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

# Begin coding here ...
# Use as many cells as you like
```

```
price_df = pd.read_csv(r'data/airbnb_price.csv')
price_df.head()
```

Hidden output

```
room_df = pd.read_excel(r'data/airbnb_room_type.xlsx')
room_df.head()
```

Hidden output

```
review_df = pd.read_csv(r'data/airbnb_last_review.tsv', sep = '\t')
review_df.head()
```

Hidden output

```
merged = pd.merge(price_df, room_df, on = 'listing_id')
merged.head()
```

Hidden output

```
merged = pd.merge(merged, review_df, on = 'listing_id')
merged.head()
```

Hidden output

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

```
merged['last_review'].dtype
```

```
dtype('<M8[ns]')
```

```
min_review = merged['last_review'].min()
min_review
```

```
Timestamp('2019-01-01 00:00:00')
```

```
max_review = merged['last_review'].max()
max_review
```

```
Timestamp('2019-07-09 00:00:00')
```

```
merged['room_type'] = merged['room_type'].str.lower()
```

Hidden output

```
private_only = merged[merged['room_type'] == 'private room']
private_count = private_only.shape[0]
private_count
```

```
11356
```

```
merged
```

Hidden output

```
#5
merged['price'] = merged['price'].str.replace(' dollars', '', regex=False).astype(float)
```

Hidden output

```
avg_price = round(merged['price'].mean(), 2)
```

avg_price

141.78

```
review_dates = pd.DataFrame({
    'first_reviewed' : [min_review],
    'last_reviewed' : [max_review],
    'nb_private_rooms' : [private_count],
    'avg_price' : [avg_price]
})

review_dates
```

	first_reviewed	last_reviewed
0	2019-01-01T00:00:00.000	2019-07-09T00:00:00.000

1 rows