

## **HDIP. DATA ANALYTICS**

CA 2 - STRATEGIC THINKING

SEMESTER 2

***AIRBNB PRICE - PREDICTIONS***

**Lecturer(s):** James Garza

**Issue Date:** 19th September 2023

**Submission Deadline Date**: 15th November 2023

**Students:** Bárbara Abreu Costa 2023099 | Juliana Cristina da Cunha Brum 2023021

**GitHub Repository:** <https://github.com/Babreucosta/CA2_Strategic_Thinking.git>

SUMMARY

1. Introduction 3
2. Business Description 4
   1. Hypothesis 4
   2. General goal 5
   3. 2.3 Success 5

1 - Introduction

2 - Business Description

The travel and lodging industries were revolutionised by Airbnb, a top online marketplace and hospitality platform, founded in 2008.

Connecting people who want to rent out their properties, typically their homes or apartments, with travellers seeking accommodations, Airbnb platform has had a transformative impact on the travel industry, disrupting traditional models and opening up new possibilities for both hosts and travellers. With its innovative platform and commitment to fostering connections and cultural exploration, Airbnb continues to shape the way people travel and experience the world.

“Airbnb is an American [San Francisco](https://en.wikipedia.org/wiki/San_Francisco)-based company operating an [online marketplace](https://en.wikipedia.org/wiki/Online_marketplace) for short- and long-term [homestays](https://en.wikipedia.org/wiki/Homestay) and experiences. The company acts as a [broker](https://en.wikipedia.org/wiki/Broker) and charges a [commission](https://en.wikipedia.org/wiki/Commission_(remuneration)) from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com. Airbnb is the most well-known company for short-term housing rentals.” (Wikipedia Contributors, 2019)

2.1 Hypothesis

Last semester, we were engaged by a consortium of investors seeking to identify suitable properties in Dublin for the establishment of Airbnb accommodations, with the objective of attaining Superhost status on the Airbnb platform. Now, those investors need to know information about the prices they could apply in the properties, to finally set up the business.

To assist us in this endeavour, we possess a comprehensive dataset encompassing information pertaining to various properties across London, a distinct European city, covering both weekday and weekend statistics. This dataset will be a useful tool for understanding the ideal contextual conditions necessary for the development of Airbnb properties in Dublin. Additionally, it will help us identify the essential qualities that such homes ought to have in order to become an excellent Airbnb choice.

2.2 General goal

The general goal of this Project is to provide insights into the Dublin Airbnb market and identify key factors that can contribute to the success of Airbnb accommodation in the city. Our analysis will help our clients make informed decisions about where to invest and what to offer to achieve optimal results, and also they will have an estimated price they could charge in short or long term rents.

2.3 Success

The success of this project will be measured by our ability to provide our clients with invaluable insights and actionable recommendations for establishing thriving Airbnb accommodations in Dublin. Armed with a comprehensive dataset from London, we are well-equipped to not only estimate optimal pricing strategies but also leverage latitude and longitude data to explore location-based price predictions. We expect that our analysis identify crucial factors that contribute to achieving Superhost status on the Airbnb platform, ensuring a high level of customer satisfaction and repeat bookings. Also enable our clients to make informed decisions about property acquisition and investment and ensuring a steady stream of satisfied guests. Ultimately, our goal is to equip our clients with the knowledge and tools they need to build a profitable and enduring presence in the competitive landscape of Airbnb hosting.

3. Data understanding

As you have already read before, the goal of this Project is making some predictions about the prices based on the Airbnb dataset to open new accommodations in Dublin. Now, you are going to understand more about the data itself and about the technologies used to do the analysis.

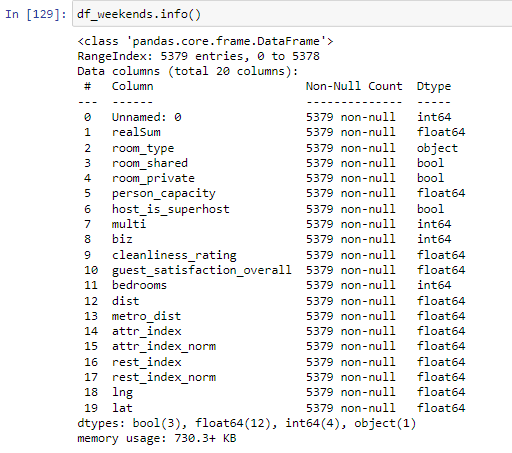
3.1 - Data overview

The data set used in this project was Airbnb Prices in European Cities, available in (www.kaggle.com, n.d.).

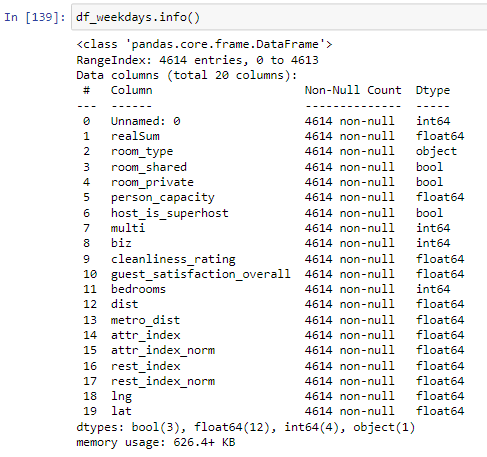
This dataset is divided into 10 European Cities, and for each city it has information about weekdays and weekends. However, for this analysis, we only used London city to work with. Because in the whole list it is the one that is closest to Dublin and also, has a similar weather.

The Airbnb dataset has in total 20 attributes around 5379 rows for weekends and around 4600 rows for weekdays. The columns have different kinds of data, such as integer, float, object and Boolean. As we can see in the images below.

Weekends:



Weekdays:



3.2 - Data preparation

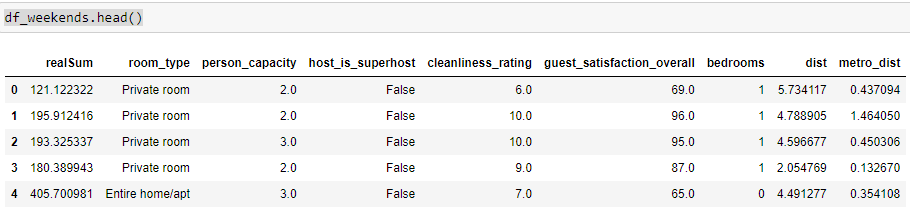
Data preparation is a Project’s step where the data might be modified in order to apply Machine Learning models.

“Data preparation is the act of manipulating (or pre-processing) [raw data](https://en.wikipedia.org/wiki/Raw_data) (which may come from disparate data sources) into a form that can readily and accurately be analysed.” (Wikipedia Contributors, 2019b).

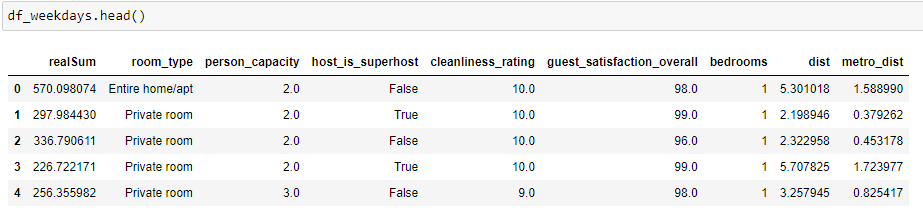
Certain unimportant columns, such as 'Unnamed: 0', 'room\_shared', 'room\_private', 'multi', 'biz', 'attr\_index', 'attr\_index\_norm', 'rest\_index', 'rest\_index\_norm', 'lng', 'lat', were exclude.

The columns below were kept.

Weekends:



Weekdays:



As the columns ‘room\_type’ has object values and the columns ‘host\_is\_superhost’ has boolean values. In order to apply Machine Learning, it was necessary to change them for numeric values using pd.get\_dummies.

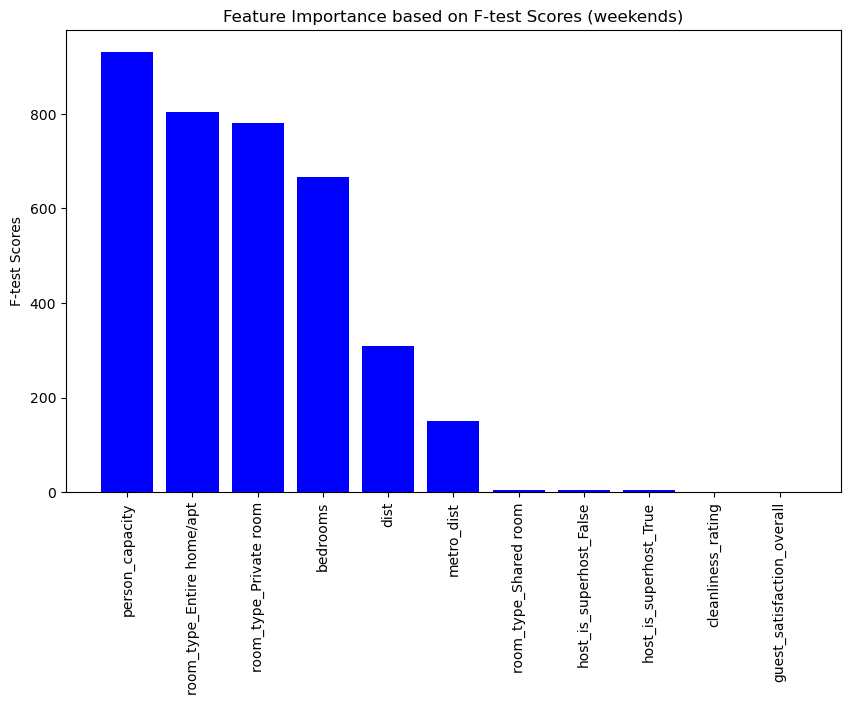
“When fitting Machine Learning algorithms (like [linear regression](https://www.statology.org/multiple-linear-regression/), [logistic regression](https://www.statology.org/logistic-regression/), [random forests](https://www.statology.org/random-forests/), etc.), we often convert categorical variables to dummy variables, which are numeric variables that are used to represent categorical data.” (Zach, 2021).

4 - Feature Selection and F-test

In order to check which features could have more relevance to help the models make predictions about the price, the F- test was applied.

“ F-Test is useful in feature selection as we get to know the significance of each feature in improving the model.” (Asaithambi, 2018).

F-test result in the weekends dataset:



F-test scores in the weekdays dataset:

