

## Team 6: Airbnb (Santa Clara County listings)

GitHub link: <https://github.com/Apurvamuchandi8/Data-Visualization>

### Dashboard 1: Neighborhood Insights

#### 1. Average Occupancy by Room Type and Neighborhood

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Group data to calculate average availability by neighborhood and room type
occupancy_data = dataset.groupby(['neighbourhood',
                                   'room_type'])['availability_365'].mean().reset_index()

# Pivot data for a stacked bar chart
occupancy_pivot = occupancy_data.pivot(index='neighbourhood',
                                         columns='room_type', values='availability_365')

# Calculate total average availability per neighborhood
occupancy_pivot['Total'] = occupancy_pivot.sum(axis=1)

# Plot bar chart
fig, ax1 = plt.subplots(figsize=(14, 8))

# Bar chart for average availability by room type
occupancy_pivot.iloc[:, :-1].plot(kind='bar', color='skyblue', stacked=True,
                                   ax=ax1, cmap='viridis', alpha=0.8)

# Line chart for total average availability
ax2 = ax1.twinx()
ax2.plot(occupancy_pivot.index, occupancy_pivot['Total'], color='skyblue',
        marker='o', linewidth=2, label='Total Average Availability')

# Add titles and labels
ax1.set_title('Average Occupancy by Room Type and Neighborhood', fontsize=16)
ax1.set_xlabel('Neighborhood', fontsize=12)
ax1.set_ylabel('Average Availability (Days)', fontsize=12)
ax2.set_ylabel('Total Average Availability (Days)', fontsize=12)

# Add legends
ax1.legend(title='Room Type', bbox_to_anchor=(1.05, 1), loc='upper left')
ax2.legend(loc='upper right')
```

```
# Rotate x-axis labels for readability
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```

## 2. Average Price and Availability by Neighborhood and Room Type

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Calculate the average price and average availability for each neighborhood and room type
heatmap_data_price = dataset.pivot_table(values='price', index='neighbourhood',
columns='room_type', aggfunc='mean')
heatmap_data_availability = dataset.pivot_table(values='availability_365',
index='neighbourhood', columns='room_type', aggfunc='mean')

# Create a combined heatmap by annotating price and availability
combined_data = heatmap_data_price.copy()
for row in heatmap_data_price.index:
    for col in heatmap_data_price.columns:
        price = heatmap_data_price.loc[row, col]
        availability = heatmap_data_availability.loc[row, col]
        combined_data.loc[row, col] = f"${price:.1f}\n({availability:.0f} days)"

# Plot the heatmap with combined annotations
plt.figure(figsize=(12, 8))
sns.heatmap(heatmap_data_price, annot=combined_data, fmt='', cmap="YlGnBu",
linewidths=0.5, annot_kws={'size': 12, 'color': 'black'})

# Add titles and labels
plt.title('Average Price and Availability by Neighborhood and Room Type',
color='white')
plt.xlabel('Room Type', color='white')
plt.ylabel('Neighborhood', color='white')
plt.tight_layout()

# Show the plot
plt.show()
```

## 3. Prices in each Neighborhood

Created in Power BI itself using the Map visualization.

## Dashboard 2: Host Analysis

### 1. Revenue by Host

```
import matplotlib.pyplot as plt
import seaborn as sns

# Calculate revenue per listing
dataset['revenue'] = dataset['price'] * dataset['availability_365']

# Group by host_name and sum the revenue
host_revenue =
dataset.groupby('host_name')['revenue'].sum().sort_values(ascending=False)

# Create a bar plot for all hosts' revenue
plt.figure(figsize=(12, 6))
sns.barplot(x=host_revenue.index, y=host_revenue.values, palette='viridis')
plt.title('Revenue by Host')
plt.xlabel('Host Name')
plt.ylabel('Total Revenue (USD)')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```

### 2. Distribution of Reviews per Month

```
import matplotlib.pyplot as plt

# Histogram of reviews per month
plt.figure(figsize=(10, 6))
plt.hist(dataset['reviews_per_month'].dropna(), bins=30, edgecolor='black',
color='lightblue')
plt.title('Distribution of Reviews per Month')
plt.xlabel('Reviews per Month')
plt.ylabel('Frequency')
plt.tight_layout()
plt.show()
```

### 3. Price Distribution by Host

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Box plot of prices by host
plt.figure(figsize=(12, 6))
sns.boxplot(x='host_name', y='price', data=dataset)
plt.xticks(rotation=90)
plt.title('Price Distribution by Host')
plt.tight_layout()
plt.show()
```

#### 4. Price Distribution: Superhost vs Non-Superhost

```
import matplotlib.pyplot as plt
import seaborn as sns

# Create a 'superhost' column based on reviews per month
dataset['superhost'] = dataset['reviews_per_month'].apply(lambda x: 'Superhost'
if x > 2 else 'Non-Superhost')

# Check the distribution of superhost status
print(dataset['superhost'].value_counts())

# Violin plot comparing price distribution by superhost status
plt.figure(figsize=(10, 6))
sns.violinplot(x='superhost', y='price', data=dataset, palette='muted')
plt.title('Price Distribution: Superhost vs Non-Superhost')
plt.show()
```

### Dashboard 3: Market Analysis

#### 1. Relationship: Key numerical features

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.pairplot(dataset[['price', 'number_of_reviews', 'availability_365',
'reviews_per_month']], palette='coolwarm')
plt.show()
```

#### 2. Treemap of Average Price by Month

```
import pandas as pd
```

```

import seaborn as sns
import matplotlib.pyplot as plt

# Convert 'last_review' to datetime
dataset['last_review'] = pd.to_datetime(dataset['last_review'], errors='coerce')

# Extract month name and calculate average price
dataset['month'] = dataset['last_review'].dt.month_name() # Extract month name
monthly_avg_price = dataset.groupby('month')['price'].mean().reset_index()

# Sort the months for proper display
monthly_avg_price['month'] = pd.Categorical(
    monthly_avg_price['month'],
    categories=["January", "February", "March", "April", "May", "June", "July",
"August", "September", "October", "November", "December"],
    ordered=True
)
monthly_avg_price = monthly_avg_price.sort_values('month')

# Create a pivot table for the heatmap
heatmap_data = monthly_avg_price.pivot_table(values='price', index='month')

# Plot heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(heatmap_data, annot=True, fmt=".2f", cmap="YlGnBu", cbar=True)

# Add titles
plt.title('Average Price by Month (Heatmap)')
plt.xlabel('')
plt.ylabel('Month')

# Show the plot
plt.tight_layout()
plt.show()

```

### 3. Price Distribution of Listings

```

import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(10, 6))
plt.hist(dataset['price'], bins=50, edgecolor='black', color='salmon')
plt.title('Price Distribution of Listings')
plt.xlabel('Price')

```

```
plt.ylabel('Frequency')  
plt.tight_layout()  
plt.show()
```