

# Econ 520 Practicum - Analyzing Vienna Accommodations

Jerry Hong

2024-03-04

## Background: Vienna, Austria Accommodations



Vienna, Austria is a popular tourist destination known for its rich history, beautiful architecture, and vibrant city culture. It is also known for its high quality of life and is often ranked as one of the best cities to live in. As a result, Vienna has a wide variety of accommodations to choose from, ranging from luxury hotels to budget hostels. Our point of interest in this report is to explore and find good deals among hotels, hospitals, and other types of accommodations across Vienna, Austria. The data set contains information, such as price, type, and customer ratings of 428 accommodations in Vienna. The data set is from the website Booking.com and was collected in 2017. The primary objective is uncover relationships between price and the other variables, in hopes gain a better understanding on finding good deals.

## Data Exploration

Let's start by loading the data set and taking a look at the first few rows and a summary of the data set. This is a processed data set where we have to modify the data types of some variables in order to do the appropriate analysis. For example, the `guestreviewsrating` variable was of type character, so we convert it to numeric type. This data set also includes a couple new variables that can complement our analysis. For example, we created a variable to better differentiate the type of accommodation.

```
## Rows: 428 Columns: 26
## -- Column specification -----
## Delimiter: ","
## chr (12): addresscountryname, city_actual, center1distance, center1label, ce...
## dbl (14): rating_reviewcount, price, starrating, rating2_ta, rating2_ta_revi...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

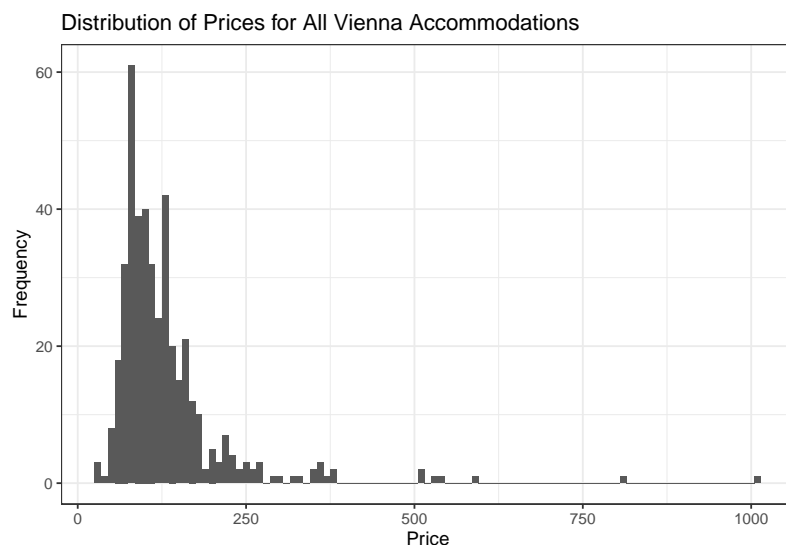
```
## # A tibble: 6 x 26
##   addresscountryname city_actual rating_reviewcount center1distance center1label
##   <chr>               <chr>                <dbl> <chr>          <chr>
## 1 Austria            Vienna                36 2.7 miles    City centre
## 2 Austria            Vienna               189 1.7 miles    City centre
## 3 Austria            Vienna                53 1.4 miles    City centre
## 4 Austria            Vienna                55 1.7 miles    City centre
## 5 Austria            Vienna                33 1.2 miles    City centre
## 6 Austria            Vienna                25 0.9 miles    City centre
## # i 21 more variables: center2distance <chr>, center2label <chr>,
## #   neighbourhood <chr>, price <dbl>, price_night <chr>, s_city <chr>,
## #   starrating <dbl>, rating2_ta <dbl>, rating2_ta_reviewcount <dbl>,
## #   accommodationtype <chr>, guestreviewsrating <dbl>, scarce_room <dbl>,
## #   hotel_id <dbl>, offer <dbl>, offer_cat <chr>, year <dbl>, month <dbl>,
## #   weekend <dbl>, holiday <dbl>, type <chr>, distancecitycenter <dbl>
```

## Exploring Relationships with Price

We will start by exploring the relationship between price and the other variables in the data set. We will use a series of different visualizations to see any relationships and trends in regards to pricing.

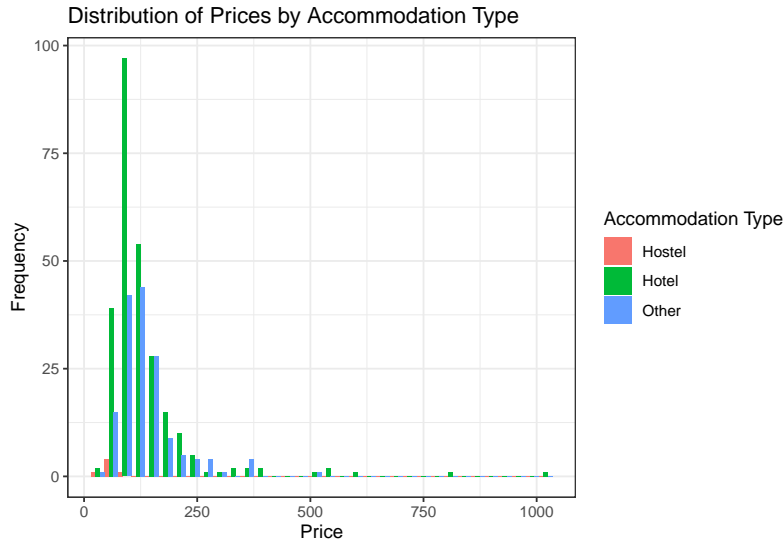
```
## # A tibble: 3 x 5
##   type    mean_price sd_price median_price count
##   <chr>      <dbl>    <dbl>      <dbl> <int>
## 1 hostel     53.7     18.2       50.5     6
## 2 hotel    130.     104.      102     264
## 3 other     137.     67.6      130    158
```

In the above table, hotels dominate throughout Vienna, with 264 in the area. Hostels are the least common, with only 6 in the area.



Based on the above distribution, we can see that the majority of the prices are between 0 and 200, with a few outliers as high as 1000. This shows that many of the accommodations are priced at a reasonable rate, thus making it an attractive location for tourists.

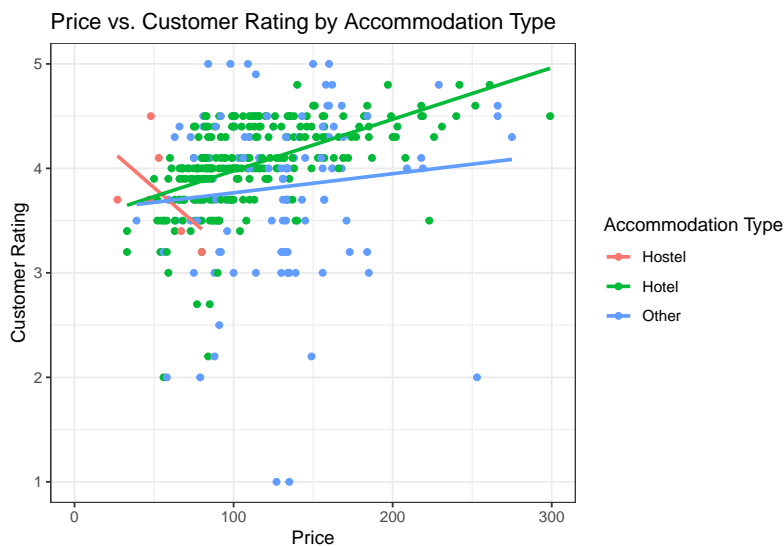
Let's explore how the price distribution varies by type of accommodation.



Below is a table showing the average price and customer rating by accommodation type. Among the three types, hotels have the highest average customer rating, though the other types are not as far behind. Additionally, hotels and other types are priced similarly per night while hostels are priced much lower. This can be a compelling option considering average customer ratings are not far behind from hotels and others.

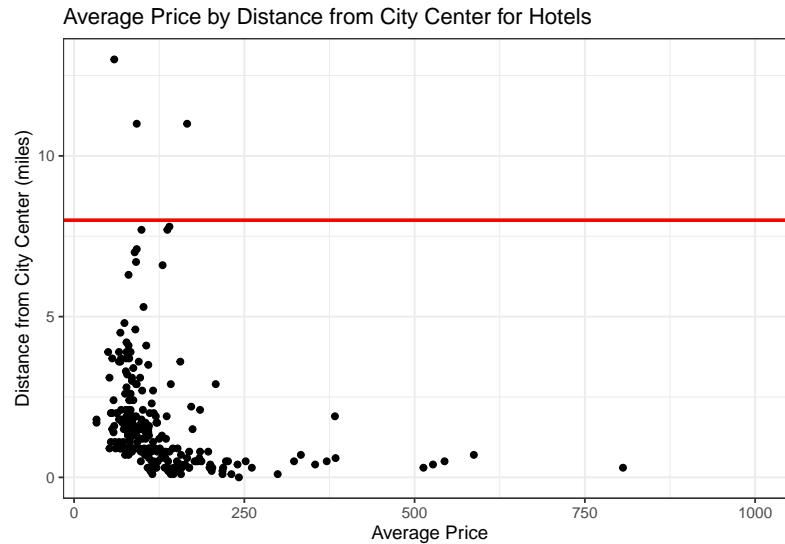
```
## # A tibble: 3 x 3
##   type    mean_price mean_rating
##   <chr>      <dbl>      <dbl>
## 1 hostel      53.7         3.77
## 2 hotel      130.         4.05
## 3 other      137.         3.82

## 'geom_smooth()' using formula = 'y ~ x'
```



Once we exclude any outliers from this scatter plot, we can uncover possible trends between price and customer rating. The relationship between the two is interesting. For hostels, there is a downward

trend, implying a higher price does not necessarily mean a higher rating. For hotels and other accommodations, there is a positive trend, implying a higher price is associated with a higher rating. However, it is not as strong of a relationship as we would expect, suggesting that there are other factors that contribute to the customer rating.



The scatter plot above shows the relationship between price and distance from the city center for hotels. The red line represents a reasonable distance from the city center given the general proximity of hotels. In this case, we set the threshold at 8 miles from the city center. We can see that the trend is generally more negative as the distance from the city center increases. This implies that hotels closer to the city center are priced higher than those further away.

## Concluding Remarks

This case study is a way to showcase the skills to take a raw data set and make the necessary cleaning techniques to make the appropriate analysis. One takeaway from this exercise is to keep track of the changes made in the data set as any visualizations may not work if the changes were not applied. Additionally, finding which visualization that can provide the best insights to the targeted audience is important with the variables that are present in the data set. Throughout this case study, there are some noteworthy relationships between price and the other variables in this data set. For example, we found that while not as strong, higher priced accommodations tend to have higher customer ratings, which demonstrates the level of quality that comes with the price. We also found that hotels closer to the city center tend to be priced more, which is expected given the convenience of being closer to the city center. Some future insights in this project are exploring what separates hotel pricing if these other variables are held constant. For instance, we may further look into amenities or services that are offered that may contribute to the price.