

# WHAT DOES TWITTER THINK?

Contrasting perceptions of Indian immigrants  
in Doha and San Francisco

Vidyut Baradwaj *for Global Society III (Spring 2022)*



# Overview



## Motivation

Why?



## Review

What?



## Settings and Context

Who, When and Where?



## Methodology and Results

How?



## Limitations and Implications

What next?





Source: Gulf News (2021)



Source: Middle East Institute (2020)



# Motivation



## *Why this topic?*

- Migration to the Gulf is one of the most unique demographic phenomena in recent times
- Large scope for research with numerous literature gaps
- Little to no research on the perceptions of Indian-Americans

# **Research Question**

**How does the dominant education level of  
Indian immigrants in San Francisco and Doha  
affect the that city's Twitter users'  
perceptions of them?**



# Goals



## **Examine existing narratives on migration flows through past literature**

Develop a contextual understanding of the selectivity of immigrants in the two locations and its consequences



## **Gauge perceptions by collecting and analyzing Tweets**

Enhance our understanding of these consequences while also developing and refining novel research methods



## **Posit theoretical models to explain differences**

Apply existing research on immigrant experiences in tandem with local cultures to expound these perceptions



# Settings and Context



## *Why Doha and San Francisco?*

- Both regional hubs for certain industries with similar cultural and financial influences
- Qatar publicly releases demographic data vis-à-vis migrant labour
- San Francisco represents one of the largest concentration of highly-skilled Indian immigrants



## *Why use Twitter?*

- Growing relevance of data-science in social science research
- Provides a comprehensive platform to gauge attitudes and behavioral trends
- Data less likely to be prone to social desirability bias<sup>^</sup>
- Fitting that Twitter's new CEO is Indian

# Review

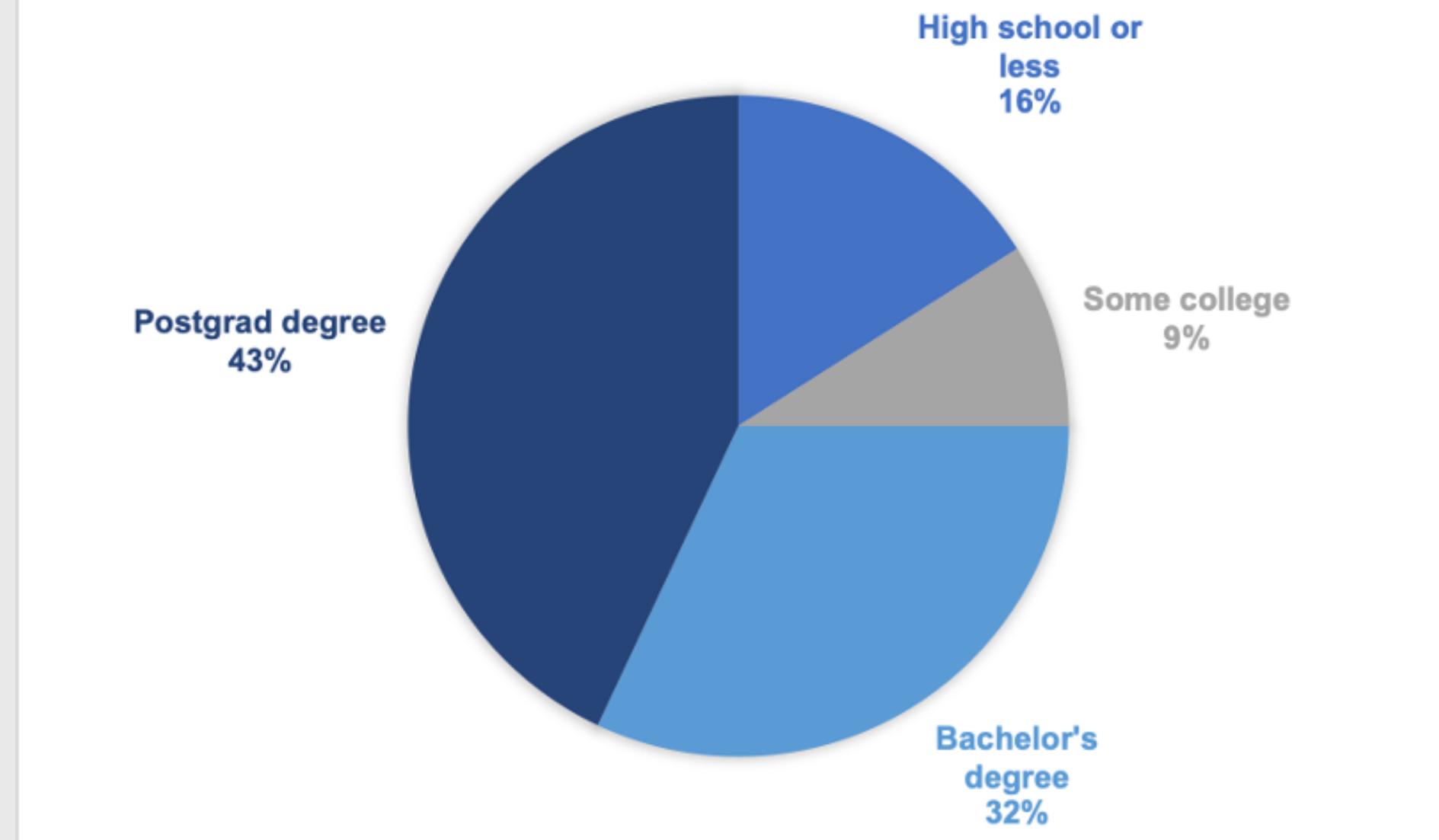
## Immigration Profile (Pew Research Center)

- Indians are the highest earning nationality group in the United States (Pew Research Center)
- Indians represent 25% of Qatar's population (Qatar Statistical

## *Determinants of Outmigration*

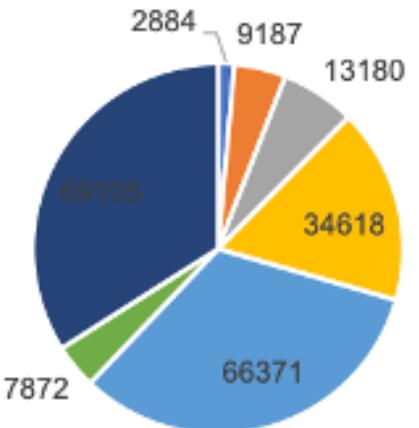
- A combination of push, pull factors and intervening obstacles (Lee's Theory of Migration)
- Region and context specific, requires rigorous empirical analysis (Douglas Massey's Theories of International Migration)
- Selectivity of Migration based on education-level:  
education --> social capital --> availability of opportunities
- Less-educated, 'mass migrants' more likely to be working-age men than families (Browning and Feidt, 1969 )

## % OF FOREIGN-BORN INDIAN IMMIGRANTS



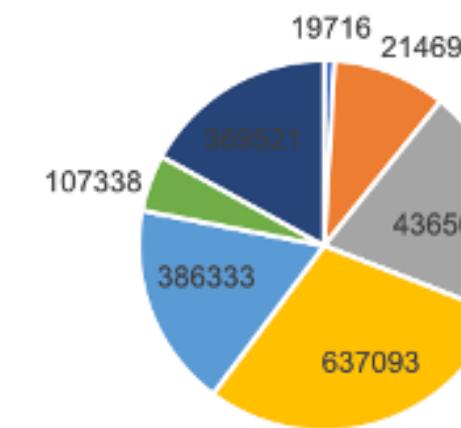
Source: American  
Community Survey

Population by education level: Qataris



- Illiterate
- Read & Write
- Primary
- Preparatory
- Secondary
- Pre U Diploma
- University +

Population by education level: Non-Qataris



- Illiterate
- Read & Write
- Primary
- Preparatory
- Secondary
- Pre U Diploma
- University +

Source: Qatar Statistical  
Authority

# Step 1: Scrape Tweets

Used Python package **Twint** to scrape data

- No restrictions unlike Twitter API
- Easy to implement with Pandas

Used Python package **Pandas** to store data

- Widely-used data analysis library
- Extremely powerful and versatile



Source: Just Understanding Data

01

**Set  
constraints**

Designed a set of keywords like 'indians', 'desi' based on review

02

**Collect  
tweets**

Built a function that takes in constraints like keywords and location and outputs dataframe

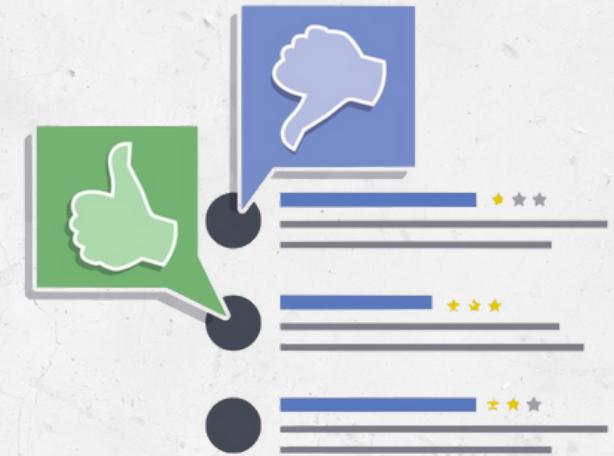
03

**Create  
dataframe**

Use Pandas to output the collected tweets into a Pandas DataFrame for ease of manipulation

# Step 2: NLP-based Analysis

Use simple algorithms and python libraries **NLTK**, **VaderSentiment** and **mathplotlib** to 'tokenize' the tweets and synthesize results on the basis of frequency and sentiment



Source: Data Science Central

## 1 Split tweets

Each tweet's text is split into a set of individual tokens (words or emojis)

## 2 Clean tokens

All emojis, links, images, punctuation and stop words are removed

## A1 Rank k-grams

$k$  no. of tokens are grouped and sorted based on tf-idf scores

## A2 Generate wordcloud

Use frequency data and wordcloud library to generate wordcloud where size  $\alpha$  frequency

## B1 Sentiment Analysis

Gauge general sentiment of tweet corpus

## Tweets

'Not just the individual stories, **Undocumented** gives us a historical and political sense of this society.' In @News9Tweets, @sinhadipanjan writes about Undocumented: Stories of Indian Migrants in the Arab Gulf by @rejitweets. <https://t.co/AFsdW6pyxl>

Its not about casteism being "allowed" or not. Subcontinental Islam has successfully adpoted and co-opted this instrument of stratification. And don't tell me heirarchy doesn't exist in Islam. Just hear the testimony of Indian Muslims who worked as migrant workers in the Gulf.

One angle of Indian #Diaspora deals with migrant laborers in #gulf nations, providing lion share of remittances are most often kept at the **bottom of the pyramid**. This matter is ignored sometimes need to be taken care of to strengthen the cooperation. @MEAIndia @NITIAayog <https://t.co/As5WhfE3Zc>

@ChinmayTumbe You must get your students and ,Äòfans.Äô (I am one of them üò) to read @rejimon,Äôs Undocumented - Stories of Indian migrants in the Arab Gulf <https://t.co/JjuYWmKdIh>

'It takes seconds to become undocumented.' In @TheLeaflet\_in, an interview with @rejitweets about his book Undocumented: Stories of Indian Migrants in the Arab Gulf. <https://t.co/4lnbuA1itQ>

Undocumented: Stories of Indian Migrants in the Arab Gulf by Rejimon Kuttappan is a critical contribution to fill this huge vacuum. @MigrantRights

## Tweets

@thinkthanked @adipalve There was a time when Uber was in the tank and I got cornered at a bar by the BBC reporter for a scoop looool. My reply, **just another Indian immigrant in the American taxi industry**

Carlson makes his audience fear that they are losing elections due to non-white immigrants; they're losing white college educated people. There's nothing inherently Democratic about immigrants. The Indian community used to be split in Silicon Valley; Bush almost split Latinos.

**Indian diaspora** is amongst the most loved and respected immigrant groups in the US. Trusted w handling some of the most sensitive aspects of American life. One of these friends is the CTO of CIA. Guess which one? An interesting weekend. <https://t.co/wyYI2RohbJ>

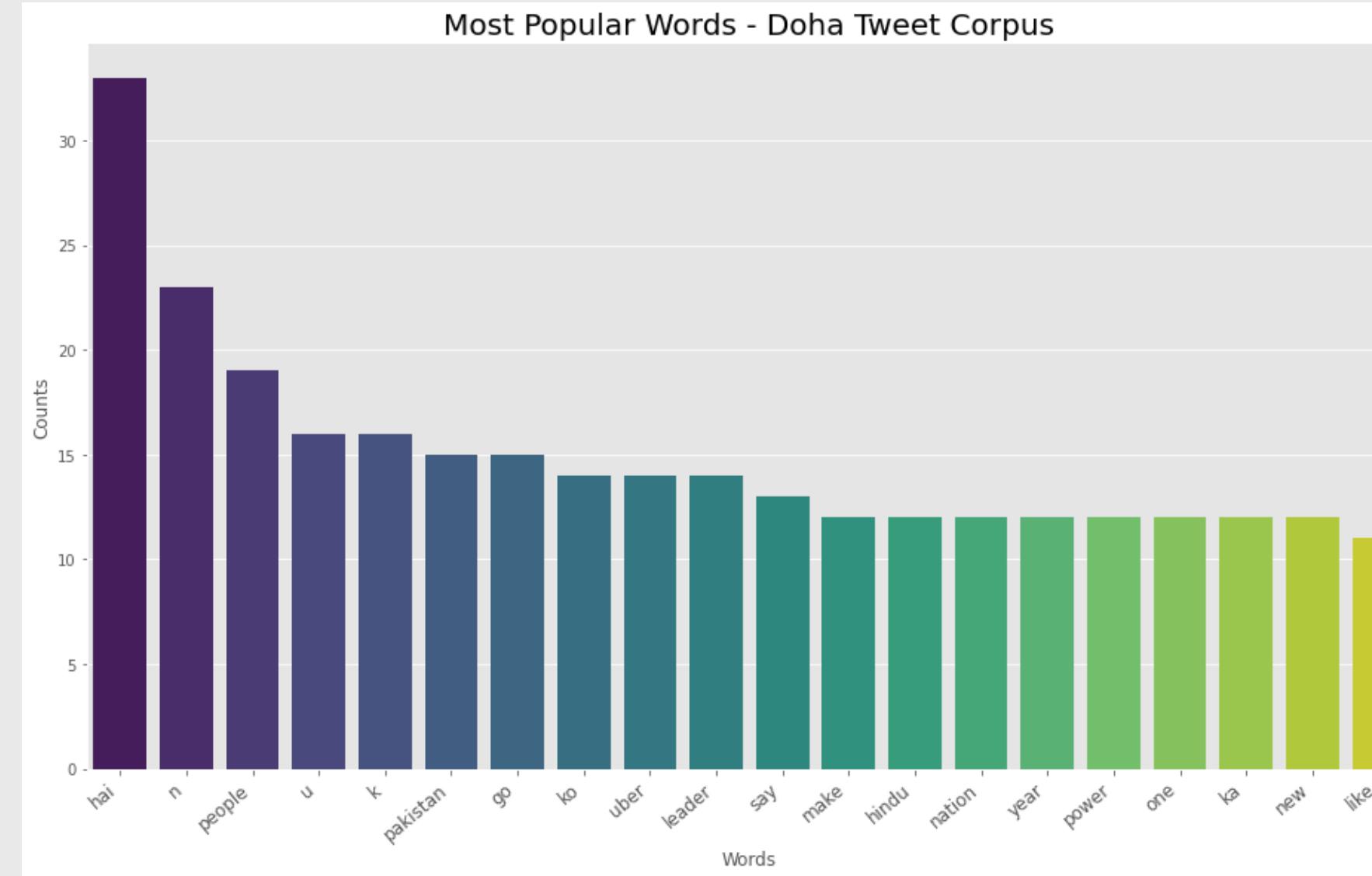
@DHSgov @USCIS - Ridiculous that Indian immigrants cannot go back to their home country for MONTHS, since you don.Äôt have visa stamping dates in India for them. Is this the price that immigrants have to pay to be in the country? @ABC @FoxNews @CNN <https://t.co/GTAktxtvzm>

Obviously the JD Vance who was a fierce critic of Trump was his voice as a husband of an Indian immigrant. The Vance who's a Senate hopeful is like Plato's animal tamer in The Republic (a demagogue following the Sophists). <https://t.co/Kgp8zn05Tr>

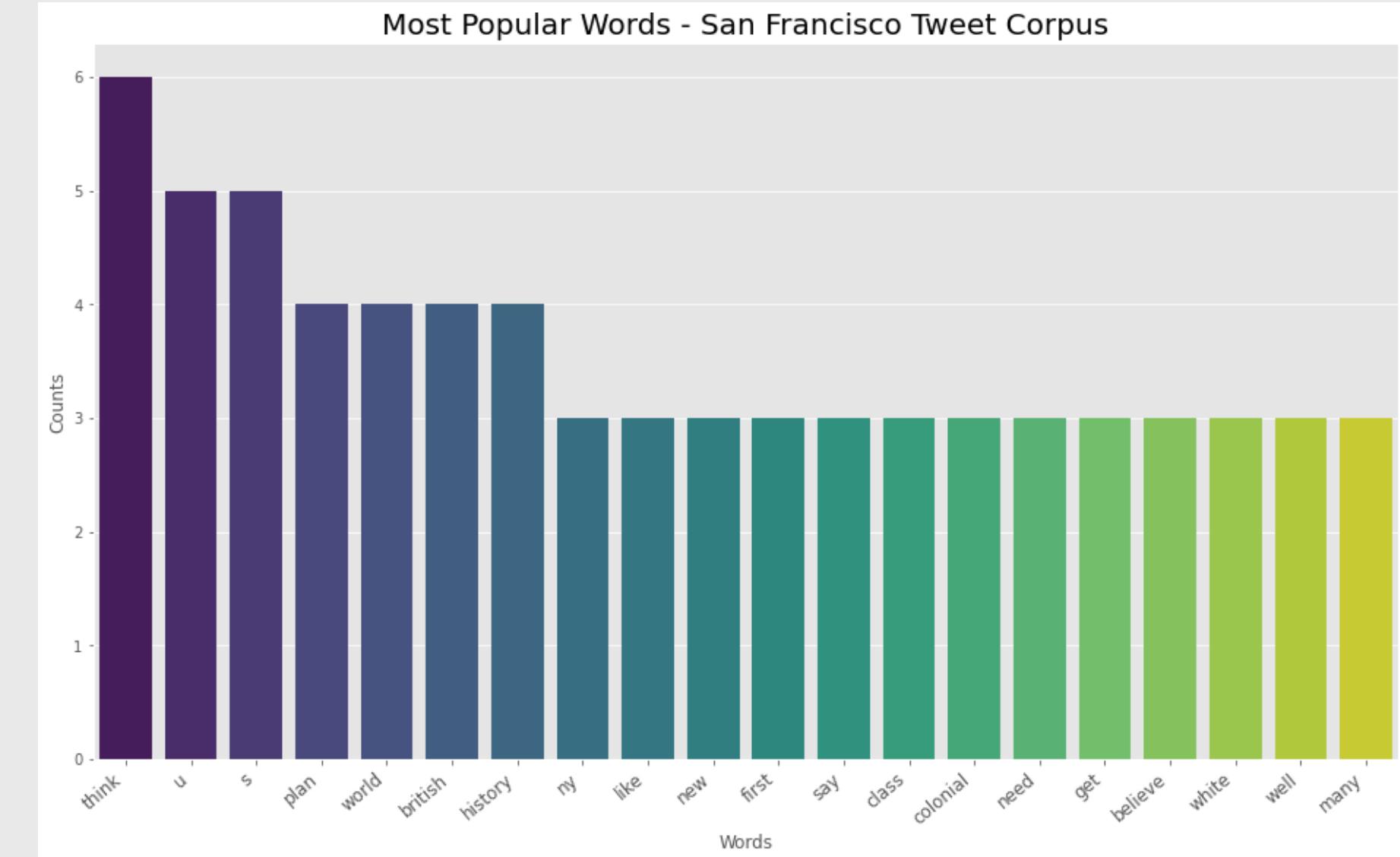
@adipalve Project Veritas come from the same sewer pits of American right wing politics as the tanton network. Getting an Indian immigrant fired in the process is just a bonus for this lot.

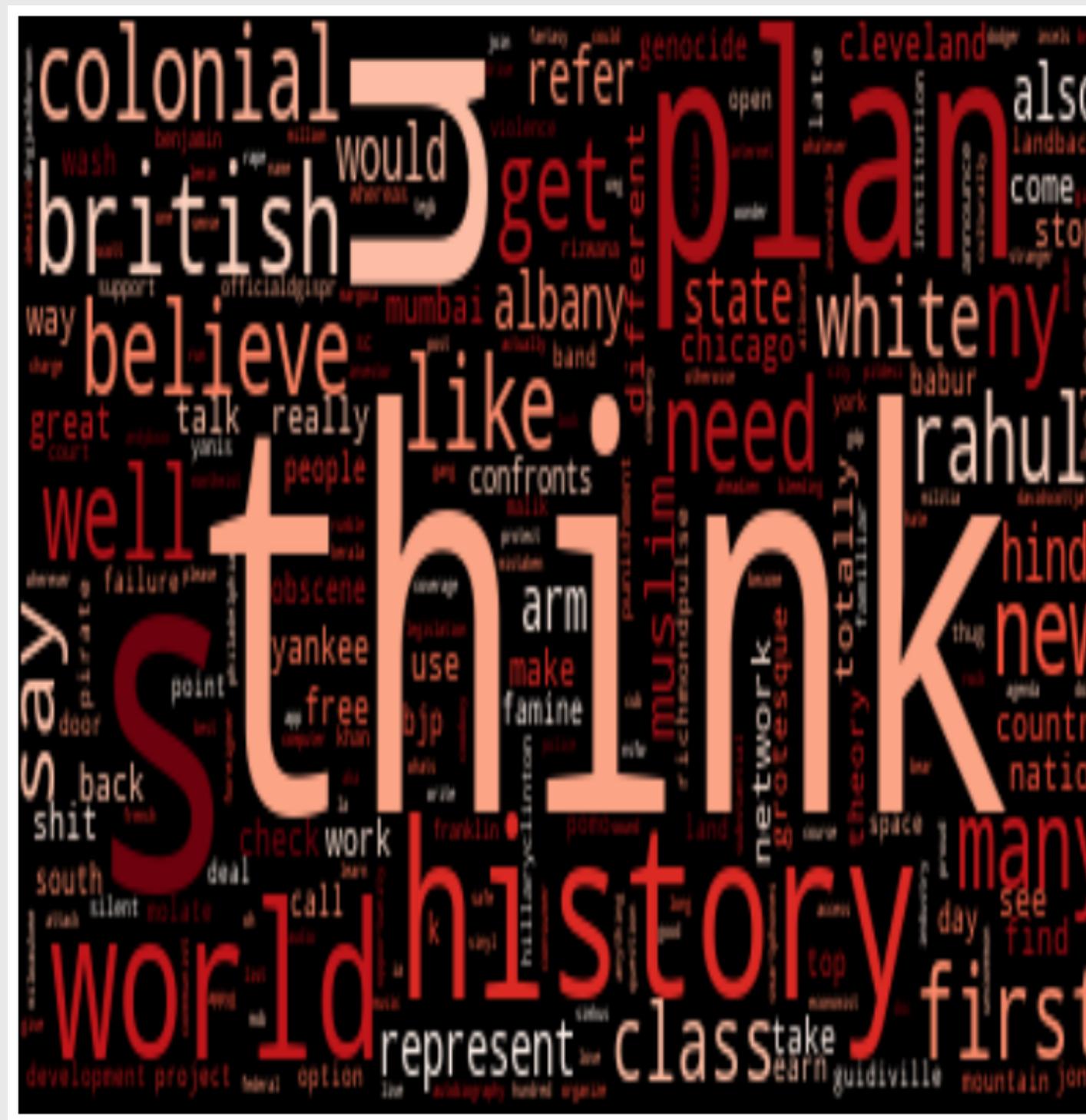
Table 2: sample tweets mentioning 'indian migrant' in Doha.

Most Popular Words - Doha Tweet Corpus



Most Popular Words - San Francisco Tweet Corpus





# Analysis



## Tweet Quality and Education-level

## Language and Assimilation

- Within the United States, Twitter users are more likely to be younger, better educated, and have higher incomes and vote Democrat than the American average → more likely to be ideologically tolerant^
- Twitter users represent a very small proportion of Dubai's population

## Demographic profile and tolerance

## Government Policies

- Can only say where the tweets are coming from, not who is tweeting them

## Relationship to Social Capital

# Limitations



## Skewed Demographics

- Within the United States, Twitter users are more likely to be younger, better educated, and have higher incomes and vote Democrat than the American average → more likely to be ideologically tolerant^
- Twitter users represent a very small proportion of Dubai's population
- English not natively spoken in the Gulf



## Existence of bot accounts and other Internet idiosyncrasies

- Generates statistical noise



## Challenging to isolate demographic identity of Tweet author

- Can only say where the tweets are coming from, not who is tweeting them



## Limited access to demographic data due to lack of transparency in the Gulf

# Implications



## Twitter analysis is a very new method of inquiry - lot of future scope

- Use more sophisticated NLP techniques that employ machine learning
- Analyze larger datasets considering other variables such as race, age etc.

## Examine other migration flows in different settings

- Regions that more actively use Twitter

# Acknowledgments and References



Davenport, James R. A., and Robert DeLine. "The Readability of Tweets and Their Geographic Correlation with Education." ArXiv.org, 23 Jan. 2014, <https://arxiv.org/abs/1401.6058>.

1 "Indian-American population is fastest-growing minority group". May 12, 2011.2

<https://www.pewresearch.org/social-trends/fact-sheet/asian-americans-indians-in-the-u-s/> 3 "Building Towers, Cheating Workers: Exploitation of Migrant Construction Workers in the United Arab Emirates" (PDF), hrw.org; accessed 27 October 2015.

4 Naaman Mor, Becker Hila, Gravano Luis. Hip and Trendy: Characterizing Emerging Trends on Twitter.

Journal of the American Society for Information Science and Technology. 2011;62:902-18. 5 Wojcik, Stefan, and Adam Hughes. "Sizing up Twitter Users." Pew Research Center: Internet, Science & Tech, Pew Research Center, 7 Jan. 2021, <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>.

## Github

<https://github.com/fedebotu/awesome-twitter-word-clouds/blob/main/2.%20Creating%20Word%20Cloud.ipynb>

<https://github.com/twintproject/twint/wiki/Pandas-integration>