

# Airbnb Bookings in New York City

# The Bionic Blobs

Much like a data binary large object (blob), this team has a diverse skillset and becomes a tremendous asset when teamed up.



Alex Reilly

---

B.S. Spanish  
Data storytelling  
enthusiast



Miko Le

---

B.S. Marketing  
PowerBI Visuals  
fanatic



Shirley Cho

---

B.S. Biochemistry  
B.S. Computer  
Science  
ETL & Database  
lover

# Scope



Airbnb bookings in New York, their prices, and factors that may influence the market



The relationship between Airbnb listings and the surrounding economy



# Project Overview

- ▶ Goals: Gain a better understanding of various external factors and their impact on Airbnb's price, bookings, and ratings.
- ▶ Data:
  - ▶ NYC Airbnb Open Data(1.1 Mil rows)
  - ▶ NYC Reported Crime(6.5 Mil rows)
  - ▶ NYC Hotels
  - ▶ Zillow Housing Data
  - ▶ Airplane Prices
  - ▶ NYC Tourist Attractions (Web scrape)
  - ▶ NYC Population





A low-angle, upward-looking photograph of several modern skyscrapers with glass facades. The sun is visible in the upper left, creating a bright lens flare and illuminating the scene. The sky is blue with scattered white clouds. The buildings converge towards the top of the frame, creating a sense of height and scale.

# Initial Questions

- ▶ What is the price range of Airbnb stays in New York?
- ▶ When is the most popular time of the year to book an Airbnb in New York?
- ▶ How long is the average stay by season/neighborhood/price?
- ▶ Does crime impact Airbnb bookings?
- ▶ Is there a relationship between Airbnb and hotels?
- ▶ How many properties does the average host have?
- ▶ What factors relate to price/review ratings?

# The ETL Process

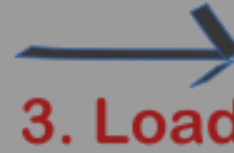
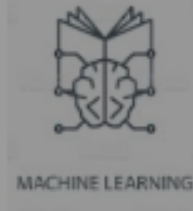
## 1. Extract unclean data



store data back as files

clean data to ml

create ml models



3. Load



dashboards/visuals



3. Load



databricks

2. Transform



store data as tables

cleaned

cleaned



automates process

# The ETL Process

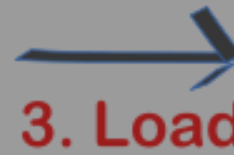
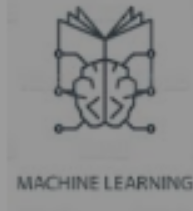
## 1. Extract unclean data



store data back as files

clean data to ml

create ml models



3. Load



dashboards/visuals



3. Load



databricks

2. Transform



store data as tables

cleaned

cleaned



automates process

# The ETL Process

## 1. Extract unclean data



store data back as files

clean data to ml

create ml models



3. Load



dashboards/visuals

3. Load



databricks

2. Transform

store data as tables

cleaned

cleaned



automates process



# The ETL Process

## 1. Extract unclean data



Azure Data Lake Storage Gen2

store data back as files

clean data to ml

create ml models



databricks  
2. Transform



3. Load

dashboards/visuals

3. Load

store data as tables




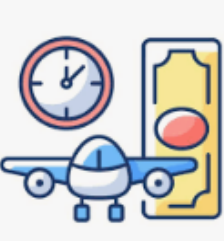
cleaned

cleaned



automates process

# Sources

Dataset	Data of interest to clean for from dataset	Extraction	File Size
 NYC Airbnb Data	Bookings, Reviews, Price, Hosts, Ratings, Location, Amenities, etc.	CSV File	336MB 6MB 82MB
NYC Crime Data	Date occurred, Offense, Type of Crime, Location	CSV File	2GB
 Zillow Housing Data	Average price on rental homes, count of Zillow listings, neighborhood, year, etc.	CSV File	1-2MB
NYC Tourist Attractions	Name, Location, Address	Web Scrape, Kafka	5KB
 NYC Hotels	Hotel Name, Location, High-end Price, Low-end Price, Location, Address, Rating, etc.	CSV File	217KB
Airplane Prices	Fare price, destination airport, total distance traveled, total duration travelled, date of departure, etc.	CSV File	30GB
 NYC Population	Population Count, Boroughs	Website	N/A

Bionic Blobs  
ERD for SQL database  
Topic: New York City AirBnBs  
By: Alex, Miko, Shirley

Legend

Candidate  
Key/Unique  
Key

Foreign  
key

Primary  
key

ZillowNYC_Housing_Counts_BB
ID int primary key identity(1,1)
NYC varchar(100) not null
Date date not null
Year int not null
Month int not null
Count int not null

Attractions_BB
ID int primary key identity(1,1)
Attraction_name varchar(100) not null
Website varchar(100) null
Location_ID bigint not null

NYCBoroughs_BB
ID int primary key identity(1,1)
Borough_name varchar(100) not null
County_name varchar(100) not null

Zillow_Rental_Homes_BB
ID int primary key identity(1,1)
NYC_Borough_ID int not null
Date date not null
Year int not null
Month int not null
Avg_Rental_Home_Price float not null

CK

Hotels_BB
ID int primary key identity(1,1)
Hotel_name varchar(100) not null
Location_ID bigint not null
Hotel_star_rating float not null
Hotel_high_rate float not null
Hotel_low_rate not null

CK

Locations_BB
ID bigint primary key identity(1,1)
Latitude float not null
Longitude float not null
Borough_ID int null
Zipcode varchar(100) null
Address varchar(100) null
Dataset varchar(100) not null

NYC_Crimes_BB
ID int primary key identity(1,1)
Complain_Number bigint not null
Date_Occured date not null
Crime_Type_ID int not null
Location_ID bigint not null

Crime_Types_BB
ID int primary key identity(1,1)
Crime_Description varchar(100) not null
Level_Of_Offense varchar(100) not null

Hosts_BB
Host_ID bigint primary key
Host_is_superhost varchar(10) null
Host_listings_count int null
Host_since date null
Host_response_time varchar(100) null
Host_response_rate varchar(100) null
Host_acceptance_rate varchar(100) null

Listings_BB
Listing_ID primary key bigint
Host_ID bigint not null
Listing_Name varchar(1000) null
About varchar(8000) null
Price_Str varchar(1000) not null
Type varchar(1000) null
Number_of_People int null
Min_n int null
Max_n int null
First_Review_Date date null
Price float null

Reviews_BB
ID bigint primary key
Listing_ID bigint null
Date date null
ReviewerID int null
Comment varchar(8000) null

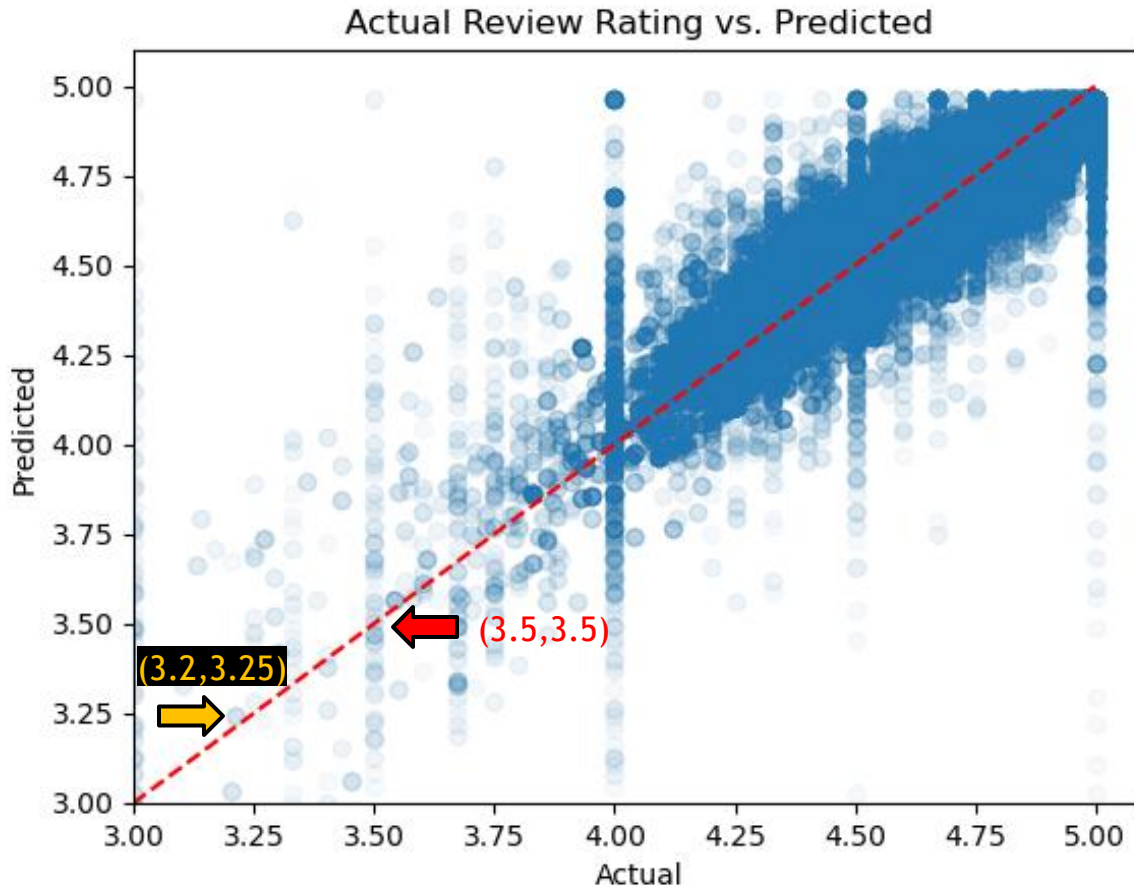
Calendars_BB
ListingID bigint primary key
Count int not null

FK

FK



# Machine Learning: Review Ratings

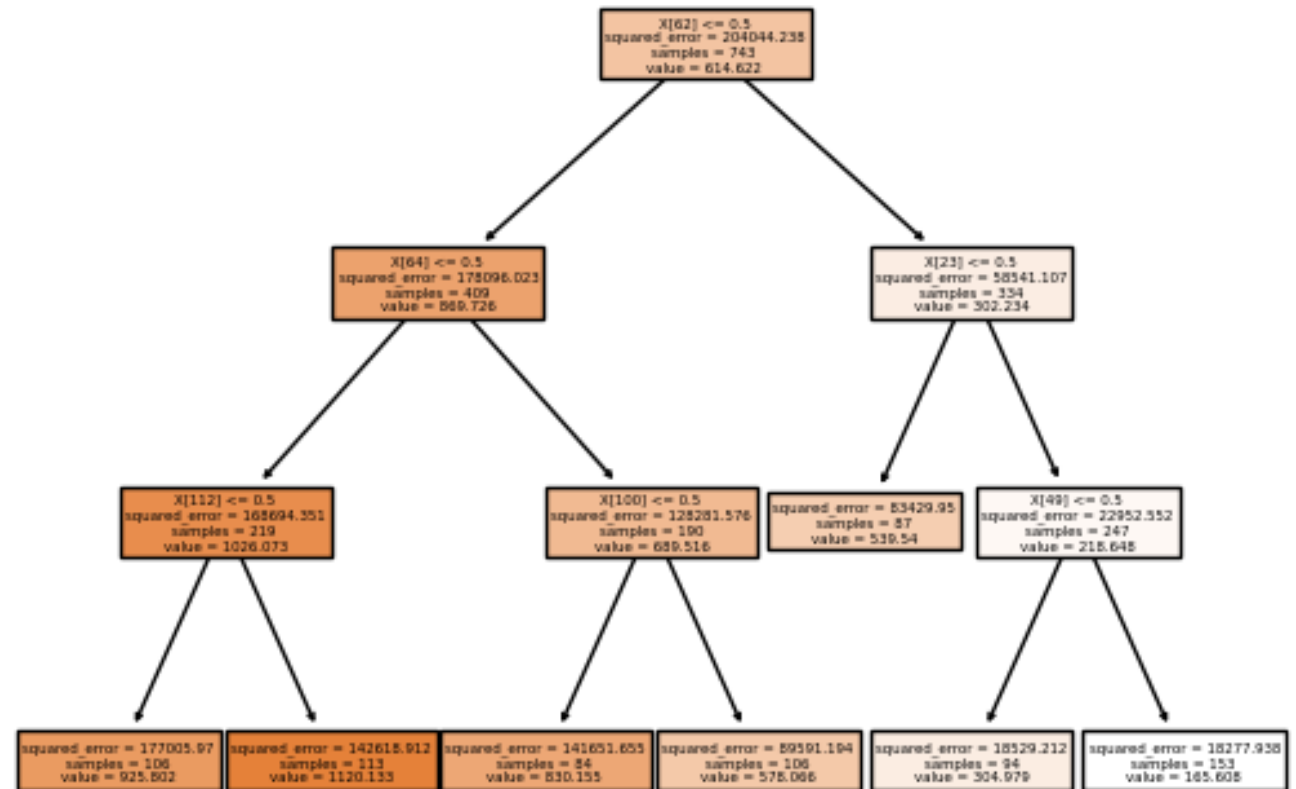


- Test Score: 0.846
- Mean squared error (MSE): 0.006
- Mean absolute error (MAE): 0.054

# Machine Learning: Amenities and Price

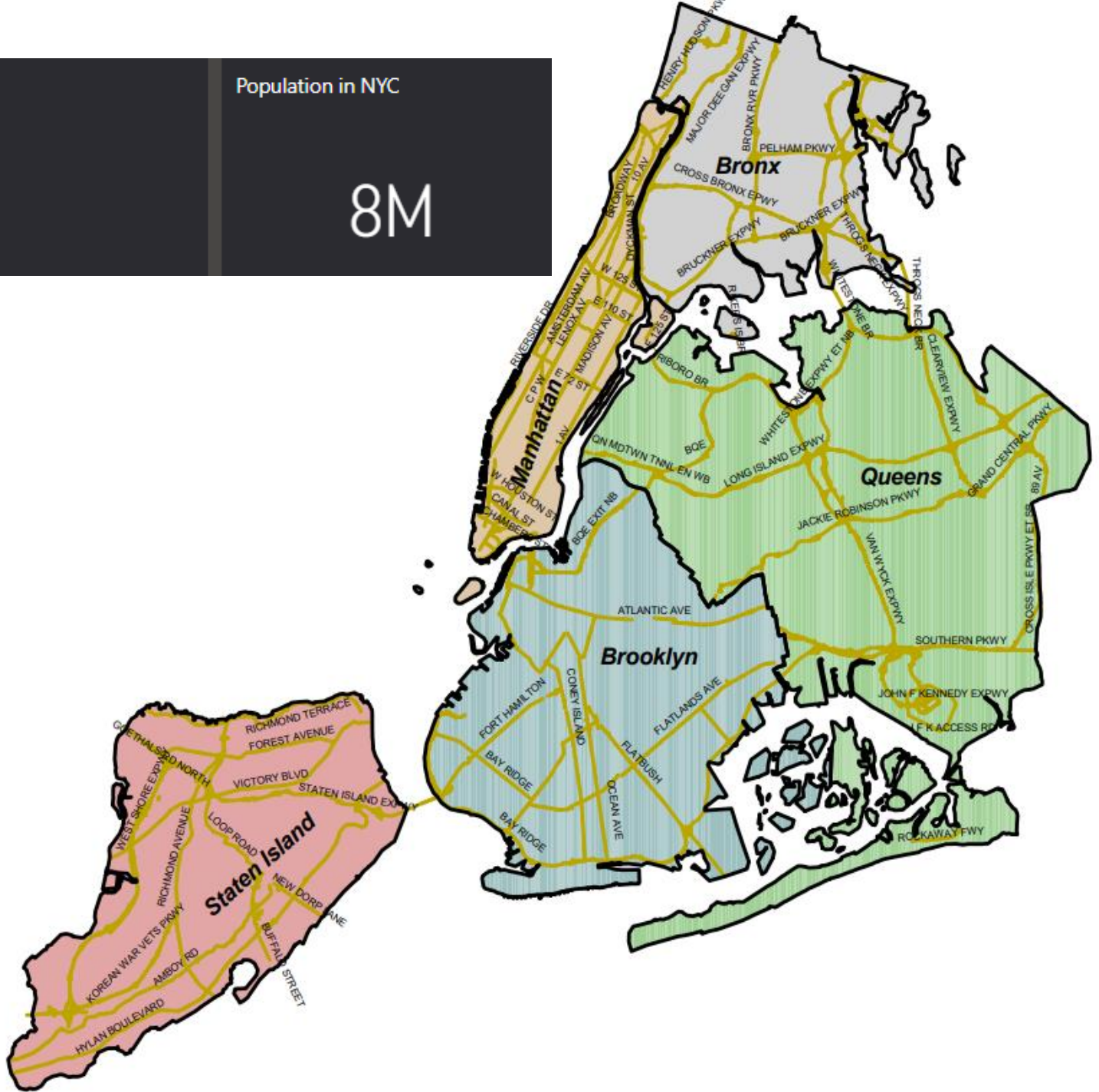
► Test score = .52

Decision tree trained on amenities



Airbnb Listings in NYC	Total Airbnb bookings in NYC 2009-2023	Population in NYC
43K	1M	8M

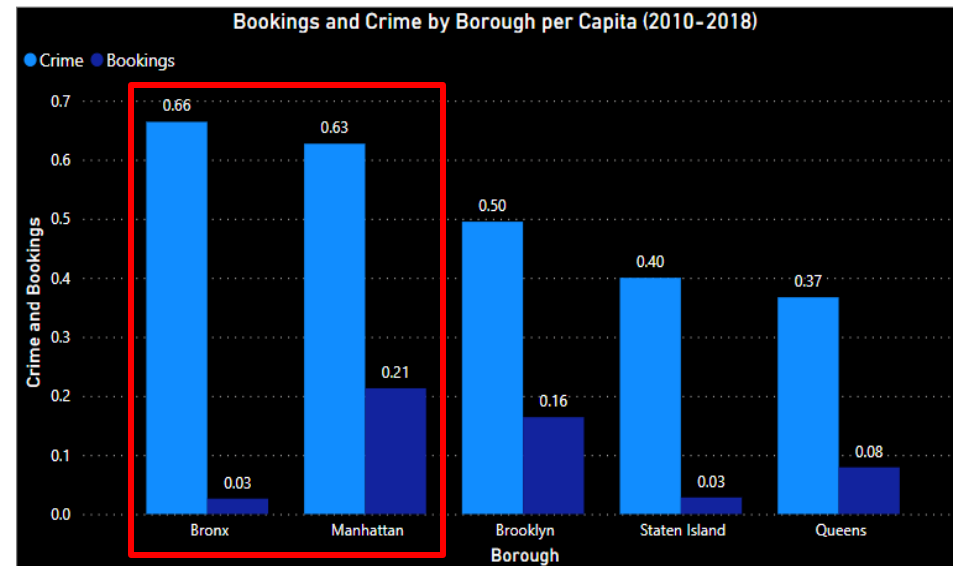
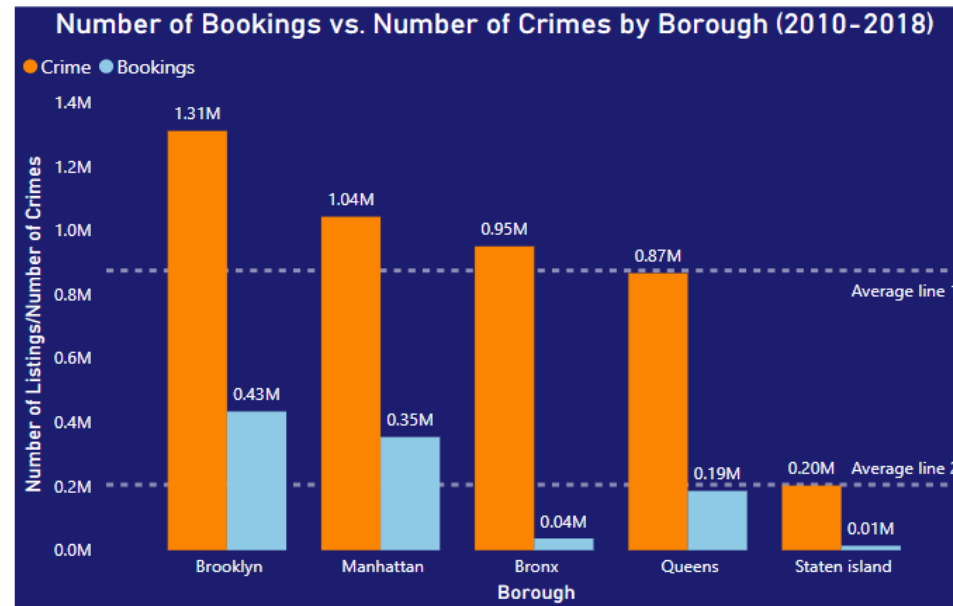
# Airbnb & New York City Info



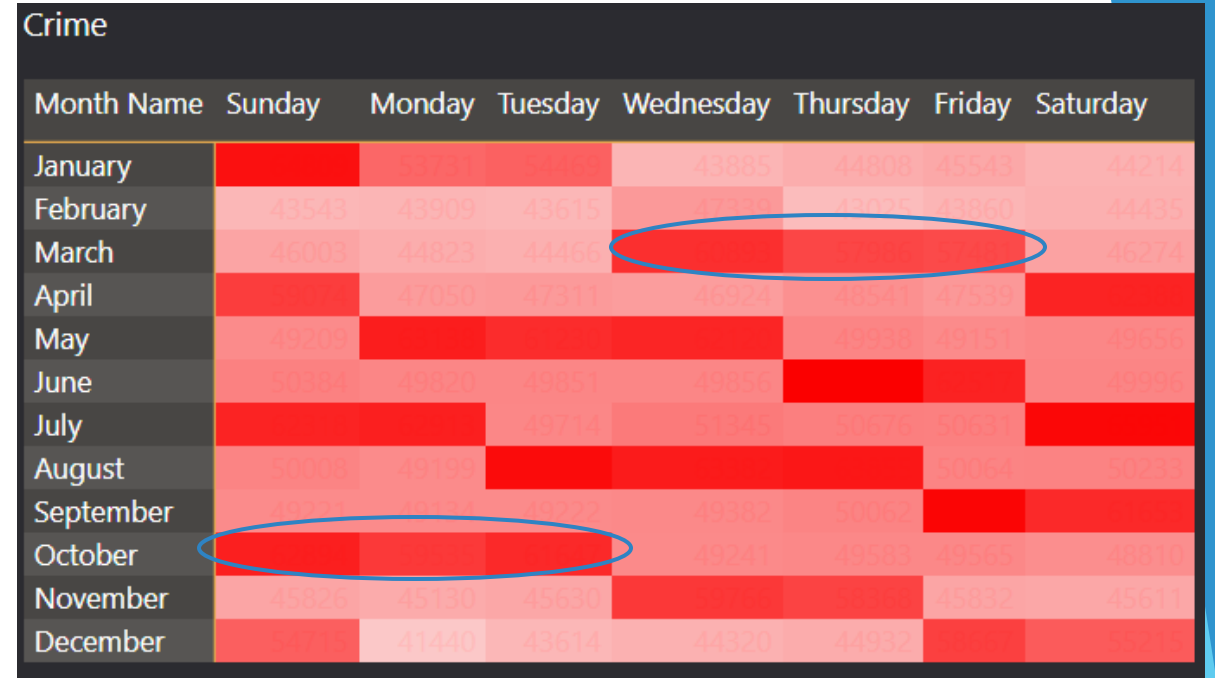
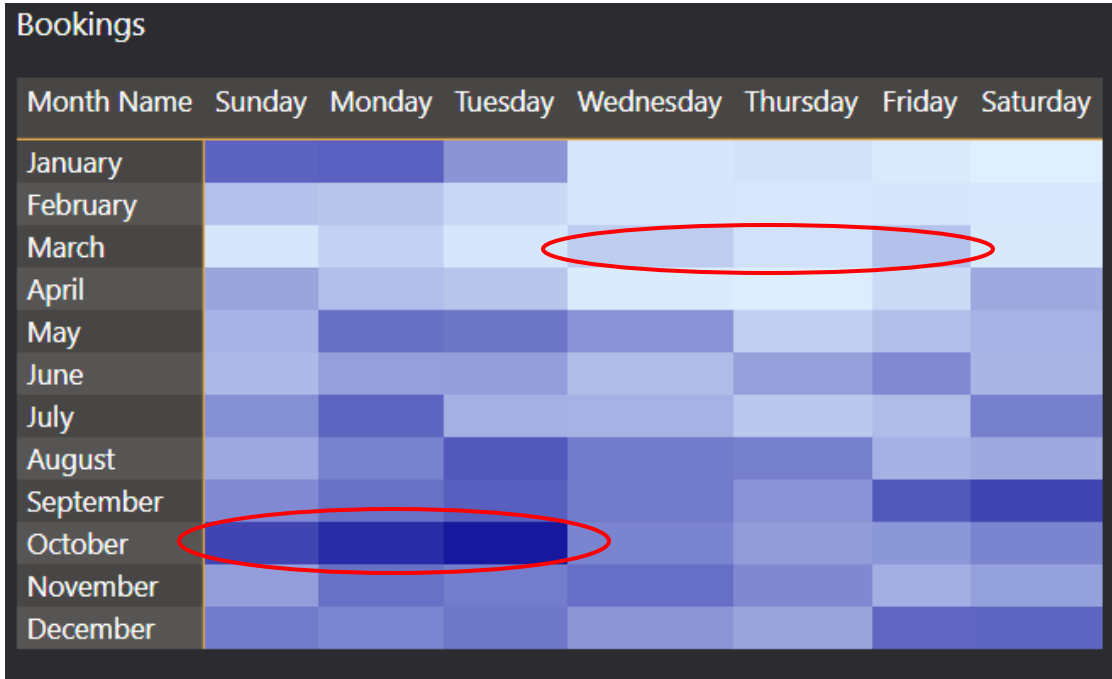


# Does crime impact Airbnb bookings?

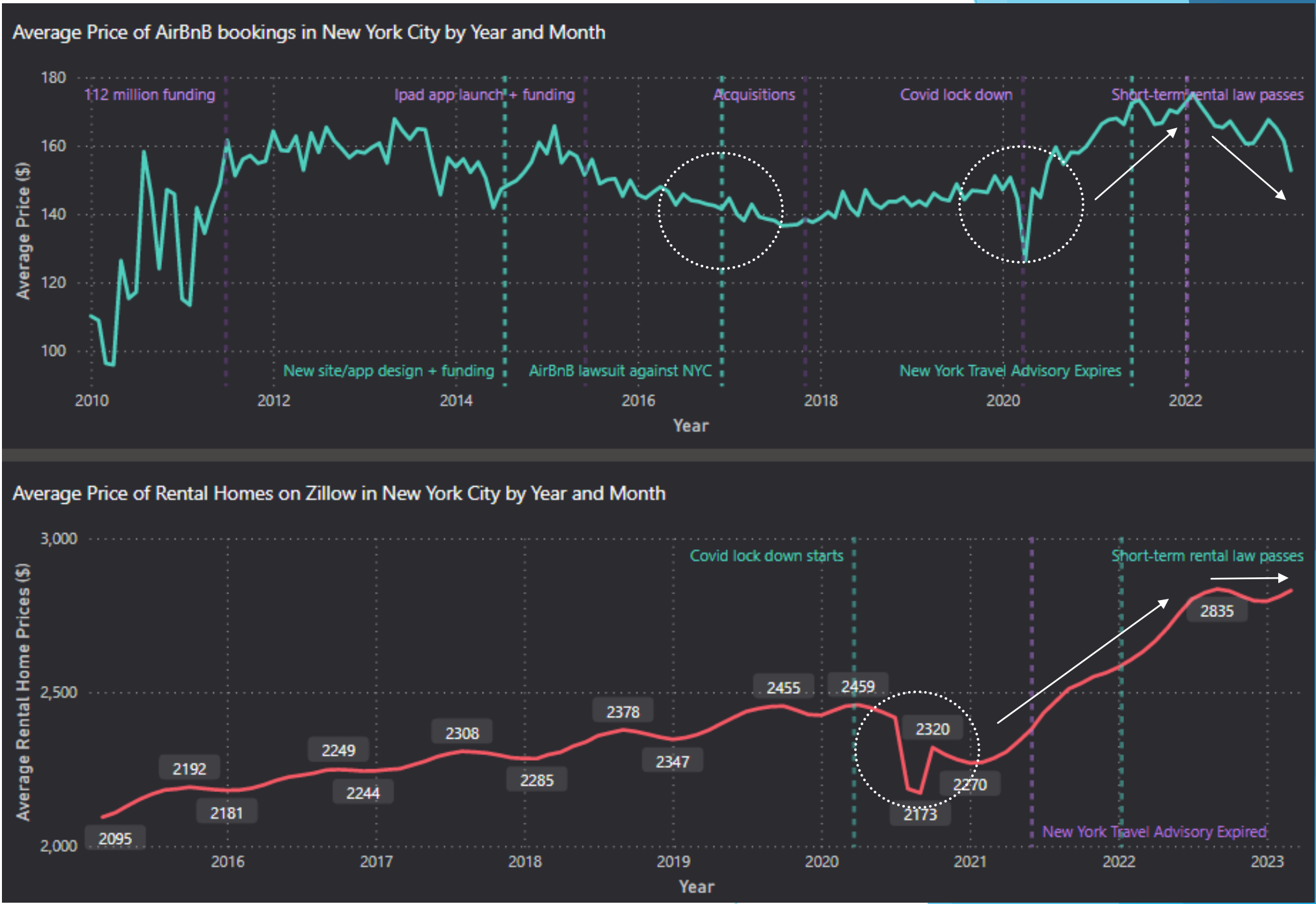
Borough	Population
Brooklyn	2648771
Queens	2358582
Manhattan	1664727
Bronx	1432132
Staten Island	505747



# Number of Bookings vs. Crime Over Time

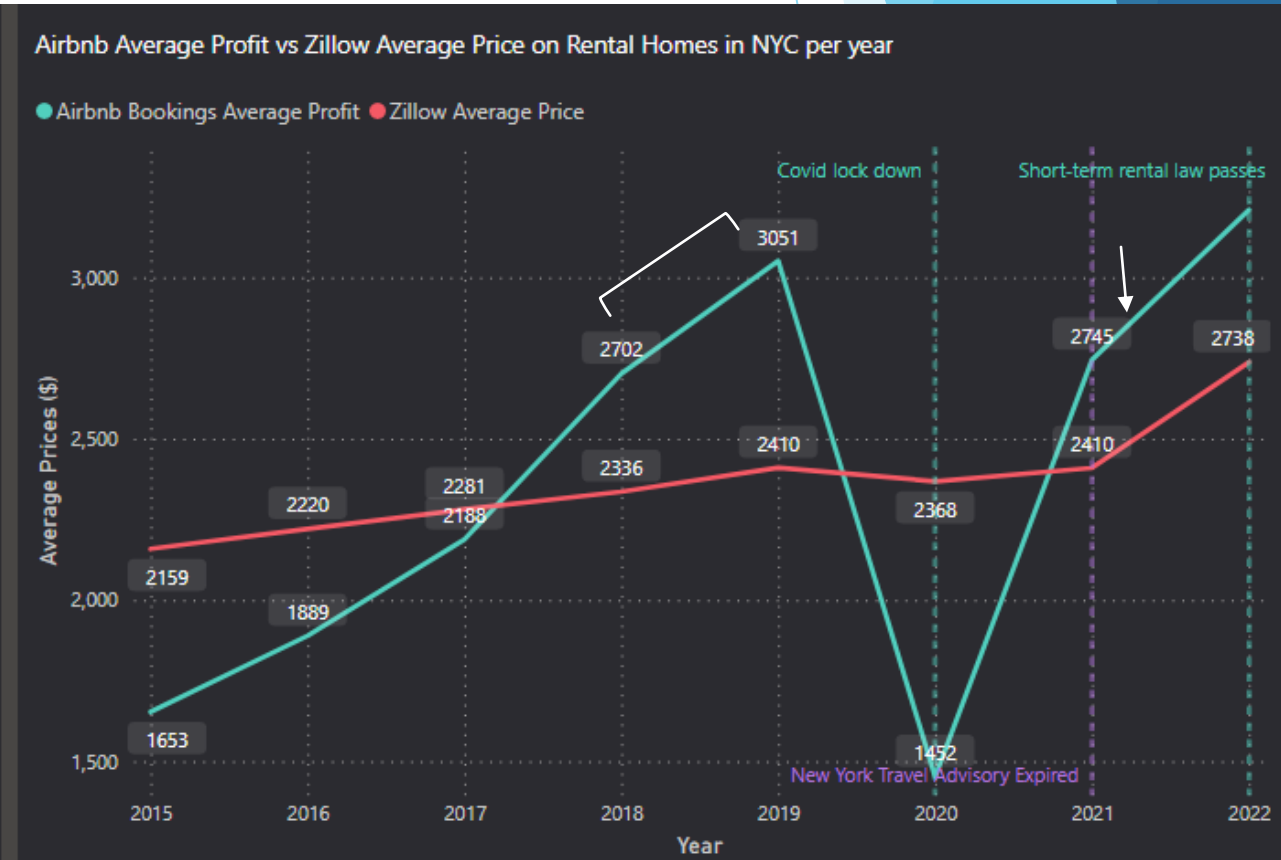
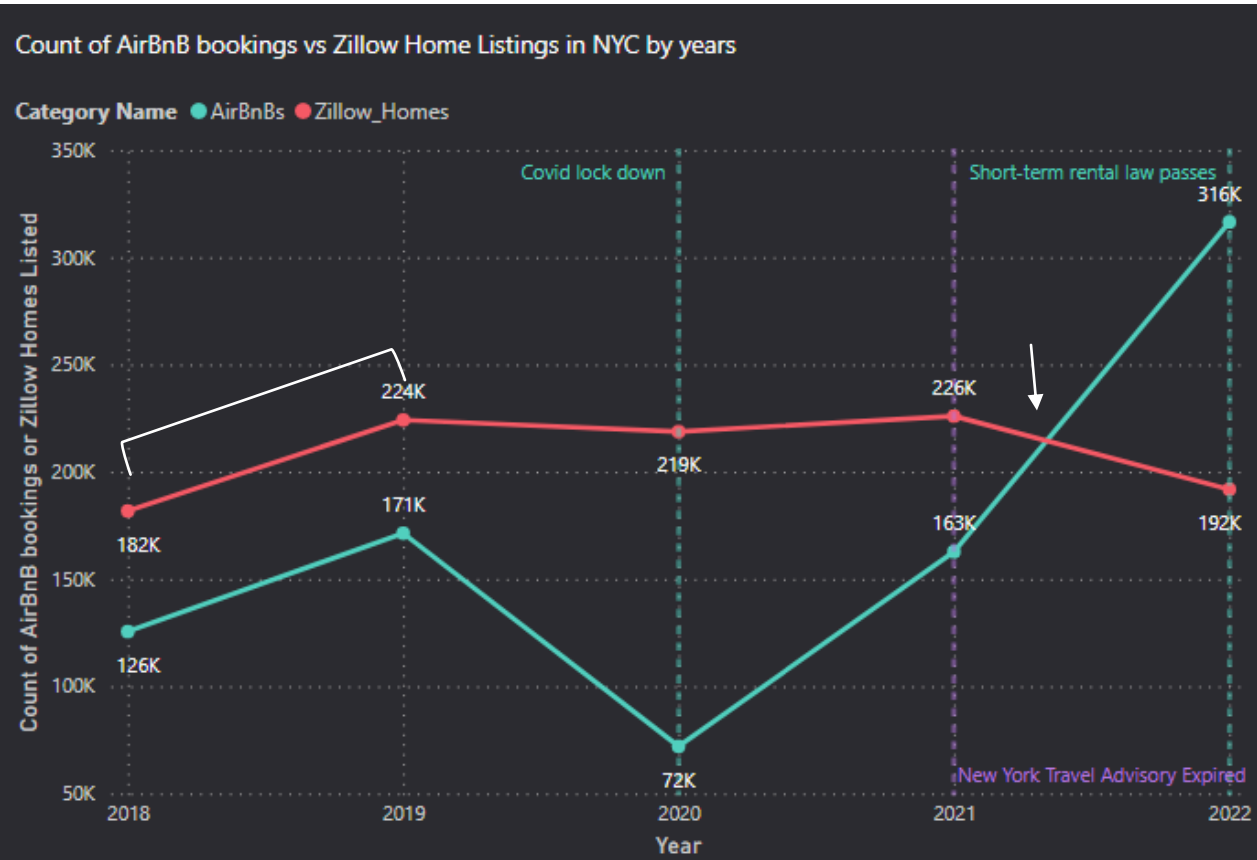


# Airbnb & Housing Market Overview

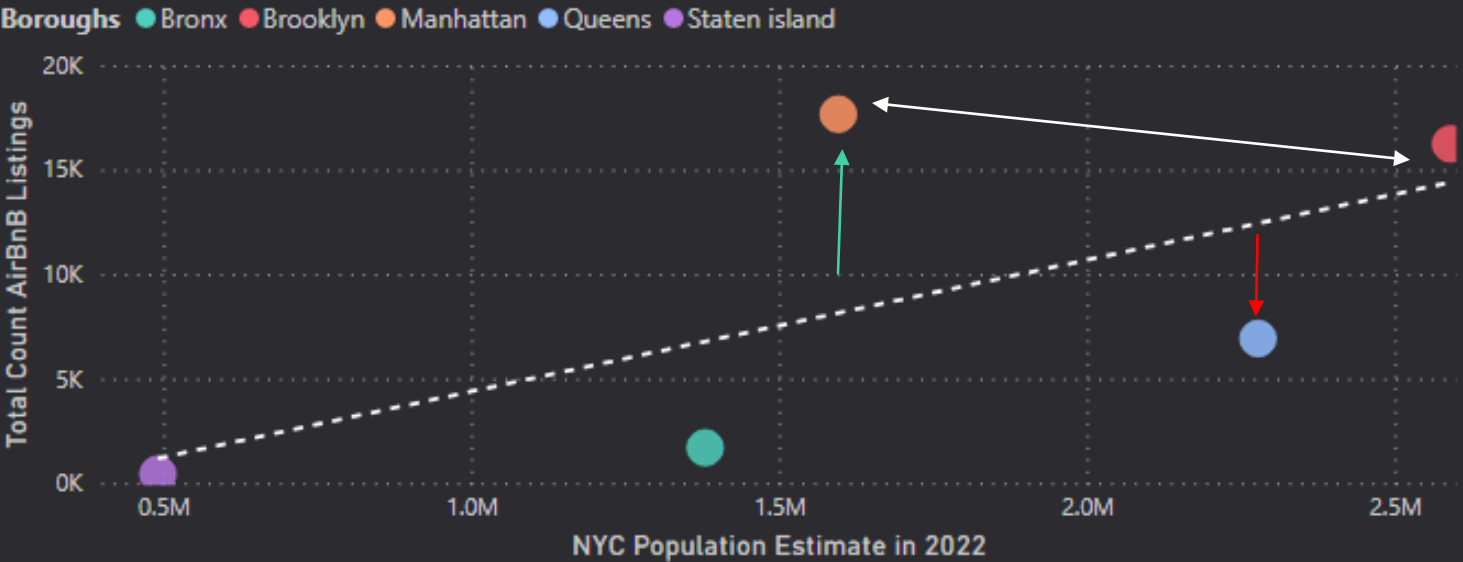




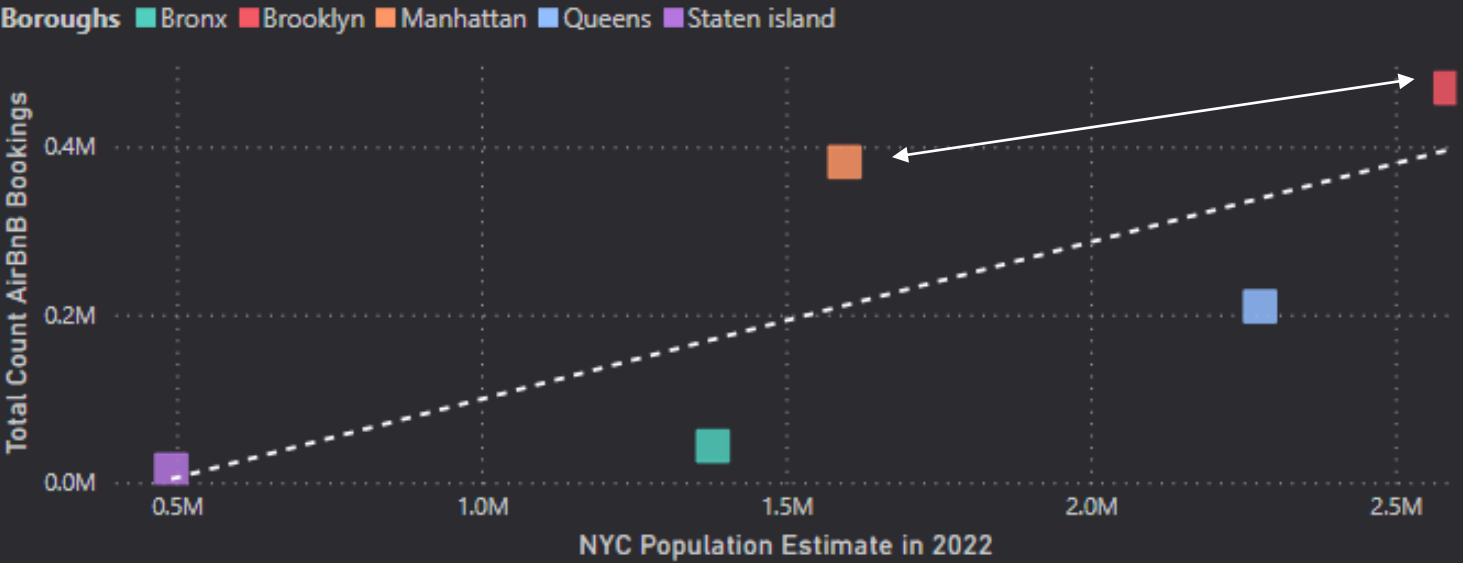
# Comparison between Airbnb and Zillow Rental Homes



Population vs Airbnb Listings in NYC by Borough

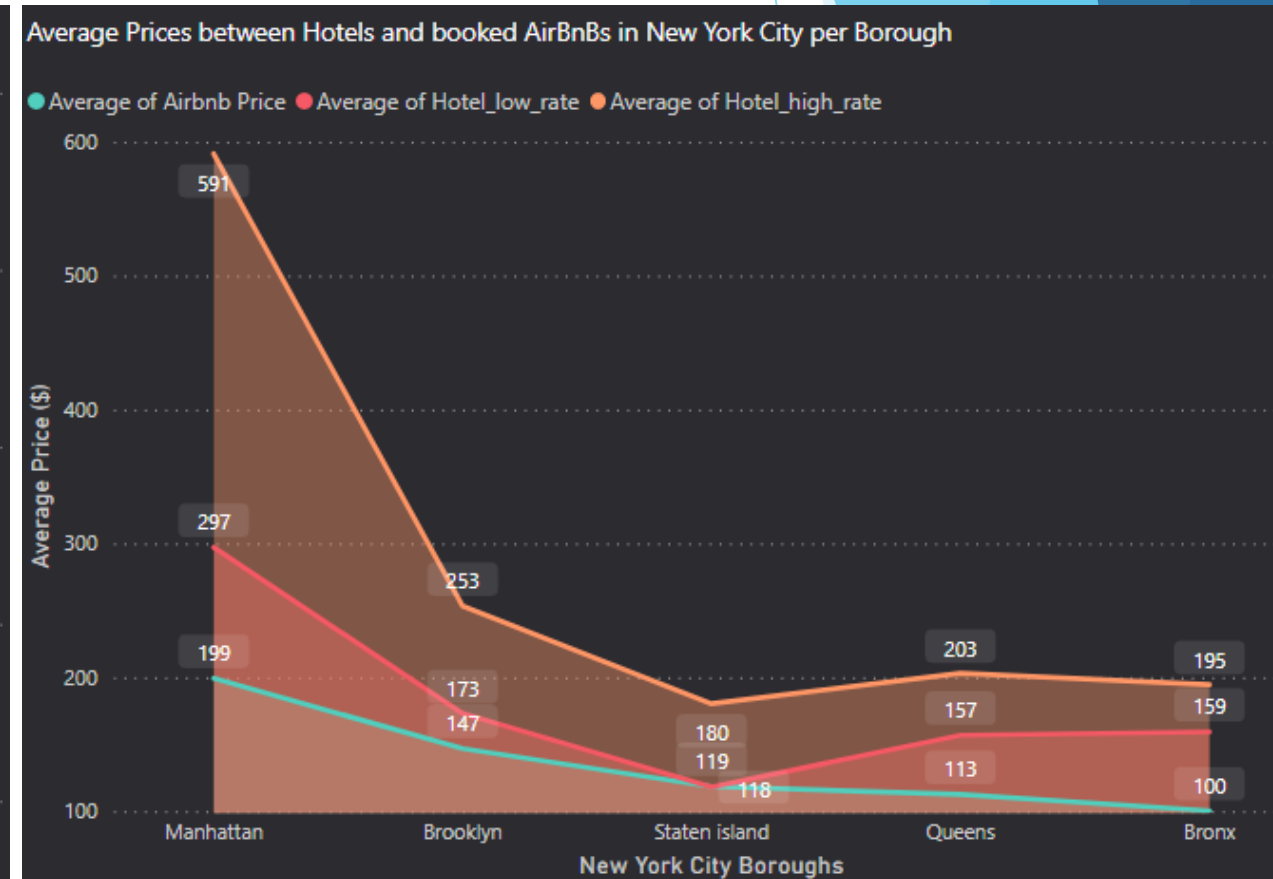
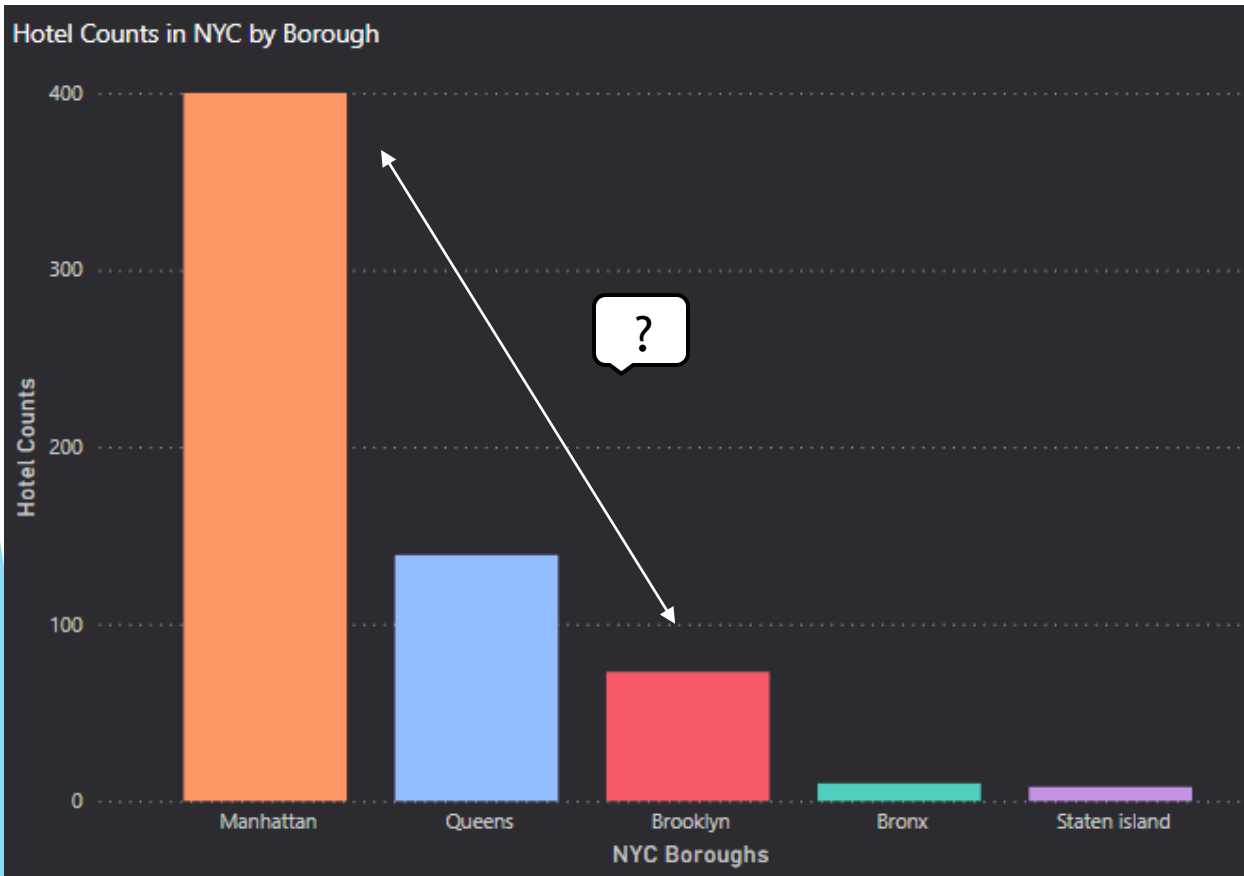


Population vs Airbnb Bookings in NYC by Borough



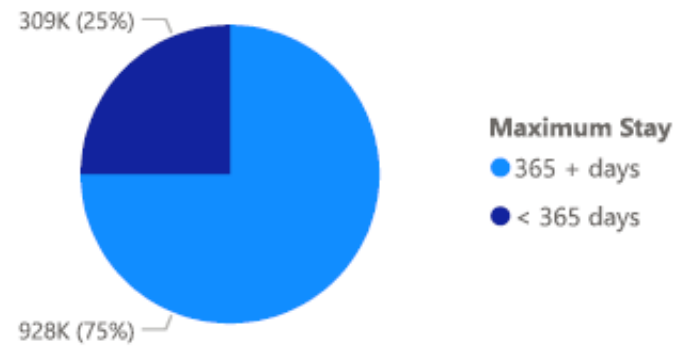
Airbnb & Population Density in Boroughs

# Hotel Effects on Airbnb

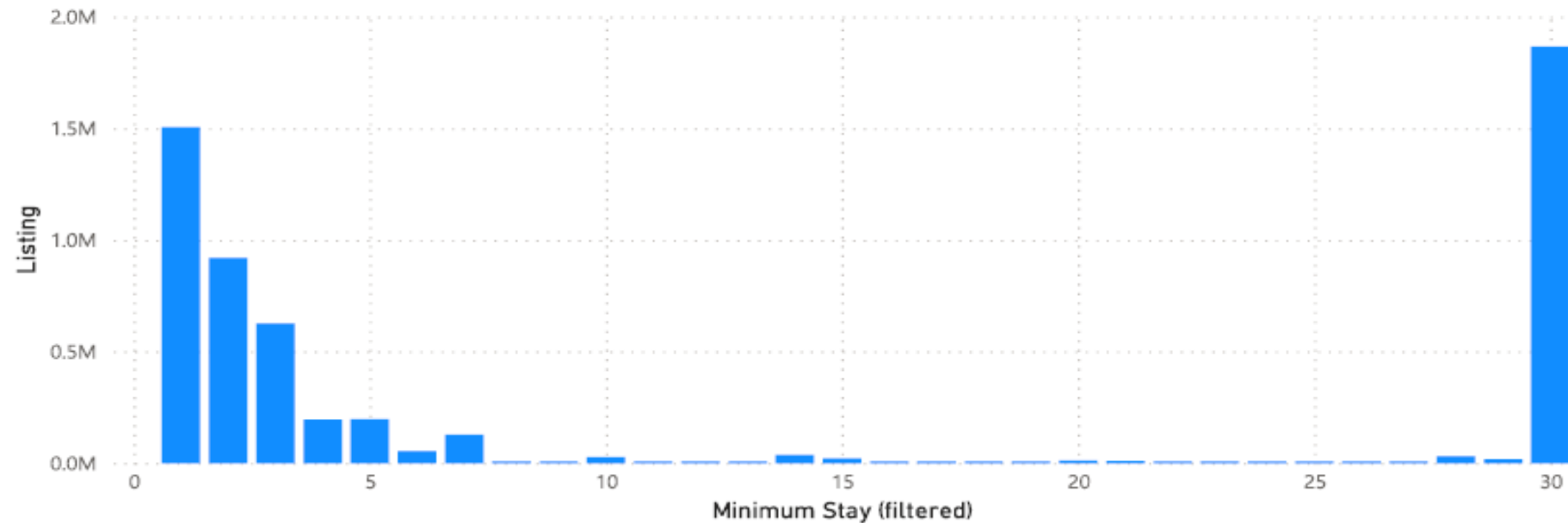




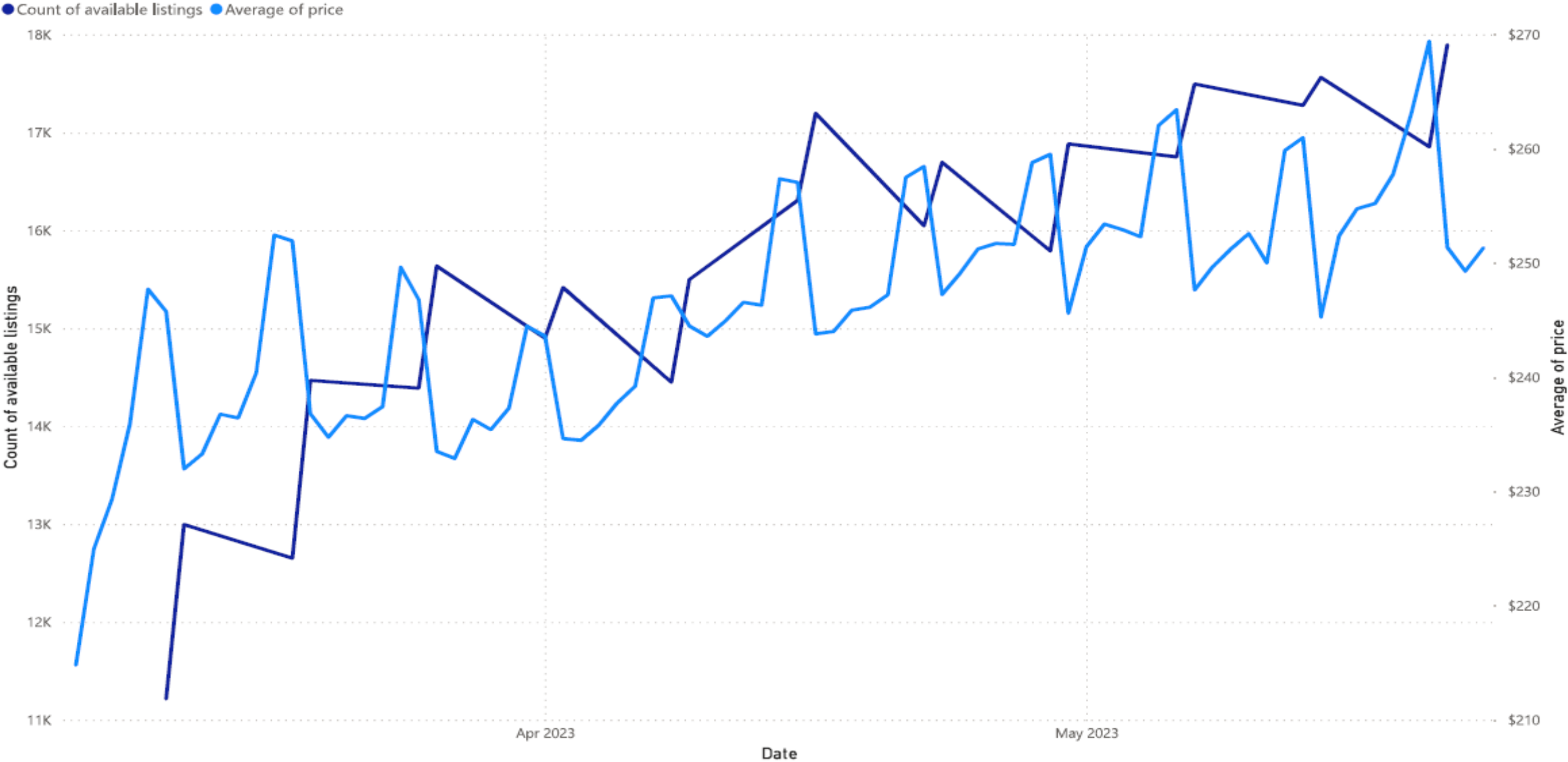
# Minimum and Maximum Stays



Listing by Minimum Stay (filtered)

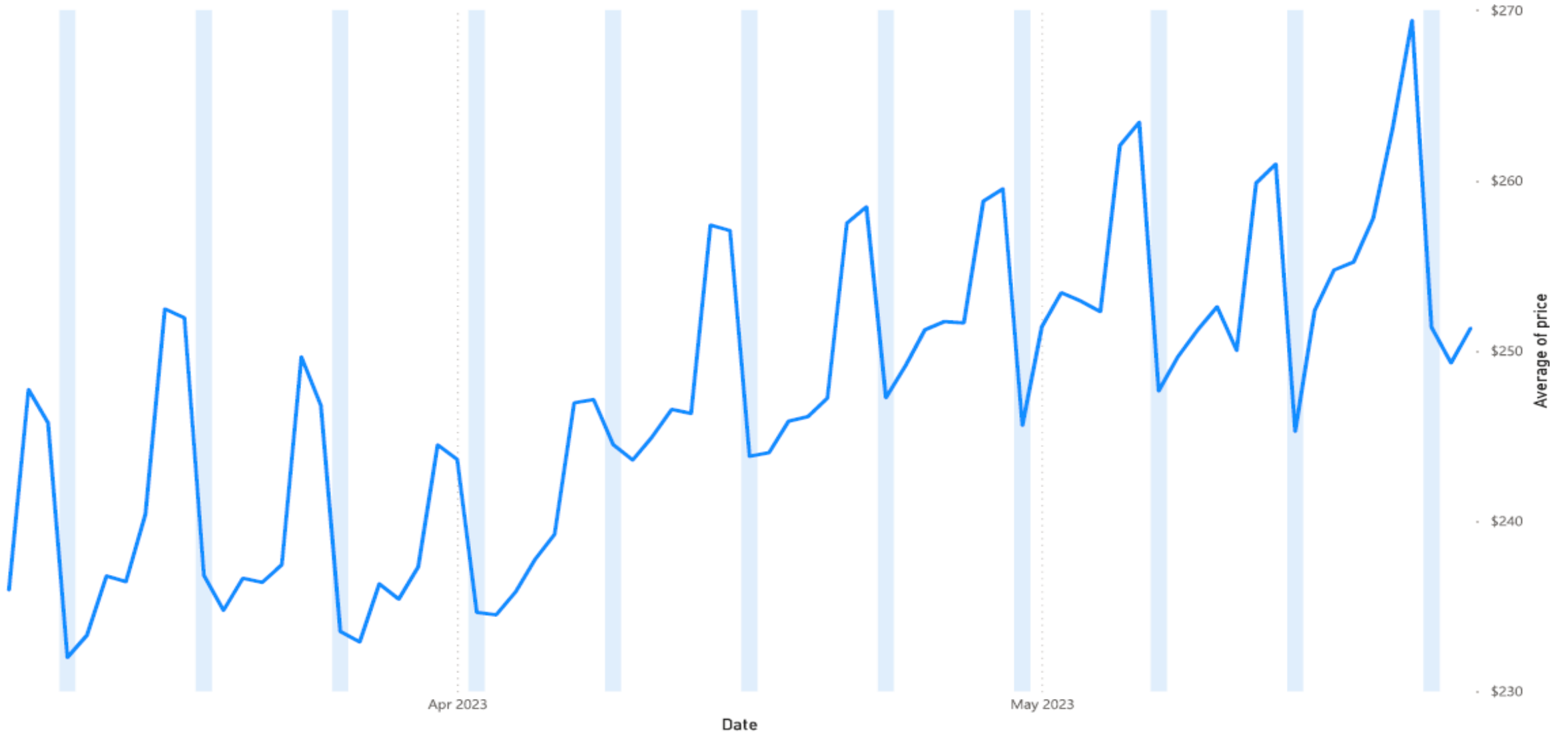


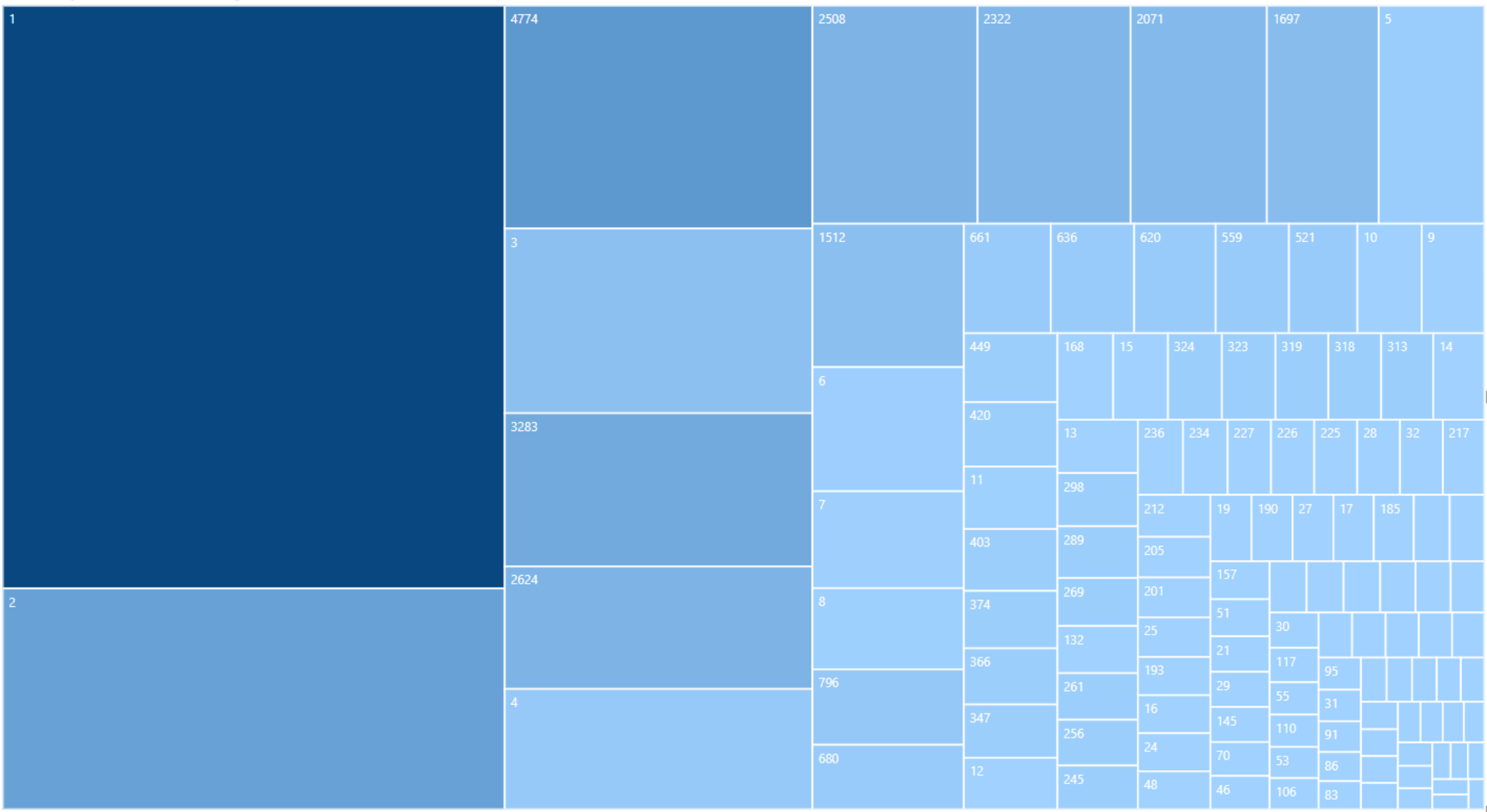
Count of available listings and Average of price by Date



## Average of price by Date

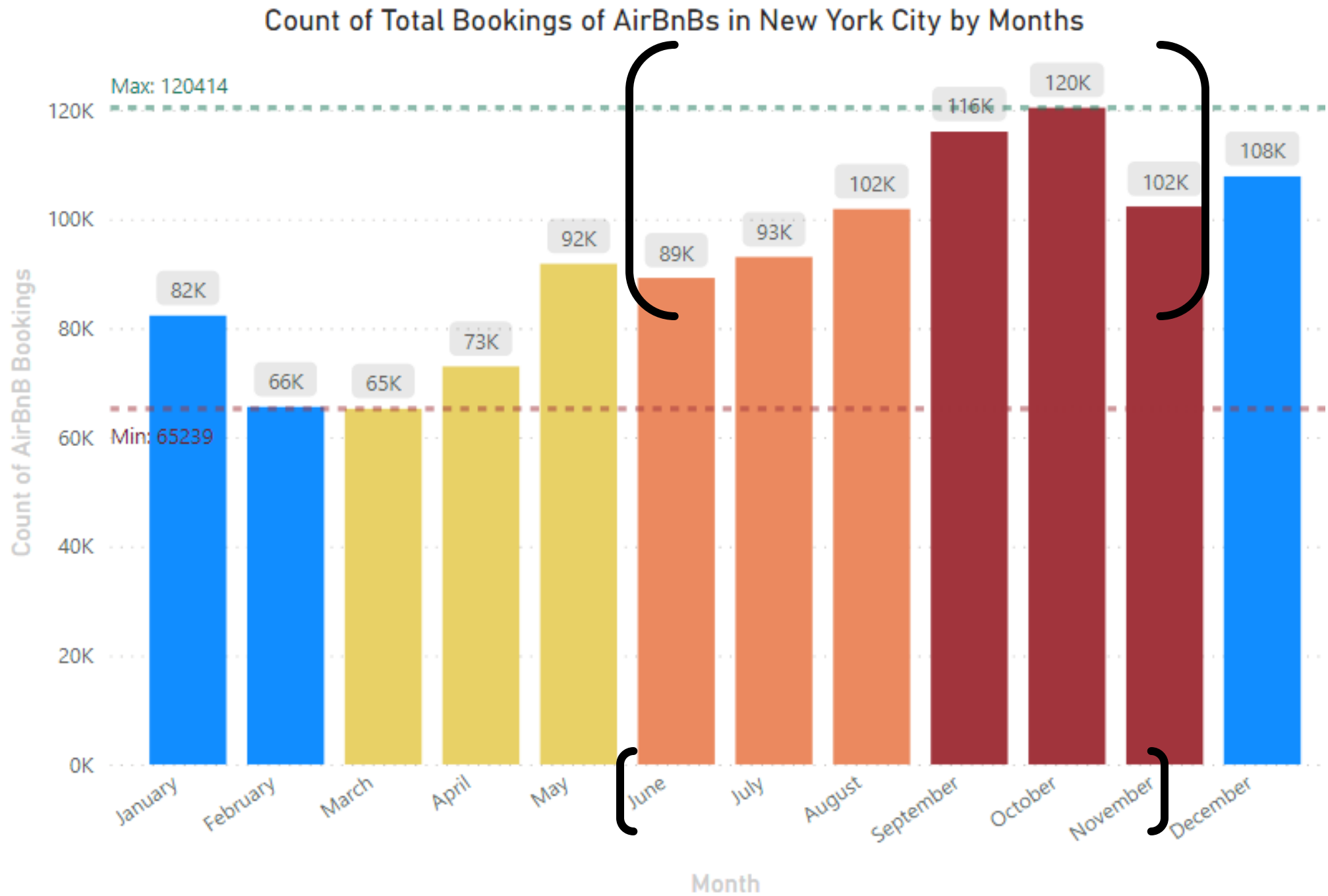
● Sundays ● Average of price





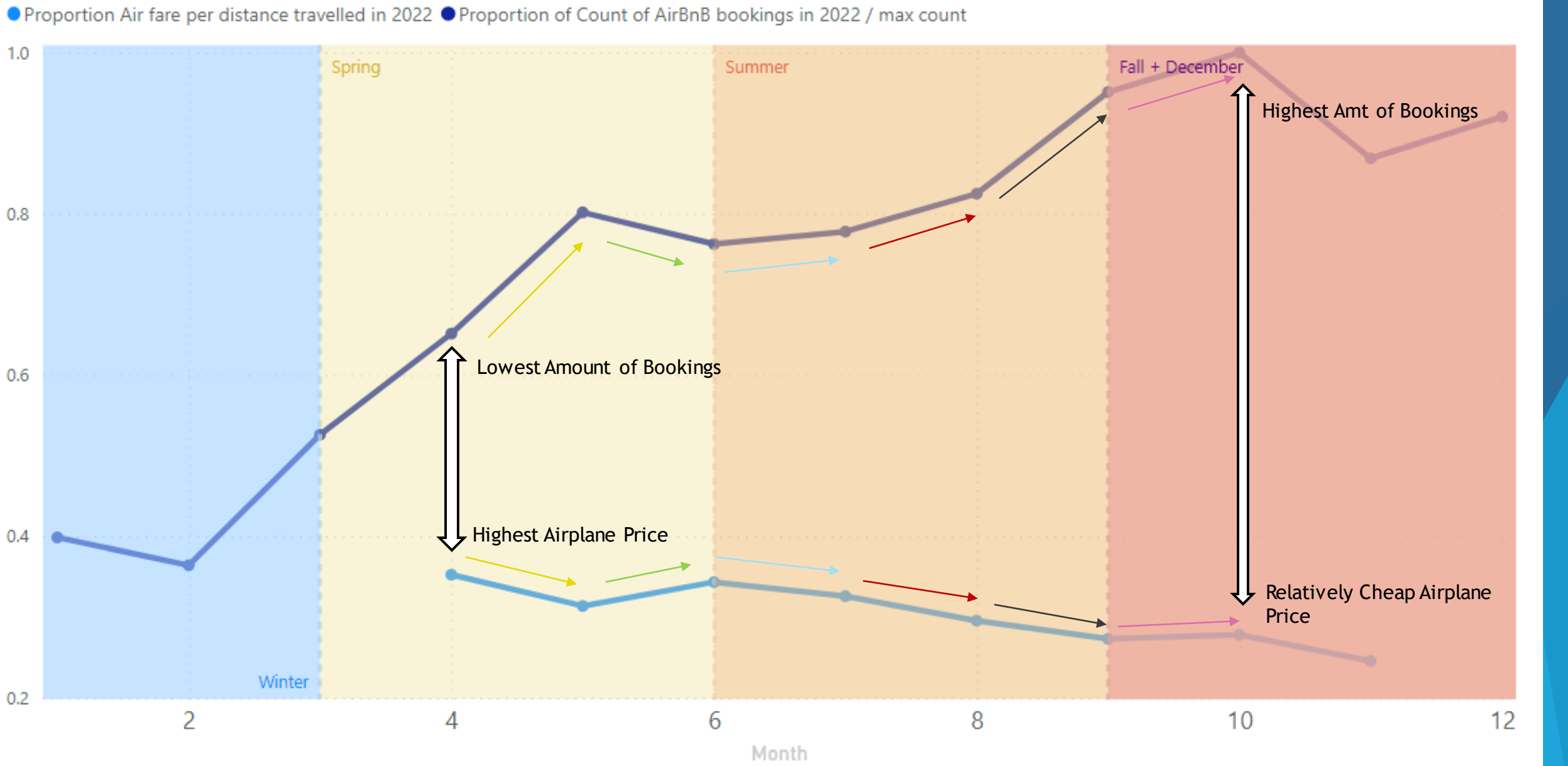


# Seasonality Trends



# Airfare Prices to NYC

Possible Inverse Relationship between NYC AirBnB Booking Count and NYC Airplane Fare in 2022 per Months



# Key Takeaways

- Fall is the most popular season for Airbnb bookings
- Airfare prices to NYC have an inverse relationship to Airbnb booking
- Housing listings decrease as Airbnb bookings have a sharp increase
- Crime does not have an impact on Airbnb bookings
- The majority of Airbnb listings are long-term stays



# Questions

Github: [https://github.com/snowwly/DataCapstone\\_Group3](https://github.com/snowwly/DataCapstone_Group3)