennedy International (JFK) and LaGuardia airports to consider the frequency of travel into NYC and to see what times airplane delays by weather are most prevalent. We discuss our sources of data, how we cleaned our data, methods used for gathering and extracting the information, data analysis, and visualizations on our cleaned datasets, and a conclusion on when the best time to travel to NYC is by airport travel factoring in weather and short-term housing costs.

Data

The primary datasets we used for our report included an Airbnb, weather, and two airline datasets that focused on available information in New York, New York. The Airbnb dataset was gathered from Kaggle which when cleaned allowed us to look at neighborhood, availability, and cost to gather an idea of what costs would be associated with somebody looking for a residence to stay at while vacationing in New York City. The airline datasets were trickier to find but were accessed through the United States Department of Transportation. We gathered a dataset of incoming flights into John F. Kennedy International (JFK) and LaGuardia airports. This data was gathered between 2022 and 2023 to allow us to view the frequency of flights into NYC by date and consider the frequency of delays from weather to those flights. The weather dataset was also gathered from Kaggle. This gave daily weather from 2016-2022 for any US city. We narrowed this down to New York, New York for 2022 to consider weather events by severity. The temperature we gathered was pulled from an API at “open-meteo.com” for 2021-2023 along with rain and snow totals. We also grouped this data by month to more easily compare to the other datasets. We used a secondary data source to chart the monthly price of hotels from the U.S. Bureau of Labor Statistics. Online

sources suggested NYC hotels are 6.7-8% greater in cost per month, so we used that assessment in our visualization.

Data sources:

[BLS Data Viewer](#)

[Detailed Statistics Arrivals \(bts.gov\)](#)

[New York Airbnb Bookings \(kaggle.com\)](#)

[US Weather Events \(2016 - 2022\) \(kaggle.com\)](#)

[NYC Hotel Info](#)

Data Cleaning

Before commencing our data analysis, we first had to clean up our datasets to ensure the datasets were refined and contained relevant information. Fortunately, the initial structure of our chosen datasets minimized the need for extensive cleaning. Our first task involved systematically reviewing each dataset column to exclude irrelevant information. Given that many datasets encompassed data beyond New York City, we employed the grouping function to consolidate data specifically by State and City, focusing solely on relevant information. Additionally, we utilized the grouping functionality to aggregate data by month, enabling us to compute average values for selected weather metrics that we used to create our visuals. In some instances, we needed to create new columns by manipulating the existing data. For instance, to get the average hotel prices in New York City, we calculated this by adjusting the national average with the percentage difference specific to NYC prices. Once we completed the cleaning process and incorporated the necessary columns, we proceeded to analyze the data and develop our visual representations.

Methods

We used online research to find our data. After sourcing through different datasets and removing ones that were not fully relevant to our report, we gained a clearer perspective of the necessary data sources we needed. We searched common datasets on Kaggle and found our Airbnb and weather data from that. We found an airline dataset but realized it did not contain enough relevant data specifically to

NYC. This led us to the United States Department of Transportation website where we gathered datasets for JFK and LaGuardia airports each with over 200,000 rows of data. Using the U.S. Bureau of Labor Statistics gave us access to the national hotel cost average. Our API was conveniently used in a previous class lab making it easy to locate.

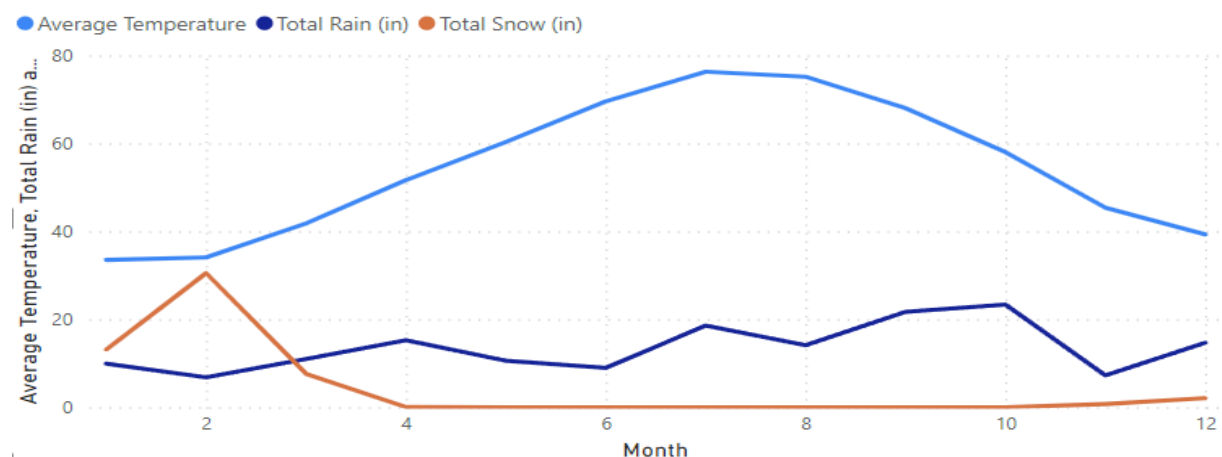
Once all our data was gathered, we worked on cleaning it by removing unnecessary columns of information and creating new columns where needed. We focused on transforming date entries into date columns to make the data easier to universally extract and compare. We also focused on creating “month” columns to organize our data by month. We made some other minor changes by renaming columns and editing weird cell values.

The final method was combining our data into two two giant CSV files that include all our information. These were separated by the JFK and LaGuardia datasets due to their sizes. We uploaded them into Power BI, and from there, ran data analysis and data visualization to reach our results.

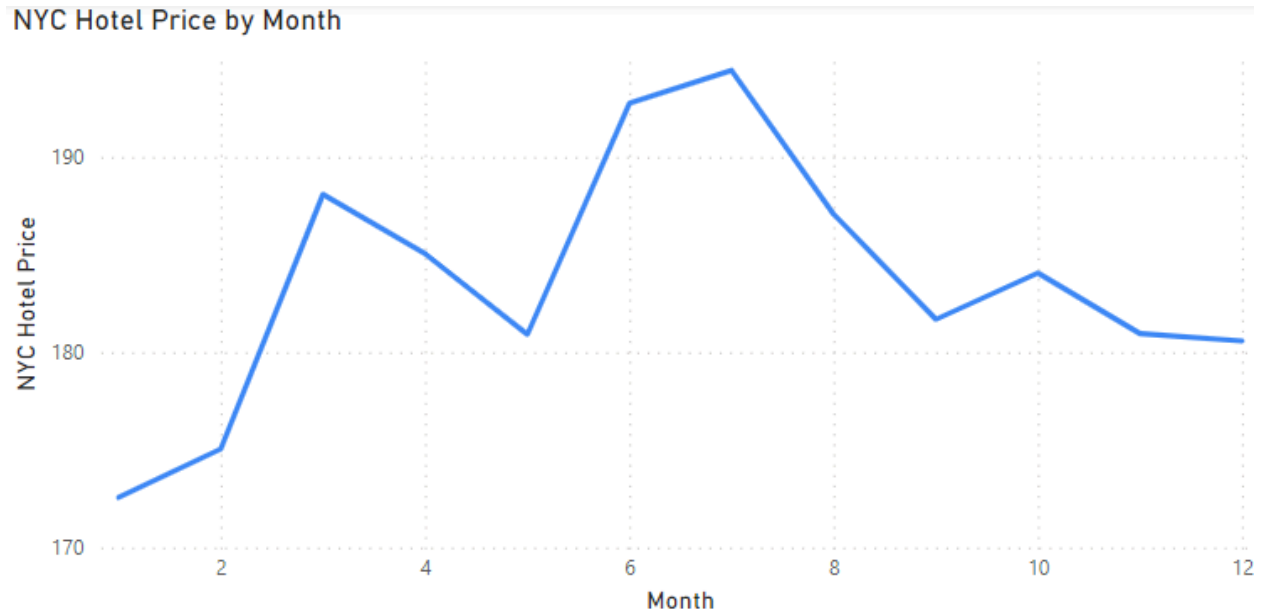
Analysis

Our analysis consisted of showing the month-to-month trends throughout the main deciding factors people look at when traveling. The charts below show the data we used to draw our conclusion. We include a brief description of each chart below each chart’s image.

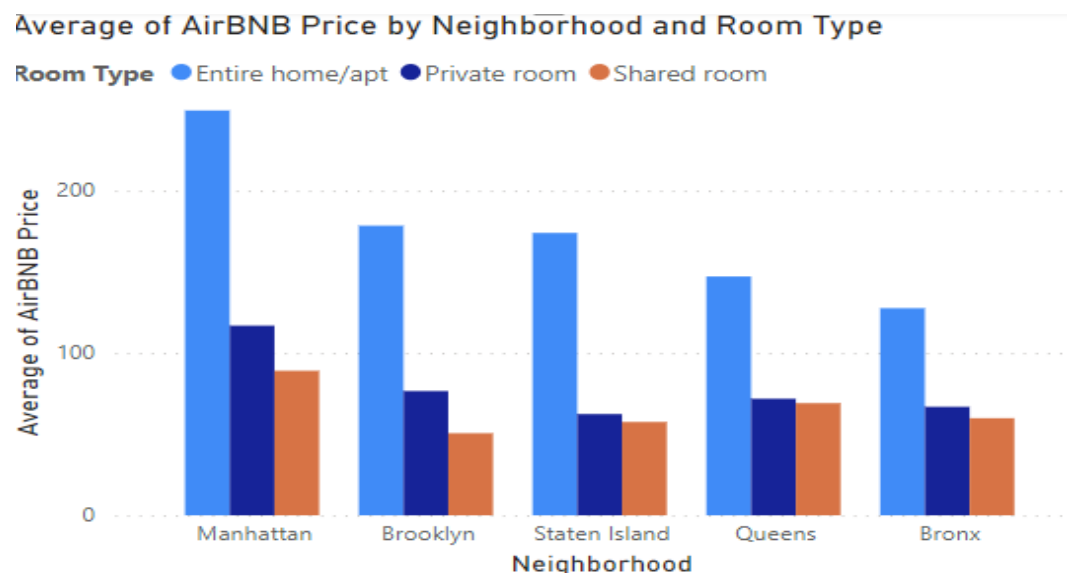
Average Temperature, Total Rain (in) and Total Snow (in) by Month



The above graph shows the average temperature, rain, and snowfall in inches by month.

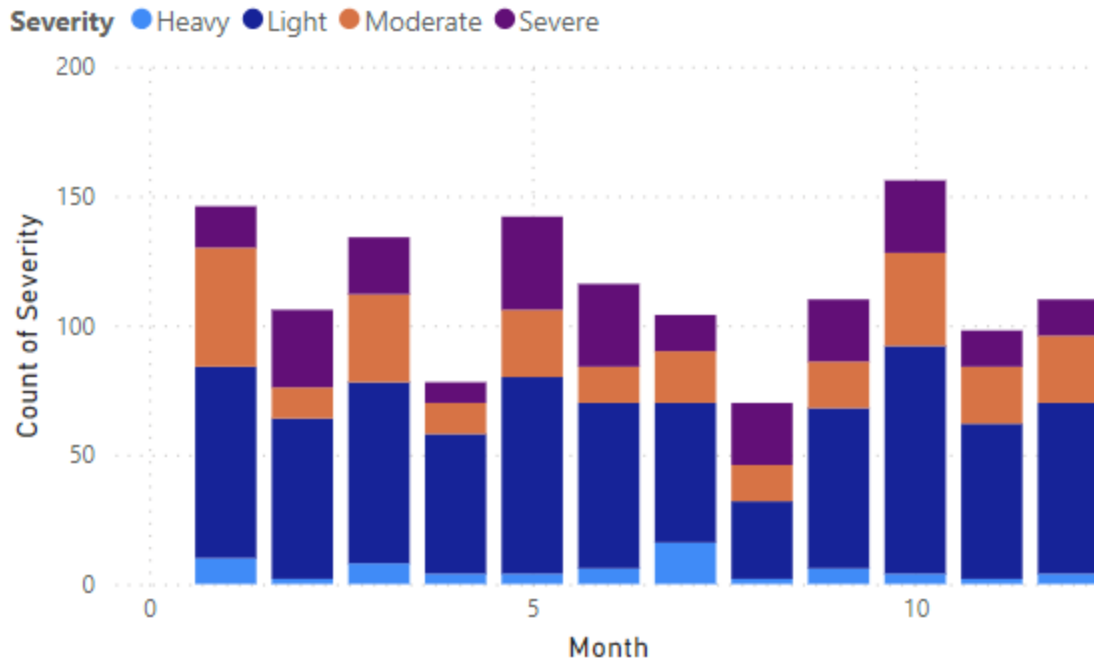


The above graph shows the hotel price trend through the months.

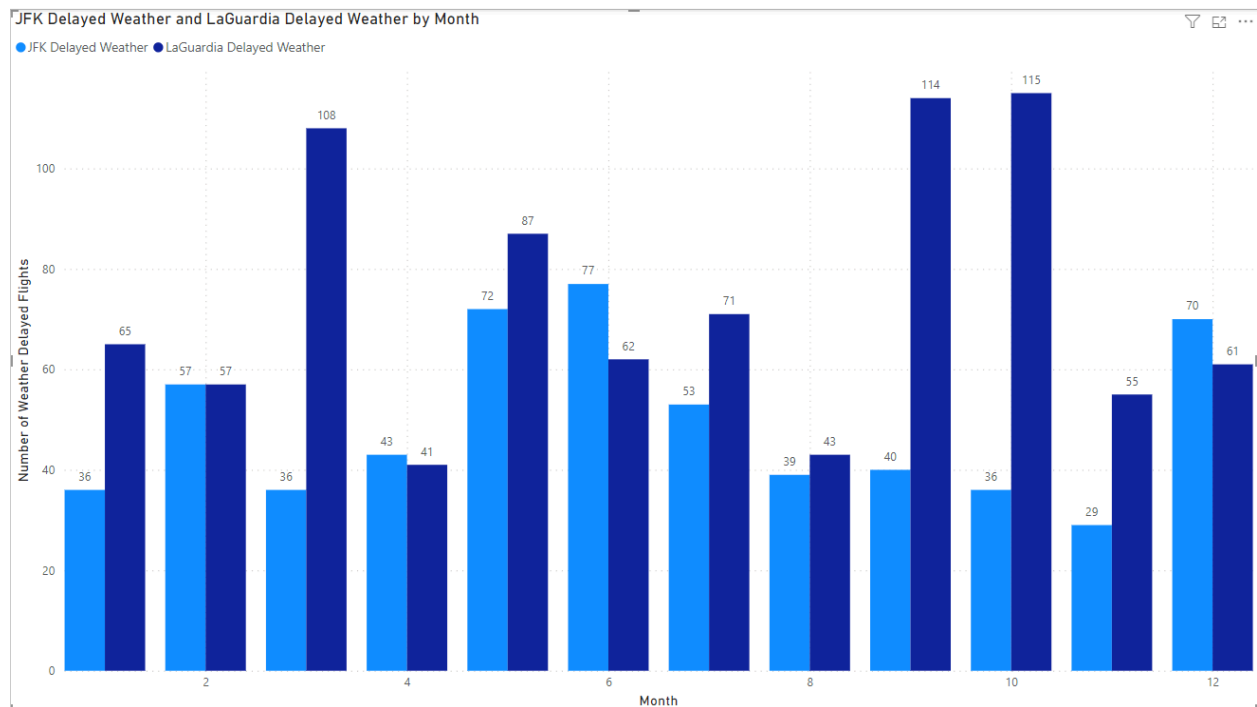


The above graph shows the costs of room types by neighborhood in New York City.

Count of Severity by Month and Severity



The above graph displays the severity counts of weather instances throughout the months.



The above chart shows the number of flights delayed by each month due to weather from the two major NYC airports.

Conclusion

After analyzing these graphs, we found that the best time to travel to New York is in late spring/early summer or late summer. Prices are just beginning to rise but the temperature is comfortably warm with the severity of weather and airport weather delays being lower relative to surrounding months. For an all-around planned trip to New York, any room type outside of Manhattan or a hotel you will save more money while planning a trip during good weather in late summer around August. If you are going the Airbnb route, you will have better luck with flights during August with slightly warmer and better weather than in June or April. If you are planning on staying in a hotel, August also presents a better time than June by cost while having a greater average temperature and less severe weather than April. Factoring in weather temperature, precipitation, weather delayed flights, Airbnb and hotel costs, we believe traveling during August is the “best” time for a vacationer to NYC.