**Questions for the hotel dataset**

**Please attempt all the questions and in case of any doubt you may mail me, but do not keep the work pending to Saturday.**

Read the data

1. As a local file
2. As a file from the google drive

Describe the data

1. Display the first 20 rows
2. Display the last 10 rows
3. List the total number of rows and columns in the dataset
4. Use the describe() function and write down the details provided by it

Data Preprocessing

1. Find the null values by running the isna().sum() function eg: if “pd” is defined as the dataframe pd.isna(). Observe the columns with missing values and enlist them in the text block
2. Use the following function to remove the null values (i.e open a code block and type the following – please mind the indentation)

def clean\_data(pd):

pd.fillna(0,inplace=TRUE)

print(pd.isnull().sum())

and run the function by typing the clean\_data(pd) in the next code block

1. Type pd.columns() in another code block and observe the columns in the data

Data Visualization

1. Use the pairplot function with hue=’country’
2. Plot the heatmap for the data – In case data is too large for the plot, use

<dataframe> = <dataframe>.drop(data.index[-50])

to drop last 50 rows or \_50 for first 50 rows etc….