



# Classification analysis of American personalities

# Project overview

- Intro
- Data
- EDA
- Sentiment Analysis
- Modeling
- Predictions
- Summary
- Future projects

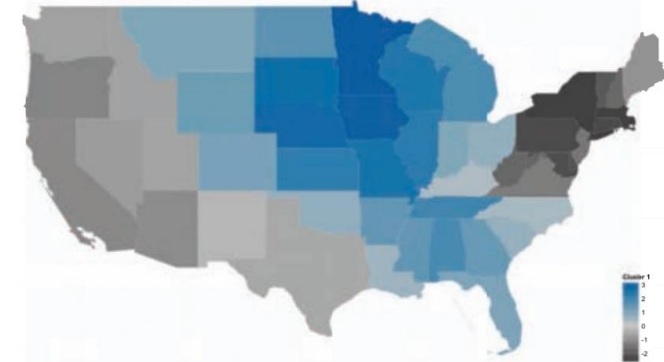


# Three Personalities of America study

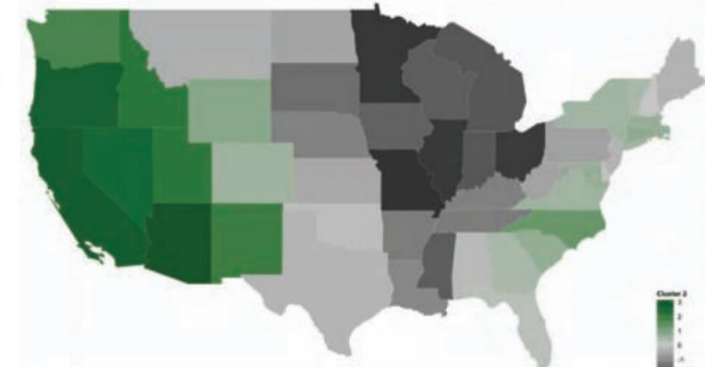
- Psychologist Jason Rentfrow
- University of Oxford - Do different places have different personalities?
  - 2003 to 2015
  - 2082 U.S. counties
  - 3 million people
  - 44 questions about habit and disposition
- The Atlantic - November 2019



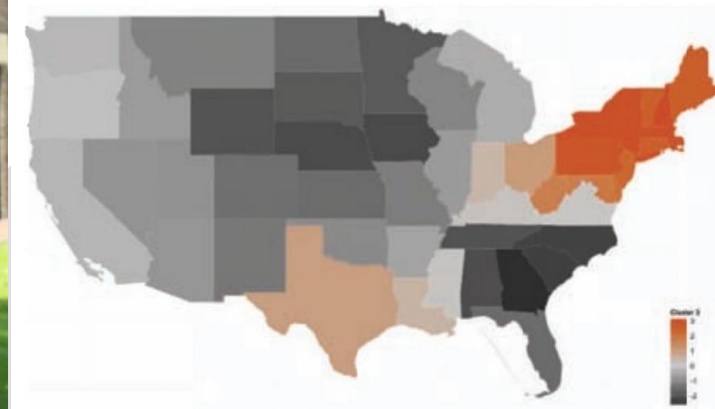
A. Cluster 1: Friendly & Conventional Region



B. Cluster 2: Relaxed & Creative Region



C. Cluster 3: Temperamental & Uninhibited Region





# r/AskNYC

- **123k members**
- Largest metro area in the world by urban area – 20 million.
- Home to the United Nations headquarters.
- One of the worlds largest natural harbors.
- Over 800 languages are spoken.

# r/AskLosAngeles

- **50.8k members**
- Second largest city after NYC.
- Second most populous after NYC.
- 13.1 million in the metro area.
- Third highest GDP in the world \$1.0 trillion.



# Sentiment analysis EDA - VADER

- Stop words removal
- Tokenizing, Lemming, Stemming
- Word count
- Unique users (6% had greater than 1 post)
- Bigram count
- Verbosity
- Most positive scoring post
- Most negative scoring posts
- Overall sentiment

**\*4000 - 5000 reddit posts**



# Verbosity...

	status_length	status_word_count
subreddit		
AskLosAngeles	363.476226	67.080330
AskNYC	313.590579	57.767226

# Top 15.....

r/AskLosAngeles  
50k members

Single words

la	1802
like	1331
know	1285
just	1164
looking	1118
ve	897
area	784
good	735
people	727
want	713
don	687
thanks	612
time	590
does	578
place	569

Bigram

los angeles	277
does know	196
santa monica	150
don know	138
thanks advance	125
amp x200b	121
don want	109
moving la	107
https www	80
feel like	73
let know	73
west hollywood	70
west la	66
la area	65
just moved	63

r/AskNYC  
123k members

Single words

just	1040
like	963
know	901
removed	818
nyc	790
ve	760
looking	685
don	657
time	579
new	548
people	514
city	511
good	476
want	475
apartment	474
thanks	459

Bigram

new york	184
does know	132
don know	117
don want	101
amp x200b	96
https www	93
thanks advance	82
central park	69
just moved	59
feel like	54
let know	54
years ago	51
just want	46
ve seen	41
imgur com	39

-Lot of shout outs in LA...?  
-Say thanks in LA more...?  
-Less looking in NYC....?  
-More moving to LA...?  
-Santa Monica is active...?  
-They like the word thanks less NYC in...



# Removed posts conundrum?

NYC = **818** of 4000 removed

```
21 [removed]
22 Hi! I'm a Latina girl in her 20's and I'm goin...
23 [removed]
24 [removed]
25 I've been living in a 20th floor midtown apart...
26 Been living here 20 years. I don't care so muc...
27 [removed]
28 If so, why? How was the food/experience? Would...
29 [removed]
30 I just signed a lease starting in mid-October,...
31
32
33 [removed]
34 Hi,\n\nLong story short: I have an eviction fr...
35 I'm 8 months deep in a lease, with only a few ...
36 [removed]
37 Hi any tips on moving to NYC from LA? Like i w...
38 [removed]
39 So after driving in from Jersey for a month I ...
40 [removed]
41 [removed]
42 Haven't seen or heard anything about new pot s...
43 [removed]
44 [removed]
45 [removed]
46 Title says it all. My boyfriend's family comes...
47 [removed]
```

-r/NYC didn't have any rules listed?  
-Each subreddit had the same amount of moderators

LA = **192** of 4000 removed

```
52 [removed]
53 Like what do you like or dislike about what th...
54 Hi so I'll be in LA for 3 nights and was wonde...
55 Hello, I'm thinking of trying to learn how to ...
56 [removed]
57 I'm due for a professional portrait and would ...
58 Has anyone here been in the past? It's this we...
59 How did you find your housing? \n\nHow much ar...
60 There seems to be many different avenues to ge...
61 Got my sneakers filthy, gotta clean them rip. ...
62 Ok, I know this is a broad question but what a...
63 I want to take my friend out for her birthday....
64 Is there a public place I could bring my guita...
65 [removed]
66 I been wanting to go for some time but during ...
67 Moving out of a one bedroom 600sqft apt and wo...
68 About to go on a 2 week trip to LA for the fir...
69 I'm falling on to my plan B because it seems l...
70 Hi everyone, I am a 24 y old male living in we...
71 Or has anyone already done it? I got an email ...
72 This past summer, I (22F) had a virtual intern...
73 Anyone in Los Angeles??👉\nPls message me ! Or ...
74 The number of ads looking for only female room...
75 [removed]
76 I'm looking for a monthly parking for rent in ...
77 Hey everyone! My girlfriend and I just moved t...
78 ?
79 What's the best flower delivery? I saw people ...
80 Any tips on living cheaply in Los Angeles? \n\...
```



- Still shout outs in LA...?
- Equal thanks...?
- Still less looking in NYC....?
- Still more moving to LA...?

# Top 15....(w/o [removed])

r/AskLosAngeles  
50k members

Single words

la	1802
like	1331
know	1285
just	1164
looking	1118
ve	897
area	784
good	735
people	727
want	713
don	687
thanks	612
time	590
does	578
place	569

Bigram

los angeles	277
does know	196
santa monica	150
don know	138
thanks advance	125
amp x200b	121
don want	109
moving la	107
https www	80
feel like	73
let know	73
west hollywood	70
west la	66
la area	65
just moved	63

r/AskNYC  
123k members

Single words

just	1372
like	1269
know	1145
nyc	1058
ve	996
looking	887
removed	871
don	845
time	742
new	713
city	684
people	659
want	622
apartment	614
thanks	608
good	583

Bigram

new york	233
does know	154
don know	144
don want	135
amp x200b	125
https www	116
thanks advance	109
central park	88
just moved	81
feel like	72
let know	66
ve seen	62
moving nyc	57
just want	57
years ago	55

# Highest count of popular words

Old Dataset  
-800 of 4000  
[removed]

New Dataset  
-871 of 5000  
[removed]

	NYC	LA
<b>Happy</b>	32	38
<b>Sad</b>	22	94
<b>Love</b>	189	306
<b>Hate</b>	55	75

\*Old dataset

	NYC w/o [removed]	LA
<b>Happy</b>	39	38
<b>Sad</b>	30	94
<b>Love</b>	257	306
<b>Hate</b>	70	75

\*New dataset

# Table of accuracies – Classification models

Baseline accuracy:  
50%

Accuracies w/ stop_words		
<b>XGBoost</b>	79.0	76.7
<b>LogReg</b>	77.2	
<b>DecTree</b>	76.6	73.3
<b>AdaBoost</b>	75.0	74
<b>ExtraTree</b>	73.6	
<b>SVC</b>	72.8	
<b>RandFor</b>	72.4	
<b>MultiNB</b>	68.0	
<b>KNN</b>	63.0	



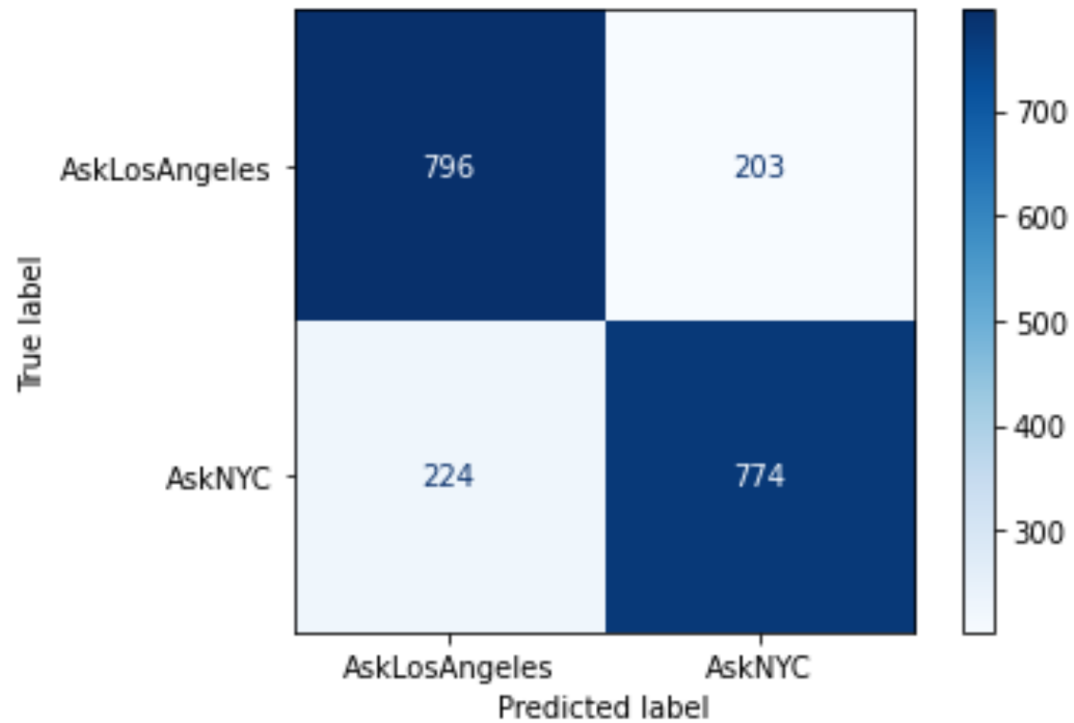
# XG Boost- parameters

Best parameters...gamma=0.4,  
max\_depth=10,  
n\_estimators=100,  
learning\_rate=0.2,  
reg\_alpha=1,  
n\_jobs=4,  
cv=5

\*gridsearch optimization

The logo for dmlc XGBoost. It features the text "dmlc" in a grey, lowercase, sans-serif font, positioned above the word "XGBoost" in a large, bold, blue, italicized sans-serif font. The background of the slide has a light purple gradient with white abstract shapes on the right side, including circles and organic, flowing lines.

# Confusion matrix



XGBoost	
Precision	79.2
Recall	78.0
Specificity	79.7
F1	78.6

# 😊 pipe.predict some stuff

Prediction	
NYPD	AskNYC
FDNY	AskNYC
Subway	AskNYC
train	AskNYC
car	AskNYC
LAPD	AskNYC
LA	AskLA
NYC	AskNYC
Los Angeles	AskLA
New York	AskNYC

Prediction	
Get away from me	AskNYC
Get out of my way	AskNYC
How are you	AskLA
Great job	AskNYC
I love you	AskLA
I hate you	AskLA

Prediction	
Happy	AskNYC
Nice	AskNYC
Smile	AskNYC
Fun	AskNYC
Sad	AskNYC
Unhappy	AskNYC
Mad	AskNYC
Angry	AskNYC





Who received the most Positive/Negative  
sentiment????

-They are pretty much same!

-LA is more positive by an extremely slim margin

- NYC is more negative by an extremely slim margin

	NYC	LA
<b>pos</b>	0.8526	0.85296
<b>neg</b>	0.1474	0.14700

# Most negative/most positive

**Most negative LA-** ThrowRABadBoi (-.09942)

'Is there anything that can be done about all the ongoing thefts? I've had multiple friends who had their converters stolen.\n\nI've heard that even the shields don't help much. Plus they're crazy expensive usually.\n\nIs there anything to be done, or are we just supposed to sort of hope for the best every night?'

**Most positive LA-** tyrelltsura (0.9995)

"Wanting to stop by there but want to be mindful of traffic and waits, because I'd presume everyone and their Mortys is going. Is there a huge line? For the uninformed, I'm talking about the Panorama City Wendy's that is now a pop-up restaurant."



# Most negative/most positive - NYC

## **Most positive NYC-** Rave-light (0.9995)

"Even locals and natives deserve a break in the BIG APPLE.\nGrab ya birks and socks, buy a new fanny pack, cuz we're going on a staycation babbyyyy.\n\nDealing with this topic a bit different than the rest:\n\nI'll be leaving some of the experience of our members. Feel free to add your own.....

## **Most negative NYC-** Arro (-0.9947)

'Suppose this is one of those "off my chest" type of posts.\n\nMoved into a place in the UWS last week. It's me a my fiancé, 34/M me and 33/F her.....

# Summary

- More expressive in LA?
- Slightly more verbose in LA?
- Looking for apartments in NYC?
- Seem to love mentioning themselves in LA (shout outs)?

# Future studies

- Study a single situation in history and see their reactions
- Add Midwest data
- More data
- Test more XGB params
- Grid search params for other models
- More Confusion Matrix comparisons
- Removed?

\*Admins, rules, or New Yorkers....