Assessed Coursework Set Front Page

**Module code:** INMR77

**Module name:** Business Intelligence and Data Mining

**Lecturer responsible:** Dr Yin Leng Tan

# Assignment Specification:

The module is assessed ***100% through this coursework assignment***.

This coursework aims to assess your knowledge of business intelligence and data mining, and your ability to perform data mining tasks by applying suitable concepts, methods and techniques learned during the lectures and practical sessions for business intelligence.

The coursework is carried out ***individuall****y*. You are required to produce a **report of 20 pages of A4 (+/- 10%),** including tables and diagrams but excluding references and appendices, based on the case as described in this coursework document.

An appendix can be used to include support materials to back up main body points where necessary. Y**ou are also required to submit the supplementary materials** of your work ***using SAS Enterprise Miner*** on blackboard by the specified deadline.

1. **The Case – Opportunities and Challenges of Sharing Economy: Airbnb and InsideAirbnb**

# Icon Description automatically generated[Airbnb - Holiday Lets, Homes, Experiences & Places](https://www.airbnb.co.uk/)

**Airbnb (Airbnb.co.uk)** is an online marketplace for arranging or *offering* ***short-term rental/lodging*** *i.e. temporary accommodation, primarily homestays, or tourism experiences*. It was founded in August 2008 by Brain Chesky and friends, and it

currently has 6,300 employees as in 2021.

On December 10, 2020, Airbnb went public with a valuation of over $100 billion, making it one of the largest IPOs (Initial Public Offerings) of 2020. It is reported that Airbnb capital market was more than the top three largest hotel chains (Marriott, Hilton, and Intercontinental) combined1. Though some are calling over evaluation, the company lacks traditional mortgages, employee fees, and maintenance fees which burden hotels. Airbnb hosts pay their own mortgage and clean their apartments, leaving the company much freer of debt, thus making it far more valuable.

### Airbnb service overview

Airbnb ***provides a platform*** for hosts to accommodate guests with *short-term lodging and tourism-related activities*. Guest can search for accommodation using filters such as location, price, specific types of home. Before booking, users must provide personal and payment information. Some hosts also require a scan of government-issues identification before accepting a reservation. Hosts provide prices and other details for their rental or listing e.g. number of guests included in the price, type of property, type of room, number of bathrooms, number of bedrooms, number of beds and type of bed, and minimum number of nights for a reservation, and amenities.

In addition, Airbnb provides **a *guest review system*** where hosts and guests can leave reviews about their experience, and rate each other after a stay. However, the truthfulness and impartiality of reviews may be adversely affected by concerns of future stays because prospective hosts may refuse to host a user who generally leaves negative reviews. Besides, the company's policy requires users to forego anonymity, which may also detract from users' willingness to leave negative reviews.

### Criticism of Airbnb

Airbnb has attracted criticism for increasing housing/residential rental prices in cities where it operates and creating nuisances and security issues etc for those living near leased properties and has negatively affects the quality of life in residential areas, and housing crisis around city in the UK, USA and Europe. The company has attracted regulatory attention from cities such as San Francisco, New York City, and the European Union over the past number years. It has also faced challenges from the hotel industry and other, similar companies.

Airbnb has made a quarter (25%) of its global workforce redundant in 2020 due to the global pandemic2. But the news was welcome by some campaigners who were fighting for soaring rents in cities with large number of Airbnb hosts. The number of ***longer-term rental properties (i.e. residential as opposed to short-term/holiday lets)*** in central Dublin was up 71% on comparable period last year, ***as landlords abandoned short-term lets through Airbnb***3.

1 <https://www.businessinsider.com/airbnb-ipo-valuation-tops-three-hotel-chains-combined-opening-day-2020-12?r=US&IR=T>

2 <https://www.theguardian.com/technology/2020/may/06/airbnb-to-make-quarter-of-its-global-workforce-redundant>

3 <https://www.rte.ie/news/ireland/2020/0417/1132149-rent-sale-property-ireland/>

# Image result for insideairbnb logoInside Airbnb – Adding Data to the Debate (insideairbnb.com)

**Inside Airbnb** (insideairbnb.com) is an independent, non-commercial set of tools and data that allows individual to explore how Airbnb is really used in cities around the world. It was set up by Murray Cox and John Morries in 2016.

Airbnb claims to be part of the “sharing economy” and disrupting the hotel industry by *offering* ***short-term rental/lodging***. However, data shows that most Airbnb listings in most cities ***are entire homes, many of which are rented all year round (i.e. illegal short-term rentals)***– disrupting housing and communities.

Most recently, New York City’s plans to crack down on illegal short-term rentals which could remove as many as 10,000 Airbnb listing is sparking fierce debates about housing, hotels, the tourist market and residents’ rights4.

By analysing publicly available information about a city’s Airbnb’s listings, Inside Airbnb provides filters and key metrics so users can see how Airbnb is being used to compete with the residential housing market. With **Inside Airbnb**, user can ask fundamental questions about Ainbnb in any neighbourhood, or across the city as a whole, such as:

* how many listings are in my neighbourhood and where are they?
* how many houses and apartments are being rented out frequently to tourists and not to long-term residents?
* how much are hosts making from renting to tourists (compare that to long-term residential rentals)?
* which host are likely running a business with a multiple listings and where are they?

These questions (and the answers) get to the core of the debate for many cities around the world, with ***Airbnb claiming that their hosts only occasionally rent the homes in which they live***. In addition, many cities or state legislation or ordinances that address residential housing, short term or vacation rentals, and zoning usually make reference to allowed use, including:

* how many nights a dwelling is rented per year
* minimum nights stay
* whether the host is present
* how many rooms are being rented in a building
* the number of occupants allowed in a rental
* whether the listing is licensed

The Inside Airbnb tool or data can be used to answer some of these questions. Some understanding of how the Airbnb platform is being used will help clear up the laws as they change.

*Further information of Airbnb, please visit:* <https://www.airbnb.co.uk/>

*Further information of Inside Airbnb, please visit:* <http://insideairbnb.com/index.html>

4 <https://www.theguardian.com/us-news/2023/jan/26/nyc-airbnb-short-term-rental-new-law>

# Coursework requirements:

The sharing economy has brought opportunities and challenges to homeowners, society, residents, communities and governments. **One of the biggest issues with Airbnb is whether hosts are renting out residential properties permanently as hotels (i.e. illegal short-term rentals) or sharing the primary residence in which they live "occasionally" (i.e. legal short- term rentals**).

Airbnb could easily answer this question but instead it is up to us to shape our communities and solve our urgent need to house tourists, housing shortage/crisis, and to address the nuisances, security and safety issues etc for those living near leased properties by Airbnb.

In this assignment, you are required to carry out data mining tasks **using data** of ***Airbnb listings of Edinburgh, Scotland, UK from InsideAirbnb***, and to report your findings as a result your data mining/analysis to address the challenges and issues of the sharing economy of Airbnb.

## DATA

The Edinburgh data (16 December 2022) is available to download from InsideAirbnb and on Blackboard as shown in Figure 1 below.

<http://insideairbnb.com/get-the-data>



Figure 1: Edinburgh Data compiled by InisideAirbnb as on 16 December 2022 (- <http://insideairbnb.com/get-the-data>

As shown in Figure 1, the data set includes:

* + 1. **Listings.csv.gz** contains **detailed listing data of Edinburgh**. The data was compiled on 16 December 2022. Each row of the data represents a single listing and

contains information about the host of the property, the property’s characteristics and overall rating of the property and its features by guests. There are 7,390 listings and 67 variables in the data set. Listing can be deleted in the Airbnb platform. The data presented is a snapshot of listings available at a particular of time as on and up to 16 December 2022.

* + 1. **Reviews.csv.gz** contains the **detailed reviews data for each listing**. The data was used for a number of derived variables in the detailed listing data e.g. number\_of\_reviews, number\_of\_review\_ltm, first\_review, last\_review, and reviews\_per\_months.
    2. **Calender.csv.gz** contains detailed calendar data i.e. the availability calendar for 365 days in the future for each listing. In addition
    3. **A data dictionary –** can be viewed and downloaded from [https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2Bo](https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit#gid%3D1322284596) [UGoNRIGa6Szc4/edit#gid=1322284596](https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit#gid%3D1322284596)

## 2.2. Your tasks

You are required to use the ***detailed listing data*** (**listings.csv)** to find meaningful pattens and rules of whether hosts in Edinburgh are renting out residential properties as hotels/business (illegal short-term rentals) or genuine sharing the primary residence in which the live

“occasionally” (legal short-term rentals).

You are expected to conduct cluster analysis (unsupervised learning) to differentiate hosts/listings that are likely to be genuine short term let (legal short-term rentals) versus hosts/listings that are likely to be operating as a business (illegal short-term rentals) based on information about the host of the property, the property’s characteristics and overall rating of the property and its features by guests etc, and to build a classification model that could differentiate hosts and/or listings that are for genuine (occasionally) short-term or vice versa. You are also expected to conduct further literature search on the issues and data exploration, data preparation for your clustering and classification tasks.

You may want to refer to the ***calendar (calendar.csv) and review data (*reviews.csv)** and derive further new variables where necessary.

## What to deliver

You are required to produce a **report** of **20 pages of A4 (+/- 10%)** including tables and diagrams but excluding references and appendices. An appendix can be used to include support materials to back up main body points where necessary. Y**ou are also required to submit the supplementary materials** of your work ***using SAS Enterprise Miner*** on blackboard by the specified deadline.

The total of 100 marks will be allocated to the following aspects of the report, which should also be used as a guideline to structure the report.

## Introduction (10%)

The introduction section should include the background and context, problems and issues of sharing economy and Airbnb, and a clear statement of the data mining goal. You are expected to justify your statements using relevant literature and sources.

## Data understanding (10%)

In this section, you are expected to conduct exploratory data analysis e.g. summaries statistics and data visualisation techniques, using suitable and relevant techniques and methods and report your key findings, including variables and measurement identified for your data preparation tasks.

## Data Preparation (20%)

In this section, you are expected to take the data identified in the previous step and prepare them for the model building. This should include:

* 1. data cleaning e.g. missing data handling
  2. data transformation e.g. creating new derive variables

You are expected to create two new derive variables “Occupancy rate” – estimating how often an Airbnb listing is being rent out, and “listing income” i.e. approximate a

listing’s income. You are expected to justify the approaches taken backup by relevant

sources or literature, and

* 1. data reduction (e.g. correlation analysis)

Make sure to include figures and tables (screenshot) to support your analyses and findings.

## Cluster Analysis and Results Interpretation (20%)

In this section, you are expected to conduct cluster analysis i.e. identifying clusters/segments of listings based a set or combination set of variables e.g. host’s characteristic,

listings/property’s characteristics and availability, and reviews from guests etc. This should include,

* 1. a list of variables and clustering techniques used with reasoning.
  2. result interpretations and comments on the characteristics of the clusters/segments obtained.

Make sure to include figures and tables (screenshots) to support your model buildings, analyses and findings. Supplement materials can be provided at the appendix section.

## Classification Model Building and Model Evaluation (20%)

In this section, you are expected to build a classification model based on the results obtained from your cluster analysis above. Since this information would be most likely to be used to differentiate those listings/hosts that are likely to be “illegal short-term rentals”, it would be more meaningful to select segment/cluster(s) that would likely be defined as “illegal short-term rental” in your classification model. This should include:

* 1. Further data preparation where necessary.
  2. Model building – a list of variable and classification methods used and provide your reasoning.
  3. Model evaluation.

Make sure to include figures and tables (screenshots) to support your model buildings, analyses and findings. Supplement materials can be provided at the appendix section.

1. **Conclusion, critical evaluation, and suggestion for model improvements (10 marks)** In this section, you should conclude the outcomes of your findings in relation to the data mining goal. Discuss the limitations of your data mining process, this might include the assessment of the suitability of data and variables, methods and techniques used, assumptions made, and provide suggestion for model improvements.

In addition, there are **10 marks allocated to the structure** (clarity of organisation and structure - addresses all components of the assignment brief with appropriate weighting across each component, logical structure to the overall argument that is easy to follow), **and presentation (**e.g. effective use of tables and diagrams, proper use of citation and referencing in an Author-Year e.g., Harvard, APA format, length/page limit).