ML technique final project

Basic information

1. Training set = 91530 - 91530 reservation within 640 days
2. Testing set = 27859
3. Train label = 640 – ranking of 640 days

Thinking flow

1. Choose the characteristic of the data that we want – what kind of data information may affect the revenue? how many characteristics we want to choose?
2. About the thinking flow(1), thinking of the relationship between characteristics and the output may help
3. How to combine the features of every day
4. Testing or validate
5. How to Decide the algorithm structure

Step

1. Load the data v
2. Choose the features

Consider:

* Adr \* days(stay\_in\_week\_night) v
* arrival\_date\_year, arrival\_date\_month , arrival\_date\_day\_of\_month v
* is\_canceled? V

Feature:

* Hotel type v
* Total\_days v
* Adult v
* Child v
* Babies v
* Meals v

1. Load the label data & plot the relationship between y & the characteristic we choose
2. Process the data we choose
3. Decide the algorithm structure & Train
4. testing