

```
1 C:\Users\Frede\Anaconda3\python.exe "C:\Program Files\  
JetBrains\PyCharm Community Edition 2018.3.3\helpers\pydev  
\pydevconsole.py" --mode=client --port=53561  
2  
3 import sys; print('Python %s on %s' % (sys.version, sys.  
platform))  
4 sys.path.extend(['C:\\Users\\Frede\\Dropbox\\Master\\DM\\  
Assignments\\2\\DM2', 'C:\\Users\\Frede\\Dropbox\\Master\\  
DM\\Assignments\\2\\DM2\\Expedia-Ranking-Competition', 'C  
:/Users/Frede/Dropbox/Master/DM/Assignments/2/DM2'])  
5  
6 Python 3.7.3 (default, Mar 27 2019, 17:13:21) [MSC v.1915  
64 bit (AMD64)]  
7 Type 'copyright', 'credits' or 'license' for more  
information  
8 IPython 7.4.0 -- An enhanced Interactive Python. Type '?'  
for help.  
9 PyDev console: using IPython 7.4.0  
10  
11 Python 3.7.3 (default, Mar 27 2019, 17:13:21) [MSC v.1915  
64 bit (AMD64)] on win32  
12 In[2]: import numpy as np  
13 ....: import pandas as pd  
14 ....: from sklearn.feature_selection import RFE  
15 ....: from sklearn.ensemble import RandomForestRegressor  
16 ....: from sklearn.ensemble import RandomForestClassifier  
17 ....: from sklearn.linear_model import LogisticRegression  
18 ....: from model_test import impute_na, get_sample,  
split_train_test  
19 ....: from feature_engineering import  
extract_train_features  
20 ....: from scoring import score_prediction  
21 ....:  
22 ....:  
23 ....: def feature_selection(data, estimator, n_features=  
None):  
24 ....:  
25 ....:     X_train = data.drop(columns=["target", "  
booking_bool", "click_bool", "position"])  
26 ....:     y_train = data["target"]  
27 ....:     selector = RFE(estimator=estimator,  
n_features_to_select=n_features)  
28 ....:     selector.fit(X_train, y_train)  
29 ....:     cols = selector.support_  
30 ....:     print(X_train.loc[:, cols].columns)
```

```
31  ....:
32  ....:
33  ....: def decreasing_features_select(data, estimator,
34  ....:     target):
35  ....:     train, test = split_train_test(data, split=4)
36  ....:
37  ....:     X_train = train.drop(columns=["target", "booking_bool", "click_bool", "position", "random_bool"])
38  ....:     y_train = train["target"]
39  ....:     X_test = test.drop(columns=["target", "booking_bool", "click_bool", "position", "random_bool"])
40  ....:     y_test = test[["target", "srch_id", "prop_id", "booking_bool", "click_bool"]]
41  ....:
42  ....:     n_features = len(X_train.columns)
43  ....:     for n in range(n_features, 10, -2):
44  ....:         print("#####")
45  ....:         print(f"Number of features used: {n}.")
46  ....:         selector = RFE(estimator=estimator,
47  ....:             n_features_to_select=n)
48  ....:         selector.fit(X_train, y_train)
49  ....:         cols = selector.support_
50  ....:         print("Features used: ")
51  ....:         print(X_train.loc[:, cols].columns)
52  ....:         reduced_train = X_train.loc[:, cols]
53  ....:         estimator.fit(reduced_train, y_train)
54  ....:         reduced_test = X_test.loc[:, cols]
55  ....:         pred = clf_to_predict(estimator,
56  ....:             reduced_test, target)
57  ....:         score_prediction(pred, y_test, to_print=True)
58  ....:
59  ....: def clf_to_predict(estimator, X_test, target):
60  ....:     if target == "book":
61  ....:         prediction = estimator.predict_proba(X_test)[:, 1]
62  ....:     elif target == "score":
63  ....:         predict_array = estimator.predict_proba(X_test)
64  ....:         predict_array[:, 2] = predict_array[:, 2] * pred_weight
65  ....:         prediction = predict_array[:, [1, 2]].sum(axis=1)
```

```
65     ...:         else:
66     ...:             print("ERROR. no using classification with
67     ...:                 score_rank!")
68     ...:
69     ...:             return prediction
70     ...:
71 Backend Qt5Agg is interactive backend. Turning
72 interactive mode on.
73 In[3]: pd.options.mode.chained_assignment = None
74     ...: data = pd.read_csv("C:/Users/Frede/Dropbox/Master/
75     ...: DM/Assignments/2/DM2/final_training_fixed_data.csv")
76     ...: impute_na(data)
77     ...: sample = get_sample(data=data, size=0.1)
78     ...: estimator = RandomForestClassifier(n_estimators=
79     ...: 250, n_jobs=-1)
80     ...:
81     ...: targets = ["book", "score"]
82     ...: max_rank = 10
83     ...: pred_weight = 3
84     ...:
85     ...: for target in targets:
86     ...:     print(f"\nCURRENT CONFIGURATION")
87     ...:     print
88     ...:     ("#####
89     ...:      Target = {target}")
90     ...:     print(f"Max_rank = {max_rank}")
91     ...:     print(f"Pred_weight = {pred_weight}")
92     ...:
93     ...:     extract_train_features(data=sample, target=
94     ...:         target, max_rank=max_rank)
95     ...:     decreasing_features_select(data=sample,
96     ...:         estimator=estimator, target=target)
97     ...:
98     ...:
```

```

98 #####
99 #####
100 Number of features used: 64.
101 Features used:
102 Index(['srch_id', 'site_id', 'visitor_location_country_id',
103         'visitor_hist_starrating', 'visitor_hist_adr_usd',
104         'prop_country_id',
105         'prop_id', 'prop_starrating', 'prop_review_score',
106         'prop_brand_bool',
107         'prop_location_score1', 'prop_location_score2',
108         'prop_log_historical_price', 'price_usd', 'promotion_flag',
109         'srch_destination_id', 'srch_length_of_stay',
110         'srch_booking_window',
111         'srch_adults_count', 'srch_children_count',
112         'srch_room_count',
113         'srch_saturday_night_bool',
114         'srch_query_affinity_score',
115         'orig_destination_distance', 'gross_bookings_usd',
116         'overall_comp',
117         'price_diff', 'star_diff', 'country_average_price',
118         'srch_average_price', 'srch_average_loc1',
119         'srch_average_loc2',
120         'srch_average_prop_review_score',
121         'country_average_star',
122         'srch_average_star', 'country_diff_price',
123         'srch_diff_price',
124         'country_diff_star', 'srch_diff_star',
125         'srch_diff_locscore1',
126         'srch_diff_locscore2',
127         'srch_diff_prop_review_score',
128         'average_num_country', 'average_num_srch',
129         'average_num_prop',
130         'median_num_country', 'median_num_srch',
131         'median_num_prop',
132         'std_num_country', 'std_num_srch', 'std_num_prop',
133         'new_customer',
134         'guest_count', 'children_bool',
135         'norm_srch_diff_price',
136         'norm_srch_diff_star', 'norm_srch_diff_locscore1',
137         'norm_srch_diff_locscore2',
138         'norm_srch_diff_prop_review_score',
139         'norm_country_diff_price', 'norm_country_diff_star'

```

```
123 ',  
124     'norm_price_usd_srch_id', 'norm_price_usd_prop_id  
,  
125     'norm_price_usd_prop_country_id'],  
126     dtype='object')  
127 Converting into submission format...  
128 Calling scoring function...  
129 Matching true booking_bool/click_bool values to  
    submission...  
130 Calculating NDCG_5 score...  
131 Finished!  
132 NDCG_5 prediction score for this prediction: 0.  
    7862947991136365.  
133 #####  
134 Number of features used: 62.  
135 Features used:  
136 Index(['srch_id', 'site_id', 'visitor_location_country_id  
,  
137     'visitor_hist_starrating', 'visitor_hist_adr_usd  
, 'prop_country_id',  
138     'prop_id', 'prop_starrating', 'prop_review_score  
, 'prop_brand_bool',  
139     'prop_location_score1', 'prop_location_score2',  
140     'prop_log_historical_price', 'price_usd', '  
    promotion_flag',  
141     'srch_destination_id', 'srch_length_of_stay', '  
    srch_booking_window',  
142     'srch_adults_count', 'srch_children_count', '  
    srch_room_count',  
143     'srch_saturday_night_bool', '  
    srch_query_affinity_score',  
144     'orig_destination_distance', 'gross_bookings_usd  
, 'overall_comp',  
145     'price_diff', 'star_diff', 'country_average_price  
,  
146     'srch_average_price', 'srch_average_loc1', '  
    srch_average_loc2',  
147     'srch_average_prop_review_score', '  
    country_average_star',  
148     'srch_average_star', 'country_diff_price', '  
    srch_diff_price',  
149     'country_diff_star', 'srch_diff_star', '  
    srch_diff_locscore1',  
150     'srch_diff_locscore2', '  
    srch_diff_prop_review_score',
```

```
151      'average_num_country', 'average_num_srch', '  
152      average_num_prop',  
153      'median_num_country', 'median_num_srch', '  
154      median_num_prop',  
155      'std_num_country', 'std_num_srch', 'std_num_prop  
, 'guest_count',  
156      'children_bool', 'norm_srch_diff_price', '  
157      norm_srch_diff_locscore1',  
158      'norm_srch_diff_locscore2', '  
159      norm_srch_diff_prop_review_score',  
160      'norm_country_diff_price', 'norm_country_diff_star  
,  
161      'norm_price_usd_srch_id', 'norm_price_usd_prop_id  
,  
162      'norm_price_usd_prop_country_id'],  
163      dtype='object')  
160 Converting into submission format...  
161 Calling scoring function...  
162 Matching true booking_bool/click_bool values to  
    submission...  
163 Calculating NDCG_5 score...  
164 Finished!  
165 NDCG_5 prediction score for this prediction: 0.  
    7858054523328822.  
166 #####  
167 Number of features used: 60.  
168 Features used:  
169 Index(['srch_id', 'site_id', 'visitor_location_country_id  
,  
170      'visitor_hist_starrating', 'prop_country_id', '  
    prop_id',  
171      'prop_starrating', 'prop_review_score', '  
    prop_brand_bool',  
172      'prop_location_score1', 'prop_location_score2',  
173      'prop_log_historical_price', 'price_usd', '  
    promotion_flag',  
174      'srch_destination_id', 'srch_length_of_stay', '  
    srch_booking_window',  
175      'srch_adults_count', 'srch_children_count', '  
    srch_room_count',  
176      'srch_saturday_night_bool', '  
    srch_query_affinity_score',  
177      'orig_destination_distance', 'gross_bookings_usd  
, 'overall_comp',  
178      'price_diff', 'star_diff', 'country_average_price
```

```

178 ',  

179      'srch_average_price', 'srch_average_loc1', '  

    srch_average_loc2',  

180      'srch_average_prop_review_score', '  

    country_average_star',  

181      'srch_average_star', 'country_diff_price', '  

    srch_diff_price',  

182      'country_diff_star', 'srch_diff_star', '  

    srch_diff_locscore1',  

183      'srch_diff_locscore2', '  

    srch_diff_prop_review_score',  

184      'average_num_country', 'average_num_srch', '  

    average_num_prop',  

185      'median_num_country', 'median_num_srch', '  

    median_num_prop',  

186      'std_num_country', 'std_num_srch', 'std_num_prop  

', 'guest_count',  

187      'children_bool', 'norm_srch_diff_price', '  

    norm_srch_diff_locscore1',  

188      'norm_srch_diff_locscore2', '  

    norm_srch_diff_prop_review_score',  

189      'norm_country_diff_price', 'norm_price_usd_srch_id  

',  

190      'norm_price_usd_prop_id', '  

    norm_price_usd_prop_country_id'],  

191      dtype='object')  

192 Converting into submission format...  

193 Calling scoring function...  

194 Matching true booking_bool/click_bool values to  

    submission...  

195 Calculating NDCG_5 score...  

196 Finished!  

197 NDCG_5 prediction score for this prediction: 0.  

    7856280250356192.  

198 #####  

199 Number of features used: 58.  

200 Features used:  

201 Index(['srch_id', 'site_id', 'visitor_location_country_id  

', 'prop_country_id',  

202      'prop_id', 'prop_starrating', 'prop_review_score  

', 'prop_brand_bool',  

203      'prop_location_score1', 'prop_location_score2',  

204      'prop_log_historical_price', 'price_usd', '  

    promotion_flag',  

205      'srch_destination_id', 'srch_length_of_stay', '

```

```

205     'srch_booking_window',
206         'srch_adults_count', 'srch_children_count', ' '
207         'srch_room_count',
208         'srch_saturday_night_bool', ' '
209         'srch_query_affinity_score',
210         'orig_destination_distance', 'gross_bookings_usd'
211         ', 'overall_comp',
212         'price_diff', 'star_diff', 'country_average_price
213         ',
214         'srch_average_price', 'srch_average_loc1', ' '
215         'srch_average_loc2',
216         'srch_average_prop_review_score', ' '
217         'country_average_star',
218         'srch_average_star', 'country_diff_price', ' '
219         'srch_diff_price',
220         'srch_diff_locscore1',
221         'srch_diff_locscore2', ' '
222         'srch_diff_prop_review_score',
223         'average_num_country', 'average_num_srch', ' '
224         'average_num_prop',
225         'median_num_country', 'median_num_srch', ' '
226         'median_num_prop',
227         'std_num_country', 'std_num_srch', 'std_num_prop
228         ', 'guest_count',
229         'norm_srch_diff_price', 'norm_srch_diff_locscore1
230         ',
231         'norm_srch_diff_locscore2', ' '
232         'norm_srch_diff_prop_review_score',
233         'norm_country_diff_price', 'norm_price_usd_srch_id
234         ',
235         'norm_price_usd_prop_id', ' '
236         'norm_price_usd_prop_country_id'],
237         dtype='object')
238 Converting into submission format...
239 Calling scoring function...
240 Matching true booking_bool/click_bool values to
241 submission...
242 Calculating NDCG_5 score...
243 Finished!
244 NDCG_5 prediction score for this prediction: 0.
245 7876142625309993.
246 #####
247 Number of features used: 56.
248 Features used:

```

```

232 Index(['srch_id', 'site_id', 'visitor_location_country_id',
233     'prop_country_id',
234     'prop_id', 'prop_starrating', 'prop_review_score',
235     'prop_location_score1', 'prop_location_score2',
236     'prop_log_historical_price', 'price_usd', 'promotion_flag',
237     'srch_destination_id', 'srch_length_of_stay', 'srch_booking_window',
238     'srch_adults_count', 'srch_children_count', 'srch_room_count',
239     'srch_query_affinity_score', 'orig_destination_distance',
240     'gross_bookings_usd', 'overall_comp', 'price_diff',
241     'star_diff',
242     'country_average_price', 'srch_average_price', 'srch_average_loc1',
243     'srch_average_loc2', 'srch_average_prop_review_score',
244     'country_average_star', 'srch_average_star', 'country_diff_price',
245     'srch_diff_price', 'country_diff_star', 'srch_diff_star',
246     'srch_diff_locscore1', 'srch_diff_locscore2',
247     'srch_diff_prop_review_score', 'average_num_country',
248     'average_num_srch', 'average_num_prop', 'median_num_country',
249     'median_num_srch', 'median_num_prop', 'std_num_country',
250     'std_num_srch', 'std_num_prop', 'guest_count', 'norm_srch_diff_price',
251     'norm_srch_diff_locscore1', 'norm_srch_diff_locscore2',
252     'norm_srch_diff_prop_review_score', 'norm_country_diff_price',
253     'norm_price_usd_srch_id', 'norm_price_usd_prop_id',
254     'norm_price_usd_prop_country_id'],
255     dtype='object')
256 Converting into submission format...
257 Calling scoring function...
258 Matching true booking_bool/click_bool values to
259 submission...
260 Calculating NDCG_5 score...
261 Finished!

```

```
259 NDCG_5 prediction score for this prediction: 0.  
7882638979783646.  
260 #####  
261 Number of features used: 54.  
262 Features used:  
263 Index(['srch_id', 'site_id', 'visitor_location_country_id',  
        'prop_country_id',  
        'prop_id', 'prop_starrating', 'prop_review_score',  
        'prop_location_score1', 'prop_location_score2',  
        'prop_log_historical_price', 'price_usd', '  
promotion_flag',  
        'srch_destination_id', 'srch_length_of_stay', '  
srch_booking_window',  
        'srch_adults_count', 'srch_query_affinity_score',  
        'orig_destination_distance', 'gross_bookings_usd  
', 'overall_comp',  
        'price_diff', 'star_diff', 'country_average_price  
' ,  
        'srch_average_price', 'srch_average_loc1', '  
srch_average_loc2',  
        'srch_average_prop_review_score', '  
country_average_star',  
        'srch_average_star', 'country_diff_price', '  
srch_diff_price',  
        'country_diff_star', 'srch_diff_star', '  
srch_diff_locscore1',  
        'srch_diff_locscore2', '  
srch_diff_prop_review_score',  
        'average_num_country', 'average_num_srch', '  
average_num_prop',  
        'median_num_country', 'median_num_srch', '  
median_num_prop',  
        'std_num_country', 'std_num_srch', 'std_num_prop  
' , 'guest_count',  
        'norm_srch_diff_price', 'norm_srch_diff_locscore1  
' ,  
        'norm_srch_diff_locscore2', '  
norm_srch_diff_prop_review_score',  
        'norm_country_diff_price', 'norm_price_usd_srch_id  
' ,  
        'norm_price_usd_prop_id', '  
norm_price_usd_prop_country_id'],  
283      dtype='object')  
284 Converting into submission format...  
285 Calling scoring function...
```

```
286 Matching true booking_bool/click_bool values to
     submission...
287 Calculating NDCG_5 score...
288 Finished!
289 NDCG_5 prediction score for this prediction: 0.
    7852026460546554.
290 #####
291 Number of features used: 52.
292 Features used:
293 Index(['srch_id', 'site_id', 'visitor_location_country_id',
     'prop_country_id',
     'prop_id', 'prop_starrating', 'prop_review_score',
     'prop_location_score1', 'prop_location_score2',
     'prop_log_historical_price', 'price_usd', 'promotion_flag',
     'srch_destination_id', 'srch_length_of_stay', 'srch_booking_window',
     'srch_adults_count', 'orig_destination_distance',
     'gross_bookings_usd',
     'overall_comp', 'price_diff', 'star_diff', 'country_average_price',
     'srch_average_price', 'srch_average_loc1', 'srch_average_loc2',
     'srch_average_prop_review_score', 'srch_average_star',
     'country_diff_price', 'srch_diff_price', 'country_diff_star',
     'srch_diff_star', 'srch_diff_locscore1', 'srch_diff_locscore2',
     'srch_diff_prop_review_score', 'average_num_country',
     'average_num_srch', 'average_num_prop', 'median_num_country',
     'median_num_srch', 'median_num_prop', 'std_num_country', 'std_num_srch',
     'std_num_prop', 'guest_count', 'norm_srch_diff_price',
     'norm_srch_diff_locscore1', 'norm_srch_diff_locscore2',
     'norm_srch_diff_prop_review_score', 'norm_country_diff_price',
     'norm_price_usd_srch_id', 'norm_price_usd_prop_id',
     '',
     'norm_price_usd_prop_country_id'],
     dtype='object')
```

```
313 Converting into submission format...
314 Calling scoring function...
315 Matching true booking_bool/click_bool values to
    submission...
316 Calculating NDCG_5 score...
317 Finished!
318 NDCG_5 prediction score for this prediction: 0.
    7888108171837258.
319 #####
320 Number of features used: 50.
321 Features used:
322 Index(['srch_id', 'site_id', 'visitor_location_country_id',
    'prop_country_id',
    'prop_id', 'prop_starrating', 'prop_review_score',
    'prop_location_score1', 'prop_location_score2',
    'prop_log_historical_price', 'price_usd', 'promotion_flag',
    'srch_destination_id', 'srch_length_of_stay', 'srch_booking_window',
    'srch_adults_count', 'orig_destination_distance',
    'gross_bookings_usd',
    'overall_comp', 'price_diff', 'star_diff', 'country_average_price',
    'srch_average_price', 'srch_average_loc1', 'srch_average_loc2',
    'srch_average_prop_review_score', 'srch_average_star',
    'country_diff_price', 'srch_diff_price', 'country_diff_star',
    'srch_diff_star', 'srch_diff_locscore1', 'srch_diff_locscore2',
    'srch_diff_prop_review_score', 'average_num_country',
    'average_num_srch', 'average_num_prop', 'median_num_srch',
    'median_num_prop', 'std_num_srch', 'std_num_prop',
    'guest_count',
    'norm_srch_diff_price', 'norm_srch_diff_locscore1',
    'norm_srch_diff_locscore2',
    'norm_srch_diff_prop_review_score',
    'norm_country_diff_price', 'norm_price_usd_srch_id',
    'norm_price_usd_prop_id',
    'norm_price_usd_prop_country_id'],
```

```
340         dtype='object')
341 Converting into submission format...
342 Calling scoring function...
343 Matching true booking_bool/click_bool values to
    submission...
344 Calculating NDCG_5 score...
345 Finished!
346 NDCG_5 prediction score for this prediction: 0.
    7890685818012965.
347 #####
348 Number of features used: 48.
349 Features used:
350 Index(['srch_id', 'visitor_location_country_id', 'prop_country_id', 'prop_id',
351         'prop_starrating', 'prop_review_score', 'prop_location_score1',
352         'prop_location_score2', 'prop_log_historical_price',
353         'price_usd',
354         'promotion_flag', 'srch_destination_id', 'srch_length_of_stay',
355         'srch_booking_window', 'orig_destination_distance',
356         'gross_bookings_usd', 'overall_comp', 'price_diff',
357         'star_diff',
358         'srch_average_price', 'srch_average_loc1', 'srch_average_loc2',
359         'srch_average_prop_review_score', 'srch_average_star',
360         'country_diff_price', 'srch_diff_price', 'country_diff_star',
361         'srch_diff_star', 'srch_diff_locscore1', 'srch_diff_locscore2',
362         'srch_diff_prop_review_score', 'average_num_country',
363         'average_num_srch', 'average_num_prop', 'median_num_srch',
364         'median_num_prop', 'std_num_country', 'std_num_srch', 'std_num_prop',
365         'guest_count', 'norm_srch_diff_price', 'norm_srch_diff_locscore1',
366         'norm_srch_diff_locscore2', 'norm_srch_diff_prop_review_score',
367         'norm_country_diff_price', 'norm_price_usd_srch_id',
368         'norm_price_usd_prop_id', '
```

```
366 norm_price_usd_prop_country_id'],
367         dtype='object')
368 Converting into submission format...
369 Calling scoring function...
370 Matching true booking_bool/click_bool values to
371 submission...
372 Calculating NDCG_5 score...
373 Finished!
374 NDCG_5 prediction score for this prediction: 0.
375 7875603244637376.
376 #####
377 Number of features used: 46.
378 Features used:
379 Index(['srch_id', 'visitor_location_country_id', 'prop_id',
380         'prop_starrating',
381         'prop_review_score', 'prop_location_score1', 'prop_location_score2',
382         'prop_log_historical_price', 'price_usd', 'promotion_flag',
383         'srch_destination_id', 'srch_length_of_stay', 'srch_booking_window',
384         'orig_destination_distance', 'gross_bookings_usd',
385         'overall_comp',
386         'price_diff', 'star_diff', 'srch_average_price',
387         'srch_average_loc1',
388         'srch_average_loc2', 'srch_average_prop_review_score',
389         'srch_average_star', 'country_diff_price',
390         'srch_diff_price',
391         'country_diff_star', 'srch_diff_star', 'srch_diff_locscore1',
392         'srch_diff_locscore2',
393         'srch_diff_prop_review_score',
394         'average_num_srch', 'average_num_prop',
395         'median_num_srch',
396         'median_num_prop', 'std_num_country',
397         'std_num_srch', 'std_num_prop',
398         'guest_count', 'norm_srch_diff_price',
399         'norm_srch_diff_locscore1',
400         'norm_srch_diff_locscore2',
401         'norm_srch_diff_prop_review_score',
402         'norm_country_diff_price', 'norm_price_usd_srch_id',
403         '',
404         'norm_price_usd_prop_id',
405         'norm_price_usd_prop_country_id'],
```

```
393         dtype='object')
394 Converting into submission format...
395 Calling scoring function...
396 Matching true booking_bool/click_bool values to
    submission...
397 Calculating NDCG_5 score...
398 Finished!
399 NDCG_5 prediction score for this prediction: 0.
    7875293442058918.
400 #####
401 Number of features used: 44.
402 Features used:
403 Index(['srch_id', 'visitor_location_country_id', 'prop_id',
    'prop_starrating',
404         'prop_review_score', 'prop_location_score1', 'prop_location_score2',
405         'prop_log_historical_price', 'price_usd', 'srch_destination_id',
406         'srch_length_of_stay', 'srch_booking_window',
407         'orig_destination_distance', 'gross_bookings_usd',
    'overall_comp',
408         'price_diff', 'star_diff', 'srch_average_price',
    'srch_average_loc1',
409         'srch_average_loc2', 'srch_average_prop_review_score',
410         'srch_average_star', 'country_diff_price',
    'srch_diff_price',
411         'country_diff_star', 'srch_diff_star', 'srch_diff_locscore1',
412         'srch_diff_locscore2', 'srch_diff_prop_review_score',
413         'average_num_srch', 'average_num_prop',
    median_num_srch',
414         'median_num_prop', 'std_num_srch', 'std_num_prop',
    'guest_count',
415         'norm_srch_diff_price', 'norm_srch_diff_locscore1',
    ,
416         'norm_srch_diff_locscore2', 'norm_srch_diff_prop_review_score',
417         'norm_country_diff_price', 'norm_price_usd_srch_id',
    ,
418         'norm_price_usd_prop_id', 'norm_price_usd_prop_country_id'],
419         dtype='object')
420 Converting into submission format...
```

```
421 Calling scoring function...
422 Matching true booking_bool/click_bool values to
    submission...
423 Calculating NDCG_5 score...
424 Finished!
425 NDCG_5 prediction score for this prediction: 0.
    7860293930122411.
426 #####
427 Number of features used: 42.
428 Features used:
429 Index(['srch_id', 'prop_id', 'prop_review_score', 'prop_location_score1',
430         'prop_location_score2', 'prop_log_historical_price',
431         'price_usd',
432         'srch_destination_id', 'srch_length_of_stay', 'srch_booking_window',
433         'orig_destination_distance', 'gross_bookings_usd',
434         'overall_comp',
435         'price_diff', 'star_diff', 'srch_average_price',
436         'srch_average_loc1',
437         'srch_average_loc2',
438         'srch_average_prop_review_score',
439         'srch_average_star', 'country_diff_price',
440         'srch_diff_price',
441         'country_diff_star', 'srch_diff_star',
442         'srch_diff_locscore1',
443         'srch_diff_locscore2',
444         'srch_diff_prop_review_score',
445         'average_num_srch', 'average_num_prop',
446         'median_num_srch',
447         'median_num_prop', 'std_num_srch', 'std_num_prop',
448         'guest_count',
449         'norm_srch_diff_price', 'norm_srch_diff_locscore1',
450         '',
451         'norm_srch_diff_locscore2',
452         'norm_srch_diff_prop_review_score',
453         'norm_country_diff_price', 'norm_price_usd_srch_id',
454         '',
455         'norm_price_usd_prop_id',
456         'norm_price_usd_prop_country_id'],
457         dtype='object')
458 Converting into submission format...
459 Calling scoring function...
460 Matching true booking_bool/click_bool values to
    submission...
```

```
448 Calculating NDCG_5 score...
449 Finished!
450 NDCG_5 prediction score for this prediction: 0.
    7854832443091901.
451 #####
452 Number of features used: 40.
453 Features used:
454 Index(['srch_id', 'prop_id', 'prop_starrating', 'prop_review_score',
455         'prop_location_score1', 'prop_location_score2',
456         'prop_log_historical_price', 'price_usd', 'srch_destination_id',
457         'srch_length_of_stay', 'srch_booking_window',
458         'orig_destination_distance', 'gross_bookings_usd',
459         'price_diff',
460         'srch_average_price', 'srch_average_loc1',
        'srch_average_loc2',
461         'srch_average_prop_review_score', 'srch_average_star',
462         'country_diff_price', 'srch_diff_price',
        'country_diff_star',
463         'srch_diff_star', 'srch_diff_locscore1',
        'srch_diff_locscore2',
464         'srch_diff_prop_review_score', 'average_num_srch',
        'average_num_prop',
465         'median_num_srch', 'median_num_prop',
        'std_num_srch', 'std_num_prop',
466         'norm_srch_diff_price', 'norm_srch_diff_locscore1',
        ,
467         'norm_srch_diff_locscore2',
        'norm_srch_diff_prop_review_score',
468         'norm_country_diff_price', 'norm_price_usd_srch_id',
        ,
469         'norm_price_usd_prop_id',
        'norm_price_usd_prop_country_id'],
470         dtype='object')
470 Converting into submission format...
471 Calling scoring function...
472 Matching true booking_bool/click_bool values to
        submission...
473 Calculating NDCG_5 score...
474 Finished!
475 NDCG_5 prediction score for this prediction: 0.
    79111726573061.
476 #####
```

```
477 Number of features used: 38.
478 Features used:
479 Index(['srch_id', 'prop_id', 'prop_review_score', 'prop_location_score1',
480         'prop_location_score2', 'prop_log_historical_price
481         ', 'price_usd',
482         'srch_destination_id', 'srch_length_of_stay', 'srch_booking_window',
483         'orig_destination_distance', 'gross_bookings_usd
484         ', 'price_diff',
485         'srch_average_price', 'srch_average_loc1', 'srch_average_loc2',
486         'srch_average_prop_review_score', 'srch_average_star',
487         'country_diff_price', 'srch_diff_price', 'srch_diff_star',
488         'srch_diff_locscore1', 'srch_diff_locscore2',
489         'srch_diff_prop_review_score', 'average_num_srch
490         ', 'average_num_prop',
491         'median_num_srch', 'median_num_prop', 'std_num_srch', 'std_num_prop',
492         'norm_srch_diff_price', 'norm_srch_diff_locscore1
493         ',
494         'norm_srch_diff_locscore2', 'norm_srch_diff_prop_review_score',
495         'norm_country_diff_price', 'norm_price_usd_srch_id
496         ',
497         'norm_price_usd_prop_id', 'norm_price_usd_prop_country_id'],
498         dtype='object')
499 Converting into submission format...
500 Calculating NDCG_5 score...
501 Finished!
502 NDCG_5 prediction score for this prediction: 0.
503 7843264926592507.
504 #####
505 Number of features used: 36.
506 Features used:
507 Index(['srch_id', 'prop_id', 'prop_review_score', 'prop_location_score1',
508         'prop_location_score2', 'prop_log_historical_price
509         ', 'price_usd',
```

```
505      'srch_destination_id', 'srch_booking_window',
506      'orig_destination_distance', 'gross_bookings_usd
', 'price_diff',
507      'srch_average_price', 'srch_average_loc1', '
srch_average_loc2',
508      'srch_average_prop_review_score', '
country_diff_price',
509      'srch_diff_price', 'srch_diff_star', '
srch_diff_locscore1',
510      'srch_diff_locscore2', '
srch_diff_prop_review_score',
511      'average_num_srch', 'average_num_prop', '
median_num_srch',
512      'median_num_prop', 'std_num_srch', 'std_num_prop',
513      'norm_srch_diff_price', 'norm_srch_diff_locscore1
',
514      'norm_srch_diff_locscore2', '
norm_srch_diff_prop_review_score',
515      'norm_country_diff_price', 'norm_price_usd_srch_id
',
516      'norm_price_usd_prop_id', '
norm_price_usd_prop_country_id'],
517      dtype='object')
518 Converting into submission format...
519 Calling scoring function...
520 Matching true booking_bool/click_bool values to
submission...
521 Calculating NDCG_5 score...
522 Finished!
523 NDCG_5 prediction score for this prediction: 0.
7887383345552681.
524 ##########
525 Number of features used: 34.
526
```