

Intuitive Decision Making Using NLP



Bastin Robins .J
Chief Data Scientist
CleverInsight

Back in 2000 , **People Magazine** PUBLISHER highlighted **Prince Williams'** PERSON style who at the time was a little more fashion-conscious , even making fashion statements at times .

Now-a-days the prince mainly wears **navy** COLOR **suits** ITEM (sometimes **double-breasted** DESIGN) , **light blue** COLOR **button-ups** ITEM with **classic** LOOK **pointed** DESIGN **collars** PART , and **burgundy** COLOR **ties** ITEM .

But who knows what the future holds ...

Duchess Kate PERSON did wear an **Alexander McQueen** BRAND **dress** ITEM to the **wedding** OCCASION in the **fall of 2017** SEASON .

Can you please come here ?

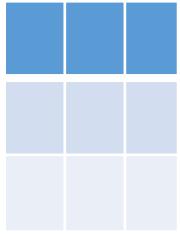


History



Word being predicted

Tabular



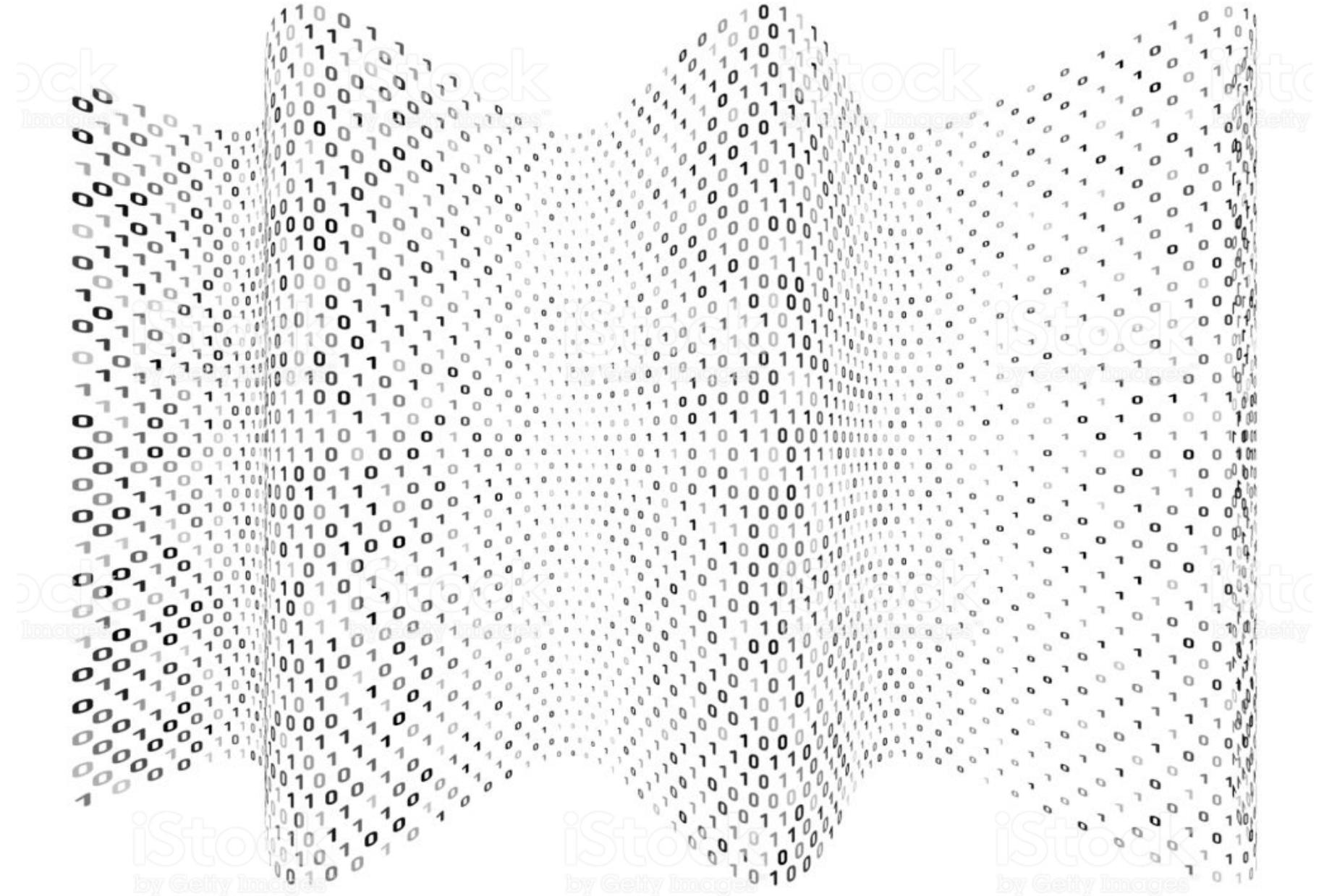
Text



Audio



Video



"The cat was a friend of rat"

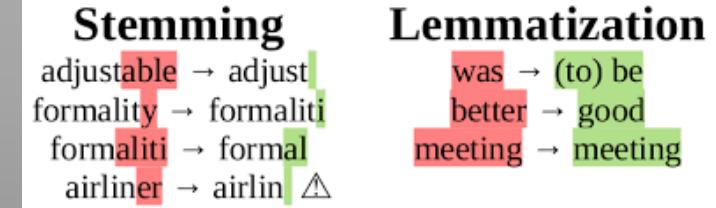
Tokenization

"the", "cat", "was", "a",
"friend", "of", "rat"

Stemming

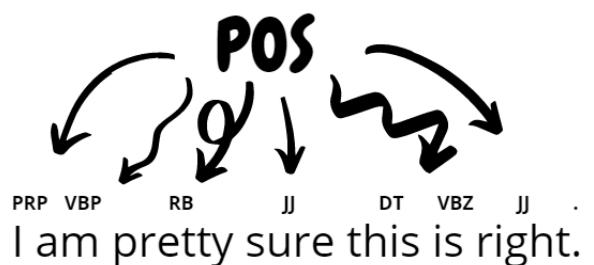
Words	Stemmed words
friend	friend
friendly	friend
friended	friend

Lemmatization



Eg: map gone, going and went into go

POS Tags



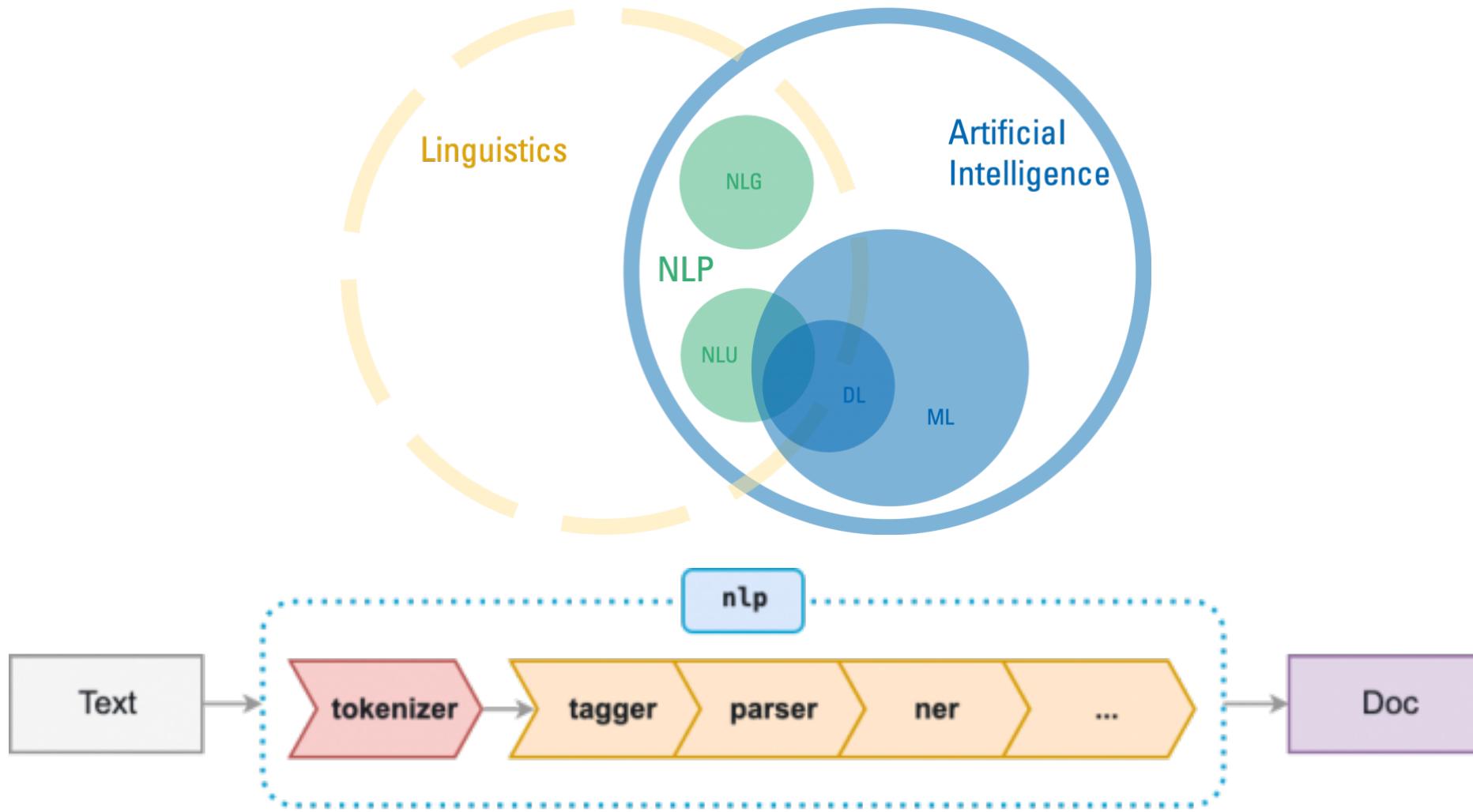
Entity Recognition

Automatically find names
of people, places, products,
and organizations in text
across many languages.

Chunking



NLP



NLP

Data > Information > Knowledge > Wisdom

this →	0.2	0.4	-0.3
movie →	0.1	0.2	0.6
has →	-0.1	0.4	-0.1
amazing →	0.7	-0.5	0.4
diverse →	0.1	-0.2	0.1
characters →	0.6	-0.3	0.8



Myers-Briggs Type Indicator





PUBLISHED JAN. 18, 2018, AT 9:00 AM

Most Personality Quizzes Are Junk Science. Take One That Isn't.

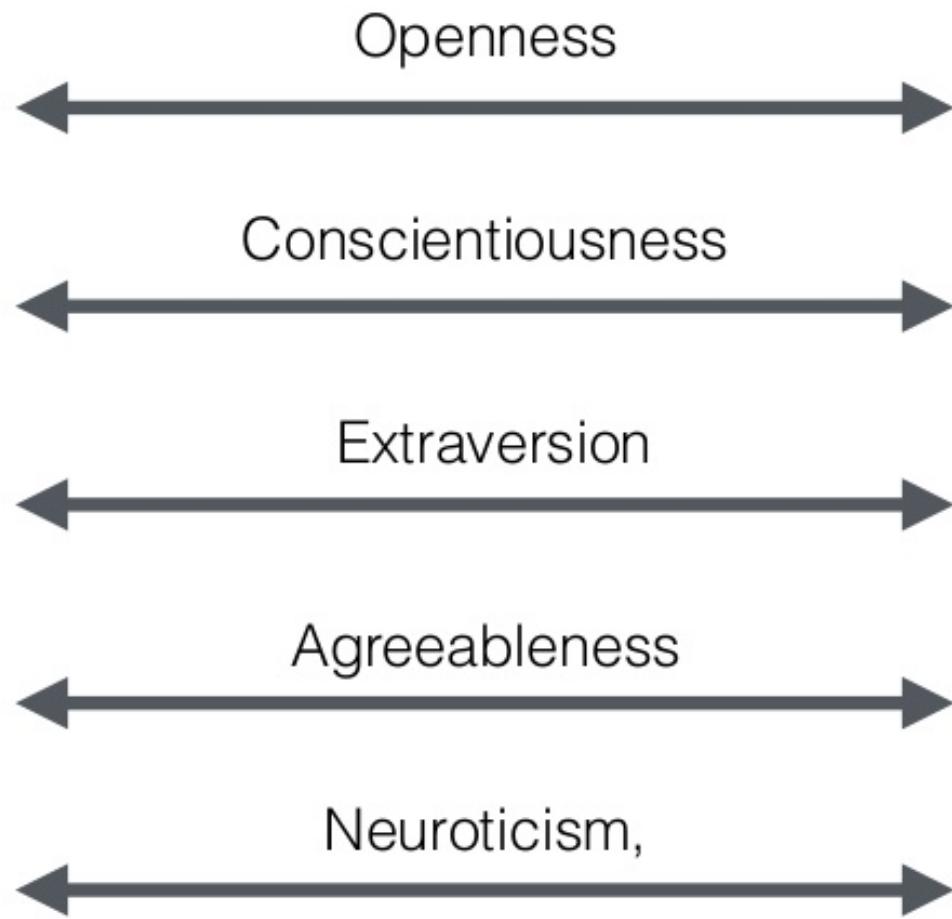
Compare your results to those of your friends and family.

By [Maggie Koerth-Baker](#) and [Julia Wolfe](#)

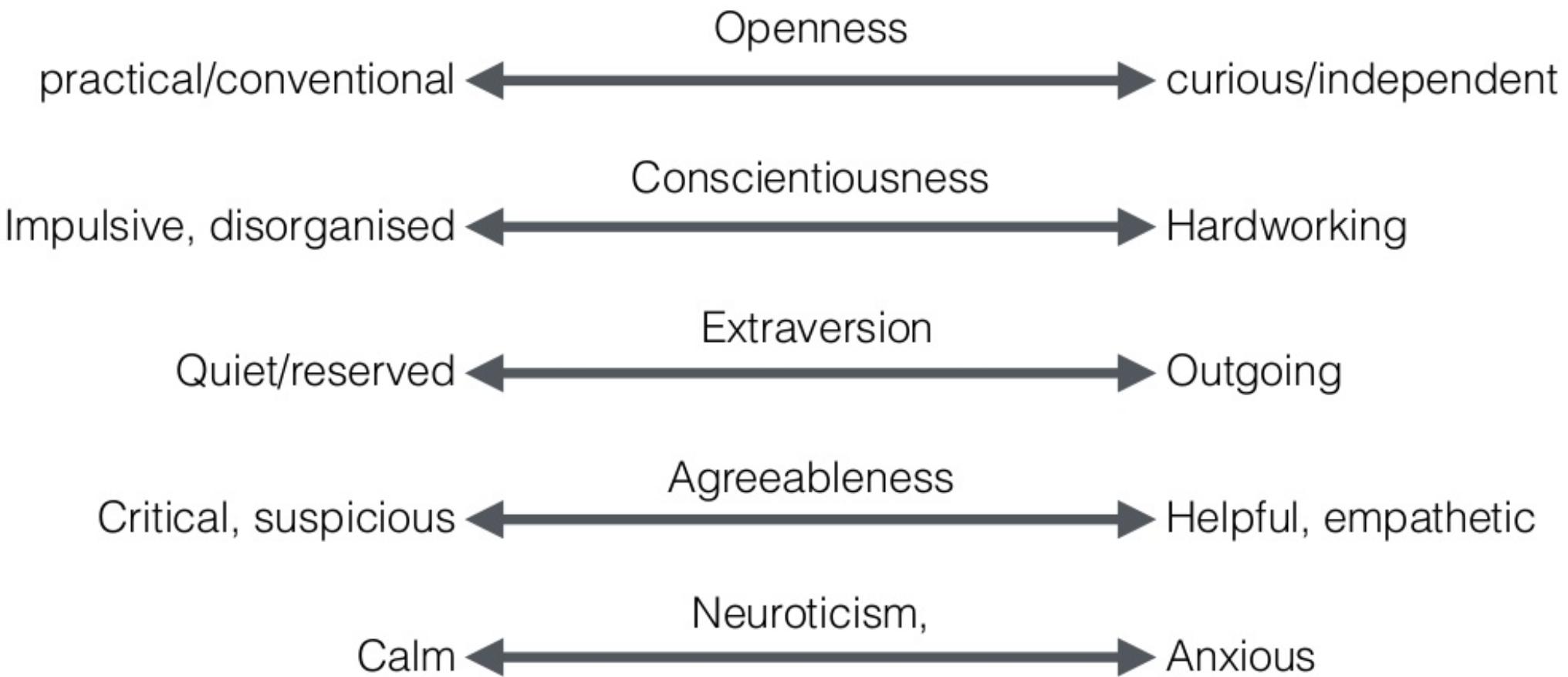
What's your personality, and what can it tell you about your true self? Those questions have launched a thousand online personality quizzes. But you can

<https://projects.fivethirtyeight.com/personality-quiz/>

Big Five Personality Traits

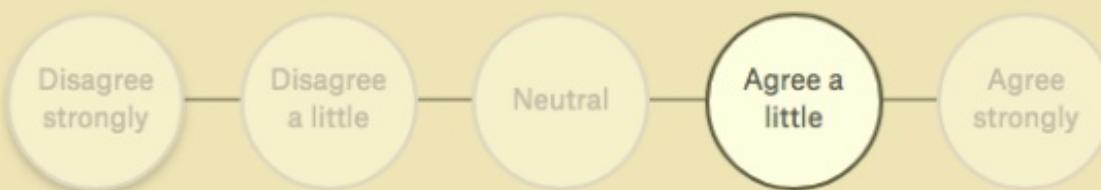


Big Five Personality Traits

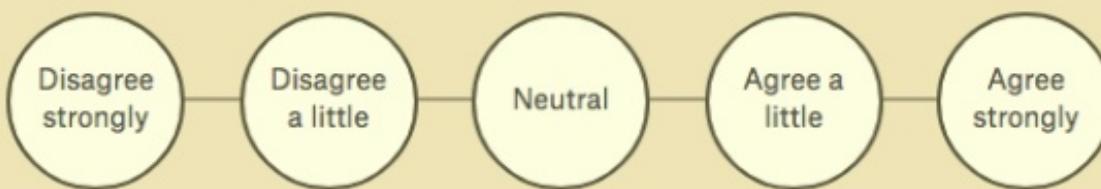


1 of 30 questions completed

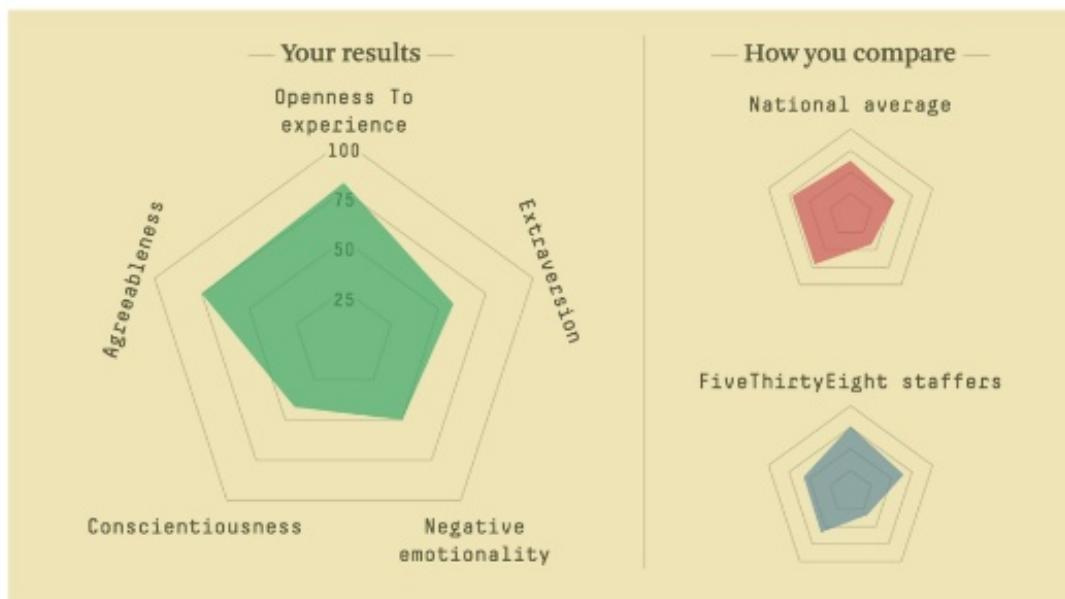
I tend to be quiet.



I am compassionate and have a soft heart.



Openness to experience 79 out of 100
Agreeableness 75 out of 100
Conscientiousness 42 out of 100
Negative emotionality 50 out of 100
Extraversion 58 out of 100





**Tools are important,
But Technique is Everything in
Data Science.**

Extraversion

100

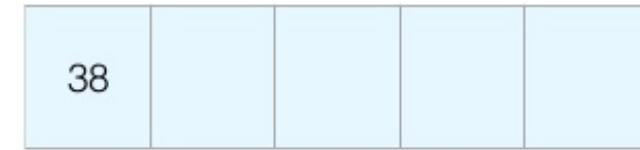


0

Introversion

Jack

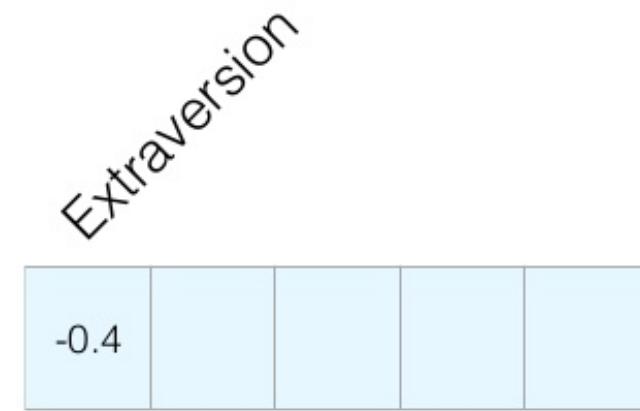
Extraversion



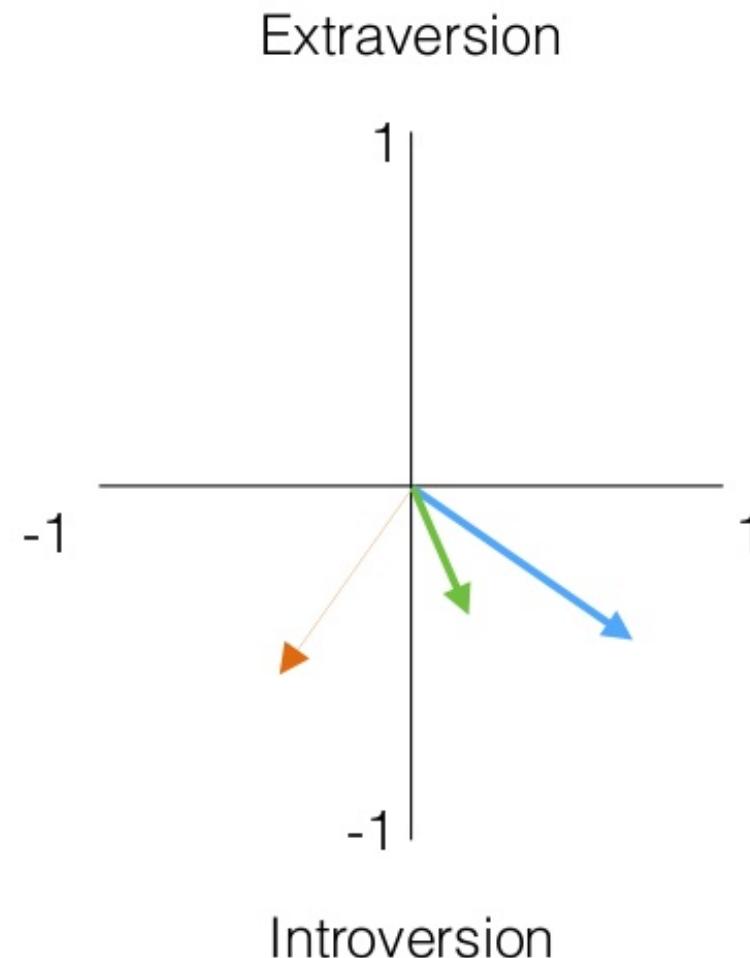
Extraversion



Jack



Introversion



Jack

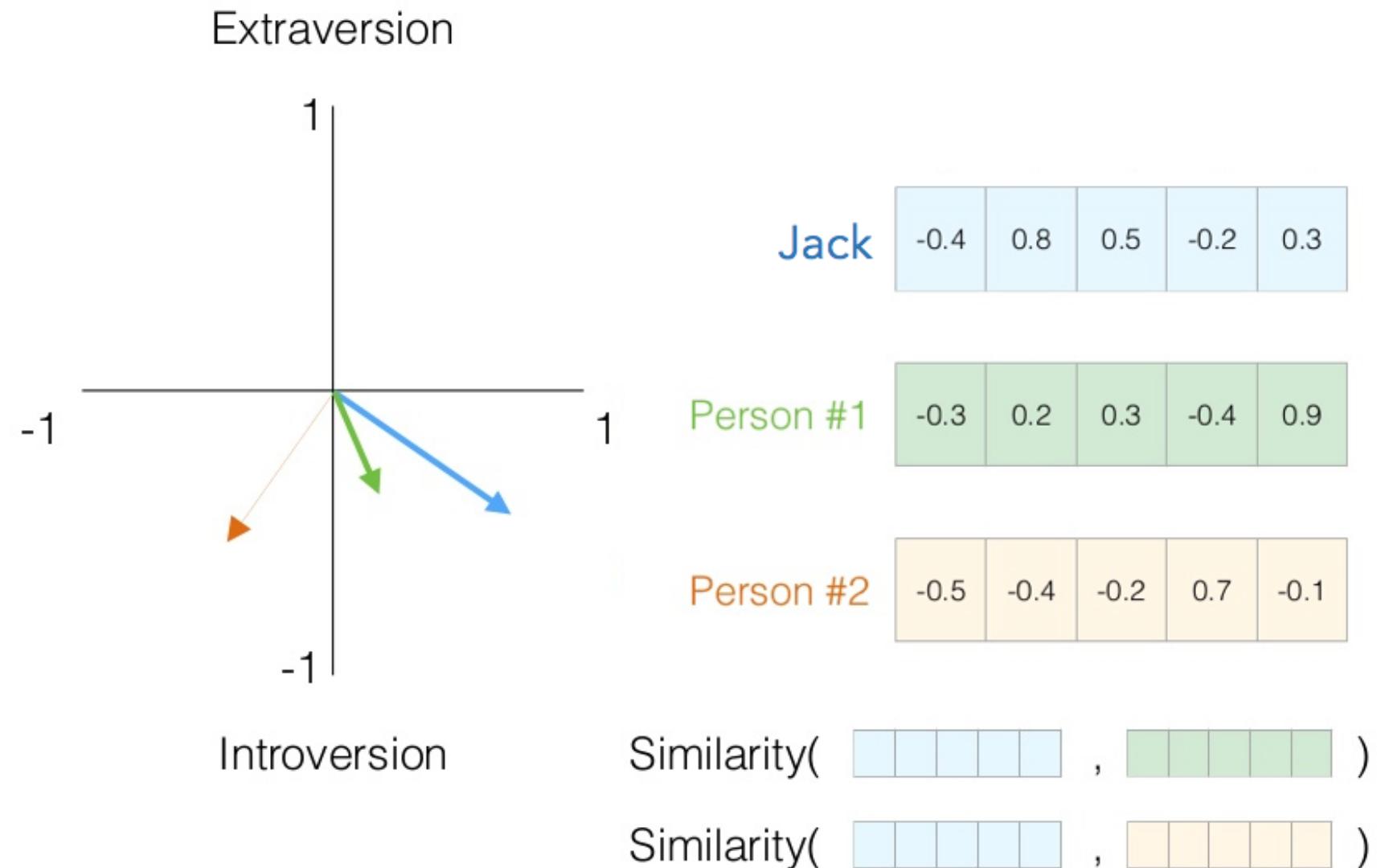
-0.4	0.8	0.5	-0.2	0.3
------	-----	-----	------	-----

Person #1

-0.3	0.2	0.3	-0.4	0.9
------	-----	-----	------	-----

Person #2

-0.5	-0.4	-0.2	0.7	-0.1
------	------	------	-----	------



$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

cosine_similarity(

-0.4	0.8
------	-----

 ,

-0.3	0.2
------	-----

) = 0.87



cosine_similarity(

-0.4	0.8
------	-----

 ,

-0.5	-0.4
------	------

) = -0.20

	Trait #1	Trait #2	Trait #3	Trait #4	Trait #5
Jack	-0.4	0.8	0.5	-0.2	0.3
Person #1	-0.3	0.2	0.3	-0.4	0.9

	Trait #1	Trait #2	Trait #3	Trait #4	Trait #5
Person #1	-0.3	0.2	0.3	-0.4	0.9
Person #2	-0.5	-0.4	-0.2	0.7	-0.1

	Trait #1	Trait #2	Trait #3	Trait #4	Trait #5
Jack	-0.4	0.8	0.5	-0.2	0.3
Person #1	-0.3	0.2	0.3	-0.4	0.9

Person #1	-0.3	0.2	0.3	-0.4	0.9
-----------	------	-----	-----	------	-----

Person #2	-0.5	-0.4	-0.2	0.7	-0.1
-----------	------	------	------	-----	------

`cosine_similarity([] , [])`

`cosine_similarity([] , [])`

1- We can represent things
(and people) as vectors of
numbers
(Which is great for machines!)

Jack

-0.4	0.8	0.5	-0.2	0.3
------	-----	-----	------	-----

- 2- We can easily calculate how similar vectors are to each other

Customers who liked Jack also liked:

	cosine_similarity ▼
Person #1	0.86
Person #2	0.5
Person #3	-0.20

Word Embedding

“king”

```
array([ 0.50451 ,  0.68607 , -0.59517 , -0.022801,  0.60046 , -0.13498 ,  
       -0.08813 ,  0.47377 , -0.61798 , -0.31012 , -0.076666,  1.493 ,  
       -0.034189, -0.98173 ,  0.68229 ,  0.81722 , -0.51874 , -0.31503 ,  
       -0.55809 ,  0.66421 ,  0.1961 , -0.13495 , -0.11476 , -0.30344 ,  
       0.41177 , -2.223 , -1.0756 , -1.0783 , -0.34354 ,  0.33505 ,  
       1.9927 , -0.04234 , -0.64319 ,  0.71125 ,  0.49159 ,  0.16754 ,  
       0.34344 , -0.25663 , -0.8523 ,  0.1661 ,  0.40102 ,  1.1685 ,  
       -1.0137 , -0.21585 , -0.15155 ,  0.78321 , -0.91241 , -1.6106 ,  
       -0.64426 , -0.51042 ], dtype=float32)
```

Embedding: GloVe

Dimensions: 50

Trained on: Wikipedia + Gigaword 5 (6B tokens)

Number of vectors: 400,000

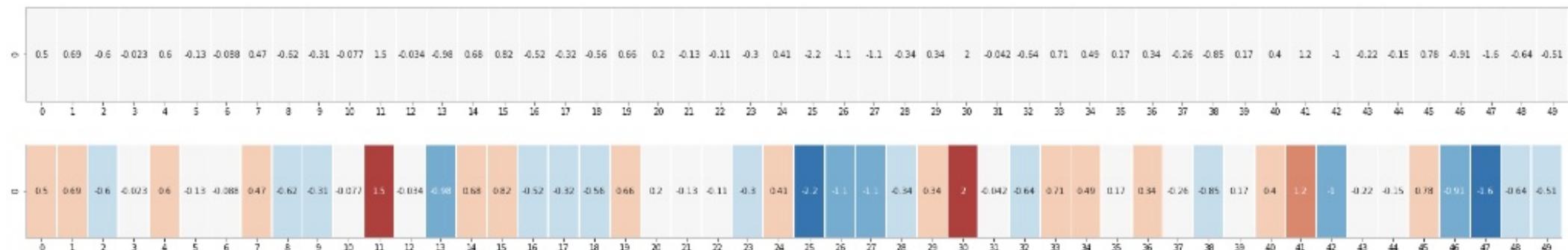
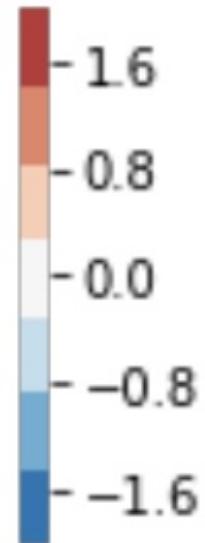
“king”

```
array([ 0.50451 ,  0.68607 , -0.59517 , -0.022801,  0.60046 , -0.13498 ,  
       -0.08813 ,  0.47377 , -0.61798 , -0.31012 , -0.076666,  1.493 ,  
       -0.034189, -0.98173 ,  0.68229 ,  0.81722 , -0.51874 , -0.31503 ,  
       -0.55809 ,  0.66421 ,  0.1961 , -0.13495 , -0.11476 , -0.30344 ,  
       0.41177 , -2.223 , -1.0756 , -1.0783 , -0.34354 ,  0.33505 ,  
       1.9927 , -0.04234 , -0.64319 ,  0.71125 ,  0.49159 ,  0.16754 ,  
       0.34344 , -0.25663 , -0.8523 ,  0.1661 ,  0.40102 ,  1.1685 ,  
       -1.0137 , -0.21585 , -0.15155 ,  0.78321 , -0.91241 , -1.6106 ,  
       -0.64426 , -0.51042 ], dtype=float32)
```



“king”

```
array([ 0.50451 ,  0.68607 , -0.59517 , -0.022801,  0.60046 , -0.13498 ,
       -0.08813 ,  0.47377 , -0.61798 , -0.31012 , -0.076666,  1.493 ,
      -0.034189, -0.98173 ,  0.68229 ,  0.81722 , -0.51874 , -0.31503 ,
      -0.55809 ,  0.66421 ,  0.1961 , -0.13495 , -0.11476 , -0.30344 ,
       0.41177 , -2.223 , -1.0756 , -1.0783 , -0.34354 ,  0.33505 ,
       1.9927 , -0.04234 , -0.64319 ,  0.71125 ,  0.49159 ,  0.16754 ,
       0.34344 , -0.25663 , -0.8523 ,  0.1661 ,  0.40102 ,  1.1685 ,
      -1.0137 , -0.21585 , -0.15155 ,  0.78321 , -0.91241 , -1.6106 ,
      -0.64426 , -0.51042 ], dtype=float32)
```



“king”



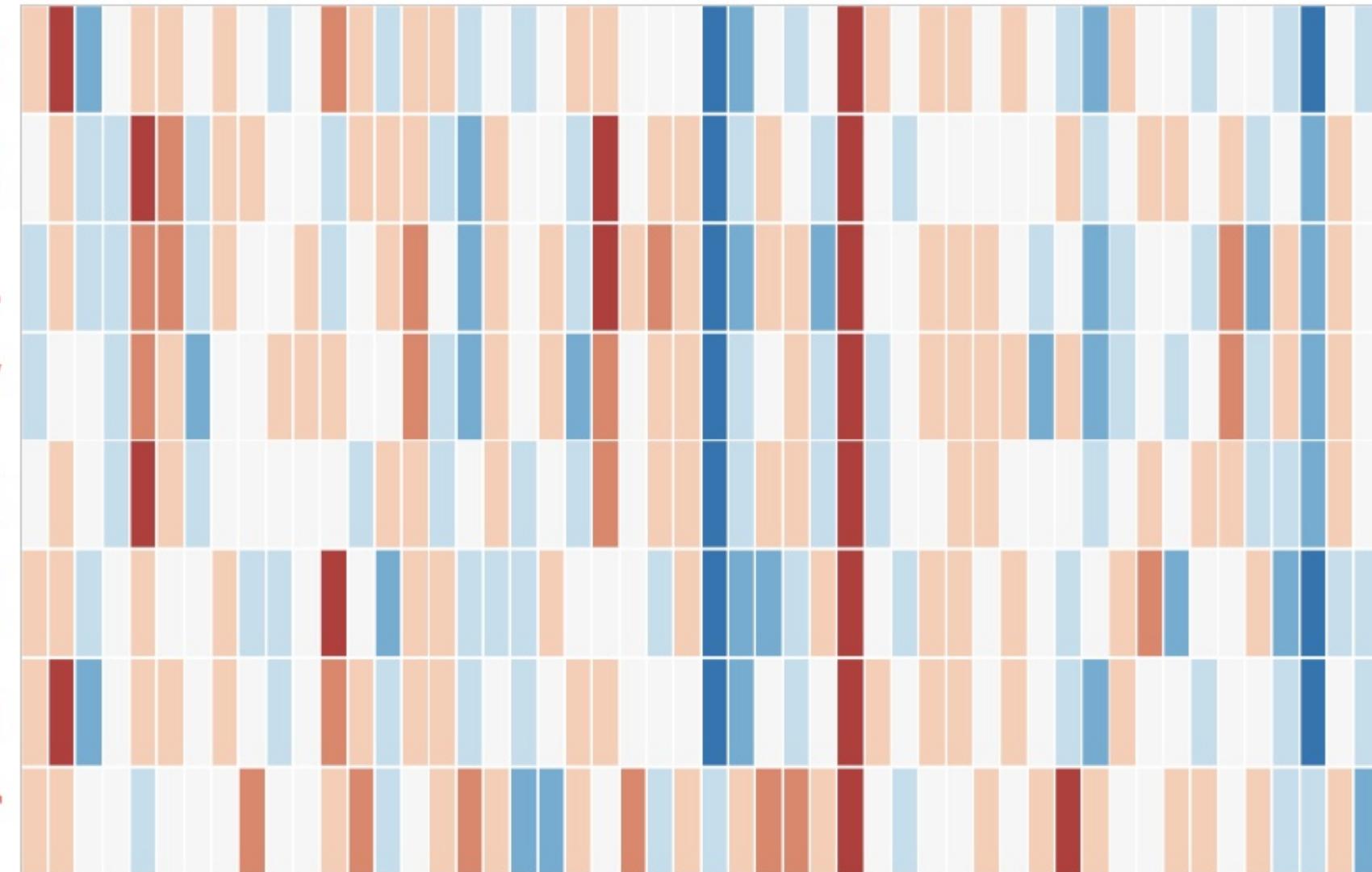
“Man”



“Woman”



queen
woman
girl
boy
man
king
queen
water



Analogies

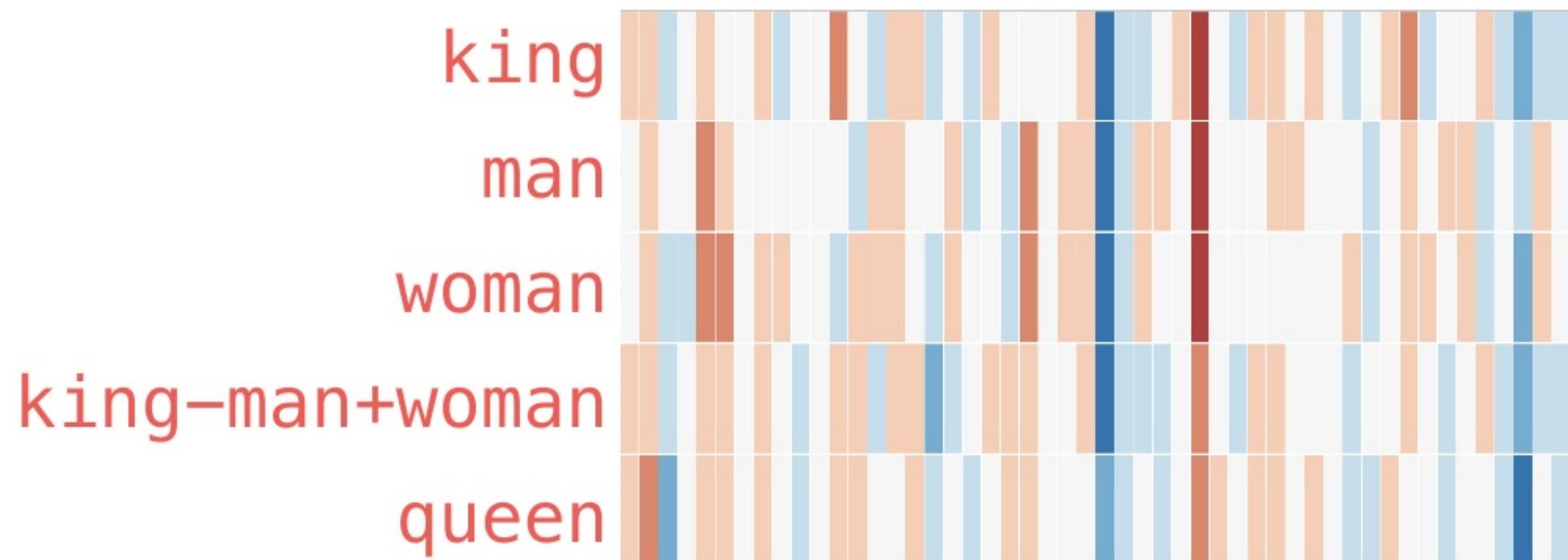
king - man + woman ~=

king - man + woman ~ queen

```
model.most_similar(positive=[ "king", "woman" ], negative=[ "man" ])
```

```
[('queen', 0.8523603677749634),  
 ('throne', 0.7664333581924438),  
 ('prince', 0.7592144012451172),  
 ('daughter', 0.7473883032798767),  
 ('elizabeth', 0.7460219860076904),  
 ('princess', 0.7424570322036743),  
 ('kingdom', 0.7337411642074585),  
 ('monarch', 0.721449077129364),  
 ('eldest', 0.7184862494468689),  
 ('widow', 0.7099430561065674)]
```

king - man + woman ~ queen

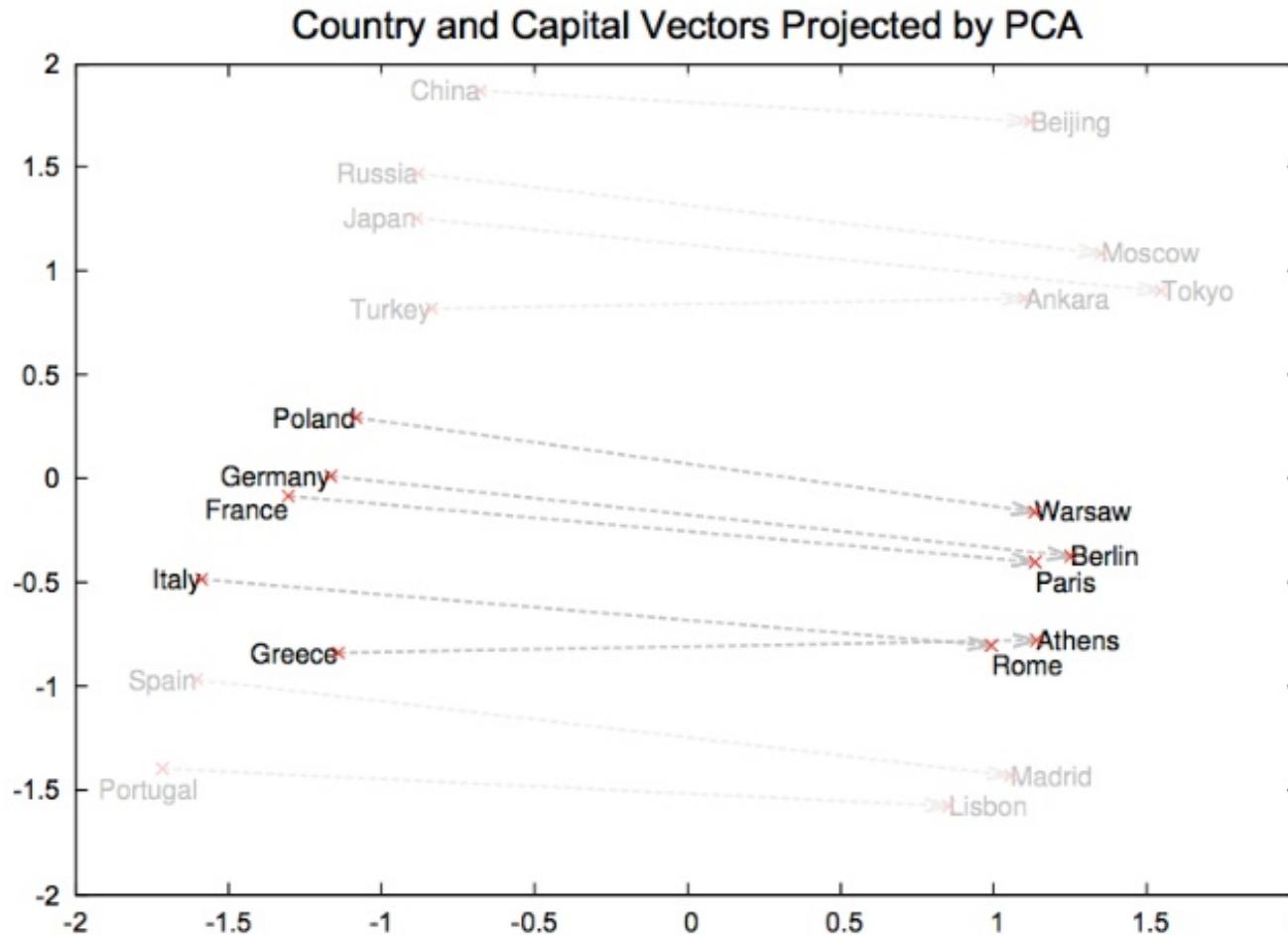


france is to paris as italy is to _____

```
model.most_similar(positive=["paris", "italy"], negative=["france"])

[('rome', 0.8465589284896851),
 ('milan', 0.77660071849823),
 ('turin', 0.7666354775428772),
 ('venice', 0.7591592073440552),
 ('madrid', 0.7565804719924927),
 ('italian', 0.7513630390167236),
 ('aires', 0.7429094910621643),
 ('naples', 0.7406353950500488),
 ('buenos', 0.7357348203659058),
 ('lisbon', 0.7244545221328735)]
```

france is to paris as italy is to



Distributed Representations of Words and Phrases and their Compositionality

Tomas Mikolov
Google Inc.
Mountain View
mikolov@google.com

Ilya Sutskever
Google Inc.
Mountain View
iya@google.com

Kai Chen
Google Inc.
Mountain View
kai@google.com

Greg Corrado
Google Inc.
Mountain View
gcorrado@google.com

Jeffrey Dean
Google Inc.
Mountain View
jeff@google.com

“Thou shalt not make a machine in the likeness of a human mind”
~Dune

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	...
------	-------	-----	------	---	---------	----	-----	-----

Dataset

input 1	input 2	output

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	...
------	-------	-----	------	---	---------	----	-----	-----

Dataset

input 1	input 2	output

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	...
------	-------	-----	------	---	---------	----	-----	-----

Dataset

input 1	input 2	output
thou	shalt	not

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	...
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	...
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not
shalt	not	make

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	...
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not
shalt	not	make
not	make	a

Thou shalt not make **a machine in** the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	...
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not
shalt	not	make
not	make	a
make	a	machine
a	machine	in

Enough of Theory ;)



homepage

**listing
#104**

Site Search

**listing
#73**

**listing
#3065**

**listing
#104**

**listing
#73**

**listing
#3065**

Click sessions



#104	#73	#3065
------	-----	-------



#100	#200	#300	#400	#500	#600	#700
------	------	------	------	------	------	------



#100	#200	#300	#400	#500	#600	#700	#800	#900	#1000	#1100	#1200
------	------	------	------	------	------	------	------	------	-------	-------	-------

Click sessions



#104	#73	#3065
------	-----	-------



#100	#200	#300	#400	#500	#600	#700
------	------	------	------	------	------	------



#100	#200	#300	#400	#500	#600	#700	#800	#900	#1000	#1100	#1200
------	------	------	------	------	------	------	------	------	-------	-------	-------

- Listing #1
- Listing #2
- Listing #3
- Listing #4
- Listing #5
- Listing #6
- Listing #7
- ...
- Listing #4000



listing #3

Similar listings:



ENTIRE CHALET · MONTE VERDE
Charming place in the Mountain
SR485 per night
★★★★★ 109



ENTIRE COTTAGE · SETOR SOCIOECONÔMICO 21
Refúgio nas nuvens
SR400 per night
★★★★★ 65



PRIVATE ROOM - SÃO JOSÉ DOS CAMPOS
Saint Barbara High SFX Chalet 1
SR745 per night
★★★★★ 53

Embedding Evaluation Tool

Search

Query Type

Listing ID

Listing ID

16486364

Search

I'm Feeling Lucky



\$236 Cabane Secrète pour 2 personnes

Location: Raray, Hauts-de-France, France

Description: Mon logement est proche de la nature et du golf. Vous appréciez mon logement pour sa tranquillité et son confort. Mon logement est parfait pour les

...

Embedding Evaluation Tool

Search

Query Type

Listing ID

Search

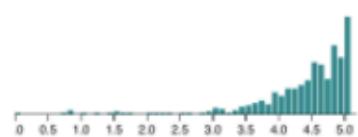
I'm Feeling Lucky



\$236 Cabane Secrète pour 2 personnes

Location: Raray, Hauts-de-France, France
Description: Mon logement est proche de la nature et du golf. Vous appréciez mon logement pour sa tranquillité et son confort. Mon logement est parfait pour les ...

Score Histogram



*Includes scores for up to the 500 nearest listings

Nearest listings (10)



\$236 Cabane Secrète pour 2 personnes

KNN: /admin/embedding_evaluation/16486364

Score: 0.00

Location: Raray, Hauts-de-France, France
Description: Mon logement est proche de la nature et du golf. Vous appréciez mon logement pour sa tranquillité et son confort. Mon logement est parfait pour les ...



\$386 Cabane Spa Origin

KNN: /admin/embedding_evaluation/16905264

Score: 0.70

Location: Raray, Hauts-de-France, France
Description: Mon logement est proche de la nature et de la forêt. Vous appréciez mon logement pour l'emplacement, les espaces extérieurs et sa tranquillité. Mon ...



\$320 Cabane SPA Cocon pour 2 personnes

KNN: /admin/embedding_evaluation/16486854

Score: 0.84

Location: Raray, Hauts-de-France, France
Description: Mon logement est proche de Paris . Vous appréciez mon logement pour son bain nordique privé. Mon logement est parfait pour les couples.



\$236 Cabane Imprenable pour 2 personnes

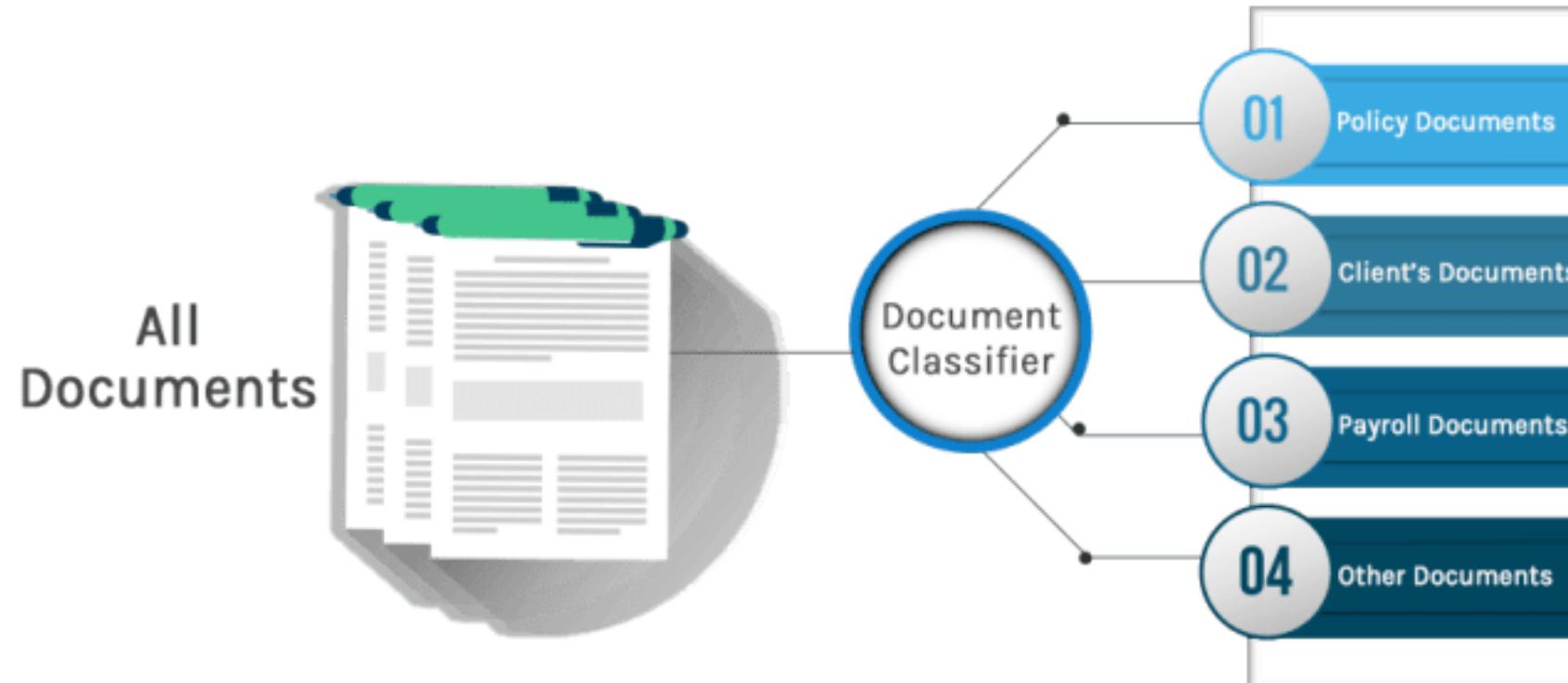
KNN: /admin/embedding_evaluation/16485735

Score: 0.87

Location: Raray, Hauts-de-France, France
Description: Mon logement est proche du golf et du château. Vous appréciez mon logement pour son calme et son confort. Mon logement est parfait pour les couples...



Document Classification



The Three Documents and Similarity Metrics

Sachin Tendulkar

Full Article: [View document](#)

"Tendulkar" indicates here the other people with the same surname, see [Tendulkar \(disambiguation\)](#).

Service like Google Scholar or the [Google Scholar](#) search engine can be used to find more information on the author's publications, reported in the following section.

MS Dhoni

Full Article: [View document](#)

MS Dhoni is an Indian international cricketer who has represented India in all formats of the game, including Test cricket, One Day International (ODI) and Twenty20 International (T20I). He is also a former Indian national team captain and has been described as one of the best fielders in the history of the sport. He is currently the coach of the Indian national team.

MS Dhoni_Small

Full Article: [View document](#)

MS Dhoni is an Indian international cricketer who has represented India in all formats of the game, including Test cricket, One Day International (ODI) and Twenty20 International (T20I). He is currently the coach of the Indian national team.

MS Dhoni

Full Article: [View document](#)

MS Dhoni is an Indian international cricketer who has represented India in all formats of the game, including Test cricket, One Day International (ODI) and Twenty20 International (T20I). He is currently the coach of the Indian national team.

MS Dhoni

Full Article: [View document](#)

MS Dhoni is an Indian international cricketer who has represented India in all formats of the game, including Test cricket, One Day International (ODI) and Twenty20 International (T20I). He is currently the coach of the Indian national team.

Considering only the 3 words from the above documents: 'sachin', 'dhoni', 'cricket'

Doc Sachin: Wiki page on Sachin Tendulkar

Dhoni	-	10
Cricket	-	50
Sachin	-	200

Doc Dhoni: Wiki page on Dhoni

Dhoni	-	400
Cricket	-	100
Sachin	-	20

Doc Dhoni_Small: Subsection of wiki on Dhoni

Dhoni	-	10
Cricket	-	5
Sachin	-	1

Document - Term Matrix (Word Counts)

Word Counts	"Dhoni"	"Cricket"	"Sachin"
<i>Doc Sachin</i>	10	50	200
<i>Doc Dhoni</i>	400	100	20
<i>Doc Dhoni_Small</i>	10	5	1



Similarity Metrics

Similarity or Distance Metrics	Total Common Words	Euclidean distance	Cosine Similarity
<i>Doc Sachin & Doc Dhoni</i>	$10 + 50 + 10 = 70$	432.4	0.15
<i>Doc Dhoni & Doc Dhoni_Small</i>	$20 + 10 + 7 = 37$	204.0	0.23
<i>Doc Sachin & Doc Dhoni_Small</i>	$10 + 10 + 7 = 27$	401.85	0.77

Data In Action



Deployment	pypl v0.0.1b3.dev0
Build Status	build passing
Metrics	coverage 28%
GitHub	contributors 8 Stars 6
License	License BSD
Community	chat on glitter

We are in beta

Cognito is an exclusive data preprocessing library that involves transforming raw data into a machine-understandable format. We at CleverInsight Open Ai Foundation took the initiative to build a better automated data preprocessing library and here it is.

Contributors



Collaborators



CHRIST (Deemed to be University) Lavasa provides developer & mentor contribution via CleverInsight Open AI Foundation Incubation Center at Christ University Lavasa since 2019.

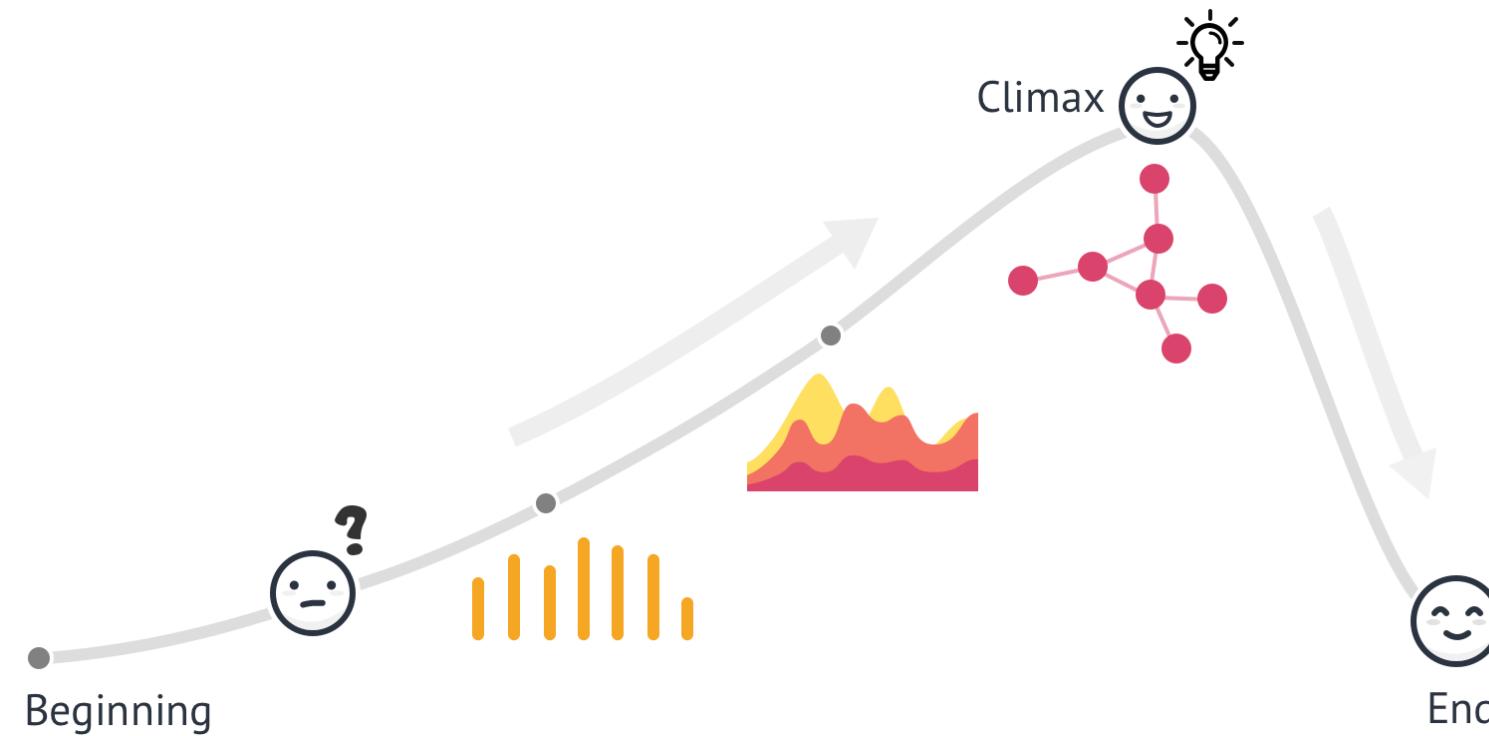
<https://pypi.org/project/cognito/>

Become a contributor

<https://github.com/Cleverinsight/cognito/issues>

Join Cognito Community

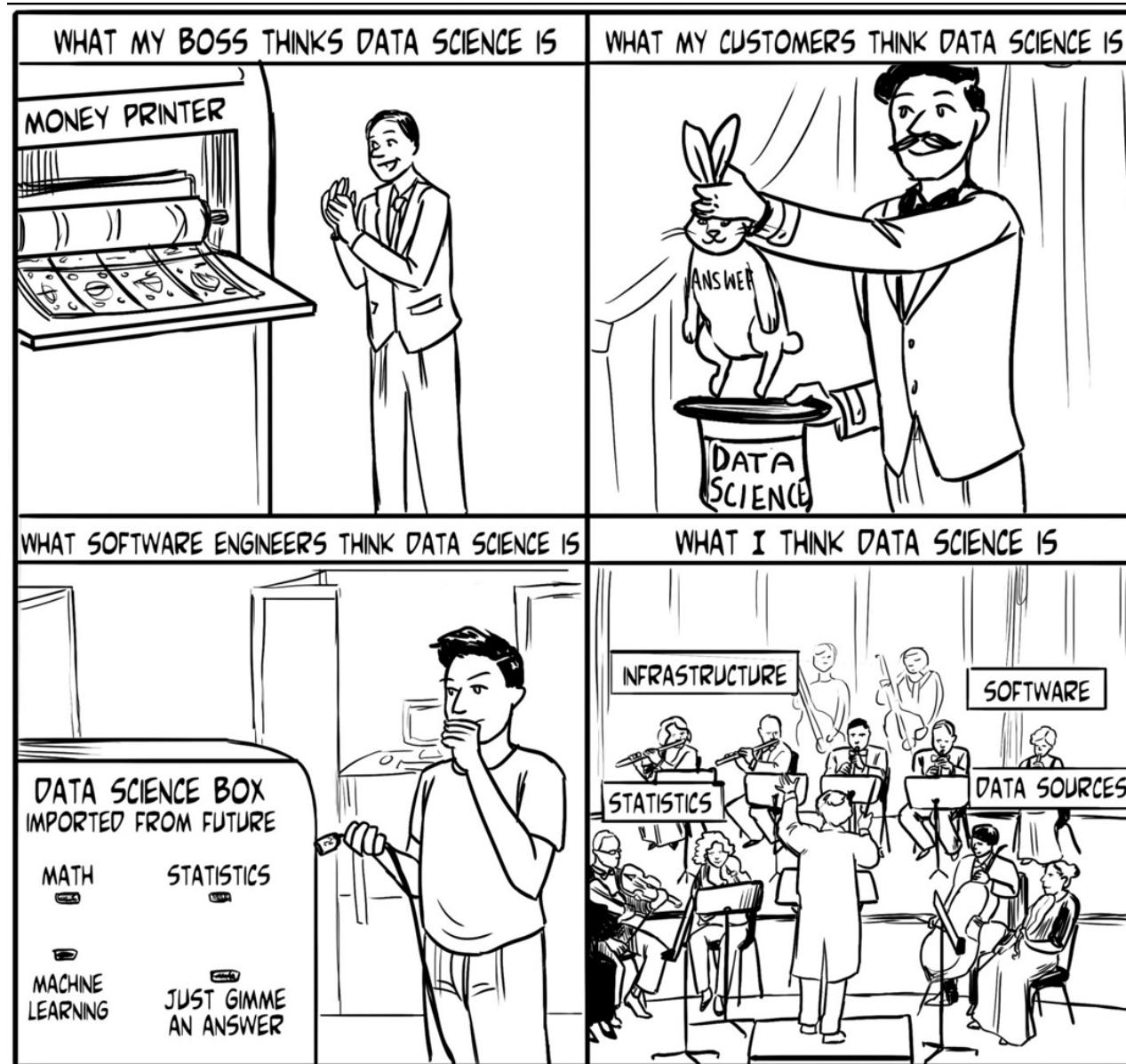
<https://gitter.im/cognito-dev/community>



**Data will talk to you
if you're willing to listen.**

In God We Trust; all others must bring DATA.

- William Edwards Deming -



Thank You

CleverInsight

CleverInsight
Open AI Foundation

149 Commonwealth Drive
Menlo Park, CA 94025
+1-408-888-4507



Bastin Robins .J

Chief Data Scientist
[@CleverInsight](https://twitter.com/CleverInsight)
robin@cleverinsight.co
<http://linkedin.com/in/bastinrobin>

Leave you Feedback
<https://bit.ly/30UDQ9A>



Time For Queries

Upcoming Bootcamps & Webinars

<https://cleverinsight.co/webinars/>