Project Airbnb Amsterdam

The purpose of this project is to identify trends within the Airbnb data and make marketing strategy recommendations. There are 6 databases provided by Kaggle that were taken from insideairbnb.com. To see the data cleaning process, see attached excel reporting.

Data Sourcing and Profiling

**Calendar Database**

|  |  |
| --- | --- |
| Data Sourcing | This data comes from within the Airbnb organization (Insideairbnb.com). Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. |
| Data Collection Methods | Administrative (internal) |
| Content | 365 records for each listing. It specifies the whether the listing is available on a particular day (365 days ahead), and the price on that day. |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Qualitative/Quantitative | Invariant/Variant | Nominal/Ordinal/Discrete/Continuous |
| Listing\_id | Qualitative | Invariant | Nominal |
| Date | Qualitative | Variant | Ordinal |
| Available | Qualitative | Invariant | Nominal |
| Price | Quantitative | Invariant | Continuous |

**Listings Database**

|  |  |
| --- | --- |
| Data Sourcing | This data comes from within the Airbnb organization (Insideairbnb.com). Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. |
| Data Collection Methods | Administrative (internal) |
| Content | Basic information regarding places that rent their property |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Qualitative/Quantitative | Invariant/Variant | Nominal/Ordinal/Discrete/Continuous |
| Id | Qualitative | Invariant | Nominal |
| Name | Qualitative | Invariant | Nominal |
| Host\_id | Qualitative | Invariant | Nominal |
| Host\_name | Qualitative | Invariant | Nominal |
| Neighbourhood\_group | Qualitative | Invariant | Nominal |
| Neighbourhood | Qualitative | Invariant | Nominal |
| Latitude | Qualitative | Invariant | Discrete |
| Longitude | Qualitative | Invariant | Discrete |
| Price | Quantitative | Invariant | Continuous |
| Minimum\_night | Quantitative | Invariant | Continuous |
| Number\_of\_reviews | Quantitative | Invariant | Discrete |
| Last\_review | Qualitative | Invariant | Discrete |
| Reviews\_per\_month | Quantitative | Invariant | Continuous |
| Calculated\_host\_listings\_count | Quantitative | Invariant | Continuous |
| Availability\_365 | Quantitative | Invariant | Continuous |

**Listings Details Database**

|  |  |
| --- | --- |
| Data Sourcing | This data comes from within the Airbnb organization (Insideairbnb.com). Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. |
| Data Collection Methods | Administrative (internal) |
| Content | Extensive information (91 variables) regarding places that rent their property |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Qualitative/Quantitative | Invariant/Variant | Nominal/Ordinal/Discrete/Continuous |
| Id | Qualitative | Invariant | Nominal |
| Listing\_url | Qualitative | Invariant | Nominal |
| Scrape\_id | Qualitative | Invariant | Nominal |
| Last\_scraped | Qualitative | Invariant | Nominal |
| Name | Qualitative | Invariant | Nominal |
| Summary | Qualitative | Invariant | Nominal |
| Space | Qualitative | Invariant | Nominal |
| Description | Qualitative | Invariant | Nominal |
| Experiences\_offered | Qualitative | Invariant | Nominal |
| Neighborhood\_overview | Qualitative | Invariant | Nominal |
| Notes | Qualitative | Invariant | Nominal |
| Transit | Qualitative | Invariant | Nominal |
| Access | Qualitative | Invariant | Nominal |
| Interaction | Qualitative | Invariant | Nominal |
| House\_rules | Qualitative | Invariant | Nominal |
| Picture\_url | Qualitative | Invariant | Nominal |
| Host\_id | Qualitative | Invariant | Nominal |
| Host\_url | Qualitative | Invariant | Nominal |
| Host\_name | Qualitative | Invariant | Nominal |
| Host\_since | Qualitative | Invariant | Nominal |
| Host\_location | Qualitative | Invariant | Nominal |
| Host\_about | Qualitative | Invariant | Nominal |
| Host\_response\_time | Quantitative | Invariant | Nominal |
| Host\_response\_rate | Quantitative | Invariant | Nominal |
| Host\_is\_superhost | Qualitative | Invariant | Nominal |
| Host\_thumbnail\_url | Qualitative | Invariant | Nominal |
| Host\_picture\_url | Qualitative | Invariant | Nominal |
| Host\_listings\_count | Quantitative | Invariant | Nominal |
| Host\_total\_listings\_count | Quantitative | Invariant | Nominal |
| Host\_verifications | Quantitative | Invariant | Nominal |
| Host\_has\_profile\_pic | Qualitative | Invariant | Nominal |
| Host\_identity\_verified | Qualitative | Invariant | Nominal |
| Street | Qualitative | Invariant | Nominal |
| Neighbourhood | Qualitative | Invariant | Nominal |
| Neighbourhood\_cleansed | Qualitative | Invariant | Nominal |
| City | Qualitative | Invariant | Nominal |
| State | Qualitative | Invariant | Nominal |
| Zipcode | Qualitative | Invariant | Nominal |
| Market | Qualitative | Invariant | Nominal |
| Smart\_location | Qualitative | Invariant | Nominal |
| Country\_code | Qualitative | Invariant | Nominal |
| Country | Qualitative | Invariant | Nominal |
| Latitude | Qualitative | Invariant | Discrete |
| Longitude | Qualitative | Invariant | Discrete |
| Is\_location\_exact | Qualitative | Invariant | Nominal |
| Property\_type | Qualitative | Invariant | Nominal |
| Room\_type | Qualitative | Invariant | Nominal |
| Accommodates | Quantitative | Invariant | Discrete |
| Bathrooms | Quantitative | Invariant | Discrete |
| Bedrooms | Quantitative | Invariant | Discrete |
| Beds | Quantitative | Invariant | Discrete |
| Bed\_type | Qualitative | Invariant | Discrete |
| Amenities | Qualitative | Invariant | Nominal |
| Square\_feet | Quantitative | Invariant | Discrete |
| Price | Quantitative | Invariant | Discrete |
| Weekly\_price | Quantitative | Invariant | Discrete |
| Monthly\_price | Quantitative | Invariant | Discrete |
| Security\_deposit | Quantitative | Invariant | Discrete |
| Cleaning\_fee | Quantitative | Invariant | Discrete |
| Guests\_included | Quantitative | Invariant | Discrete |
| Extra\_people | Quantitative | Invariant | Discrete |
| Minimum\_nights | Quantitative | Invariant | Discrete |
| Maximum\_nights | Quantitative | Invariant | Discrete |
| Calendar\_updated | Qualitative | Invariant | Nominal |
| Has\_availability | Qualitative | Invariant | Nominal |
| Availability\_30 | Qualitative | Invariant | Discrete |
| Availability\_60 | Qualitative | Invariant | Discrete |
| Availability\_90 | Qualitative | Invariant | Discrete |
| Availability\_365 | Qualitative | Invariant | Discrete |
| Calendar\_last\_scraped | Qualitative | Invariant | Nominal |
| First\_review | Qualitative | Invariant | Nominal |
| Last\_review | Qualitative | Invariant | Nominal |
| Review\_scores\_rating | Quantitative | Invariant | Nominal |
| Review\_scores\_accuracy | Quantitative | Invariant | Nominal |
| Review\_scores\_cleanliness | Quantitative | Invariant | Nominal |
| Review\_scores\_checkin | Quantitative | Invariant | Nominal |
| Review\_scores\_communication | Quantitative | Invariant | Nominal |
| Review\_scores\_location | Quantitative | Invariant | Nominal |
| Review\_scores\_value | Quantitative | Invariant | Discrete |
| Requires\_license | Qualitative | Invariant | Nominal |
| License | Qualitative | Invariant | Nominal |
| Jurisdiction\_names | Qualitative | Invariant | Nominal |
| Instant\_bookable | Qualitative | Invariant | Nominal |
| Is\_business\_travel\_ready | Qualitative | Invariant | Nominal |
| Cancelation\_policy | Qualitative | Invariant | Nominal |
| Require\_guest\_profile\_picture | Qualitative | Invariant | Nominal |
| Require\_guest\_phone\_verification | Qualitative | Invariant | Nominal |
| Calculated\_host\_listings\_count | Quantitative | Invariant | Discrete |
| Reviews\_per\_month | Quantitative | Invariant | Discrete |

**Neighbourhood Database**

|  |  |
| --- | --- |
| Data Sourcing | This data comes from within the Airbnb organization (Insideairbnb.com). Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. |
| Data Collection Methods | Administrative (internal) |
| Content | Lists the names of the neighborhoods (only 2 columns) |

**Reviews Database**

|  |  |
| --- | --- |
| Data Sourcing | This data comes from within the Airbnb organization (Insideairbnb.com). Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. |
| Data Collection Methods | Survey (internal) |
| Content | Listing ID and date of review (only 2 columns) |

**Reviews Details Database**

|  |  |
| --- | --- |
| Data Sourcing | This data comes from within the Airbnb organization (Insideairbnb.com). Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. |
| Data Collection Methods | Survey (internal) |
| Content | Reviews from patrons of listings |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Qualitative/Quantitative | Invariant/Variant | Nominal/Ordinal/Discrete/Continuous |
| Listing\_id | Qualitative | Invariant | Nominal |
| Id | Qualitative | Invariant | Nominal |
| Date | Qualitative | Variant | Continuous |
| Reviewer\_id | Qualitative | Invariant | Nominal |
| Reviewer\_name | Qualitative | Invariant | Nominal |
| comments | Qualitative | Invariant | Nominal |

Data Bias and Limitations

On the insideairbnb.com website, they state that they clean, aggregate, and filter their data. This leaves room for bias because we do not know what they are omitting or decided not to take into consideration.

The limits in this data are lack of demographics. There is no data on gender, race, age, first time users, etc.

Exploratory Questions

* Seasonality?
* Where is the most popular area?
* Does the price of an Airbnb affect how frequently it’s rented?
* What are frequently used terms in reviews?
* Does host response rate factor into renting?
* What factors contribute to rentals?
* How do different scores affect renting?