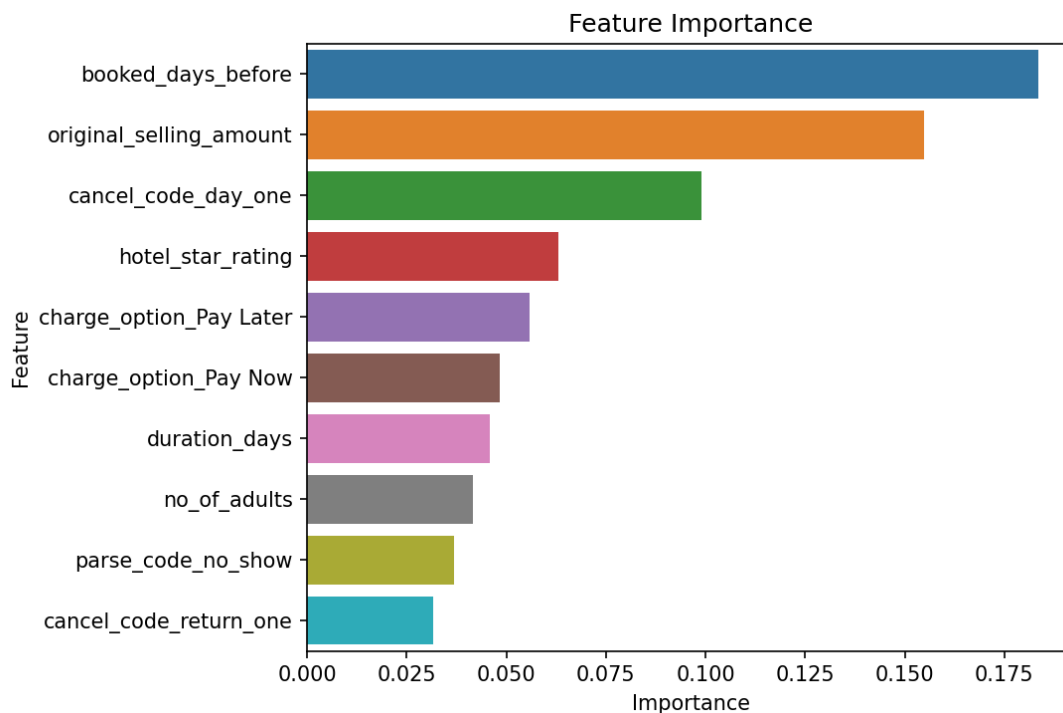


## Task 3: Churn prediction Model

ה-features הכי חשובים כדי לחזות ביטולים:

באמצעות שימוש ב-Random Forest על הדאטה לאחר processing, מצאנו ש-10 ה-features הכי חשובים הם:

```
clf = RandomForestClassifier(n_estimators=100, random_state=0)
clf.fit(X, y)
importances = clf.feature_importances_
```



בשלב הראשון עברנו על ה-features, ובדקנו קורלציה עם ביטולי הזמנות. מקבלים תוצאות דומות:

cancellation	1.000000
booked_days_before	0.309535
duration_days	0.137351
original_selling_amount	0.118875
is_user_logged_in	0.094142
request_largebed	-0.075190
no_of_children	0.053339
guest_is_not_the_customer	-0.046645
request_highfloor	0.043517

לפי ערכי הקורולציה הגבוהים ביותר בערך מוחלט בחרנו את המאפיינים שאיתם נרצה לאמן את המודל, והם הערכים שיעזרו לנו לחזות האם ההזמנה תבוטל או לא בצורה הטובה ביותר.

כדי לא לחזור על עצמנו ולהוסיף בשנית את אותם גרפים, נציין כי בקובץ `project.pdf` מצורפים גרפים עבור המשתנים `is_user_logged_in` ו-`charge_option`. ובקובץ שאלה 4 מופיעים גרפים עבור `duration_days` ו-`booked_days_before`.