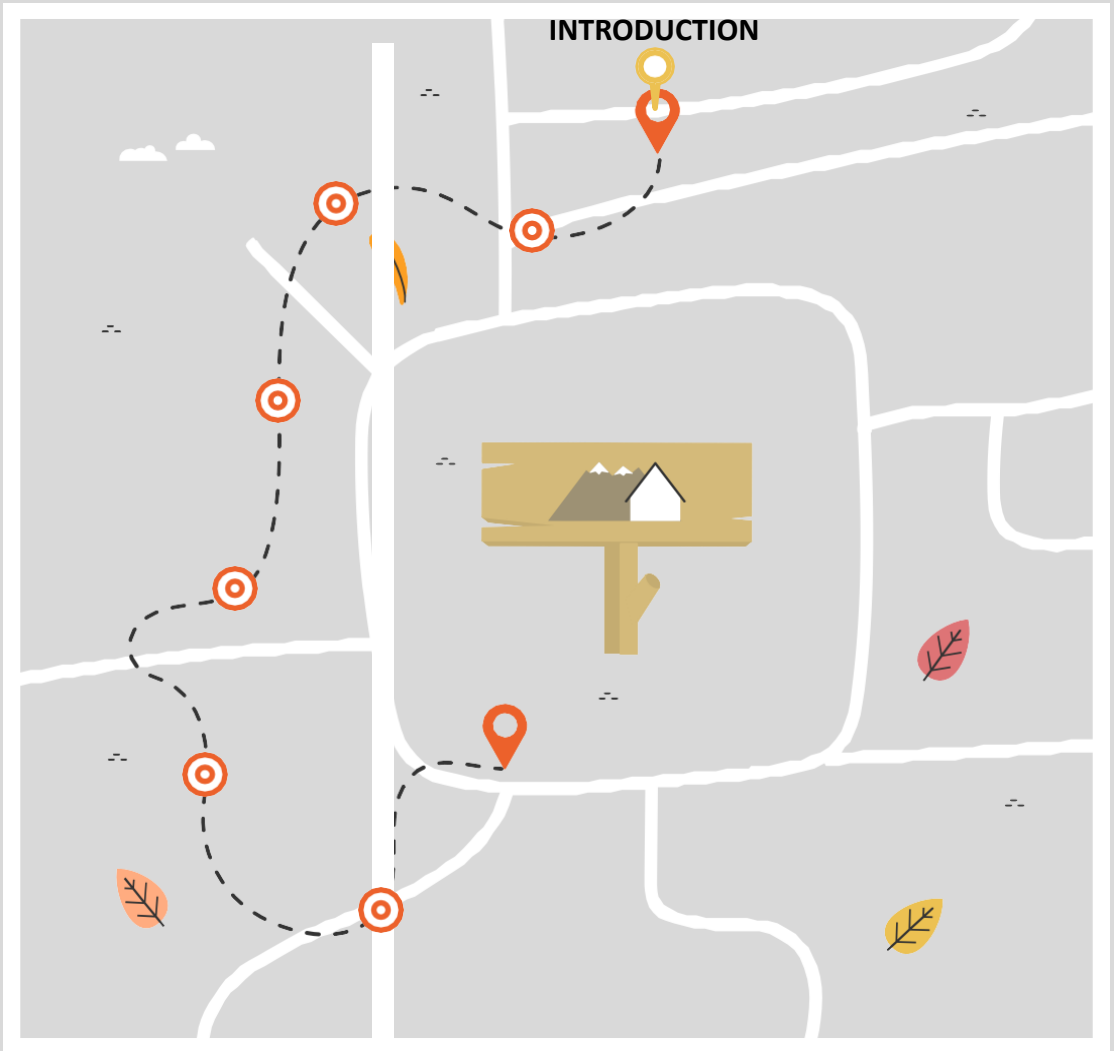
Hotel Review

Booking.com

Aishah Alanazi, Hala Alenazi, Hussain Alsultan
Yaqeen Alhawaj and Shaikha saud

TABLE OF CONTENTS

01	02	03
INTRODUCTION	PREPROCESSING	DATA ANALYSIS
04	05	06
TOPIC MODELING	CLUSTERING	RECOMMENDATION SYSTEM
07	08	09
DATA ANALYSIS	MODELS	CONCLUSION



01

INTRODUCTION

Backstory

Reviews on websites as simple it sounds; but it has a significant impact on how tourists choose their accommodation. It is vital for hotels to have reputation for quality and it would work as trustable verification for the hotel quality. A recent Barclays study showed there is potential of an extra £3.2 billion, by just paying more attention to online reviews. The dataset that was collected from Booking.com.

The **goal** of this project is to build unsupervised Natural Language Processing (NLP) machine learning models that decide whether a text review is positive review or negative review. This project, will help hotels to determine the category of text review and cluster them automaticity to improve their services.

DATASET

Data from Booking.com

Hotel Reviews Data in Europe (Kaggle)

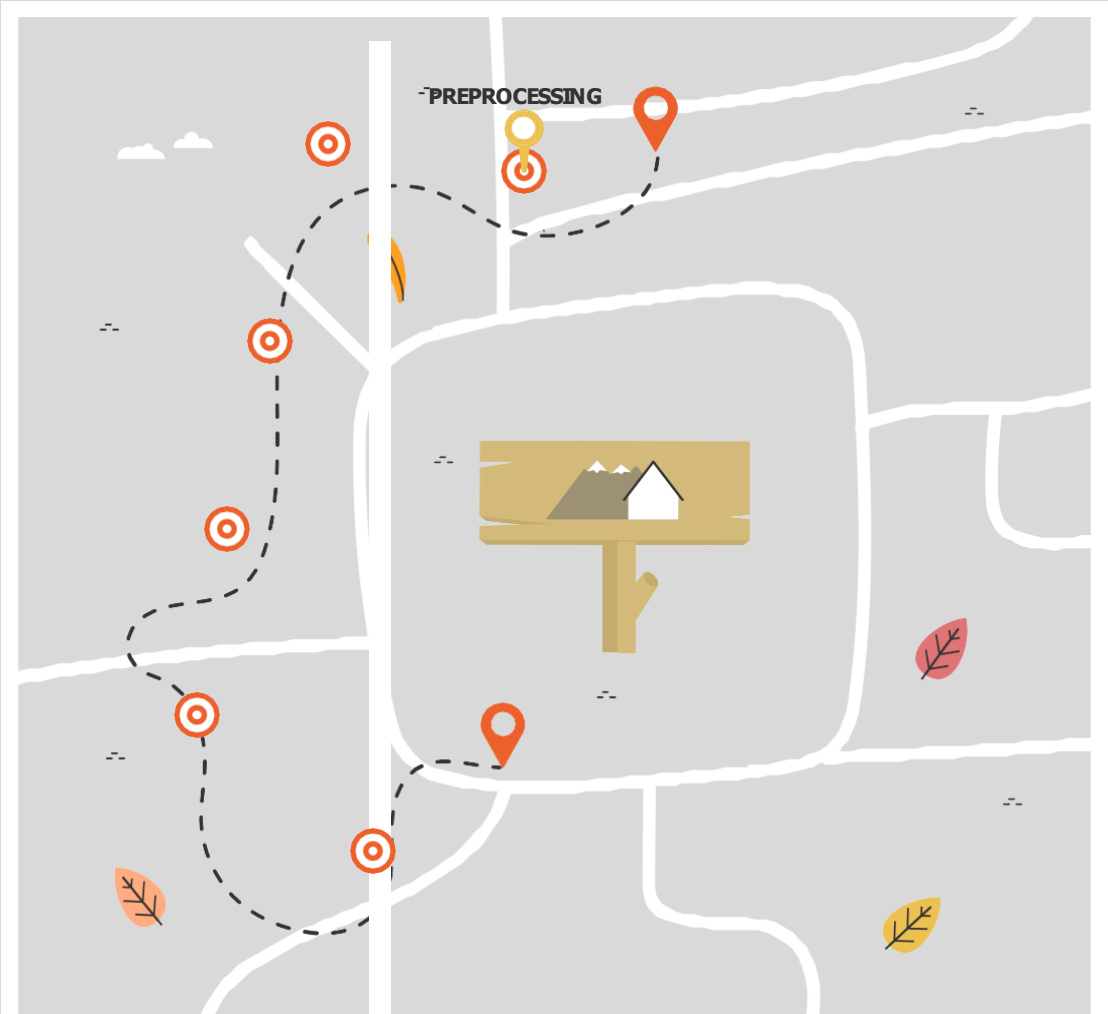
Before: 515,738 rows , 17 columns

After: 939,223 rows × 6 columns

Data Sample 100,000

Tools

- Pandas
- NumPy
- Matplotlib
- Seaborn
- Sklearn
- NLTK
- CorEx
- WordCloud



02

PREPROCESSING

PREPROCESSING

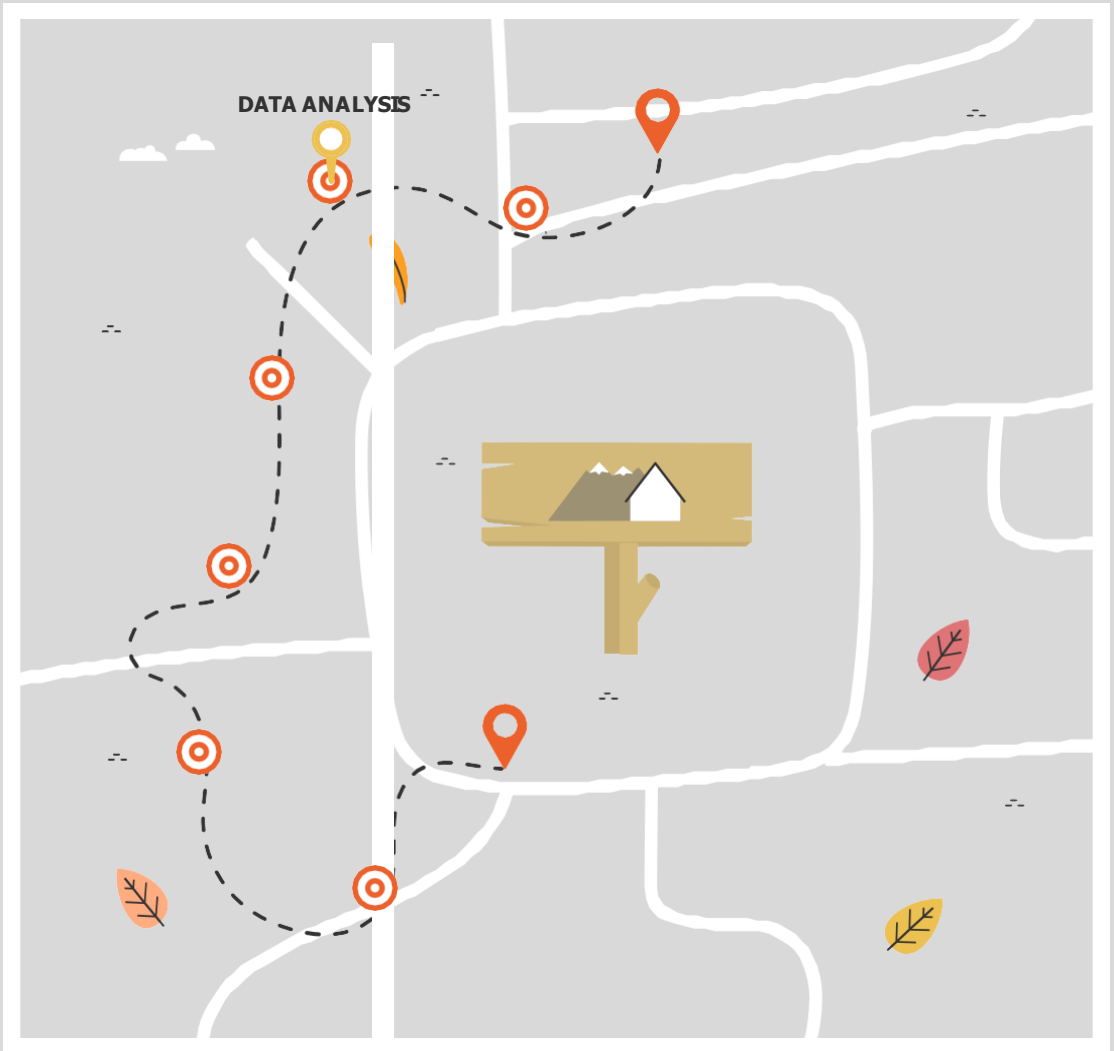
Data Cleaning

Spelling Correction

Vectorization

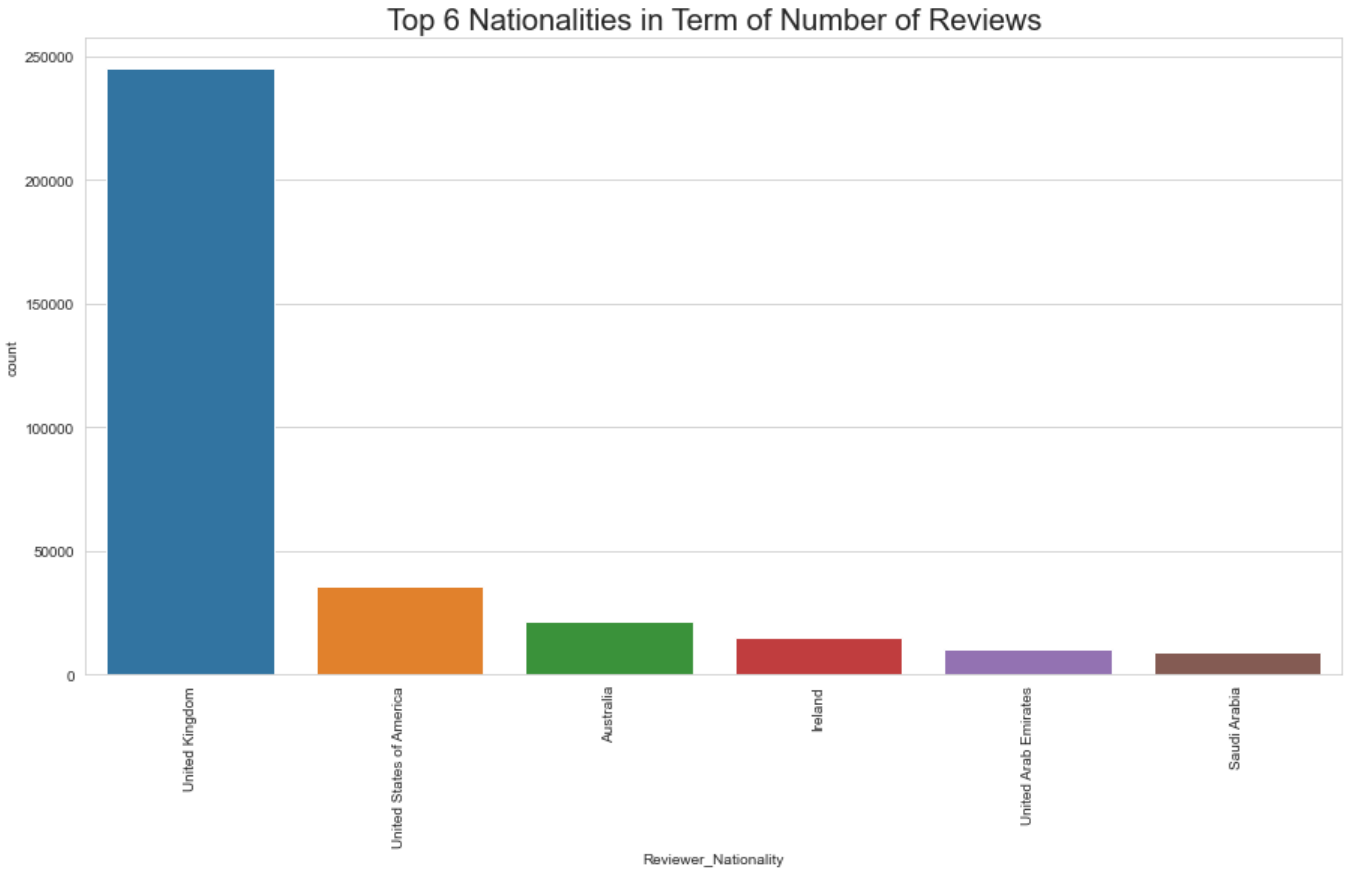
Lemmatization

NLP pipe



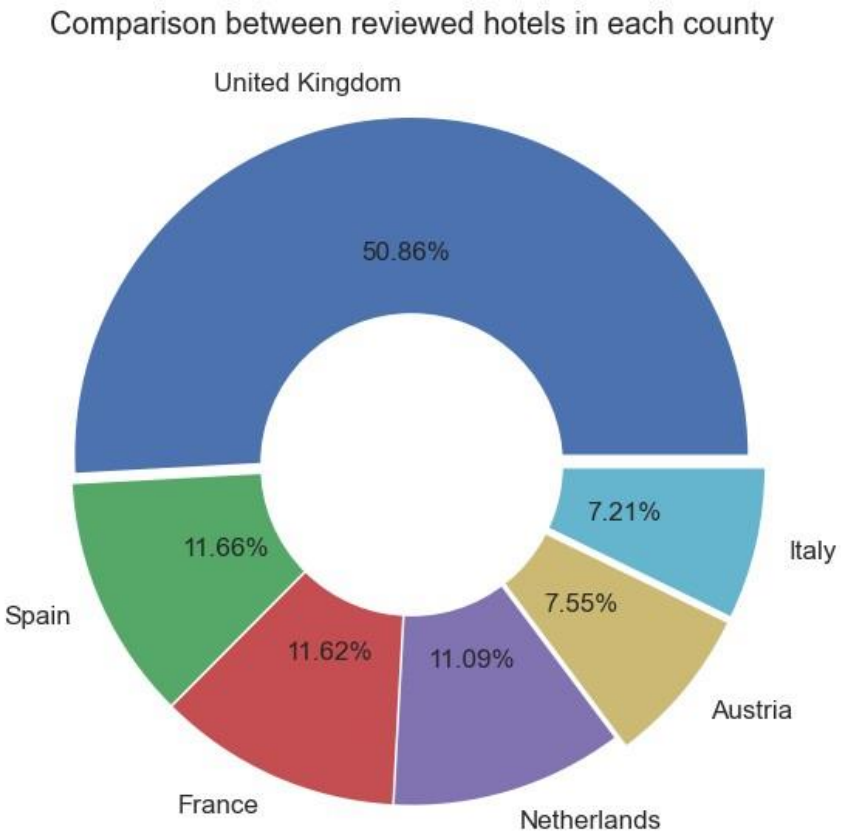
03

DATA ANALYSIS



Top 6 Nationalities

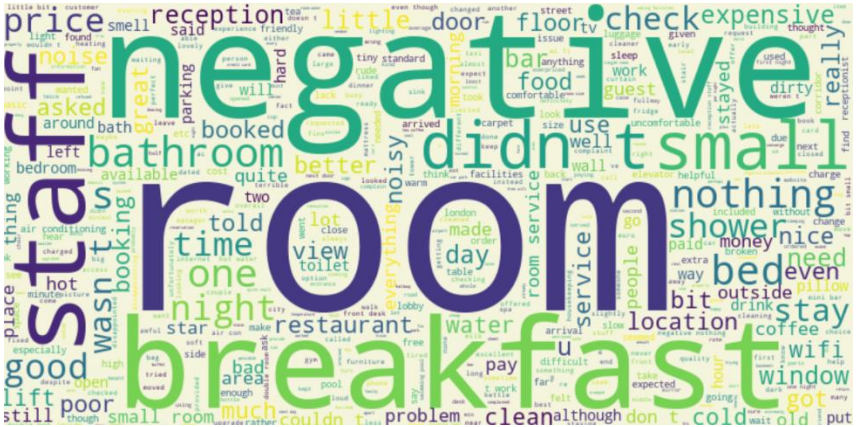
Hotels in Each Country plot



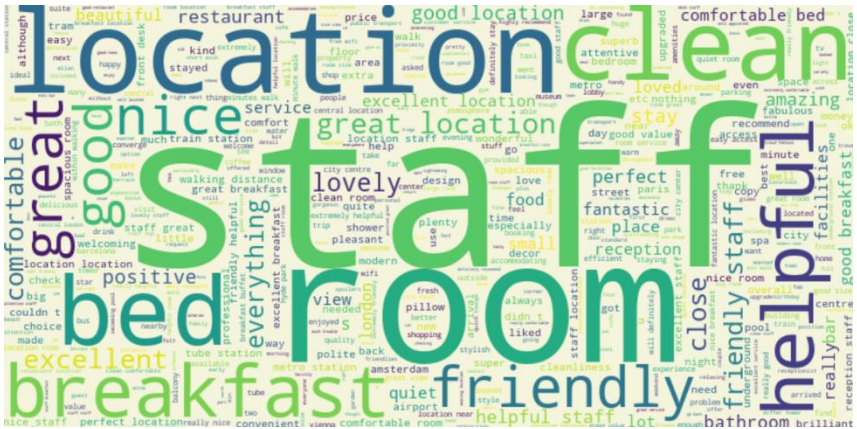
Hotels in Each Country

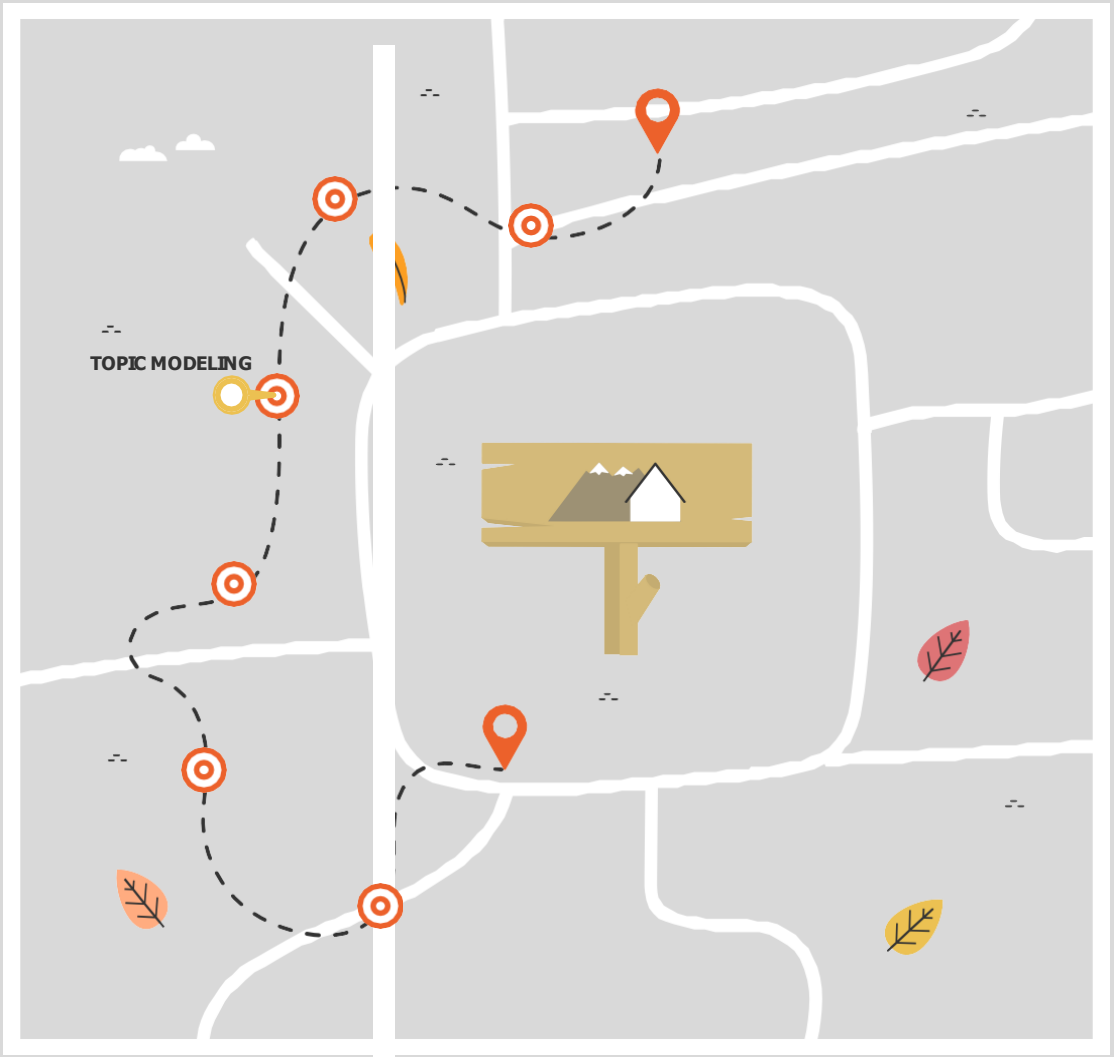
WordCloud Most Common Words

Negative Reviews



Positive Reviews





04

TOPIC MODELING

TOPIC MODELING

LSA

Topic 0

Hotel services

room, staff, location, friendly, helpful, clean, nice, comfortable, excellent, bed.

Topic 1

Room services

room, bed, comfortable, clean, nice, bathroom, spacious, view, big, small.

Topic 2

Hotel Location

location, excellent, station, perfect, close, central, walk, near, tube, nice.

TOPIC MODELING

NMF

Topic 0

Staff service

staff, friendly, helpful, excellent, lovely, clean, reception, extremely, comfortable, amazing

Topic 1

Room description

room, comfortable, clean, bed, lovely, bathroom, spacious, view, quiet, small

Topic 2

location

location, excellent, perfect, station, close, central, walk, near, tube, London

Topic 3

Transportation

nice, station, close, restaurant, bed, walk, mero, area, minute, clean

TOPIC MODELING

CorEx

Topic 0

Transportion

station,close,walk,metro,walking,restaurants,tube,distance,near,walking distance

Topic 1

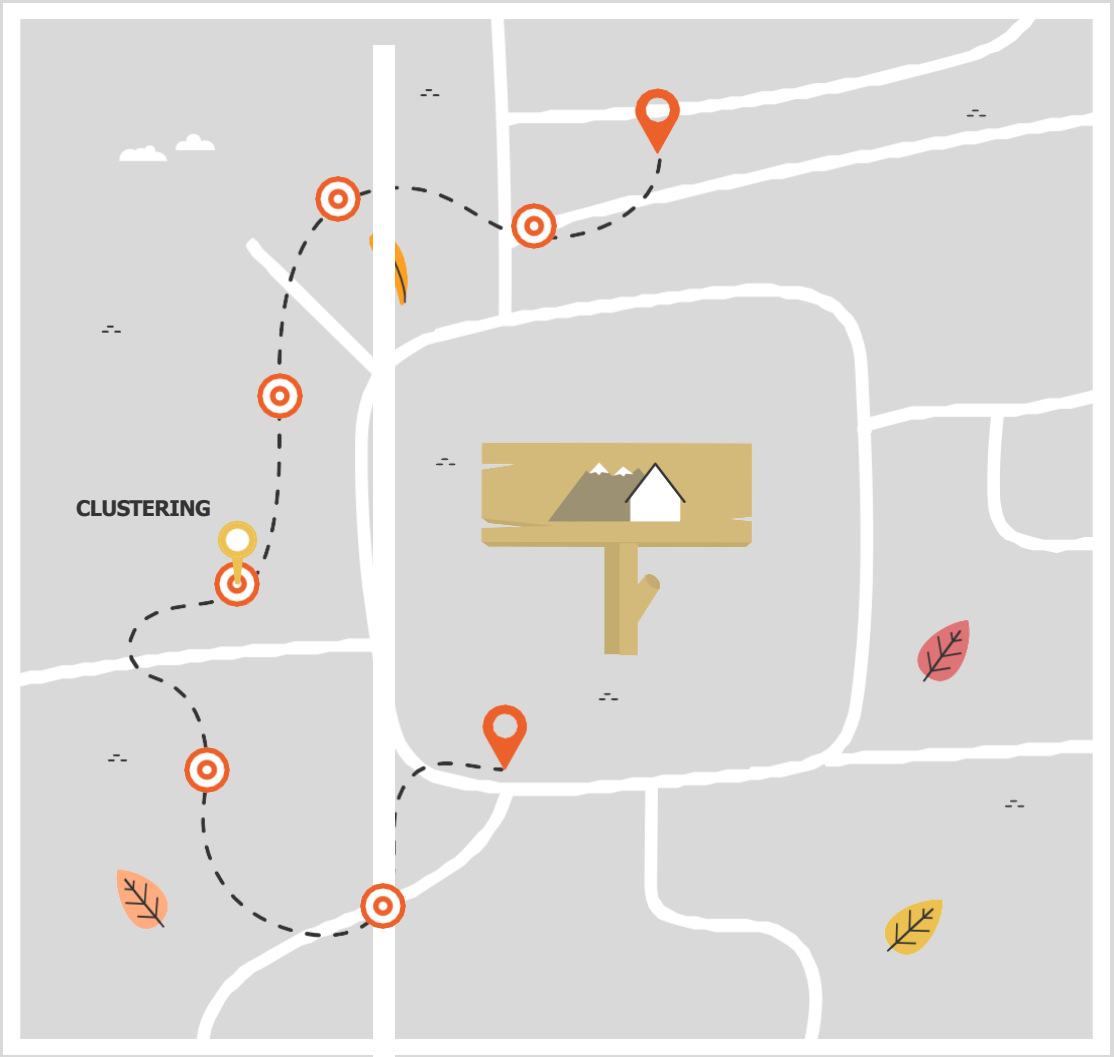
Room service

room,comfortable,clean,bed,rooms,spacious,bathroom,nice,size,room clean

Topic 2

staff service

staff,friendly,helpful,staff friendly,friendly staff,friendly helpful,staff helpful,helpful staff,positive,location staff



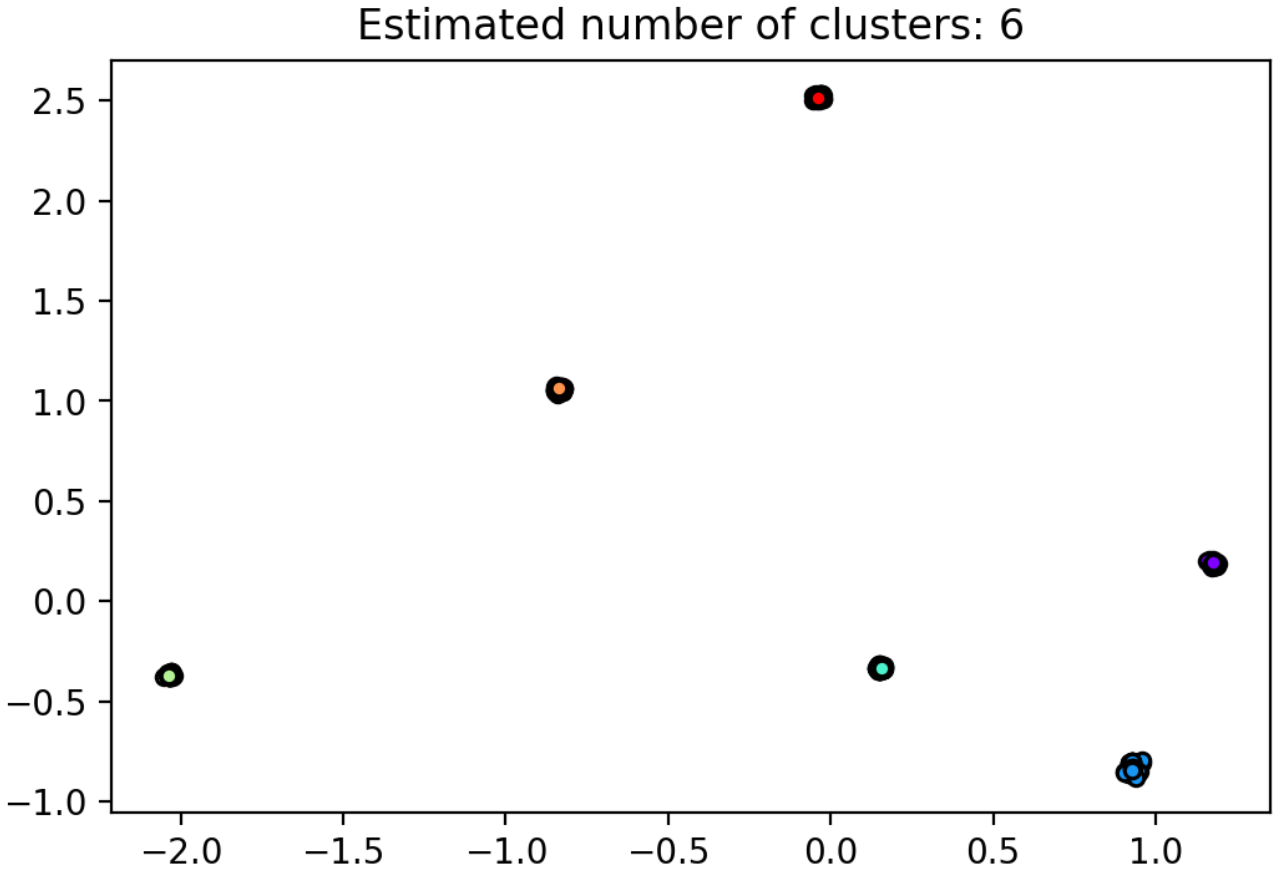
05

CLUSTERING

DBSCAN

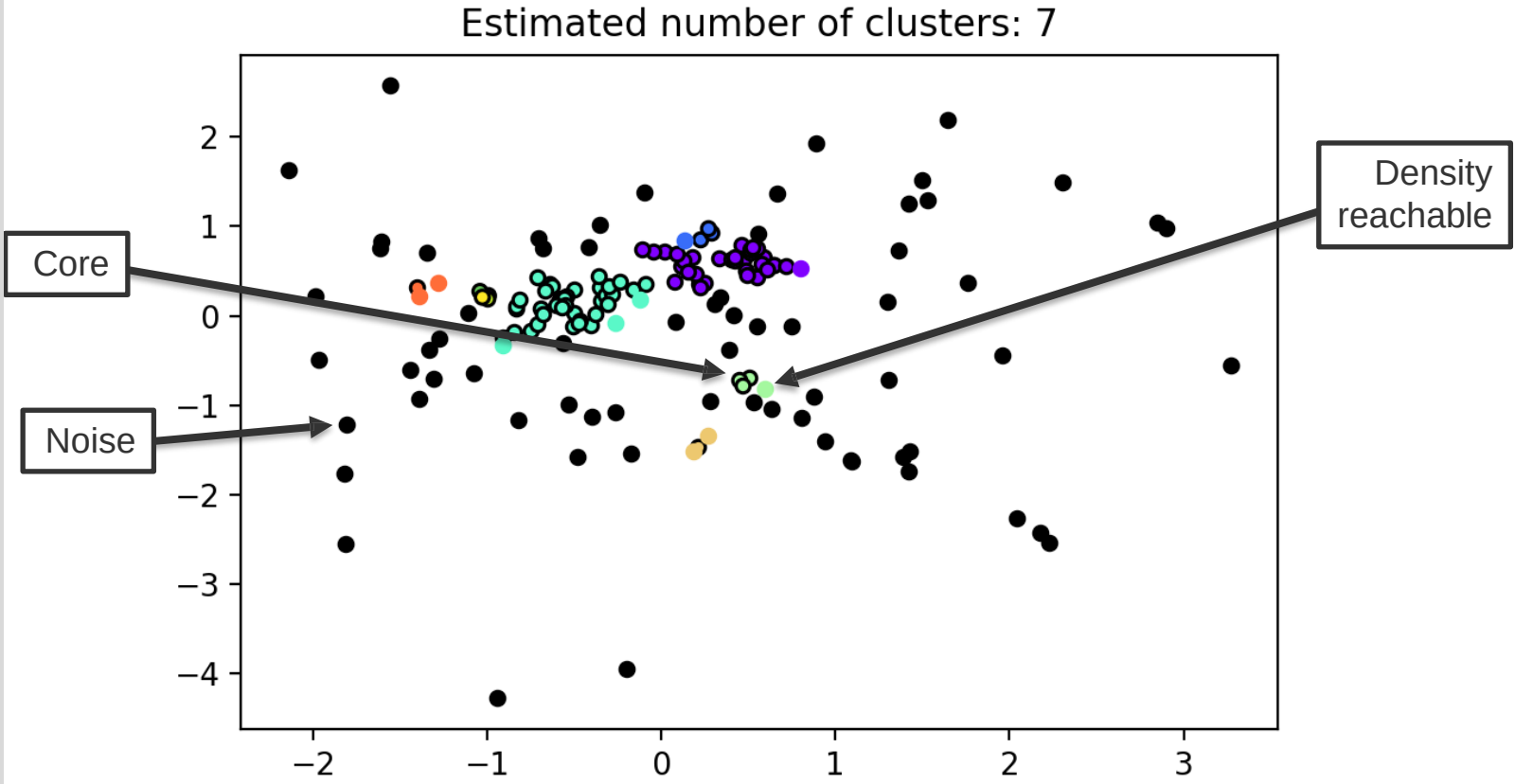
Eps = 0.15

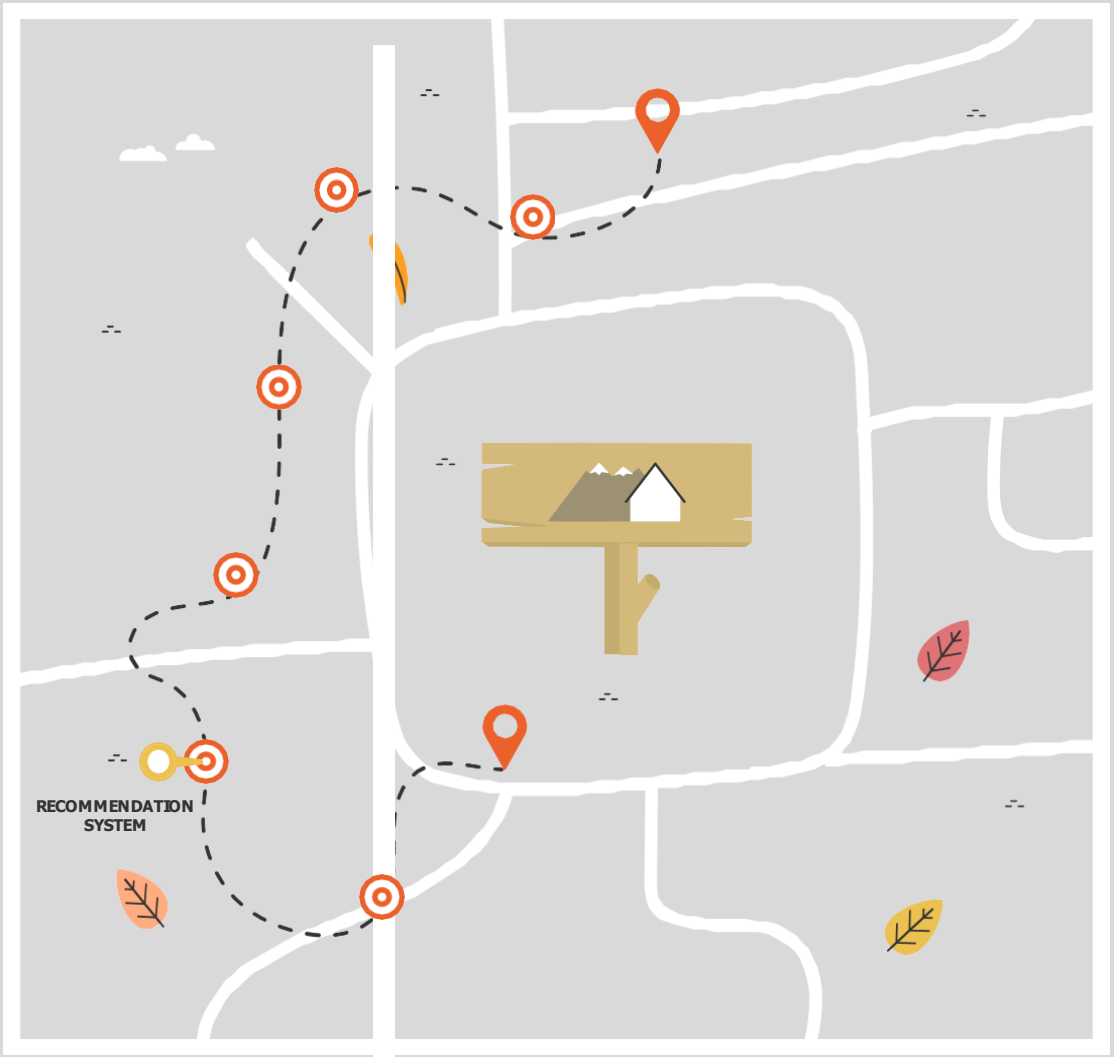
Min samples = 3



Clustering using Lat & Lng For All Data

Clustering using Lat & Lng (Italy)





06

RECOMMENDATION SYSTEM

SVD Recommendation

=========People from Saudi Arabia=========

===Recommended hotels in United Kingdom===
Users from Saudi Arabia are most similar to users from Estonia.
There are 17 hotels that people from Saudi Arabia did not visit, they might like

10 Hotels for people from Saudi Arabia to check out:
['Andaz London Liverpool Street', 'Canary Riverside Plaza Hotel', 'Chiswick Rooms', 'DoubleTree By Hilton London Excel', 'DoubleTree by Hilton London Islington', 'Goodenough Club', 'Grange Fitzrovia Hotel', 'Grange Wellington Hotel', 'Holiday Inn London Oxford Circus', 'Holiday Inn London Whitechapel']

===Recommended hotels in Spain===
Users from Saudi Arabia are most similar to users from Mauritius.
There are 1 hotels that people from Saudi Arabia did not visit, they might like

1 Hotels for people from Saudi Arabia to check out:
['Hotel Granados 83']

Recommendation

Similar Hotels

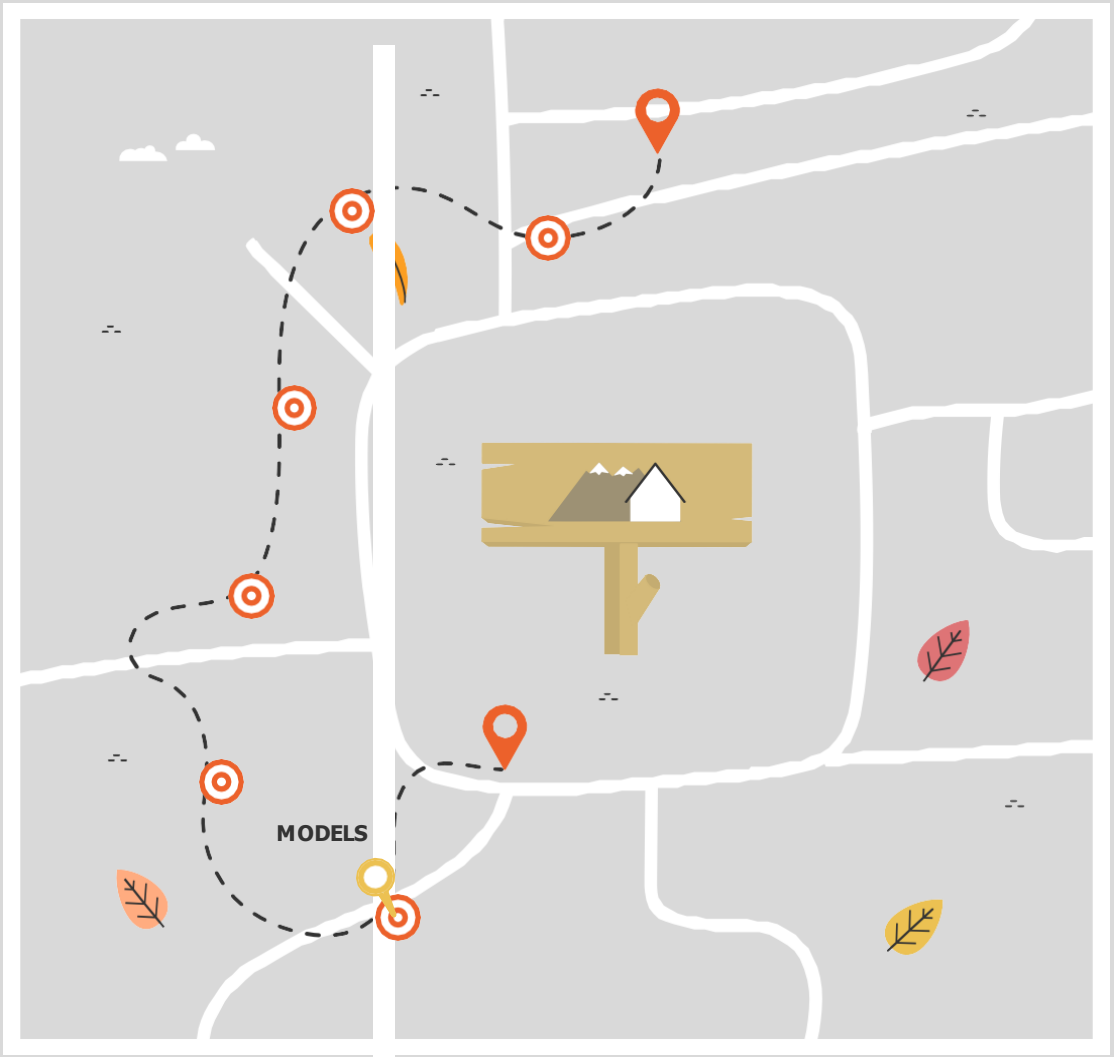
The Kensington Hotel:
Park Grand Paddington Court
Park Plaza Westminster Bridge London
Best Western Premier Hotel Couture

Similar Users

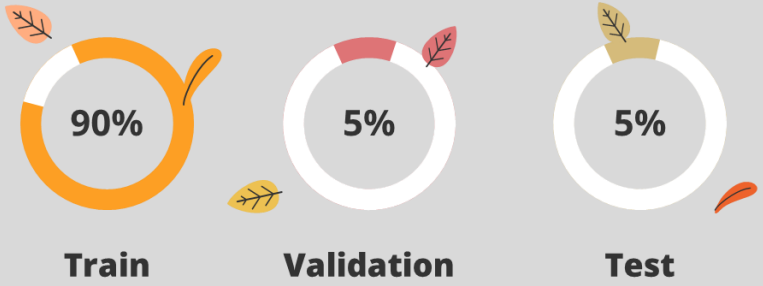
Kuwait:
United Arab Emirates
Saudi Arabia
Canada

Recommended Hotel

Saudi Arabia:
Park Plaza Westminster Bridge London
The Student Hotel Amsterdam City
Best Western Premier Hotel Couture

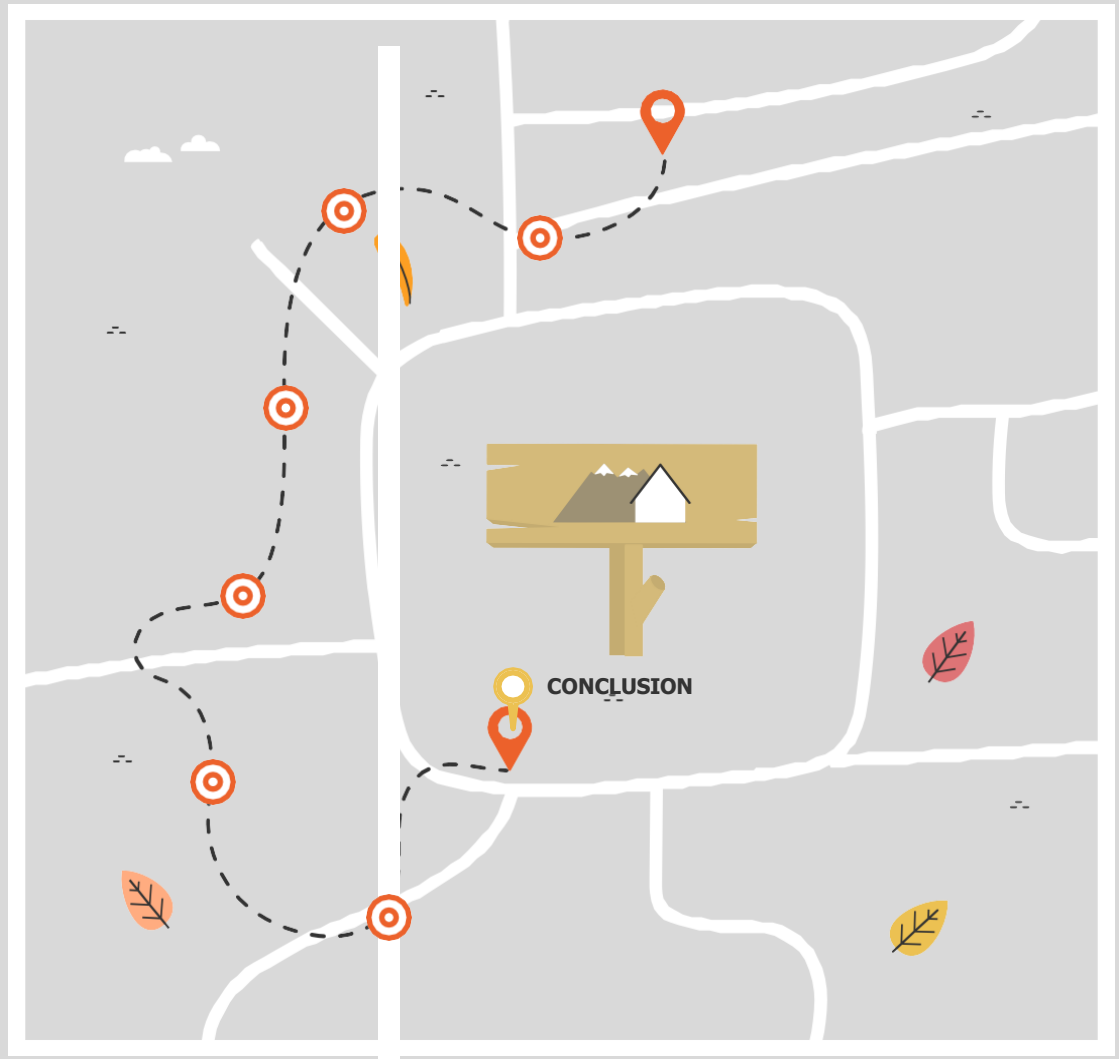


07
MODELS



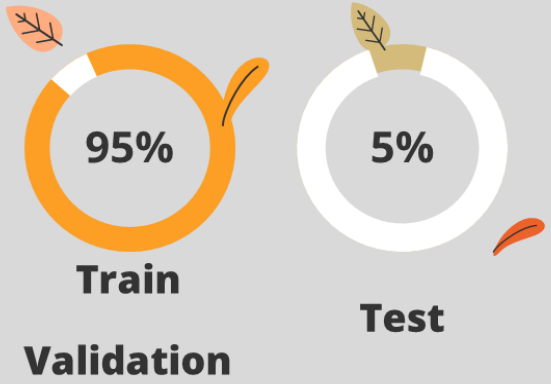
Classification Models

Model	Accuracy	
	Train	Validation
Counter Vectorization		
Logistic Regression	0.952	0.941
Bernoulli NB	0.889	0.877
Decision Tree	0.991	0.907
TF-IDF		
Logistic Regression	0.948	0.942
Bernoulli NB	0.889	0.877
Decision Tree	0.991	0.906



08

CONCLUSION



CONCLUSION

Logistic Regression TF-IDF

Accuracy

Train & validation :	0.948
Test :	0.942

Thank You!

Any Question?