Task 6.1

Hotel Bookings Data Set

Data source:

This data came from Kaggle. The community consists of data scientists and developers. With over one million registered users, this is the largest data science community in the world. Members can discuss development models, study data sets or network across 194 separate countries around the world.

Data Set and content

This dataset is from DatAfrik and consists of hotel bookings between January 1st, 2010, to December 31st, 2019 from a company who help international travellers around the world secure hotel in destination countries. It includes many details about bookings, including origin country, origin states, destination country, destination states, age, gender, room type, check-in and check-out dates, number of room occupant, booking fees, taxes (GST), hotel rating and customers payment methods for the bookings.

Limitations & Ethics

Due to privacy limitations the original source of data (the company has not been revealed) DatAfrik , Yet looking at data set in detail and also the organisation which is an educational institute for data related courses and boot camps ; the data could be consider moderate reliable. Since The data set was mainly picked for portfolio analytical purpose it could be used and it is Important to run consistency check of data and clean if it is necessary.

Relevance

Data set meets all requirements of the achievement criteria.

Data profile

Variable		Profile		
	Time-variant / -invariant	**Structured / Unstructured**	**Qualitative / Quantitative**	**Qualitative: Nominal / Ordinal** **Quantitative: Discrete / Continuous**
Booking ID	Time invariant	Structured	Qualitative	Ordinal
Date of Booking	Time variant	Structured	Quantitative	Continuous
Time	Time variant	Structured	Quantitative	Continuous
Customer id	Time invariant	Structured	Qualitative	Ordinal
Gender	Time invariant	Structured	Qualitative	Nominal
Age	Time variant	Structured	Quantitative	Continuous
Origin country	Time invariant	Structured	Qualitative	Nominal
State	Time invariant	Structured	Qualitative	Nominal
Destination country	Time invariant	Structured	Qualitative	Nominal
Destination city	Time invariant	Structured	Qualitative	Nominal
No of people	Time invariant	Structured	Quantitative	Discrete
No of days	Time variant	Structured	Quantitative	Discrete
No of rooms	Time invariant	Structured	Quantitative	Discrete
Hotel name	Time invariant	Structured	Qualitative	Nominal
Hotel rating	Time invariant	Structured	Quantitative	Discrete
Payment mode	Time invariant	Structured	Qualitative	Nominal
Bank name	Time invariant	Structured	Qualitative	Nominal
Booking price_Sin_$	Time invariant	Structured	Quantitative	Discrete
Discount	Time invariant	Structured	Quantitative	Discrete
Profit margin	Time invariant	Structured	Quantitative	Discrete

Questions to explore.

1. Which countries have the highest hotel booking?
2. What´s the favourite destination country for different nations?
3. How does demographical elements impact hotel book?
4. How different hotels are preferred among customers from the region?
5. What are the durations of stays during trips?
6. Which customers have booked most recurring bookings and their origins?
7. What´s the relationship between hotel rating, price and booking frequency?