Air BNB Analyze for Pillow Palooza

Video link-
https://drive.google.com/file/d/1Dr4FOclHGeyba_ZFs8rqqoucyboBfnkn/view?usp=sharing

Author: Yishai Ben Dov, data analyst at the company
Date: 31.5.23

Summary:

For having good answer for what we check we need to add more data for the prices in neighbourhood in NYC, and the revenue that we want to get.
We can see that the prices are changes dramatically between borough and neighborhoods.

Context:

When I approached to the data I wanted to see if we have some criteria that will help us decide easily if we want the apartment in our company or not.
It wasn't clear so I decided that the owner of the house is the one who rent the apartment so we need to show him if he will get much more profit in short term rental then in long term that it will justify the effort.
Although we can play with some data a lot of critical data was missing, like:

- What the avg. price by borough and by neighborhood?
- What the avg. price to sq. ft.?
- How much people can sleep in shared room and private room?
- What the extra profit that will make more of our owners to want to change ftom short term to long term?

All of those numbers are numbers that I'm sure we can find, like in this website-
https://www.rentcafe.com/average-rent-market-trends/us/ny/manhattan/
It's critical data to be able to give answer.
What I did now is to clean and arange the data so now when we will have the extra data it will be very easy to give the answer we looking for.

Results:

There is a big difference between borough and neighborhood and the room type. Also I'm sure that there is a big difference in the long term rental between the neighborhoods. That why we can't give right now an answer about what the criteria that we want in apartment in our firm and how to increase our revenue.

Recommendation:

Add the extra data, see that he is accurate and clean and then do the analyze again.

Appendix:

Tableau link-
https://public.tableau.com/views/airbnbproject_16854313797030/AirBNBProject?:language=en-US&:display_count=n&:origin=viz_share_link

Python:

```
import numpy as np
import pandas as pd
import datetime as dt

# Step 1. Importing the data

# Load airbnb_price.csv, prices
prices = pd.read_csv("data/airbnb_price.csv")

# Load airbnb_room_type.xlsx, xls
xls = pd.ExcelFile("data/airbnb_room_type.xlsx")

# Parse the first sheet from xls, room_types
room_types = xls.parse(0)

# Load airbnb_last_review.tsv, reviews
reviews = pd.read_csv("data/airbnb_last_review.tsv", sep="\t")


# Step 2. Cleaning the price column

# Remove whitespace and string characters from prices column
prices["price"] = prices["price"].str.replace(" dollars", "")
```

```
# Convert prices column to numeric datatype
prices["price"] = pd.to_numeric(prices["price"])


# Step 3. Calculating average price

# Subset prices for listings costing $0, free_listings
free_listings = prices["price"] == 0

# Update prices by removing all free listings from prices
prices = prices.loc[~free_listings]

# Calculate the average price, avg_price
avg_price = round(prices["price"].mean(), 2)


# Step 4. Comparing costs to the private rental market

# Add a new column to the prices DataFrame, price_per_month
prices["price_per_month"] = prices["price"] * 365 / 12

# Calculate average_price_per_month
average_price_per_month = round(prices["price_per_month"].mean(), 2)
difference = round((average_price_per_month - 3100),2)


# Step 5. Cleaning the room_type column

# Convert the room_type column to lowercase
room_types["room_type"] = room_types["room_type"].str.lower()

# Update the room_type column to category data type
room_types["room_type"] = room_types["room_type"].astype("category")

# Create the variable room_frequencies
room_frequencies = room_types["room_type"].value_counts()
```

```
# Step 6. What timeframe are we working with?

# Change the data type of the last_review column to datetime
reviews["last_review"] = pd.to_datetime(reviews["last_review"])

# Create first_reviewed, the earliest review date
first_reviewed = reviews["last_review"].dt.date.min()

# Create last_reviewed, the most recent review date
last_reviewed = reviews["last_review"].dt.date.max()


# Step 7. Joining the DataFrames

# Merge prices and room_types to create rooms_and_prices
rooms_and_prices = prices.merge(room_types, how="outer", on="listing_id")

# Merge rooms_and_prices with the reviews DataFrame to create airbnb_merged
airbnb_merged = rooms_and_prices.merge(reviews, how="outer", on="listing_id")

# Drop missing values from airbnb_merged
airbnb_merged.dropna(inplace=True)


# Step 8. Analyzing listing prices by NYC borough

# Extract information from the nbhood_full column and store as a new column, borough
airbnb_merged["borough"] = airbnb_merged["nbhood_full"].str.partition(",")[0]

# Group by borough and calculate summary statistics
boroughs = airbnb_merged.groupby("borough")["price"].agg(["sum", "mean", "median",
"count"])

# Round boroughs to 2 decimal places, and sort by mean in descending order
boroughs = boroughs.round(2).sort_values("mean", ascending=False)
```

Step 9. Price range by borough

# Create labels for the price range, label_names
label_names = ["Budget", "Average", "Expensive", "Extravagant"]

# Create the label ranges, ranges
ranges = [0, 69, 175, 350, np.inf]

Insert new column, price_range, into DataFrame
airbnb_merged["price_range"] = pd.cut(airbnb_merged["price"], bins=ranges, labels=label_names)

# Calculate occurence frequencies for each label, prices_by_borough
prices_by_borough = airbnb_merged.groupby(["borough", "price_range"])["price_range"].count()

Sql business question:

1. What are the most popular neighborhoods for short-term rentals in New York City?

   Query:

   select neighbourhood, count(booked_days_365) as num_days

   from reviews r

   join prices p

   on p.listing_id=r.listing_id

   group by 1

   order by 2 desc

   Answer: Bedford-Stuyvesant, 2206 nights

2. What is the average rental price for short-term rentals in New York City, and how does it vary by neighborhood and property type?

   Avg price: 141.8173160860249187,

   query:select avg(price)

   from prices

   Very change between neighborhoods. Query: select neighbourhood, avg(price)

   from prices

   group by 1

   order by 2

   Very change by room types. query:select room_type, avg(price)

   from prices p

   join room_types r

   on p.listing_id=r.listing_id

   group by 1

   order by 2

3. What are the most commonly rented property types on Airbnb in New York City, and how does this vary by neighborhood?

   Query: select room_type, sum(booked_days_365)

   from reviews re

   join room_types r

   on re.listing_id=r.listing_id

   group by 1

order by 2

By neighborhood: select neighbourhood, room_type, sum(booked_days_365)

from reviews re

join room_types r

on re.listing_id=r.listing_id

join prices p

on p.listing_id=r.listing_id

group by 1,2

order by 1, 2

4. What is the average length of stay for short-term rentals in New York City, and how does this vary by neighborhood and property type?

   I cant know it from this data. I need the data about number of nights for each property

5. How has demand for short-term rentals in New York City changed over time, and are there any seasonal trends that could impact business decisions?

   I cant know this i need to know the dates of the nights

1. What is the most common room type in NYC Airbnb listings?

Query: select room_type, sum(booked_days_365)

from reviews re

join room_types r

on re.listing_id=r.listing_id

group by 1

order by 2

2. What is the average price of a listing by room type?

   select room_type, avg(price)

   from prices p

   join room_types r

   on p.listing_id=r.listing_id

   group by 1

   order by 2

3. Which borough has the highest average price per month?

   select borough, avg(price_per_month)

   from prices

   group by 1

   order by 2 desc

   ;

4. How many listings of each room type are in each borough?

```
select borough, room_type, count(p.listing_id)
from prices p
join room_types r
on p.listing_id=r.listing_id
group by 1, 2
order by 1, 3 desc
;
```

5. How many listings in each room type category have a price of over $500 per night?

```
select room_type, count(p.listing_id)
from prices p
join room_types r
on p.listing_id=r.listing_id
where price > 500
group by 1
order by 1
;
```

6. What is the distribution of listing prices by neighborhood?

```
select neighbourhood, max(price), round(avg(price),2) as avg, min(price)
from prices p
group by 1
order by 1
```

;

7. What is the estimated amount of revenue generated by hosts in each borough?

```
select borough, sum(booked_days_365*price)
from reviews r
join prices p
on p.listing_id=r.listing_id
group by 1
order by 2 desc
;
```

8. What is the average price per month for listings in each neighborhood?

```
select neighbourhood, avg(price_per_month)
from prices p
group by 1
order by 2 desc
;
```

9. How many listings have no reviews?

```
select (count(p.listing_id)-count(r.listing_id)) as diff
from prices p
right join reviews r
on p.listing_id=r.listing_id
```

;

10. How do the estimated book days correlate with the price of an Airbnb listing in New York City?

    select corr(price, booked_days_365)

    from prices p

    join reviews r

    on p.listing_id=r.listing_id

    ;

11. What is the average price per room type for listings that have at least 100 reviews and are available more than 200 days a year?

    select room_type, avg(price)

    from prices p

    join room_types r

    on p.listing_id=r.listing_id

    join reviews re

    on p.listing_id=re.listing_id

    where number_of_reviews >= 100

    and availability_365 > 200

    group by 1

    ;

12. How many hosts have more than one listing, and what's the maximum number of listings by a single host name?

    ```
    select count(calculated_host_listings_count),
    max(calculated_host_listings_count)

    from reviews

    where calculated_host_listings_count >= 2

    ;
    ```

13. Determine the top 5 hosts who have the highest price_per_month for their listings, considering only hosts who have at least 10 listings.

    ```
    select host_name, price_per_month

    from prices p

    join reviews r

    on p.listing_id=r.listing_id

    where calculated_host_listings_count >=10

    order by 2 desc

    limit 5
    ```

14. Find the neighborhood(s) that have the highest variance in listing prices.

    ```
    select neighbourhood, (max(price)-min(price)) as diff

    from prices

    group by 1

    order by 2 desc
    ```

15. Calculate the average price_per_month for each neighborhood, taking into account only listings where the host has a minimum_nights value that is higher than the average minimum_nights value across all listings.

```
select neighbourhood, avg(price_per_month)

from prices p

join reviews r

on p.listing_id=r.listing_id

where minimum_nights > 4.898

group by 1

order by 2 desc

;
```