

CODE OF THE TABLES CREATION

```
CREATE TABLE if not exists Users(  
  ID_U INT NOT NULL AUTO_INCREMENT,  
  user_uuid int not null,  
  subscription_date DATE,  
  language varchar(40),  
  country varchar(40),  
  platform varchar(40),  
  cooking_date DATE,  
  chef_id int not null,  
  challenge_id int not null,  
  primary key (ID_U)  
);
```

```
CREATE TABLE if not exists Recipes(  
  ID_R INT NOT NULL AUTO_INCREMENT,  
  recipe_id int not null,  
  level varchar(40),  
  category varchar(40),  
  stars int not null,  
  chef_id int not null,  
  challenge_id int not null,  
  primary key (ID_R)  
);
```

CODE TO IMPORT THE DATA OF THE DATASET

```
LOAD DATA INFILE 'C:\food_data_clean.csv'
INTO TABLE users
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY '"'
IGNORE 1 LINES
(@col1, @col2, @dummy1, @col3, @col4, @dummy2, @col5, @col6, @col7, @col8)
SET cooking_date = @col1, user_uuid = @col2, chef_id = @col3,
language = @col4, challenge_id = @col5, country = @col6, subscription_date = @col7, platform = @col8;
```

```
LOAD DATA INFILE 'C:\food_data_clean.csv'
INTO TABLE recipes
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY '"'
IGNORE 1 LINES
(@dummy1, @dummy2, @col1, @col2, @dummy3, @col3, @col4, @dummy4, @dummy5, @dummy6,
@col5, @col6)
SET category = @col1, chef_id = @col2, level = @col3,
challenge_id = @col4, recipe_id = @col5, stars = @col6;
```

CODE OF ALL THE QUERIES

Min e max subscription year in the dataset

```
SELECT
MIN(subscription_date) AS first_year,
MAX(subscription_date) AS last_year
FROM users;
```

The years in the dataset are 2020, 2021, and 2022.

Let's calculate the amount of users who enrolled each year.

```
SELECT country, COUNT(DISTINCT user_uuid) AS TotalUsers
FROM users
WHERE EXTRACT(YEAR from subscription_date) = 2020
GROUP BY country;
```

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

```
SELECT country, COUNT(DISTINCT user_uuid) AS TotalUsers
FROM users
WHERE EXTRACT(YEAR from subscription_date) = 2021
GROUP BY country;
```

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

```
SELECT country, COUNT(DISTINCT user_uuid) AS TotalUsers
FROM users
WHERE EXTRACT(YEAR from subscription_date) = 2022
GROUP BY country;
```

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

There has been a drop of new subscriptions in 2022.

The majority of the users comes from Italy.

Let's move on the recipes, checking how much they grew during the years.

```
SELECT  
  
MIN(cooking_date) AS first_year,  
MAX(cooking_date) AS last_year  
FROM users;
```

The years when the recipes have been posted are 2021, 2022, 2023.

```
SELECT  
  
    EXTRACT(YEAR FROM u.cooking_date) AS year,  
    COUNT(DISTINCT r.recipe_id) AS TotalRecipes,  
    COUNT(DISTINCT u.challenge_id) as total_challenges,  
    count(distinct u.user_uuid) as total_users  
FROM  
    users u  
INNER JOIN  
    recipes r ON u.challenge_id = r.challenge_id  
GROUP BY  
    EXTRACT(YEAR FROM u.cooking_date)  
ORDER BY  
    year DESC;
```

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

Let's now calculate how many recipes received 4 and 5 stars.

```
SELECT COUNT(DISTINCT recipe_id) AS TotalRecipes
```

FROM recipes

WHERE stars >=4;

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).

Let's analyze more in detail which recipes are per level and category.

```
SELECT category, level, COUNT(DISTINCT recipe_id) AS TopRecipes
```

```
FROM recipes
```

```
WHERE stars >= 4 AND level = 'basic'
```

```
GROUP BY category, level;
```

	category	level	TopRecipes
1	ethnic	basic	39
2	breakfast	basic	25
3	snack	basic	14
4	dinner	basic	12
5	lunch	basic	11
6	smoothie	basic	11
7	vegan	basic	9

```
SELECT category, level, COUNT(DISTINCT recipe_id) AS TopRecipes
```

```
FROM recipes
```

```
WHERE stars >= 4 AND level = 'intermediate'
```

```
GROUP BY category, level;
```

	category	level	TopRecipes
1	ethnic	intermed...	29
2	lunch	intermed...	20
3	breakfast	intermed...	11
4	dinner	intermed...	8
5	smoothie	intermed...	8
6	vegan	intermed...	8
7	snack	intermed...	4

```

SELECT category, level, COUNT(DISTINCT recipe_id) AS TopRecipes
FROM recipes
WHERE stars >= 4 AND level = 'advanced'
GROUP BY category, level;

```

	category ▲	level ▲	TopRecipes ▼
1	ethnic	advanced	48
2	lunch	advanced	29
3	breakfast	advanced	24
4	vegan	advanced	18
5	dinner	advanced	17
6	smoothie	advanced	13
7	snack	advanced	12

Quick check of the total recipes published per category.

```

SELECT category, COUNT(DISTINCT recipe_id) AS TotalRecipes
FROM recipes
GROUP BY category;

```

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

The top categories are then ethnic, breakfast, lunch. To find out who are the most popular chefs, I group only the most successful categories by chef.

```
SELECT category, chef_id, COUNT(DISTINCT recipe_id) AS TopRecipes
FROM recipes
WHERE stars >= 4 AND category IN ('ethnic', 'breakfast', 'lunch')
GROUP BY category, chef_id
ORDER BY TopRecipes DESC
LIMIT 10;
```

	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

Now it's time to analyze the users. Let's sort them by country of origin per each category and level of recipes.

```
SELECT country, category, level, COUNT(DISTINCT user_uuid) AS UserOrigin
FROM recipes
INNER JOIN users ON recipes.chef_id = users.chef_id
WHERE stars >= 4 AND country = 'Italy'
GROUP BY category;
```

	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

```

SELECT country, category, level, COUNT(DISTINCT user_uuid) AS UserOrigin
FROM recipes
INNER JOIN users ON recipes.chef_id = users.chef_id
WHERE stars >= 4 AND country = 'France'
GROUP BY category;

```

	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

```

SELECT country, category, level, COUNT(DISTINCT user_uuid) AS UserOrigin
FROM recipes
INNER JOIN users ON recipes.chef_id = users.chef_id
WHERE stars >= 4 AND country = 'UnitedKingdom'
GROUP BY category;

```


	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

Let's move on in analyzing the users regardless of the successful recipes.

```
SELECT country, category, level, COUNT(DISTINCT user_uid) AS UserOriginTotal
FROM recipes
INNER JOIN users ON recipes.chef_id = users.chef_id
GROUP BY category, level;
```

	country ▲	category ▲	level ▲	UserOriginTotal ▼		country	category	level	UserOriginTotal
1	Italy	dinner	advanced	242	11	UnitedKingdom	vegan	advanced	237
2	Italy	ethnic	advanced	242	12	UnitedKingdom	lunch	basic	236
3	UnitedKingdom	ethnic	basic	242	13	Italy	smoothie	basic	231
4	France	breakfast	advanced	241	14	UnitedKingdom	smoothie	advanced	228
5	Italy	breakfast	basic	241	15	UnitedKingdom	breakfast	intermediate	227
6	UnitedKingdom	ethnic	intermediate	241	16	UnitedKingdom	snack	basic	225
7	France	lunch	advanced	241	17	Italy	smoothie	intermediate	224
8	UnitedKingdom	lunch	intermediate	241	18	France	dinner	intermediate	222
9	Italy	dinner	basic	240	19	UnitedKingdom	vegan	basic	220
10	Italy	snack	advanced	240	20	UnitedKingdom	snack	intermediate	199
					21	Italy	vegan	intermediate	194

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, let's analyze also the preferences of the users during the years of subscription.

```
SELECT cooking_date, country, category, level, COUNT(DISTINCT user_uuid) AS UserPreferenceTime
FROM recipes
INNER JOIN users ON recipes.chef_id = users.chef_id
GROUP BY category, level;
```

	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.

Eventually, let's analyze the challenges per country of origin and platform used.

```
SELECT u.challenge_id, u.platform, u.country, COUNT(DISTINCT u.user_uuid) AS UserChallenge
FROM recipes r
INNER JOIN users u ON r.challenge_id = u.challenge_id
GROUP BY country, platform, challenge_id;
```

	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.

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.