STUDY OF GENDER DISTRIBUTION ACROSS JOBS IN NEW YORK CITY

Manushi Majumdar (mam1612@nyu.edu), Github handle - ManushiM

ABSTRACT

The project focuses on seeing the gender distribution across areas of New York cities, defined by zip codes, and focuses on seeing this distribution specific to employees belonging to different industries. I check for correlation between employee count of various industries and the ratio of men to women in each industry. Based on this, I then perform clustering to see what conclusions can be drawn from the data. The results obtained show a slight gender dominance in a few industries and few industries being very prominent in few neighborhoods of New York City.

INTRODUCTION

New York City is considered as the financial capital of the world. With a population of about 8.5 million, New York City has about 3.9 million employees as recorded in the year 2013 [1]. The same data also reports number of male employees (1969961) and number of female employees (1950301) to be almost at par. The question I wish to answer through this analysis revolves around the gender distribution of men and women across jobs in New York City, down to the zip code level. I attempt to see if there is any relation between the industry type and the gender of people across different areas across New York City. A paper 'Occupational gender segregation Trends and explanations', written by Jo Anne Preston, discusses different methods of measuring occupational segregation and the implications of each. It also presents the results of studies attempting to measure the degree of occupational segregation by gender over time and across nations. A third section of the paper focuses on some current explanations for occupational segregation including the feminization of specific occupations. Finally, it takes up the social and economic implications of occupational segregation such as its relationship to the lack of authority and the inferior rewards for women [2].

DATA

My data source was the LODES LEHD data for the year 2013[1] (most recent data available). From this dataset I used the following two files:

1. Workforce area Characteristics (WAC) for New York State 2013 for all employees

2. Geography Crosswalk File for New York State

The Workforce Area Characteristics file provides for total number of employees as well as number of employees based on age, income, race, gender and educational background of the employees per workplace census block code, and number of employees based on industry type, industry size (based on employee count) and age of the industry, for the same workplace census block code. The Geography Crosswalk file is available for New York State, which gives based on block code, the state, county, block group name, fips code, as well as other geographical details. Based on the counties of New York City (New York, Kings, Queens, Bronx and Richmond), I created a geographical crosswalk file for New York City from the crosswalk file of the state. I then merged this file with the WAC file for the entire state to

limit my workforce data to only New York City. Once this file was created, I aggregated on the blocks falling within a zip code, and calculated number of employees for industry and gender based on zip code. Based on this dataset, I then created a condensed dataset that would hold only total number of employees and employee count for each industry and each gender. I got the sum of each column (total number of employees per industry) and picked only those industries which had employees greater than 100,000. This filtering was based on the assumption that in a city like New York City, only those industries with comprising of a certain number of people should be used to observe gender distribution. Also certain industry types (Agriculture, for instance) may not obviously have a large number of workers in New York City. I then picked 14 out of 20 industries which satisfied that employee count limit. I then created a data frame which held the zip code and number of employees in those 14 industries and number of male and female employees. This was the dataframe that was used for the analysis.

Additionally, I used the Census Block data shapefile [3], as well as a New York City Zip code geojsonfile [4] to visualize the results over the map of New York City.

METHODOLOGY

I ran a correlation on the ratio of number of male employees to number of female employees, to the number of employees per industry. Based on the correlation coefficients and the scatter matrix generated, I got a rough estimate of industries where I could expect a certain gender dominance. To ascertain this assumption, I performed a KMeans Clustering on the ratio of male to female employees and number of people in all the industries to see the relation I obtain from this. A linear regression would have given a better understanding of the effect of the ratio to genders to number of people in each industry, but it would have also not been the most feasible tool to apply across 14 industry types.

Results obtained are as follows:

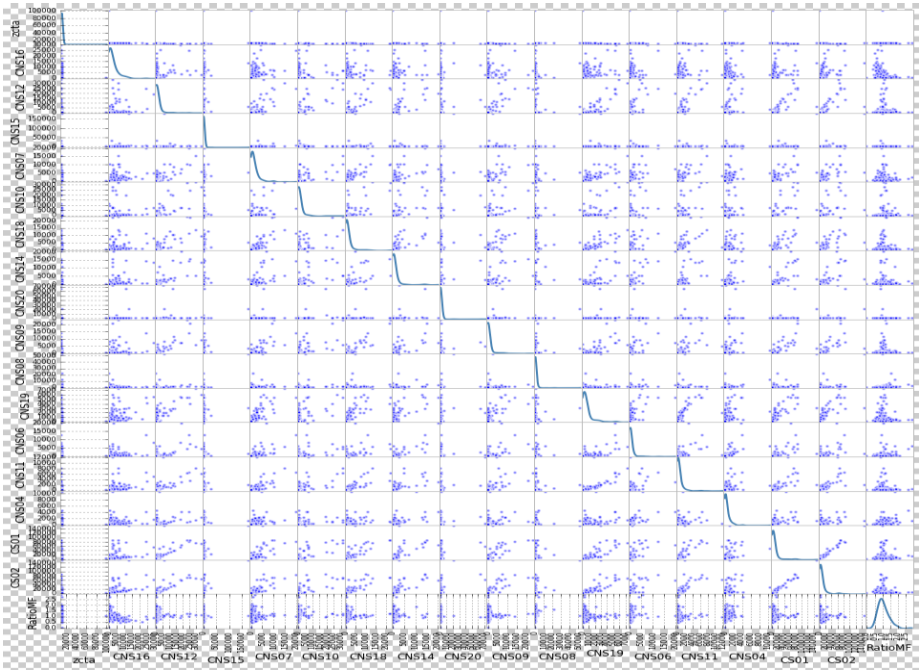

Fig 1. Scatter matrix of correlation between all variables

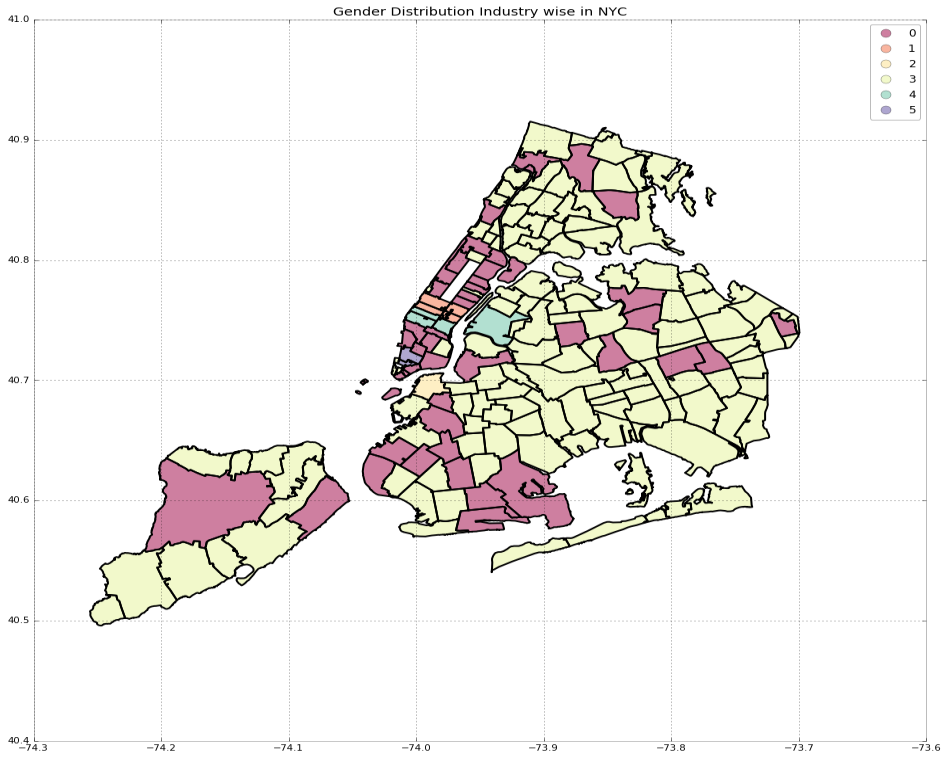

Fig 2. Gender Distribution Industry wise obtained after clustering

CONCLUSION

The results obtained show a higher number of women than men in the Health Care Industry, and a good number of women in Finance, especially around Midtown Manhattan. Men seem to be concentrated in industries such as Finance and Insurance, Professional, Scientific and Technical services, Educational Services, as well as the Health Care and Social Assistance. Downtown Brooklyn seems to be focused more on Educational Services. Parts of Downtown Manhattan seem to be centered on Professional, Scientific and Technical services, as well as Health Care and social Assistance. I desired to check results for the gender-industry distribution down to the block level in the city, but that yielded a result as follows, which would need further analysis and refined plotting techniques. I plan to further this study, by further testing a granular visualization and analysis of this study. And also seeing if there are other factors, such as age or race, for instance that affect this gender-industry distribution.

Figure 3. Gender-Industry distribution on Census Block Shapefile (Further analysis needed)

BIBLIOGRAPHY

1. LEHD Data - http://lehd.ces.census.gov/data/lodes/LODES7/ny/wac/

2. Preston, Jo Anne. "Occupational gender segregation trends and explanations." *The Quarterly Review of Economics and Finance* 39.5 (1999): 611-624.

3. Census Block Data - ftp://ftp2.census.gov/geo/tiger/TIGER2013/TABBLOCK/

4. NYC Zip Codes Shape File - http://data.nycprepared.org/dataset/nyc-zip-code-tabulation-areas/resource/0c0e14e9-78e1-404e-97b0-c2fabceb3981