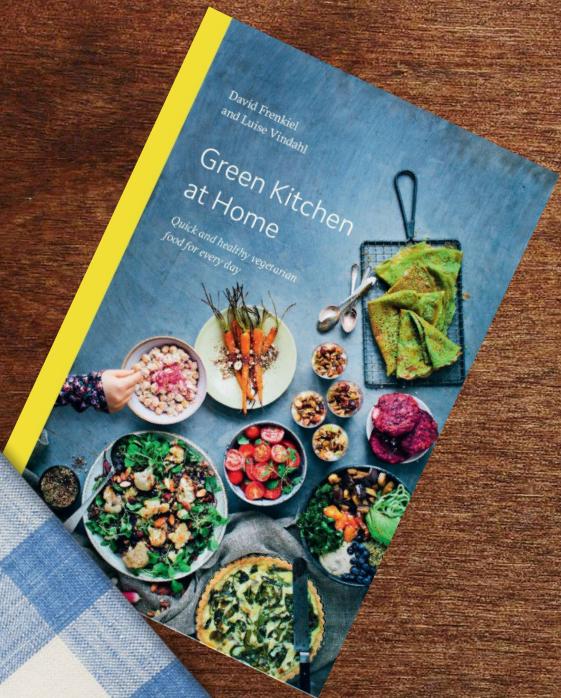
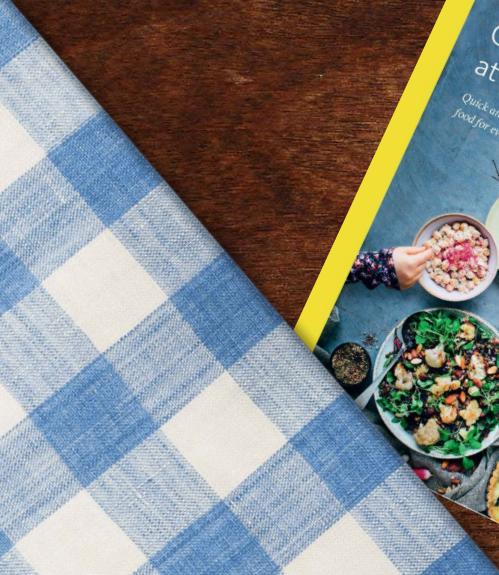
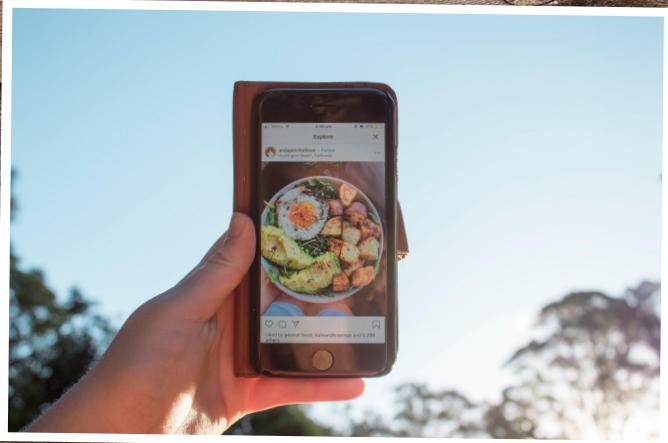


Trend analysis: Evolution of healthy food over the last decade in western countries



Marc Pou
Data Analytics 2020



The Project

Overview of the entire process



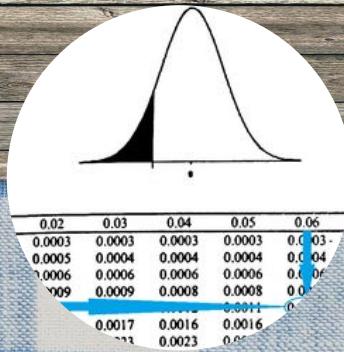
Food.com

Data selected, Processing
(Cleaning, Wrangling, EDA)
and analysis



Google Trends

First attempt, scraping
Google trends, keywords
used, exploration and
visualizing



Analysis

Defining our H₀, applying
inferential statistics, Tests



Conclusions

Conclusions extracted from
the analysis



HOT THIS WEEK: THE TOP 50 RECIPES OF 2020, FAN FAVES

Food.com

- Data set from Kaggle
- Clean

Selected Time: 2008 to 2018

Removed outliers: "How to preserve a husband"

Minutes: 20 up to 180

Convert nutritional values to columns

Understand Nutritional Values (PDV)

	name	id	minutes	submitted
109624	how to preserve a husband	447963	1051200	2011-02-01

	name	id	minutes	submitted	tags	n_steps	steps	description	ingredients	n_ingredients	year	calories	total_fat	tags	n_steps	steps	description	ingredients	n_ingredients	year	calories	total_fat	sugar	sodium	protein	saturated_fat	carbohydrates
115	1 in canada chocolate chip cookies	453467	45	2011-04-11	['60-minutes-or-less', 'time-to-make', 'cuisin...', ...]	12	['pre-heat oven the 350 degrees f.', 'in a mixi...']	this is the recipe that we use at my school ca...	['white sugar', 'brown sugar', 'salt', 'margar...']	11	2011	595.1	46.0	['60-minutes-or-less', 'time-to-make', 'cuisin...', ...]	12	['pre-heat oven the 350 degrees f.', 'in a mixi...']	this is the recipe that we use at my school ca...	['white sugar', 'brown sugar', 'salt', 'margar...']	11	2011	595.1	46.0	211.0	22.0	13.0	51.0	26.0
125	2000 meatloaf	475785	90	2012-03-06	['time-to-make', 'course', 'main-ingredient', ...]	17	['pan fry bacon, and set aside on a paper tow...']	ready, set, cook special edition contest entr...	['meatloaf mixture', 'unsmoked bacon', 'goat c...']	13	2012	267.0	30.0	['time-to-make', 'course', 'main-ingredient', ...]	17	['pan fry bacon, and set aside on a paper tow...']	ready, set, cook special edition contest entr...	['meatloaf mixture', 'unsmoked bacon', 'goat c...']	13	2012	267.0	30.0	12.0	12.0	29.0	48.0	2.0
135	blepandekager danish apple pancakes	503475	50	2013-07-08	['danish', '60-minutes-or-less', 'time-to-make...', ...]	10	['beat the eggs lightly and add the milk', 'co...']	this recipe has been posted here for play in z...	['eggs', 'milk', 'flour', 'sugar', 'salt', 'cr...']	10	2013	358.2	30.0	['danish', '60-minutes-or-less', 'time-to-make...', ...]	10	['beat the eggs lightly and add the milk', 'co...']	this recipe has been posted here for play in z...	['eggs', 'milk', 'flour', 'sugar', 'salt', 'cr...']	10	2013	358.2	30.0	62.0	14.0	19.0	54.0	12.0
136	Iplermagrone	522861	50	2015-07-25	['60-minutes-or-less', 'time-to-make', 'course...']	10	['a variant on the theme in the swiss canton ...']	alplermagronethe name doesn't translate perfec...	['milk', 'salt', 'macaroni', 'cheese', 'fres...']	8	2015	1003.8	72.0	['60-minutes-or-less', 'time-to-make', 'course...', ...]	10	['a variant on the theme in the swiss canton ...']	alplermagronethe name doesn't translate perfec...	['milk', 'salt', 'macaroni', 'cheese', 'fres...']	8	2015	1003.8	72.0	21.0	103.0	69.0	143.0	37.0



Food.

[RECIPES](#)[POPULAR](#)[HOLIDAY CENTRAL](#)

HOT THIS WEEK: THE TOP 50 RECIPES OF 2020, FAN FAVES

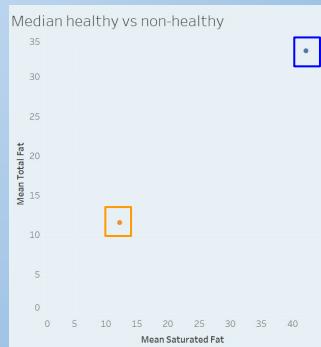
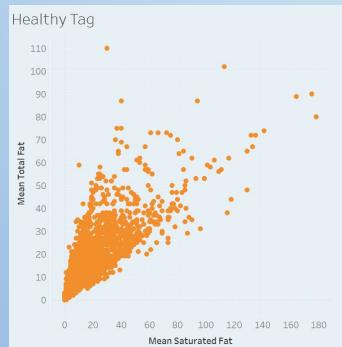
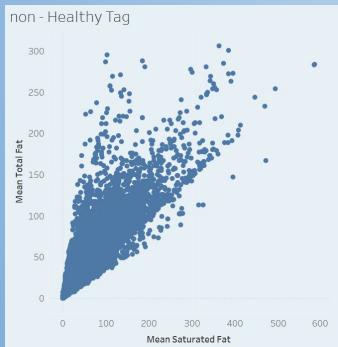
Food.com

- Exploring healthy as tag:

```
recipes['healthy'] = ['healthy' in tag for tag in recipes['tags']]
recipes_tags = recipes.drop(columns=['name', 'tags', 'steps', 'description', 'ingredients'])
recipes_tags.head(2)
```

	id	minutes	submitted	n_steps	n_ingredients	year	calories	total_fat	sugar	sodium	protein	saturated_fat	carbohydrates	healthy	
	115	453467	45	2011-04-11	12	11	2011	595.1	46.0	211.0	22.0	13.0	51.0	26.0	False
	125	475785	90	2012-03-06	17	13	2012	267.0	30.0	12.0	12.0	29.0	48.0	2.0	False

1. Healthier recipes means lower levels of saturated and total fat ?





Food.

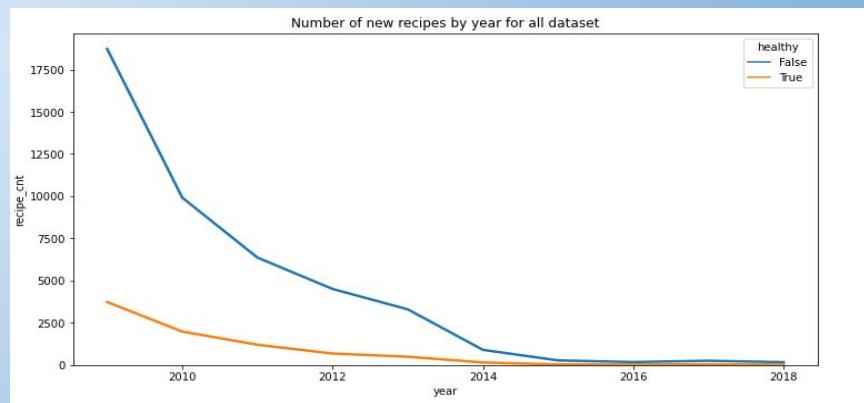
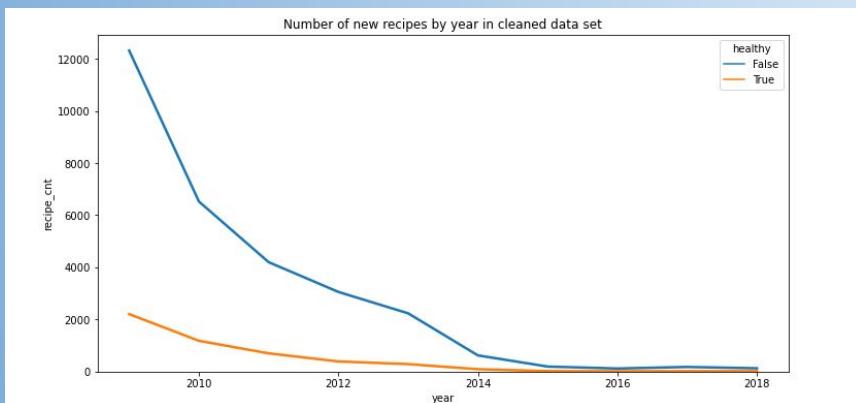
[RECIPES](#)[POPULAR](#)[HOLIDAY CENTRAL](#)

HOT THIS WEEK: THE TOP 50 RECIPES OF 2020, FAN FAVES

Food.com

- Exploring healthy as tag:

2. How is the trend of new recipes per year?





Food.

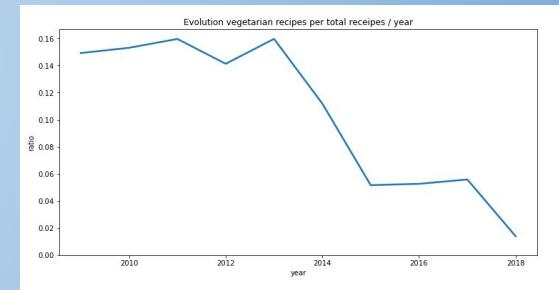
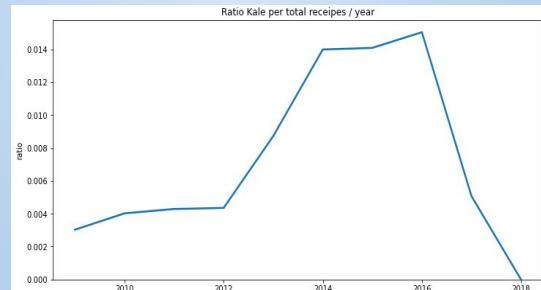
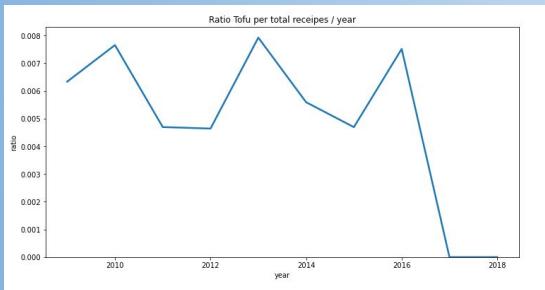
[RECIPES](#)[POPULAR](#)[HOLIDAY CENTRAL](#)

HOT THIS WEEK: THE TOP 50 RECIPES OF 2020, FAN FAVES

Food.com

- Exploring other tags:

3. Possible other tags and tendencies



- What is happening?

- Healthy food becoming less popular and I am completely wrong
- Recipe webs are losing market share and new platforms such as social media or video-sharing are becoming more popular



Taylor Swift • Kim Kardashian

World Cup

Football • American football

Google Trends

- What is ? Is it possible to scrape ? What did I use?

Pytrends

Interest Over Time

- Attempt 1: Generate a list of recipes names with Tofu as an ingredient and analyze its google trend (2008 to 2018, for US)
- Attempt 2: With my key words, create a list and scrape Google Trends (2008 to 2018, for US)

```
In [24]: M directory = "/Users/marcop/Desktop/Ironhack/Clases/FINAL PROJECT/Project 2/Google Trends scrapping 2"
ku_list = ['tofu','kale','vegetarian','healthy food']

In [25]: M category=''
time='2008-01-01 2018-12-31'
loc='US'

pytrends.build_payload(ku_list, cat=0, timeframe=time, geo=loc, gprop='')

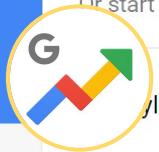
df_time = pytrends.interest_over_time()
df_time.reset_index(inplace=True)

In [26]: M df_time
Out[26]:
   date      tofu      kale vegetarian healthy food  isPartial
0 2008-01-01  27     10       70        17    False
1 2008-02-01  25     10       70        15    False
2 2008-03-01  25     10       65        16    False
3 2008-04-01  25      9       62        17    False
4 2008-05-01  23      9       62        16    False
...
127 2018-08-01  36     40       77        30    False
128 2018-09-01  35     39       75        29    False
129 2018-10-01  34     38       75        28    False
130 2018-11-01  33     40       82        25    False
131 2018-12-01  33     38       81        24    False
132 rows × 6 columns
```

Simplified code from Attempt 1, that was inspired in a Github Google Trends Analysis

Or start with an example

HIDE



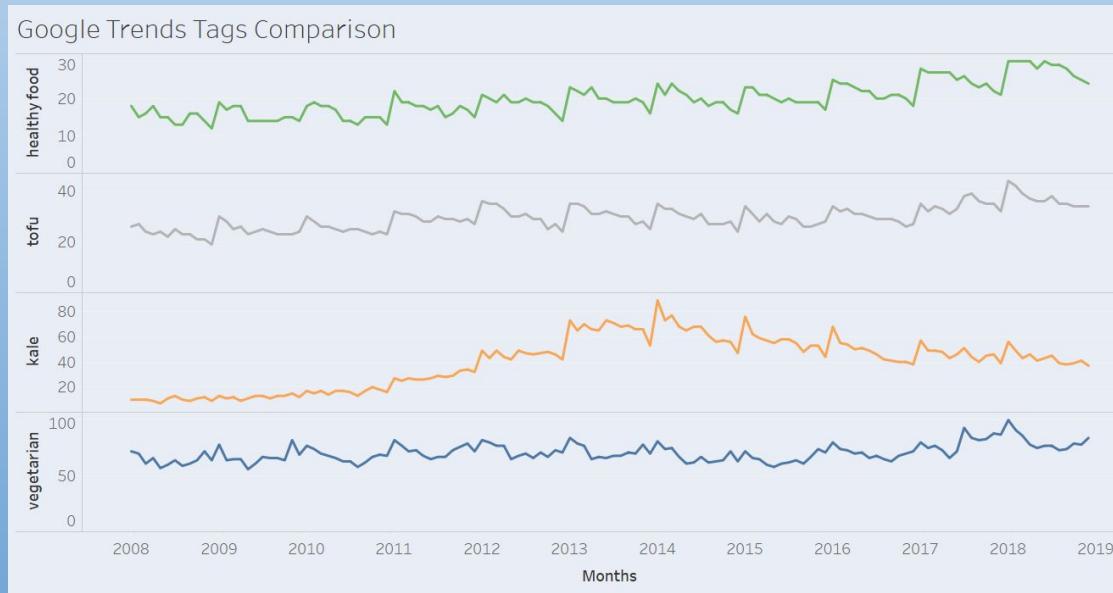
Taylor Swift ● Kim Kardashian

● World Cup

● Football ● American football

Google Trends

- Comparing the evolution of our keywords



Statistical Analysis

- Applying inferential statistics to evaluate our H0

Tofu's trends results for the analysis

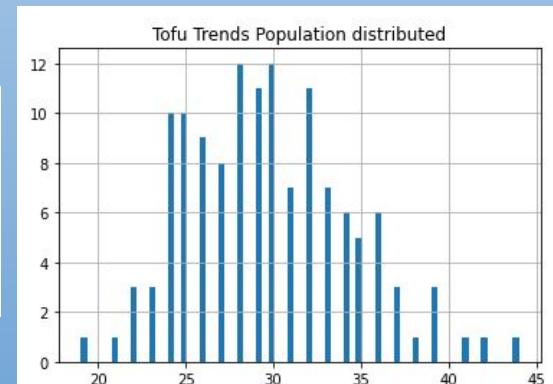
Hypothesis: Between the years of 2008 and 2018 the number of mentions of tofu increased significantly in the US

H0: mean sample = mean population

Question = Is the group significantly different from the regular population ?

Model: z-statistic with 95% of confidence -- 1.96 rejection area

	Population	(1) Random Sample All	(2) Random Sample 1m/y	(3) Sample_2017	(4) Sample_2018
n	132	12	12	12	12
Mean	29,68	32,17	29,45	34,33	37,33
stdv	4,69	5,44	4,93	2,67	3,60
z -statistic	-1,96 < > 1,96	1,84	-0,17	3,43	5,65
Reject H0		No	No	Yes	Yes





CONCLUSIONS

- Analytical Conclusions
- Project Management
- Improvise, Adapt and Overcome
- Future steps





THANKS !

Does anyone have any questions?