Kyoto AirBnB Analysis

Marina Huang

2023-04-19

Exploratory Data Analysis

Which Type of Listings Are There in the Cities?

We do an analysis to find out the type of listings that are common to a particular city.

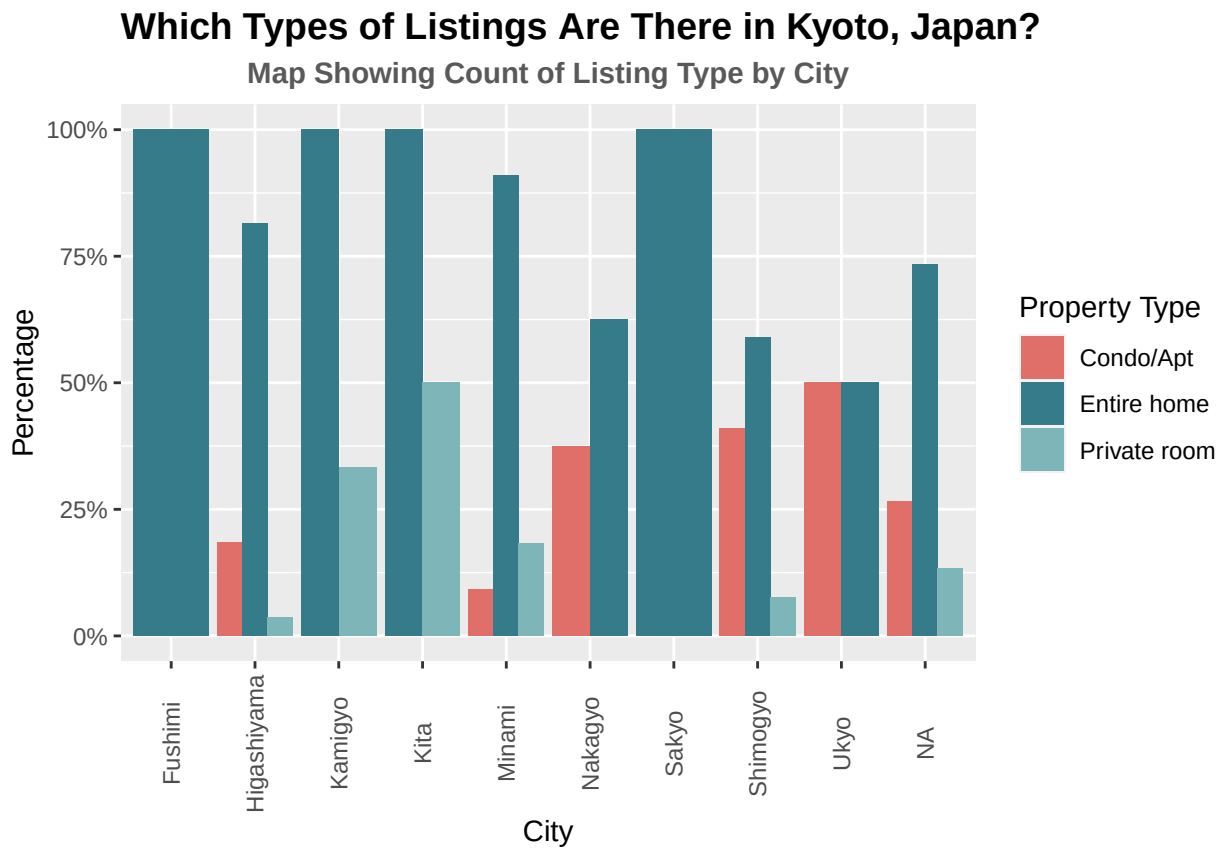
```
property_df <- airbnb %>%
  group_by(city, property) %>%
  summarize(Freq = n())

total_property <- airbnb %>%
  filter(property %in% c("Entire home", "Condo/Apt", "Private Room")) %>%
  group_by(city) %>%
  summarize(sum = n())

property_ratio <- merge(property_df, total_property, by = "city")

property_ratio <- property_ratio %>%
  mutate(ratio = Freq/sum)

ggplot(property_ratio, aes(x = city, y = ratio, fill = property)) +
  geom_bar(position = "dodge", stat = "identity") +
  xlab("City") + ylab("Count") +
  scale_fill_discrete(name = "Property Type") +
  scale_y_continuous(labels = scales::percent) +
  ggtitle("Which Types of Listings Are There in Kyoto, Japan?",
          subtitle = "Map Showing Count of Listing Type by City") +
          theme(plot.title = element_text(face = "bold", size = 14)) +
          theme(plot.subtitle = element_text(face = "bold",
                                             color = "grey35", hjust = 0.5)) +
          theme(plot.caption = element_text(color = "grey69")) +
          scale_color_gradient(low = "#d3cbcb", high = "#852eaa") +
          scale_fill_manual("Property Type", values = c("#e06f69","#357b8a", "#7db5b8",
                                                        "#59c6f3", "#f6c458")) +
          xlab("City") + ylab("Percentage") +
          theme(axis.text.x = element_text(angle = 90,
                                           vjust = 0.5,
                                           hjust = 0.5))
```



Observations

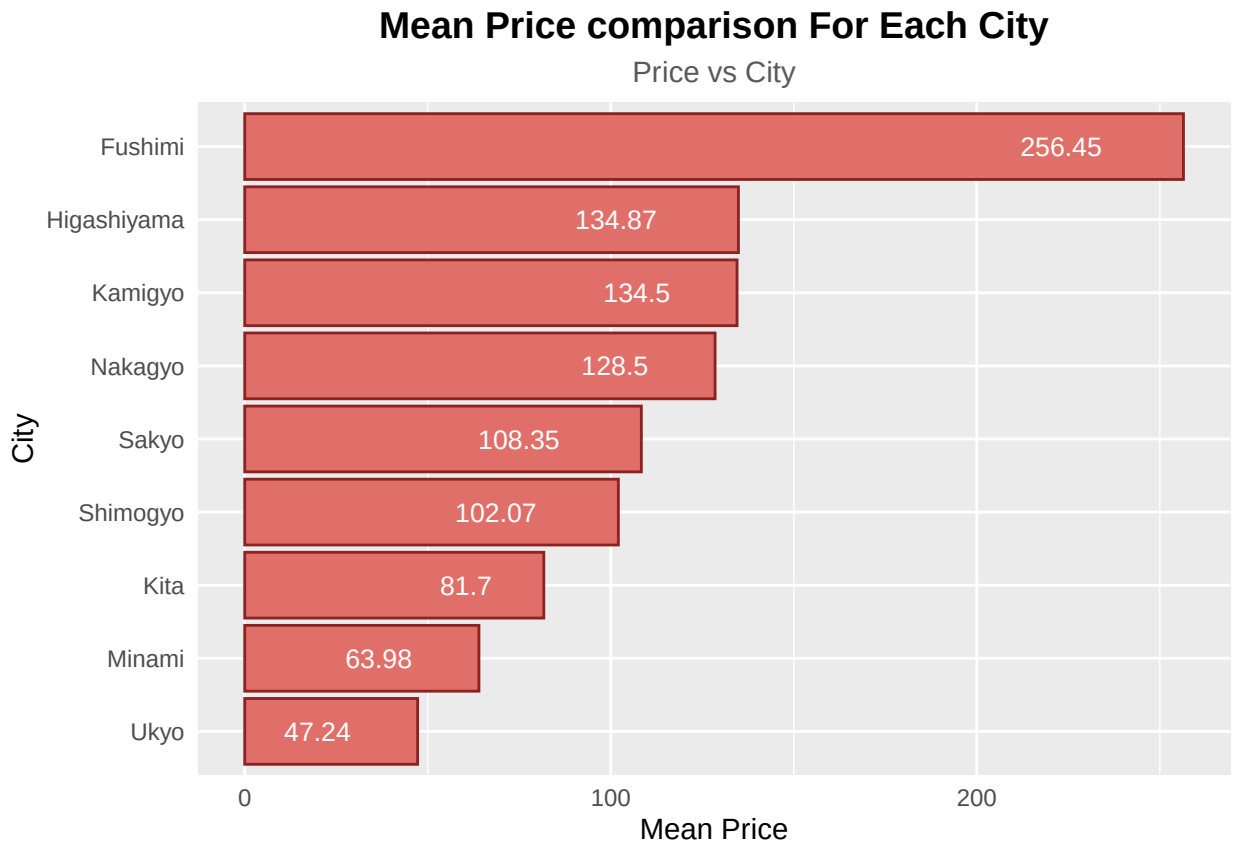
1. Entire homes is the most common listing type in all cities except Ukyo, where Entire homes and Condo/Apt are equally the most common (one of each).
2. There are about equal numbers of Condo/Apt and Private Rooms.

Mean Price Comparison For Each City Group

Obtain the average prices of listings in every city.

```
airbnb %>%
  filter(!(is.na(city))) %>%
  filter(!(city == "Unknown")) %>%
  group_by(city) %>%
  summarize(mean_price = mean(per_night, na.rm = TRUE)) %>%
  ggplot(aes(x = reorder(city, mean_price), y = mean_price, fill = city)) +
  geom_col(stat = "identity", color = "brown4", fill = "#e06f69") +
  coord_flip() +
  theme_gray() +
  labs(x = "City", y = "Price") +
  geom_text(aes(label = round(mean_price, digit = 2)),
            hjust = 2.0, color = "white", size = 3.5) +
  ggtitle("Mean Price comparison For Each City", subtitle = "Price vs City") +
```

```
xlab("City") +
ylab("Mean Price") +
theme(legend.position = "none",
      plot.title = element_text(color = "black",
                                size = 14, face = "bold", hjust = 0.5),
      plot.subtitle = element_text(color = "grey35", hjust = 0.5),
      axis.title.y = element_text(),
      axis.title.x = element_text(),
      axis.ticks = element_blank())
```



Observations

1. Average price of listings is the highest for Fushimi (256.45 USD) followed by Higashiyama (134.87 USD), which is only 0.37 USD more expensive than Kamigyo (134.50 USD).
2. There was only 1 listing in Fushimi, and that should be taken under consideration. If this one listing for Fushimi were to be neglected as an outlier, Higashiyama and Kamigyo would be the top 2 in average prices.
3. Ukyo has the cheapest listings with an average price of 87.5 USD.

Mean Price Comparison For Each Room Type

Obtain the average prices of listings by accommodation type.

```r
airbnb %>%
  filter(!(is.na(property))) %>%
  filter(!(property == "Unknown")) %>%
  group_by(property) %>%
  summarise(mean_price = mean(per_night, na.rm = TRUE)) %>%
  ggplot(aes(x = reorder(property, mean_price), y = mean_price, fill = property)) +
  geom_col(stat = "identity", color = "brown4", fill = "#e06f69") +
  coord_flip() +
  theme_gray() +
  labs(x = "Accomodation Type", y = "Price") +
  geom_text(aes(label = round(mean_price, digit = 2)),
            hjust = 2.0, color = "white", size = 3.5) +
  ggtitle("Mean Price comparison For Each Accomodation Type",
          subtitle = "Price vs Accomodation Type") +
  xlab("Accomodation Type") +
  ylab("Mean Price") +
  theme(legend.position = "none",
        plot.title = element_text(color = "black",
                                  size = 14, face = "bold", hjust = 0.5),
        plot.subtitle = element_text(color = "grey35", hjust = 0.5),
        axis.title.y = element_text(),
        axis.title.x = element_text(),
        axis.ticks = element_blank())
```

Observation

1. Average price is highest for Entire Homes, followed by Condo/Apt, which is expected since entire homes tend to have larger rooms and multiple stories.

Modelling

Data Splitting