**Logo, company name

Description automatically generated**

**Project 2: Commercial Analysis**

Prepared for: Inside Airbnb

**TABLE OF CONTENTS**

1. Executive Summary 3
2. Problem Statement 3
3. Methodology4
4. Results & Findings 14
5. Conclusions & Recommendations 21
6. References 22
7. Appendix 23

**1. EXECUTIVE SUMMARY**

Airbnb hosts are potentially treating rentals as businesses by using their property portfolios as permanent holiday rentals. The project focuses on creating a hands-on visualization tool for a broad interested audience that highlights some of the key commercial aspects and statistics pertaining to the current listings of Airbnb across the United States of America & Italy. Through the tool, the objective is to show the nature of Airbnb with metrics that differentiate "home sharing" vs more commercial use, e.g., entire home rentals, property portfolios etc.

This tool highlights 5 dashboards each giving an in-depth view of the different metrics that have been implemented across different levels of geographical granularity. 1) The Executive dashboard gives a high-level overview of the various metrics for any user to understand the commercial nature of Airbnb from scratch; 2) The Property Portfolio dashboard gives insights into the listing activities aggregated at the host level; 3) The Listings Dashboard gives insights into the Commercial Nature by providing general statistics around count, listing properties, room types and the revenue generated; 4) Commercial Segmentation dashboard helps to understand the Frequency of renting, the key drivers of revenue generated and indicators of commercial properties; 5) The licensing dashboard provides intelligent insights into the Short-term rental Licensing laws at city-level and the current licensing distribution across room types.

The delivered tool is built keeping in mind future scalability i.e., incorporation of additional countries. Dashboards capture all the key commercial metrics and other indicators that illustrate the potential commercial usage of Airbnb listings.

**2. PROBLEM STATEMENT**

**The Situation** – High number of commercial properties in a region pushes the rent prices up for the locals. When a host converts long terms rentals to short term rentals its impact comes on the residents and families. This creates an influence on the local housing markets effecting an increase in prices and reduces supply of housing. This effect is like gentrification where it gradually increases the value of an area which is detrimental to the residents thereby pushing them out due to financial constraints.

**The Problem** - Drilling down to the problem, its primary cause can be identified by the type of listing. Listings can be either of Commercial use or Home Sharing motivated. Many of the listings being shown as home shares can be converted to hotel suites as earnings from them can be far more than then the revenue generated from home Sharing.

**3. METHODOLOGY**

This section contains the detailed steps taken to create the “Airbnb – Commercial Analysis” visualization tool. Sub-sections elaborate on the plan, approach, and each subsequent step taken for data preparation and the individual dashboard design and function.

**3.1 THE PLAN**

The plan is to develop a visualization tool on Tableau to create self-explanatory visuals that users can interact with using basic dropdowns, filters, and searches. The dashboards are aimed to provide actionable insights into any potential commercial utilization of Airbnb listings by the hosts.

The dashboards for the initial tool are based on Inside Airbnb's USA & Italy datasets, which can then be expanded for other countries to ensure scalability. The dashboards created include the following:

1. **Interactive Geographical maps** showing listings categorized by **different room types** and **commercial metrics** across the geographical hierarchies: (Region/State, Province/County, City, Zip Code)
2. **Host-level data** – Top hosts, listings per host ratio – for differentiating between commercial and home-sharing properties
3. **Property Statistics** **and Revenue Analysis** – Across Commercial Segmentations
4. **Property Licensing** – Algorithm-based deduction of licensing requirements

The final tool lets the user/audience:

1. **Scan through different geographical granularities** like country, region/state, province/county, city, and zip code and see the region-wise number of listings, revenue generated, and host portfolios
2. Analyze **rental frequency information** and get a better context of the type of listing (whether commercial or semi-commercial or home-sharing)
3. Focus on the **behavior of the short-term rentals** based on the commercial metrics
4. **Understand the licensing laws** enforced across different geographic regions
5. Compare and analyze the metrics for a region with the **state and national average**

**3.2 APPROACH**

The approach has been categorized into four key sections:

* 3.2.1 Data Preparation
* 3.2.2 Algorithm Development
* 3.2.3 Incorporating More Countries into the Dashboard (Ensuring Scalability)
* 3.2.4 Tableau Implementation

3.2.1 DATA PREPARATION

Exploring and compiling the Datasets

1. The first step in developing the tool is to collate the relevant datasets provided to include all the crucial commercial metrics. These commercial metrics help distinguish between commercial and home-sharing listings.
2. Inside Airbnb's USA dataset and U.S. Census Data have been used as preliminary datasets. A custom algorithm has been chosen to build the *licensing dataset* since the laws enforced for short-term rentals are implemented at a city level and based on research, such a dataset is not yet available in public records.

Missing Value Imputation

1. There were 5.6% and 1.4% of listings with missing values for ‘Beds’ and ‘Bedrooms’.
2. These columns are crucial for the calculation of “Commercial Segments” [Covered later in the document]. Hence, they have been imputed with Median Values for those respective columns.

Feature Engineering: Deriving Additional Metrics

1. ***Short-term vs Long-term Rental***: A new column ‘STR\_vs\_LTR’ has been added where - If the minimum number of nights of stay is < 30 days, the listing is classified as a *Short-term rental*, else a *Long-term rental.*
2. ***Number of Bookings (Last 12 months)*:** A new column ‘*estimated\_booking\_ltm*’ derived from ‘*number\_of\_reviews\_ltm’* has been added. **Formula used:**

Review Rate = 0.5

Estimated Bookings = Number of reviews / Review Rate



1. ***Number of Nights Booked (Last 12 months)*:** A new column estimated\_nights\_ltm derived from estimated\_booking\_ltm and minimum nights of stay has been added.

**Formula used:**

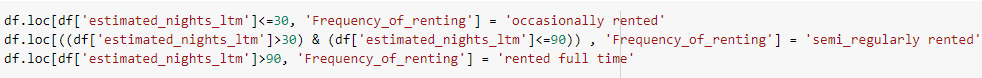
Maximum Occupancy = 255, Average Night Stay =3

…

1. ***Revenue Generated (Last 12 months)*:** A new column estimated\_revenue\_ltm derived from estimated\_booking\_ltm and price has been added. Formula used:



1. ***Frequency of Renting*:** A new column Frequency\_of\_renting derived from estimated\_nights\_ltm has been added. Listings have been split across “Occasionally rented”, “Semi-Regularly Rented” and “Rented Full-time”. Formula used:



1. ***Commercial Nature of Listing*:** This column has been derived from a combination of variables such as 'room\_type', 'calculated\_host\_listings\_count', 'beds' and 'Frequency\_of\_renting' and represents the commercial nature of the listing. **Formula used:**

**Home-Sharing:**

df.loc[(df['room\_type']=='Shared room') & (df['calculated\_host\_listings\_count\_shared\_rooms']==1),'Commercial\_Nature']='Home Sharing'

df.loc[(df['room\_type']=='Shared room') & (df['calculated\_host\_listings\_count\_shared\_rooms']>1) & (df['beds']<=3),'Commercial\_Nature']='Home Sharing'

df.loc[(df['room\_type']=='Private room') & (df['calculated\_host\_listings\_count\_private\_rooms']<=3),'Commercial\_Nature']='Home Sharing'

df.loc[(df['room\_type']=='Entire home/apt') & (df['calculated\_host\_listings\_count\_entire\_homes']==1) & (df['Frequency\_of\_renting']=='occasionally rented'),'Commercial\_Nature']= 'Home Sharing'

**Semi-Commercial:**

df.loc[(df['room\_type']=='Shared room') & (df['calculated\_host\_listings\_count\_shared\_rooms']>1) & (df['beds']>3),'Commercial\_Nature']='Semi-Commercial'

df.loc[(df['room\_type']=='Private room') & (df['calculated\_host\_listings\_count\_private\_rooms']>3) & (df['calculated\_host\_listings\_count\_private\_rooms']<=6),'Commercial\_Nature']='Semi-Commercial'

df.loc[(df['room\_type']=='Entire home/apt') & (df['calculated\_host\_listings\_count\_entire\_homes']==1) & (df['Frequency\_of\_renting']=='semi\_regularly rented'),'Commercial\_Nature']= 'Semi-Commercial'

**Commercial:**

df.loc[(df['room\_type']=='Hotel room'),'Commercial\_Nature'] ='Commercial'

df.loc[(df['room\_type']=='Entire home/apt') & (df['calculated\_host\_listings\_count\_entire\_homes']>1),'Commercial\_Nature'] ='Commercial'

df.loc[(df['room\_type']=='Entire home/apt') & (df['calculated\_host\_listings\_count\_entire\_homes']==1) & (df['Frequency\_of\_renting']=='rented full time'),'Commercial\_Nature']= 'Commercial'

df.loc[(df['room\_type']=='Private room') & (df['calculated\_host\_listings\_count\_private\_rooms']>6),'Commercial\_Nature']='Commercial'

1. ***Zip Code*:** Derived using the ‘geopandas’ and ‘*USA ZCTA*’ dataset using latitude and longitude as common points.
2. ***City*:** Derived using ‘georef-united-states-of-america-zc-point.csv’ data with zip code as the common point. The required variables have been extracted from ‘georef-united-states-of-america-zc-point.csv’ file and saved to ‘city.csv’ file (which has been used in the pre-processing file).

3.2.3 ALGORITHM DEVELOPMENT

Licensing is a key commercial metric but has partial data in the form of license numbers in the main dataset. Owing to the unavailability of a collated licensing law document/dataset for the major cities in the US, an algorithm has been developed for the tool to directionally indicate the presence of an STR law for Airbnb listings in a region.

1. The STR laws are known to be implemented at the “City” level. Hence, the algorithm checks the following at any CITY level:

IF COUNT (UNIQUE (License Nos.)) > **5%** of COUNT (UNIQUE (Listings IDs)

“License Required”

Else

“License Not Required”

END

The following Python code has been used to derive the variable that shows whether there is licensing requirement for a city.

Text

Description automatically generated

1. Through this new column “requires\_license\_5”, a new variable ‘licensed\_5’ has been derived that categorizes the listings as:
   1. Licensed – License No. available
   2. Not Licensed – License No. not available in a city with STR Licensing law
   3. Exempt – License No. not available in a city without STR Licensing law

Chart

Description automatically generated

***Figure*** *3.1: Licensing: District of Columbia*

The following Python code has been used to derive the aforementioned variable.



1. The 5% threshold can be updated based on the observed results i.e., if the directional indicators have good accuracy for determining the licensing laws for STRs.

3.2.3 INCORPORATING MORE COUNTRIES INTO THE DASHBOARD (SCALABILITY)

**Diagram, timeline

Description automatically generated**

* Under step 2 of the Approach, Zip Code level data has been included in the geographical hierarchy of USA but not for Italy. In this step, ZCTA file with the Zip Code shapefile has been used to generate the Zip Codes for the listings.
* Since the Inside Airbnb dataset for USA had listings with missing values for ‘City’, the Zip Codes have been used to generate the ‘city’ data for the listings in the USA using the ‘city.csv’ file. Since there were no missing values for City for Italy data provided by Inside Airbnb, no additional city-level data has been generated. Moreover, Zip Code level data has not been incorporated into Italy dataset for the analysis.
* To run the pre-processing file for a country with geo. hierarchy similar to that of the USA, external datasets with Zip Codes need to be imported to extract Zip Codes from the Latitude and Longitude of the listings*.* If the dataset for the chosen country has missing values for City data, external datasets which include City & Zip Code data need to be imported to derive City-level data from the Zip Codes generated.
* To run the pre-processing file for a country with geo. hierarchy similar to that of Italy, no external datasets have to be imported.
* Once the pre-processing is done for the chosen set of countries & the cleaned datasets are available, the cleaned CSV files are merged using ‘Team5\_Merged.ipynb’ file. The cleaned datasets must be imported separately as different data frames to perform the merge operation.
* The final merged CSV file is then uploaded to Tableau. To replace an existing data source with a new file in Tableau, the following steps must be followed.

Graphical user interface, application

Description automatically generated

* In addition to the 5% threshold, the algorithm has been replicated for a 10% threshold as well. This can be easily implemented in the Tableau file.

*Map

Description automatically generated*

3.2.4 TABLEAU IMPLEMENTATION

Geographical Hierarchy

The COMMERCIAL ANALYSIS dashboard has been designed to be highly intuitive and user-friendly. The idea is to minimize the effort a user would need to make to visualize the listings on the map along with the relevant commercial metrics.

The most crucial step is to create a **geographical hierarchy** that could automatically ZOOM-IN to the user selection without the requirement of manual pan and zoom. For the US dataset, this is a two-step process –

1. *Renaming COLUMNS*: The Inside Airbnb dataset had the three higher levels of geographical hierarchy as Region Parent Parent Name, Region Parent Name and Region Name. For the US, Region Parent Name represents ‘*Region/State’* and Region Name represents ‘*Province/County’* respectively. For Italy, Region Parent Parent Name represents ‘*Region/State’,* Region Parent Name represents ‘*Province/County’* and Region Name represents *‘City’* respectively. These were renamed within Tableau.
2. *Assigning the right geographical roles*: For the US, the correct geographical roles were assigned to newly created columns “City” and “Zip Codes”.

Graphical user interface, application

Description automatically generated

***Figure*** *3.2: Assigning Geographical Roles*

1. *Creating the geographical hierarchy*: A hierarchy (“Geo Hierarchy”) has been created and all four levels (Region / State, Province / County, City, Zip Codes) have been assigned to it. This helps Tableau understand the “FILTER” dynamics and automatically captures the ZOOM functionalities upon selection on a map.

Graphical user interface, text, application, chat or text message

Description automatically generatedText

Description automatically generated with medium confidence

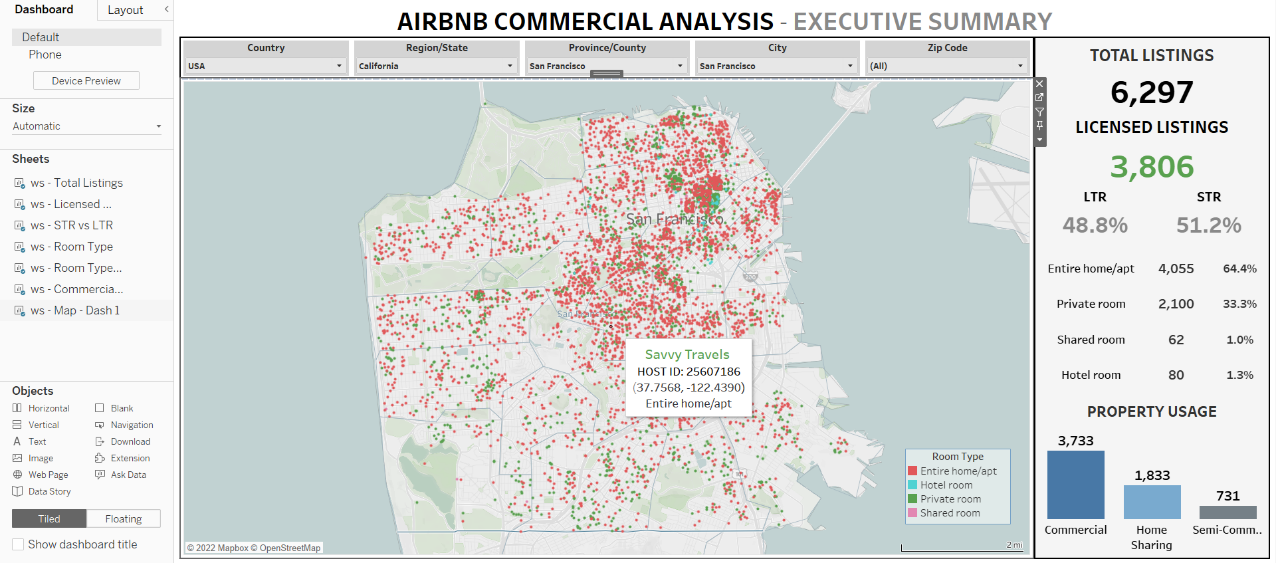
***Figure*** *3.3: Creating Geographical hierarchy*

Tableau Design

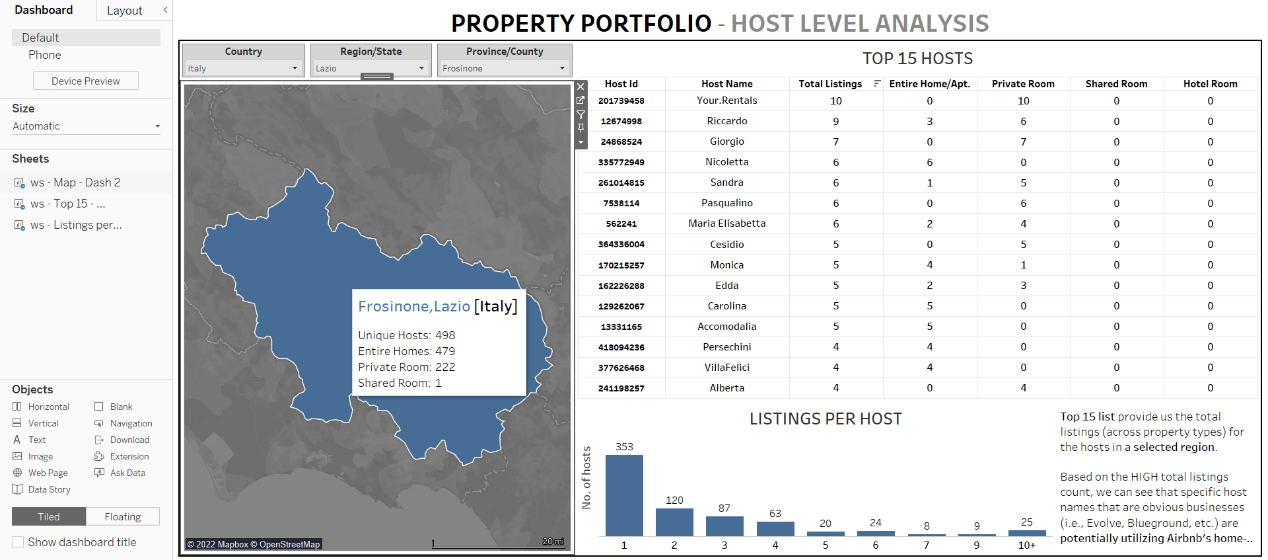
Data preparation has been completed through a Python code to avoid tableau data modifications

1. Each dashboard view has **multiple worksheets** **[labeled: “ws –”]** which can be traced on the **left-hand side through the sheets view**.
2. Dashboard 2 contains **“Calculated Fields [labeled: “zz –”]** to inform host level metrics

**Dashboard 1: Executive Summary**



**Dashboard 2: Property Portfolio**



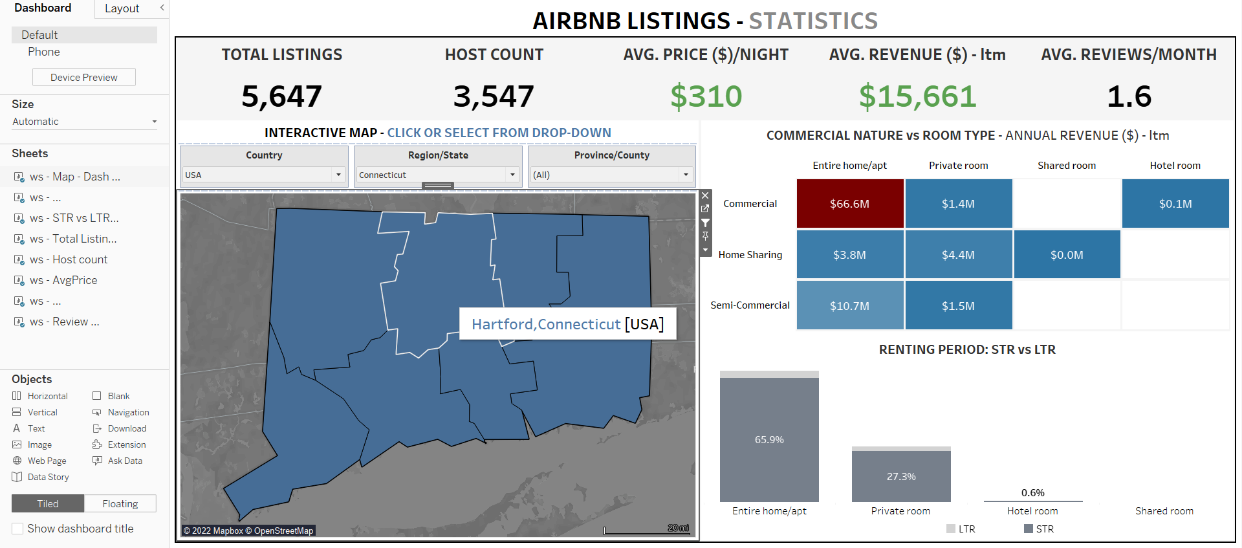
*Calculated Fields:*

Text

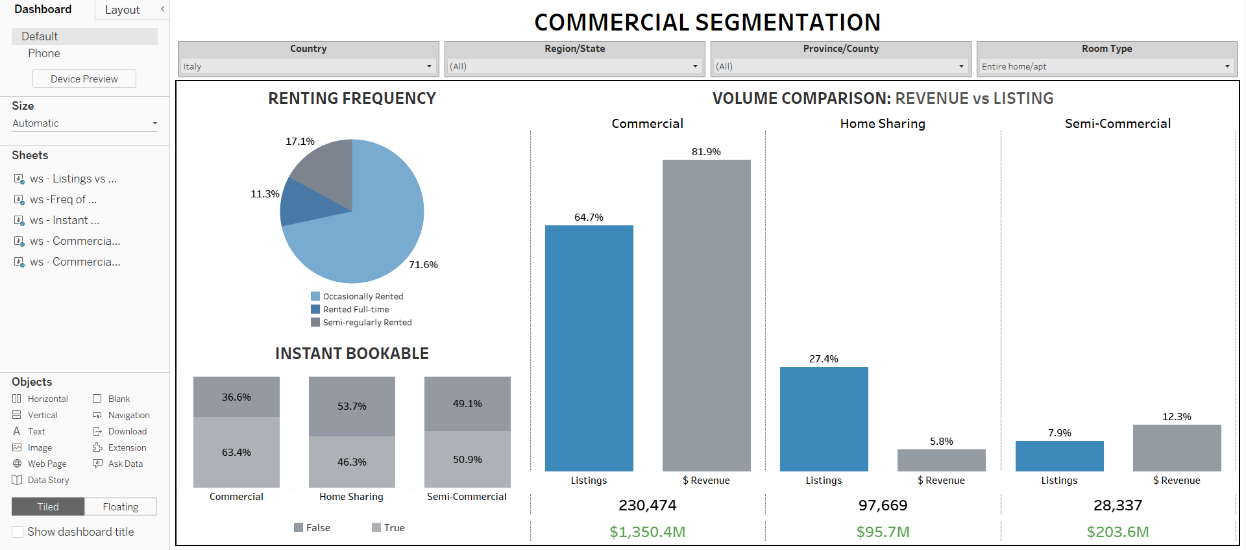
Description automatically generated A picture containing text

Description automatically generated

**Dashboard 3: Listing - Statistics**



**Dashboard 4: Commercial Segmentation**

****

**Dashboard 5: Licensing Analysis**

**Graphical user interface, application

Description automatically generated**

**4. RESULTS AND FINDINGS**

Results cater towards generating useful insights for any useful for understanding the commercialization of Airbnb Business Model. Details have been provided elaborating on the potential insights across each dashboard:

**Dashboard 1: Executive Summary**

Potential Insights

1. Key Commercial Metrics at different granularities for QUICK COMPARISON
   * USA: State – County – City – Zip Code
   * ITALY: Region – Province - City

**California, USA** **ITALY**

Table

Description automatically generatedTable

Description automatically generated

1. Visual representation of listings on the map – color segregation based on **ROOM TYPE**

Chart, map

Description automatically generated

**Dashboard 2: Host-Level Analysis**

Potential Insights

1. Top 15 hosts at selected granularity: Country – Region/State – Province/County
   * Helps to shortlist the top hosts that are listing properties across each room type

Table

Description automatically generated with medium confidence

1. Listings per host – Distribution of no. of hosts based on number of properties listed

**Chart

Description automatically generated**

1. Interactive Province/County level map – With High-level tool-tip data for quick reference

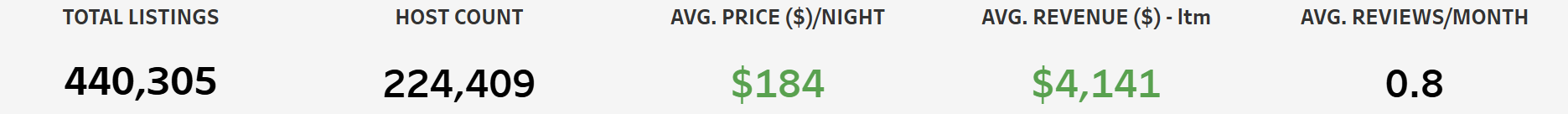
A picture containing diagram

Description automatically generated

**Dashboard 3: Listings Level Analysis – Statistics**

Potential Insights

1. High-level quick summary statistics based on geographical selection
   * Country – Region/State – Province/County



1. Interactive Geographical Map – Click to update all visualizations on the dashboard

Map

Description automatically generated

1. Revenue across commercial segments and ROOM types

Table, timeline

Description automatically generated with medium confidence

1. Renting period – Distribution of STR and LTR property across ROOM types

Graphical user interface, application

Description automatically generated

**Dashboard 4: Commercial Segmentation**

Potential Insights

1. Total Revenue and Listings count across the commercial segmentations and ROOM type
   * Country – Region/State – Province/County

Chart, bar chart

Description automatically generated

1. Renting Frequency – Split for listings based on how often they’re rented
   * High number of “Rented Full-time” listings could indicate high commercial usage

Chart, pie chart

Description automatically generated

1. Instant Bookable – Efficient booking method for short term rentals
   * High distribution of instantly bookable listings could indicate commercial usage

Chart, waterfall chart

Description automatically generated

**Dashboard 5: Licensing Analysis**

Potential Insights

1. Algorithm based STR License Enforcement predictor (CITY) based on 5% licensing cap

Map

Description automatically generated

1. Distribution and count of licensed properties across ROOM types – Since some CITIES enforce licensing laws based on the property type

**A picture containing text, electronics, iPod, compact disk

Description automatically generated**

**5. CONCLUSIONS AND RECOMMENDATIONS**

* The tool is a comprehensive and user-friendly dashboard sequence that could assist any end user gain actionable insights on the commercial aspects of Airbnb
* The highly interactive dashboards give users the flexibility for observing data at various granularities
* All the key commercial metrics provided by INSIDE AIRBNB were successfully incorporated into the dashboard
* The delivered tool is built keeping in mind future scalability i.e., incorporation of additional countries.
* Through the detailed schematic and steps, any prospective user can use the existing pre-processing and tableau files for pulling new country data.
* Dashboards capture all the key commercial metrics and other indicators that illustrate the potential commercial usage of Airbnb listings.
* Any new data source/column can be pulled into the tool as new worksheets or dashboards -New calculated fields which may provide clarity on the subject can also be created.

**6. REFERENCES**

* *Inside Airbnb Data Sources for the USA and Italy*
  + [*http://data.insideairbnb.com/uconn/msbapm/capstone/2022/fall/listings-united-states-2022-07.csv.gz*](http://data.insideairbnb.com/uconn/msbapm/capstone/2022/fall/listings-united-states-2022-07.csv.gz)*(USA)*
  + [*http://data.insideairbnb.com/uconn/msbapm/capstone/2022/fall/listings-italy-2022-08.csv.gz*](http://data.insideairbnb.com/uconn/msbapm/capstone/2022/fall/listings-italy-2022-08.csv.gz) *(Italy)*
* *External Data Sources for deriving zip codes from Latitude and Longitude for USA*
  + [*https://drive.google.com/file/d/16iv4VggXpiqM-nu\_RtkOjRtGblRN\_lw8/view?usp=share\_link*](https://drive.google.com/file/d/16iv4VggXpiqM-nu_RtkOjRtGblRN_lw8/view?usp=share_link) *(tl\_2017\_us\_zcta510.zip)*
* *External CSV file used to generate city-level data from the Zip Codes for the USA data. This file has 2 columns Zip Code and City derived from ‘georef-united-states-of-america-zc-point.csv’.*
  + [*https://drive.google.com/file/d/10V47ehEtDUiEm\_Uf1RD57AqzVeBfWEv/view?usp=share\_link*](https://drive.google.com/file/d/10V47ehEtDUiEm_Uf1RD57AqzVeBfWEv/view?usp=share_link) *(city.csv)*
* *Consolidated file based on the research on state-level licensing requirement*
  + [*https://docs.google.com/spreadsheets/d/1dc\_o2hPQfAY62jRNTf0rsfwtfWU1F9ft/edit?usp=share\_link&ouid=111884287115992792490&rtpof=true&sd=true*](https://docs.google.com/spreadsheets/d/1dc_o2hPQfAY62jRNTf0rsfwtfWU1F9ft/edit?usp=share_link&ouid=111884287115992792490&rtpof=true&sd=true) *(Licensing\_Final.xlsx)*

**7. APPENDIX**

