



# Mass tourism can kill a place

## Six reasons why mass tourism is unsustainable

**Global tourism is destroying the environment and cultural identities - and doesn't make good business sense, argues Anna Pollock**



**THE EFFECTS OF MASS TOURISM (HOW OVERTOURISM IS DESTROYING 30+ DESTINATIONS)**

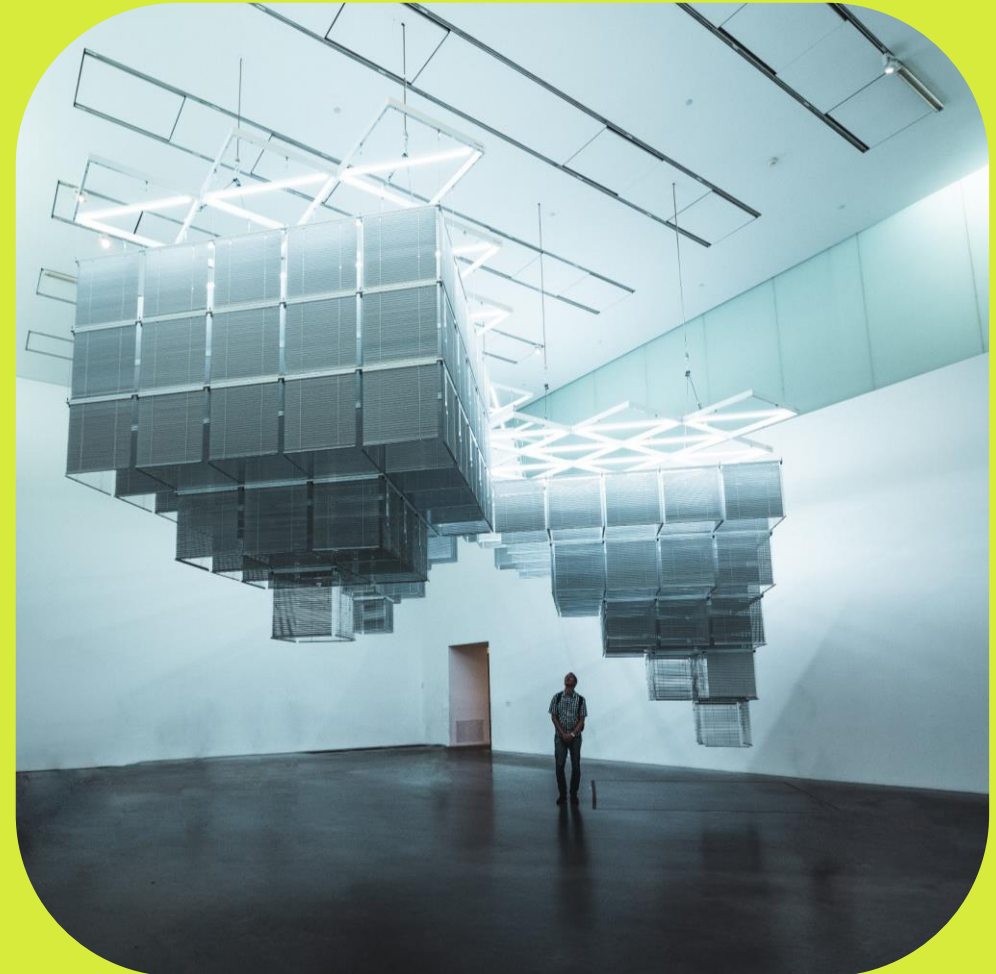


**Pól Ó Conghaile: 'Tourism can kill a place'... so here are three ways we can all travel better in 2023**

**Independent.ie** 



Internet recommends what  
everyone likes



But what if it could recommend  
what **you might love**?



# Mood Travel

SPAIN



## Final Project

Ainara Guerra



# Index

---



The **problem**



The **solution** created



The **data** used



**Obstacles** along the way



The **model**



MVP **examples**



Next **steps**

# The problem

The Spanish government wants to create **a more sustainable tourism paradigm**

They know they can achieve this by promoting the country's tourism among its citizens....

Mass tourism

but... how?

Distributed tourism



# The solution: Put the citizen in the heart of the tourism activity

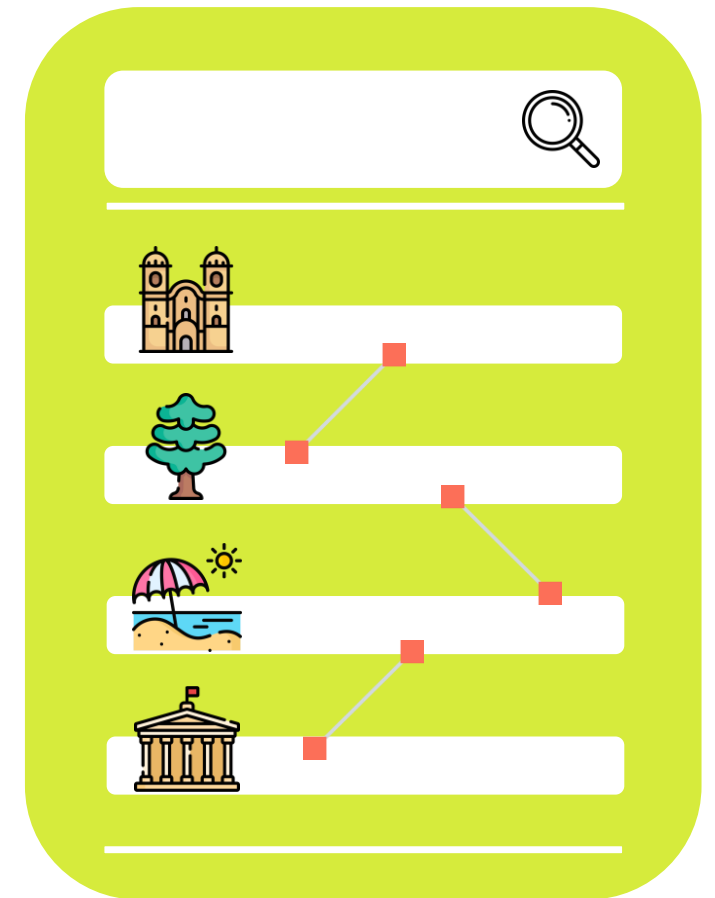


Mass tourism

Listening to what they like and recommending that to them

Distributed tourism

Build a public online platform that will contain a **personalized recommender** that the citizen can use to discover sites for their next trip, according to **their age**, how are they travelling and how are they feeling about this trip.



# Your Personalized Journey Companion

Revolutionary data-driven approach to travel recommendations

## Mission

Connect locals with unforgettable experiences across Spain, tailored to their unique preferences and their mood

“

*Embark on a seamless and soulful journey with Mood Travel Spain and embrace every moment of your adventure in your country, guided by our personalized recommendations*

## Mood Travel Spain

*Embark on a seamless and soulful journey*





# First step: an MVP

The government wants to create an MVP for this recommender with data from some regions







# How did I create it: the dataset

In this dataset, every row is a different site.

Id	Name	Category	Category reduced	Average sentiment	Latitude	Longitude	Address	Rating	Average Age	Average way of travel
Retrieved from datos.gob.es					Retrieved using Google Maps API				Synthetic data	

# The data

Id	Name	Category	Category reduced	Average sentiment	Latitude	Longitude	Address	Rating	Average Age	Average way of travel

Retrieved from [datos.gob.es](https://datos.gob.es)

Spanish Open Data Platform

**No standards when creating tourism data:** Fragmented data and various types of data qualities depending on the region

## Vigo

- all\_vigo.xlsx
- poi-albergues-es.csv
- poi-alojamientos-es.csv
- poi-areas-es.csv
- poi-artesanos-es.csv
- poi-edificios-es.csv
- poi-galerias-es.csv
- poi-hoteles-es.csv
- poi-islas-cies-es.csv
- poi-jardines-es.csv
- poi-lugares-es.csv
- poi-mercados-es.csv
- poi-monumentos-es.csv
- poi-museos-es.csv
- poi-pensiones-es.csv
- poi-pforestais-es.csv
- poi-pit-es.csv
- poi-playas-es.csv

## Cáceres

- monumentos.csv

## Madrid

- Madrid\_xml
- alojamientos\_mad.csv
- comercios\_mad.csv
- deporte\_mad.csv
- idiomas\_mad.csv
- informacion\_turismo\_mad.csv
- nocturno\_mad.csv
- oficina\_turismo\_mad.csv
- poi\_mad.csv

# The data

Id	Name	Category	Category reduced	Average sentiment	Latitude	Longitude	Address	Rating	Average Age	Average way of travel
----	------	----------	------------------	-------------------	----------	-----------	---------	--------	-------------	-----------------------

Unified after reviewing how every city or region categorize their sites

Categories created after reviewing types of tourism

Categories reduced to serve de ML model

## Result

More than 5.000 sites and tourist attractions





# The data

Id	Name	Category	Category reduced	Average sentiment	Latitude	Longitude	Address	Rating	Average Age	Average way of travel
----	------	----------	------------------	-------------------	----------	-----------	---------	--------	-------------	-----------------------

Retrieved using Google Maps API

+5.000 sites and tourist attractions

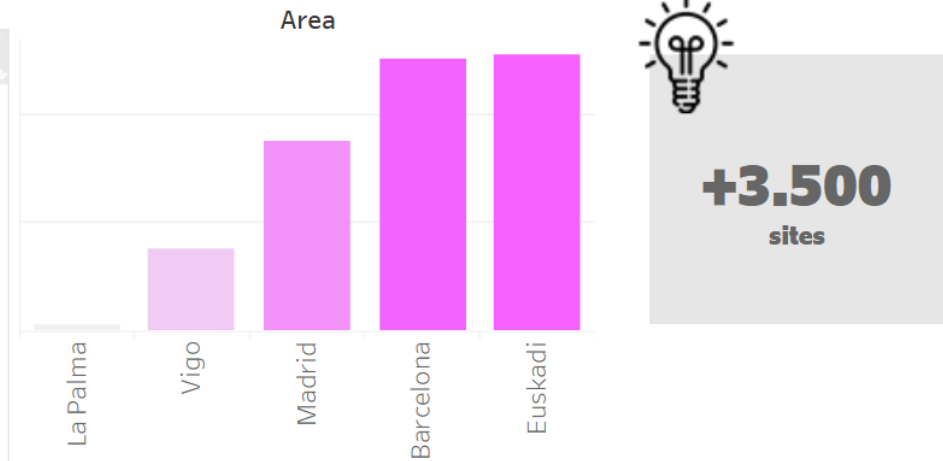
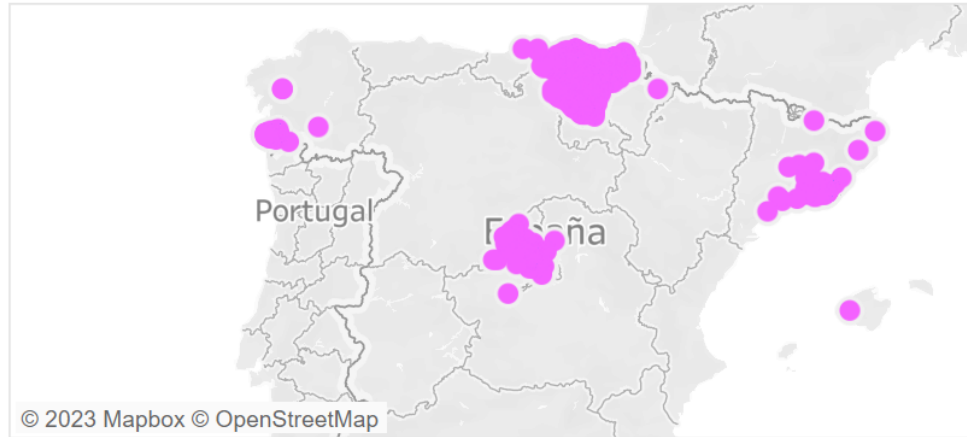
Google Maps API

**+3.500 sites and tourist attractions**

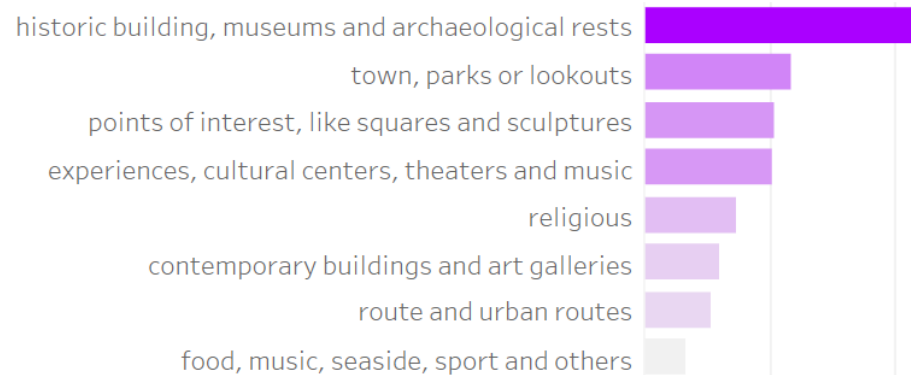


## Main data of the personalized recommender for tourism in Spain

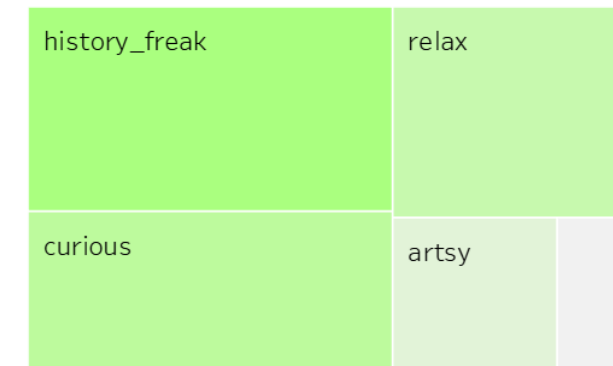
### Distribution per area



### Distribution per category of the site



### Distribution per sentiment in the site






# The data


Id	Name	Category	Category reduced	Average sentiment	Latitude	Longitude	Address	Rating	Average Age	Average way of travel
									Synthetic data	

## PROBLEM


**Current travel recommenders are not very personalized...**




**Palacio Real de Madrid**  
●●●●○ 35.274 opiniones  
Plaza De Oriente S/n, Madrid, España  
2.919 menciones de madrid  
"...Para mi una visita imprescindible si vas a Madrid..."



**SANDEMANs New Europe - Madrid**  
●●●●○ 21.026 opiniones  
Plaza Mayor, Madrid, España  
6.428 menciones de madrid  
"...quien nos ha acompañado con mucho entusiasmo en nuestro paseo por el Madrid..."



**Madrid Metro**  
●●●●○ 9.066 opiniones  
Madrid, España  
2.031 menciones de madrid  
"...El estadio precioso, actualmente en remodelación Hasta pronto Madrid..."



**Parque Warner Madrid**  
●●●●○ 7.029 opiniones  
Autovia A-4, Exit 22, San Martín de la Vega, España  
543 menciones de madrid  
"...Al igual que el Parque de Atracciones de Madrid, se me queda un poco corto..."

When you are preparing a trip, you encounter random rankings and “one size fits all” results

# The data

Id	Name	Category	Category reduced	Average sentiment	Latitude	Longitude	Address	Rating	Average Age	Average way of travel
----	------	----------	------------------	-------------------	----------	-----------	---------	--------	-------------	-----------------------

To make this recommender personalized we really needed features of the user, that we couldn't scrape in the web because this information:

- Doesn't exist.
- If exists, isn't public

## Three options to create it:

### Random values

✗ Useless for ML

### Heavily imbalanced data

● 0.95 of accuracy in some ML models  
✗ Not ethical, not close to reality

### Clustered and slightly imbalanced data

● 0.78-0.81 of accuracy in best ML models for this case  
● Closer to reality

# The model

## Input.



## Model & outputs.

**TARGET:** categories of sites

**1.**

Selects the category of sites that better suits these features based on multiclass ML models.

For this, it is used **Random Forest (0.78 accuracy)**.

**2.**

Selects top three sites of this category based on rating in the destination the user selected.



Project by Nara Guerra 🧐

## Find the best places with this personalised browser

Enter your preferences to get the site category you can't missed in Spain and top places to visit.

Select your mood for this travel:

Select your age group:

Select your way of travel:

Select your destination in Spain:

Rate for this experience (1-5):

3

15

Get Recommendations

Spain made for you

Discover it!



# MVP next steps



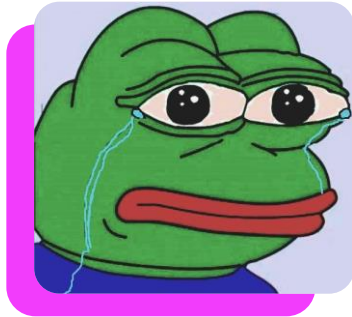
Improve the  
model



Make it usable  
for NGOs related  
to tourism and  
accessibility







# Thank you, Ironhack

This experience is unforgettable  
I'll miss u all