

### 1. What is the problem you want to solve?

Use Airbnb listing data from major metropolitan areas segmented by different continents (America, Europe, South America, Asia and Australia) with economic data indicative of the levels of Economic Development in those areas to develop models that demonstrate trends in listings occupancy rate, availability and pricing.

The model can be used to predict behavior of listing inventory and user behavior in certain geographic areas based on economic indicators.

(At this point, the specifics of the correlations that I will be analyzing are to be further specified and more refined as I progress through the Machine Learning module of the course)

### 2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

Airbnb is leading the burgeoning 'sharing economy' with its peer to peer real estate rental business model. According to CNBC, the United States makes up only 16 percent of Airbnb's listings globally. The company has been investing in the overseas market, pitting themselves specifically against traditional online rental sites. This investment has appeared to have paid off; growth in those markets is much higher, ranging from 170 percent to 180 percent. Foreign markets offer more opportunity for organic growth because penetration is lower and internet adoption is rising.

Analyzing the relation of economic to listing data will provide valuable information into how customers are booking and hosts are listing inventory using Airbnb. These behaviors can be predicted based on the economic health and development by using the models developed in this project.

Specifically, the insights can be used to determine what kind of marketing needs to be employed to help bolster and maintain growth over traditional rental sites based on how users are going to act based on economic indicators in certain geographic areas.

### 3. What data are you going to use for this? How will you acquire this data?

The data available for Airbnb listings from websites <http://insideairbnb.com/index.html> and <http://tomslee.net/airbnb-data>:

Room\_id, Host\_id, Room\_type, City, Neighborhood, # of Reviews, Reviews, Overall\_Satisfaction/Rating, # Accommodates, # Bedrooms, # Bathrooms, Price per night

Minimum Required Stay, Latitude, Longitude

#### Metrics that can be calculated based on this data:

- **Occupancy rate:** Use the minimum stay, price and number of reviews to estimate the occupancy rate, the number of nights per year a listing is booked
- **Activity of Listings:** Based on the number, frequency reviews
- **Average price per night, Income per month or year, Availability** (based on min. Required stay), **Listings per Host, Type of Listing**

Economic Indicators from OECD(

<http://www.oecd.org/std/oecdmaineconomicindicatorsmei.htm>), World Bank

(<http://data.worldbank.org/>) and UN (<http://hdr.undp.org/en/content/human-development-index-hdi>) :

- **GDP:** Strength of Economy
- **GDP Growth:** Economic Growth
- **GDP/Capita:** Strength on Economy (good for comparison of countries)

- **Economic Structure:** Shows the division of a country's economy between primary, secondary and tertiary industries.
- **Major Industries:** based on exports, largest sectors of the GDP (based on major exports, sectors of GDP)
- **Human Development Index:** To balance inaccuracies, indices tend to be an amalgamation of many different indicators. The United Nations Human Development Index (HDI) is a weighted mix of indices that show life expectancy, knowledge (adult literacy and education) and standard of living (GDP per capita). As Vietnam has a higher literacy rate and life expectancy than Pakistan, it has much higher HDI value even though it has a similar per capita GDP.

**4. In brief, outline your approach to solving this problem (knowing that this might change later).**

- First, segment the cities geographically and analyze economic trends and standing. Choose which and how many cities to use from major continents ((America, Europe, South America, Asia and Australia) Within each continent, cities can be segmented by economic strength, major industries etc...
- Look closely to determine the best economic indicators across continents. Possibly develop an index that combines multiple different indicators.
- Develop the Airbnb statistics that I want to look at to analyze user and host behavior based on data available from website.
- Pair the two sets of datas to look for interesting trends, correlations, outliers. Perhaps a predictive model based on economic indicators.
- Create visualizations, possible geographic maps that demonstrate the areas based on how the cities have been segmented.

**5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

My deliverables are going to be a python notebook, paper and a slide deck.