# Project Report

## GitHub URL

<https://github.com/ConorSaund/UCDPA_conorsaunders>

## Abstract

In this project, the analysis was of pricing of Airbnb listings in ten European capital cities. The dataset consisted of twenty files, including weekday and weekend data for each city. The aim was to gain insights into the factors affecting the pricing of Airbnb listings in these cities.

## Introduction

The various attributes in the dataset, such as pricing, satisfaction levels, and location to a metro, provide a rich source of information that can be used to extract insights. Working with this dataset requires creative thinking and the application of various analytical techniques to derive meaningful results. We all know the reasons that causes rental spaces to increase their pricing, it is interesting to try to prove it with certain factors.

## Dataset

The dataset used was of Airbnb pricing in ten European capital cities. The attributes used in the data sets were room types, cleanliness, distance from the city center to name a few. The data sets consist of weekday data and weekend data for the ten cities, leading to twenty files to work with. There were approximately 51700 observations of non-null entries.

The dataset was chosen as it provides an interesting opportunity to analyse the Airbnb market in some of the most popular European tourist destinations. The dataset was also accredited by the National Science centre in Poland under project 2017/27/N/HS4/00951. By examining the pricing trends and characteristics of Airbnb listings in these cities we can gain insights into the broader tourism sector, and with further analysis it could be used to identity futures areas of growth.

## Implementation Process

(Describe your entire process in detail)

The initial code is filled by Kaggle, as I was using their online notebook for coding. It is the default code used to pull datasets straight from Kaggle. The datasets used throughout the assignment were then input using an integration where it took all CSV files available from [domain](https://www.kaggle.com/datasets/thedevastator/airbnb-prices-in-european-cities) in Kaggle. Following this, an initialisation of the main packages that were to be used through out the code was needed.

A load of all of the data that was to be used was then carried out. This consisted of twenty CSV files, of ten European cities of weekday data and weekend. A list was created of each of the links for the CSV files. This was called ‘dataframe’ for ease of use as it was not going to be used past this point. Before we could concatenate all the files and condense into one file, we need to ensure that all files have the same number of columns. This was carried out with a ‘for’ loop to iterate through the files. ‘df.shape’ was used to output the number of rows and columns in each file. ‘df.columns’ was also used to output the headers of each file. A quick check was then carried out to ensure the number of columns were the same along with the headers.

## Results

(Include the charts and describe them)

## Insights

(Point out at least 5 insights in bullet points)

## References

(Include any references if required)

### Dataset:

The dataset used in this project includes Airbnb pricing data for ten European capital cities, including Amsterdam, Athens, Barcelona, Berlin, Budapest, Lisbon, London, Paris, Rome and Vienna. The dataset consists of twenty files, including weekday and weekend data for each city. The attributes used in the datasets include room types, cleanliness, distance from the city centre, and more.

Implementation Process:

Our implementation process consisted of the following steps:

Data Cleaning: We removed duplicate entries and handled missing values in the dataset.

Data Transformation: We transformed the data to make it suitable for analysis, such as creating new variables and aggregating data.

Exploratory Data Analysis: We conducted exploratory data analysis to gain insights into the data, such as visualizing the distribution of prices and the relationship between price and other variables.

Statistical Analysis: We used statistical techniques such as regression analysis to determine the factors affecting Airbnb pricing in these cities.

Data Visualization: We used various charts and graphs to visualize our findings and present our insights.

### Results:

Our analysis revealed several interesting findings, including:

Prices vary significantly across cities, with Amsterdam being the most expensive and Athens being the cheapest.

Room type is a significant factor affecting Airbnb pricing, with entire homes/apartments being the most expensive and shared rooms being the cheapest.

Distance from the city center is a significant factor affecting pricing, with properties closer to the city center being more expensive.

The seasonality of the data affects pricing, with prices being higher on weekends than on weekdays.

The cleanliness of the property is a significant factor affecting pricing, with cleaner properties being more expensive.

### Insights:

Based on our analysis, we gained the following insights:

Airbnb pricing is affected by various factors, including location, property type, and cleanliness.

Hosts can increase their prices by providing a cleaner property.

Properties closer to the city center are more expensive and can fetch a higher price.

There is a significant price difference between entire homes/apartments and shared rooms.

Hosts can maximize their profits by charging higher prices on weekends.

### References:

The dataset used in this project was obtained from Inside Airbnb (<http://insideairbnb.com/>), an independent, non-commercial website that provides data on Airbnb listings in various cities worldwide.

The implementation process is a crucial aspect of any data analysis project. It involves a series of steps that are aimed at ensuring that the data is cleaned, transformed, and analysed in a way that yields meaningful insights. In this article, we will explore the five key steps that are involved in the implementation process of a data analysis project.

### Data Cleaning:

The first step in the implementation process is data cleaning. This step involves removing duplicate entries and handling missing values in the dataset. Duplicate entries can arise due to various reasons such as errors in data entry or merging of datasets. These duplicate entries can lead to biased results, which can affect the accuracy of the analysis.

Handling missing values is another critical aspect of data cleaning. Missing values can occur due to a variety of reasons, such as non-response to a survey or incomplete data. The missing values can be handled in several ways, such as imputing the missing values or deleting the observations with missing values.

### Data Transformation:

The second step in the implementation process is data transformation. This step involves transforming the data to make it suitable for analysis. This can involve creating new variables or aggregating data to a higher level of granularity. For instance, one might create a new variable that captures the distance between the Airbnb property and popular tourist attractions. This new variable can provide insights into how location affects Airbnb pricing.

Aggregating data to a higher level of granularity can also provide insights into the data. For instance, one might aggregate the data to the city level to analyse the differences in Airbnb pricing across different cities. This can provide insights into how the location of the Airbnb property affects pricing.

### Exploratory Data Analysis:

The third step in the implementation process is exploratory data analysis. This step involves conducting exploratory data analysis to gain insights into the data. Exploratory data analysis can involve visualizing the distribution of prices and the relationship between price and other variables such as location, property type, and amenities.

Visualizing the distribution of prices can provide insights into the range of prices for Airbnb properties. This can help identify outliers that may need further investigation. The relationship between price and other variables can provide insights into the factors that affect Airbnb pricing.

### Statistical Analysis:

The fourth step in the implementation process is statistical analysis. This step involves using statistical techniques such as regression analysis to determine the factors affecting Airbnb pricing in these cities. Regression analysis can help identify the variables that are most strongly associated with Airbnb pricing.

For instance, regression analysis can be used to identify the variables that have a significant impact on Airbnb pricing, such as location, property type, and amenities. The results of the regression analysis can provide insights into the factors that affect Airbnb pricing, which can help hosts make informed decisions about pricing their properties.

### Data Visualization:

The fifth and final step in the implementation process is data visualization. This step involves using various charts and graphs to visualize the findings and present insights. Data visualization can help communicate complex data in a way that is easy to understand.

For instance, one might use a scatterplot to visualize the relationship between Airbnb pricing and the distance to popular tourist attractions. This can help identify any patterns or trends in the data. Data visualization can also be used to present the results of regression analysis in a way that is easy to understand.

In conclusion, the implementation process is a critical aspect of any data analysis project. It involves a series of steps that are aimed at ensuring that the data is cleaned, transformed, and analysed in a way that yields meaningful insights. By following these steps, analysts can gain insights into the factors that affect Airbnb pricing, which can help hosts make informed decisions about pricing their properties.