

ACME Corporation

Team D

Milan, Umut, Tony, Eugen, Dennis, Alexandre, Vimal

AcME
CORPORATION

A woman with long dark hair tied back is wearing a white lace-trimmed blouse and blue jeans. She is holding a silver bowl filled with a colorful salad or smoothie bowl containing various fruits like strawberries, kiwi, and granola. The background shows a striped wall and a green cloth.

**FOR A
HEALTHIER
EATING HABIT**



The Spanish Market

THE DATASET

Generated from world.openfoodfacts.org



Open Food
Database

Scanned
products from
+150 countries

THE CHALLENGES



LACK OF STANDARDIZATION

Data inputs do not follow a standardized schema



DATA CLEANING

Data cleaning was required to set a certain structure and have relevant output



CSV NOT PARSED CORRECTLY

Lack of structure in the open dataset led to Spark confusion while parsing comma separated fields into columns



DATA QUALITY

Existence of a lot of null values in relevant features of interest has an impact on the quality of the report

The Oldest Product

The oldest product in the sortiment:

```
+-----+-----+-----+
|created_datetime |code          |product_name   |
+-----+-----+-----+
|2012-03-20 16:34:12|8.712566246465E12|moutarde fine de dijon|
+-----+-----+-----+
only showing top 1 row
```



"created_date_time" feature contains the timestamp when the product was added to the website.

To obtain the Newest Product, the data was sorted by "created_date_time" in ASCENDING order, showing the product name and product code.

The Newest Product

The newest product in the sortiment:

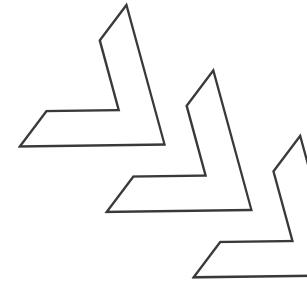
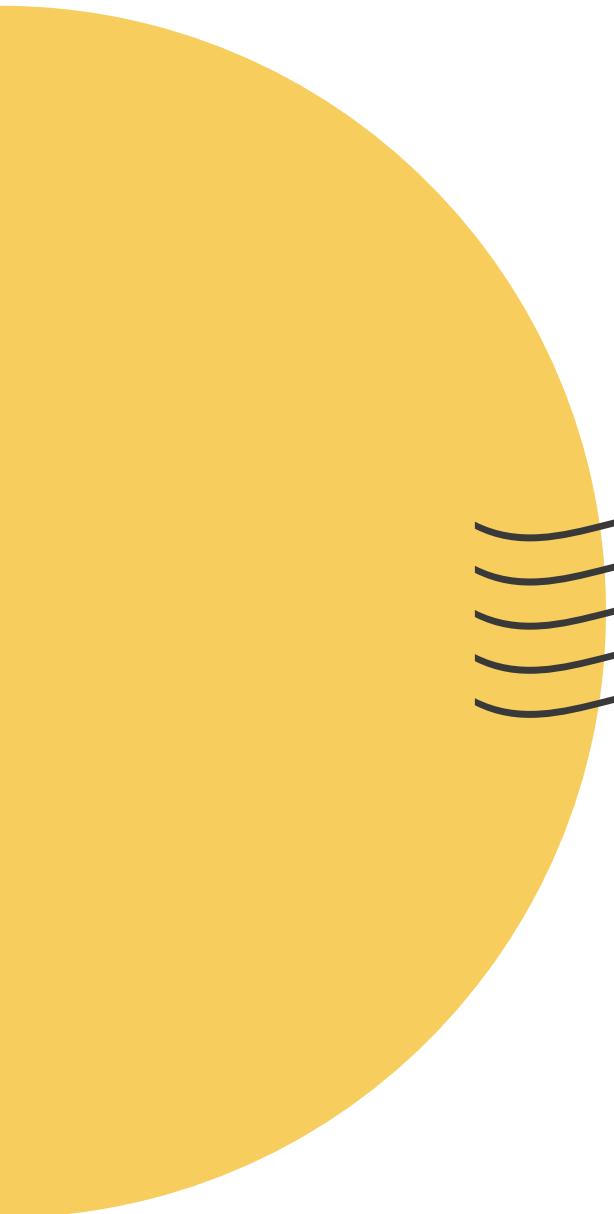
```
+-----+-----+
|created_datetime |code          |product_name
+-----+-----+
|2020-09-13 02:49:51|8.412540019282E12|palometa austral ahumada en aceite|
+-----+-----+
only showing top 1 row
```



"created_date_time" feature contains the timestamp when the product was added to the website.

To obtain the Newest Product, the data was sorted by "created_date_time" in DESCENDING order, showing the product name and product code.

The Average Product Age



420
days

Age was defined as a new feature,
referring to time passed (in days)
since the addition of the product to
the website

Average Product Age

Top 10 oldest products

code	product_name	Avg_days
8.712566246465E12	moutarde fine de dijon	3125.0
3.179732348913E12	perrier	3122.0
4.008400811727E12	mon chéri	3121.0
3.270190008613E12	couscous grain moyen	3116.0
5.4492387E7	coca cola light	3107.0
5.410041040807E12	pim's framboise	3103.0
5.6004083E7	pedras	3097.0
5.606293100127E12	rucula	3093.0
5.449000169327E12	coca cola zéro sans caféine	3077.0
7.6840600038E10	chocolate fudge brownie	3069.0

The Oldest Products have been uploaded to the platform
8.5 years ago.



Average Product Age

Top 10 youngest products

code	product_name	Avg_days
8.436575091648E12	vitamins hair/skins/nails	26.0
8.412540019282E12	palometa austral ahumada en aceite	26.0
8.414121190014E12	garbanzos con alga kombu	26.0
8.480017225016E12	patatas bravas	26.0
8.437008675084E12	almogrote	26.0
8.424465231045E12	donuts chocolate	26.0
5.9712490452000205E17	lomo de cerdo ibérico	27.0
8.436594380099E12	bento box	27.0
4.316734084596E12	miso light	27.0
2.902721002241E12	petxuga gall indi	27.0

The Newest Products have been uploaded on the platform in the last 26 days.

ACME Corporation can investigate recently uploaded items further to observe the new trends:

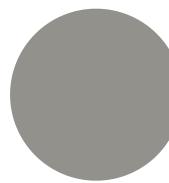
Can the most recent product be categorized as healthier? Are they prepacked prepared food?



Product Related Column Profiling



177,221 Distinct Products



692 Countries Combinations Where the Products Are Sold



2,405 Distinct Packaging Tags



1,042 Distinct Traces



4,175 Distinct Additives



More than 100,000 Null Values
Traces, Packaging, Categories & Additives

More focus should be put on the quality of the data added to the website
Some features such as Additives and Traces are important for the users as they can be detrimental to one's health

List of Countries

Where the products are sold to

The majority of the products sold in Spain are also sold in France.

These 2 countries seem to have similar eating habits among their people.

The overall health state situation of these countries will tend to be similar.

country	n
france	12147
germany	1107
belgium	886
switzerland	694
usa	646
portugal	488
italy	474
uk	358
netherlands	272
mexico	228
argentina	206
morocco	136
canada	121
austria	119
norway	106
venezuela	84
cyprus-ελληνικά	82
croatia-hrvatski	82
moldova-roman	82
guatemala	82

only showing top 20 rows

List of top countries (other than Spain)

List of Product Categories

Sorted in descending order

The majority of the top products
are ready-to-eat, pre-packed /
canned products



More additives?



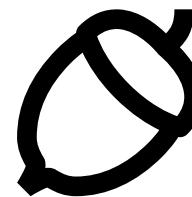
Dangerous?

main_category_en	product_name_count
extra-virgin olive oils	999
cheeses	960
biscuits	880
prepared meats	737
salty snacks	551
dark chocolates	534
breakfast cereals	515
dietary supplements	511
dairies	497
fruit-jams	429
bread	389
yogurts	376
sauces	343
cow milks	332
smoked salmons	329
canned fishes	323
sheep's-milk cheeses	319
honeys	315
biscuits and cakes	312
candies	311

List of Product Traces

Sorted in descending order

Top identified traces in Spain:



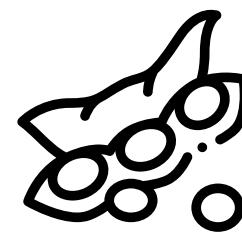
Nuts



Gluten



Milk



Soybeans

*Different combination of these traces
(products that have more than one trace in
their composition)*

traces	traces_count
nuts	3158
milk	2350
soybeans	2093
eggs	1299
sesame seeds	1266
gluten	1249
peanuts	990
celery	586
mustard	442
fish	366
sulphur dioxide and sulphites	332
crustaceans	300
molluscs	218
lupin	91
ca:llet	18
ca:sesam	15
es:ninguna	12
es:soja-y-leche-lactosa	9
es:pipas-de-girasol	8
ca:ou	8

Data Quality Analysis

On the fields of interest

1,234 Products with complete information

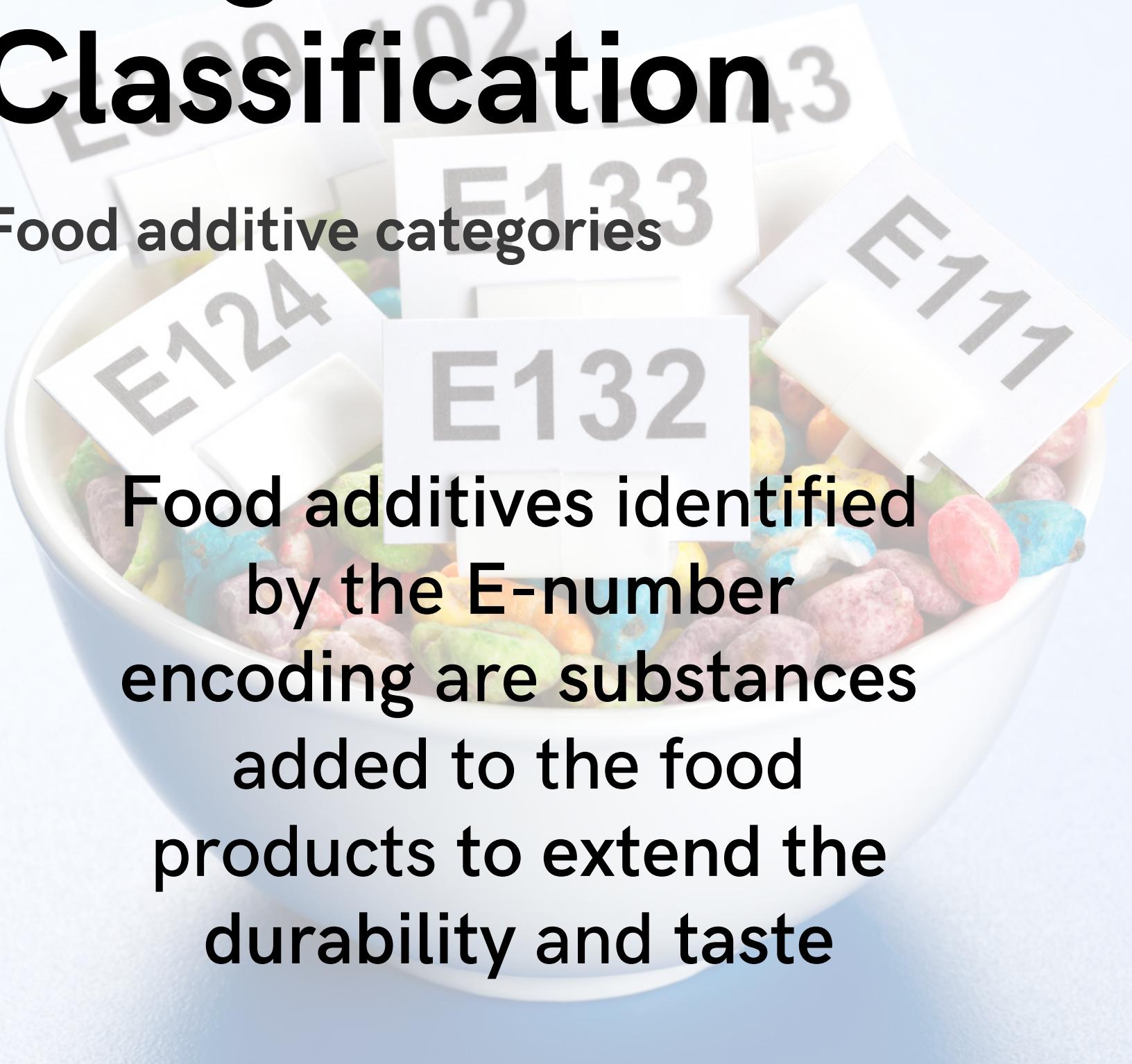
26% of the products are without complete analysis per 100g

94% of the products are without additive tags info

96% of the products are without traces info

Dangerous Product Classification

Food additive categories



Food additives identified by the E-number encoding are substances added to the food products to extend the durability and taste

Natural

Made from natural sources without chemicals processes

Synthetic

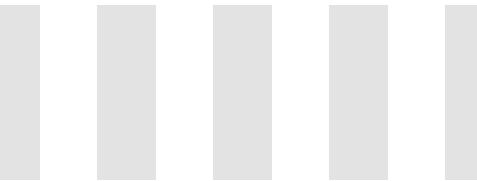
Does not occur in nature and have to be produced in laboratories

Nature Identical

Artificially produced using the same composition of natural additives

Dangerous Product Identification

E-numbers categorization



E100-E199
Colors

E400-E499
Thickeners, Stabilisers

E700 - E799
Antibiotics

E200-E299
Preservatives

E500-E599
Acidity Regulators

E900-E999
Glazing Agents,
Sweeteners

E300 - E399
Antioxidants

E600-E699
Flavor Enhancers

E1000-E1599
Additional Additives

Dangerous Product Identification

Categories defined

Safe

- Natural Additives
- Approved in the EU
- Miscellaneous

Dangerous

- Digestive Disorders
- Nervous Disorders
(Hyperactivity)
- Respiratory Problems
- Skin Problems:
Eczema

Cancer

- Hyperactivity
- Headaches
- Intestine Upset
- Skin Disorders
- Allergies & Intolerance alerts

Not Approved in the
EU

Asthmatic adverse
reactions

The data included 72% safe and 28% dangerous additives

Dangerous Product Identification



55%

Dangerous

Out of the 9,787 food products with full nutritional food facts added, 5,385 products had AT LEAST ONE DANGEROUS additive in its composition

List of the Most Dangerous Products



**SOME
PRODUCTS
HAVE UP TO
38
ADDITIVES!**

In terms of Additives

product_name	FoodSafety	n
mayonesa	Dangerous	38
ketchup	Dangerous	28
pizza barbacoa	Dangerous	21
pechuga de pavo	Dangerous	18
chorizo iberico	Dangerous	17
turrón yema tostada	Dangerous	15
filete de lomo adobado familiar	Dangerous	15
mortadela con aceitunas	Dangerous	15
turrón de yema tostada	Dangerous	14
pizza maxi barbacoa	Dangerous	13
pizza barbacoa	Dangerous	13
dulce de membrillo	Dangerous	13
paté de perdiz	Dangerous	13
patatas ali oli	Dangerous	12
pizza jamon y queso	Dangerous	12
ensalada de cangrejo	Dangerous	12
pizza carbonara	Dangerous	12
rosquillas	Dangerous	12
empiñonadas	Dangerous	11
flamenquines	Dangerous	11

List of the Most Dangerous Products

In Terms of Packaging

product_name	packaging_tags	FoodSafety
pechuga de pavo	plastic	Dangerous 18
dulce de membrillo	tarrina-de-plastico	Dangerous 13
pizza barbacoa	plastic,hdpe	Dangerous 13
chorizo extra	plastic	Dangerous 11
empiñonadas	caja,carton,plastico	Dangerous 11
bloomer de pavo braseado	bossa	Dangerous 9
queso semicurado	plastico,envasado-al-vacio	Dangerous 9
mazapán de soto	caja,carton,plastico	Dangerous 9
fromage fondu tranches	sachet-plastique,sachet-plastique-individuel	Dangerous 9
pizza atún y bacon	product	Dangerous 9
jamón cocido extra	plastic	Dangerous 8
césar american style	plastic	Dangerous 8
magro de cerdo cocido	laton	Dangerous 8
salchichón ibérico extra, loncheado	plastic	Dangerous 7
filete de lacón a la gallega	plastic	Dangerous 7
salsa barbacoa	plastic	Dangerous 7
chorizo de pavo	plastic	Dangerous 7
yum yum grilled chicken flavour	sachet	Dangerous 7
dulce de membrillo	tarrina,plastico	Dangerous 7
tigretón	plastic	Dangerous 7



Plastic Packaging seems to require additional additives to preserve the food

Recommendations

For a healthier Spain



Raise awareness about the danger of additives

- The danger of consuming dangerous additives in the long term have health implications.
- The danger of some products are in direct link with the usage of additives (pavo) / type of packaging



Relevant team involvement for quality report analysis

- Backend team : Forcing the user to input the nutritional facts by making these fields mandatory
- Analytics team: to handle data formalization for more quality reports.

THE TEAM



ALEXANDRE BOUAMAMA



DENNIS PEDERSEN



EUGEN WETTSTEIN



MILAN MEDINA



TONY MATTÀ



UMUT VAROL



VIMAL KRISHNAN