# 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

# WHAT IS A GOOD PLACE TO EAT?

the #1 at any restaurant site is a saf
e bet

This is about the #119 or the #314 th
at's where quirkiness live

# SO HOW DO WE FIND THEM?

# 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

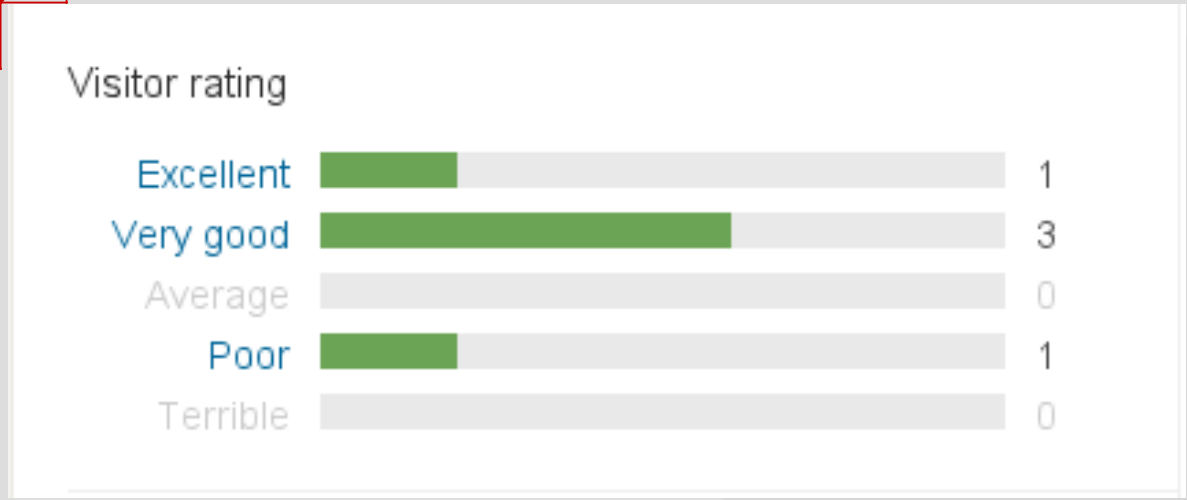
ABSOLUTELY
NOTHING

# NOT REALLY ABSOLUTELY NOTHING, BUT NOT MUCH

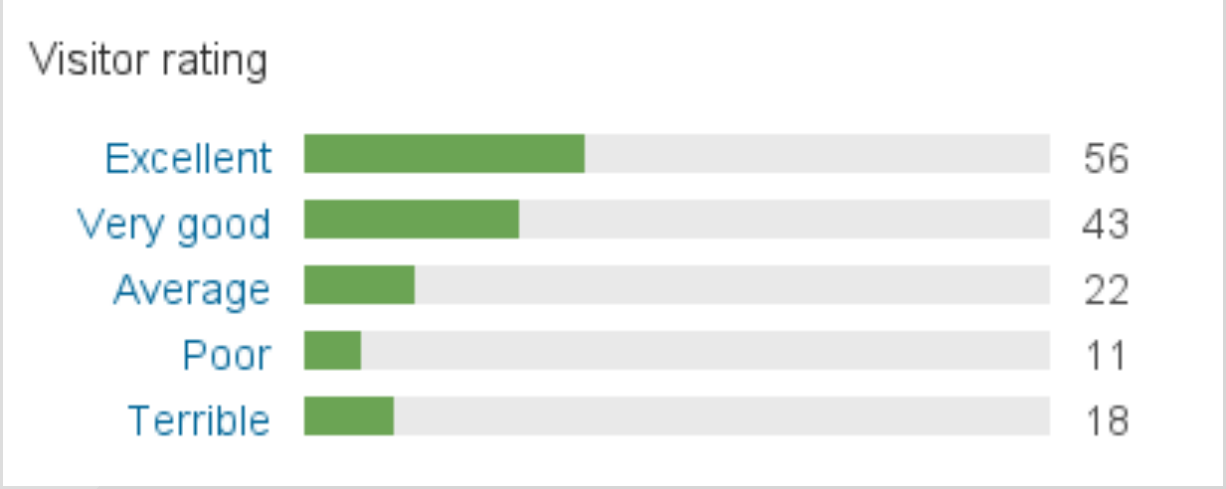
# RATINGS ALONE JUST DON'T CUT IT

# NUMBER AND DISTRIBUTION OF RATINGS

# RESTAURANT A




# RESTAURANT B



# 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!

# LINEAR ALGEBRA!

# WHEN IS THIS TRUE?

$$AxB \neq BxA$$

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

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

$$AxB = \begin{vmatrix} 22 & 28 \\ 49 & 64 \end{vmatrix}$$

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

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

# MATHEMATICAL!

# THE DIMENSION VANISHING TRICK

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

# 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

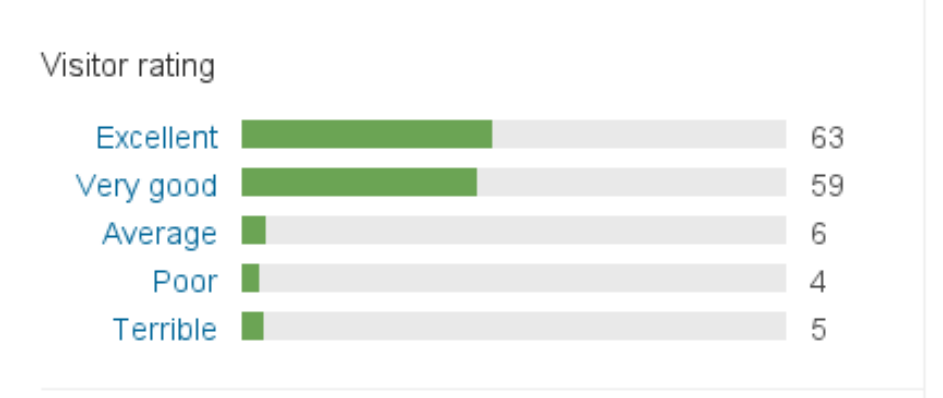
# HYPOTHESIS:

People that have the same backgro
und judge food using the same sta
ndards

## 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

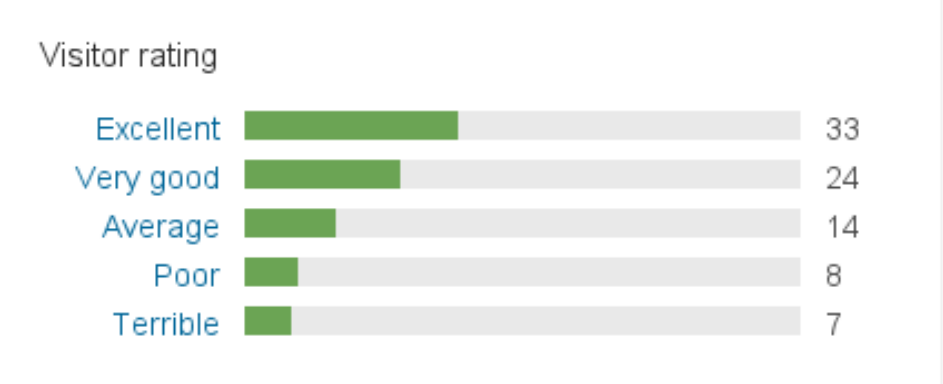
Jardins | Pizza, Pizza & Pasta




## True Pizza Co

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

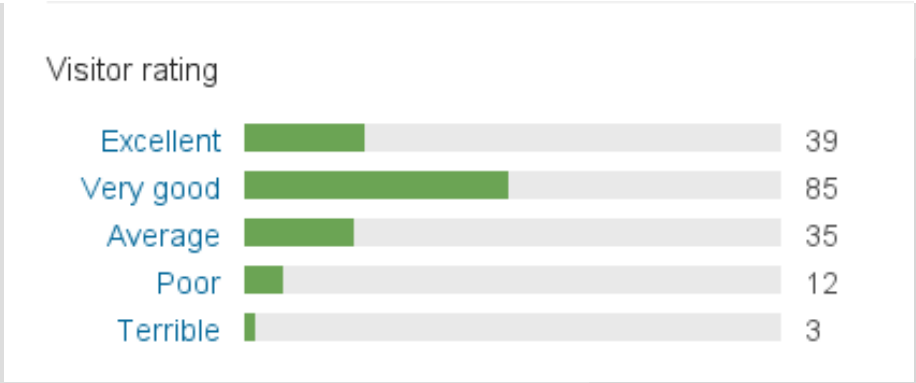
££ | Italian, French, Pizza



# Rio Oja

174 Reviews | #119 of 781 Restaurants in Bilbao | #130 of 820 Places to Eat in Bilbao

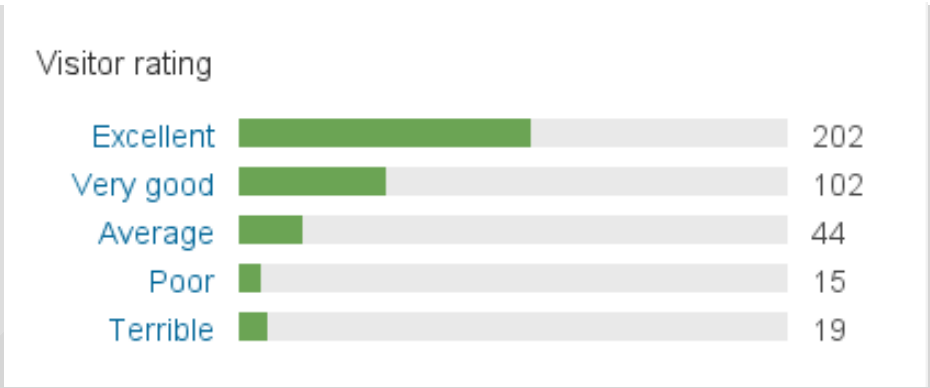
Bilbo Zaharra | + Add cuisine




# The Tailend

382 Reviews | #578 of 1,806 Restaurants in Edinburgh | #655 of 2,025 Places to Eat in Edinburgh

Scottish



# THUS PROVING MY HYPOTHESIS!




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 ea
t is deeply related to an individual b
ackground

# 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

# LINEAR ALGEBRA WITH FOOD

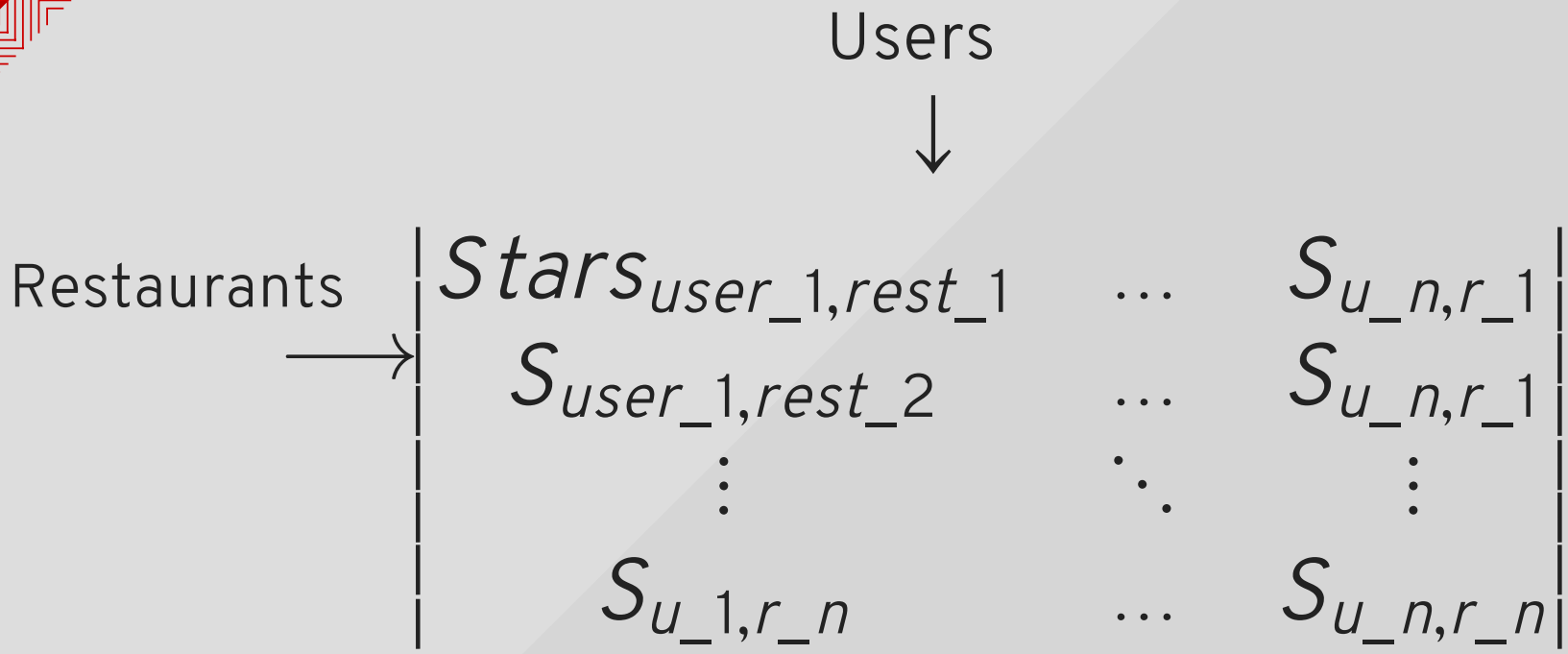
# SUPPOSE I HAVE A LOAD OF DATA LIKE THIS

**Pilar L** gave **4** stars to **Restaurante Ambigu**

**ピーターサム** gave **4** stars to **Rio Oja**

....

{{ **user_u** }} gave {{ **s** }} starts to {{ **restaurant_r** }}


Users
Restaurants
$Stars_{user\_1,rest\_1}$ … $S_{u\_n,r\_1}$
$S_{user\_1,rest\_2}$ … $S_{u\_n,r\_1}$
$S_{u\_1,r\_n}$ … $S_{u\_n,r\_n}$

# LET'S CALL IT M

*M*restaurants,users

Just to be dench*

* sorry it is a UK thinguie

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 weigthing data in automatic categories

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

# 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

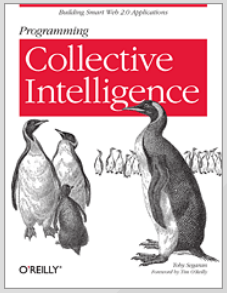
```
def factorize(v, pc=10, iteractions=50):
    # Initialize the R and U matrices with random values

    # Perform operation a maximum of iter times
    for i in range(iteractions):

        # 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 doin g when you're doing some stupid, fiddly lit tle 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 met a-task."

http://projects.csail.mit.edu/gsb/old-archive/gsb-archive/gsb2000-02-11.html

# 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