**Final Year Project Plan**

**Supervisor: Giacomo Livan**

Fazaan Hassan

09/11/2019

**Project Title**

The Use of Statistical Natural Language Processing Techniques to Rationalise the Evolution of Airbnb Reviews

**Aims & Objectives (A &O)**

Aims

* Understand how NLP models are used to infer topics and what impact this has had on our findings
* Answering the following: Can we see a change in the complexity of language overtime, and if so, is this also prevalent on a granular scale such as boroughs.

Objectives:

* Review such models such as LDA, NFM and BOW and understand their working in the results obtained
* Find suitable metrics to qualitatively assess the underlying numerical interpretation
* Evaluate the success of both the models, and NLP statistical metrics with regards to their definition.
* Analyse the above by comparing and contrasting the results with respect to subsets of the data

**Expected Outcomes/Deliverables**

* Literature Survey summarising previous technologies used and results obtained.
* An analysis of new, yielded results and their significance in relation to both social economical uses of the excavated source.
* An analysis of the various models implemented in terms of performance, scalability and robustness.
* Future work and how improvements can be made to the analysis conducted. This includes the models, metrics and data used.

**Work Plan**

* Start of October – End of November (Wait for ethics approval)
  + First 4 Weeks
    - Read Papers and understand various NLP methods used
    - Collate published results
  + Last 4 Weeks
    - Start defining requirements
    - Plan what will be implemented and how the NLP techniques will be used
* Start of December – Mid-January (When ethics approved, download dataset and start first iteration)
  + First 4 Weeks
    - Understand data and collect statistical results
    - See how results change with respect to time
  + Last 6 Weeks
    - Concentrate on specific boroughs and properties
* Mid-January – Mid-February
  + Research and try to implement robust classification techniques for types of properties
  + Using the above, use NLP models to infer topics and draw statistical correlations
* Mid-February – End of March
  + Work on final report