Inceptez Private Ltd.,

Let us now assume inceptez has started a new venture called Inceptez Private Ltd. Inceptez helps users find good properties in the least amount of time. Properties receive interactions on Inceptez Platform. One interaction is defined as one user requesting an owner contact on a property. A property can receive 0 to many interactions. This research will focus on studying and modelling the interactions received by properties. We are interested in studying and statistically modelling property interactions. At Inceptez, we are building the best-in-class Data Science, Machine Learning & Artificial Intelligence platform for bringing real value to our customers. Do you have what it takes to be a part of the best-in-class Data Scientist Team at Inceptez? Let's see!

Note: This is a real-life data set Brand name and others details are changed to Inceptez concerning privacy.

Part 1: Data Analysis

The Inceptez Data team has been tasked with predicting what it takes to identify whether a created property post would be Enquired about / Interacted with. To help explore this question, we have provided the attached csvs of the properties' performance.

We would like you to use this data set to help understand what factors are best at predicting Property clicks, and offer suggestions to operationalize those insights to help Inceptez.

See below for a detailed description of the dataset. Please include any code you wrote for the analysis. Please also call out any data related assumptions or issues that you encounter.

- Perform any merging, cleaning, exploratory analysis, and/or visualizations to use the provided data for this analysis

Part 2: Data Modelling

We are interested in studying and statistically modelling property interactions.

- We would like to have a predictive model that would say the number of interactions that a property would receive in a period of time. For simplicity let's say we would like to predict the number of interactions that a property would receive within 3 days of its activation and 7 days of its activation

However, this part is open ended and you could bring your own time intervals into the problem. This is the part of your artistry in data science. In the end we need to profess the number of interaction that a certain kind of property would receive within a given number of days.

Part 3: Presentation

We DO NOT look just at your final model and its performance, rather we look for the research mindset in you, your curiosity in data. Therefore, we urge you to present whatever you do with standards followed among the data science community. Keep an open-ended eye on the problem and feel free to approach the data in whatever way you think suits the problem.

- We urge you to try out different methodologies and present your results.

Data Sets

Unzip the resources.zip file to find the following 3 data sets:

a. Dataset.csv :

■ Properties data containing various features like activation_date, BHK type, locality, property size, property age, rent, apartment type etc.

■ activation_date is the date property got activated on Inceptez. Fields like lift, gym etc are binary valued - 1 indicating presence and 0 indicating absence. All other fields are self-explanatory.

■ You may use these along with the rest of the data sets to engineer the features that you would use in your study

b. Photos.tsv :

■ Data containing photo counts of properties

■ photo_urls column contains string values that you have to parse to obtain the number of photos uploaded on a property

■ Each value in the photo_url column is supposed to be a string representation of an array of json [in python terms a list of dictionaries] where each json object represents one image. However due to some unforeseen events, these values got corrupted and lost their valid json array representation. You could see this if you observe the data closely. Hint: There is a missing " before 'title' for the first json object in each value. There is also an additional " at the end of each value. Also you must remove all the \\ to get a valid json representation.

■ Your objective is to get the number of photos uploaded for a property. For this you should correct the corrupt string and make it a valid json. Once you have a valid json string, you can get the length of this array, which would be the number of photos uploaded on the property.

■ Also note that these are not images, but just names that we use to point to images.

You are NOT given the images nor do we expect you to have them. All that you are expected to do it get the number of photos on each property by cleaning up the corrupt invalid json array string.

■ NULL/NaN values indicate absence of photos on the property, ie; photo_count = 0

c. Interactions.csv :

■ Data containing the timestamps of interaction on the properties.

■ Each 'request_date' value represents the timestamp of a unique valid interaction on a property

(contact owner happened and a user received the owner contact phone number)

   ■ Therefore if you count the number of times each property has appeared in this table, it tells you the number of interaction received on this property

   ■ You will use this request_date along with the activation_date in our first table and other features in our study

Note: property_id is the unique property identifier in all the three data sets.

ALL THE BEST!