Travel Aggregator Analysis

Certification Project

Scenario

Travelling has significantly changed thanks to technology. In the earlier days, it was not easy to travel as booking journeys used to be a hassle, where people needed to contact travel agents, get prices and then do a lot of paperwork.

But, with the advent of online apps, it becomes easier to book tickets. However, one challenge is still there. Due to many available apps which offer bookings, there can be a lot of comparisons for getting the best price.

A new Indian start-up, "MyNextBooking" is an aggregator on top of the available top platforms (Yatra, MMT, Goibibo). It helps the user compare prices for their upcoming journeys on its platform and the customer is redirected to the desired platform after looking at the price.

However, they need to do a lot of analysis of the data which they have collected over a while.

As a newly hired analyst and proficiency in Python, your role is to help them perform an in-depth analysis.

Objective: Please provide answers to the following questions:

1) Find the number of distinct bookings, sessions, and searches from the given data sets.

2) How many sessions have more than one booking?

3) Which days of the week have the highest number of bookings? Also, draw a pie chart to show the distribution for all days of the week?

4) For each of the service names, display the total number of bookings and the total Gross Booking Value in INR.

5) For customers who have more than 1 booking, which is the most booked route (from_city to to_city)?

6) Which are the top 3 departure cities from where customers book mostly in advance, provided that there have been at least 5 departures from that city?

7) Plot a heatmap displaying correlations of the numerical column and report which pair of numerical columns in the bookings data set, have the maximum correlation?

8) For each service, which is the most used device type for making bookings on the platform?

9) Plot the trends at a quarterly frequency for the number of bookings by each of the device types, that is, plot a time series for each year and quarter showing the number of bookings performed by each device type.

10) Consider the following example:

12 customers performed a total of 100 searches but only a few of them performed 10 bookings in total.

Searches = 100 bookings = 10 customers = 12
The overall booking to search ratio, Ober, is 10/100 = 0.1 Using the above information/context, answer the following:

- What is the average oBSR for each month of the year?
- What is the average oBSR for each day of the week?
- Plot a time series of oBSR on all the given dates.

About the Data: The data provided consists of the following two files:

1) Bookings.csv
   - customer_id
   - booking_id
   - from_city

- from_country
- to_city
- to_country
- booking_time
- device_type_used
- INR_Amount
- service_name
- no_of_passengers
- days_to_departure
- distance_km

2) Sessions.csv
   - session_id
   - search_id
   - search_time
   - session_starting_time
   - booking_id