# Analysis of Asheville AirBnB Data

Dylan Henegar

2024-07-27

## Asheville AirBnB Neighborhood Demand Report: Jan 2020-Mar 2023

### Introduction

This is a report to study AirBnB data for Asheville, NC from January 2020-2023. This report will highlight findings about the high-demand neighborhoods in the Asheville market with key metrics.

"Asheville is a city in western North Carolina's Blue Ridge Mountains. It's known for a vibrant arts scene and historic architecture, including the dome-topped Basilica of Saint Lawrence. The vast 19th-century Biltmore estate displays artwork by masters like Renoir. The Downtown Art District is filled with galleries and museums, and in the nearby River Arts District, former factory buildings house artists' studios." — Google

"Western North Carolina is the region of North Carolina which includes the Appalachian Mountains; it is often known geographically as the state's Mountain Region. It contains the highest mountains in the Eastern United States, with 125 peaks rising to over 5,000 feet in elevation."–Wikipedia

### Objectives

1. Identify high-demand neighborhoods.
2. Analyze optimal property types.
3. Develop a competitive pricing strategy.
4. Assess occupancy and availability trends.

### Data Preparation

The dataset includes columns such as id, name, host_id, host_name, neighborhood_group, neighborhood, latitude, longitude, room_type, price, minimum_nights, number_of_reviews, last_review, reviews_per_month, calculated_host_listings_count, availability_365, number_of_reviews_ltm, and city. Data cleaning involved handling missing values and outliers

### Methodology

Data is from AirBnB's open source data, it was downloaded from Kaggle here. The data was loaded into R Studio, the data for Asheville was extracted for analysis and then filtered for only the years 2020 to 2023. The data does not go past March 2023.

**Key Metrics used to analyze data:**

- **Total Listings**: Total number of listings per neighborhood
- **Average Reviews**: Average number of reviews per listing
- **Average Price**: Average price per night
- **Occupancy**: Average availability (days available per year)

## Neighborhood Analysis

```
top_neighborhoods <- asheville_data %>%
  group_by(neighbourhood) %>%
  summarize(
    count = n(),
    avg_price_USD = mean(price, na.rm = TRUE),
    max_price_USD = max(price, na.rm = TRUE),
    min_price_USD = min(price, na.rm = TRUE),
    avg_avail = mean(availability_365, na.rm = TRUE),
    reviews_per_month = mean(reviews_per_month, na.rm = TRUE),
    most_common_rental_type = room_type[which.max(table(room_type))]
    ) %>%
  arrange(desc(count)) %>%
  top_n(10,count)
print(top_neighborhoods)
```

```
## # A tibble: 8 x 8
##   neighbourhood count avg_price_USD max_price_USD min_price_USD avg_avail
##   <chr>         <int>         <dbl>         <dbl>         <dbl>     <dbl>
## 1 28806           851          148.          6588            10      195.
## 2 28801           600          204.          1941            33      210.
## 3 28803           411          163.          1365            21      209.
## 4 28804           377          182.          1129            36      201.
## 5 28805           281          169.           779            40      194.
## 6 28704           202          163.           999            43      230.
## 7 28715            87          160.           600            42      222.
## 8 28732            73          170.           825            44      242.
## # i 2 more variables: reviews_per_month <dbl>, most_common_rental_type <chr>
```

The above table outlines each neighborhood by highlighting the key metrics used to determine high-demand neighborhoods. The table shows the total number of listings, average price, max/min price, average availability, average reviews per month, and most common rental type.

### Top Neighborhoods

- **28801 (Downtown Asheville)**: this area has the highest overall price, second most listings and is near a lot of desirable amenities.
- **28806 (West Asheville)**: this area has the most listings, the most reviews, and affordable pricing; this area is also nearest to the mountains
- **28803 (South Asheville)**: this area has the third most listings, moderate pricing, and is near the Biltmore Estate a highly desirable tourist attraction in the area.

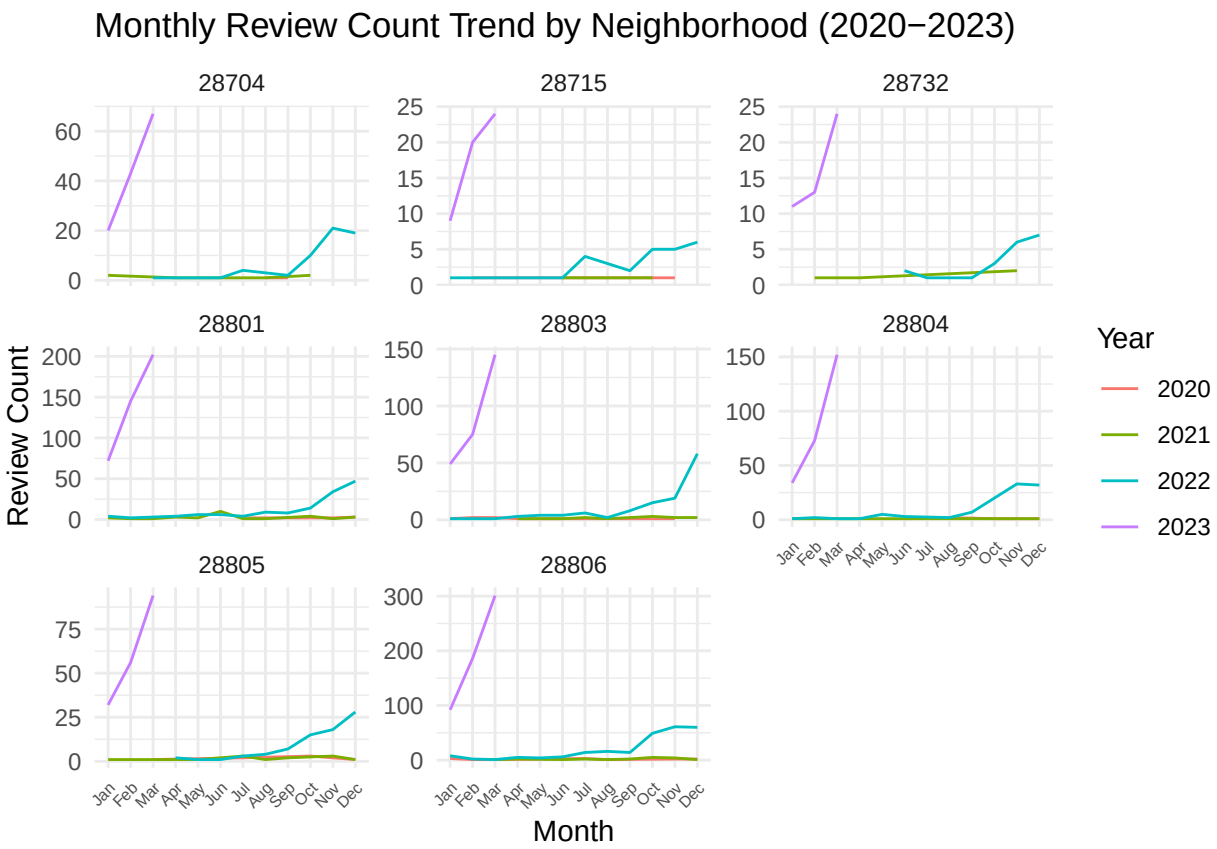
The table below details total reviews, average minimum stay, average price, and average reviews per month. This table shows that 28806 has the highest number of reviews, followed by 28801 and 28803.

```
reviews_count <- asheville_data %>%
  group_by(neighbourhood) %>%
  summarize(
    total_reviews = sum(number_of_reviews, na.rm = TRUE),
    avg_min_stay = mean(minimum_nights, na.rm = TRUE),
    avg_price_USD = mean(price, na.rm = TRUE),
    review_per_month = mean(reviews_per_month, na.rm = TRUE)
    ) %>%
  arrange(desc(total_reviews))
print(reviews_count)
```

```
## # A tibble: 8 x 5
##   neighbourhood total_reviews avg_min_stay avg_price_USD review_per_month
##   <chr>                 <dbl>        <dbl>         <dbl>            <dbl>
## 1 28806                 93098         4.30          148.             2.96
## 2 28801                 70289         5.28          204.             2.47
## 3 28803                 42532         5.73          163.             2.59
## 4 28804                 41061         3.24          182.             2.75
## 5 28805                 28361         4.65          169.             2.28
## 6 28704                 13062         2.43          163.             2.22
## 7 28715                  6737         2.22          160.             2.24
## 8 28732                  5575         2.68          170.             2.30
```

The below charts show review count for each neighborhood by month over the time period. This illustrates the exponential growth in reviews from mid-2022 through 2023 across the entire region, indicating this area is highly desirable for rentals.

```
ggplot(monthly_review_count, aes(x = month, y = review_count, group = year, color = as.factor(year))) +
  geom_line() +
  facet_wrap(~ neighbourhood, scales = "free_y") +
  labs(title = "Monthly Review Count Trend by Neighborhood (2020-2023)",
       x = "Month",
       y = "Review Count",
       color = "Year") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))
```




## Optimal Room Type

The rental listings in the region are divided into 4 categories:

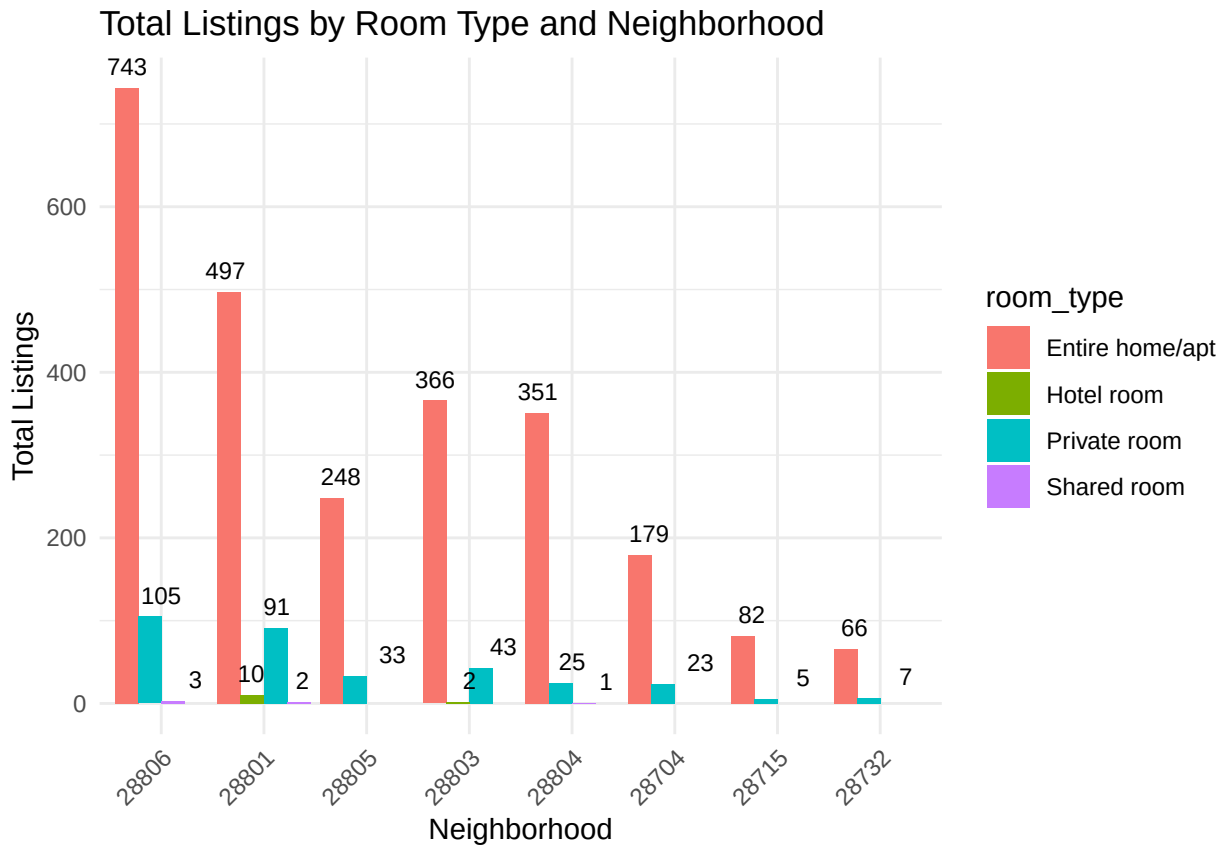
- Entire Home/Apt: the entire home/apt is available to use
- Private Room: room is private inside shared residence
- Shared Room: room is shared with other guests
- Hotel Room: room is private inside building with other rentals

### Top Room Types

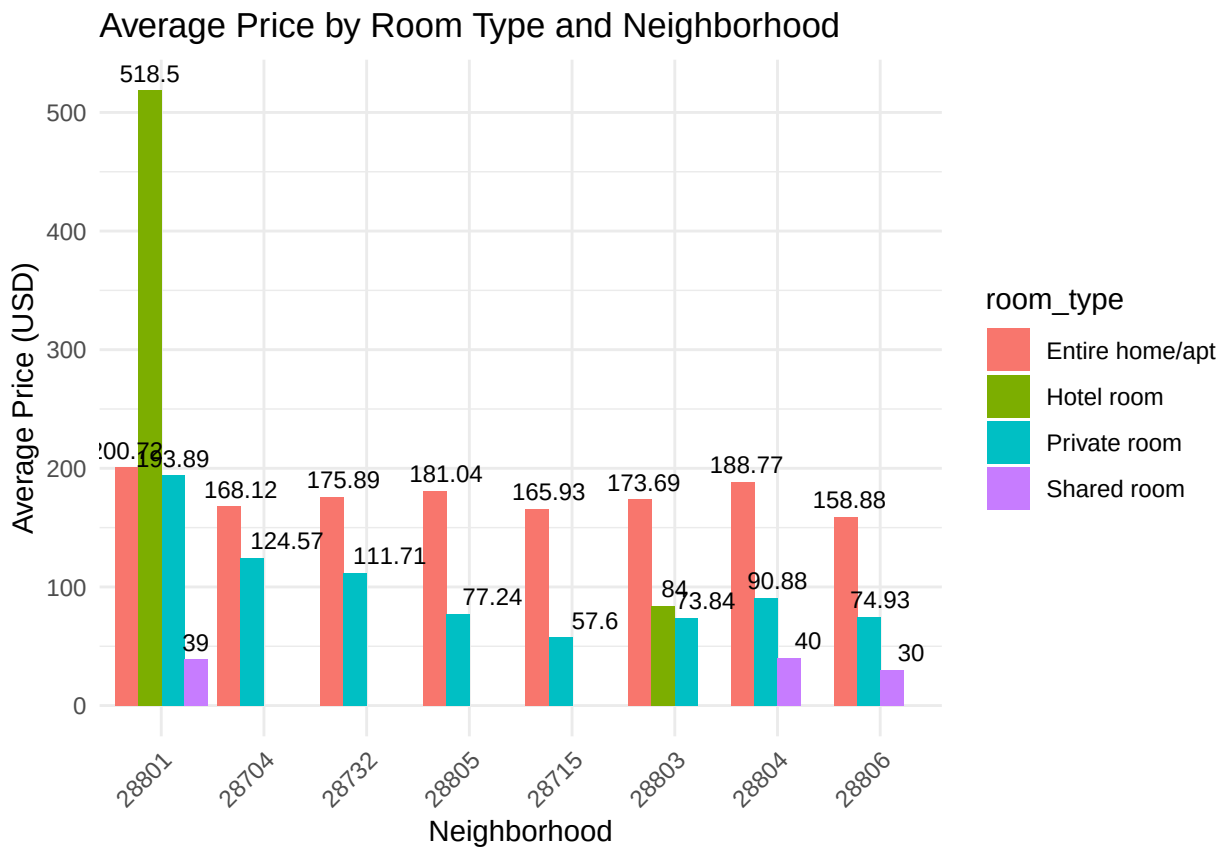
1. Entire Home/Apt: Most profitable, Highest overall price
2. Private Room: Most economical, high occupancy rate
3. Shared Room: Low cost, lower occupancy rate

```
ggplot(rental_type_analysis, aes(x = reorder(neighbourhood, -total_listings), y = total_listings, fill =
  geom_bar(stat = "identity", position = position_dodge(preserve = "single")) +
  geom_text(data = rental_type_analysis, aes(x = reorder(neighbourhood, -total_listings), y = total_lis
            position = position_dodge(width = 1), vjust = -.75, color = "black", size = 3) +
  labs(title = "Total Listings by Room Type and Neighborhood",
```

```
       x = "Neighborhood",
       y = "Total Listings") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(rental_type_analysis, aes(x = reorder(neighbourhood, -avg_price_USD), y = avg_price_USD, fill = r
  geom_bar(stat = "identity", position = position_dodge(preserve = "single")) +
  geom_text(aes(label = round(avg_price_USD, 2)),
            position = position_dodge(width = 0.9), vjust = -0.5, color = "black", size = 3) +
  labs(title = "Average Price by Room Type and Neighborhood",
       x = "Neighborhood",
       y = "Average Price (USD)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

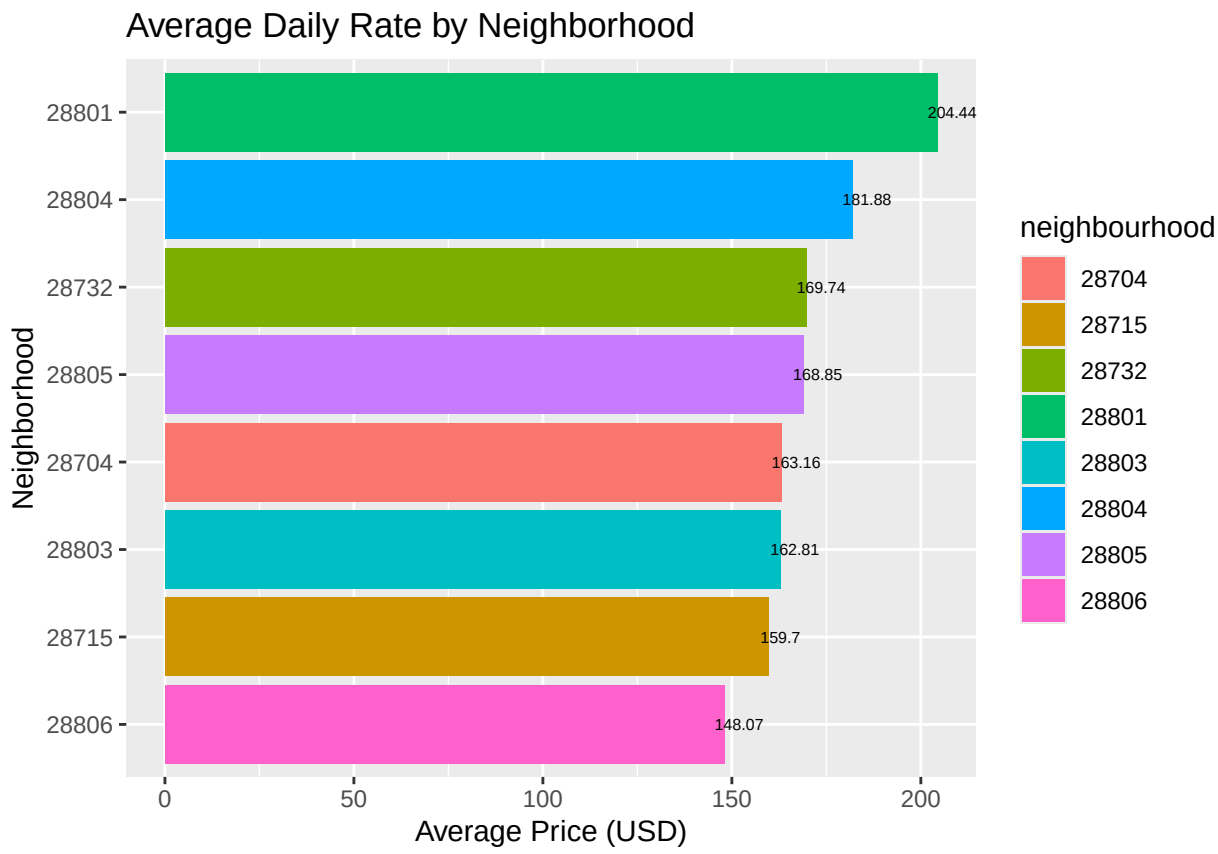



*the Hotel room type has very high daily price but it is the smallest group in terms of number of listings*

## Pricing

Examining the average daily rental price shows that the range of averages across the region is $218.27 - $148.34, the average across the area is $175.03. The top 5 areas for average rental price are 28801 ($218.27), 28732 ($193.15), 28804 ($186.11), 28805 ($170.65), and 28803 ($163.30).

```
ggplot(avg_daily_rate, aes(x = reorder(neighbourhood, avg_price_USD), y = avg_price_USD, fill = neighbou
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(avg_price_USD, 2)), hjust = 0.2, color = "black", size = 2) +
  coord_flip() +
  labs(title = "Average Daily Rate by Neighborhood", x = "Neighborhood", y = "Average Price (USD)")
```

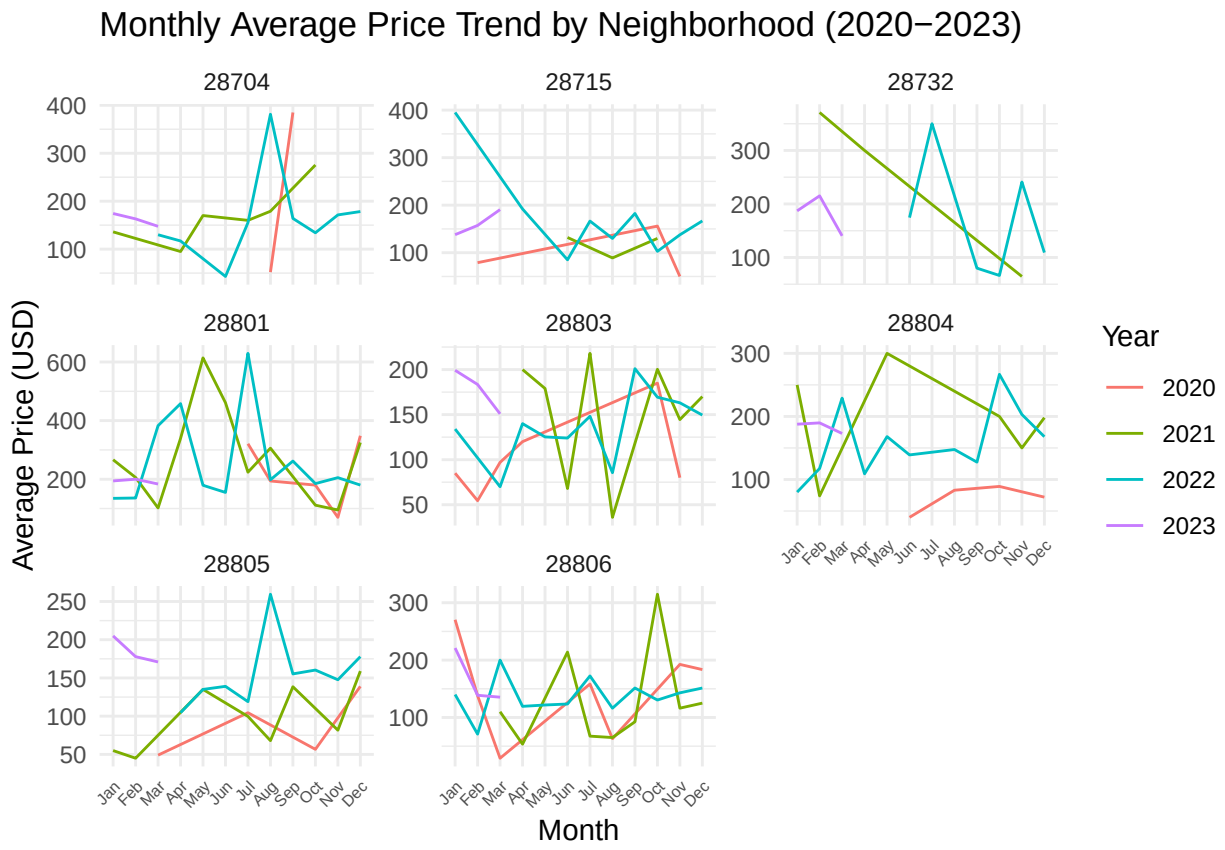



```
avg_daily_rate <- asheville_data %>%
  group_by(neighbourhood) %>%
  summarize(avg_price_USD = mean(price, na.rm = TRUE)) %>%
  arrange(desc(avg_price_USD))
print(avg_daily_rate)
```

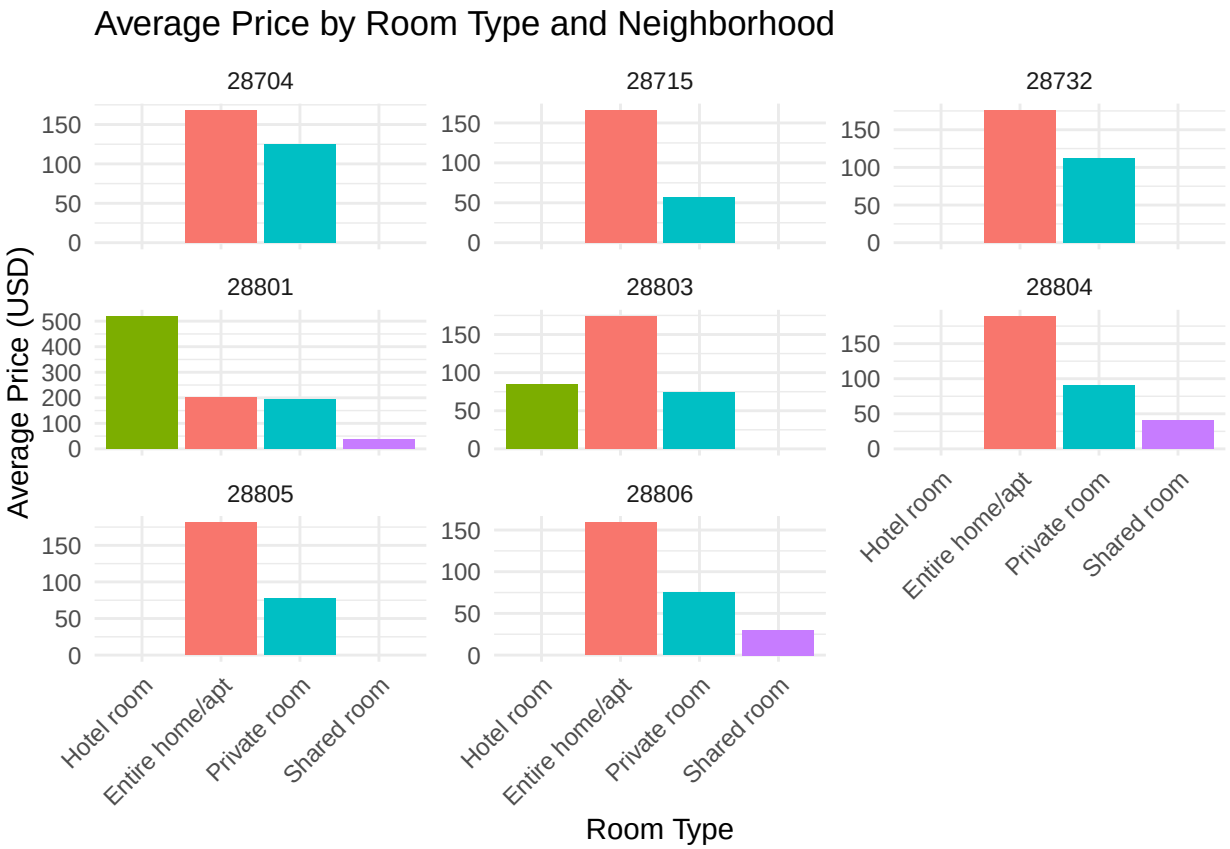
```
## # A tibble: 8 x 2
##   neighbourhood avg_price_USD
##   <chr>                 <dbl>
## 1 28801                  204.
## 2 28804                  182.
## 3 28732                  170.
## 4 28805                  169.
## 5 28704                  163.
## 6 28803                  163.
## 7 28715                  160.
## 8 28806                  148.
```

```
ggplot(monthly_avg_price, aes(x = month, y = avg_price_USD, group = year, color = as.factor(year))) +
  geom_line() +
  facet_wrap(~ neighbourhood, scales = "free_y") +
  labs(title = "Monthly Average Price Trend by Neighborhood (2020-2023)",
       x = "Month",
       y = "Average Price (USD)",
       color = "Year") +
```

```
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))
```




```
ggplot(rental_type_analysis, aes(x = reorder(room_type, -avg_price_USD), y = avg_price_USD, fill = room_
  geom_bar(stat = "identity") +
  facet_wrap(~ neighbourhood, scales = "free_y") +
  labs(title = "Average Price by Room Type and Neighborhood",
       x = "Room Type",
       y = "Average Price (USD)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none")
```




## Occupancy and Seasonal Trends

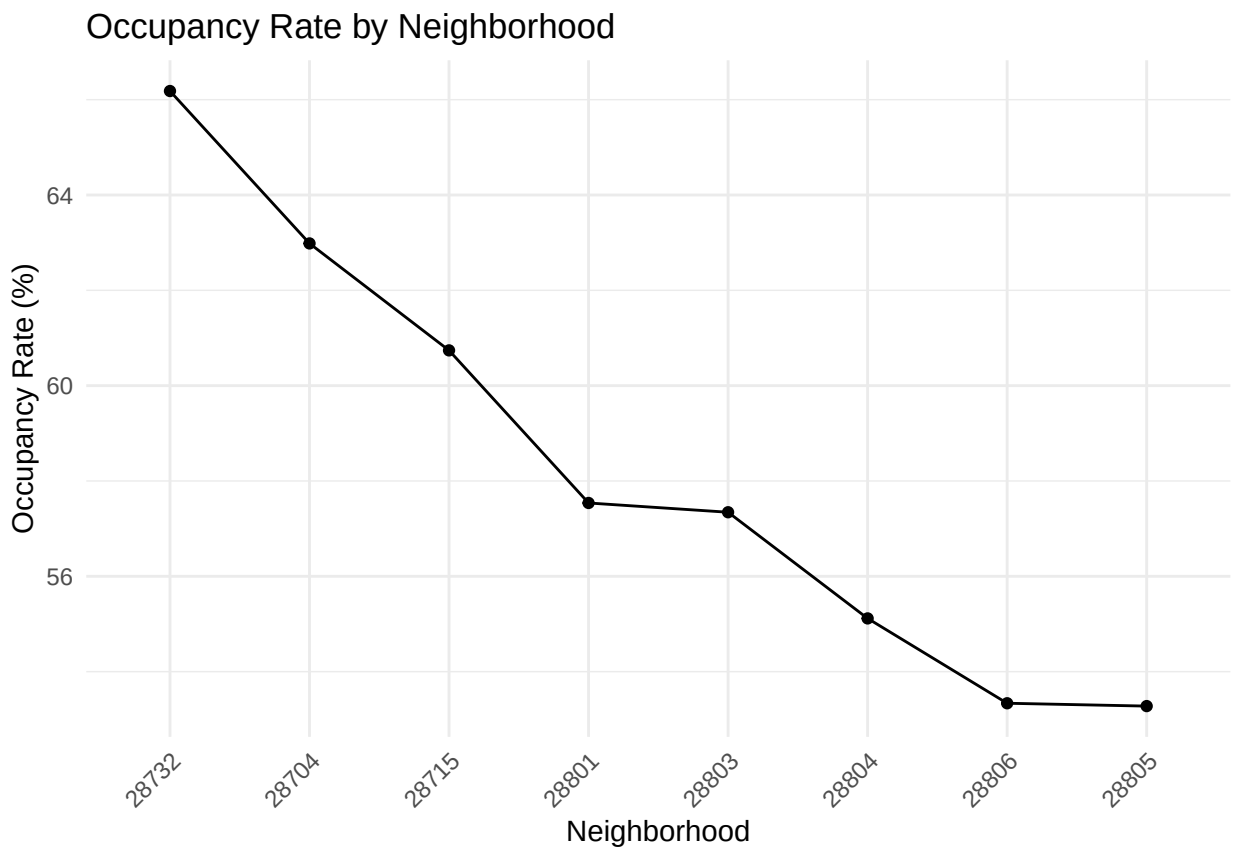
### Higest Occupancy

1. **28732 (Fletcher, Nc)**: 66.18%
2. **28704 (Arden, NC)**: 62.98%
3. **28715 (Candler, NC)**: 60.74%

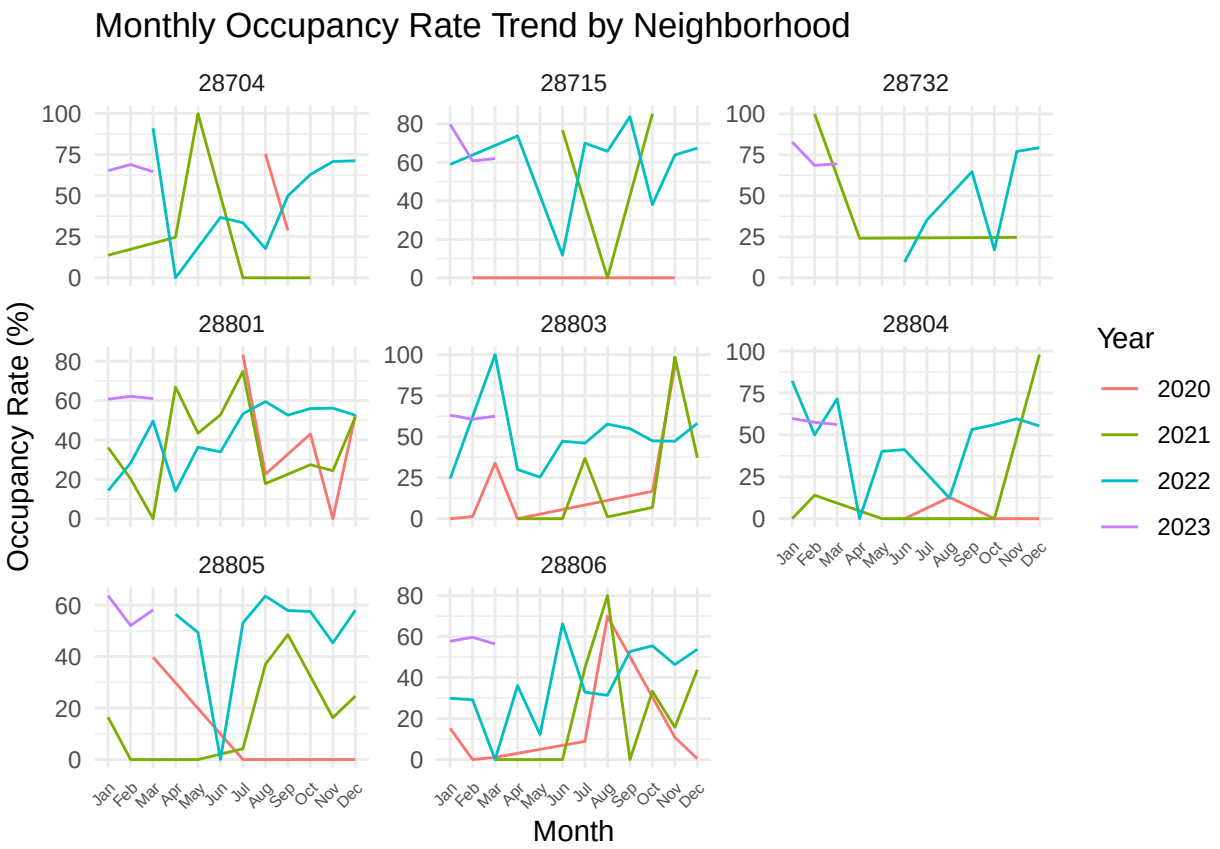
Analyzing the occupancy rate by neighborhood show that the neighborhood 28732 (Fletcher, NC) has the highest occupancy rate 66.18%, followed by 28704(Arden, NC) at 62.98%, 28715(Candler, NC) at 60.74%. This indicates that these areas have the most time during the year booked. However, they do not have the largest number of rental listings(73, 202, 87) as other areas. The area with the largest number of listings is neighborhood 28806 (West Asheville, NC) with a total of 851 and an occupancy rate of 53.33%. Neighborhood 28801 (Downtown Asheville, NC) is second in total listings with 600 and an occupancy rate of 57.54%. Third in total listings is 28803(Asheville/Biltmore Forest,NC) with 411 listings and an occupancy rate of 57.34%. This area is most popular during the late spring, summer, and fall, the winter is not as popular for rentals.

```
ggplot(occupancy_rate, aes(x = reorder(neighbourhood, -occupancy_rate), y = occupancy_rate)) +
  geom_line(group = 1) +  # group = 1 to ensure the line connects all points
  geom_point() +
  labs(title = "Occupancy Rate by Neighborhood",
       x = "Neighborhood",
```

```
       y = "Occupancy Rate (%)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(monthly_occupancy, aes(x = month, y = occupancy_rate, group = year, color = as.factor(year))) +
  geom_line() +
  facet_wrap(~ neighbourhood, scales = "free_y") +
  labs(title = "Monthly Occupancy Rate Trend by Neighborhood",
       x = "Month",
       y = "Occupancy Rate (%)",
       color = "Year") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))
```




```
occupancy_rate <- asheville_data %>%
  group_by(neighbourhood) %>%
  summarize(
    total_booked_days = sum(availability_365, na.rm = TRUE),
    total_listings = n(),
    occupancy_rate = total_booked_days / (total_listings * 365) *100
  ) %>%
  arrange(desc(occupancy_rate))
print(occupancy_rate)
```

```
## # A tibble: 8 x 4
##   neighbourhood total_booked_days total_listings occupancy_rate
##   <chr>                     <dbl>          <int>          <dbl>
## 1 28732                     17634             73           66.2
## 2 28704                     46438            202           63.0
## 3 28715                     19288             87           60.7
## 4 28801                    126010            600           57.5
## 5 28803                     86025            411           57.3
## 6 28804                     75844            377           55.1
## 7 28806                    165673            851           53.3
## 8 28805                     54644            281           53.3
```

## Conclusions

Looking at this area overall, 28806, 28801, and 28803 are the top 3 areas which are good to invest in the rental market. 28732 and 28715 are emerging areas which do not have as many listings but experience high occupancy and are located in desirable areas that are close to the popular attraction/amenities. The Asheville area is home to the Biltmore House(specifically 28801) and also situated in the Western North Carolina area which is mountainous and a great area to hike, bike, camp, etc. This area would be a great investment due to the popular activities year-round.

### Key Takeaways

- 28806, 28801, 28803 are top markets
- Most popular rental type is Entire home/apt
- 28801 has the highest average daily rental
- 28732 is the market outside of the top 3 that would be the best investment