

Welcome to New York City, one of the most-visited cities in the world. There are many Airbnb listings in New York City to meet the high demand for temporary lodging for travelers, which can be anywhere between a few nights to many months. In this project, we will take a closer look at the New York Airbnb market by combining data from multiple file types like `.csv`, `.tsv`, and `.xlsx`.

Recall that **CSV**, **TSV**, and **Excel** files are three common formats for storing data. Three files containing data on 2019 Airbnb listings are available to you:

data/airbnb_price.csv This is a CSV file containing data on Airbnb listing prices and locations.

- `listing_id`: unique identifier of listing
- `price`: nightly listing price in USD
- `nbhood_full`: name of borough and neighborhood where listing is located

data/airbnb_room_type.xlsx This is an Excel file containing data on Airbnb listing descriptions and room types.

- `listing_id`: unique identifier of listing
- `description`: listing description
- `room_type`: Airbnb has three types of rooms: shared rooms, private rooms, and entire homes/apartments

data/airbnb_last_review.tsv This is a TSV file containing data on Airbnb host names and review dates.

- `listing_id`: unique identifier of listing
- `host_name`: name of listing host
- `last_review`: date when the listing was last reviewed

```
# Import necessary packages
import pandas as pd
import numpy as np
```

```
# Import the necessary datasets
prices = pd.read_csv("data/airbnb_price.csv")
room_types = pd.read_excel("data/airbnb_room_type.xlsx")
last_reviews = pd.read_csv("data/airbnb_last_review.tsv", sep="\t")
```

```
prices.head()
```

index	...	listing_id	...	price	...	nbhood_full
0		2595		225 dollars		Manhattan, Midtown
1		3831		89 dollars		Brooklyn, Clinton Hill
2		5099		200 dollars		Manhattan, Murray Hill
3		5178		79 dollars		Manhattan, Hell's Kitchen
4		5238		150 dollars		Manhattan, Chinatown

Rows: 5

Convert the `price` attribute to the more suitable datatype, `int`. Before converting the datatype, remove the string component, `dollars`.

```
print('Prices before conversion: ')
print(prices['price'].head())
prices['price'] = prices['price'].str.strip(' dollars').astype(int)
print('+*40)
print('Prices after Conversion: ')
print(prices['price'].head())
```

```

Prices before conversion:
0    225 dollars
1     89 dollars
2    200 dollars
3     79 dollars
4    150 dollars
Name: price, dtype: object
*****
Prices after Conversion:
0    225
1     89
2    200
3     79
4    150
Name: price, dtype: int64

```

```
avg_price = prices['price'].mean()
```

```
room_types.head()
```

index	...	↑↓	listing_id	...	↑↓	description	...	↑↓	room_type
		0			2595	Skylit Midtown Castle			Entire home/apt
		1			3831	Cozy Entire Floor of Brownstone			Entire home/apt
		2			5099	Large Cozy 1 BR Apartment In Midtown East			Entire home/apt
		3			5178	Large Furnished Room Near B'way			private room
		4			5238	Cute & Cozy Lower East Side 1 bdrm			Entire home/apt

Rows: 5 [↓](#)

```
room_types['room_type'] = room_types['room_type'].str.lower().astype('category')
```

```

nb_private_rooms = room_types['room_type'].value_counts()['private room']
apt = room_types['room_type'].value_counts()['entire home/apt']
shared = room_types['room_type'].value_counts()['shared room']

```

```
last_reviews.head()
```

index	...	↑↓	listing_id	...	↑↓	host_name	...	↑↓	last_review
		0			2595	Jennifer			May 21 2019
		1			3831	LisaRoxanne			July 05 2019
		2			5099	Chris			June 22 2019
		3			5178	Shunichi			June 24 2019
		4			5238	Ben			June 09 2019

Rows: 5 [↓](#)

```
last_reviews['last_review'] = pd.to_datetime(last_reviews['last_review'])
```