



# Data analysis of a food community

---

Maria Luisa Balacco

# Index

---

- [Dataset description](#)
- [Data cleaning](#)
- [Database structure](#)
- [Database legend](#)
- [SQL queries](#)
- [SQL analysis conclusions](#)
- [EDA with Python](#)
- [Conclusions EDA with Python](#)
- [Dashboard overall trend](#)
- [Buyer Persona](#)
- [Final conclusion](#)
- [Links to the full code](#)



# Dataset description

---

The dataset I analyzed has the insights of an online community that targets plant-based cuisine lovers.

Every day expert chefs share plant-based recipes which are easy to do at home. The users can repeat those recipes and publish the results on their page to inspire the other members of the community.

Users can also take part in themed challenges, where they gain points per each recipe they cooked and shared. The challenges lasts from 4 to 12 weeks and can be done at any moment, giving flexibility and fun in improving cooking skills and attracting more interest in a sustainable and aware approach to food.

# Data Cleaning

---

On Jupiter Notebook:

- null values have been checked,
- the unique values have been identified,
- entries names have been changed from the column country for a better legibility,
- the columns cooking\_date and subscription\_date have been changed in DATE format.

The dataset was now ready to be imported in a relational database and then start the analysis with the first queries.

The clean data have been imported on the multiplatform web server Xampp on which has been launched MySQL: a database service opensource for relational databases. The editor Beekeeper Studio has been used to create the relational database and the data analysis.

# Database structure

## Users

ID_U	user_uuid	subscription_date	language	country	platform	cooking_date	chef_id	challenge_id
🔑								

## Recipes

ID_R	recipe_id	level	category	stars	chef_id	challenge_id
🔑						



# Database legend

---

cooking\_date

Date when the users posted their recipe version

user\_uuid

User identification number

category

Recipes category

chef\_id

Chef identification number

language

User language

level

Recipes difficulty

challenge\_id

Challenge identification number

country

User nationality

subscription\_date

Day of subscription of the user

platform

Platform from which the user watched the video

recipe\_id

Recipe identification number

stars

Rating given to the recipe  
(from 1 to 5)

# SQL queries

Years when the users subscribed = 2020, 2021, 2022

Total number of users who subscribed each year per country of origin:

2020

	country ▲	TotalUsers ▼
1	Italy	64
2	UnitedKingdom	63
3	France	49

2021

	country ▲	TotalUsers ▼
1	Italy	71
2	UnitedKingdom	67
3	France	54

2022

	country ▲	TotalUsers ▼
1	Italy	65
2	UnitedKingdom	54
3	France	43

In 2021 there was a growth, but in 2022 subscriptions decreased.  
The majority of the users comes from Italy.



# SQL queries

Years when the recipes have been posted = 2021, 2022, 2023

The user interaction with the recipes has been stable during the years, even though the number of users increased

	year ▲	TotalRecipes ▲	total_challenges ▲	total_users ▲
1	2023	593	4	216
2	2022	593	4	131
3	2021	593	4	62

The recipes which received a rating of 4 and 5 stars are 327, on a total of 593 recipes (according to the unique values calculated with Pandas).

All the recipes posted per category

	category ▲	TotalRecipes ▼
1	ethnic	198
2	breakfast	126
3	lunch	119
4	dinner	95
5	snack	66
6	vegan	65
7	smoothie	62



# SQL queries

Here is the total number of recipes with 4 and 5 stars grouped by category and level. The most successful categories are: ethnic, breakfast, lunch.

	category ▲	level ▲	TopRecipes		category ▲	level ▲	TopRecipes		category ▲	level ▲	TopRecipes ▼
1	ethnic	basic	39	1	ethnic	intermed...	29	1	ethnic	advanced	48
2	breakfast	basic	25	2	lunch	intermed...	20	2	lunch	advanced	29
3	snack	basic	14	3	breakfast	intermed...	11	3	breakfast	advanced	24
4	dinner	basic	12	4	dinner	intermed...	8	4	vegan	advanced	18
5	lunch	basic	11	5	smoothie	intermed...	8	5	dinner	advanced	17
6	smoothie	basic	11	6	vegan	intermed...	8	6	smoothie	advanced	13
7	vegan	basic	9	7	snack	intermed...	4	7	snack	advanced	12

# SQL queries

These are the top 10 chef who posted more recipes and with the best rating in the most popular recipe categories.

	category ▲	chef_id ▲	TopRecipes ▼
1	ethnic	108	13
2	ethnic	105	12
3	ethnic	120	12
4	ethnic	118	10
5	ethnic	112	9
6	ethnic	100	7
7	breakfast	107	7
8	breakfast	106	7
9	ethnic	121	7
10	ethnic	101	7

We can now analyze the users who watched the recipes with the highest rating. These are the favourite categories of the users from Italy.

	country ▲	category ▲	level ▲	UserOrigin ▼
1	Italy	ethnic	basic	200
2	Italy	lunch	intermediate	200
3	Italy	dinner	advanced	194
4	Italy	smoothie	basic	191
5	Italy	breakfast	basic	190
6	Italy	vegan	intermediate	189
7	Italy	snack	advanced	172



# SQL queries

And those of users from France.

	country ▲	category ▲	level ▲	UserOrigin ▼
1	France	ethnic	intermediate	146
2	France	lunch	basic	146
3	France	dinner	basic	136
4	France	smoothie	advanced	134
5	France	vegan	intermediate	133
6	France	breakfast	advanced	130
7	France	snack	advanced	123

And those of users from United Kingdom.

	country ▲	category ▲	level ▲	UserOrigin ▼
1	UnitedKingdom	ethnic	advanced	184
2	UnitedKingdom	lunch	intermediate	184
3	UnitedKingdom	dinner	advanced	177
4	UnitedKingdom	breakfast	advanced	170
5	UnitedKingdom	vegan	basic	168
6	UnitedKingdom	smoothie	intermediate	165
7	UnitedKingdom	snack	basic	158

These same queries have been repeated including the user languages: in all the dataset languages the results are the same of the previous ones which took into account only country and not country + language.

# SQL queries

---

We deduct that the most active users come from Italy, followed by United Kingdom and France. All the country origins share the same favourite recipes per rating in the same categories with this order, but for different levels: ethnic, lunch, dinner.

- In Italy the preferences for advanced recipes are only for dinner
- In France there are not favourite advanced recipes in the first 3 categories but only for the others: smoothie, breakfast, snack
- In United Kingdom both ethnic recipes and dinner stand out in the advanced level for the amount of preferences



# SQL queries

Also users preferences have been analyzed taking into account all recipes posted, regardless of the rating.

	country ▲	category ▲	level ▲	UserOriginTotal ▼
1	Italy	dinner	advanced	242
2	Italy	ethnic	advanced	242
3	UnitedKingdom	ethnic	basic	242
4	France	breakfast	advanced	241
5	Italy	breakfast	basic	241
6	UnitedKingdom	ethnic	intermediate	241
7	France	lunch	advanced	241
8	UnitedKingdom	lunch	intermediate	241
9	Italy	dinner	basic	240
10	Italy	snack	advanced	240

11	UnitedKingdom	vegan	advanced	237
12	UnitedKingdom	lunch	basic	236
13	Italy	smoothie	basic	231
14	UnitedKingdom	smoothie	advanced	228
15	UnitedKingdom	breakfast	intermediate	227
16	UnitedKingdom	snack	basic	225
17	Italy	smoothie	intermediate	224
18	France	dinner	intermediate	222
19	UnitedKingdom	vegan	basic	220
20	UnitedKingdom	snack	intermediate	199
21	Italy	vegan	intermediate	194

# SQL queries

---

Also from the overall statistics we deduct that the most active users come from Italy, followed by United Kingdom and France.

But here the 3 favourite categories are dinner, ethnic, breakfast.

Even though in United Kingdom the favourite recipes were the ethnic advanced, without taking account the rating are the ethnic basic.

For a in-depth study, the data of this same query have been reordered with the addition of the cooking\_date to analyze the preferences during the subscription years.



# SQL queries

	cooking_date ▲	country ▲	category ▲	level ▲	UserPreferenceTime ▲						
1	2021-08-01	UnitedKingdom	vegan	basic	220	12	2023-04-25	France	snack	basic	225
2	2022-02-04	Italy	dinner	intermediate	222	13	2023-05-22	Italy	breakfast	basic	241
3	2022-04-09	France	snack	intermediate	199	14	2023-08-11	Italy	vegan	intermediate	194
4	2022-06-14	Italy	ethnic	basic	242	15	2023-08-15	UnitedKingdom	vegan	advanced	237
5	2022-11-19	France	ethnic	advanced	242	16	2023-08-20	Italy	dinner	basic	240
6	2022-12-17	UnitedKingdom	dinner	advanced	242	17	2023-09-19	Italy	smoothie	intermediate	224
7	2023-01-14	UnitedKingdom	breakfast	advanced	241	18	2023-09-27	Italy	breakfast	intermediate	227
8	2023-02-11	Italy	smoothie	basic	231	19	2023-10-01	Italy	lunch	advanced	241
9	2023-02-18	UnitedKingdom	snack	advanced	240	20	2023-10-09	UnitedKingdom	ethnic	intermediate	241
10	2023-03-27	Italy	smoothie	advanced	228	21	2023-11-28	France	lunch	basic	236
11	2023-04-24	Italy	lunch	intermediate	241						

During the years the same preferences per nationality changed especially in United Kingdom. But the users from Italy and United Kingdom have always been the most active ones.

# SQL queries

Here are the statistics of the challenges per platform used in each country.

	challenge_id ▲	platform ▲	country ▲	UserChallenge ▼
1	2	tv	Italy	52
2	3	tv	Italy	50
3	1	tv	UnitedKingdom	45
4	3	pc	Italy	42
5	0	tv	Italy	42
6	1	tv	Italy	39
7	3	tv	UnitedKingdom	38
8	0	tv	UnitedKingdom	34
9	2	pc	Italy	33
10	1	pc	Italy	31
11	1	tv	France	29
12	3	pc	UnitedKingdom	29
13	1	mobile	Italy	28
14	3	mobile	Italy	27
15	2	pc	UnitedKingdom	27

Also here, we deduct that the most active users come from Italy and United Kingdom.

The most chosen platform is the tv.



# SQL analysis conclusions

---

The most active users in the interaction with recipes come from Italy and United Kingdom for the recipes with any rating (but are also the subscribers with the highest number).

The preferences per category of recipes changed during the years especially in United Kingdom.

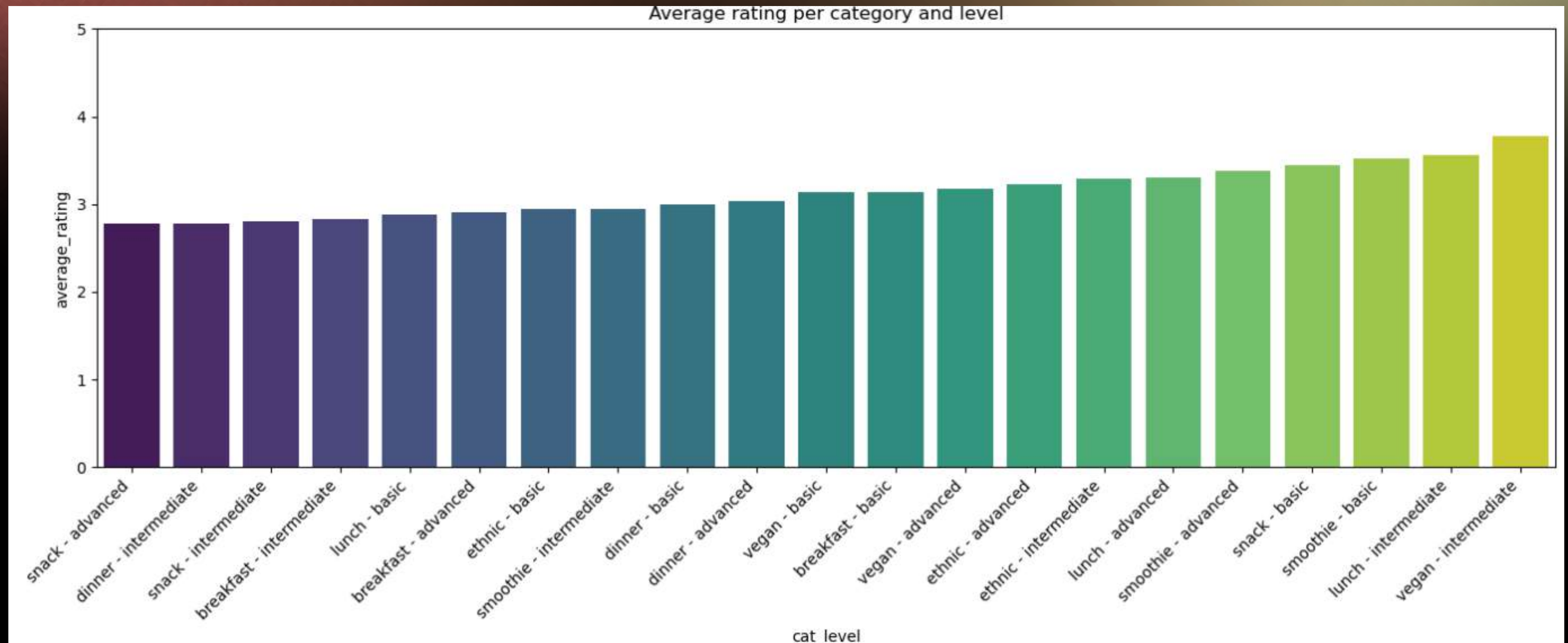
The most published categories of recipes are ethnic and breakfast, which are also the most successful together with the lunch category.

The rating per difficulty of recipe changes according to the country of origin.

The only stable data throughout the whole analysis is the one of the tv as the most chosen channel for the challenges.

# EDA with Python

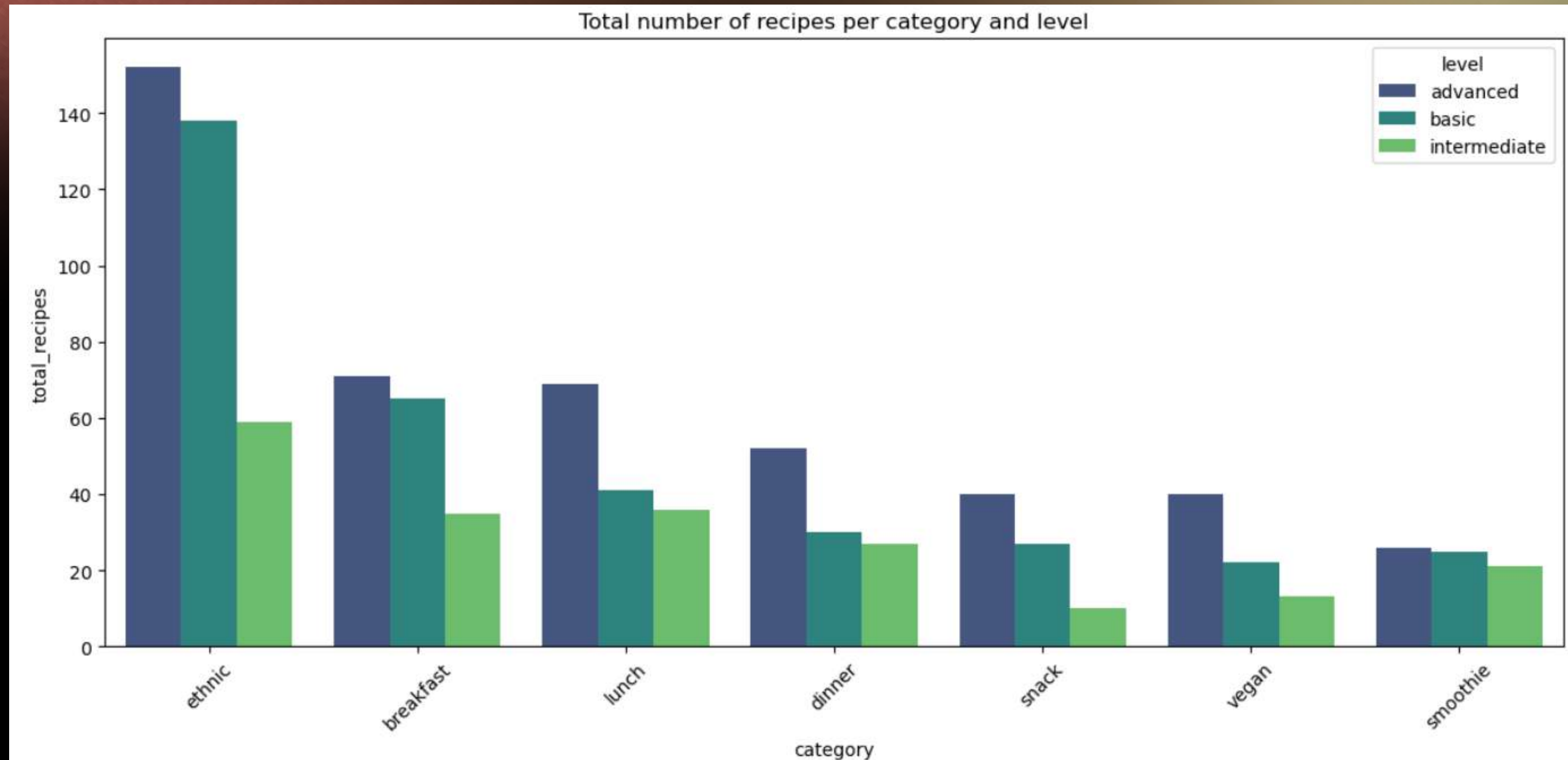
Analysis of the average rating per category and level





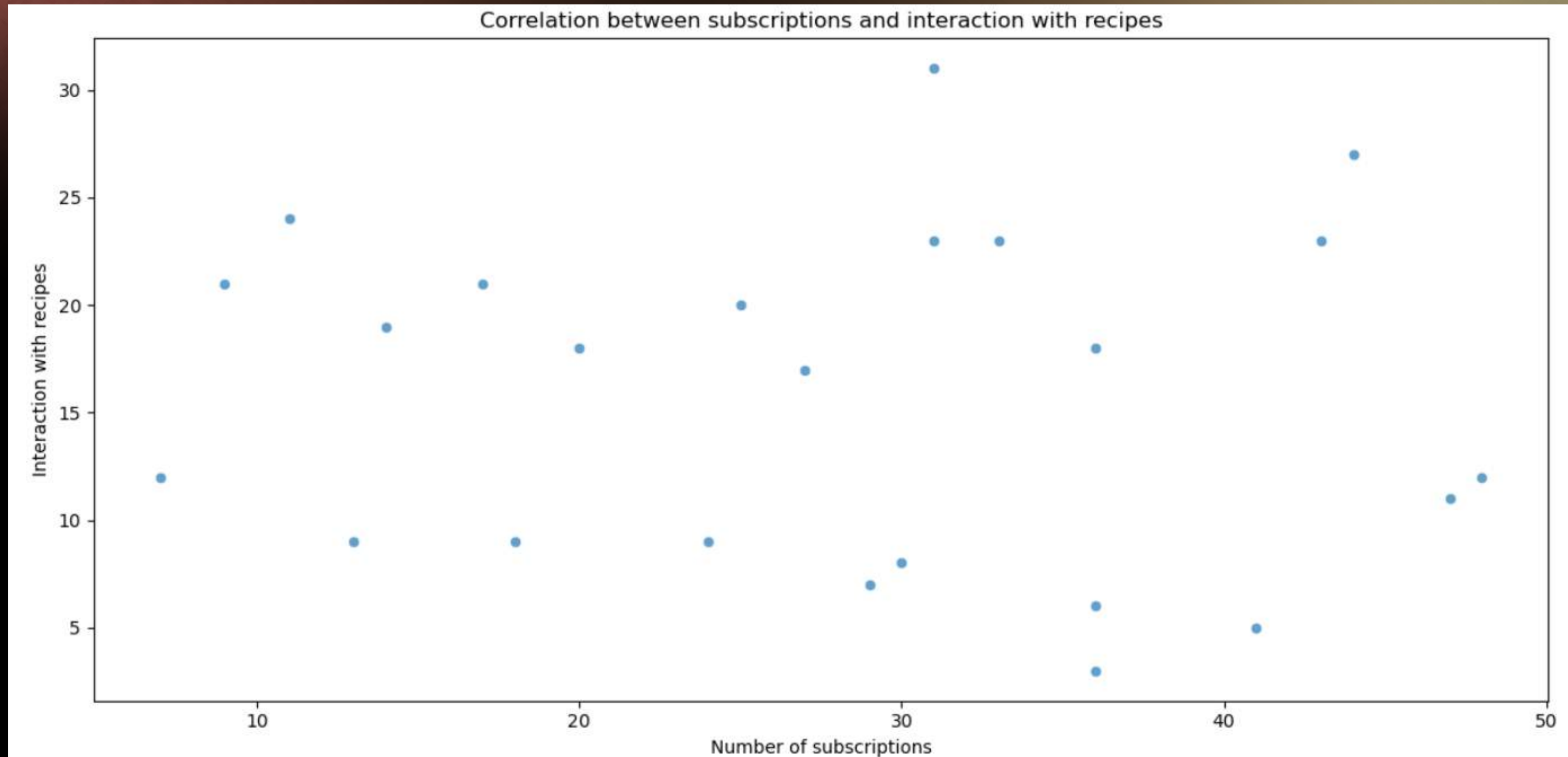
# EDA with Python

Total number of recipes per category and level



# EDA with Python

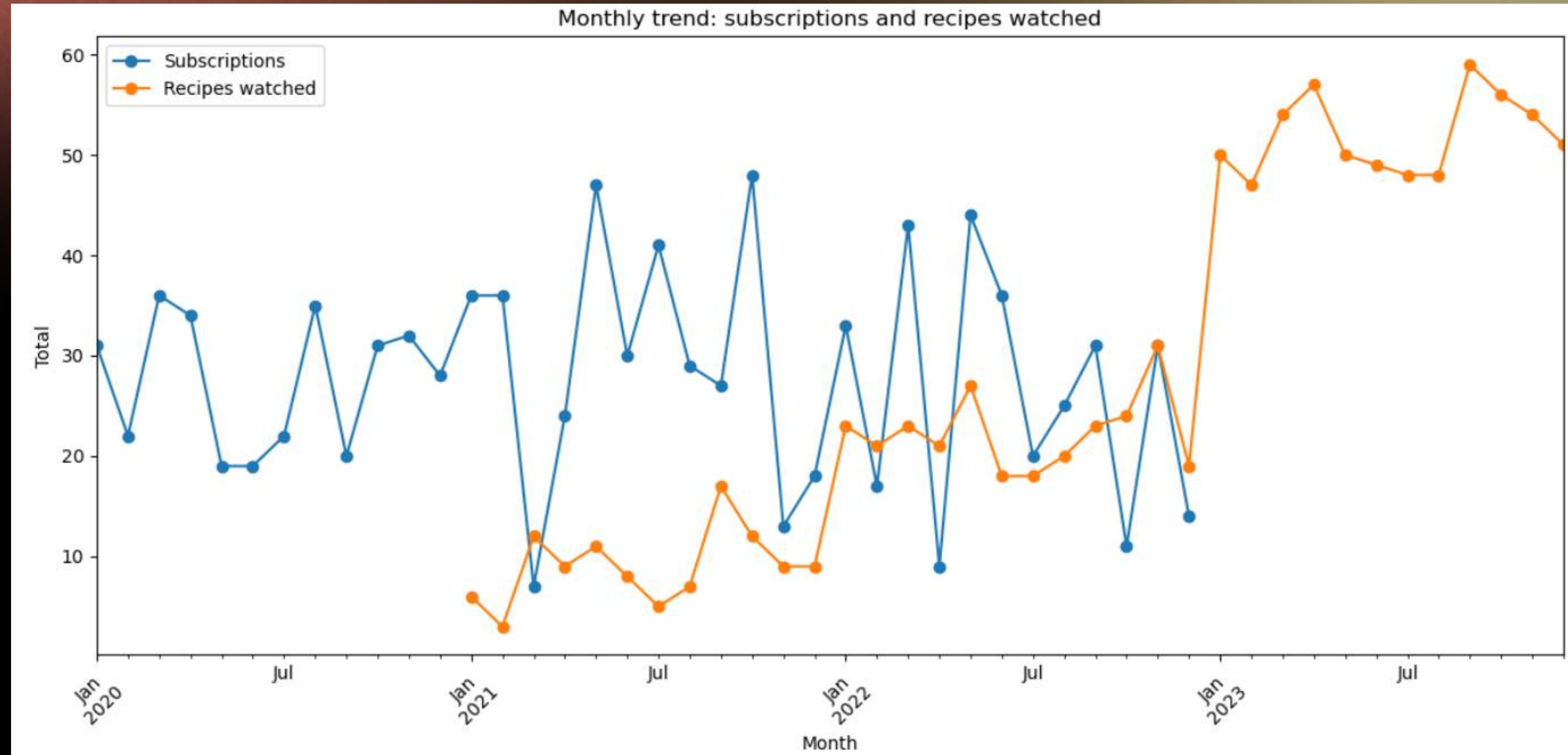
Scatterplot to check if during the months with more subscriptions there are also more interactions with the recipes → there is no correlation





# EDA with Python

Graphic representation of the monthly trend of subscriptions and recipes watched





# EDA with Python – Conclusions

---

This analysis shows a new insight for the categories vegan and smoothie: are the ones with the lowest number of posted recipes but also the ones the highest average rating in all the categories.

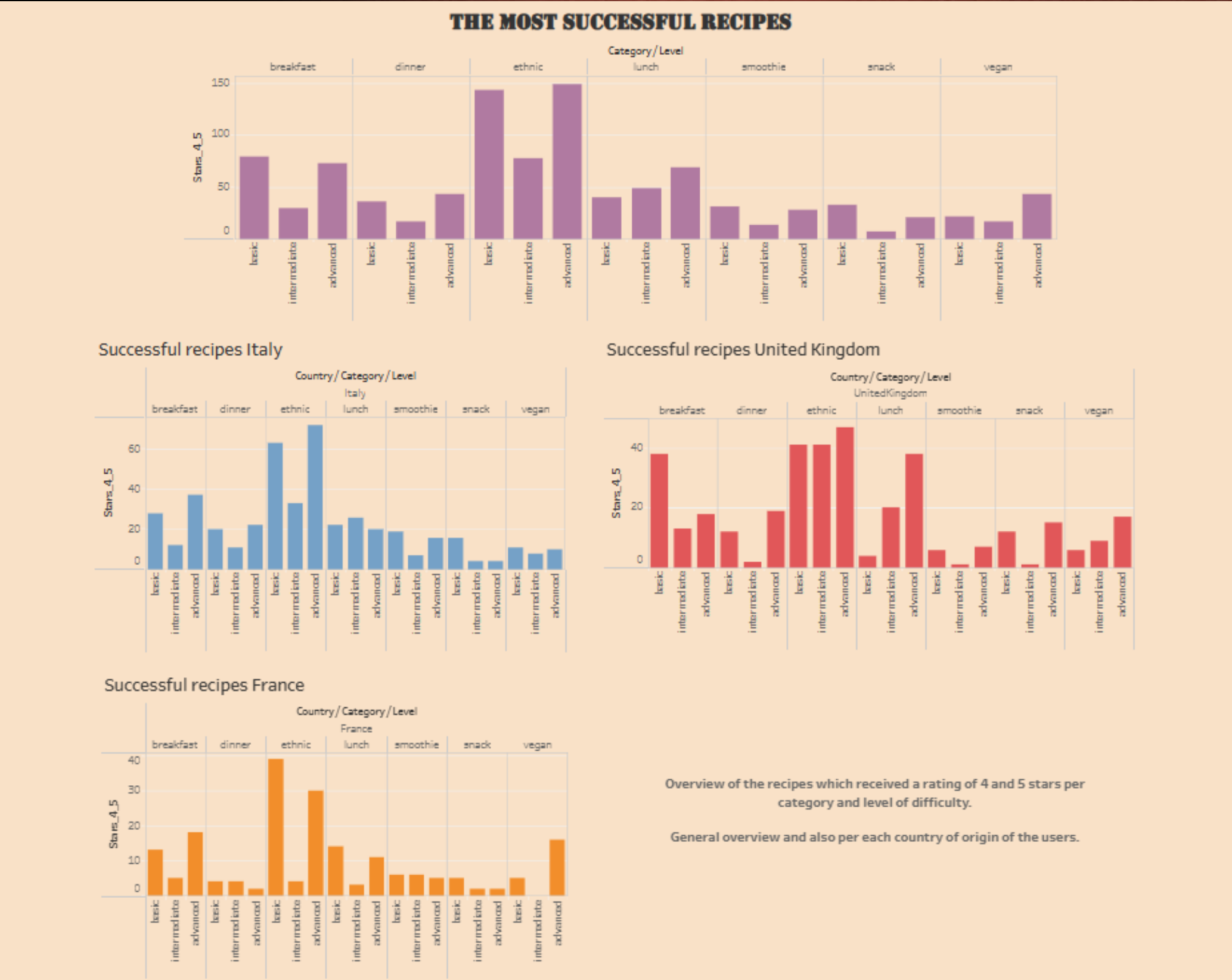
The ethnic category is once again the favourite one at the levels intermediate and advanced.

During the months with more subscriptions there is not more interaction with the recipes. It might be that the old users are the most active and the new users take more time to start engaging.

In the dataset there is a gap of information. The subscriptions record stop at 2022, whereas the interaction with the recipes continues until 2023. Also, apparently in the first year of the platform (2020) there has been no interaction at all.



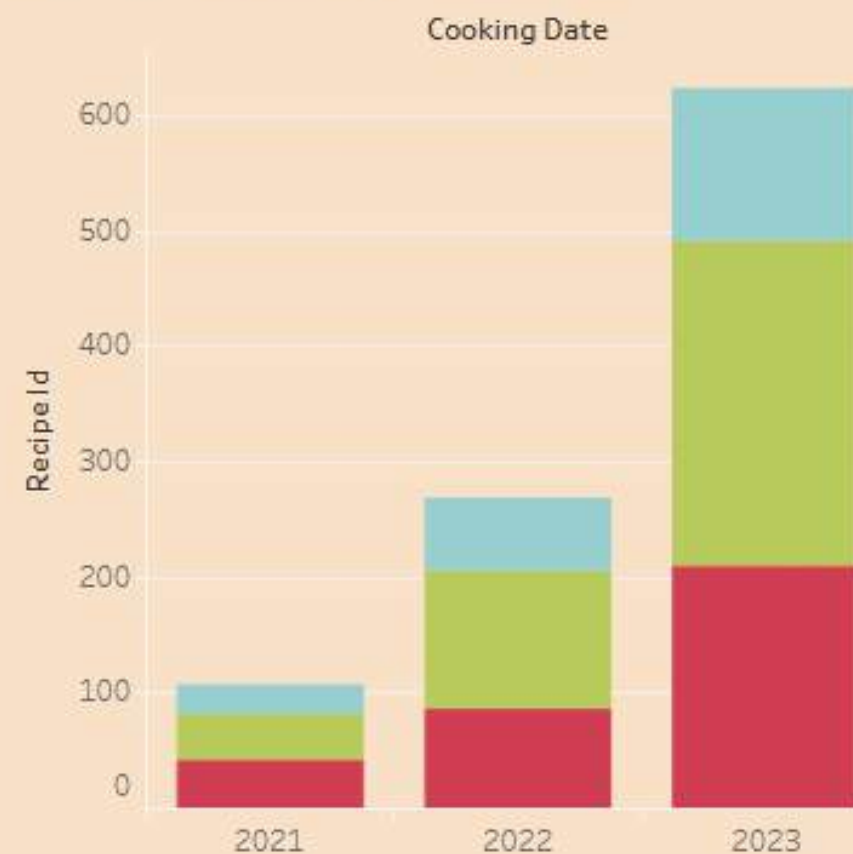
# Successful recipes | Dashboard URL



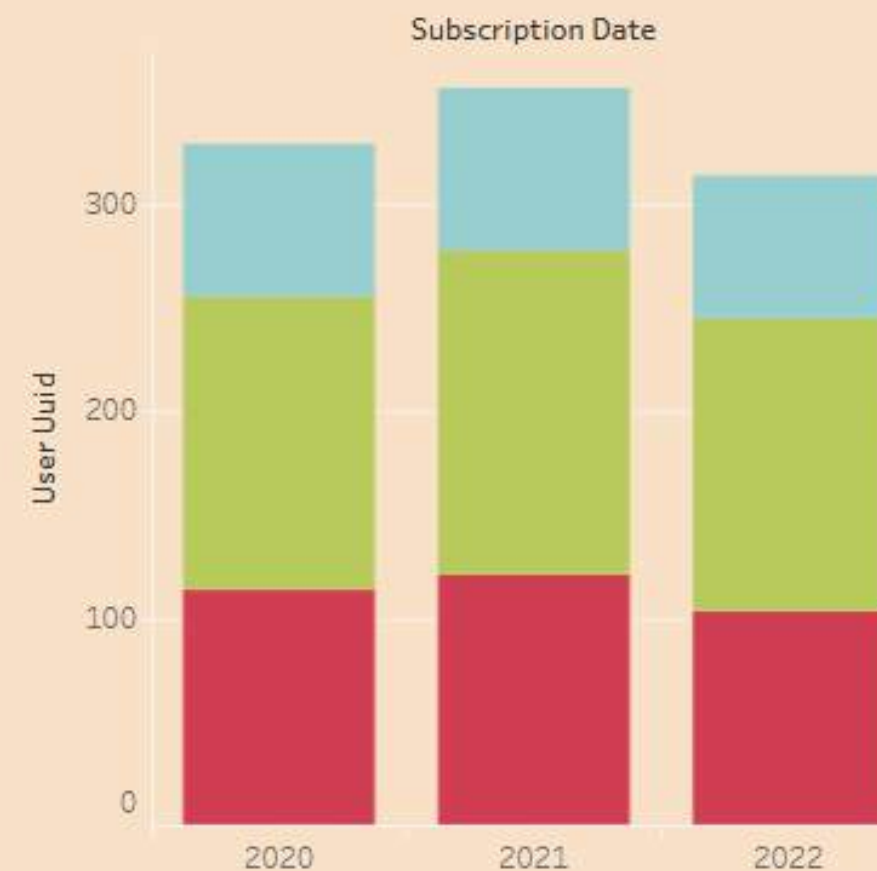
# Platform growth | Dashboard URL

## PLATFORM GROWTH

Recipes growth



Subscriptions growth



Interactions



Country

- France
- Italy
- UnitedKingdom

Overview of how the platform grew during the years.

The recipes growth have been increasing year after year, whereas there has been a drop of subscriptions in 2022.

The interactions have been growing year after year.

The users come especially from Italy, followed by United Kingdom.



# Buyer Persona

Some clusters of users have been selected according to the favourite recipes, to better target their interests.



**Carlotta Perego**

Users who cook only vegan recipes  
and are interested especially in  
healthy food.  
[vegan and smoothie recipes]



**Aïcha**

Curious and adventurous users who  
want to try new cuisines and to  
experiment.  
[ethnic recipes]



# Buyer Persona

Some clusters of users have been selected according to the favourite recipes, to better target their interests.



**Tara McCloud**

Users keen on sport who are looking for quick recipes but also something nutritious and energetic to eat.  
[breakfast and snack recipes]

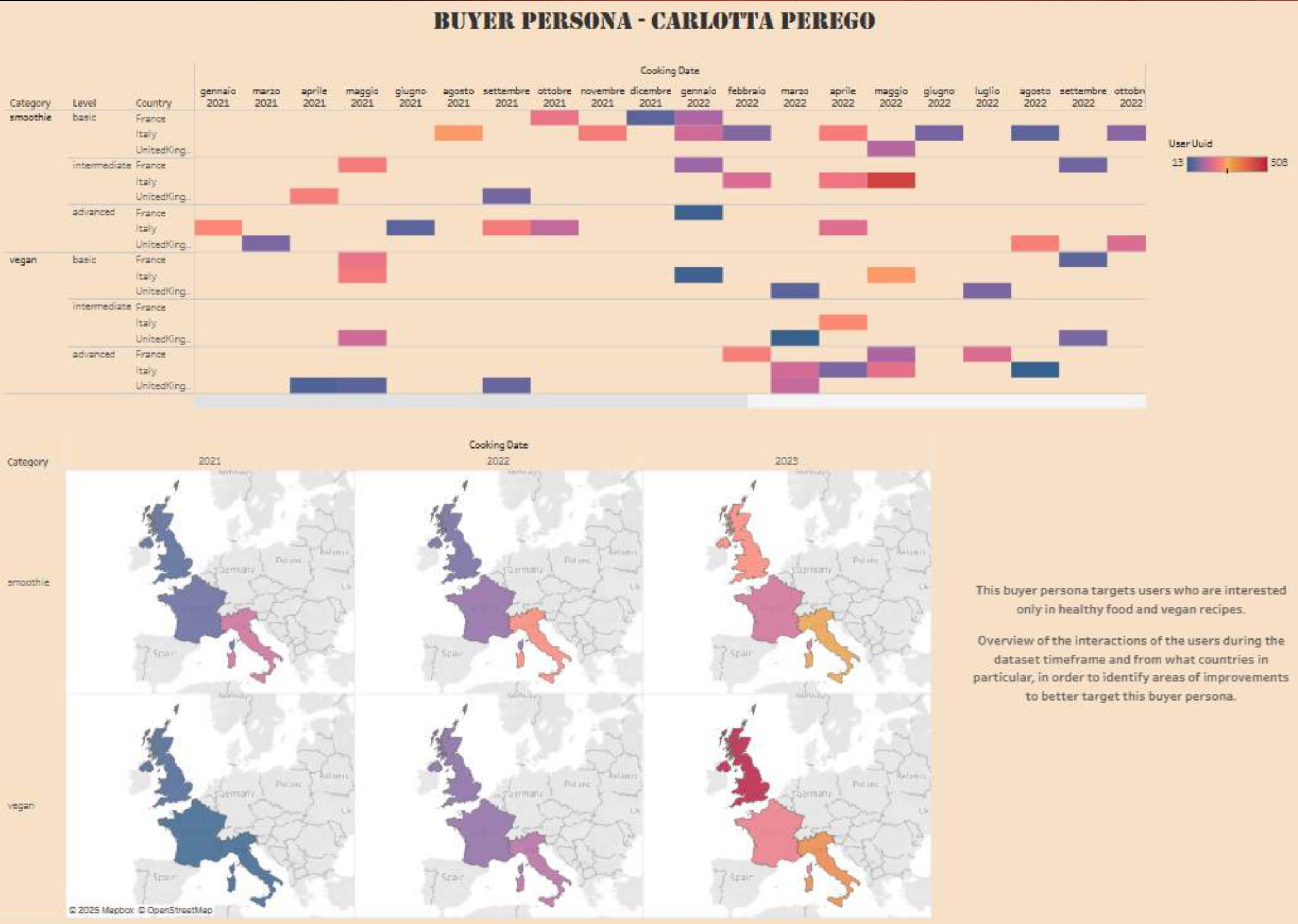


**Nunzia Caputo**

Users who spend most of their time at home and prefer to cook all their meals from there.  
[lunch and dinner recipes]

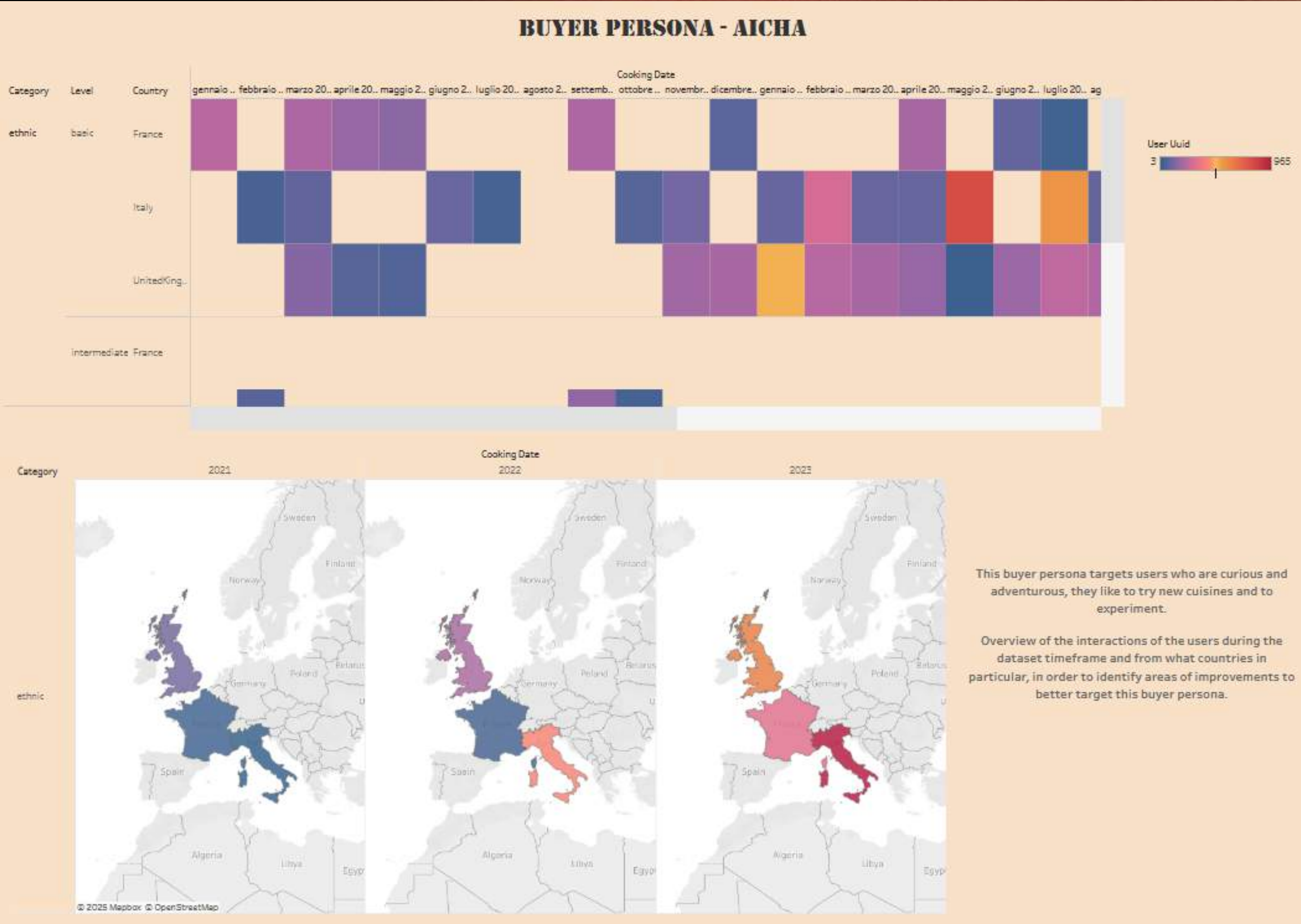


# Buyer Persona – Carlotta | Dashboard URL



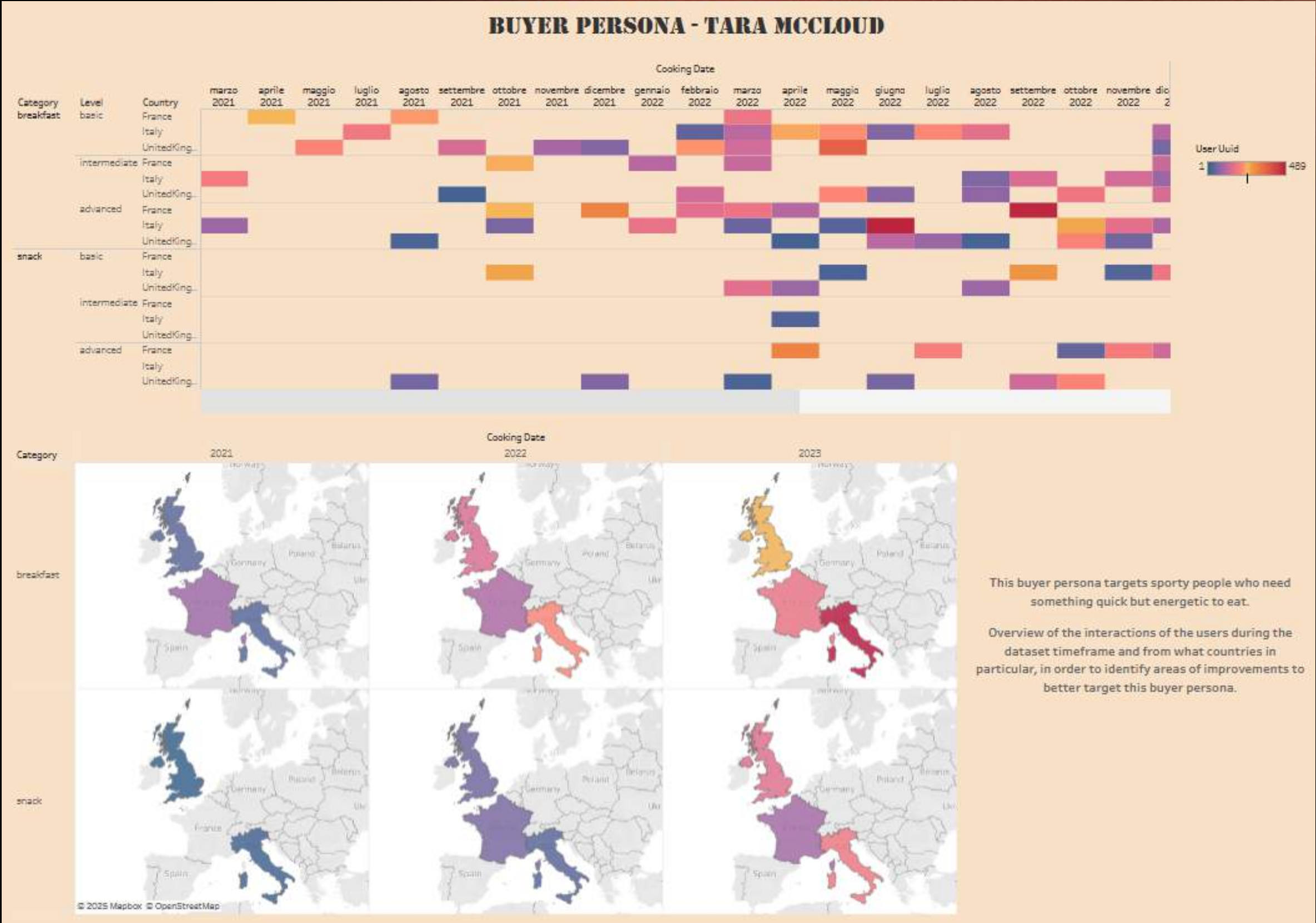


# Buyer Persona – Aicha | Dashboard URL



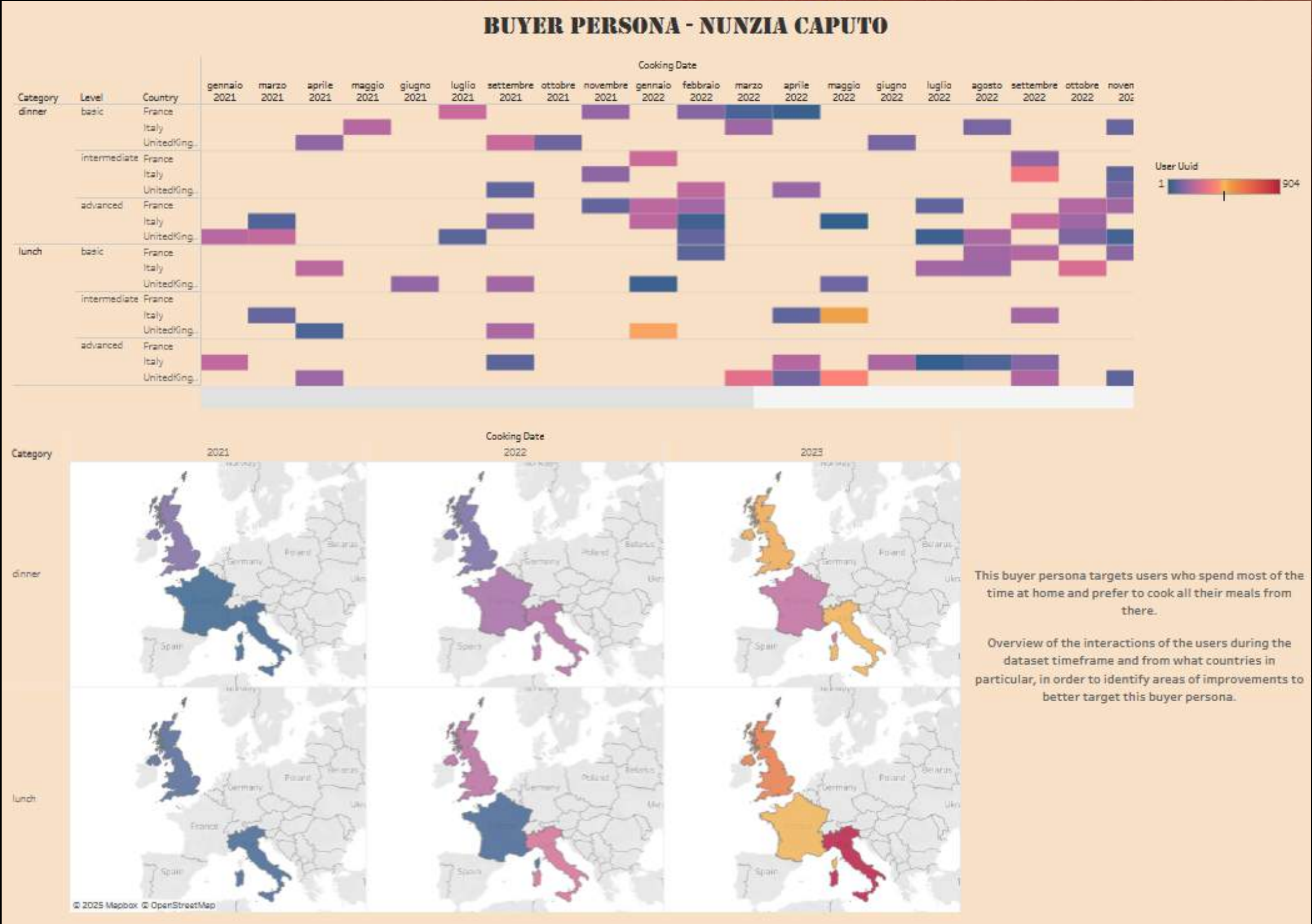


# Buyer Persona – Tara | Dashboard URL





# Buyer Persona – Nunzia | Dashboard URL





# Final conclusion

---

The platform has registered a stable growth during the years, but there has been a slight drop in the subscriptions in 2022.

During the months with more subscriptions there have not been more interactions with the recipes, it might be that new users need more time to be retained and be really engaged.

The recipes with the highest rating are not the ones with more interactions (like the breakfast category), but it is also because they are also the ones published more on the platform.

There are good growth prospects according to the buyer persona on which invest more. Another strategy could also be to target users from other countries, instead of increasing the subscriptions from France which are the users with the lowest numbers but also the less active ones.

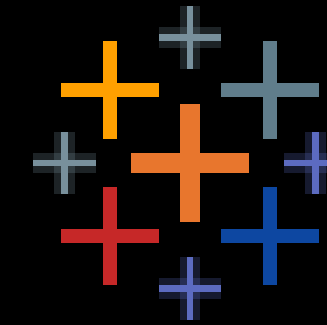
SQL queries URL



Jupyter Notebook URL



Tableau Public URL



---

[Back to index](#)