

```
In [249.. #pandas importeren
import pandas as pd

# dataset inlezen
ingredients = pd.read_csv('ingredients.csv')
recipes = pd.read_csv('recipes.csv')
tags = pd.read_csv('tags.csv')
nutrititions = pd.read_csv('nutrititions.csv')
```

```
In [241.. #ingredients dataset inspecteren
ingredients.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 71809 entries, 0 to 71808
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0    71809 non-null   int64
1   recipe        71809 non-null   object
2   ingredient     71809 non-null   object
3   quantity      71809 non-null   float64
4   unit          57509 non-null   object
dtypes: float64(1), int64(1), object(3)
memory usage: 2.7+ MB
```

```
In [242.. # de eerste 5 rijen van de dataset tonen
ingredients.head()
```

```
Out[242]:
```

	Unnamed: 0	recipe	ingredient	quantity	unit
0	0	Kruidnoten met choco-discodip	melkchocolade	100.0	g
1	1	Kruidnoten met choco-discodip	kruidnoten	100.0	g
2	2	Kruidnoten met choco-discodip	discodip	2.0	el
3	3	Kruidnoten in marsepein	blanke marsepein	150.0	g
4	4	Kruidnoten in marsepein	ongezouten roomboter	15.0	g

```
In [243.. # de kolom 'Unnamed: 0' verwijderen en de dataset opnieuw inspeteren
ingredients = ingredients.drop(['Unnamed: 0'], axis = 1)
ingredients.head()
```

```
Out[243]:
```

	recipe	ingredient	quantity	unit
0	Kruidnoten met choco-discodip	melkchocolade	100.0	g
1	Kruidnoten met choco-discodip	kruidnoten	100.0	g
2	Kruidnoten met choco-discodip	discodip	2.0	el
3	Kruidnoten in marsepein	blanke marsepein	150.0	g
4	Kruidnoten in marsepein	ongezouten roomboter	15.0	g

```
In [244.. # groeperen per recept en alle ingredienten bij elkaar joinen
ingredients = ingredients.groupby('recipe', as_index = False, sort = False).agg({'ingredient': ', '.join})
ingredients.head()
```

```
Out[244]:
```

	recipe	ingredient
0	Kruidnoten met choco-discodip	melkchocolade,kruidnoten,discodip
1	Kruidnoten in marsepein	blanke marsepein,ongezouten roomboter,kruidnot...
2	Kruidnoten met chocodips	pure chocolade,ongezouten roomboter,kruidnoten...
3	Pepernotentartaart met marsepeinstrik	ongezouten roomboter,witte basterdsuiker,vanil...
4	Perencake	middelgroot ei,Griekse yoghurt 10%,fijne krist...

```
In [245.. # de kolom "recipe" uit de dataset gefilterd
recipe = ingredients['recipe']
recipe = pd.DataFrame(recipe)
recipe.head()
```

```
Out[245]:
```

	recipe
0	Kruidnoten met choco-discodip
1	Kruidnoten in marsepein
2	Kruidnoten met chocodips
3	Pepernotentartaart met marsepeinstrik
4	Perencake

```
In [246.. # de split methode gebruikt om de in dekolom in meerdere kolom te splitsen
recipe_ingredients = ingredients['ingredient'].str.split(',', expand = True)
recipe_ingredients = pd.DataFrame(recipe_ingredients)
recipe_ingredients.head()
```

```
Out[246]:
```

	0	1	2	3	4	5	6	7	8	9	...	17	18	19	20	21	22	23	24	25	26
0	melkchocolade	kruidnoten	discodip	None	None	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None	None
1	blanke marsepein	ongezouten roomboter	kruidnoten	nougatine in bakje	None	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None	None
2	pure chocolade	ongezouten roomboter	kruidnoten	cacaopoeder	hagelslag puur	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None	None
3	ongezouten roomboter	witte basterdsuiker	vanille-extract	zout	middelgroot ei	tarwebloem	bakpoeder	poedersuiker	aardbeienjam	rode marsepein	...	None	None	None	None	None	None	None	None	None	None
4	middelgroot ei	Griekse yoghurt 10%	fijne kristalsuiker	vanille-extract	zonnebloemolie	tarwebloem	bakpoeder	zout	Conference peer	None	...	None	None	None	None	None	None	None	None	None	None

5 rows × 27 columns

```
In [225.. ingredienten_lijst = []
for x in range(len(ingredients)):
    rij_ingredienten = [i for i in ingredients.iloc[x]]
    for t in rij_ingredienten:
        if t not in ingredienten_lijst:
            ingredienten_lijst.append(t)

ingredienten_lijst.insert(0, 'recipe')
```

```
In [247.. # de dataframes recipe en recipe ingredient bij elkaar voegen
ingredients = pd.concat([recipe, recipe_ingredients],axis = 1, sort = False)
ingredients.head()
```

```
Out[247]:
```

	recipe	0	1	2	3	4	5	6	7	8	...	17	18	19	20	21	22	23	24	25	26
0	Kruidnoten met choco-discodip	melkchocolade	kruidnoten	discodip	None	None	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None
1	Kruidnoten in marsepein	blanke marsepein	ongezouten roomboter	kruidnoten	nougatine in bakje	None	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None
2	Kruidnoten met chocodips	pure chocolade	ongezouten roomboter	kruidnoten	cacaopoeder	hagelslag puur	None	None	None	None	None	...	None	None	None	None	None	None	None	None	None
3	Pepernotentartaart met marsepeinstrik	ongezouten roomboter	witte basterdsuiker	vanille-extract	zout	middelgroot ei	tarwebloem	bakpoeder	poedersuiker	aardbeienjam	...	None	None	None	None	None	None	None	None	None	None
4	Perencake	middelgroot ei	Griekse yoghurt 10%	fijne kristalsuiker	vanille-extract	zonnebloemolie	tarwebloem	bakpoeder	zout	Conference peer	...	None	None	None	None	None	None	None	None	None	None

5 rows × 28 columns

```
In [197.. # de eerste 5 rijen van de dataset tonen
recipes.head()
```

```
Out[197]:
```

	Unnamed: 0	title	persons	time	calories	stars	url	image
0	0	Kruidnoten met choco-discodip	4	25	260	0	https://www.ah.nl/allerhande/recept/R-R1195893...	https://static.ah.nl/static/recepten/img_RAM_P...
1	0	Kruidnoten in marsepein	4	25	265	0	https://www.ah.nl/allerhande/recept/R-R1195892...	https://static.ah.nl/static/recepten/img_RAM_P...
2	0	Kruidnoten met chocodips	4	25	335	0	https://www.ah.nl/allerhande/recept/R-R1195891...	https://static.ah.nl/static/recepten/img_RAM_P...
3	0	Pepernotentartaart met marsepeinstrik	10	30	560	0	https://www.ah.nl/allerhande/recept/R-R1195887...	https://static.ah.nl/static/recepten/img_RAM_P...
4	0	Perencake	12	10	265	0	https://www.ah.nl/allerhande/recept/R-R1195790...	https://static.ah.nl/static/recepten/img_RAM_P...

```
In [198.. # de kolom 'Unnamed: 0' verwijderen en de dataset opnieuw inspeteren
recipes = recipes.drop(['Unnamed: 0'], axis =1)
recipes.head()
```

```
Out[198]:
```

	title	persons	time	calories	stars	url	image
0	Kruidnoten met choco-discodip	4	25	260	0	https://www.ah.nl/allerhande/recept/R-R1195893...	https://static.ah.nl/static/recepten/img_RAM_P...
1	Kruidnoten in marsepein	4	25	265	0	https://www.ah.nl/allerhande/recept/R-R1195892...	https://static.ah.nl/static/recepten/img_RAM_P...
2	Kruidnoten met chocodips	4	25	335	0	https://www.ah.nl/allerhande/recept/R-R1195891...	https://static.ah.nl/static/recepten/img_RAM_P...
3	Pepernotentartaart met marsepeinstrik	10	30	560	0	https://www.ah.nl/allerhande/recept/R-R1195887...	https://static.ah.nl/static/recepten/img_RAM_P...
4	Perencake	12	10	265	0	https://www.ah.nl/allerhande/recept/R-R1195790...	https://static.ah.nl/static/recepten/img_RAM_P...

```
In [199.. #tags dataset inspecteren
tags.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 46040 entries, 0 to 46039
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0    46040 non-null   int64
1   recipe        46040 non-null   object
2   tag          46040 non-null   object
dtypes: int64(1), object(2)
memory usage: 1.1+ MB
```

```
In [200.. # de eerste 5 rijen van de dataset tonen
tags.head()
```

```
Out[200]:
```

	Unnamed: 0	recipe	tag
0	0	Kruidnoten met choco-discodip	hollands
1	1	Kruidnoten met choco-discodip	gebak
2	2	Kruidnoten met choco-discodip	gebak
3	3	Kruidnoten met choco-discodip	sinterklaas
4	4	Kruidnoten met choco-discodip	sinterklaasavond

```
In [248.. # de kolom 'Unnamed: 0' verwijderen en de dataset opnieuw inspeteren
tags = tags.drop(['Unnamed: 0'], axis = 1)
tags.head()
```

```
Out[248]:
```

	recipe	tag
0	Kruidnoten met choco-discodip	hollands
1	Kruidnoten met choco-discodip	gebak
2	Kruidnoten met choco-discodip	gebak
3	Kruidnoten met choco-discodip	sinterklaas
4	Kruidnoten met choco-discodip	sinterklaasavond

```
In [249.. # groeperen per recipe en de tags bij elkaar joinen
tags = tags.groupby('recipe', as_index = False, sort = False).agg({'recipe': 'first', 'tag': ', '.join})
tags.head()
```

```
Out[249]:
```

	recipe	tag
0	Kruidnoten met choco-discodip	hollands,gebak,gebak,sinterklaas,sinterklaasavond
1	Kruidnoten in marsepein	hollands,gebak,gebak,sinterklaas,sinterklaasavond
2	Kruidnoten met chocodips	hollands,gebak,gebak,sinterklaas,sinterklaasavond
3	Pepernotentartaart met marsepeinstrik	oven,hollands,gebak,gebak,sinterklaas,sinterkl...
4	Perencake	vooraf te maken,oven,gebak,gebak

```
In [253.. #dataset ingredients en tags mergen op recipe
data = tags.merge(ingredients, on = "recipe")
data.head()
```

```
Out[253]:
```

	recipe	tag	0	1	2	3	4	5	6	7	...	17	18	19	20	21	22	23	24
0	Kruidnoten met choco-discodip	hollands,gebak,gebak,sinterklaas,sinterklaasavond	melkchocolade	kruidnoten	discodip	None	None	None	None	None	...	None	None	None	None	None	None	None	None
1	Kruidnoten in marsepein	hollands,gebak,gebak,sinterklaas,sinterklaasavond	blanke marsepein	ongezouten roomboter	kruidnoten	nougatine in bakje	None	None	None	None	...	None	None	None	None	None	None	None	None
2	Kruidnoten met chocodips	hollands,gebak,gebak,sinterklaas,sinterklaasavond	pure chocolade	ongezouten roomboter	kruidnoten	cacaopoeder	hagelslag puur	None	None	None	...	None	None	None	None	None	None	None	None
3	Pepernotentartaart met marsepeinstrik	oven,hollands,gebak,gebak,sinterklaas,sinterkl...	ongezouten roomboter	witte basterdsuiker	vanille-extract	zout	middelgroot ei	tarwebloem	bakpoeder	poedersuiker	...	None	None	None	None	None	None	None	None
4	Perencake	vooraf te maken,oven,gebak,gebak	middelgroot ei	Griekse yoghurt 10%	fijne kristalsuiker	vanille-extract	zonnebloemolie	tarwebloem	bakpoeder	zout	...	None	None	None	None	None	None	None	None

5 rows × 29 columns

```
In [113.. #nutrititions dataset inspecteren
nutrititions.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 58765 entries, 0 to 58764
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0    58765 non-null   int64
1   recipe        58765 non-null   object
2   nutrition     58765 non-null   object
3   value        58765 non-null   object
dtypes: int64(1), object(3)
memory usage: 1.8+ MB
```

```
In [114.. # de eerste 5 rijen van de dataset tonen
nutrititions.head()
```

```
Out[114]:
```

	Unnamed: 0	recipe	nutrition	value
0	0	Kruidnoten met choco-discodip	energie	260 kcal
1	1	Kruidnoten met choco-discodip	koolhydraten	34 g
2	2	Kruidnoten met choco-discodip	waarvan suikers	22 g
3	3	Kruidnoten met choco-discodip	natrium	200 mg
4	4	Kruidnoten met choco-discodip	eiwit	3 g

```
In [115.. # de kolom 'Unnamed: 0' verwijderen en de dataset opnieuw inspeteren
nutritions = nutrititions.drop(['Unnamed: 0'], axis = 1)
nutritions.head(10)
```

```
Out[115]:
```

	recipe	nutrition	value
0	Kruidnoten met choco-discodip	energie	260 kcal
1	Kruidnoten met choco-discodip	koolhydraten	34 g
2	Kruidnoten met choco-discodip	waarvan suikers	22 g
3	Kruidnoten met choco-discodip	natrium	200 mg
4	Kruidnoten met choco-discodip	eiwit	3 g
5	Kruidnoten met choco-discodip	vet	12 g
6	Kruidnoten met choco-discodip	waarvan verzadigd	7 g
7	Kruidnoten met choco-discodip	vezels	1 g
8	Kruidnoten in marsepein	energie	265 kcal
9	Kruidnoten in marsepein	koolhydraten	43 g

```
In [116.. nutritions = nutrititions.groupby('recipe', as_index = False).agg({'recipe': 'first', 'nutrition': ', '.join })
```

```
In [117.. nutritions.head()
```

```
Out[117]:
```

	recipe	nutrition
0	Andijviestamppot met gebakken salami	energie, koolhydraten, natrium, eiwit, vet, wa...
1	Bietensoep met selderijroom	energie, koolhydraten, natrium, eiwit, vet, wa...
2	Libanese wraps met appel-koolsalade	energie, koolhydraten, natrium, eiwit, vet, wa...
3	Plaattaart met gegrilde groenten en witte kaas	energie, koolhydraten, waarvan suikers, natriu...
4	Sticky cauliflower (gegrilde bloemkoolroosjes...	energie, koolhydraten, natrium, eiwit, vet, wa...

```
In [118.. nutritions.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8706 entries, 0 to 8705
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   recipe      8706 non-null   object
1   nutrition   8706 non-null   object
dtypes: object(2)
memory usage: 136.2+ KB
```

```
In [ ] :
```