Airbnb CompSet Viewer

Team Members: Carmen Varón & Carlos Naranjo

Project Summary V0 Updated: May 3rd 2018

**Intro and challenge definition**

We are currently on dates of historical records on tourism. At this juncture, two disruptive innovations has changed the tourism market:

* Irruption of the Internet (with a radical change in distribution models)
* Incipient explosion of the collaborative economy (room & home sharing)

This **project will focus on the collaborative economy at the Hotel Industry**. The war is served at the internet, with millions of new players acting as hotel managers with their own property. These new players are non-hoteliers, so their capabilities to price their properties are reduced.

Revenue Management

Revenue Management Systems are based on both, internal and external data, and each day these become more and more complex.



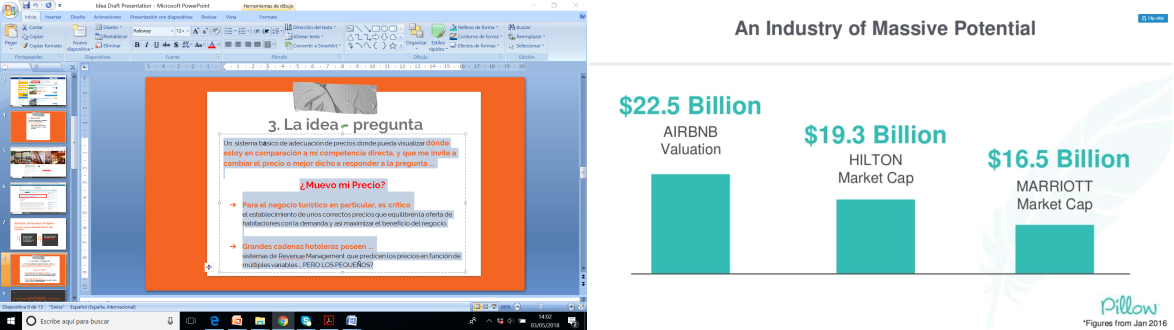
Project Goal

This Project will try to **explain the key factors for succeed on the total revenue income for rental properties, mainly at the pricing, based on a comparative analysis with competitors**.

Road Map

1. Identify what are the main factors affecting success (Attributes Analysis)
2. Establish who are my real competitors (Clustering)
3. Define the right price to maximize income (Recommended price for a new or existing property)

Scope

The scope will be the **Airbnb site** properties, considered one of the biggest players at the pitch. And more specifically, **Madrid** as the selected city for the analysis, with over than 16K listed properties at the site.

**Data first approach**

The data for this project will be obtained from the Inside Airbnb Initiative. In their words: “Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world.” (**Inside Airbnb website**).

The data that wil be used is based on a snapshot of date Compiled at 17 January, 2018, for Madrid. These are the files provided:

|  |  |
| --- | --- |
| File Name | Description |
| [listings.csv.gz](http://data.insideairbnb.com/spain/comunidad-de-madrid/madrid/2018-01-17/data/listings.csv.gz) | Detailed Listings data for Madrid |
| [calendar.csv.gz](http://data.insideairbnb.com/spain/comunidad-de-madrid/madrid/2018-01-17/data/calendar.csv.gz) | Detailed Calendar Data for listings in Madrid |
| [reviews.csv.gz](http://data.insideairbnb.com/spain/comunidad-de-madrid/madrid/2018-01-17/data/reviews.csv.gz) | Detailed Review Data for listings in Madrid |
| [listings.csv](http://data.insideairbnb.com/spain/comunidad-de-madrid/madrid/2018-01-17/visualisations/listings.csv) | Summary information and metrics for listings in Madrid (good for visualisations). |
| [reviews.csv](http://data.insideairbnb.com/spain/comunidad-de-madrid/madrid/2018-01-17/visualisations/reviews.csv) | Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing). |
| [neighbourhoods.csv](http://data.insideairbnb.com/spain/comunidad-de-madrid/madrid/2018-01-17/visualisations/neighbourhoods.csv) | Neighbourhood list for geo filter. Sourced from city or open source GIS files. |
| [neighbourhoods.geojson](http://data.insideairbnb.com/spain/comunidad-de-madrid/madrid/2018-01-17/visualisations/neighbourhoods.geojson) | GeoJSON file of neighbourhoods of the city. |

The main file is the listing.csv, which has not just the property feautures, but integrates other fields calculated form the Calendar and Reviews files.

This is the **general scheme of data reviewed** at the moment:

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Field** | **Extracted** | **Format** |
| Identifier - Lisiting | **id** | Scrapped | Number |
| Listing Description (Text) | **name** | Scrapped | Text |
| Listing Description (Text) | **summary** | Scrapped | Text |
| Listing Description (Text) | **space** | Scrapped | Text |
| Listing Description (Text) | **description** | Scrapped | Text |
| Listing Description (Text) | **neighborhood\_overview** | Scrapped | Text |
| Listing Description (Text) | **notes** | Scrapped | Text |
| Listing Description (Text) | **transit** | Scrapped | Text |
| Listing Description (Text) | **[access]** | Scrapped | Text |
| Listing Description (Text) | **interaction** | Scrapped | Text |
| Listing Description (Text) | **house\_rules** | Scrapped | Text |
| Location | **is\_location\_exact** | Calculated | Boolean |
| Location | **latitude** | Scrapped | Number |
| Location | **longitude** | Scrapped | Number |
| Location | **zipcode** | Calculated | Number |
| Location | **street** | Calculated | Text |
| Location | **neighbourhood** | Calculated | Text Category |
| Location | **neighbourhood\_cleansed** | Calculated | Text Category |
| Location | **city** | Scrapped | Text Category |
| Location | **state** | Scrapped | Text Category |
| Location | **market** | Calculated | Text Category |
| Location | **smart\_location** | Calculated | Text Category |
| Location | **country\_code** | Calculated | Text Category |
| Location | **country** | Scrapped | Text Category |
| Location | **neighbourhood\_group\_cleansed** | Calculated | Text Category |
| Performance - Future Activity | **availability\_30** | Calculated | Number |
| Performance - Future Activity | **availability\_60** | Calculated | Number |
| Performance - Future Activity | **availability\_90** | Calculated | Number |
| Performance - Future Activity | **availability\_365** | Calculated | Number |
| Performance - Past Activity | **first\_review** | Scrapped | Date |
| Performance - Past Activity | **last\_review** | Scrapped | Date |
| Performance - Past Activity | **number\_of\_reviews** | Scrapped | Number |
| Performance - Past Activity | **reviews\_per\_month** | Calculated | Number |
| Performance - Quality | **review\_scores\_accuracy** | Scrapped | Number |
| Performance - Quality | **review\_scores\_cleanliness** | Scrapped | Number |
| Performance - Quality | **review\_scores\_checkin** | Scrapped | Number |
| Performance - Quality | **review\_scores\_communication** | Scrapped | Number |
| Performance - Quality | **review\_scores\_location** | Scrapped | Number |
| Performance - Quality | **review\_scores\_value** | Scrapped | Number |
| Performance - Quality | **review\_scores\_rating** | Scrapped | Number |
| Phisic Listing Features | **amenities** | Scrapped | List of Text Categories |
| Phisic Listing Features | **square\_feet** | Scrapped | Number |
| Phisic Listing Features | **bathrooms** | Scrapped | Text Category |
| Phisic Listing Features | **bedrooms** | Scrapped | Text Category |
| Phisic Listing Features | **beds** | Scrapped | Text Category |
| Phisic Listing Features | **property\_type** | Scrapped | Text Category |
| Phisic Listing Features | **room\_type** | Scrapped | Text Category |
| Phisic Listing Features | **bed\_type** | Scrapped | Text Category |
| Pricing | **price** | Scrapped | Number |
| Pricing | **weekly\_price** | Scrapped | Number |
| Pricing | **monthly\_price** | Scrapped | Number |
| Pricing | **security\_deposit** | Scrapped | Number |
| Pricing | **cleaning\_fee** | Scrapped | Number |
| Pricing | **extra\_people** | Scrapped | Number |
| Rental Charasteristics | **jurisdiction\_names** | ¿? | ¿? |
| Rental Charasteristics | **instant\_bookable** | Scrapped | Boolean |
| Rental Charasteristics | **is\_business\_travel\_ready** | Scrapped | Boolean |
| Rental Charasteristics | **requires\_license** | Scrapped | Boolean |
| Rental Charasteristics | **require\_guest\_profile\_picture** | Scrapped | Boolean |
| Rental Charasteristics | **require\_guest\_phone\_verification** | Scrapped | Boolean |
| Rental Charasteristics | **guests\_included** | ¿? | Number |
| Rental Charasteristics | **minimum\_nights** | Scrapped | Number |
| Rental Charasteristics | **maximum\_nights** | Scrapped | Number |
| Rental Charasteristics | **license** | Scrapped | Text |
| Rental Charasteristics | **cancellation\_policy** | Scrapped | Text Category |
| Rental Charasteristics | **accommodates** | Scrapped | Text Category |
| Renting Features | **experiences\_offered** | Scrapped | Text |
| Scrapping Info | **last\_scraped** | Calculated | Date |
| Scrapping Info | **calendar\_last\_scraped** | Calculated | Date |
| Scrapping Info | **scrape\_id** | Calculated | Number |
| url | **listing\_url** | Scrapped | Text |
| url | **thumbnail\_url** | Scrapped | Text |
| url | **medium\_url** | Scrapped | Text |
| url | **picture\_url** | Scrapped | Text |
| url | **xl\_picture\_url** | Scrapped | Text |
| Calendar Info | **has\_availability** | Calculated | Boolean |
| Calendar Info | **calendar\_updated** | Scrapped | Date |
| Host Info | **host\_is\_superhost** | Scrapped | Boolean |
| Host Info | **host\_has\_profile\_pic** | Scrapped | Boolean |
| Host Info | **host\_identity\_verified** | Scrapped | Boolean |
| Host Info | **host\_since** | Scrapped | Date |
| Host Info | **host\_verifications** | Scrapped | List of Text Categories |
| Host Info | **host\_id** | Scrapped | Number |
| Host Info | **host\_listings\_count** | Scrapped | Number |
| Host Info | **calculated\_host\_listings\_count** | Calculated | Number |
| Host Info | **host\_response\_rate** | Scrapped | Number |
| Host Info | **host\_acceptance\_rate** | Scrapped | Number |
| Host Info | **host\_total\_listings\_count** | Scrapped | Number |
| Host Info | **host\_url** | Scrapped | Text |
| Host Info | **host\_name** | Scrapped | Text |
| Host Info | **host\_about** | Scrapped | Text |
| Host Info | **host\_thumbnail\_url** | Scrapped | Text |
| Host Info | **host\_picture\_url** | Scrapped | Text |
| Host Info | **host\_location** | Scrapped | Text Category |
| Host Info | **host\_neighbourhood** | Scrapped | Text Category |
| Host Info | **host\_response\_time** | Scrapped | Text Category |