**MSDS Kaggle Project**

# Introduction

The purpose of this paper is to provide a predictive statistical analysis of the New York Airbnb Rental price using New York Airbnb database. The first part of the analysis provides an answer to a question of relationship between rental price and living space, and the potential influence from the relationship of a subset of the neighborhoods provided by the dataset.

The second part of the analysis offers an attempt to provide a statistical model that would offer the best pricing prediction. This time, all neighborhoods in are included.

## Data Description

(Where did the data come from? How big is it? How many observations? Where can we find out more? What are the specific variables that we need to know with respect to your analysis?)

Data came from the American Statistical Association website.   
<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/downloads/new-york-city-airbnb-open-data.zip/3>

The data set is comprised of ? observations, and ?exploratory variables that describe numerous aspects of Airbnb in New York City. All these exploratory variables are collected with the goal of predicting the selling price of those residential homes.

* NeighborhoodGroup = (Bronx, Brooklyn, Manhattan, Queens, Staten Island)
* 𝜇{RentalPrice | GrLIvArea, NeighborhoodGroup} = + GrLIvArea + Bronx + Brooklyn + Manhattan + Queens+ Staten Island + GrLIvArea \* Bronx, + GrLIvArea \* Brooklyn + + GrLIvArea \* Staten Island
* 𝜇{RentalPrice | GrLIvArea, NeighborhoodGroup = Brooklyn } = ( + ) + GrLIvArea
* 𝜇{RentalPrice | GrLIvArea, NeighborhoodGroup = Manhattan} = ( + ) + GrLIvArea
* 𝜇{RentalPrice | GrLIvArea, NeighborhoodGroup = Queens } = ( + ) + GrLIvArea
* 𝜇{RentalPrice | GrLIvArea, NeighborhoodGroup = Staten Island } = ( + ) + GrLIvArea