



Hello!

I am Alex Egorenkov

- Can we predict who
will retweet
#personalvictory



RECAP

● GA COMMUNITY

○ I ran into...

- Jeff
- www.personalvictories.com
- Displays #personalvictory tweets



2

THE APPROACH

THE APPROACH



user1	Random	tweet keyword1	tweet keyword1
user2	Random	tweet keyword1	tweet keyword2
⋮			
user2000	Retweeter	tweet keyword1	tweet keyword1
user2001	Retweeter	tweet keyword1	tweet keyword2

● THE APPROACH

○ It's important to ...

- Prevent leakage
 - Personal victory keyword
 - Personal victory tweets on user timelines



3

DATA COLLECTION

● TWITTER API

○ I used ...

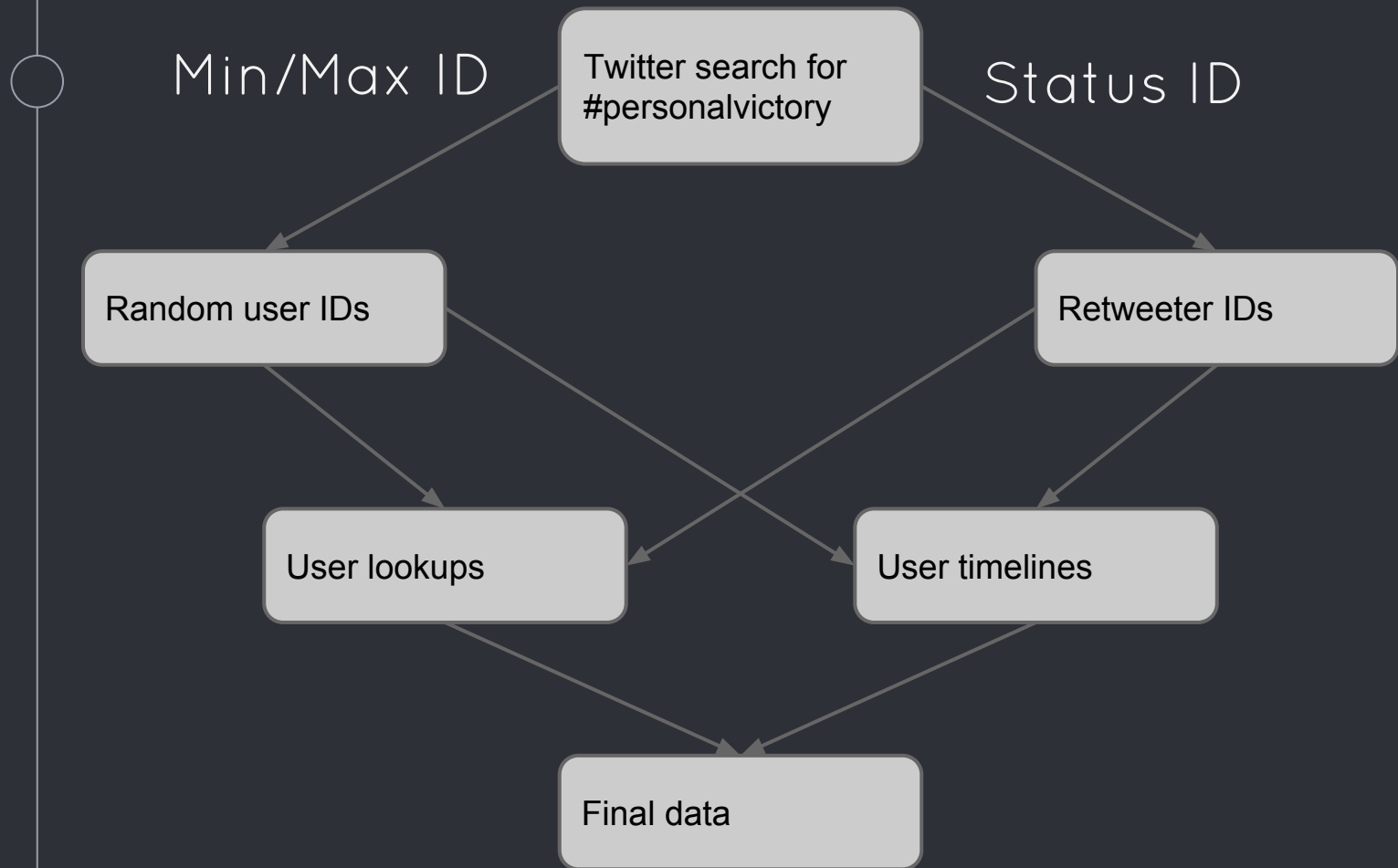
- GET application/rate_limit_status
- GET statuses/retweets/:id
- GET statuses/user_timeline
- GET users/lookup

EACH URL COMES WITH A RATE LIMIT

Title	Requests / 15-min window
GET application/rate_limit_status	180
GET statuses/retweets/:id	15
GET statuses/user_timeline	180
GET users/lookup	180

- The rate limit is specific to each resource
- The fastest way to collect data is to access multiple urls at a time

● FLOW OF DATA





4

EXPLORATION

● COLLECTED DATA

○ I collected ...

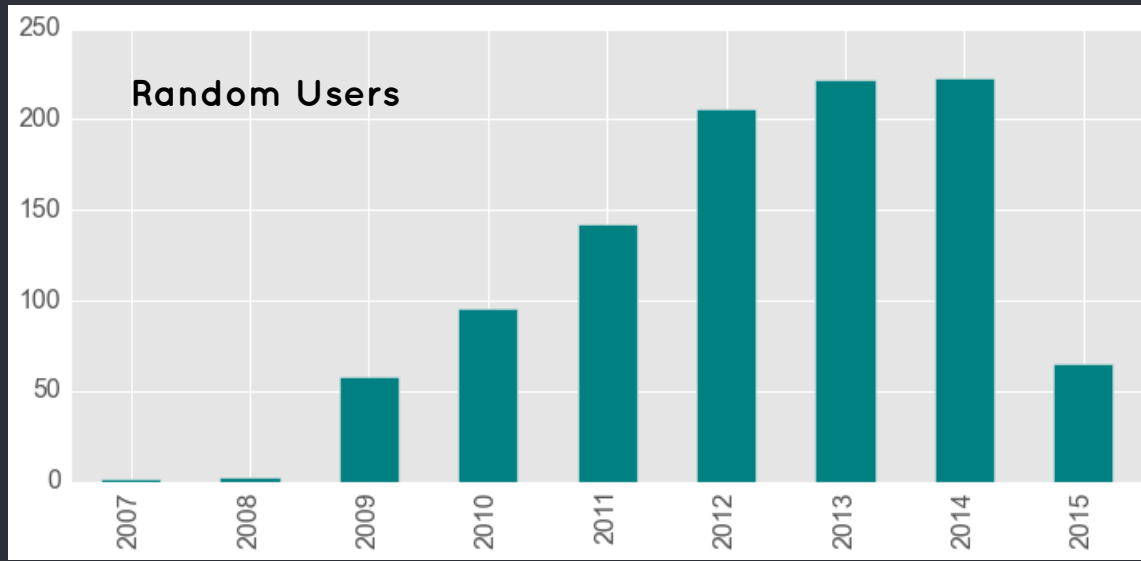
- 3,672 personal victory tweets
- 285 retweeters
- 1,014 random users
- Total of 1,299 users (22% prevalence)
- 126,211 tweets on the timelines of these users

USER CHARACTERISTICS

User Type	Favourites Count	Followers Count	Friends Count
Random	241.706070	182.233909	97.958595
Retweeter	4600.053381	2056.151408	1421.263158

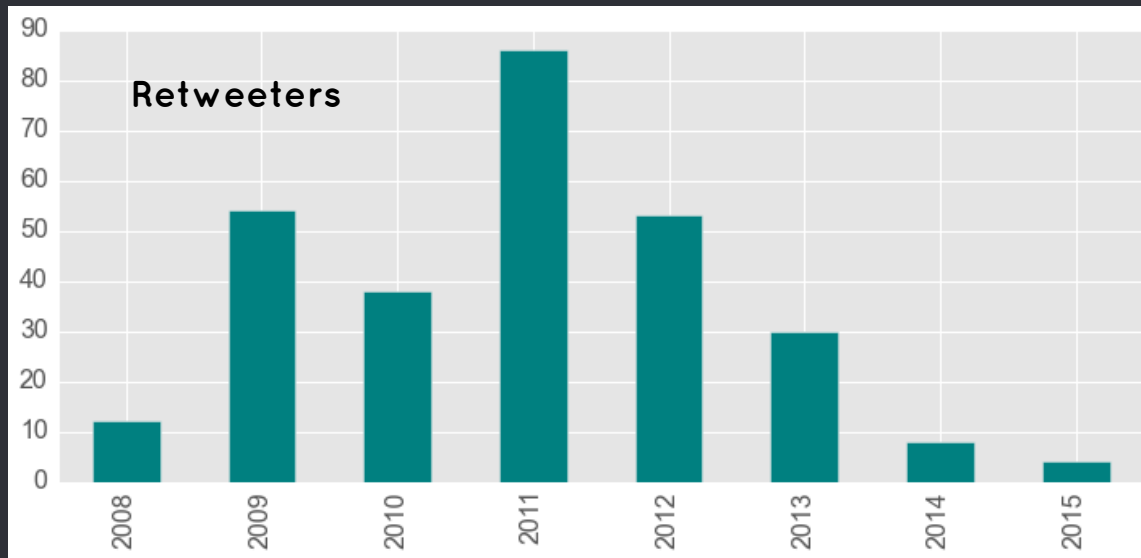
ACCOUNT CREATION YEAR

COUNTS



YEAR

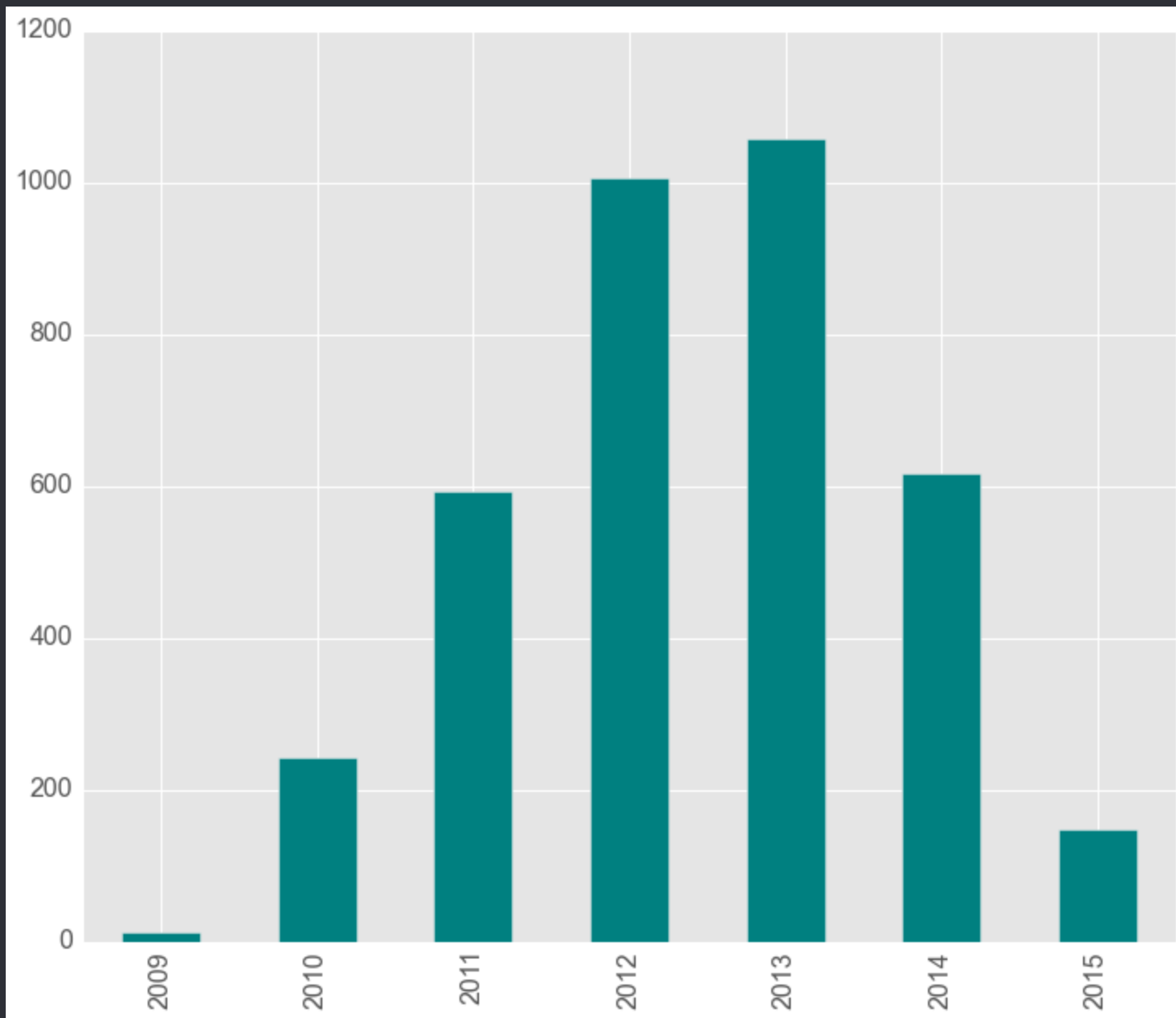
COUNTS



YEAR

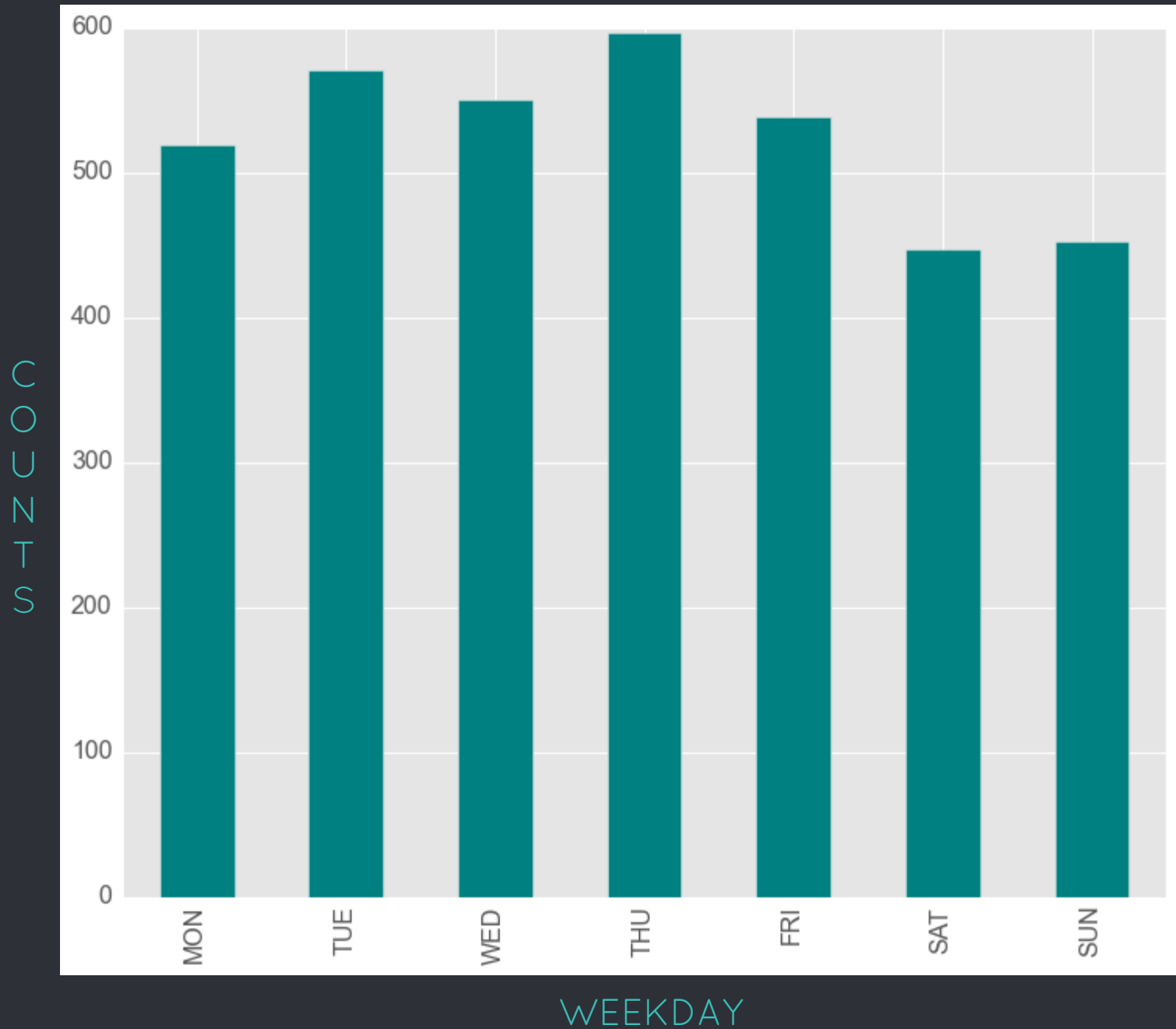
COUNT OF PERSONAL VICTORY TWEETS BY YEAR

COUNTS



YEAR

COUNT OF PERSONAL VICTORY TWEETS BY WEEKDAY





50 MOST COMMON TOKENS IN PERSONALVICTORY TWEETS

personalvictory

just

today

got

time

com

finally

didn

day

ve

http

twitter

did

like

pic

good

tonight

work

went

actually

rt

know

night

right

lol

week

class

yes

morning

proud

managed

getting

feel

lost

going

people

don

finished

home

way

instagram

won

think

year

run

minutes

took

win

love

game



5

RESULTS

OVERVIEW

- We can definitely predict which users will retweet a #personalvictory tweet
- Used a simple Naive Bayes with a document term matrix that looks at keyword presence
- Null accuracy 52% (balanced test set)
- Accuracy: 80%
- AUC: .94

● ELABORATION

- Modeling is pretty rudimentary
 - Result is based on keywords only
- Model works best when using relatively few features <150
 - Eliminated all victory tweets
 - Eliminated all instance of personal and victory keywords
 - Eliminated all retweets
 - Eliminated users who have few friends
 - Similar results

50 MOST DISCRIMINATING TOKENS IN AMONG RETWEETERS

http

rt

https

just

like

amp

love

don

day

time

new

today

good

people

know

http rt

life

want

ve

think

make

best

night

got

really

happy

need

going

great

ll

lol

right

tonight

thank

thanks

come

work

game

way

let

man

video

follow

gt

world

la

live

en

que

el

30 MOST DISCRIMINATING TOKENS AMONG RETWEETERS USING A LARGE QUANTITY OF TOKEN (AUC: .88)

Imao	basically
drunk	fitness
makeup	af
entire	relationship
point	energy
missing	opportunity
literally	fault
pictures	interesting
thoughts	mind
sports	extra
thankful	eat
fell	pretty sure
automatically	gun
tbt	glad
conversation	
drive	



6

APPLICATIONS

● APPLICATIONS

- Customer acquisition
 - Run scraper + predictive model to identify users who are likely to spend time reading through personal victory messages
- Customer segmentation
 - Difficult since there were only 285 retweeters

Thanks!

SUGGESTION ARE WELCOME!

● CREDITS

○ Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by [SlidesCarnival](#)
- Photographs by [Unsplash](#)