Data:

- The data shared has 8 lakh rows of sessions happened on UC customer app. from Dec 17,2018 to Jan 30,2019
- These are the sessions landed on the UC slots page. (This is the page where we ask customer to select booking time preference)
- Our hypothesis is conversion from slots page to request placed is dependent on 1) category 2) day of week 3) no of slots shown to customer

Column Definitions:

record_id	Session Id of Customer visit. All visits by customer in a 30 min slot in a category is one record_id
city_key	City of customer
event_time	Time of visit on slots page
weekday	Weekday of event time
category_key	category of customer
rptcatg	Supercategory (Supercategory can have multiple categories)
req_id	If the session has placed the request then request_id is present else blank
session_group	4 digits (d0_possible , d0_group, d1_group, d2_group) D0 - Same day as day of visit D1 - Next day from day of visit D2 - Next to next day from day of visit

Detailed explanation of session group:

Day is categorized into following slots

- 0-10 hours
- 10-12 hours
- 12-15 hours
- 15-18 hours
- 18-20 hours
- 20+ hours

To simplify further Slots shown in a day are counted as a variable with 4 possible values

- 0 (No slots shown)
- 1 (1 Slot shown)
- 2 (2 or 3 slots shown)
- 3 (4 or 5 or 6 slots shown)

Our understanding is if everything is same, all sessions of a given session group have similar conversion.

For Ex: session group is 2123 then (d0_possible=2 , d0_group=1 , d1_group=2, d2_group=3)

D0_possible = Slot configuration possible in D0 (possible values 0,1,2,3 as explained above). If we open all slots in D0 this is what we get
D0_group = Slot configuration shown for D0
D1_group = Slot configuration shown for D1
D2_group = slot configuration shown for D2

*Note : D1_possible and D2_possible are always 3.

So for this user if urbanclap showed all slots then possible configuration could be 2233

Objective:
Build a model using 3 weeks of historical data (Dec 17 to Jan 13) to predict what is the probability a given session would result in a request ?

Testing Criteria:
Use the model to predict the probability of request for test period (Jan 14 - Jan 20) sessions,

Test cases:

1. Total Orders predicted should be close to Total Actual orders (model should have more accuracy for super categories where more sessions are available)
2. Sessions can be classified based on predicted probability. For each segment, predicted requests to be close to actual requests (Eg : All sessions where expected probability between 20% and 25%. Actual conversion from these sessions also be in 20% and 25%)

Please share the final result with working code.

The dataset is confidential. Please do not share with anyone.