SPEARS SCHOOL OF BUSINESS

MANAGEMENT SCIENCE & INFORMATION SYSTEMS

MSIS 5673 – DESCRIPTIVE ANALYTICS AND VISUALIZATION

EFFECT OF POPULATION ON LITERACY & LEVEL OF EDUCATION IN INDIA

SPRING 2018

ISHAN MALPOTRA

A20104861

ishan.malpotra@okstate.edu

Introduction

India, officially known as The Republic of India is a country in South Asia. It is the seventh-largest country by area, the second most populous country (with over 1.2 billion people) and the most populous democracy in the world. India is a federal republic governed under a parliamentary system and consists of 29 states and 7 union territories. India is a fastest developing country in the world. India, with its diversified culture, civilization, natural resources, technology and huge skilled human resources, is also a fastest growing economy in the world. But at the same time, there are several problems plaguing the modern India which is affecting its growth and development. We will focus on some of the problems faced in the top 500 cities of India such as overpopulation, no. of graduates, literates, and literacy rates, etc. using our descriptive visualizations techniques. The data is taken from 2011 census of India. The census data includes those cities with a population greater than 100,000 in India.

Exploratory Analysis

I have had a wonderful opportunity to travel almost every state of India in the past 25 years majorly because my father works for the Indian Air Force and he gets transferred to a new city after every 3 years. I studied in 7 different schools in 7 different cities from 7 different states during my childhood. My parents were always curious to know about which state is the best in terms of literacy rate and which state stands out of all the other states in terms of most graduates and many other factors since they had to choose the best education for me. Unfortunately, I neither knew about the data visualization tools and techniques in the past nor did I find any relevant dataset to answer my parent's questions. When I started searching for a dataset for the project for this class, I found it quite interesting when I imagined how helpful it would be if those questions could be answered with various visualizations instead of just simple text. Hence, when I came across a dataset containing various details about the top 500 cities of India, I decided to do my project my project on the same.

MSIS 5673 – DESCRIPTIVE ANALYTICS AND VISUALIZATION

This dataset contains almost all relevant and important information that can answer any question about the growth and major problems faced by the major cities of India. This dataset contains following attributes:

Header	Description
Name_of_city	Name of the city
State_code	State code for the respective cities
State_name	Name of the state
Dist_code	District code for the respective cities
Population_total	Total population of the city
Population_male	Total male population in the city
Population_female	Total female population in the city
0-6_population_total	Total population in the age group 0-6 years
0-6_population_male	Total male population in the age group 0-6 years
0-6_population_female	Total female population in the age group 0-6 years
Literates_total	Total number of literates in the city
Literates_male	Total number of male literates in the city
Literates_female	Total number of female literates in the city
Sex_ratio	Ratio of males to females in a population
Child_sex_ratio	Ratio of males to females in a population in the age group 0-6 years
Effective_literacy_rate_total	Total Effective literacy rate
Effective_literacy_rate_male	Male Effective literacy rate
Effective_literacy_rate_female	Female effective literacy rate

Location	Location of the city
Total_graduates	Total number of graduates in the city
Male_graduates	Total number of male graduates in the city
Female_graduates	Total number of female graduates in the city

This dataset was already cleaned when I got it from www.kaggle.com. Hence, I really didn't need to do any data massaging or cleaning. However, I had to convert the .csv file obtained from the source to .xlsx file. The data hardly had any missing values and it has more of a continuous data than the categorical data. Moreover, blending and combining additional data would definitely help in the better analysis if we have wanted to do a more detailed analysis. I am working only on this dataset only since I'm focused on only the factors mentioned in the data to create a general awareness amongst the people regarding the basic developing factors in the top cities of India.

Explanatory Analysis

The visualization could be of interest to

- Education officials who plan to bring the literacy rate to a better level in their respective cities and states.
- The government who is working hard for the development of the nation. These visualizations would help them a lot in understanding the major concerning areas of development.
- Parents who need to choose the best place to educate their kids based on the literacy level and the best place to live at based on other development factors.
- Students, traveler's and NGO organizations who are either curious to know about the current scenario or are working for the betterment of the country.

The audience of this presentation could be government officials whom I am indirectly related to as I've seen my father who works for the betterment of the country. I am working
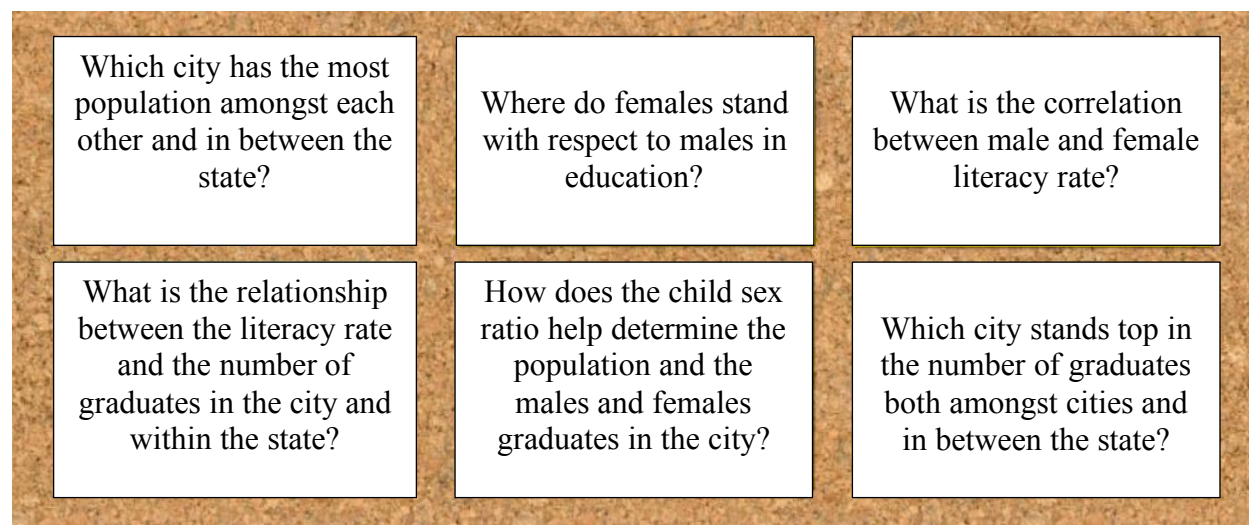
on creating visualizations for the same which would help them in an indirect way. As mentioned above, the audience can be education officials, travelers, tourists, and NGO organizations as well. I have never had any direct interaction with the education officials but I did struggle a bit before coming to OSU for my masters in choosing the perfect curriculum, location and other factors of a graduate college. I hope that my visualization would do good for some chunk of the education officials in improving the education status of their cities and their states.

Moreover, I have traveled a lot since my childhood as mentioned above and I'm pretty sure that my visualizations would help tourists and travelers a lot in getting a basic idea about the top cities of India.

Storyboarding

Various people need help with many different aspects such as the government officials need help regarding the factors involved with the development of the nation, education officials with improving their education system and improving the literacy rate in their city or state, parents in choosing the best city to live based on several factors such as sex ratio, literacy rate, population and several other factors. Tourists and NGO's would need guidance as well in getting to know the basic details of any new city or state they're visiting or working for a good cause.

Which city has the most population amongst each other and in between the state?	Where do females stand with respect to males in education?	What is the correlation between male and female literacy rate?
What is the relationship between the literacy rate and the number of graduates in the city and within the state?	How does the child sex ratio help determine the population and the males and females graduates in the city?	Which city stands top in the number of graduates both amongst cities and in between the state?

An analysis or visualization would serve as a boon that can guide all of the concerned people mentioned above. Given above is a storyboard which gives a more detailed explanation of the visual analysis performed on this dataset.

Final Thoughts

This visualization would help the government agencies, education officials, parents, travelers, and NGO organizations with all the details required for the betterment of the nation, education institutions and give basic details required for them. The relationship between the population of states/cities with the number of literates, number of graduates and the effective literacy rate is also seen using these visualizations.

References

A few of the references from where the dataset was obtained and the basic details were gathered are mentioned below:

- https://www.kaggle.com/zed9941/top-500-indian-cities/version/4
- https://www.kaggle.com/umeshnarayanappa/exploring-top-500-indian-cities?scriptVersionId=556575