Growing a Family Hotel Business with Data Science & Customer Experience

1. Introduction

1.1 Background

As an aspiring data scientist graduating IBM Coursera Data Science Certification I offered my services to help a friend owning a family managed business hotel in Barcelona.
They want to increase revenue by offering additional products to customer to improve their journey by harnessing important touchpoints.
Touchpoints we want to fill are:

- Free time between trips and business for those customers traveling for business
- The need to discover the city with a guide for tourists visiting Barcelona

1.2 Problem

The business owner wants to increase revenues by selling tours or making visitors return to his hotel as follows:

- For new visitors he offers a tour for a very enticing price (Eur 15/person) if a group of minimum 3 customers is formed;
- For returning visitors he offers a tour for free

During these tours, guided by one of the managers they know the customers better, they can make a closer relationship with them and invite them to recommend the hotel to friends or by posting favorable reviews on special sites or by sharing the experience on social media. There is also an additional PR advantage for the hotel to offer these tours.

After brainstorming in the family (management of the business) using the Foursquare data they selected 4 types of tours:

- Tapas & Wine with Tapas Restaurants & Wine Bars
- Cultural with: Museum, Church, Cultural Center, Historic Site
- Walk & Shop: Stores, Neighborhood, Road
- Outdoor: Park, Plaza, Market

Further, after already offered these tours for 3 months, they want to optimize the offer and management of the tours by predicting for each new customer, based on the reservation data, which would be the preferred tour. They keep evidence with all the former customers and their choice and wants to leverage it to predict the preference of the coming customers.

1.3 Interest

My audience, in this phase, is the management team of Casa Kessler Hotel in Barcelona. Our goal is to obtain a 10% growth in their profit by offering these tours and optimizing the offer through the machine learning algorithm I recommend.

Later, depending on the success of this project, I will make similar analyses for other small businesses.

2. Data Description

To design an optimum variety of tours to cover most of the preferences we used Foursquare and extract all locations in Barcelona. For optimizing the offer we use an internal file.

2.1. Foursquare for getting data about places to visit in Barcelona and group them to obtain specific tours meant to cover a wide range of preferences.

categories	name	lat	lng
Tapas Restaurant	11	11	11
Hotel	7	7	7
Spanish Restaurant	7	7	7
Bookstore	5	5	5
Coffee Shop	5	5	5
Wine Bar	4	4	4
Plaza	4	4	4
Cocktail Bar	4	4	4
Mediterranean Restaurant	4	4	4
Japanese Restaurant	3	3	3
Sandwich Place	3	3	3
Donut Shop	3	3	3
Ice Cream Shop	2	2	2
Pizza Place	2	2	2
Fish & Chips Shop	2	2	2

2.2 Internal customer file (first data) for determining patterns and predicting customer preferences. Once the project is launched the hotel keeps evidence of the preferred tours.

The internal file contains demographic data collected usually from the internal reservations system, such as: gender, Entering day of the week, Staying days, Country, Number of persons. Other data are collected by the reception clerk at check-in: Travel type, Tourpref and returning times from the crm database of the hotel.
The CRM software classifies automatically the countries in 4 groups:
Spain - 1
Western Europe - 2
Eastern Europe - 3
Outside Europe -4

No	Age	Gender	Entering Day of Week	Staying days	Number of persons	Country	Country Code	Returning times	Travel Type	Tourpref
1	34	M	Mon	2	1	Spain	1		Tourist	
2	76	M	Mon	1	1	France	2		Tourist	Tapas&Wine
3	24	F	Wed	2	2	Spain	1		Tourist	
4	47	F	Fri	2	1	Germany	2		Tourist	Tapas&Wine
5	51	M	Tue	3	2	Hungary	3		Tourist	Cultural
6	34	M	Sat	4	2	Spain	1	1	Tourist	
7	32	M	Tue	2	3	Croatia	3		Business	Cultural
8	55	F	Wed	1	1	Austria	2		Business	
9	21	F	Mon	1	1	Ukraine	3		Tourist	Cultural
10	66	F	Tue	2	2	United Ki	2		Tourist	Walk&Shop
11	59	M	Fri	3	2	Spain	1	1	Tourist	Tapas&Wine
12	44	M	Mon	1	1	Germany	2		Business	
13	28	M	Tue	2	1	Austria	2		Tourist	Tapas&Wine
14	31	M	Wed	1	3	Spain	1		Tourist	
15	46	F	Tue	3	2	Spain	1	2	Tourist	
16	32	M	Mon	2	2	Italy	2		Tourist	Tapas&Wine
17	38	M	Wed	2	1	Netherlar	2		Business	Cultural
18	57	F	Fri	1	1	Spain	1	1	Tourist	
19	39	F	Tue	3	2	Spain	1		Tourist	
20	24	M	Mon	2	2	Andorra	2		Tourist	Tapas&Wine

Looking at the data some cleaning operations are necessary:

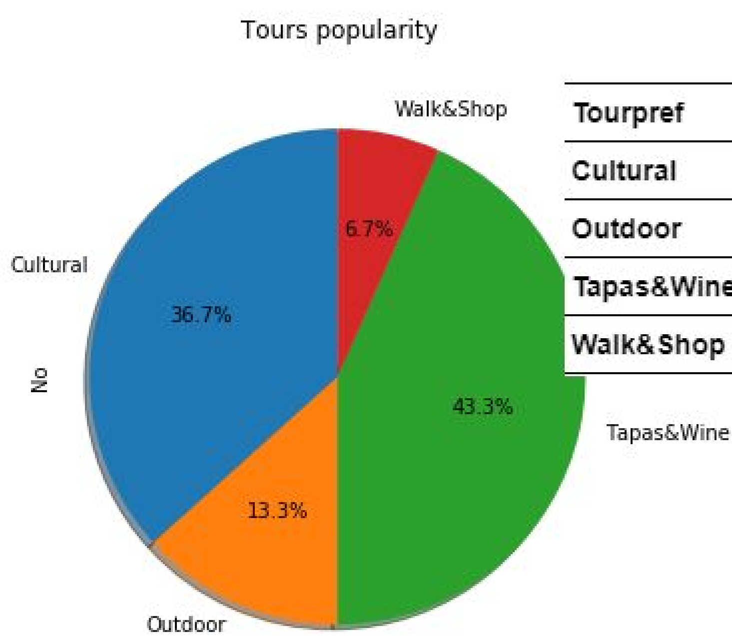
- columns “Returning times” and “Tourpref” contains null values, we will fill them with 0
- We have categorical values like “Gender”, “Country code”, “Entering Day”, “Travel Type” and the target variable -> we will cast them to numerical values

We can also see a redundant feature - Country. I will not use it as I already have a category for it.

3. Methodology

3.1 Exploratory Data Analysis

3.1 A - We analyze first the popularity of the tours to understand if the offer was attractive enough, if they are unpreferred tours or to see some insights visible with the naked eye:

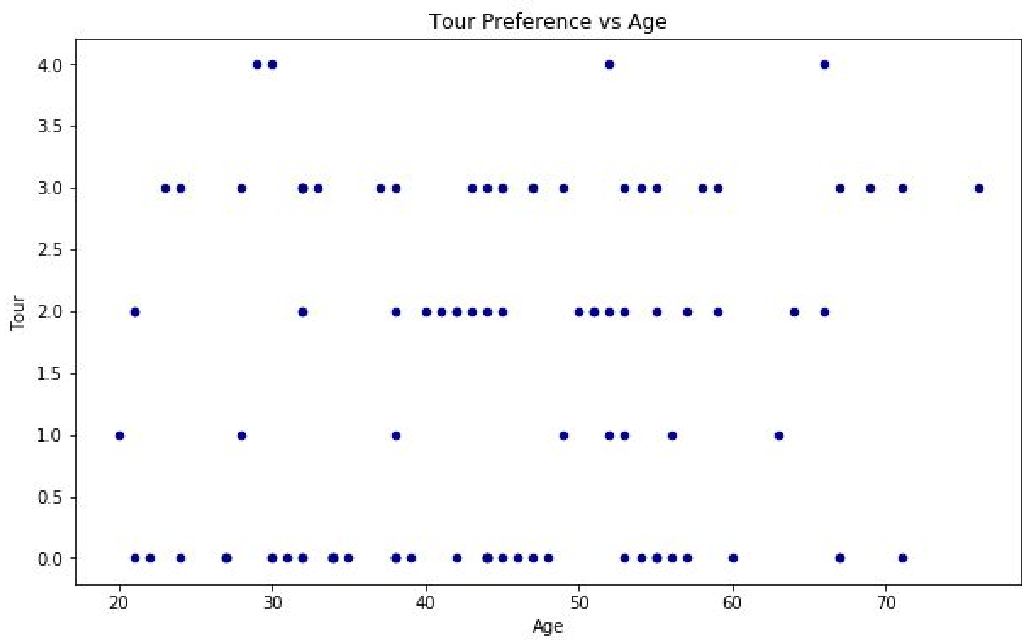


	No
Tourpref	
Cultural	22
Outdoor	8
Tapas&Wine	26
Walk&Shop	4

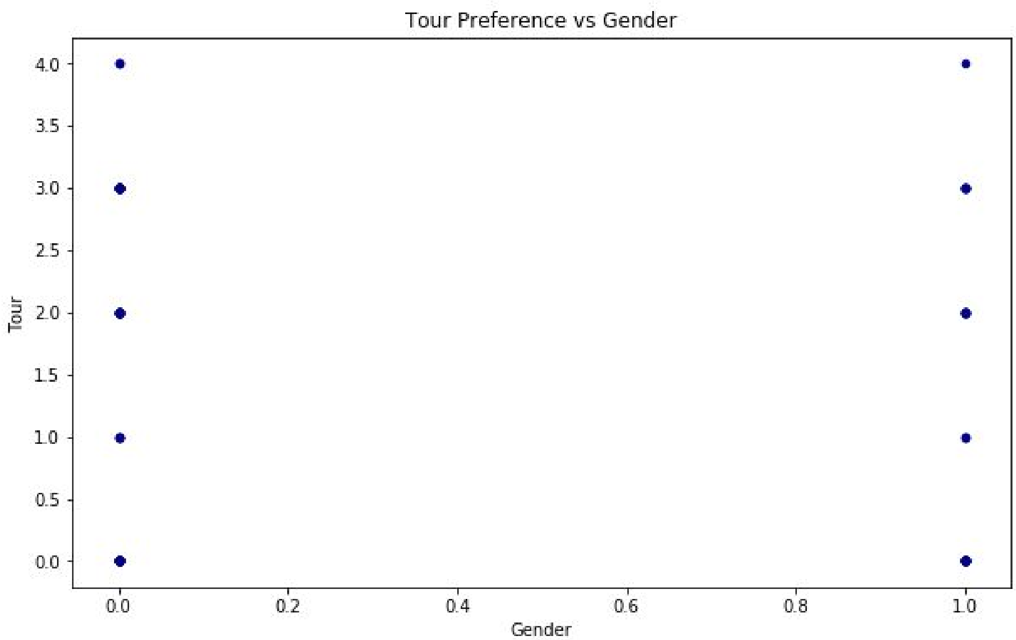
Tour Name	Tour Code
Outdoor	1
Cultural	2
Tapas	3
Walk & Shop	4

3.1.B The target variable is Tourpref and we have to decide which features influence the target. We will visualize the target variable correlation with each of the features.

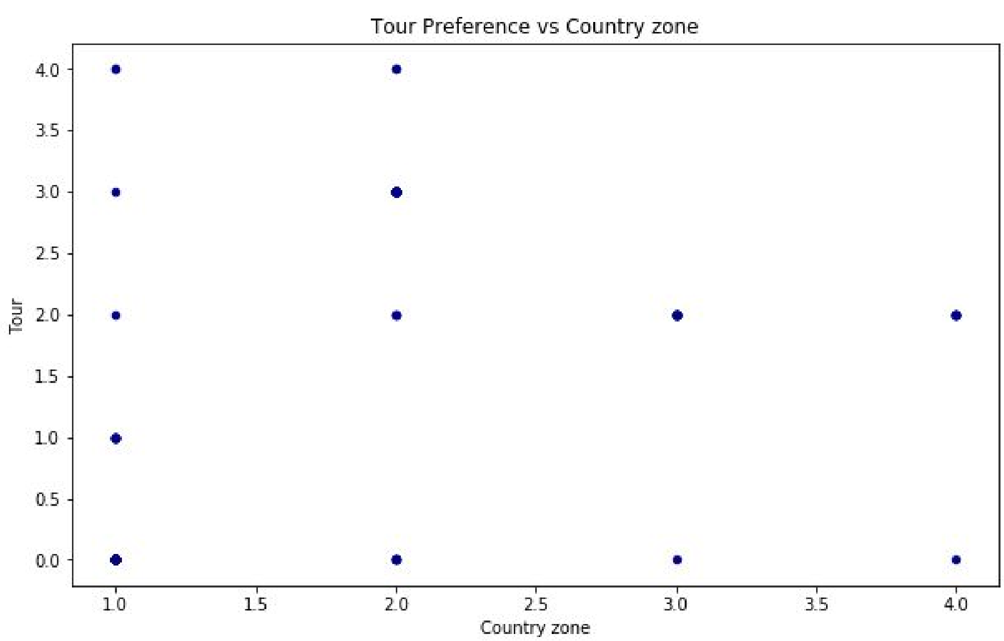
Tourpref vs Age - some tours are preferred equally by all ages while others are more popular among certain age. For example, Walk & Shop have no customer between 30 and 50 years old while Cultural is very popular among people in this age range.



Tourpref vs Gender - all tours are equally preferred by men and women, and this is the reason not to use gender as a parameter when modeling as it does not influence the option.



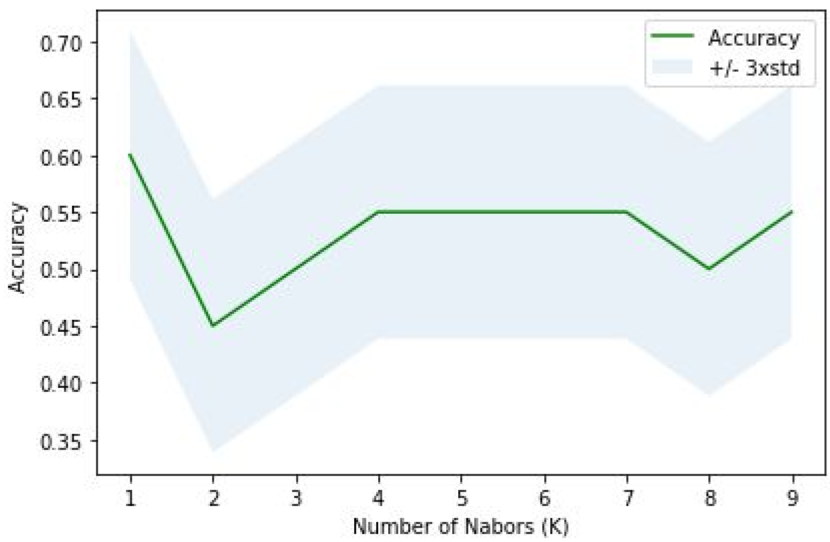
Tourpref vs Country - it is obvious that the origin country of the customer matters.



We will also keep all the other feature in the model.

3.2 Machine learning algorithm - kNN

The hotel segments its customers by tour preference into 4 groups. They want to predict group membership of future customers based on demographic data. This is a classification problem and we will use kNN (K Nearest Neighbour) method. This is a classification algorithm that takes a bunch of labeled points and use them to learn how to label new points. The algorithm classifies new customers based on their similarity with other customers. In order to decide the value of k we break the existing set into train and test. Out of 100 customers we will use 80 to train the model and 20 to test the model Simulating for k in range 1-10 I got the maximum accuracy (0,6) with k=1. In other words we obtain the best accuracy when comparing with only 1 neighbour.



Applied on the test set the accuracy was 0.91

4. Results

The first prediction is for a set of 5 customers expected to come.

	No	Age	Gender	Entering Day of Week	Staying days	Number of persons	Country	Country Code	Returning times	Travel Type
0	1	44	0	1	2	1	Spain	1	0.0	0
1	2	22	1	1	1	1	France	2	0.0	0
2	3	31	1	1	2	2	Spain	1	1.0	1
3	4	50	0	2	3	1	Germany	2	0.0	0
4	5	55	0	2	2	2	Hungary	3	0.0	0

The hotel expects them to arrive on Monday and Tuesday and wants to organize everything as efficient as possible.
Applying the model on these clients we get the following prediction for them:

	No	Age	Gender	Entering Day of Week	Staying days	Number of persons	Country	Country Code	Returning times	Travel Type	Tourpref
0	1	44	0	1	2	1	Spain	1	0.0	0	0
1	2	22	1	1	1	1	France	2	0.0	0	Cultural
2	3	31	1	1	2	2	Spain	1	1.0	1	0
3	4	50	0	2	3	1	Germany	2	0.0	0	Tapas&Wine
4	5	55	0	2	2	2	Hungary	3	0.0	0	Cultural

In other words, the hotel knows that on Monday they only have to prepare for a Cultural tour while for Tuesday and/or Wednesday there will be opportunities for Cultural and Tapas. This way they can manage their resources more efficient and they also can predict the expected revenue from these tours.

5. Discussion

The idea of offering a complementary product generated additional sales. Out of 100 customers in the first 3 months from the implementation, the hotel sold 61 tours out of which 17 were for loyalty (free tour as a promo for returning visitors). They made an additional revenue of $660 from paid tours and gained 17 returning customers.

The cost of organizing these tours will be minimal due to predictions. They will manage the daily activities in such a way that one of the family members working there to be able to organize everything in the usual labor time, with no additional cost or at an insignificant one.

I also recommend them to increase the price of a tour to Eur 25 based on a comparison with taxi fares and other similar offers on the market including so-called “Free tours” which, in fact, are not free.

As we can see in the Tours’ Popularity Analysis, Cultural and Tapas & Wine are the most popular ones. Besides this, Walk & Shop and Outdoor are very similar to each other. I recommend them to think about reshaping the offer to contain only 3 tours by combining Walk & Shop with Outdoor into one single product.

The accuracy of the machine learning algorithm used is still influenced by the small amount of training data we have now. This could be the reason for k=1 being the most accurate k. In time, after acquiring more customers I hope to increase the accuracy, and I will also have to reassess the value of k. For the moment, having a small number of clients and a great diversity among them cause the best accuracy to be for k=1.

6. Conclusion

In this study, I analyzed the impact of offering additional products during the customer journey to increase revenue and customer satisfaction & loyalty. I came up with an idea, I brought Foursquare data to nail the additional offer then, I applied kNN machine learning on a set of labeled customers to predict the preference of new customers based on demographic data similarity.

My model was adopted, generated revenue in 3 months from its kick-off, and we discovered a few ways to improve it. Monitoring and fine-tuning it in the very next period could build us trust to meet the initial goal of a 10% increase in annual revenue.