Cover Letter

Dear Migration Policy Institute,

Attached to this cover letter you will find three different documents, a problem statement regarding migration in China, and literature review on the studies that were reviewed in order to gain a further understanding on migration, and a research proposal created by yours truly in order to search for a solution to this issue. I am very passionate about this issue as I am the child of two Chinese immigrants and have visited China multiple times, witnessing firsthand just how cramped the city has become and how clogged the roads are everyday. I believe that the problem stems from migration of people from rural to urban areas, also called rural flight. My research questions aim to find a reason for this and to help prepare the country so that they can make quick and proper adjustments to keep migration , regardless of whether it is everyday commuting or long-distance travel, smooth and with no issue. Big data plays a large role in my plan to accomplish this study as I can generate large amounts of data very fast that is also accurate based on location data from big Internet companies like Tencent. I believe that my topic is one of the most important topics to be considered due to just how large scale migration is, since every country has it occur within their boundaries. China is not the only large country that has issues with migration and I am sure other large countries could use better infrastructure when it comes to migration. Any advancement on migration in China can be quickly adjusted and adapted for other countries to use, so this study could transition from just a country to globally, possibly being beneficial to the entire world as opposed to one small area.

Sincerely,

Owen Zhang

Problem Statement

Migration is an extremely important and relevant topic in society today. Everyone must travel around for something, whether it be for a job, school, shopping, or visiting people. There is large-scale migration, which is traveling long distances or purposes such as a new business opportunity, while there is small scale migration, which involves things like the commute to work or a drive to a restaurant in order to eat a meal. Regardless of large or small, people migrate everyday with different intentions and purposes, causing them to interact and communicate with others, which then affects how those people can act or interact sometimes creating a complex web of interactions, making migration in general a complex topic. Migration involves many different sustainable development goals, but the ones that are most relevant to urban migration are economic growth, good health and well-being, sustainable cities, and building good infrastructure as these are all necessary to prevent difficulties and harms for appearing as a result of migration. These sustainable goals allow for a healthy and good city environment to be maintained.

Although migration is prevalent in every country, the countries that have high urbanization are the ones most affected by migration due to the high populations that most urban cities boast. A higher population means more interactions and more people that are moving about, increasing the level of small-scale migration as people go about their days doing everyday things. There is also more large-scale migration in countries that have large urban areas, as people will be constantly migrating to the urban areas in search of a better or more modern place to live and a job that pays more or is more convenient for them. China has the largest population in the world and due to the popularity and population in urban centers, it is a good country to examine for trends in migration (Pan and Lai 2019). This is especially true in China, where people often move out from rural areas and instead travel or migrate to larger urban cities like Beijing, Shanghai, and Hong Kong. Migration presents a lot of harms if not controlled for or planned, such as urban sprawl and loss of the environment or the displacement and competition between those who live in the city for living space. As the population of a country increases, there will always be a need for expansion in order to house all the people as well as create more infrastructure and other different facilities in order to support an increase in population. However. There must be a meaning and a purpose for expansion so that governments don't end up meaninglessly expand the urban boundaries of a city. Urban sprawl is when urban areas are allowed to grow without any thought put into infrastructure or urban planning (Deng et al., 2019). This creates areas that are not needed and increases the need to travel and have small-scale migration as people who live in those areas must travel even farther to get to a desired place. The environment is also damaged in the process as vegetation and other land must be removed and adjusted so that roads and buildings can be created. There is no need to sacrifice trees and other parts of the environment when there would not be much use and only an increase in pollution for an extra area of land. As with most cities, Chinese cities do not only just build outwards, but instead also build upwards as there are many large skyscrapers and other tall buildings in China, reducing the need to expand the city boundaries to some extent. However, there is also a need for urban expansion in some cases, as a failure to have available experience makes living conditions and cities themselves cramped places to be. China is experiencing this right now as in its capital city of Beijing and other cities, there is a limit on who can own a car due to the crazy amount of cars on the highways. The Chinese government actually set up a drawing system to decide who can buy a car, and there are also regulations that say certain cars cannot be on the road on certain days in an effort to keep pollution down and roads clearer and less congested. The streets are often crowded as the large population travels individually through bikes, walking, taxis, subways in addition to cars due to the space being limited. Housing is also increasingly expensive for increasingly smaller areas because the demand for housing near more urbanized areas is so high. What could buy you a home in the United States can only buy you a small apartment due to the demand being so high. Cramping of areas and a heavy reliance on public transportation is also an issue as it allows diseases to travel quicker. Because of all the close person-to-person contact in buses and other means of transportation as well as the crowding on the streets mixed with the long-distance travels of many to different cities, it is very easy to spread a disease in China. These are not harms that are unique to China because these issues can happen anywhere around the world, however due to the fact that China has such a large population and such a large amount of people concentrated in its urban areas, these issues are just more prominent and on a larger scale in China than in other countries. If answers to these issues can be found in as large an area as China and its urban centers, then it can surely be applied to countries where these issues happen on a smaller scale.

Tracking the migration patterns of an individual is an easy task. Tracking the patterns of ten people is a slightly more difficult. Tracking a whole population is considerably more difficult to pull off. When you have to track a population of a little more than one billion people while taking into account all the different methods of travel such as high speed rail, car, or subways and planes and look into all the cities and all the distance that people have to travel, tracking everyone becomes nearly impossible. The difficulty in tracking so many people makes migration in China a complex system as there are so many different levels and different intentions that must be taken into consideration. Because of this complexity only general trends can be seen while studying how the population migrates, but even those are only useful to some extent. China used to be an underdeveloped country that was late to the trend of urbanization. However, in recent decades that changed and now the urban centers of China are extremely large and well-developed, changing from about 18% urbanized in 1970 to about 50% now, and new cities are becoming more and more developed with each passing day (Gaughan et al., 2016). The population has also boomed even further, with about 54% of the population living in urban centers and having that number be predicted to grow up to 70% in about ten years (Gaughan et al., 2016). Although there are still considerable amounts of people living in rural areas, the increase in urbanization has attracted more people to come to urban centers for jobs as one does not need a high degree of study to do cheap labor, which urban centers have a lot of opportunities for. Because of pull factors like this and push factors such as a desire for better living conditions, it can be expected that people would move increasingly from rural areas to urban areas in an attempt to find better opportunities for themselves and for their family. This is just one trend out of many that can be examined while exploring China further.

The main methods of transportation in China are walking, the taxi, the subway, and public buses. All of these are affordable and very easy to use, allowing for quick travel which also reduces the amount of people in cars on the roads. For long distance travel, there are planes and high speed rail, which also allows for fast travel and makes it easier for people to migrate to different cities regardless of purpose (Zhang et al., 2019). These methods of transportation allow for cities to be more connected to each other. Migration in China must be smooth due to the large number of people and just how large of a scale migrations are in China, especially during holidays. A mistake or an issue such as a blocked point of migration can eventually become widespread and cause major issues due to the connection between cities. One major holiday that causes an increase in temporary migration is the Lunar New Year. Because it is the start of a new year, it is a common tradition for people to leave from wherever they are and return to their families. Because of this, the week before the holiday, migration numbers increase by a lot due to all the workers in different cities and people who live elsewhere making the journey back to their hometowns (Hu, 2019). Good urban planning and management would greatly benefit the accommodations that would need to be made for either an influx of people in rural areas or hometowns and an outflux of people out of urban cities in the process of returning home. Different events in China such as the Lunar New Year can also cause random changes in migration which must be accounted for in the overall consideration of migration. With all the different methods of transportation, difference in rural and urban areas, holidays and events causing migration, and population of the country as a whole, China is a very complex country in terms of migration which can be looked into further.

Literature Review

Throughout my research I looked through many different articles in an attempt to learn more about my topic and the data methods that were used to conduct studies. The most used and more accurate data that was pulled from a dataset were data that was used by Tencent. Tencent is a large Internet-based company that releases a lot of different products. They range from online video games such as League of Legends or Call of Duty to different communication apps such as Wechat or QQ. Both Wechat and QQ are a part of the few select communication or social media apps that are allowed by the "Great Firewall of China". This is a firewall in China that effectively prevents social media apps that are popular in the United States from working, such as Facebook, Instagram, Snapchat and Twitter from working. It also prevents Youtube and Google from working, basically forcing anyone who wants to use such apps while in China to install a VPN. Because other communication apps and social media are blocked, Chinese citizens don't have much of a choice when it comes to using apps to communicate, having to use Wechat. In fact, Wechat has many features that are installed within its citizens to do many different things, such as pay electronically at grocery or receive money from friends on New Years after going through an online verification process to make sure you are a Chinese citizen. Because of Wechat's popularity, it has a widespread influence and use has allowed it to accumulate 963 million active users monthly, while QQ has 850 million (Pan and Lai, 2019). Tencent is able to collect data from the locations of when the app is used, and then store that location-based data allowing it to be used in different studies (Ma et al., 2018). Because the data is collected in real time and it is constantly updated, Tencent's big data is one of the more accurate and reliable sources to pull data from to use in a study or to conduct research on. Taking the entire population of China into account in comparison to the number of active users it has monthly, there is only a small proportion of the country that either does not have phones or do not have any Tencent related application causing them to not be covered by Tencent (Zhang et al., 2020). As compared to census data, big data would be more reliable especially in the aspect of transportation as it can be updated in real time and it can also be used to get lots of data quickly, but census data also contains information that big data can't record, like childbirth and other factors. Census' are also not conducted that often, meaning there is more variation in years that can't be seen because the government does not conduct censuses yearly. Some of my studies used census data while some of my studies used big data and datasets and some incorporated both into their studies. Overall, I think it is acceptable to use census data, but there should be an effort to include big data as well.

The two best methods that I came across were the random tree model and the complex analysis model while many other models or studies I looked into were also useful in researching more about my topic. Essentially, a random tree model consists of individual decision trees which make predictions. Then these individual decision trees are combined together into an ensemble prediction, which is more accurate than any individual decision trees. The researchers pulled the data from the National Bureau of Statistics and matched with their GIS administrative boundaries between years of 1990, 2000, and 2010. The data was then aggregated to the Global Administrative Unit Layer, which is a spatial database. They then used covariates that weren't affected much by time to reduce the chances of differing data between decades such as rivers, elevation, lights, and census data. They also added a covariate for each year called the distance-to-built layer so that the model could take changes in settlement into account as it performed its calculations. Multiple covariates were important and they had to be time-invariant so as to not affect how the time changed in between censuses.They then used the random tree model to predict population density (Gaughan et al., 2016). The researchers were then able to create a map of predicted population density. After that, the administrative units were disaggregated which allowed them to produce 3 different gridded dataset which show the predicted number of people per hectare. It shows an increase in population in urban areas, probably due to cheap labor and the necessity for jobs in urban areas as well as a general trend of rural to urban migration. The second geospatial data method that was used was complex network analysis, which is a topological relationship between nodes and nodes. They used data from Tencent as mentioned before due accuracy and large dataset to pull data from. The researchers then constructed a bidirectional matrix and then found the net inflow from the inflow and outflows. They used network analysis indicators such as in-degree and out-degree, which is used to show the mobility through cities (Pan and Lai, 2019). The in-degree shows how attractive a city is and is based on human weight. Centrality is another indicator, and it shows the influence and control of a city in correlation with other cities near it. It is also used to find the shortest path between node cities and how frequently they are used. The higher the centrality, the more important the city. The last indicator is the cluster coefficient, which shows the connection of cities. If cities are close enough together, they can interact and form a complex community. From the results of this study, the researchers were able to map out major migration routes and find cities with the highest net inflow and outflow population as well as compare different results involving centrality and other indicators and factors (Pan and Lai, 2019). The results show that almost all of the cities that had the highest population flow were cities that had a well-developed economy and high levels of administration, meaning they were sophisticated. There were also various studies that I looked into read in order to look for inspiration and help out my research further. One such study was where transportation and migration was a part of an overall study on “production-life ecology”, and the researchers took that as part of an assessment of a province in China (Zong et al., 2018). Although it was not that related to the information that I wanted to find, it helped in that it showed me that transportation was linked with many other things and it was not a standalone topic. There were multiple studies that were conducted during the week of the Lunar New Year, leading me to believe that the “golden week” before the actual Chinese New Year is an interesting and a good place to conduct a study as migration seems to spike during that time period. Most of the studies that I studied seem to conclude and confirm common trends that were to be expected, such as the study about inflow and outflow of cities where the large urban cities had the highest amount of flow. Some of the methods were researched for their background information while others were searched for specific data on the geospatial data science method that was conducted.

A gap in the literature that I found was the lack of subgroups and the variety of subgroups used in studies. The closest I could find to the use of subgroups was in a study where they wanted to compare the migration of different kinds of people in China, ranging from tourists, which were people who stayed in a city for less than 10 days, to permanent residents, who stayed in the city all the time (Wu et al., 2020). I thought that the use of subgroups was interesting as it allowed for an even more specific stream of information and narrowed the focus of the study down. However, there weren’t many studies conducted on subgroup differences that I could find as most just wanted to compare the scale of migration of the Chinese people in general as opposed to specific groups of people. However, I believe that different and specific subgroups of people are important to conduct studies on as different and specific trends can be witnessed. If one were to consider the socioeconomic status of subgroups, is it possible to look into the trend of impoverished or low-income earners moving to urban areas to confirm the theory of looking for better job opportunities? Or is it possible to see if a person who was earning little money saw an increase in salary when moving from a more rural to urban area? There are many different ways to categorize people, opening up the way for many different possibilities for studies that researchers would like to conduct. Looking into subgroups, I wanted to create a research question that looked into two subgroups, the people who migrated into a different city from an urban city and people who migrated to a city from a rural area. This allows researchers to look more into seeing if more people are coming from a rural area to urban cities or just from another city based on their socioeconomic status. The closest study I could find to this research study was a study conducted on rural flight. The study used the level of lights to classify grid cells as urban or rural, and then also incorporated census data in their study to estimate for rural flight. The researchers then compared the locations requests on social media of rural areas during Chinese New Year when all residents are expected to be at their hometowns with their relatives as compared to other times of the year where they would be working in another city or traveling somewhere else. From this study I concluded that using the Chinese New Year numbers would be very beneficial as everyone will be at their hometown for sure to honor the tradition of spending time with one’s relatives. Because of this, the numbers then can act as a benchmark when looking into rural flight. Overall, I considered and reviewed many different methods in an effort to gain a broader understanding on my topic as well as the data science methods that one could use to investigate studies related to my topic.

Research Proposal

For my research idea I was planning mixing some data science methods around in an effort to make obtaining results more accurate. The question that I have identified was looking into rural migration based on how where the person came from, whether it be from a rural area or from another urban area. A more specific look in the plan that I have is using a combination of census data and big data, as previous studies have shown that they can cover for each other's unreliable areas. We could then use complex network analysis to study the inflows and outflows of cities from specifically rural areas as well as specifically from urban areas, and using the data found in censuses conducted earlier to classify different urban or rural areas into wealthy and poorer areas, adding a socioeconomic factor into the study as well, which is something that studies from before have been lacking. This allows for a deeper understanding of where exactly the migration is starting from and offers researchers a pinpointed location to focus their studies. It also makes the reasoning for why this subgroup of people would want to migrate, as migration could mean that they want to look for better opportunities or they want to earn more money in a different place. There are multiple benefits to this study especially if you narrow the time period down to just Chinese New Year. Narrowing the time period down allows you to see the extent of migration throughout the week and also find out exactly where migration hits its peak and the population numbers in cities. From this, it is possible to find out around what time most people travel back home and increase the availability and accommodation of different methods of transportation, reduce the time needed to travel and also reduce the cramped space during this rush to return to one’s hometown. Understanding migration in general would also be useful as it would play a big role in urban planning. If the migration trend has been on a downswing lately in the city, they would know not to keep a high level of urban expansion out as that would just lead to issues like urban sprawl, and instead focus resources on other things related to infrastructure. Likewise, if a city was experiencing a large increase in migration, they would know in advance and be able to improve accommodations and prep living areas in advance as well as adjust the accessibility and frequency of public transportation and properly adjust to expanding the urban area. Knowing migration patterns and trends would also help everyday transportation as well as transportation that is used a lot during the "return to hometown period". The purpose of this inquiry is to gain information that can eventually be used in other studies and furthered, as it could provide a benchmark for studies that would be conducted later on. One the major obstacles would be the limitations of big data. Big data is very accurate and fast as it is used in real time and it is also updated frequently while also being capable of obtaining a large amount of data almost instantly, but it has its own weaknesses. Although census data can cover big data to some extent by using it to look at salary earned, using big data is also not super accurate as some people do not have access to a phone. This is especially important because as my study plans on looking into rural to urban migration and rural areas, those without a phone could be presumed to come from rural areas which could potentially mess with the data as some citizens may not be wealthy enough to use a phone and may be left out by the rural to urban migration study. Another method would be the accuracy of classifying socioeconomic status based on the urban or rural areas that were taken into consideration, as a blanket assumption cannot be done since not everyone in that rural area will be rich or poor and there will be variation, so just using a generalized "poor area" or "rich area" would be a lot less accurate. Because of this, a way to specify and make the distinctions clear would be needed. This study would be beneficial to

China as a whole, as it would possibly lead to general improvements on infrastructure and allow for better urban planning and more preparedness on the topic of migration, especially when it becomes time for Chinese citizens to start their annual migration back to their hometowns for Chinese New Year. This way of thinking can be applied to the whole world as migration occurs everywhere and in cities like New York where the city itself is also cramped and there is no not much space to maneuver, similar urban planning steps can be taken to help with that issue and deal with small-scale migration. Not many countries have the same level of large-scale migration from other parts of the country like China does but the same steps could be applied in case such a thing ever occurs in the future.

I think that my topic is one of the more important topics to consider because of just how large-scale my topic as a whole. Although my focus is on China, these studies and the methods used can be taken and applied anywhere around the world, instead of just having to keep things in China. No matter what city or country we live in, migration on any scale is bound to occur. Compared to epidemics or other diseases which primarily occur in one area and are restrained to that area, migration is everywhere, from third-world countries to booming wealthy powerhouse countries. Diseases such as malaria are not common anywhere in developed countries and are prevalent often only in Africa, which is only a small part of the world and does not encompass any other areas of the world. Although people may not necessarily die from problems with migration not being affected, many lives will still be negatively affected and that could eventually chain with other problems to eventually become a much larger issue that governments would then have to deal with. Not only is my topic that encompasses the entire world, it also is very affordable and would not take much resources to conduct. It is primarily focused on the process of exploring and learning more as opposed to finding a direct solution because the realm of migration is in a half-explored state. Although progress has been made through previous studies, there are still many different angles that haven’t been discovered and that have yet to be studied. Although some may argue that because there is a lack of a real solution, this plan and topic are not as important as other proposals, however, I recognize that I am not completely skilled and well equipped enough to conduct a more specific study to the fullest, so using this exploratory method allows me to put in max effort without failure, while also paving the way for future studies to use and improve the data that I have found through this research proposal, adding multiple opinions and studies to eventually create a firm concrete base of results and ideas.

Throughout the year, there would be two studies conducted. One would be conducted during the week leading up to Chinese New Year, as according to previous studies and Chinese traditions, people would be returning hope so that week would be the most accurate in collecting information to get an understanding of what the rural populations are like. Tencent would be ideal due to how often it is used and the success it has had in previous studies. A complex network analysis can be used for this portion of the study. Once that data is taken into account, a daily observation of migration from city to city and from rural and urban areas can be taken while taking the socioeconomic status of rural and urban areas can also be recorded. This is where census data would come in handy as it is hard for big data to calculate and predict how much money someone makes and classify them as having a high socioeconomic status or high economic status based on that. It would not cost that much money to push this research plan through as most of it can be done for free with minimal expenses, it would just be a long process as daily observations would have to be taken, a method to simplify things would be to just take it weekly, as instead of 365 observations there would just be 52 observations instead. Money might have to be put into manpower, as although getting the data may not cost much, people would still be needed to analyze the data and judge whether an area of the country is to be deemed poor or not. Overall, I think that this is a very feasible plan and it would require a minimum amount of money while accomplishing a lot, and helping start a new exploratory path on the topic migration that could influence many other studies in the future.

Works Cited

Gaughan, A. E., Stevens, F. R., Huang, Z., Nieves, J. J., Sorichetta, A., Lai, S., … Tatem, A. J. (2016, February 16). Spatiotemporal patterns of population in mainland China, 1990 to 2010. Retrieved from https://www.nature.com/articles/sdata20165

Pan, J., & Lai, J. (2019, June 5). Spatial pattern of population mobility among cities in China: Case study of the National Day plus Mid-Autumn Festival based on Tencent migration data. Retrieved from https://www.sciencedirect.com/science/article/pii/S0264275118311703

Deng, Y., Qi, W., Fu, B., & Wang, K. (2019, July 26). Geographical transformations of urban sprawl: Exploring the spatial heterogeneity across cities in China 1992–2015. Retrieved from https://www.sciencedirect.com/science/article/pii/S0264275119300307?via=ihub

Ma, T., Lu, R., Zhao, N., & Shaw, S.-L. (2018). An estimate of rural exodus in China using location-aware data. *PLoS ONE*, *13*(8), 1–14. https://doi.org/10.1371/journal.pone.0201458

Fan, C. C. (2005). Interprovincial Migration, Population Redistribution, and Regional Development in China: 1990 and 2000 Census Comparisons. *Professional Geographer*, *57*(2), 295–311. https://doi.org/10.1111/j.0033-0124.2005.00479.x

Hu, M. (2019). Visualizing the largest annual human migration during the Spring Festival travel season in China. *Environment and Planning A: Economy and Space*, *51*(8), 1618–1621. https://doi.org/10.1177/0308518X19845908

Wu, Y., Wang, L., Fan, L., Yang, M., Zhang, Y., & Feng, Y. (2020, March 5). Comparison of the spatiotemporal mobility patterns among typical subgroups of the actual population with mobile phone data: A case study of Beijing.

https://www.sciencedirect.com/science/article/pii/S0264275119316725

Zhang, W., Chong, Z., Li, X., & Nie, G. (2020, February 13). Spatial patterns and determinant factors of population flow networks in China: Analysis on Tencent Location Big Data. https://www.sciencedirect.com/science/article/pii/S0264275119311862

Zong, W., Cheng, L., Xia, N., Jiang, P., Wei, X., Zhang, F., … Li, M. (2018, August 22). New technical framework for assessing the spatial pattern of land development in Yunnan Province, China: A "production-life-ecology" perspective.

https://www.sciencedirect.com/science/article/pii/S0197397518302674

Zhang, G., Zheng, D., Wu, H., Wang, J., & Li, S. (2019, November 21). Assessing the role of high-speed rail in shaping the spatial patterns of urban and rural development: A case of the Middle Reaches of the Yangtze River, China.

https://www.sciencedirect.com/science/article/pii/S0048969719353926