

YAK SHAVING A GOOD PLACE TO EAT USING NON NEGATIVE MATRIX FACTORIZATION

EUROPYTHON 2015

Adriano Petrich

petrich@gmail.com

@fractal

DISCLAIMER

My opinions not Red Hat's opinions

Mostly because I don't know Red Hat's opinions about
restaurants or linear algebra

I'm just lazy at creating styles for slides
and this looked ok

All this talk was created after having multiple
bad pizzas from multiple places with 5 stars
reviews of "Best pizza in <foo>" and a good
book

I don't know Red Hat's opinion about pizza also.



WHO AM I







A Brazilian pythonista that moved to Scotland

FOR THE WEATHER












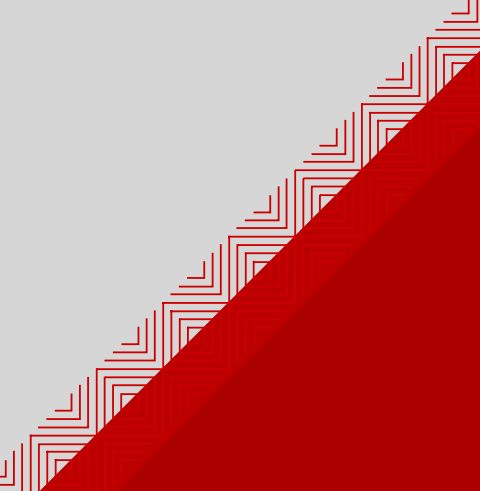
The image features a light gray background with two large, diagonal red geometric patterns in the corners. These patterns consist of nested, stepped lines that create a sense of depth and movement. The text is centered in the middle of the frame.

**WHAT IS A GOOD
PLACE TO EAT?**



the #1 at any restaurant site is a safe bet

This is about the #119 or the #314 that's where quirkiness lives



The image features a light gray background with two large, diagonal red geometric patterns in the corners. These patterns consist of nested, stepped lines that create a sense of depth and movement. The text is centered in the middle of the frame.

**SO HOW DO WE
FIND THEM?**

The background features a light gray gradient with two prominent red geometric patterns in the corners. The top-left corner has a solid red triangle with a series of white, nested, stepped lines extending from its hypotenuse. The bottom-right corner has a similar pattern, with a solid red triangle and white stepped lines extending from its hypotenuse.

STARS AND RATINGS



WHAT DOES THIS TELL US?

RESTAURANT A

Sakrow gave 5 stars

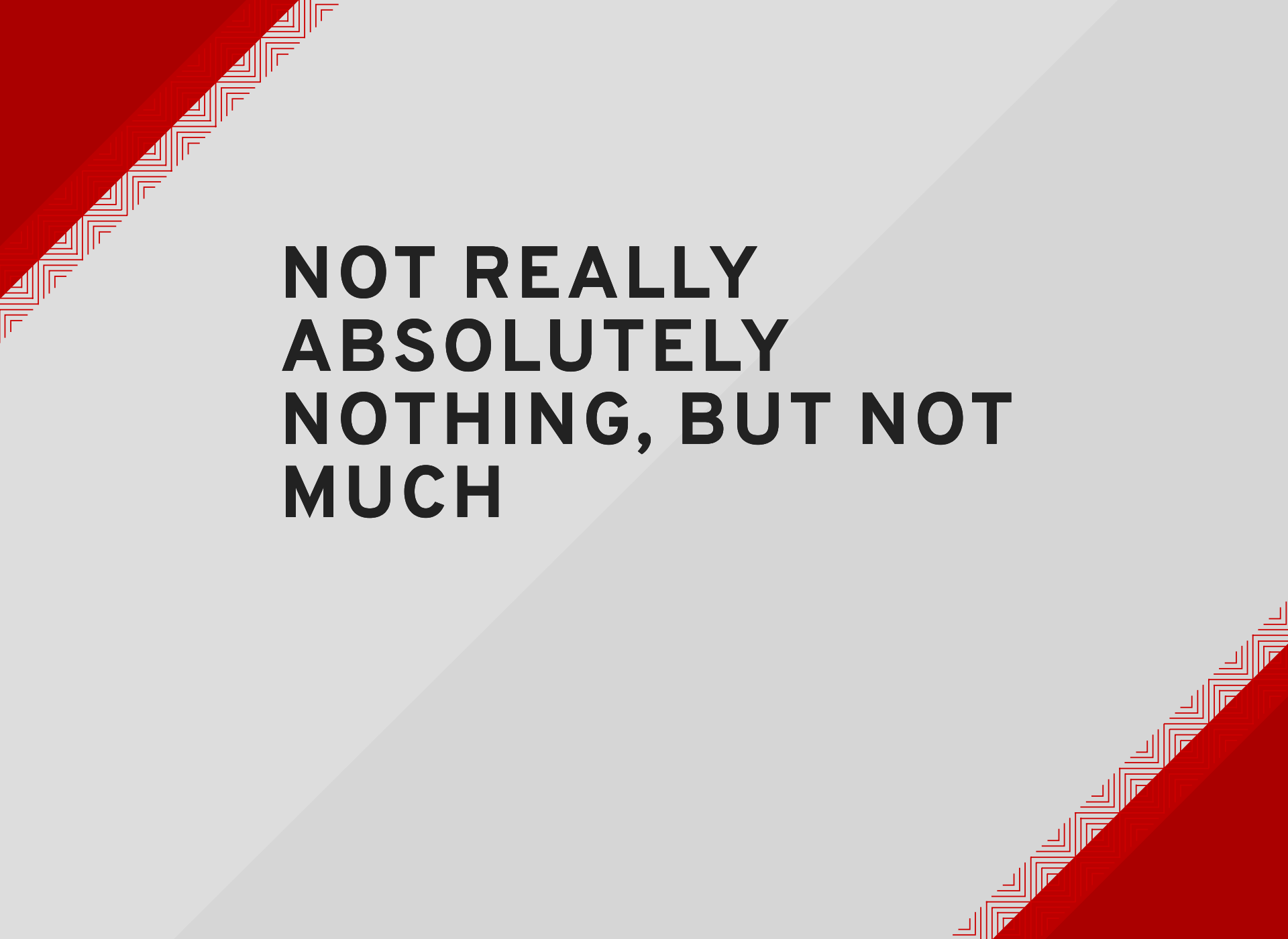
RESTAURANT B

ピーターサム gave 4 stars





**ABSOLUTELY
NOTHING**



**NOT REALLY
ABSOLUTELY
NOTHING, BUT NOT
MUCH**



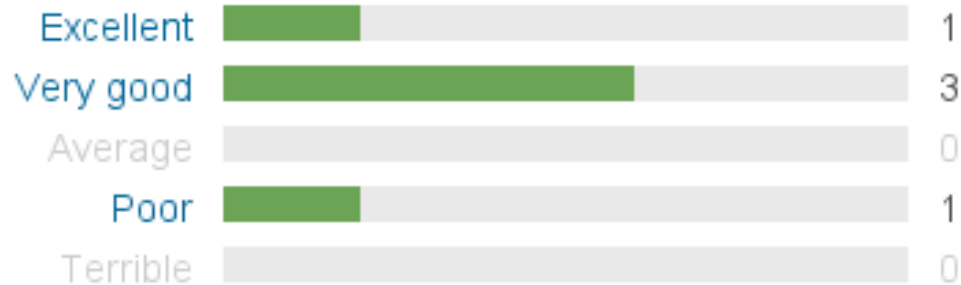
**RATINGS ALONE
JUST DON'T CUT
IT**

The slide features a light gray background with two large red triangular shapes in the corners. The top-left triangle is solid red, while the bottom-right triangle is also solid red but contains a white geometric pattern of nested squares and lines. The title text is centered in the middle of the slide.

NUMBER AND DISTRIBUTION OF RATINGS

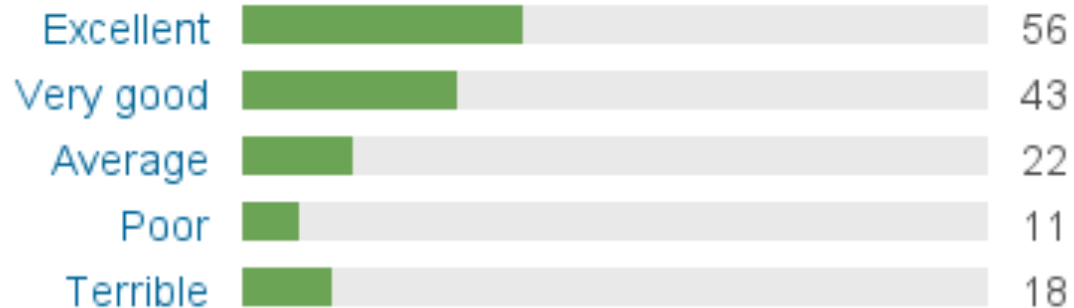
RESTAURANT A

Visitor rating



RESTAURANT B

Visitor rating



The image features a light gray background with two large, diagonal red geometric patterns in the corners. These patterns consist of nested, stepped lines that create a sense of depth and movement. The text is centered in the middle of the frame.

**THE AMOUNT OF RATINGS
HELP BUT DOESN'T SOLVE
THE PROBLEM**



**AND NOW
FOR SOMETHING**

**COMPLETELY
DIFFERENT!**





**A MAN WITH
LINEAR ALGEBRA
UP HIS NOSE!**

The background features a light gray diagonal split. In the top-left and bottom-right corners, there are red triangular areas. These red areas are decorated with a series of concentric, stepped lines that create a 3D, architectural effect, resembling a staircase or a series of nested squares.

LINEAR ALGEBRA!

WHEN IS THIS TRUE?

$$A \times B \neq B \times A$$



YES!

DIRTY CLOTHES!

Washing \times Drying \neq Drying \times Washing

IS NOT COMMUTATIVE

Thanks wikipedia for the weirdest latex math equation I've ever typed!

https://en.wikipedia.org/wiki/Commutative_property

Also matrix multiplication

$$A_{2,3} = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{vmatrix}$$

$$B_{3,2} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{vmatrix}$$

$$A \times B = \begin{vmatrix} 22 & 28 \\ 49 & 64 \end{vmatrix}$$

$$B \times A = \begin{vmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \\ 29 & 40 & 51 \end{vmatrix}$$

OBS: I'm using "x" to represent multiplication from now on

The image features a light gray background with two large, diagonal, light gray triangular sections. In the top-left and bottom-right corners, there are red geometric patterns consisting of nested, stepped lines that create a sense of depth and movement.

MATHEMATICAL!

THE DIMENSION VANISHING TRICK

$$\begin{array}{c} A_{x,n} \times B_{n,y} = C_{x,y} \\ \uparrow \quad \uparrow \end{array}$$

THE VANISHING TRICK

$$A_{x, \text{🍕}} \times B_{\text{🍕}, y} = C_{x, y}$$



**HE WHO CAN
DESTROY A THING,
CONTROLS A
THING.**

Paul Muad'Dib

São Paulo City in 1893

| Year | Immigrants | Percentage of the City [8] |
|------------|------------|----------------------------|
| Italians | 45,457 | 35% |
| Portuguese | 14,437 | 11% |
| Spanish | 4,818 | 3.7% |

Immigrants established in São Paulo state in 1940

| Immigrants | Population [45] |
|------------|-----------------|
| Italians | 694.489 |
| Spanish | 374.658 |
| Portuguese | 362.156 |
| Japanese | 85.103 |
| Germans | 50.507 |
| Austrians | 33.133 |

https://en.wikipedia.org/wiki/Immigration_to_Brazil

HYPOTHESIS:

People that have the same background and judge food using the same standards

Pizza Na Mao

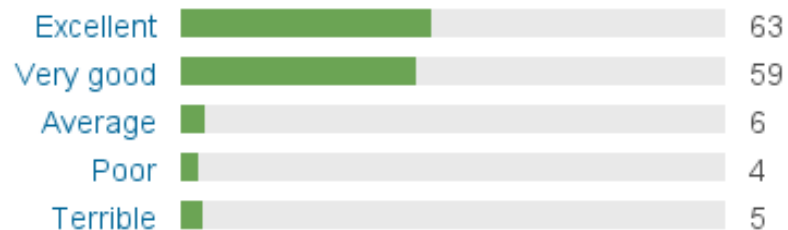
●●●●● 137 Reviews | #396 of 32,634 Restaurants in Sao Paulo | #445 of 35,018 Places to Eat in Sao Paulo | 🏆 Certificate of Excellence



Jardins

Pizza, Pizza & Pasta

Visitor rating



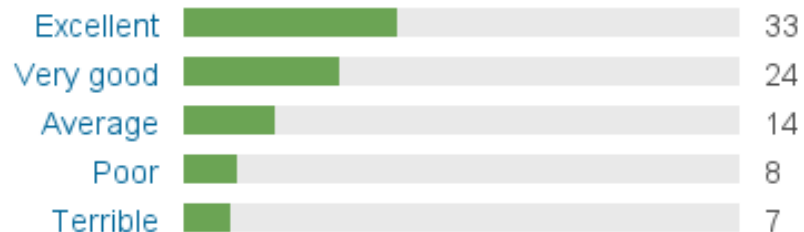
True Pizza Co

●●●●● 86 Reviews | #89 of 282 Restaurants in Dundee | #101 of 305 Places to Eat in Dundee

££

Italian, French, Pizza

Visitor rating



Rio Oja



174 Reviews

#119 of 781 Restaurants in Bilbao

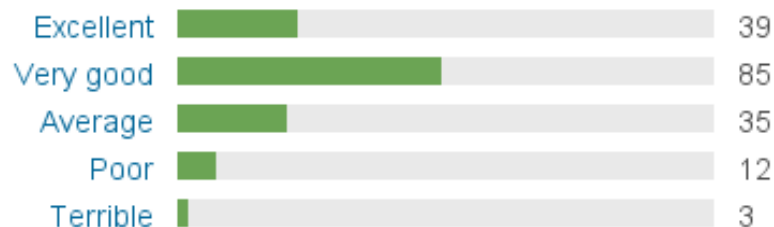
#130 of 820 Places to Eat in Bilbao



Bilbo Zaharra

+ Add cuisine

Visitor rating



The Tailend



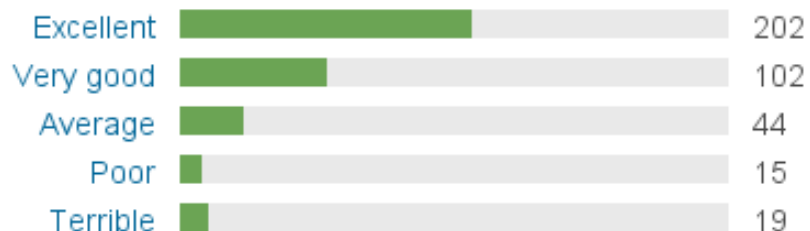
382 Reviews

#578 of 1,806 Restaurants in Edinburgh

#655 of 2,025 Places to Eat in Edinburgh

Scottish

Visitor rating



The background features a light gray gradient with two prominent red geometric patterns in the corners. The top-left corner has a solid red triangle with a series of white, nested, stepped lines extending from its hypotenuse. The bottom-right corner has a similar pattern, with a solid red triangle and white stepped lines extending from its hypotenuse.

**THUS PROVING MY
HYPOTHESIS!**

HOW TO MAKE A SCIENTIST'S HEAD EXPLODE:

ANECDOTAL EVIDENCE
ISN'T VALID.

YES IT IS! I ONCE
USED AN ANECDOTE AS
EVIDENCE, AND LATER
IT TURNED OUT I
WAS RIGHT!



© smbc-comics.com

<http://www.smbc-comics.com/index.php?db=comics&id=2159/>



This talk is not really data science.

I didn't want to go into a Gaussian
analysing rampage.

this is based on a hunch



THIS HUNCH:

What constitutes a good place to eat is deeply related to an individual background

AND THIS ONE:

The more that you agree with multiples reviews from
SAKROW

The more likely that you might like another
restaurant that she likes

The background features a light gray gradient with two prominent red geometric patterns in the corners. The top-left corner has a solid red triangle with a series of white, nested, stepped lines extending from its hypotenuse. The bottom-right corner has a similar pattern, with a solid red triangle and white stepped lines extending from its hypotenuse.

LINEAR ALGEBRA WITH FOOD

SUPPOSE I HAVE A LOAD OF DATA LIKE THIS

Pilar L gave 4 stars to Restaurante Ambigu

ピーターサム gave 4 stars to Rio Oja

....

{{ user_u }} gave {{ s }} stars to {{ restaurant_r }}

Users



Restaurants



| | | |
|---------------------------|----------|-----------------|
| $Stars_{user_1,rest_1}$ | ... | S_{u_n,r_1} |
| $S_{user_1,rest_2}$ | ... | S_{u_n,r_1} |
| \vdots | \ddots | \vdots |
| S_{u_1,r_n} | ... | S_{u_n,r_n} |

LET'S CALL IT M

M_{restaurants,users}

Just to be dench*

* sorry it is a UK thingie

<http://www.urbandictionary.com/define.php?term=Dench>

CREATE A DIMENSION

The same way you can vanish one

$$C_{x,y} = A_{x, \text{🍕}} \times B_{\text{🍕}, y}$$

where 🍕 is any emoji number that I want

And create another matrix C so that

$$M_{r,u} \approx C_{r,u}$$

In a way that I can calculate R and U

$$C_{r,u} = R_{r, categories} \times U_{categories, u}$$

$$R_{r, categories}$$

Is a matrix where each restaurant is weighted in each category

$$U_{categories, u}$$

The same way that U a matrix where each user is weighted in each category



NON NEGATIVE MATRIX FACTORIZATION!

Is a way of weighing data in
automatic categories

$$M_{r,u} \approx C_{r,u} = R_{r, \text{categories}} \times U_{\text{categories}, u}$$

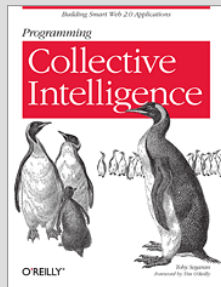
THE NON NEGATIVE PART OF IT

- It facilitates a lot of the methods that we can use to generate R and U
- in our case it is good that stars are `range(6)`



THE NEXT PAGE HAS CAT.GIF

```
def factorize(v, pc=10, interactions=50):  
    # Initialize the R and U matrices with random values  
  
    # Perform operation a maximum of iter times  
    for i in range(interactions):  
  
        # Calculate the current difference  
  
        # Terminate if the matrix has been fully factorized  
  
        # Update R matrix  
        # Update U matrix  
    return R, U
```



Programming the collective intelligence

Toby Seagran

<http://shop.oreilly.com/product/9780596529321.do>



<http://akamaicovers.oreilly.com/images/9780596529321/cat.gif>

DEMO TIME!

http://localhost:8888/notebooks/EP2015_yakshaving_with_nmf.ipynb#

OBS: I'm not publishing the dataset because I'm not sure about scraped data, but talk to me if you want it

YAK SHAVING

"You see, yak shaving is what you are doing when you're doing some stupid, fiddly little task that bears no obvious relationship to what you're supposed to be working on, but yet a chain of twelve causal relations links what you're doing to the original meta-task."

THANK YOU

Adriano Petrich

petrich@gmail

@fractal

slides at:

<http://redhat.slides.com/apetrich/yak-shaving-a-good-place-to-eat-using-non-negative-matrix-factorization>

notebook at [github/frac/ep2015](https://github.com/frac/ep2015)