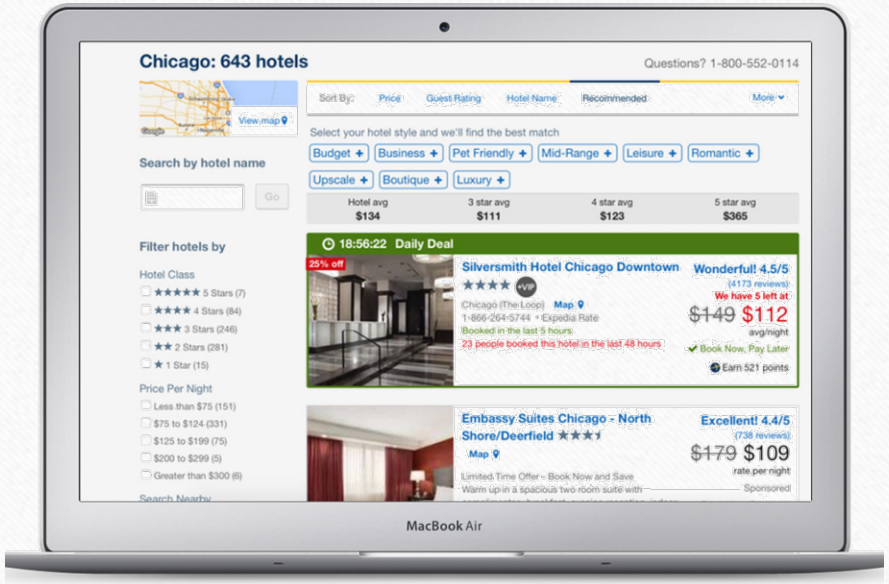
Predicting Conversion of a search to a booking by a travel enthusiast

By Vamsi Krishna Teegavarapu

Credits: Kaggle for the data set , SRI KANAJAN for the incredible support

Hypothesis

- The Hypothesis for the problem at hand is that we would be able to predict based on conversion of a search into booking by using search parameters.



DATA Exploration

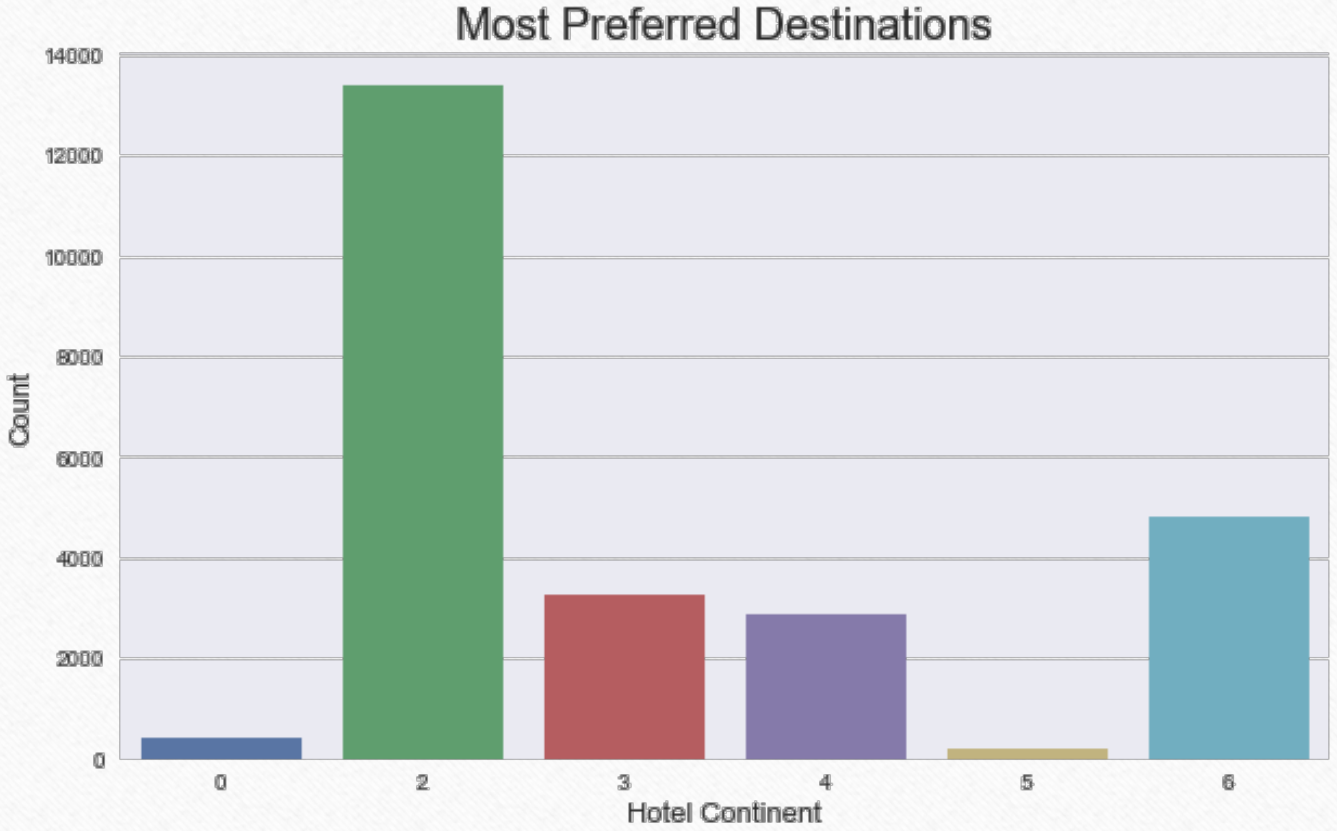
1.1

Quick view of the features available in the dataset

Column name	Description	Data type
date_time	Timestamp	string
site_name	ID of the Expedia point of sale (i.e. Expedia.com, Expedia.co.uk, Expedia.co.jp, ...)	int
posa_continent	ID of continent associated with site_name	int
user_location_country	The ID of the country the customer is located	int
user_location_region	The ID of the region the customer is located	int
user_location_city	The ID of the city the customer is located	int
orig_destination_distance	Physical distance between a hotel and a customer at the time of search. A null means the distance could not be calculated	double
user_id	ID of user	int
is_mobile	1 when a user connected from a mobile device, 0 otherwise	tinyint
is_package	1 if the click/booking was generated as a part of a package (i.e. combined with a flight), 0 otherwise	int
channel	ID of a marketing channel	int
srch_ci	Checkin date	string
srch_co	Checkout date	string
srch_adults_cnt	The number of adults specified in the hotel room	int
srch_children_cnt	The number of (extra occupancy) children specified in the hotel room	int
srch_rm_cnt	The number of hotel rooms specified in the search	int
srch_destination_id	ID of the destination where the hotel search was performed	int
srch_destination_type_id	Type of destination	int
hotel_continent	Hotel continent	int
hotel_country	Hotel country	int
hotel_market	Hotel market	int
is_booking	1 if a booking, 0 if a click	tinyint
cnt	Numer of similar events in the context of the same user session	bigint
hotel_cluster	ID of a hotel cluster	int

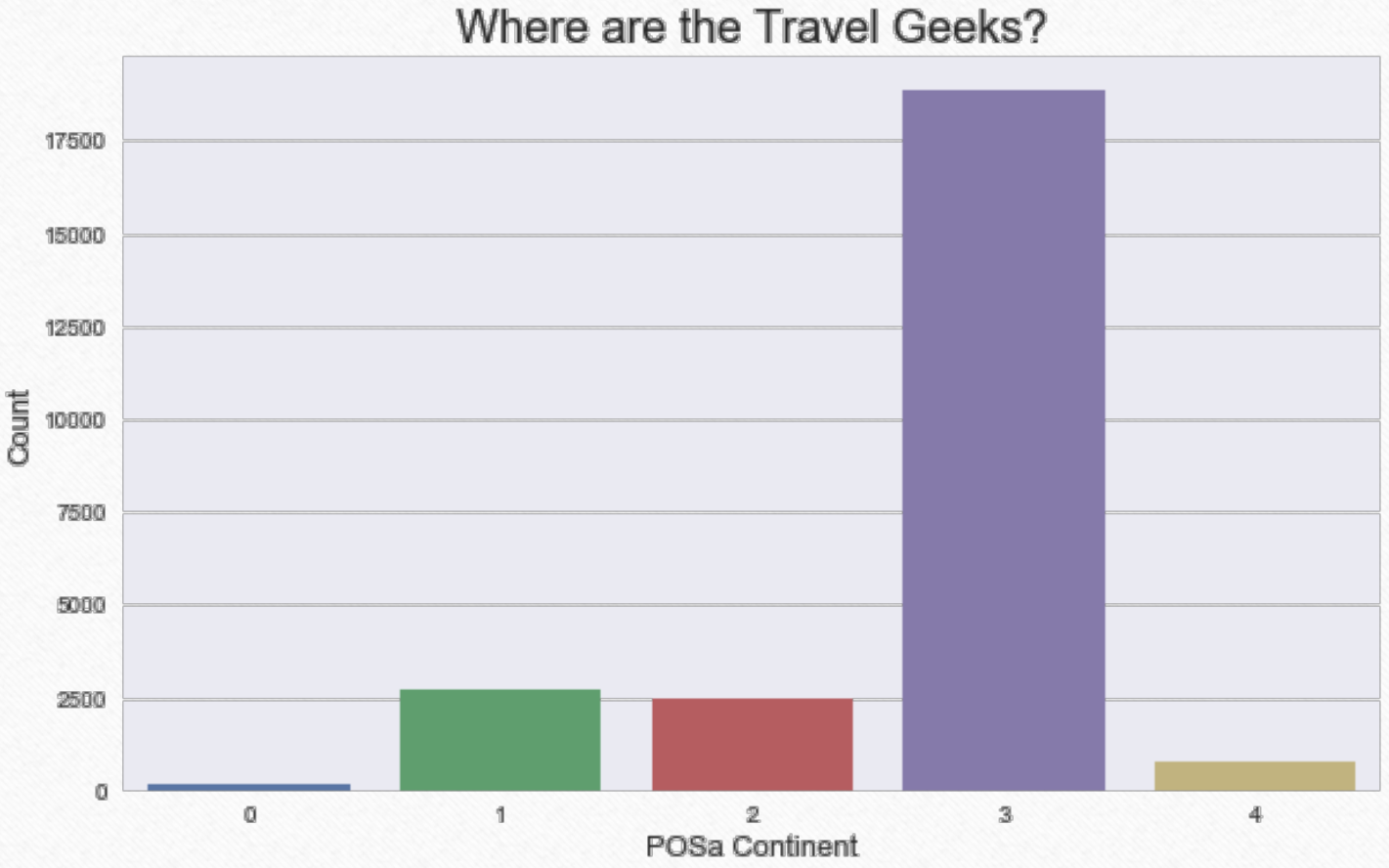
DATA Exploration

Exploring the dataset to evaluate which variables can be influence search and conversions



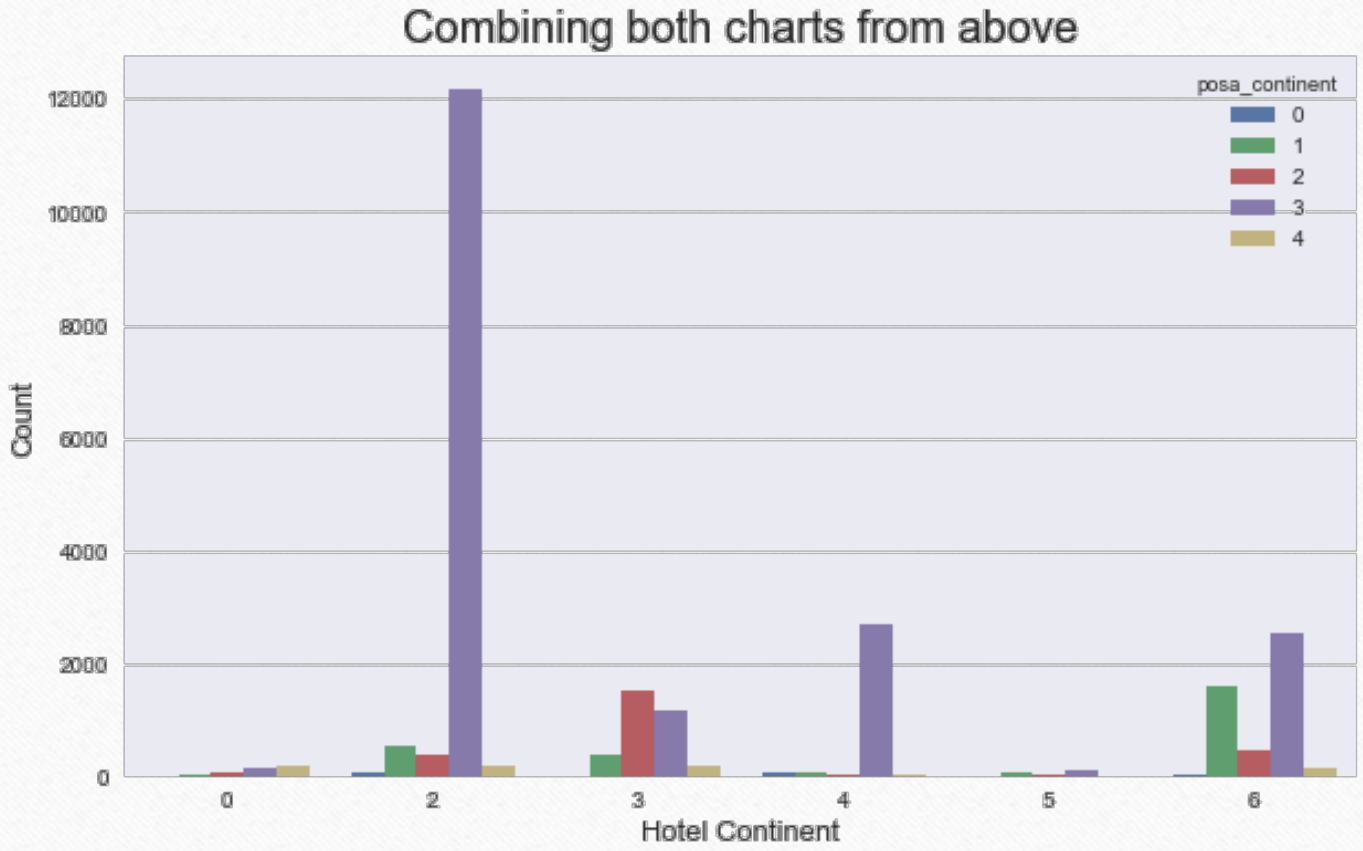
Continent 2 has the most preferred destination i.e., most search destinations are in continent 2

DATA Exploration



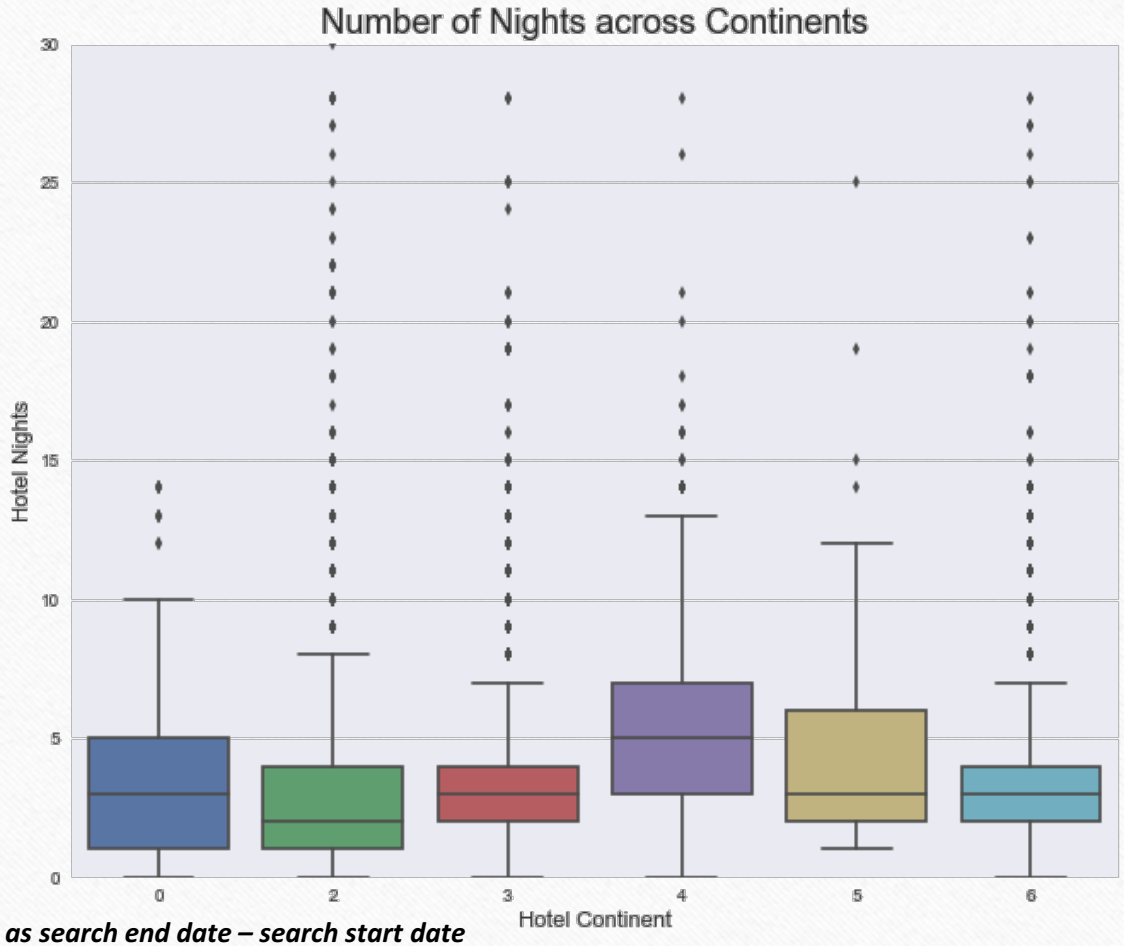
Most user who intend to travel are from continent 3

DATA Exploration



DATA Exploration

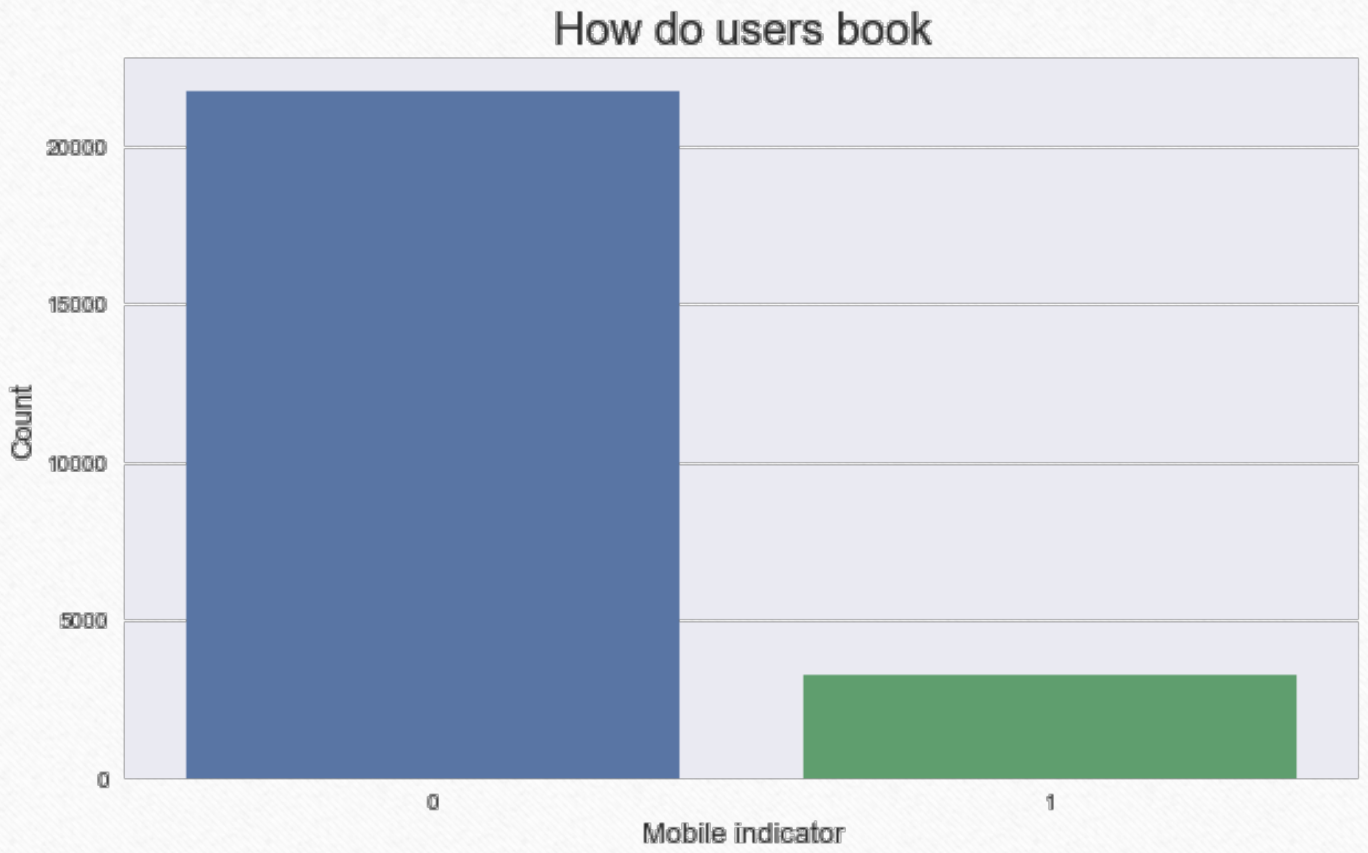
Hotels in continent 4 have higher median hotel nights* compared to others



**** Hotel nights is defined as search end date – search start date***

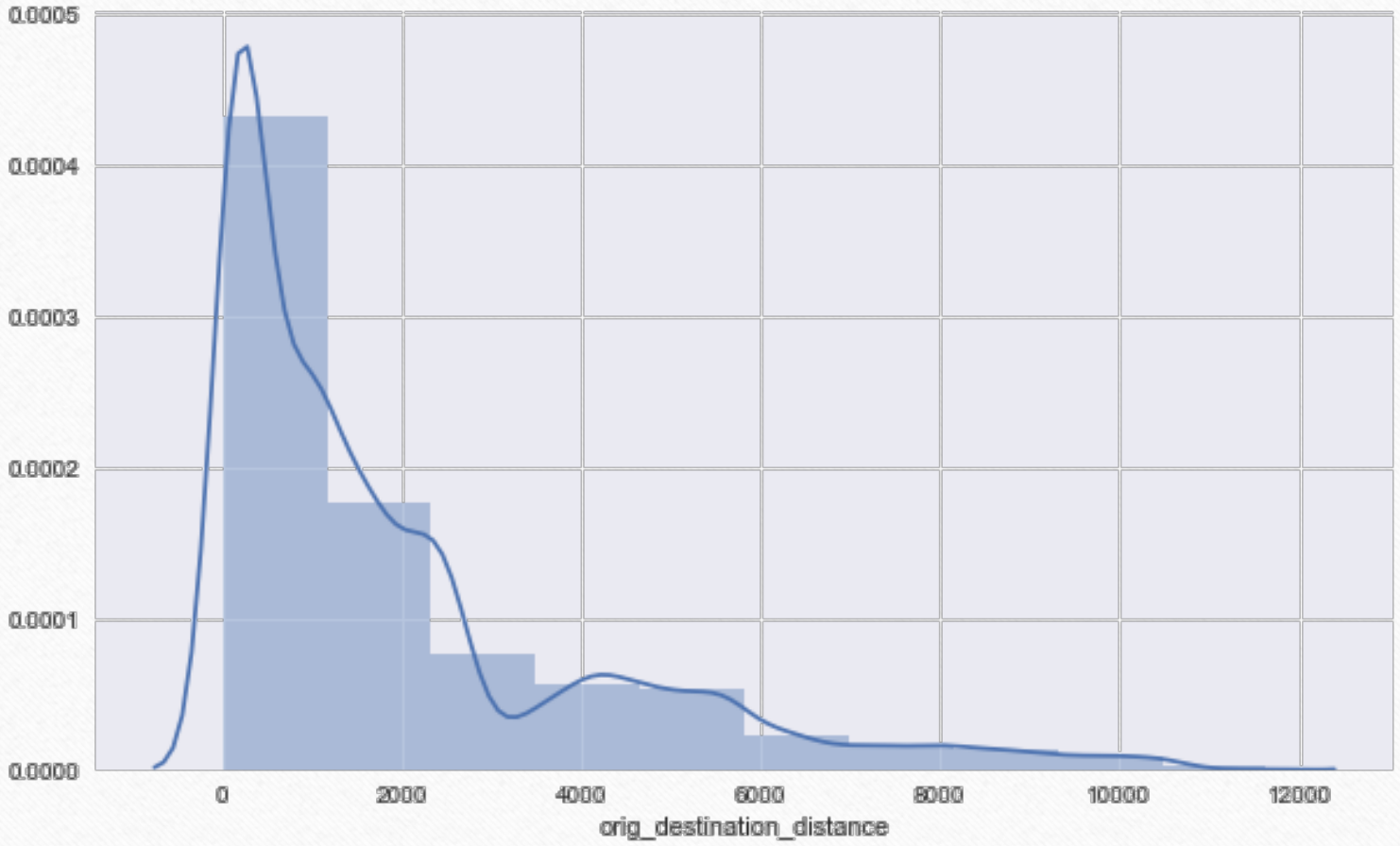
DATA Exploration

Most users search/book for their trip via Desktops



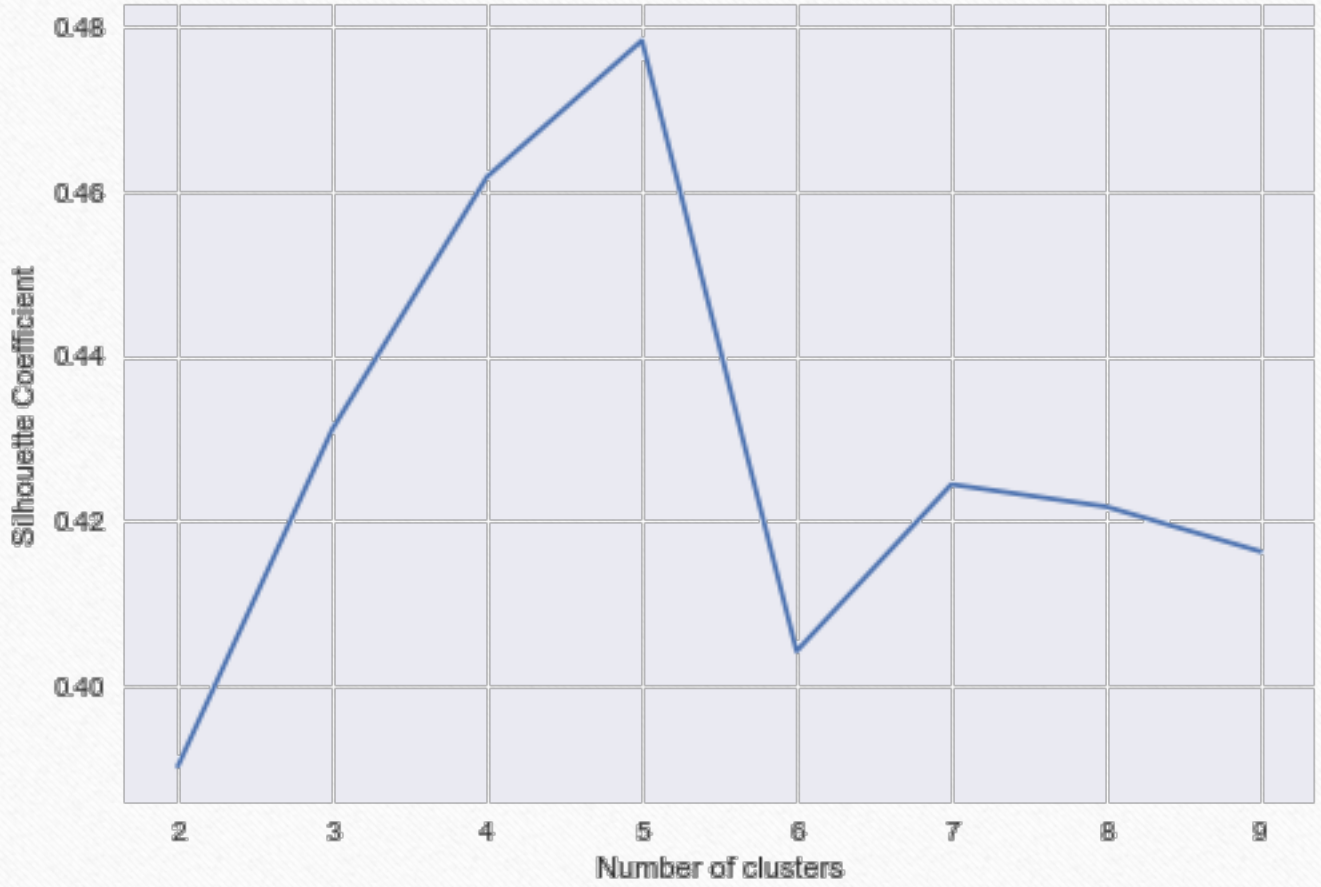
DATA Exploration

Users search for destinations with median distance 1950 miles

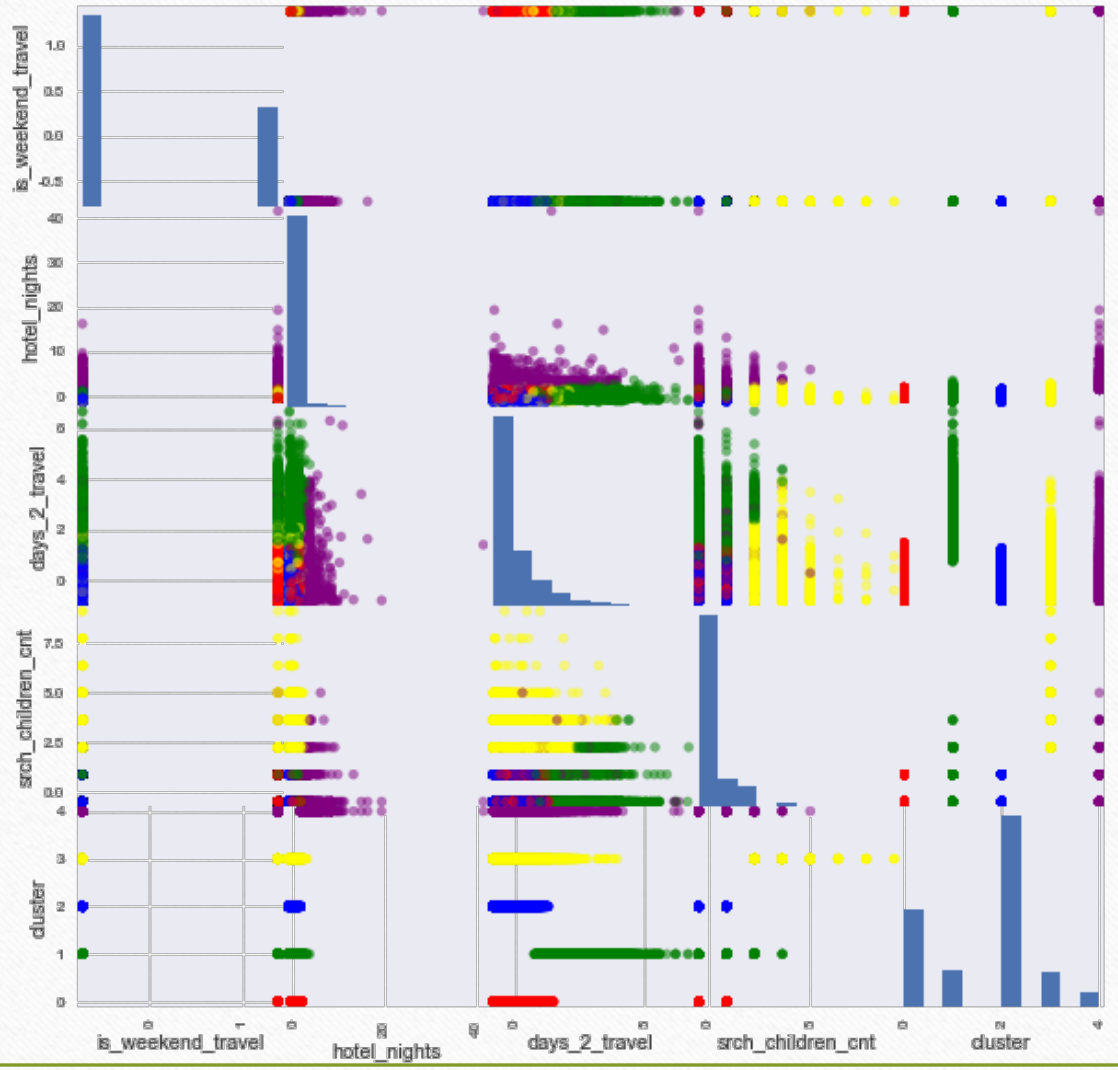


Other Results

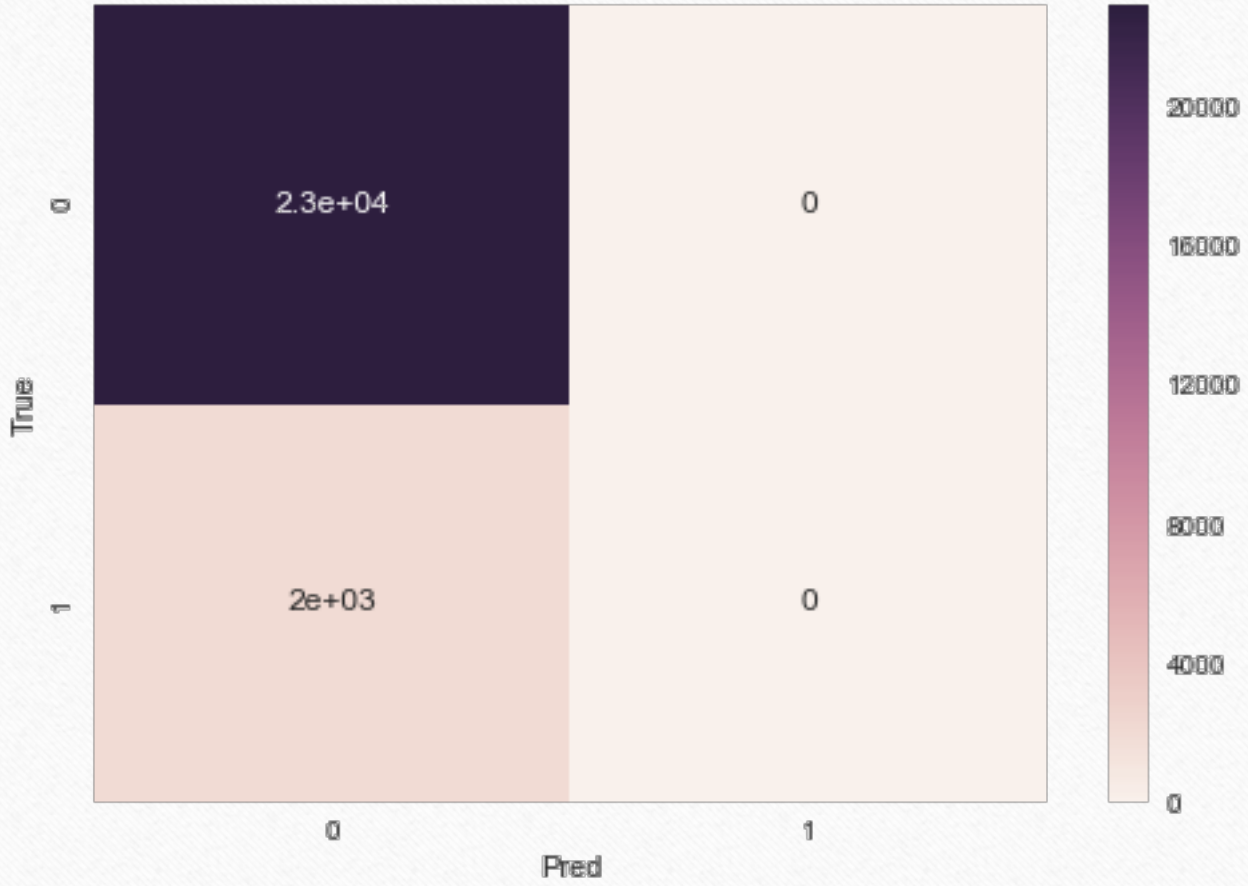
Segmenting customer based on type of travel to improve model



Other Results



Confusion Matrix from the model



RoC Curve