Greetings from my side.

First of all thank you for providing me the opportunity and finding me suitable for the role. I have tried my level best to do justice with the expectations.

I would like to give a brief overview of the task which I have done and its further extension which might be taken upon.

First, I shall share my understanding on the dataset.

The dataset contained information regarding the booking history of various users along with the specifications of their source and destination and vehicle booked alongside and the mode of booking.

Since there were three different types of travel\_type\_ids, thus the very first thing to strike my mind was that this problem needs to be approached in three steps- wherein in each step I tackle each type of travel types.

So I made 3 dataframes namely ‘long’, ’hourly’, ‘point’ and explored each of them separately.

I started off with **point to point** travel type.

Here most of the columns of to\_city\_id and to\_date were missing. That meant that we cannot compute the travel journey length and other specifications and to\_city\_id but looking from the data made me realise that in general point-to-point journeys were mostly intra-city and so both the above attributes were not much needed.

Using various data visualization tools, many inferences were drawn from questions that came to my mind and all of the same have been mentioned in the i-python notebook.   
The primal objective with which I was observing the dataset was considering myself a part of the company and what all aspects would I need to check in order to maximize my company’s working efficiency and profit.

The notebook in itself explains with figures and charts, various conclusions which the company can look into.

In **Long** travel type, destination longitude and latitudes were missing along with to\_area. But this was evident that people would generally use long travel when they need to move in between different cities and thus the questions to which I tried to find answers were in this context. I found some interesting patterns and facts which came from the dataset which I feel can help the company.

Similarly in **Hourly Package** travel type, we had different packages and thus I had to draw inferences in that approach so that I find most beneficial package and also in what condition, which package is generally chosen and other such questions.

I feel that in all the three, the conclusions (which have been listed in bold in the notebook below the charts or their numerical data proof) can actually help the company gain a lot of insight about its customers and can help them serve better and when this happens, automatically the company would start gaining more customers.

**Some limitations** on which I could not work upon were missing longitudes and latitudes data in some cases and I was unable to implement for the data present. One could have mapped the longitude and latitudes with area or city\_id using the relations between latitude and longitude and this could have further enhanced the observations.

Also, the price factor is very important. Had the prices been given, I could have tried to find some relations between prices and bookings and also this could have helped working more on the price model which in turn could amplify the company’s working and growings.

In the end, I’d like to point out that I have tried to make the attached ipython notebook self explanatory and all the conclusions are present in it.

Please go through the work and do notify how I could improve myself and this. I’d love to be part of your esteemed organization and work on these kinds of real life problems.

I really enjoyed working on this dataset.

Looking forward to working and learning from you.

Thanks and warm regards

Vipul Vaibhav