Solent University

Faculty of Business, Law and Digital Technologies

**The Data Project**

**(Software Artefact)**

Author : Nwanneka Ngozichukwu Okoli

Course Title : Programming for Problem-Solving

Course Code : COM 728

Module Leader : **Jarutas Andritsch**

Date : April, 2023

Table of Contents

[1. Overview 3](#_Toc127897102)

[2. Project Implementation 4](#_Toc127897103)

[2.1 Project structure 4](#_Toc127897104)

[2.2 Modules/ Functions 5](#_Toc127897105)

[2.2.1 Module/ function name 5](#_Toc127897106)

[2.2.2 Module/ function name 5](#_Toc127897107)

[2.2.3 Module/ function name 5](#_Toc127897108)

[2.2.4 … 5](#_Toc127897109)

[3. GitHub Repository Evidence 6](#_Toc127897110)

Table of Figures

Figure 1 Customer Behaviour of Car Sale Market Project Structure (Andritsch, J., 2022) 4

Figure 2 Code to perform size checks on entities 5

Index of Tables

Table 1: Task Completion 4

# 1. Overview

The font style to use throughout your document is Trebuchet MS as this the font style recommended by Solent University. The font size for text in the body of your document should be 12 pt. Headings should have a suitable font size e.g., 18 pt for main section headings, 16 pt for sub-headings and 14 pt for sub-sub-headings.

In this section, you should explain the aim and objectives of the project. You should provide a summary of the data set briefly discussing the main features of the data. You should also detail any interesting observations regarding the data set.

You should also include a table summarising what requirement have been achieved.

Table 1: Requirement Completion

|  |  |
| --- | --- |
| **Requirement** | **Status** |
| Load the data from a CSV file into memory using the CSV module |  |
| Retrieve a name of listing, host\_name, description, host\_location, and the date the host was created for an individual host by host\_id |  |
| Retrieve host\_name, property\_type, price, minimum\_nights, and maximum\_nights of all Airbnb listing for a specified location |  |
| Retrieve room\_type, accommodates, bathrooms, bedroom, and beds of all Airbnb listing for a specified property type |  |
| Retrieve specific columns of your choice related to an individual host by location |  |
| Load the data from a CSV file into memory using the pandas module |  |
| Identifying the most popular amenities or features that Airbnb guests are looking for |  |
| Analyse the average price of stay in each location |  |
| Analyse the average review scores rating for each location |  |
| Analyse to get insightful information based on your own selection |  |
| Display the proportion of number of bedrooms of Airbnb listing using pie chart |  |
| Display the number of listings for each room type using bar chart |  |
| Display the relationship between accommodates and price using scatter plot |  |
| Display Airbnb prices from 2019 - 2022 with line chart using subplots (one year per plot) |  |
| Display a visualisation of your choice to present customer usage behaviour on Airbnb |  |

**Status options:** Completed/ Partially Completed/ Not Attempted

# 2. Project Implementation

A heading should not be followed directly by another heading. There should be some connecting text. Thus, you should briefly describe the structure of this section here. Then, you can explain the detail in the individual section below. Your explanation should be clear, concise and justified.

#### 2.1 Project structure

Explain the structure of your project. Your explanations should include appropriate detail and terminology.

It would be useful to include a diagram showing your project structure. The diagram can be included in this section as a figure or, if it is large, in the appendices and cited in this section.

You should ensure that any figures or tables are appropriately label following suitable conventions. Generally, the figure or table is centred, and figure captions are placed below a figure and table captions are placed above a table.

The example below shows a figure with a caption.

|  |
| --- |
|  |

Figure 1 Customer Behaviour of Car Sale Market Project Structure (Andritsch, J., 2022)

The example of a figure caption above includes a reference. If an image is taken without any modifications from another source, then it should include a reference as shown in the example above. If the image has been adapted then this should be stated in the caption e.g., Figure 1 – Car Market Analysis Project (adapted from Andritsch, J., 2022)

#### 2.2 Modules/ Functions

Explain the implementation of each module/function in your project. You can present each module/function using sub-section topic. More complicated modules/functions should have greater detail whilst smaller and similar functions can be grouped together in your explanation. You may wish to include relevant code snippets to support your explanations. However, you should ensure that you do not simply restate or summarise what is shown in the code but explain how it works (what, how and why).

You should include code snippets as figures with appropriate captions. An example of a code snippet is shown below:

|  |
| --- |
|  |

Figure 2 Code to perform size checks on entities

The code snippet should be short, concise and appropriate formatted. If the code snippet is from another source, then you should also include a suitable reference.

##### 2.2.1 Module/ function name

Explain the functionality implemented in this module/function. Relevant guidelines mentioned in the previous sections should also be followed here.

##### 2.2.2 Module/ function name

Explain the functionality implemented in this module/function. Relevant guidelines mentioned in the previous sections should also be followed here.

##### 2.2.3 Module/ function name

Explain the functionality implemented in this module/function. Relevant guidelines mentioned in the previous section should also be followed here.

##### 2.3 Airbnb\_tui

The text user interface module of this software contains and implements ten(10) functions. The random module was also used in the implementation of this module. The functionalities of the functions are explained below:

1. Name\_of\_user: The name\_of\_user function is the first point of contact of the user with the software. It requests for the name of the use and welcomes the user to use the software.
2. Start: Start take in one argument which is message and informs the user about the start of the program and states which program the user is running.
3. End: End function informs the user when the program or operation selected by the user has been completed.
4. Error: The error function displays error message to the user when the user input a wrong value or option.
5. Exit\_msg: Exit message function uses the choice function of the random module to display a random exit message to the user. It also reassures the user that the program successfully exit after the user has inputted exit.
6. Users\_choice\_of\_program: This function asks the user about the choice of programme he wants to perform with respect to question a, b and c of the assessment requirement. The options include A, B, C, or EXIT. A performs tasks with the CSV module, B queries and analyses data using the Pandas module, C visualizes the airbnb dataset using the matplotlib module. EXIT quits the user from the software. The function converts the user’s selection to capital letter. Should the user input non alphabets the function continues to present and ask the user to make selections or exit. The values error exception was also employed to avoid the interruption of the software due to the wrong value.
7. Menu1, menu2, menu3, menu4: This function menus tasks of the assessment requirement and asks the user to make selection based on 1,2,3,4, (5), yes, no, and exit options. If – else statement is used to match the user's response and the user is informed of the selection he has made. If the user inputs wrong option, the function continues to present and ask the user to make selections or exit. The values error exception was also employed to avoid the interruption of the software due to the wrong value.

##### 2.4 Airbnb\_plot

The airbnb\_plot module interacts with airbnb\_tui module and supports the matplotlib.pyplot and pandas module. It has 6 own defined functions.

1. Proportion\_bedrooms funtion: This function takes in the airbnb data in dataframe as an argument. The function groups the airbnb data by the number of bedrooms, returns the number of occurrences of each number of bedrooms and sorts them in ascending order. The number of bedrooms as a category and the frequency is cast into a list and plotted using a pie chart.
2. Num\_listing\_roomtype: There are two sections in the function, the data preparation and the plotting section.

Data preparation: The airbnb dataset is first grouped by room\_type. Using the .size() and sort\_values() functions of the pandas module counts the number of listings of each room type and sorts the value in ascending order. The room types and the frequency of each room type is converted to a list using the tolist function of the pandas module.

Plotting: the room type is plotted against the number of listings of each room type using the bar chart. The x and y axes are labelled. The yticks is stated to range from 0 to 10000 with the step of 500 to show a more meaningful reading of the data due to smaller number of listings of shared room and hotel room types. The lower step size is not advised as it clusters the axis and harms readability of the graph. The plot is titles and the show function of the matplotlib is used to show the graph.

1. Accommodates\_and\_price function: this function slices the accommodates and price columns of the airbnb dataset and plot accommodates against the price of the airbnb using a scatter plot. The figure containing the graph is titled and the x and y axes labelled.
2. Prices\_per\_year: The prices per year function plots the sum of the daily prices in each month of year 2019 to 2022. The following steps was used in the execution of the function.
3. The host\_since column of the Airbnb dataset was sliced and converted into datetime.
4. Each year beginning from 2019 is filtered from the dataset.
5. The filtered years are grouped by month and the price column sliced and summed for each year.
6. 4 by 1 subplot is created, and each subplot would represent the Airbnb prices per month for each year.
7. For each year the index (months) is assigned to the (x axis) and the sum of the price made for each month assigned to y (axis).
8. Using the plot function of the matplotlib, line graphs is plotted and the labels, titles, suptitle and adjustments set properly.
9. Plot\_top10\_amenities: The function is a branch of the top\_10\_amenities function which plots the top ten (10) amenities provided by the hosts using a horizontal bar graph (why horizontal bar) to show how frequently the host provides each amenity.
10. Own\_selection\_c: This function displays the customer usage behaviour of the guests using four-row subplots. First, it plots using line graph to understand the variability of how the customers rate their experiences at the listings over the years and might suggest to future guests what month to use Airbnb or use other temporary accommodation options. For example, from the graph reveal a drop in the review rates in March over the last four years.

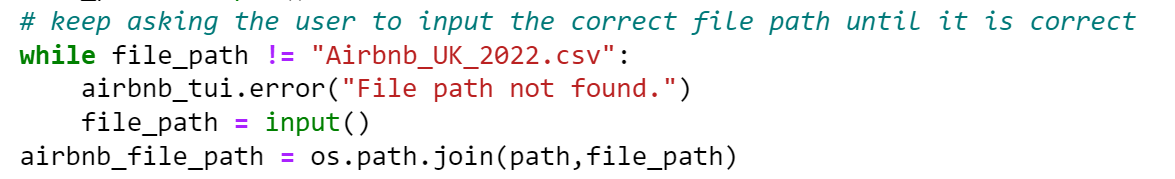
The second and third rows of the subplots uses a scatter plot to show the relationship between the number of minimum nights customers spends with respect to the review scores of accuracy, cleanliness, check-in, communication and location. At face view, the graphs depict that with increase in the review ratings, the minimum nights that customers spend at the listings also increase.

##### 2.4 Airbnb\_process

The airbnb\_process module implents functions to complete the tasks including working with the os, csv and pandas module to query and analyse airbnb data. This module also interacts with an own defined module, airbnb\_tui. The functionality of the functions in this

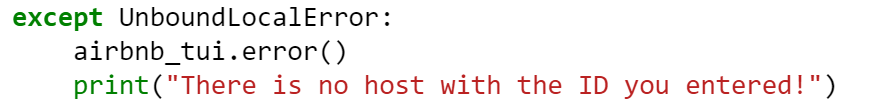
2.4.1 get file path function

The dataset used for the software is stored as Airbnb\_UK\_2022.csv in a parent directory called Data. This function gets the name of this Airbnb dataset from the user and uses the path.join() method of the os module to join the file path. The function also uses the while loop to check the user’s input to ensure the correct name of the file is entered.



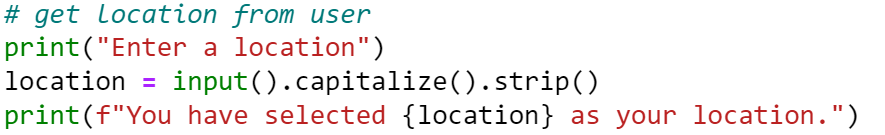
2.4.2 get by host id function

The function is designed to return the name of listing, host name, the date the host was created, and the description of the listing based on the host id entered by the user. This function takes in one argument which is the airbnb dataset and the function starts by prompting the user to enter a host id, it then loops through the data to search the host id and prints the necessary information regarding that host. In situations where the ID the user is looking for is not found, the function excepts the UnboundLocalError and the airbnb\_tui error function informs the user about the error and that the host ID was not found.



2.4.3 get by location function

The function takes in the airbnb dataset and loops through the dataset to retrieve information about hosts based on the location inputted by the user. The capitalize and strip function ensures only the first letter of the user’s input is in capital letter and is stripped of white spaces to ensure the format of the location entered by the user matches with that of the dataset. If the user enters a location that is not in the dataset, it prompts the airbnb\_tui error function and informs the user that the location was not found.



2.4.4 get by property type function

The function takes in the airbnb dataset and loops through the dataset to retrieve information about hosts based on the property type inputted by the user. The capitalize and strip function ensures only the first letter of the user’s input is in capital letter and is stripped of white spaces to ensure the format of the property type entered by the user matches with that of the dataset. If the user enters a property type is not in the dataset, it prompts the airbnb\_tui error function and informs the user that the property type was not found.

2.4.5 by\_location\_superhost function

This function gets information from the dataset based on the whether the host is superhost, the location and the review score rating the user is interested in. The user inputs the choices of selection and at each point the function puts the responses in the correct format and strips the responses of white spaces. This function is designed to enable the user to have a quick overview of make a choice of a host in any location and be able to decide to pick a listing based of the review score rating, the review score of the location, the value for money and whether the host can be booked instantly.

2.4.6 top\_10\_amenities function

This function displays the ten most popular amenites provided by the hosts. It take the dataset as an argument and slices the amenities column of the data and converts it into a list. Due to the format of the data in this column, a list of list, the function creates an empty list and adds all the amenities in the list of lists into the empty list using the extend method. The list is then converted to a pandas series and uses the value\_count function to count the items in the series and calls the first ten (10) most occurred amenities. The function then displays the output nicely to the user.

2.4.7 average\_price function

This function is implemented to analyse the average daily price of the listings for each location. The function first groups the data based on location and then slices the price, minimum and maximum nights and takes their mean. It the displays the output nicely to the user.

2.4.8 average\_review\_score function

This function is implemented to analyse the average review score of the listings for each location. The function first groups the data based on location and then slices the review score rating column and takes the mean. It the displays the output nicely to the user.

2.4.9 best\_rating\_price function

This function takes location from the user and returns the room types available in that location. The function will also calculate and display the overall rating for each room type available and the average price for each room type.

The function takes the average of all the review score ratings in the dataset to generate a new column called overall\_rating. The overall rating would determine the general view of airbnb guests about the listings. The function however makes a copy of the airbnb dataset to avoid changing the oroiginal dataset.

The function then groups the data by location and room type and slices the overall rating and price column. It then takes the location of interest from the user and returns the information to the user.

In situations where the location entered by the user doesnot exit or does not have any values the function informs the user that the lacation was not found.

The function also uses the background gradient methos of pandas style function to highlight the overall ratings for emphasis and order.

2.4.10 merge function

Explain the functionality implemented in this module/function. Relevant guidelines mentioned in the previous sections should also be followed here.

When you explain in a detail for each module/function a proper example or justification should present. For example, if you implemented your project using main module to control the logical work of your project and utilises the other modules, it may be useful to include an example or two of how the main module utilises the other modules to deliver a function. You may, for example, present this as a diagram or a series of annotated function calls or if you explain your implementation for visualisation, you can discuss how you have derived with specific selection of your own choice for visualisation (what and why). You should include suitable screenshots of the final visualisations.

# 3. GitHub Repository Evidence

A screen shot of your private Git repository. The screen shot need to clearly show your history of your commit of your project implementation. You need to click on the clock symbol on the right conner of your repo.

Graphical user interface, text, application, email

Description automatically generated

Sample screen shot of your commit history:

A screenshot of a computer

Description automatically generated