A Migrant's Tale of Two Cities

Examining how the education level of Indian immigrants affects perceptions of them

Vidyut Baradwaj

Abstract:
India has one of the richest histories when it comes to immigration and emigration. The country is the world's biggest sender of migrants, with one in every twenty migrants being of Indian origin. More recently, the United States and the Gulf emerge as the destinations of choice for Indians[1], with around 13 million distributed between the two. However, the public image of Indians in either place couldn't be more different. While several multibillion-dollar American enterprises bear Indian-born CEOs, the working and living conditions of Indians in the Gulf face continual scrutiny for human rights violations, highlighted by the thousands of reported deaths of those involved in building infrastructure for Qatar's World Cup. Thus I explain how immigrants leaving the same country have different lives and why they're perceived differently is a function of both their education levels and those of the host country using sentiment analysis of Twitter, considering the specific case of the San Francisco Bay Area and Doha.

Introduction

With Parag Agarwal being named the CEO of Twitter in late 2021, India added another name to its list of some of the world's biggest companies that have an Indian immigrant as its CEO. At the same time, there are a slew of human rights reports exposing violations committed by employers in Gulf nations like Qatar, Saudi Arabia and the United Arab Emirates. I questioned whether the lives of Indian immigrants could be any more different, or whether this was simply my perception.

---

[1] Connor, Phillip. "India Is a Top Source and Destination for World's Migrants." *Pew Research Center*, Pew Research Center, 31 May 2020, https://www.pewresearch.org/fact-tank/2017/03/03/india-is-a-top-source-and-destination-for-worlds-migrants/.

I realized that there was a clear literature gap in comparing different immigrant groups originating from the same country. Additionally, none of them used Twitter as a means to gather data on perceptions of immigrants.

First, I will examine existing narratives on the two contrasting migration flows in the Settings and Context section. I'll follow this by developing a contextual understanding on the selectivity of migrants in both locations and in consequences through past literature in the Review section. In particular, I will build a model based on *migration selectivity* and *social capital* to forge a connection between education level and perception to form a hypothesis. Consequently, I demonstrate how I collected and analyzed two tweet corpora for each city to achieve two intents. First, to operationalize and gauge perceptions of migrants to either support or challenge the hypothesis. Second, to further enhance our understanding of the consequences of education levels on both the immigrant experience and perceptions of them while refining novel research methods. Ultimately, I will evaluate and expound upon our model using theories on assimilation and the *'readability'* of tweets while providing further implications for research in this field. In essence, I want to evaluate how useful tweets are in gauging public perception while confirming or denying the validity of our model.

With regards to the structure of the paper itself, I will establish a contextual understanding of the settings being studied in this paper: Indian nationals residing in Doha and San Francisco in the 21st century and how narratives of these immigrants reflect their demographic profile, why I chose these two cities in particular for comparison and to develop a demographic profile of Indian immigrants in either city with the aim to prove why our research is worthwhile. In the review section that follows, I will discuss prior academic explorations on the determinants of outmigration followed by a more specific look at selectivity of migration on the basis of education-level and social capital, connecting this to perception research, and migration research focused on the Gulf and the US. There will also be a subsection examining past research using a Twitter-based methodology to gauge perceptions. Accordingly, I will formulate a model based on these notions developed from our academic and contextual understanding to hypothesize what the contrasts in Twitter perceptions will look like. With this, I get into the data and methods section, in which I explain why I used Twitter as a data source and how I proceeded using a selection of Python packages to develop my corpora, clarifying the constraints I used and measures that were taken to make the data easier to analyze. Consequently, I present my results through graphs and word clouds and analyze them under the framework of my model, evaluating it based on the limitations of our research and looking at future implications in the conclusion section.

Settings and Context

My study will consider the case of Indian immigrants in the cities of the United States and Qatar specifically. As such, this section will provide a brief demographic context of both these countries, followed by the history and present landscape of Indian migration into these two countries. Consequently, I will focus on our Twitter case study, elaborating on why I chose to restrict my search to the cities of Doha and San Francisco in this study.

A brief overview of migration between South Asia and Qatar reveals many statistical and systematic quirks that indicate why it is one of the most unique recent demographic phenomena. For example, Qatar's population tripled in the decade leading up to to 2011, with 88% of its total population is migrant workers[2]. Saudi Arabia has the greatest number of foreign born residents, amounting to 13.1 million in 2019. Examining Qatar's population, there is a clear dichotomy between its residents. Non-Qataris residing in the country are consistently less educated than citizens, which is congruent with the fact that the vast majority of these migrant workers are unskilled to semi-skilled men working in either construction or the service industry, originating from Africa and other Asian countries like the Phillipines and the Indian subcontinent. Moreover, unlike any other nation in the world, nearly 1.5 million residents live in labor camps, as opposed to the 1 million that live in households or public housing.[3]

In stark contrast, Indians are the richest nationality group in the United States earning around $120,000/yr, in addition to being amongst the most educated, as per the Pew Research Center. Another study undertaken for the Kauffman Foundation on entrepreneurship reported that immigrants were founding fewer companies, with the striking exception being Indian Americans. While there is no similar report examining incomes of Indians in Qatar, it can be asserted that they are amongst the lowest earners in the country on the basis of (1) high number salary-related complaints to the Indian embassy[4] (2) job listings in the classifieds section of Indian newspapers, which are close to the minimum wage of QAR 1,000.

Why focus on the San Francisco Bay Area and Doha specifically? The two cities are appropriate settings for our research due to their unique combination of similarities and differences. In addition to being important global hubs for technology and natural gas respectively, the broader San Francisco Bay Area has been compared to Qatar in the past, wherein they would

[2]Qatar Statistics Authority

[3] Qatar Statistical Authority

[4] "Indian Migrant Workers Are Not Really Benefiting from Qatar's Labour Reforms." *The News Minute*, 30 Dec. 2020,https://www.thenewsminute.com/article/indian-migrant-workers-are-not-really-benefiting-qatar-s-labour-reforms-140533.

hypothetically be the two richest countries in the world based on per-capita GDP.[5] Most importantly, they both share large percentages of foreign-born, specifically Indian populations (as seen in Table 2) in the present-day, having seen a significant increase between 2000 and 2010.

	San Francisco Bay Area[6]	**Qatar**
GDP per Capita (PPP, 2017 est.) (Source: World Bank)	$128,308	$128,647
%-Indian-born population	237,672 (2010 census)	~700,000 (2017 est.)

Table 2: Comparing San Francisco Bay Area and Qatar

Literature Review

Education-level, migration networks and selectivity of Outmigration

While migration may be looked at as a macro sociological force, I examine it from the micro-level, i.e. what are the driving factors that make an individual more likely to migrate. Traditionally, education and migration share a positive relationship (IMF, 2016), though results were more regionally heterogeneous for highly-skilled and educated migrants. Van Dalen et. Al. (2005)'s study revealed that higher educated individuals were more likely to migrate in Ghana and Egypt, though the opposite was the case in Morocco. On the other hand, multiple studies, Beine et.al., 2011, Bertoli, 2010; McKenzie & Rapoport, 2010) all suggest that home countries with more dense migration networks lead to greater selectivity of low-educated migrants. Migration networks are defined to be sets of interpersonal ties between incoming, present, and former migrants in origin and destination areas through ties of kinship, friendship, and shared

[5] "If Silicon Valley Were a Country, It Would Be among the Richest on Earth." The Guardian, Guardian News and Media, 30 Apr. 2019,https://www.theguardian.com/technology/2019/apr/30/silicon-valley-wealth-second-richest-country-world-earth.

[6] The 'Bay Area' was considered to include only San Francisco and Santa Clara counties in this study for clarity

community origin (Lundquist & Massey, 2005). As these existing gateways allow for larger social networks, they significantly lower both the financial and social cost of relocation. Grogger and Hanson (2011)'s study focusing on OECD countries reveal positive-selectivity based on education, as they settle in countries with better compensation for their skill-level. In the context of my research, the presence of the well-established *kafala* system and the large-scale industrial operation to relocate migrants from India into the Gulf revealed by Gardner (2011)'s ethnographic study indicates the presence of a strong network in place, which supports the theory that there is a selectivity for low-skilled migrants. Conversely, the United States being an OECD member confirms the other side, where selectivity is positive for the highly-educated.

Several studies reveal the vital role of these social networks at every stage of the migration journey (Palloni, Massey et. Al., 2001), (Schapendonk, 2012 & 2015), (Schapendonk & van Moppes, 2007) as this facilitates the movement of important information from the destination to the origin and vice versa, lowering the anticipated risk and inertia to relocate. There are also feedback mechanisms, especially present in conflict zones, that prompt the perpetuation of a culture that strengthens the migration flow (Davenport et. All, 2003), (Barthel & Neumayer, 2015). On the effect of negative information flow on migration, studies like Van Mot et. al. 2018, and Fussel & Massey 2004 show that reports of hardships in the host country, including hostility, restrictive policies, and other hardships can reduce migration. The Gulf serves as a counterexample on this front, as migration remains strong despite negative perceptions.

Social networks, trust, and perceptions

Literature on immigrant perceptions also crucially reveal the important role of social networks. Cheong, Edwards et. al (2007)'s study of the UK demonstrates how social capital, which is value-based on largely defined on the basis of ethnic identity, social trust and cohesion, is linked to immigration attitudes. Herreros and Craido (2009) suggests that societies with greater social capital and cohesion exhibit more positive attitudes towards immigration. Helliwell and Putnam (1999) reveal the role of education in positively improving social trust. Thus, I also predict that Doha, which has a lower education level than the Bay Area (Figures 4.1 and 4.2), has lower cohesion levels, and is thus more likely to show negative perceptions than its American counterpart.

Thus, one can see how social networks of both the migrant and the perceiver play an intermediary role at both the relocation and settlement stage, allowing us to generalize a model that connects migrant education-level and perception. With this I build a model where education-level and perceptions, via social networks and cohesion, are positively correlated.

The book 'The Other One Percent: Indians in America' by Sanjoy Chakravorty, Nirvikar Singh, Devesh Kapur in 2016 was also instrumental in providing a contextual background for the processes of selection, assimilation and entrepreneurship of Indian immigrants in America.

Use of Twitter in perception research

My primary methodology employs Twitter, which has been used in the past to gauge public perception on a variety of topics, ranging from education policy (Saini, Singh et. all, 2020), education during COVID-19 (Mujahid, Rustam et. all, 2021), and general personality trait research (Lin Qiu, Han Lin, Jonathan Ramsay, Fang Yang, 2012).

Overview of Variables

The independent variable that allows us to characterize the two scenarios will be the dominant education-level. The dependent variable will be operationally defined as a combination of polarity scores, frequently occurring terms, and additional qualitative analysis of two random samples.

Data

The data used in this study is a compilation of census and survey data in tandem with a corpora of tweets, all viewed under the lens of my model. I chose to prioritize data that was published most recently as that best represents the broader trends examined in the Settings and Context section.

In particular, data used to measure the education levels of Indian immigrants in the United States was sourced from the US Census Bureau (USCB) published report "The Asian Population (2010)" and the Pew Research Center's analysis of American Community Surveys (ACS) ranging from 2015-2019. Data examining experiences of Indian Americans was primarily sourced from the 2020 Indian American Attitudes Survey (IAAS) conducted by the Carnegie Endowment for International Peace in conjunction with the 'Immigrants in California Factsheet' and 'Statewide Survey: Californians and the Government' published by Public Policy Institute of

California (PPIC) in March 2021 and January 2021 respectively. I believe this combination of data provides a strong basis for analysis due to its balance of objective demographic statistics (USCB) and more subjective collection of opinions

While the USCB data is a factual representation of demographic information like identity, education, and employment etc, the ACS and Pew data provide crucial insight into opinions and attitudes.

For Qatar, this data was sourced from the Annual Economic Report of the country's Ministry of Economy, giving us a breakdown of the education-level for non-Qataris in 2015. I support this statistical data with more qualitative insight from existing literature focused on the region, including an ethnographic account by and other reports published by the Human Rights Watch[7], and Amnesty International, allowing us to further develop my theoretical model.

It is worth noting that there was little , something I note in my analysis.

Methodology

My primary research approach encompasses the development and analysis of two corpora of tweets that each focus on the two cities that form the setting of my study.

I scraped and organized these tweets using Twint and Pandas - both Python-based packages used to scrape tweets and conduct data analysis respectively - in a Python notebook. The utility of Twint is in its ease of use and lack of restriction, providing an alternative to the Twitter Application Programming Interface, which has those drawbacks.

[7] "Building Towers, Cheating Workers: Exploitation of Migrant Construction Workers in the United Arab Emirates" (PDF), hrw.org; accessed 27 October 2015.

Tweets were chosen on the basis of geographical and topical relevance. First, I formulated a set of keywords that would constrain my search to only those relevant tweets that mention them. I then programmed a scraper function (as shown in Figure 1) using Twint to constrain to tweets that had both

- their geotag parameter associated with either Doha or San Francisco, i.e they were posted in that given city
- at least one of the keywords in the keyword set

The advantage of using only geotagged tweets is that they're inversely related to follower count. In other words, users with more followers are more likely to post geotagged tweets. The benefit is two-fold - this eliminates a large proportion of non-human actors and bots with no followers and also skews towards more visible accounts with greater influence.

Once these tweets were collected, they were stored into a Pandas DataFrame, a two-dimensional tabular data structure that is convenient for sophisticated data analysis and manipulation.

```
import twint, pandas as pd
keyword_set = set()

def tweet_scraper(keyword_set, city):
   c = twint.Config()
   for word in keyword_set:
       c.Search = word #topical constraint
       c.Pandas = True
       c.Near = city #geographical constraint
       c.Pandas_clean = True
       twint.run.Search(c)
   return twint.output.panda.Tweets_df[["username","tweet"]] #store into Pandas DataFrame
```

Figure 3.1: Representative scraper function written in Python

The data frame was then 'cleaned' using a cleaner function (Figure 4.4) before analysis. This cleaner was used to eliminate

- elusive properties like multimedia, emojis, and links to only focus on the tweet's text for clarity.
- Stop words or commonly-used words like 'the', 'is','and'[8]
- Case inconsistencies by making all the words lower case

```
def cleaner(dataframe):
   # Add whitespace to the end of every tweet
   dataframe['cleaned_tweet'] = dataframe.tweet.map(lambda x: x + " ")
   # Remove http links
   dataframe.cleaned_tweet = dataframe.cleaned_tweet.map(lambda x: re.sub(r'http.*', '', x))
   # Remove special characters and numbers
   dataframe.cleaned_tweet = dataframe.cleaned_tweet.map(lambda x: re.sub(r"[^a-zA-Z#]", '
', x))
   # Lowercase all tweets
   dataframe.cleaned_tweet = dataframe.cleaned_tweet.map(lambda x: x.lower())
   #Tokenize tweets and remove stop words
   stopword_list = stopwords.words('english')
   for i in tqdm(range(len(dataframe.cleaned_tweet))):
       tokens = word_tokenize(dataframe.cleaned_tweet[i])
       clean_tokens = [w for w in tokens if w not in stopword_list]
       dataframe.cleaned_tweet[i] = clean_tokens
```

Figure 3.2: Representative cleaner function written in Python

I then used basic natural language processing (NLP) techniques to conduct analysis on my corpora. Perceptions were operationally defined as a combination of the sentiment breakdown and term frequency, both widely-used natural language processing techniques. Sentiment breakdown indicates the polarity of each tweet from -1 to 1, where -1 represents negativity and +1 represents positivity, by computationally identifying and designating individual words with a certain emotional value.

Practically, I used the packages textblob, Natural Language Toolkit (NLTK), and matplotlib for sentiment analysis, term frequency, and data visualization respectively. For sentiment analysis, each tweet is put through the textblob-powered engine which produces the polarity score

[8] This list was refined and enhanced to include words more specific to my research over the course of analysis to generate better results

described above. Accordingly, calculate the overall distribution for each corpus to produce a final representation of each city's sentiment on the same plane for ease of comparison.

Concurrently, I compute the term frequencies using the NLTK package. First, I split the text of each tweet into individual words or 'tokens'. Following this, each word is attributed to its part of speech and is then 'lemmatized' or stemmed to its root form ('prettier' becomes 'pretty'). Finally, I collate and identify the most frequently occurring terms or phrases for each corpus by setting a k-value (1-gram is a word, 2-gram is a two-word phrase etc.), which I represent in both wordcloud and bar graph form.

I also plan on manually analyzing the sentiment of a random sample of 7 tweets from both corpora to add a dimension of qualitative analysis to my research.

I believe this methodology brings novelty to my research for several reasons. Generally, data-science oriented digital methodologies are rapidly growing as a paradigm of social science in the last decade. As sophisticated computational tools have made it easier to analyze Big Data, there are a growing number of APIs developed by organizations associated with large-scale data banks including World Bank, NYTimes etc. Moreover, Twitter itself is such a powerful yet accessible platform for social sciences to harness real-time behavioral data without the logistical challenges of collecting survey data or interviews, which are more resource-intensive. Opinions on Twitter are also relatively immune from social desirability bias compared to surveys or interviews as opinionated tweets are often authentic and are known to provide an accurate illustration of a population's preferences and activities in real time.[9] Thus, Twitter offers a more

[9] 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.

novel, contemporary insight into a demographic's opinions and attitudes that surveys necessarily cannot.

While I believe the data mining of social media platforms like Twitter is set to be a prominent part of social science research in the future, there are quite a few limitations in terms of scraping tweets, both in principle and my methodology. Firstly, it is important to recognize that the demography of Twitter users in both Doha and San Francisco is skewed relative to the cities themselves. Within the United States,

Though there have been studies that attempt to develop an accurate framework to identify the demographic backgrounds of Twitter users solely from what they tweet, my methodology doesn't distinguish between the identities of the tweet posters themselves. Hence, tweets made by immigrants themselves are also included in our corpora. Additionally , any data scraped from Twitter is far from 'clean'. The platform has numerous oddities and idiosyncrasies that produce statistical noise. With the current level of sophistication in NLP, I am unable to filter out such unwanted noise from bad-actors and non-human accounts (or bots) that could contaminate results.

Moreover, the demographics of Twitter is skewed. 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.[10] As per Ganzach and Schul (2020), this makes them more ideologically tolerant than their conservative counterparts who are not as well-represented on Twitter.

This demographic bias is much stronger in Doha, as Twitter is nowhere near being the most popular social media platform in the region, with only 11% of the population deemed to be 'heavy users'.[11]

---

[10] 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/.

[11] Ministry of Information and Communication (2015)

Results

I first present my findings on the education-level of immigrants in both the SF Bay area and Doha before those on my research of Twitter data.

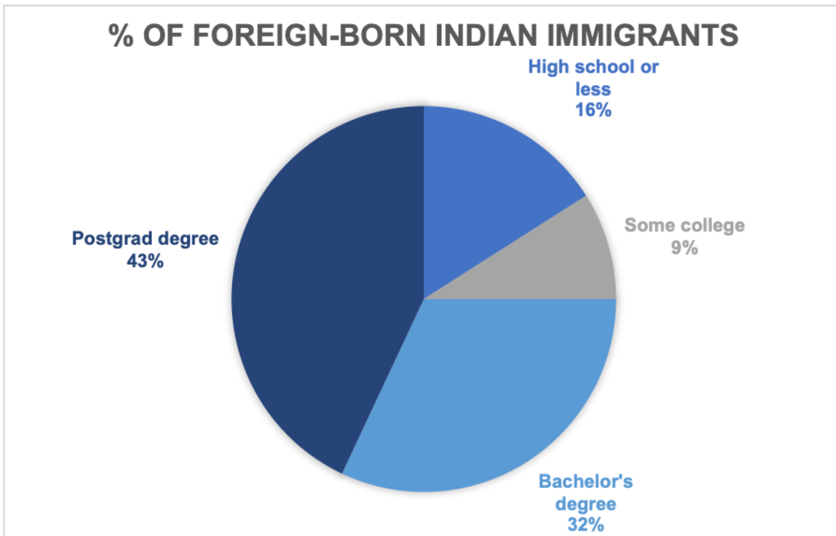

Figure 4.1: (Source: Pew Research Center analysis of 2017-2019 American Community Survey (IPUMS)

Education Level	**Percentage**
High school graduate or higher	88.65%
Bachelor's degree or higher	56.15%

Table 1: Education levels of persons age 25+, avg. of San Francisco City and Santa Clara county

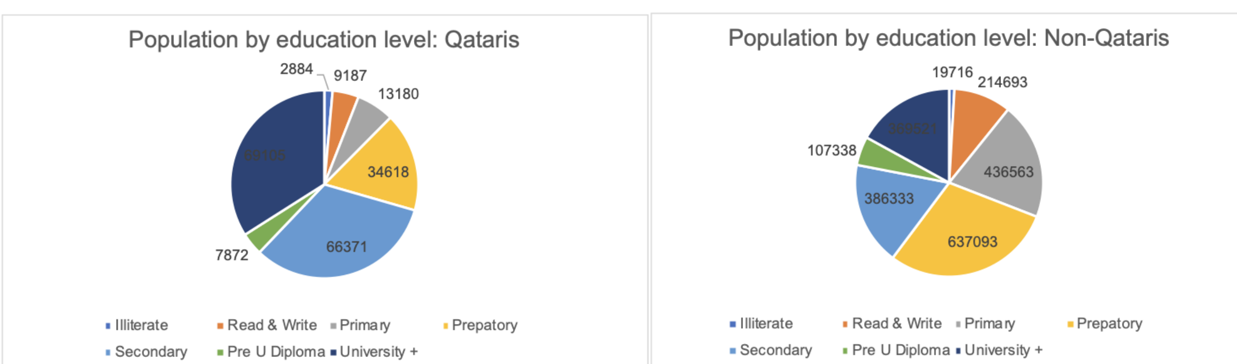

Figure 4.2: Comparing education levels on the basis of citizenship status, Qatar (Source: Qatar Statistics Authority)

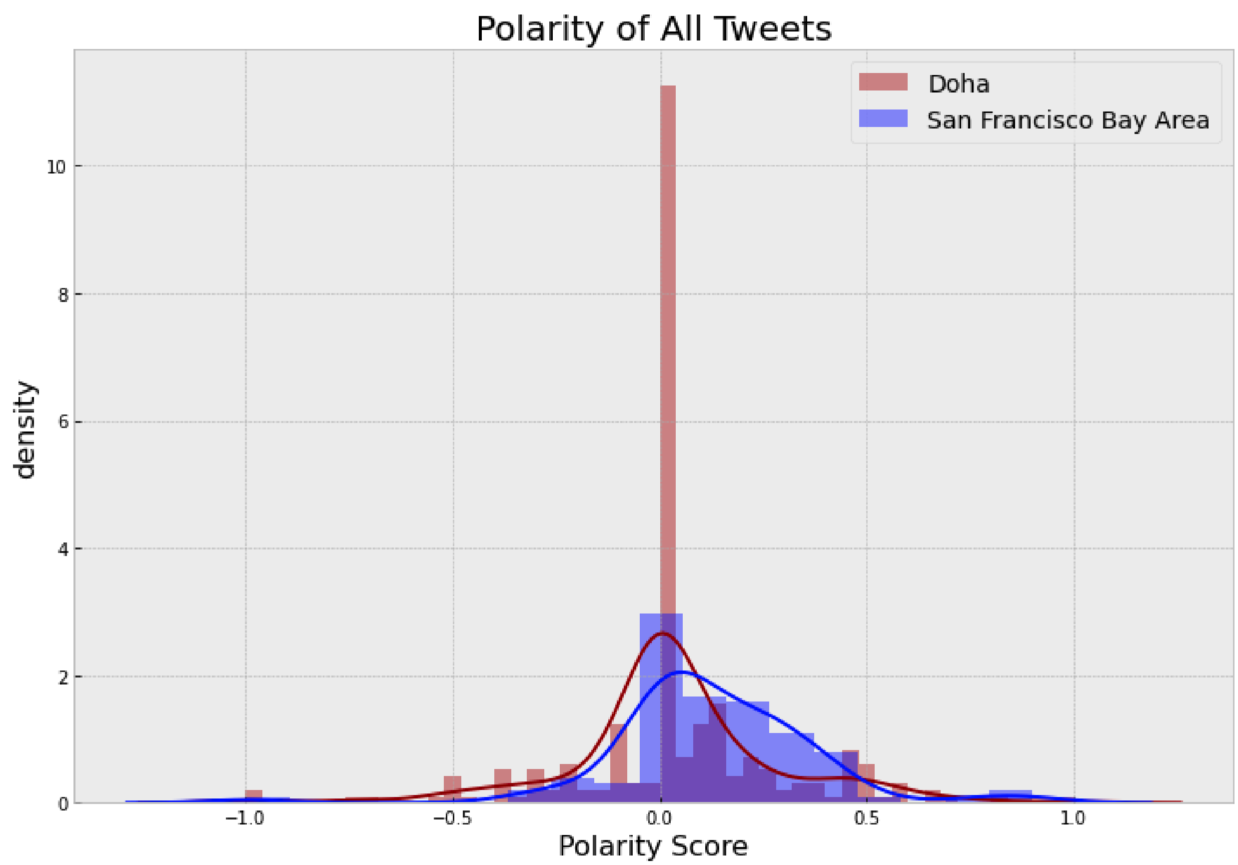

Figure 4.3: Polarity-density graph for the two corpora. +1 indicates positive, -1 indicates negative

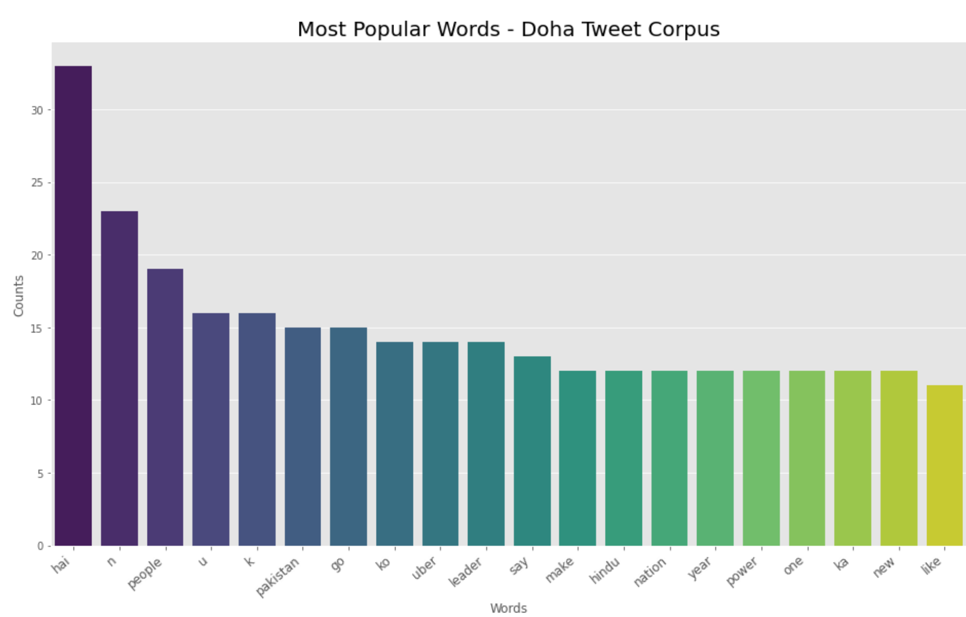

Figure 4.4: Most frequently-occuring words, Doha corpus

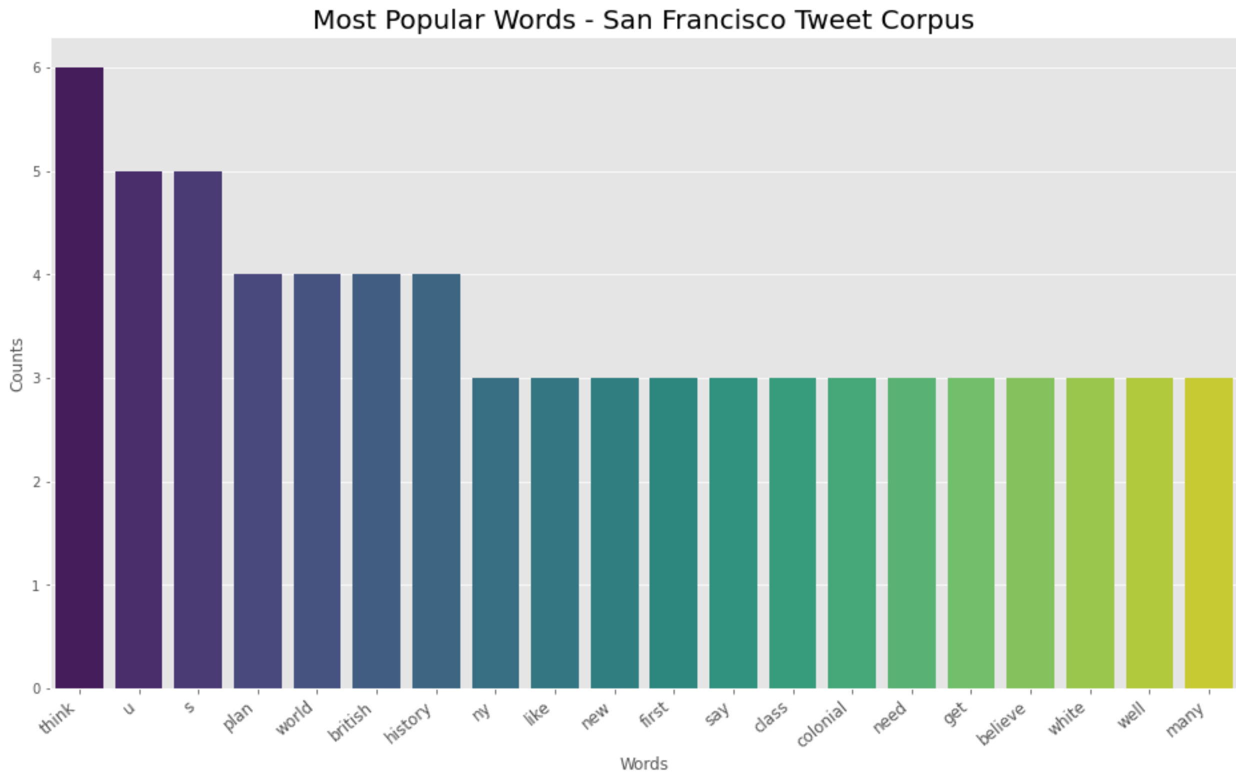

Figure 4.4: Most frequently-occuring words, San Francisco corpus

Figure 4.5a and b: WordClouds for San Francisco Corpus and Doha Corpus respectively

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: Random sample of 7 tweets mentioning 'indian immigrant' from Bay Area Tweet Corpus.

Tweets
@LethalLafanga Extremely privileged. They're used to doing 2 hrs of work and making quadruple the amount of money Pakis/Indians do. Goray earn around the same as them but a little less. FOR THE SAME AMOUNT OF WORK.
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
I have a question regarding the English language. Is it possible that native English speakers don't understand English spoken by Nigerians, Filipinos, Arabs,Indians....?🤔
Female Nurse space is dominated by Philipinas & Indians in GCC. They have a degree while ours have diplomas. Could earn several times more than maids. But yet we are to yet give them degree status. Is it cos doctors oppose ? Just don't get it 🤦‍♂️ @nanayakkara77 @RW_UNP @GotabayaR
Undocumented: Stories of Indian Migrants in the Arab Gulf by Rejimon Kuttappan is a critical contribution to fill this huge vacuum. @MigrantRights

Table 3: Random sample of 7 tweets mentioning 'indian' from Doha Tweet Corpus

Analysis

This study has provided considerable insight into how differently the two immigrant groups are perceived, generally and more specifically in these two cities. My aim is to compare and contrast the results from our two corpora, which include the word clouds, frequency charts, sentiment breakdowns, and following which I argue that the findings support the model. From there, I will analyze the role of other confounding variables that play a significant role in the immigrant experience, both in terms of influencing social capital, and perceptions.

Education Level

In Figure 4.1, I observe that the dominant education group of immigrants are those with Post-grad or Master's degrees, with about 84% immigrants having attended college of some sort. In contrast, in Figure 4.2, I observe that only 17% of non-Qataris are known to have received education at the college level. It is worth noting the added nuance in the data collection within school education. Thus, it is clear that Qatar attracts a clearly different demographic of the Indian workforce from the Bay based on what I predicted in the review section. Additionally, I see a larger proportion of Bay Area residents to be college educated than both the Qatari and non-Qatari populations.

Twitter Corpus

While there is not much I can discern from comparing the frequency charts and the word clouds, there is a distinct difference in *readability* of the Doha corpus, which is defined to be the complexity of language and ease of understanding, is lower than the Bay Area corpus.This is exhibited by the greater occurrence of typos like 'pakista', 'millio', and 'hidu' in addition to several single character-words, which I can also attribute to posters typing in other languages as English is less prevalent in Qatar[12]. Based on Davenport and Deline (2014), one can observe the positive correlation of readability to education level.

The polarity scores were more indicative of a contrast between the corpora. Tweets from the SF corpus showed greater proximity to +1 i.e. were more positive than those from Doha, which were mostly neutral in nature. The other reason for high neutrality is that engine was unable to assign a polarity score due to the quality of the tweet

Table 2 reveals some key takeaways in how American-Indians are perceived on the internet. One can broadly attribute that these tweets exhibit emotions of respect and empathy for the Indian-American community. While the first tweet refers to the popular stereotype of Indians driving cabs, the third tweet is very telling in its direct declaration of Indian-Americans being the 'most loved and respected communities' in America. The use of the word 'trust' exemplifies the 'social trust' notion, strengthening its basis in my model. The second tweet asserts that Indians living in Silicon Valley used to vote both Republican and Democrat, reflecting political parties to be a strong driver of division for Indian Americans, as confirmed by the 2020 IAAS. The last three tweets are more emphatic, referring to the (alleged) anti-immigrant rhetoric of American right-wing politics and the bureaucratic hardships faced by Indian immigrants. In all, I can assert that the overall sentiment for Indian Americans based on the San Francisco Tweet Corpus is strongly positive or neutral.

Table 3 is also greatly insightful. The. It is unclear who 'they' refers to, but it clearly communicates the poor compensation of '*Indians/Pakis*' as compared to *'Gorai'*, or the Hinglish equivalent of 'white'. Upon first look, the phrase 'bottom of the pyramid', and 'undocumented' are noted. A closer inspection reveals that the tweet refers to the recent publishing of a book with

[12] I chose to refrain from filtering the single-characters as this would let me comment on the readability

the same title, compiling stories of Indian Migrant Workers in the Gulf. There is also a reference to education level, the role of islam in the 'stratification' and Indians being at 'the bottom of the pyramid'. In all, these tweets are expressive in their dissatisfaction for the status quo, and contemplative of why that is the case.

Social Capital model

I believe these results were largely successful in demonstrating a strong contrast in perception, thus exhibiting the correlation predicted between education-level and perceptions of Indian immigrants by my model, which uses social capital as the intermediary variable.

Still, there are other significant confounding factors that impact how immigrants in both nations are perceived beyond the role of social trust and capital

Language assimilation

A key difference between Qatar and the United States is the dominant language. While the former speak Arabic, it is primarily English for the latter. Per Abrar ul Hassan (2010), English plays a key role in the formation of the social identity for many South Asians, including Indians. This can broadly be attributed to the colonial presence of the British empire in the region up until the mid-20th century, which played an indispensable role in revolutionizing the linguistic framework of the subcontinent. Moreover, formative research on migration, including a Pew factsheet on Assimilation and Language published in 2004 suggests that there is a positive relationship between acquisition of English and assimilation into English-speaking countries. Supporting this, the 2017-19 ACS data suggests that 77% of Foreign-born Indian Americans are proficient in English, which is the highest of all the other Asian-American subgroups.

Social conservatism and tolerance

The role of the host nation's social values is beyond crucial, a characteristic that is so different for the two settings. San Francisco (and California) is a largely liberal-leaning region, and has always represented a large heterogeneity in immigrant populations. This strongly contrasts the largely conservative nature of Qatar's social environment.

Government intervention

As discussed in the Settings and Context section, migrants face institutionalized barriers such as labor camps and the *kafala* system that prevent them from fully assimilating with the local population. Clearly, this is the largest factor in what differentiates the two migration groups

Implications and Conclusions.

I believe perception research has important implications in public policy implementation, both on the sending and host country side. Given how Indian-Americans have been established as being such an influential demographic, governments may be more incentivized to improve visa related barriers that continue to hampen migration flows. On the Gulf side of this matter, there are important questions regarding the economic and social sustainability of its current *kafala*-backed model, especially with growing moral concerns over its treatment of its workers.

In conclusion, this paper set out to achieve a lot, establish migration narratives, explain why they are the way they are by developing a model, employ a novel method of researching perceptions, and analyzing them on the basis of the model.

While I think I was broadly successful in its ambition, this paper is simply a starting point for any perception-related research that uses Twitter as its data source, especially research focused on other migration flows in different settings, especially those that use Twitter more actively . Of course, I can use more sophisticated NLP techniques where I use machine learning to train a model and then apply it to our corpus to produce more accurate results. Our corpora was limited to just a few hundred tweets, so larger datasets that consider other identity parameters like race, gender, and age would also be useful.

Regardless, one can see how Indian immigrants arguably form a clear socio-economic backbone in both the countries I've explored, albeit in very, very different manners. And herein, lies my account of a migrant's tale of two cities.

Literature Cited

Discussion Paper 9/2019 - DIE_GDI. https://www.die-gdi.de/uploads/media/DP_9.2019.pdf.

Helliwell, John F., and Robert D. Putnam. "Education and Social Capital." *NBER*, 1 May 1999, https://www.nber.org/papers/w7121.

"If Silicon Valley Were a Country, It Would Be among the Richest on Earth." *The Guardian*, Guardian News and Media, 30 Apr. 2019, https://www.theguardian.com/technology/2019/apr/30/silicon-valley-wealth-second-richest-country-world-earth.

"Indian Migrant Workers Are Not Really Benefiting from Qatar's Labour Reforms." *The News Minute*, 30 Dec. 2020, https://www.thenewsminute.com/article/indian-migrant-workers-are-not-really-benefiting-qatar-s-labour-reforms-140533.

Pattisson, Pete, and Niamh McIntyre. “Revealed: 6,500 Migrant Workers Have Died in Qatar since World Cup Awarded.” *The Guardian*, Guardian News and Media, 23 Feb. 2021, https://www.theguardian.com/global-development/2021/feb/23/revealed-migrant-worker-deaths-qatar-fifa-world-cup-2022.

Singh, Ruchi. “Origin of World's Largest Migrant Population, India Seeks to Leverage Immigration.” *Migrationpolicy.org*, 9 Mar. 2022, https://www.migrationpolicy.org/article/india-migration-country-profile.

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.

Connor, Phillip. “India Is a Top Source and Destination for World's Migrants.” *Pew Research Center*, Pew Research Center, 31 May 2020, https://www.pewresearch.org/fact-tank/2017/03/03/india-is-a-top-source-and-destination-for-worlds-migrants/.