Luther Nicholaus
CIS 607 – Unit 2 Assignment
Park University

Software Used

Python is the software used for this analysis.

Description of Dataset

The dataset used for analysis is the "Airbnb Property Data". The dataset contains data of more than 18,259 property listings. The data fields for each property defined in the dataset are the respective rate per night, city, and geographical coordinates, among other things. Used Python (pandas) to drop irrelevant columns for the test remaining with columns: "property_ID", "average_rate_per_night", "bedrooms_count", and "city".

Column	Data Type Object	Variable Type
property_ID	int64	Discrete
average_rate_per_night	float64	Continuous
bedrooms_count	int64	Discrete
city	object	Categorical

Hypothesis 1 Tested

Airbnb publishes that average global booking value per night in 2020 was $174 (Airbnb, 2021). This claim was published on its annual report intended for their investors. Since the claim contains the statement of equity, it will be the null hypothesis. Therefore, the alternative hypothesis will be: the average global booking value per night in 2020 was not $174. The absence of population standard deviation commands the use of the t test. The level of significance for this test is 0.05.

```
Ttest_1sampResult(statistic=12.19699971462573, pvalue=4.378965441991893e-34)
```

Since the two-sided p value of 4.37e-34 is significantly less than the significance level of 0.05, the null hypothesis is rejected. In other words, there is compelling evidence from the sample to reject the null hypothesis of the test. Therefore, at a 5 % level of significance, there is enough evidence to conclude that the average global booking value per night in 202 was not $174.

Hypothesis 2 Tested.

The second test tests for the mean number of bedrooms of properties listed on Airbnb. A company director claims the average number of bedrooms per property listed is 2. This time, we will use the sample module on Python to select a small random sample of 20 properties from the dataset. The t-test will be applied on this test since the standard deviation is unknown. Five sample tests are performed to guarantee concrete evidence on whether or not to reject the null hypothesis. The level of significance for this test is 0.05.

```
In [12]: runfile('C:/Users/18166/OneDrive/Documents/untitled2.py', wdir='C:/Users/18166/OneDrive/Documents')
Ttest_1sampResult(statistic=0.0, pvalue=1.0)

In [13]: runfile('C:/Users/18166/OneDrive/Documents/untitled2.py', wdir='C:/Users/18166/OneDrive/Documents')
Ttest_1sampResult(statistic=-inf, pvalue=0.0)

In [14]: runfile('C:/Users/18166/OneDrive/Documents/untitled2.py', wdir='C:/Users/18166/OneDrive/Documents')
Ttest_1sampResult(statistic=-2.4494897427831788, pvalue=0.07048399691021988)

In [15]: runfile('C:/Users/18166/OneDrive/Documents/untitled2.py', wdir='C:/Users/18166/OneDrive/Documents')
Ttest_1sampResult(statistic=-0.9999999999999998, pvalue=0.373900966300059)

In [16]: runfile('C:/Users/18166/OneDrive/Documents/untitled2.py', wdir='C:/Users/18166/OneDrive/Documents')
Ttest_1sampResult(statistic=0.342997170285018, pvalue=0.7488684500235263)
```

Since all the two-sided p values from five random samples are considerably greater than the significance level of 0.05, the test fails to reject the null hypothesis. At a 5% level of significance, there is no sufficient evidence to reject the claim that the average number of bedrooms per property listed is 2.

Conclusions for Airbnb and its Investors.

Investors can use the test results for a solid and diligent assessment of the company's operational performance and predict future cash flows. As they draw up their investment decisions, due diligence in this particular test informs an investor to reject the company's claim about the average booking value per night. For the company, the average bedrooms per property metric aids the strategic planning process. With the failure to reject the claim of an average of 2 bedrooms per property, the company can confidently explore and prescribe effective marketing strategies. For example, the company can deploy aggressive promotional campaigns for property owners in areas where visitors tend to visit in large groups (demanding more than 5 rooms) and there aren't enough properties with more than 5 rooms.

Python Code

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import scipy
from scipy import stats as st
```

```
a = pd.read_csv('C:/Users/18166/Downloads/Airbnb.csv')
a.drop(['latitude', 'longitude', 'date_of_listing', 'description', 'title', 'url'], axis=1, inplace=True)
a.average_rate_per_night.astype(float)
a.bedrooms_count.astype(int)

test1 = st.ttest_1samp(a.average_rate_per_night, 174)
print(test1)

c = a.sample(5, replace=False)
test2 = st.ttest_1samp(c.bedrooms_count, 2)
print(test2)
```

References

Airbnb, Inc. (2021). 2020 Annual report. *Airbnb, Inc. (Online)*. Retrieved from https://investors.hess.com/static-files/f0375fc7-bcbb-4b34-a984-c17c26cde53a

Kaggle. (2020). Airbnb Property Data from Texas. *Kaggle.com*. Retrieved from https://www.kaggle.com/PromptCloudHQ/airbnb-property-data-from-texas/version/1