



GROWING A FAMILY HOTEL BUSINESS WITH DATA SCIENCE

CASA KESSLER BARCELONA

Capstone Project - Mirela Iancu

April 2019



Business Problem

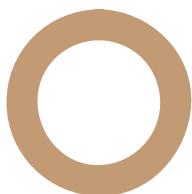
As an aspiring data scientist graduating IBM Coursera Data Science Certification I offered my services to help a friend owning a family managed business hotel in Barcelona - Casa Kesler Barcelona. They want to increase revenue by offering additional products to customers to improve their journey by harnessing essential 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



Business Problem



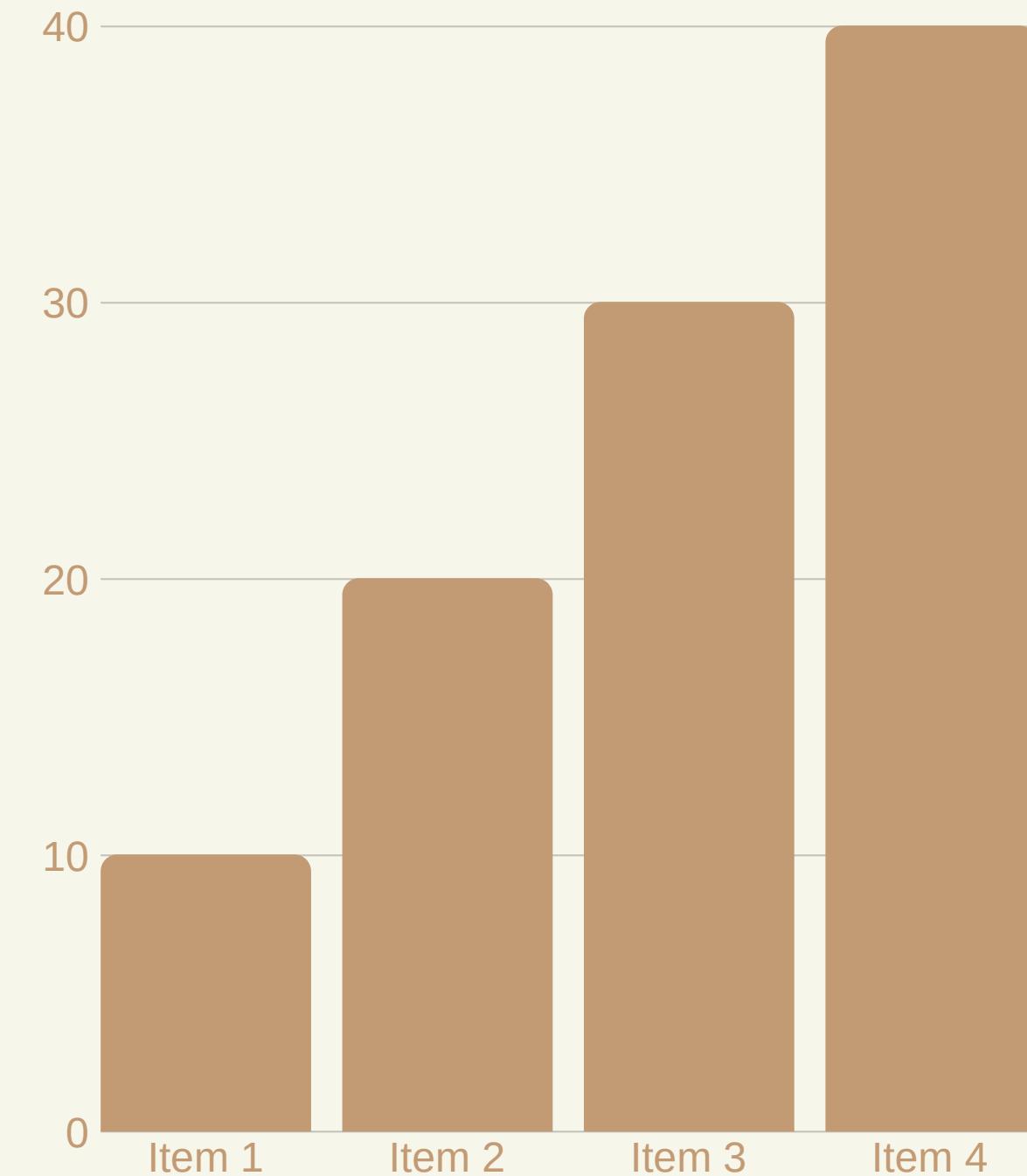
The business owner wants to increase revenues by selling tours or making visitors return to his hotel as follows:

- 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 provides a tour for free

During these tours, guided by one of the managers they know the customers better, they can make closer relationships 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.

DATA DESCRIPTION & USE

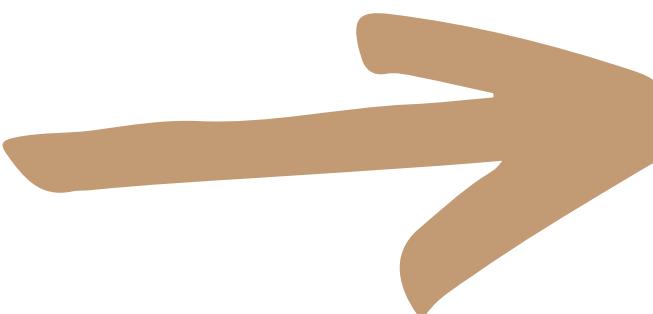
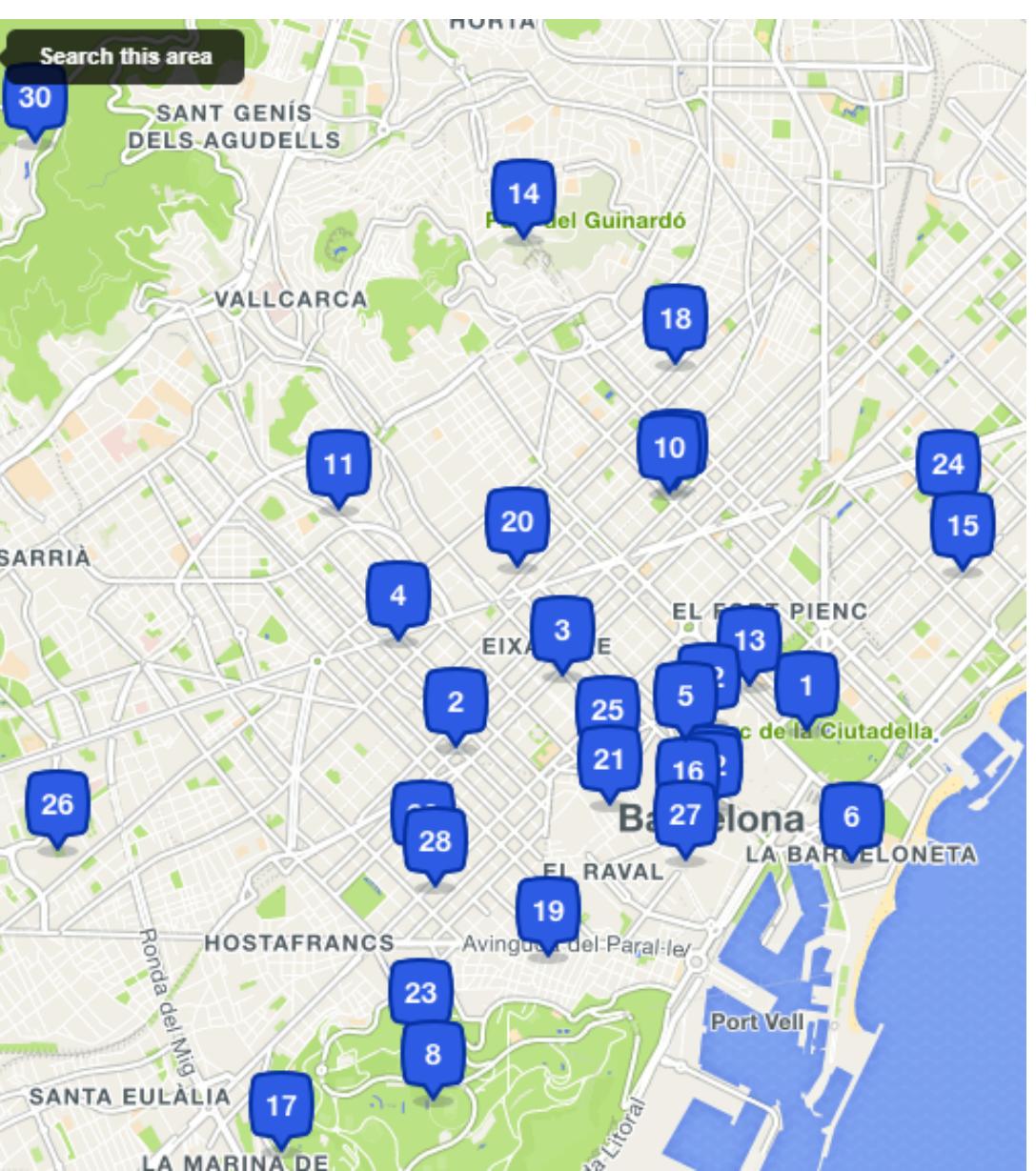
FOURSQUARE & INTERNAL FILE



WHAT I'LL BE DOING

DATA DESCRIPTION

To design an excellent variety of tours to cover most of the preferences we used Foursquare and extract all locations in Barcelona.



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

DATA DESCRIPTION

We will use KNNs method to make a prediction, for each new customer to know what his choice would be.

K-nearest neighbors is a pattern recognition method we will discuss in the next pages.

To conclude the data description topic, I resume the data sources and model:

1. Foursquare for getting data about places to visit in Barcelona
2. Internal customer file (first data) for determining patterns and predicting customer preferences
3. KNN algorithm applied on internal file to optimize the offer

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

DATA DESCRIPTION

After brainstorming in the family (management of the business) using the Foursquare data we selected four 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 three months, he wants 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. He kept evidence with all the former customers and their choice and wants to leverage it.

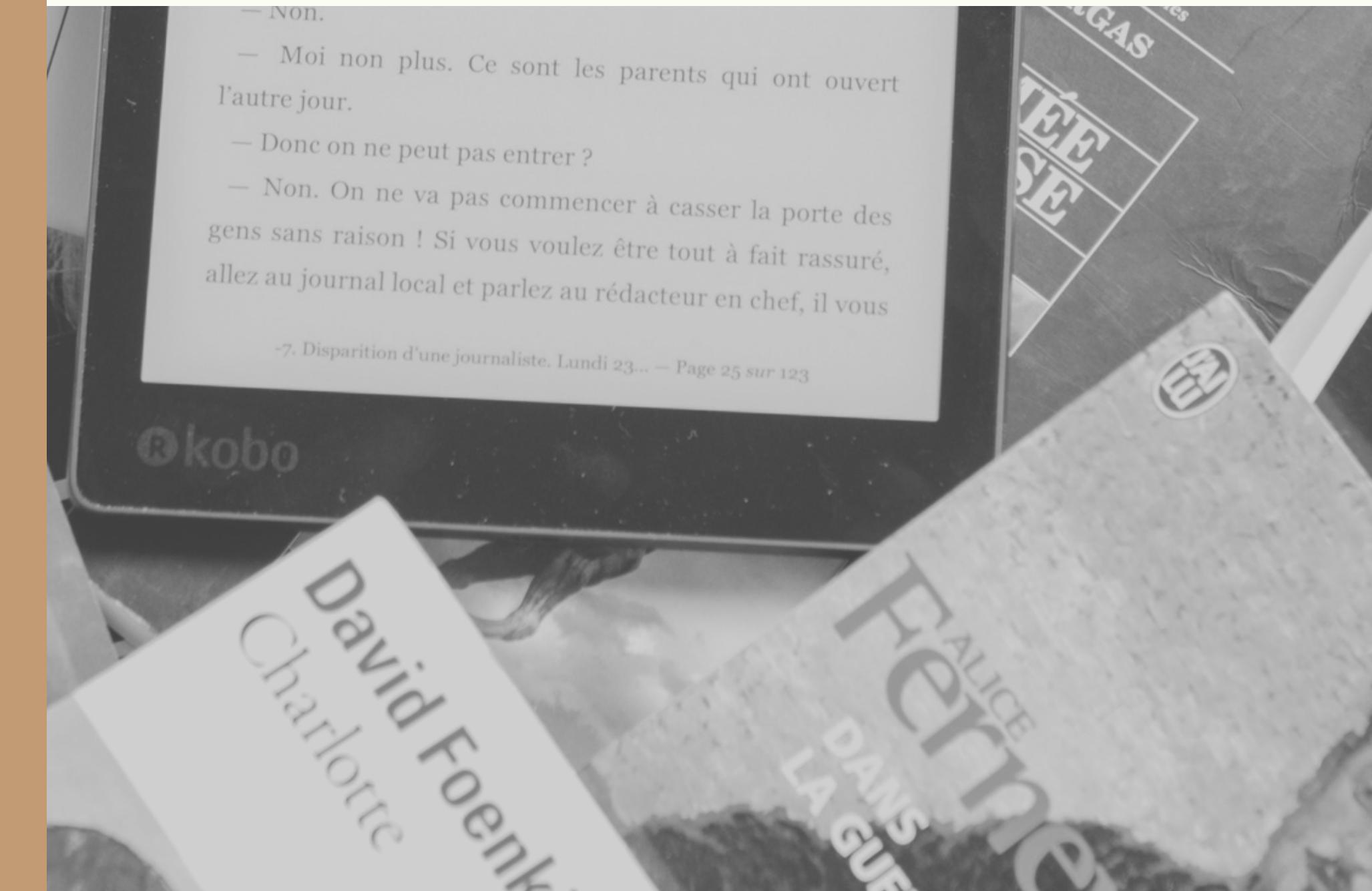
METHODOLOGY

What we'll be doing...

EXPLORE DATA AND APPLY
MACHINE LEARNING

— Non.
— Moi non plus. Ce sont les parents qui ont ouvert l'autre jour.
— Donc on ne peut pas entrer ?
— Non. On ne va pas commencer à casser la porte des gens sans raison ! Si vous voulez être tout à fait rassuré, allez au journal local et parlez au rédacteur en chef, il vous

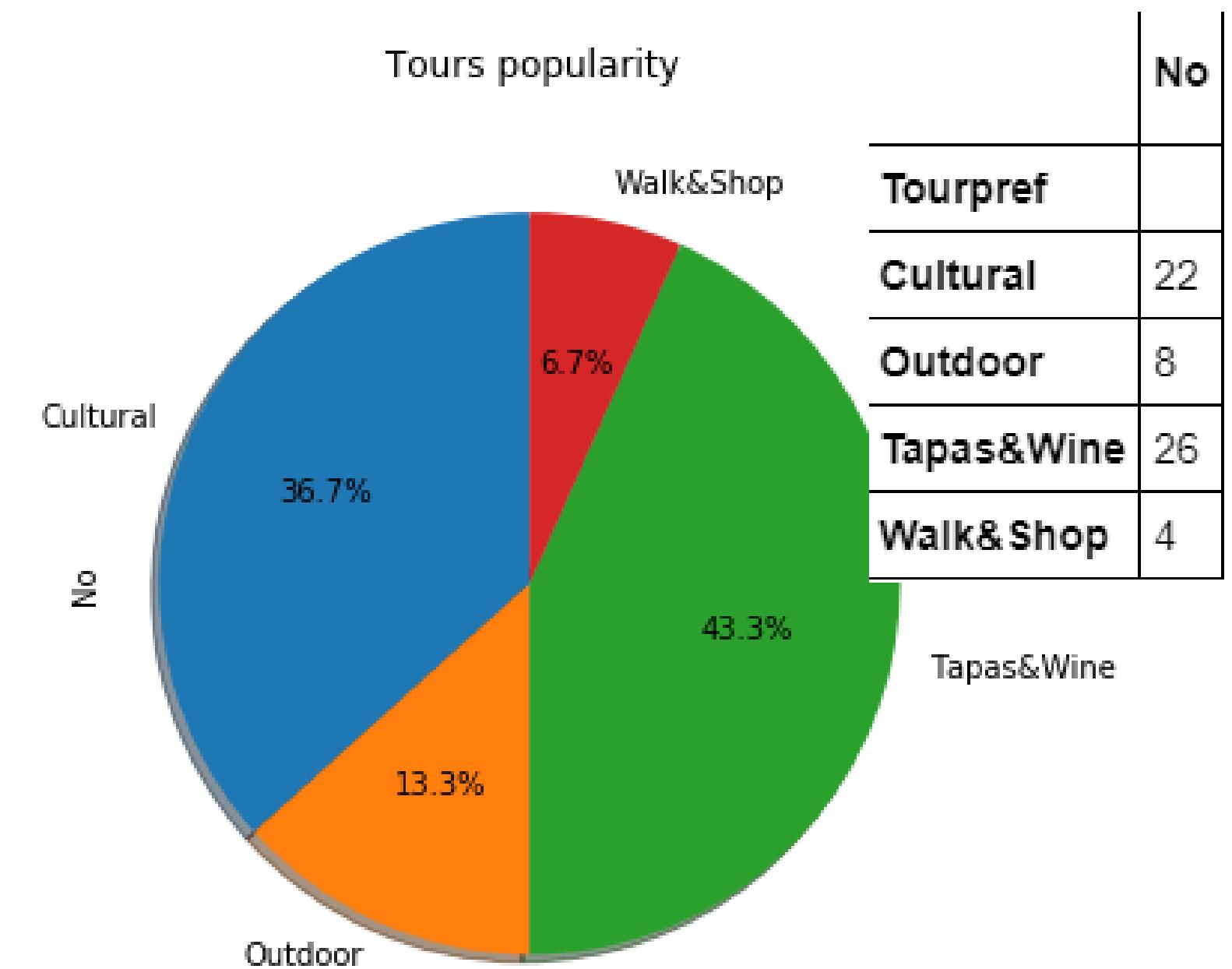
—7. Disparition d'une journaliste. Lundi 23... — Page 25 sur 123



METHODOLOGY

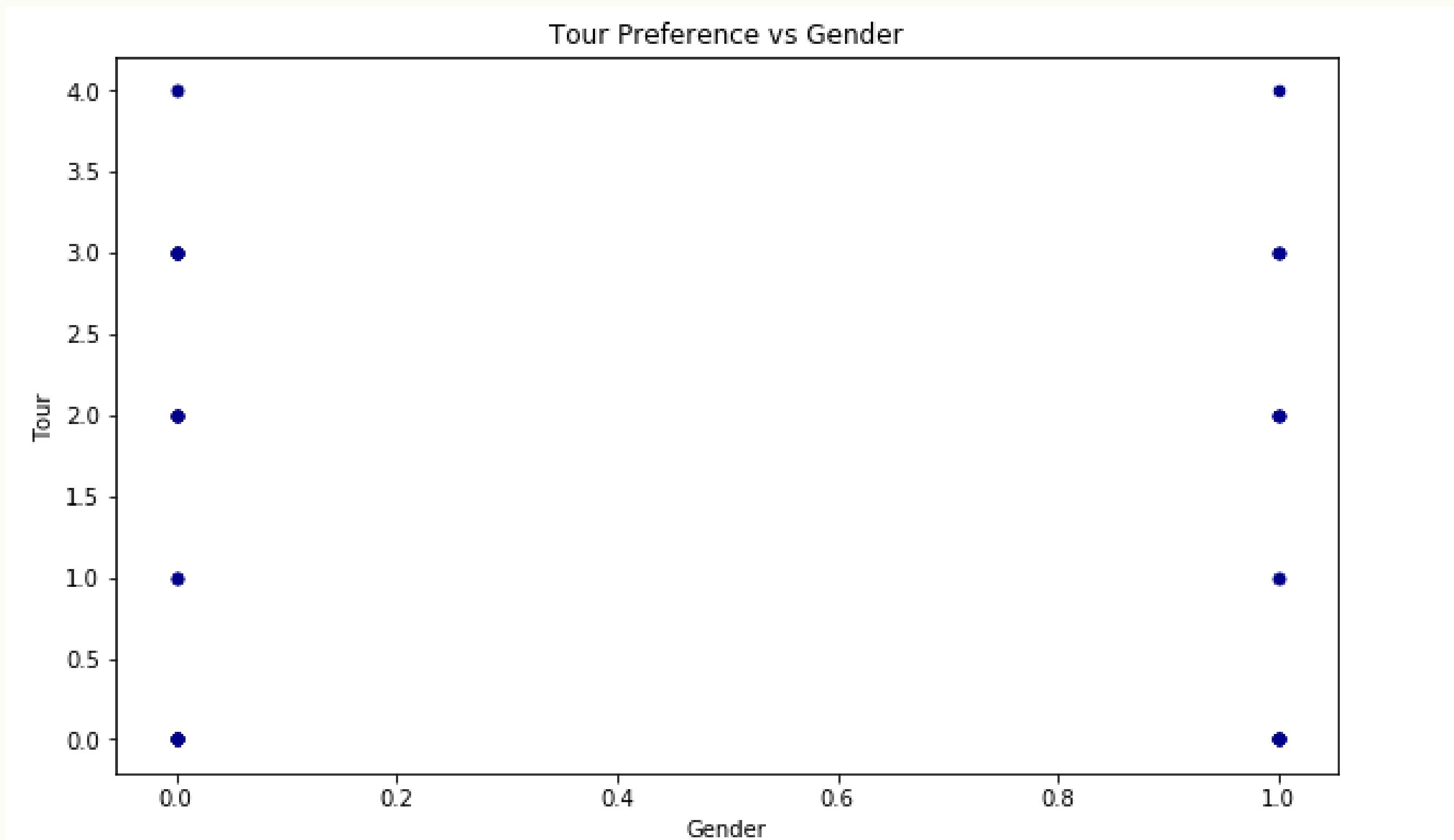
The Foursquare data helped us in designing 4 types of tours.

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

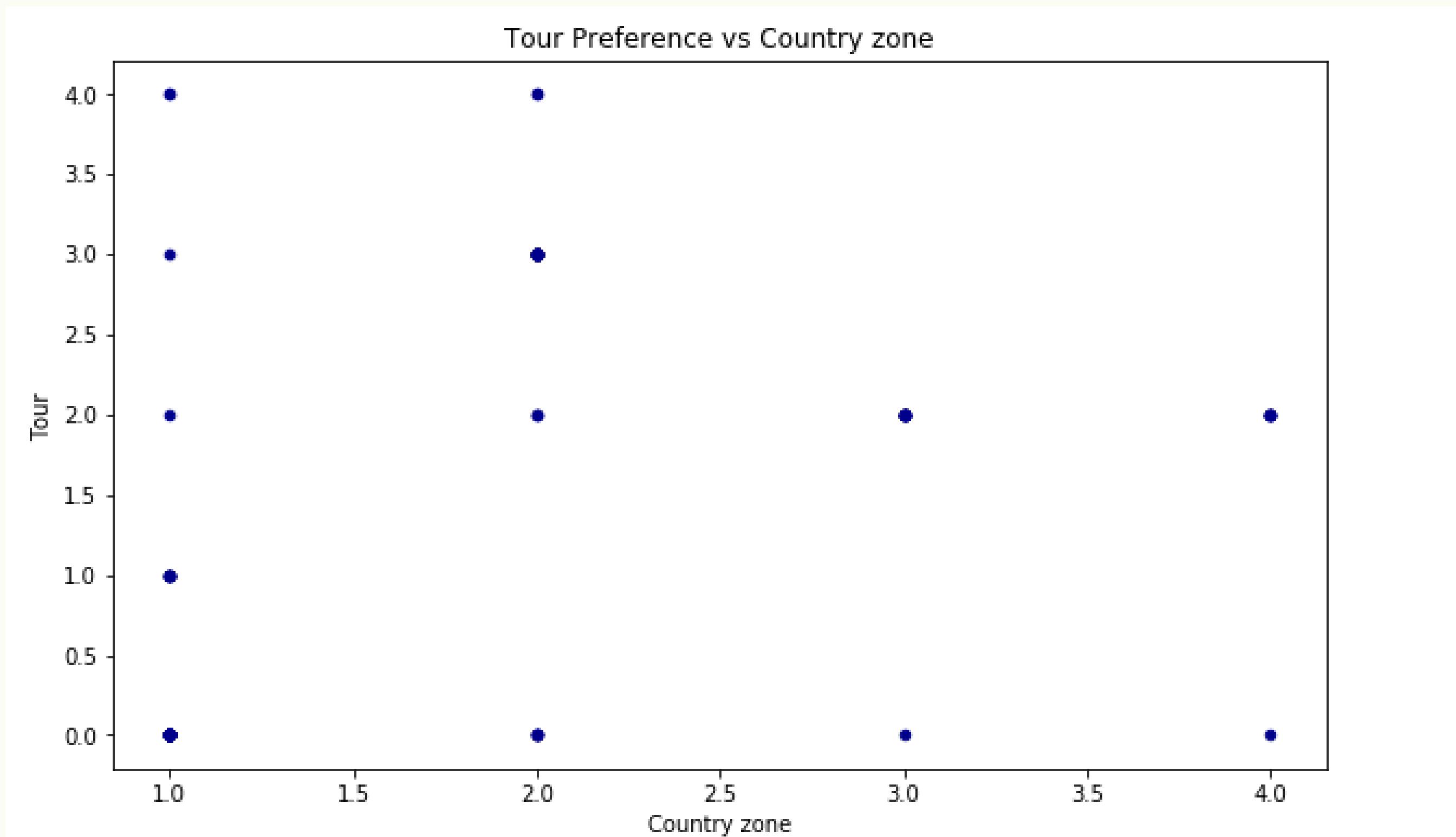


METHODOLOGY

Tourpref VS Essential Features

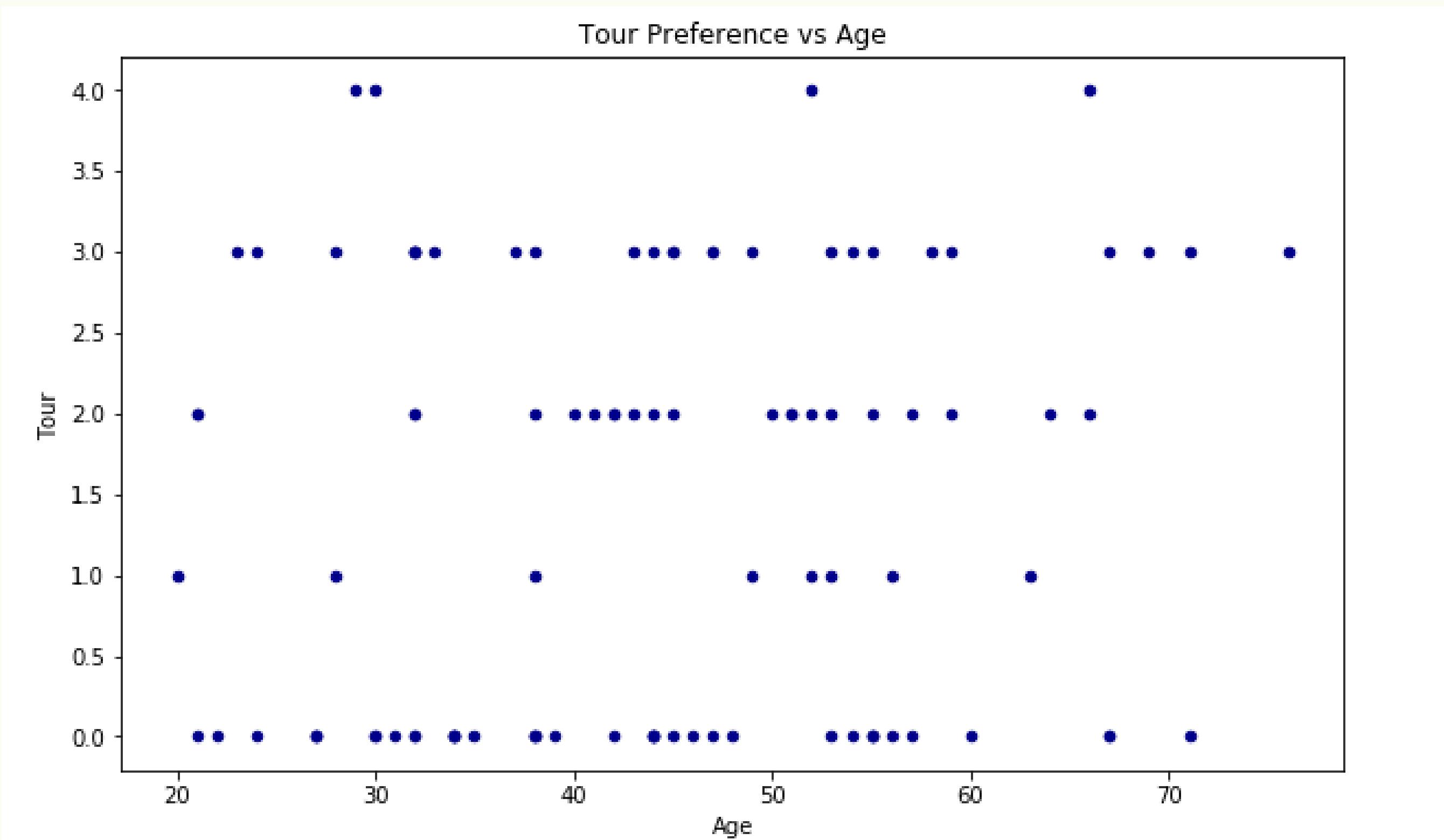


Tourpref VS Essential Features



METHODOLOGY

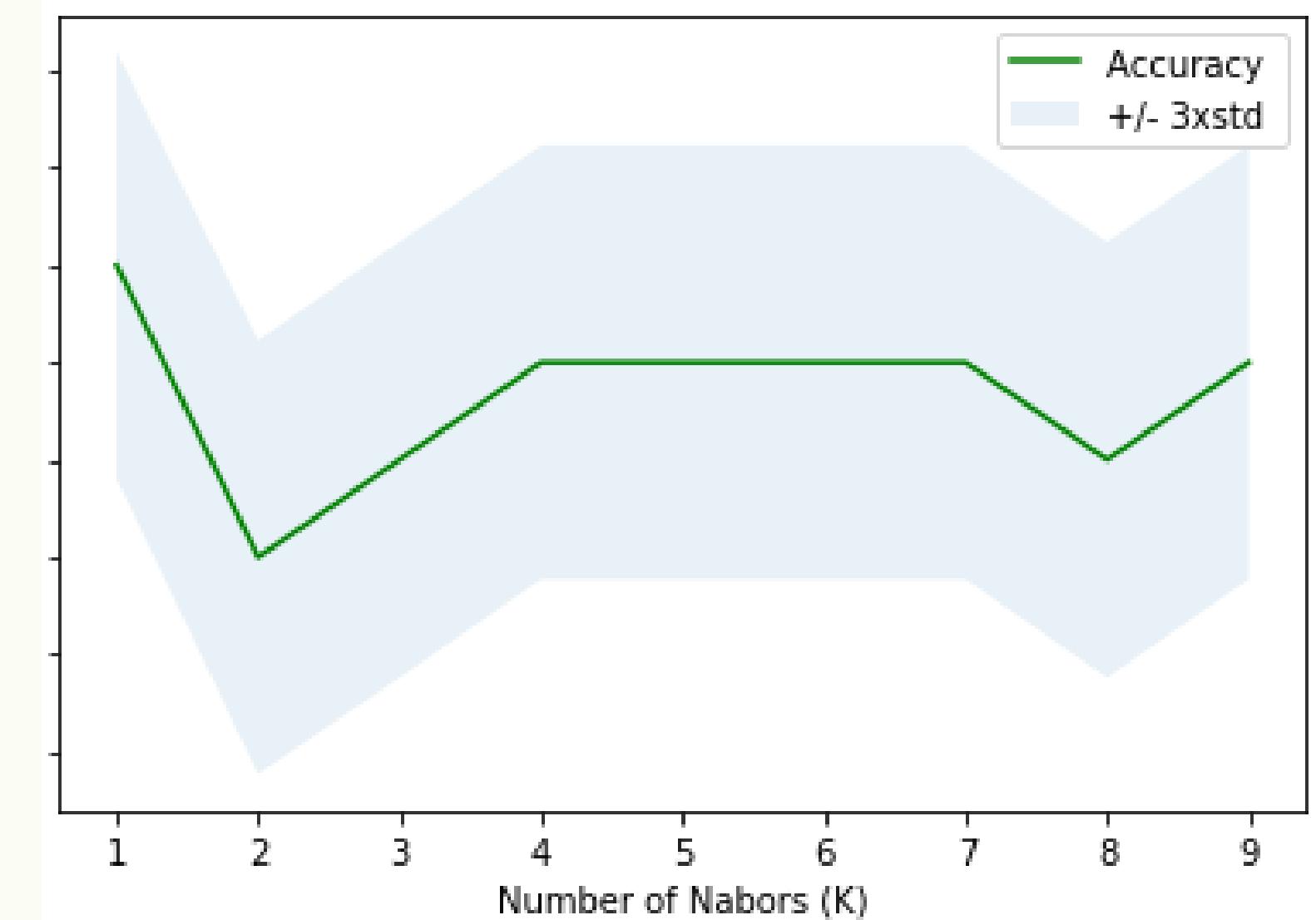
Tourpref VS Essential Features



METHODOLOGY

Machine learning algorithm - kNN

- We will use kNN algorithm to classify new customers based on their similarity with other customers.
- From all the features in the internal customer file we will not use 'Country' as it is redundant and Gender which do not impact the dependent variable. we will use all the other features
- For the moment k=1 determines the best accuracy



	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

RESULTS

DISCUSSION



- Generated sales: 61 tours
- Returning visitors: 17
- Recommendations: to narrow the offer to only 3 tours and raising the price to Eur 25

ALWAYS REMEMBER

“

Numbers have an important story to tell. They rely on you to give them a voice

Stephen Few - Innovator