**Hotel Reservations Exercise**

**About this file**

The file contains the different attributes of customers' reservation details. The detailed data dictionary is given below.

**Data Dictionary**

* **Booking\_ID**: unique identifier of each booking
* **no\_of\_adults**: Number of adults
* **no\_of\_children**: Number of Children
* **no\_of\_weekend\_nights**: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
* **no\_of\_week\_nights**: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
* **type\_of\_meal\_plan**: Type of meal plan booked by the customer:
* **required\_car\_parking\_space**: Does the customer require a car parking space? (0 - No, 1- Yes)
* **room\_type\_reserved**: Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels.
* **lead\_time**: Number of days between the date of booking and the arrival date
* **arrival\_year**: Year of arrival date
* **arrival\_month**: Month of arrival date
* **arrival\_date**: Date of the month
* **market\_segment\_type**: Market segment designation.
* **repeated\_guest**: Is the customer a repeated guest? (0 - No, 1- Yes)
* **no\_of\_previous\_cancellations**: Number of previous bookings that were canceled by the customer prior to the current booking
* **no\_of\_previous\_bookings\_not\_canceled**: Number of previous bookings not canceled by the customer prior to the current booking
* **avg\_price\_per\_room**: Average price per day of the reservation; prices of the rooms are dynamic. (in euros)
* **no\_of\_special\_requests**: Total number of special requests made by the customer (e.g. high floor, view from the room, etc)
* **booking\_status**: Flag indicating if the booking was canceled or not.
  1. **It is recommended to start from the following libraries:**

import numpy as np *# linear algebra*

import pandas as pd *# data processing, CSV file I/O (e.g. pd.read\_csv)*

from sklearn import metrics

import matplotlib.pyplot as plt

import seaborn as sns

Read the file (change the file directory according your)

df=pd.read\_csv('/input/hotel-reservations-classification-dataset/Hotel Reservations.csv')

Or other option to read the file:

from google.colab import files

uploaded = files.upload()

import io

df = pd.read\_csv(io.BytesIO(uploaded['FileName.csv']))

* 1. **Investigate the data and add the data preprocessing methods (check the materials of the previous session).**

df.info()

df.nunique()

df

* 1. **Check the number of visitors:**

df.groupby('no\_of\_adults')['Booking\_ID'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='pie',autopct='**%1.2f%%**',subplots=True,title='Adults',figsize=(9,9))

df.groupby('no\_of\_children')['Booking\_ID'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='pie',autopct='**%1.2f%%**',subplots=True,title='Children',figsize=(9,9))

* 1. **Nights:**

df.groupby('no\_of\_weekend\_nights')['Booking\_ID'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='bar',figsize=(9,9))

df.groupby('no\_of\_week\_nights')['Booking\_ID'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='barh',figsize=(9,9))

df.groupby('no\_of\_weekend\_nights')['no\_of\_week\_nights'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='pie',autopct='**%1.2f%%**',subplots=True,title='Weekend nights VS week nights',figsize=(9,9))

* 1. **Meal plans:**

df.groupby('type\_of\_meal\_plan')['Booking\_ID'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='pie',autopct='**%1.2f%%**',subplots=True,title='Meal',figsize=(9,9))

* 1. **Parking spaces:**

df.groupby('required\_car\_parking\_space')['Booking\_ID'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='pie',autopct='**%1.2f%%**',subplots=True,title='Required car parking space',figsize=(9,9))

* 1. **Room type:**

df.groupby('room\_type\_reserved')['Booking\_ID'].agg(['count']).sort\_values(by='count',ascending=False).plot(kind='barh',figsize=(9,9))

* 1. **Use “groupby” function to investigate the data further.**
  2. **Correlation and correlation matrix:**

df.corr()

sns.heatmap(df.corr())

* 1. **Colour change in heatmap:**
* sns.heatmap(df.corr(),cmap="BrBG")

**other type examples**: ‘hot’, ‘YlGnBu’, ‘jet’, 'gray', ‘inferno', ‘RdBu’, ‘HSV’, ‘Accent’, etc.