Applied Data Science Capstone
Final Project

COFFEE CULTURE: A COMPARISON OF GLOBAL METROPOLES

Oytun Altasli Widmer, 24 July 2020

INTRODUCTION: Coffee Culture

o In the last 20 years, coffee shops have spread in almost all cities becoming ever-integrated in our culture

    o Cafes have emerged as places where business meetings take place

    o With the availability of wireless internet, it has evolved into work place and study hall for many people

    o Our consumption preferences and habits have been shaped by the ubiquitous coffee shop

IDEA: How do metropoles compare?

o Cafes are everywhere, but there bound to be differences across global metropoles in terms of prevalence and penetration of chains

o Intensity or number of coffee houses in city centres can point to cultural and economic activity of cities

o It is also interesting and useful to see where coffee chains dominate

AUDIENCE: Investors and managers

o The outcome of this project should be interesting for entrepreneurs who would like to open a new venue in city centre

o It should also be useful for coffee chain managers who can utilise it to optimise brand presence as well as explore new geographies

o Finally, I think it is a fun information for coffee-loving travellers.

DATA: FourSquare

o FourSquare venue search function was utilised for this project

o After experimenting with different search items such as coffee, café or both in my code, I settled on 'coffee' for the search

o There is no single search item that is perfect for this project as there seems to be categorical variations by city. Additionally, a search for 'café' yielded more than 50 venues for most city centres (even with a reduced 200m radius). As 50 venues is the limit set by FourSquare, this meant 50 venues for most cities which made comparison across cities impossible
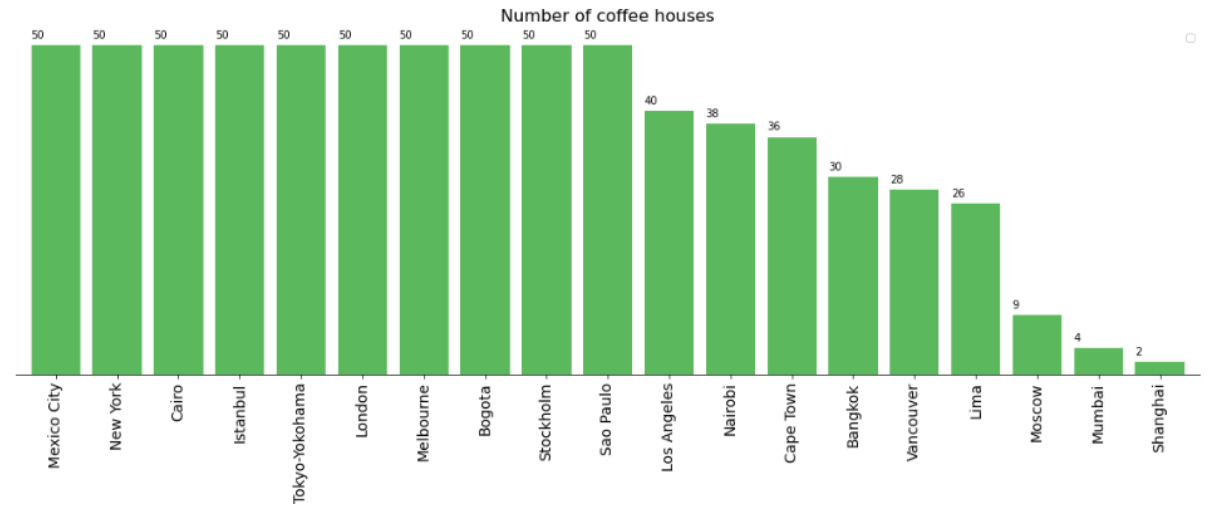
DATA: FourSquare

○ A comparison of search for 'coffee' and 'café' with FourSquare API

'Coffee' search
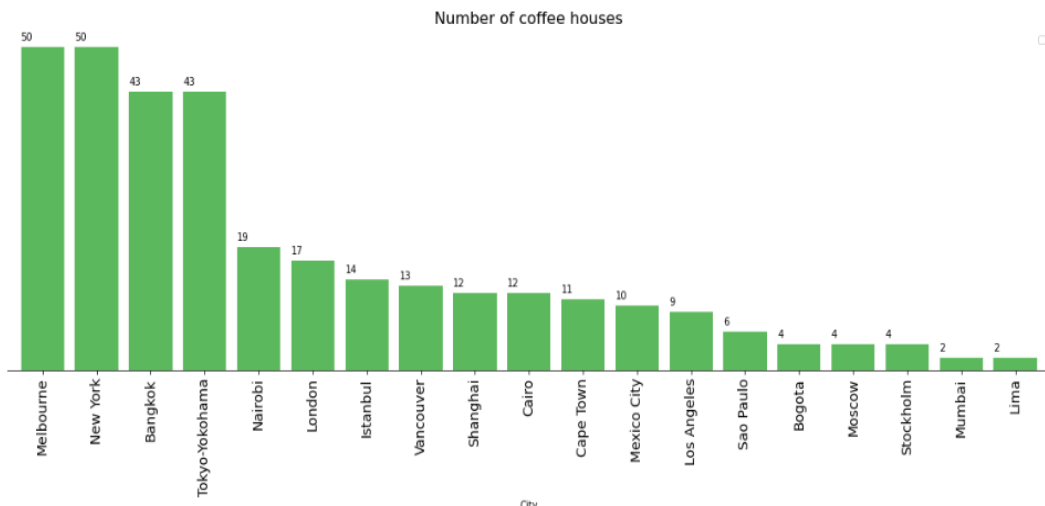
'Cafe' search

DATA: Geolocator and Databank

o For location data, I used Geolocator

o I gathered additional city density data from London Datastore

o After quality-checking data for individual cities, my final list of cities included 20 metropoles globally

o The main dataset was based on 20 cities, 'coffee' as search item in a 500m radius in city centre

o There was limited data cleansing required
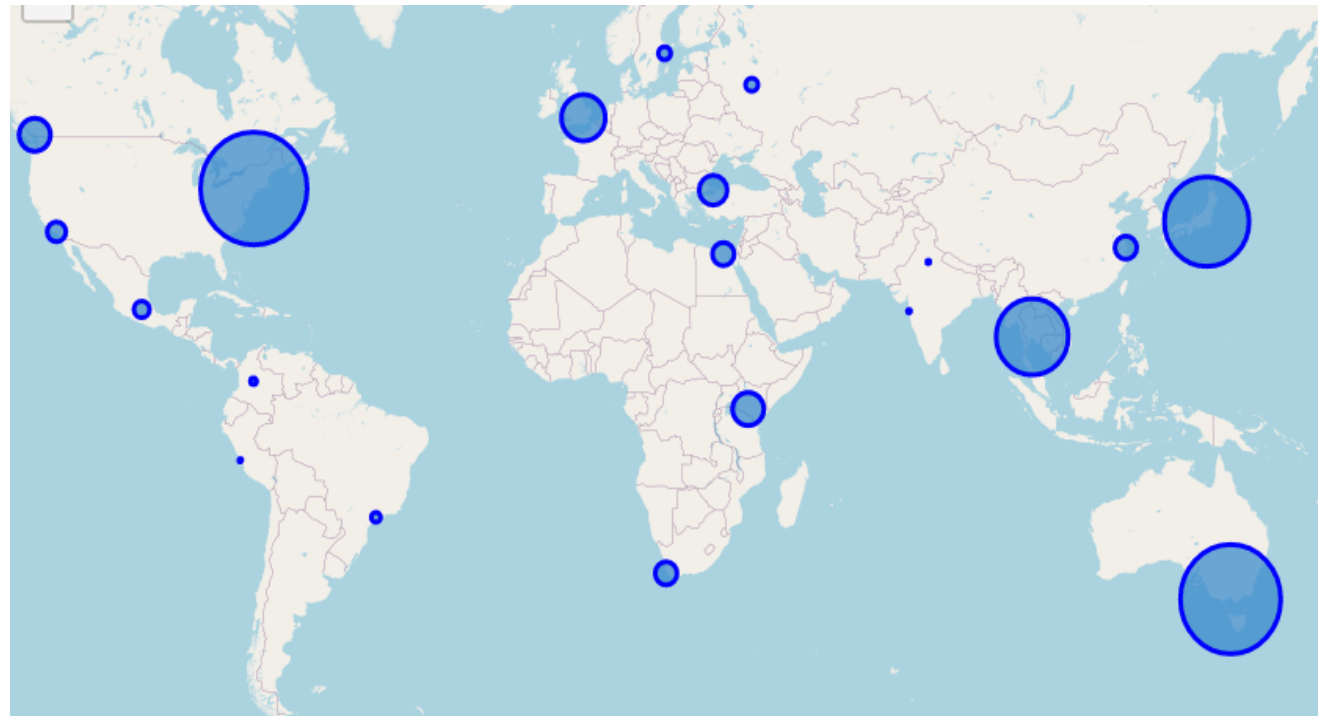
METHODOLOGY: Maps and charts

o After extracting and cleansing city data, I utilised

- o Bubble maps to see how cities compare in terms of coffee house penetration, with and without adjustments for population density in cities
- o Bar charts to visualise how cities rank in terms of number of coffee houses in their centre
- o Histogram to see frequency of number of coffee houses in a 500m city centre
- o Bar charts to see how cities rank in terms of penetration of coffee houses in city centres
- o Scatter plots to see the relation of population density vs. number of coffee houses
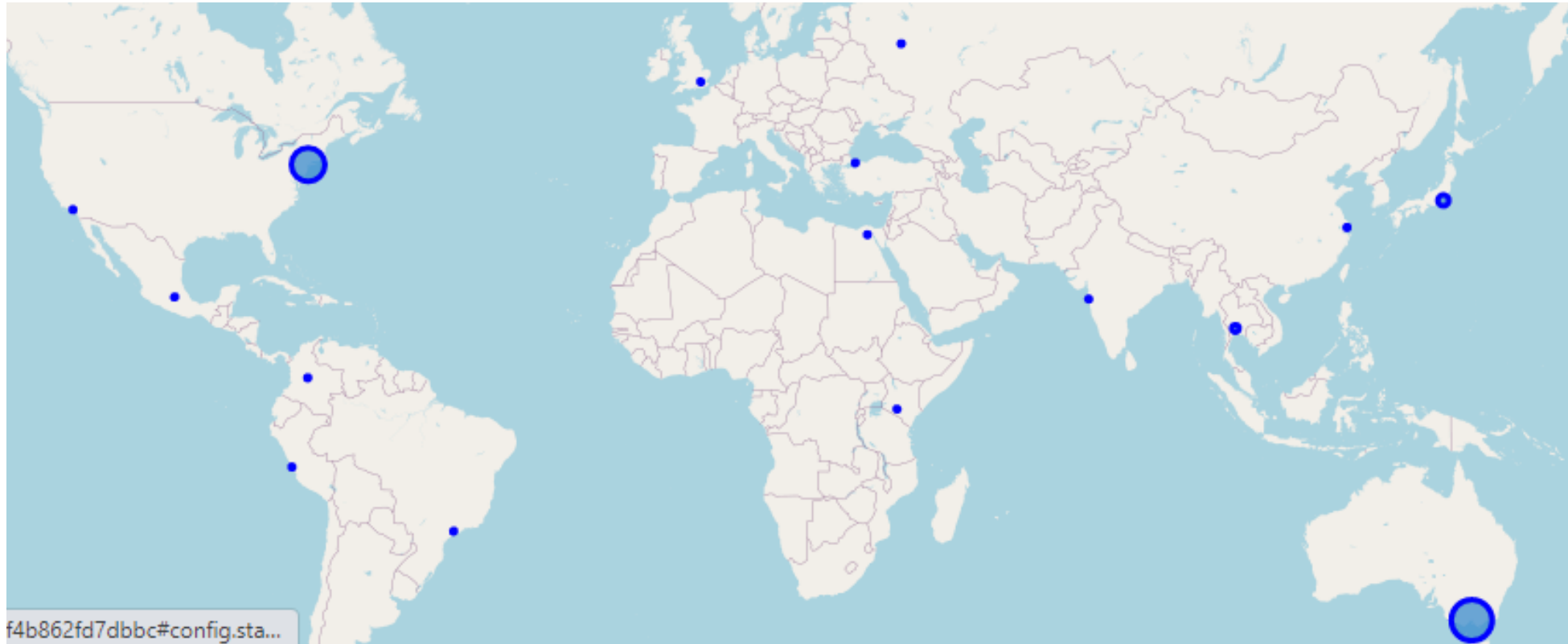
METHODOLOGY: Bubble charts

○ For a visual overview, I utilised Folium Maps with bubbles. Bubble sizes show the relative number of coffee houses in each city
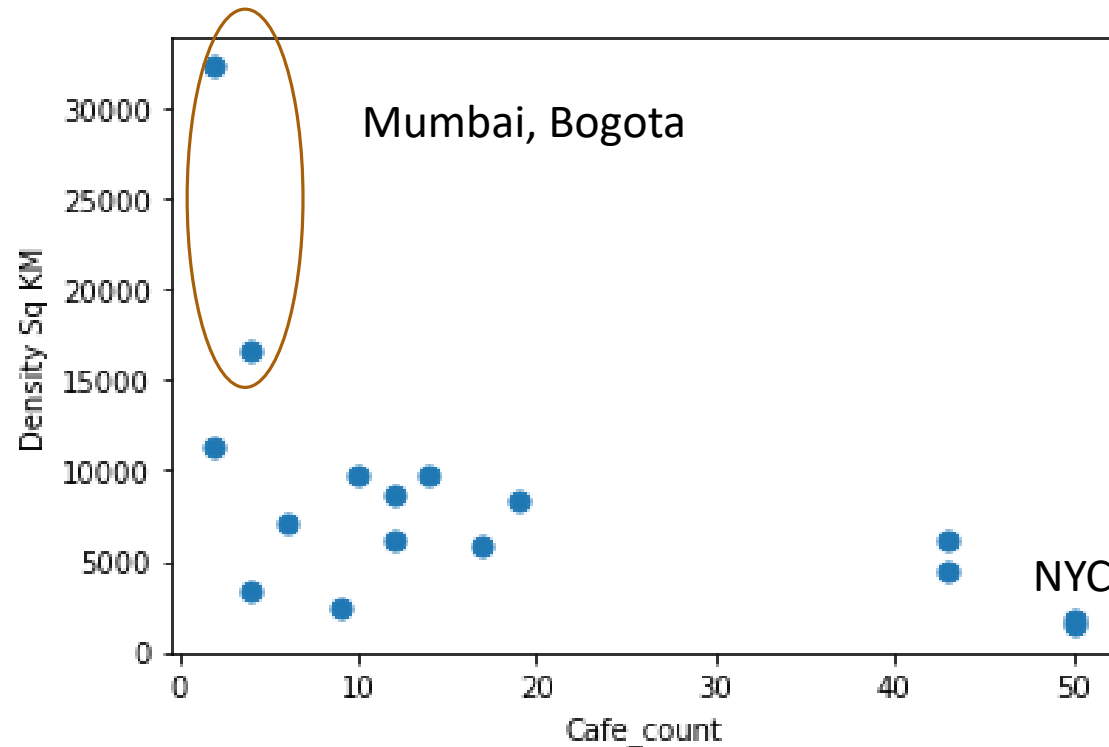
METHODOLOGY: Bubble charts

o After adjusting for population density, differences are less significant

METHODOLOGY: Correlation

o Excluded for outliers, the number of coffee shops and population density seem to be positively correlated

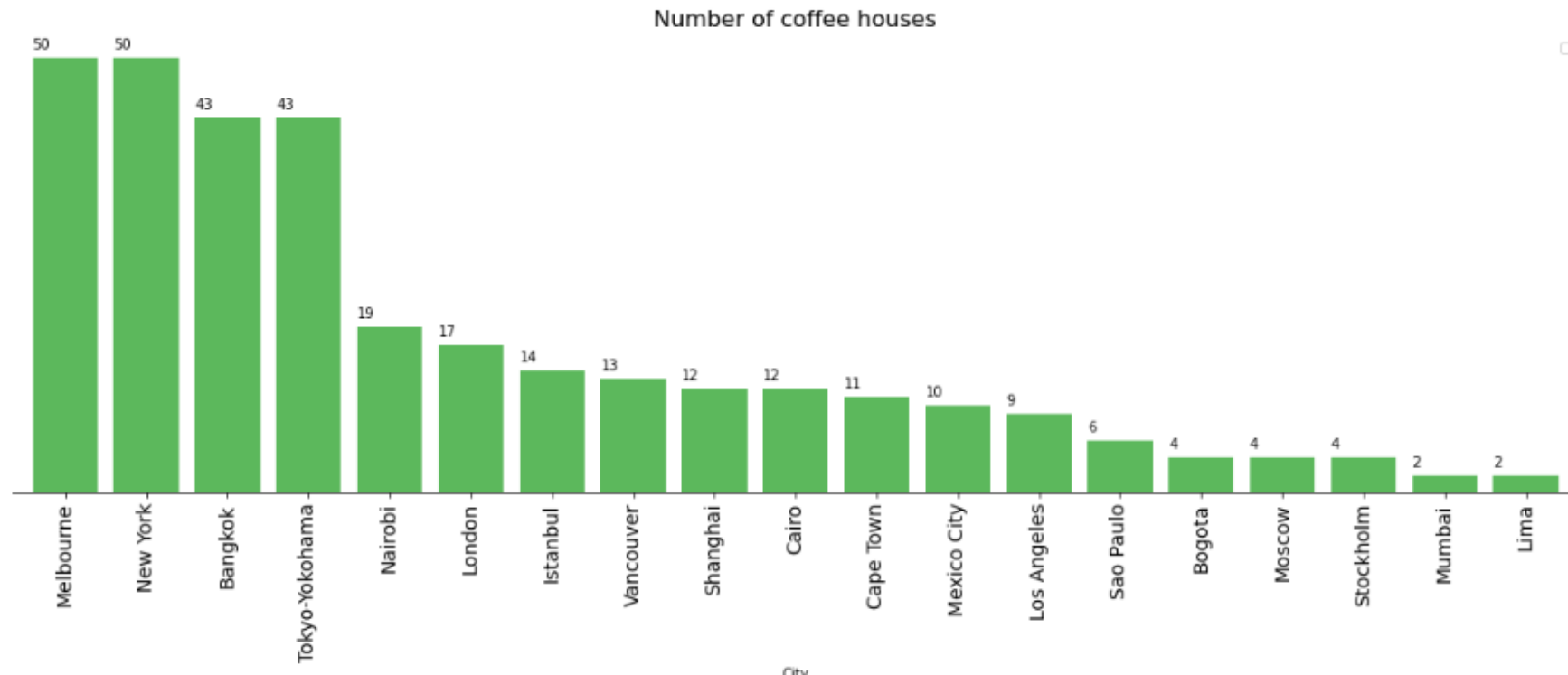Mumbai, Bogota

NYC

Density Sq KM

Cafe_count

- The outliers are Mumbai and Bogota which has suspiciously low number of venues, which we discuss in more detail later. New York's density is also underestimated as it includes a bigger metropolitan area
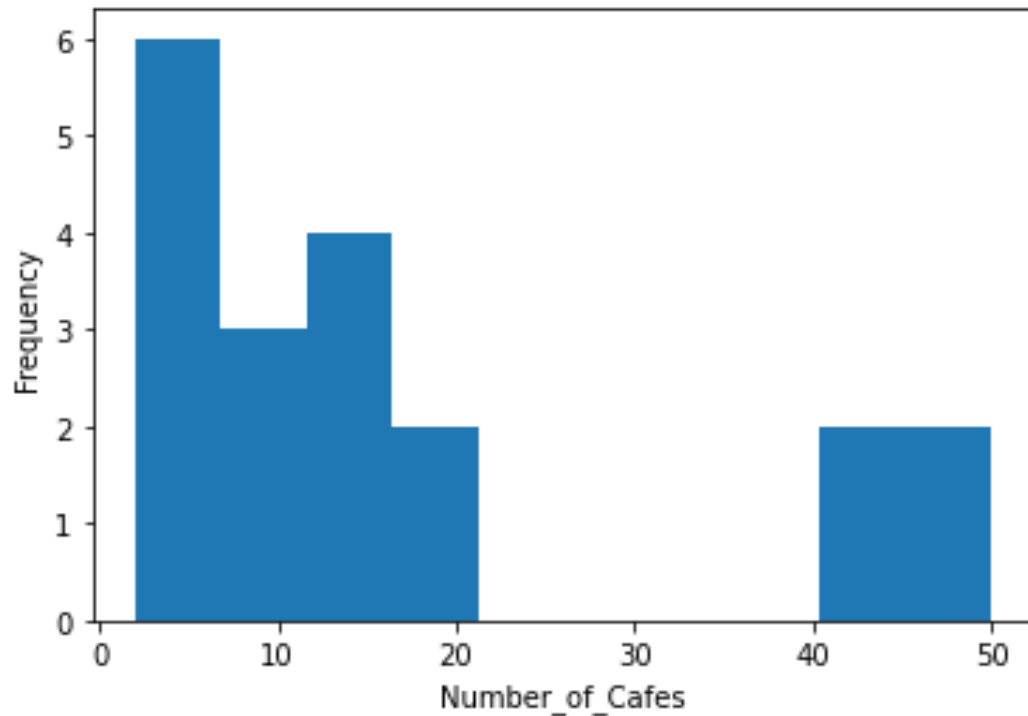
METHODOLOGY: Rankings

o Melbourne and New York come at the top of list in terms of coffee shops
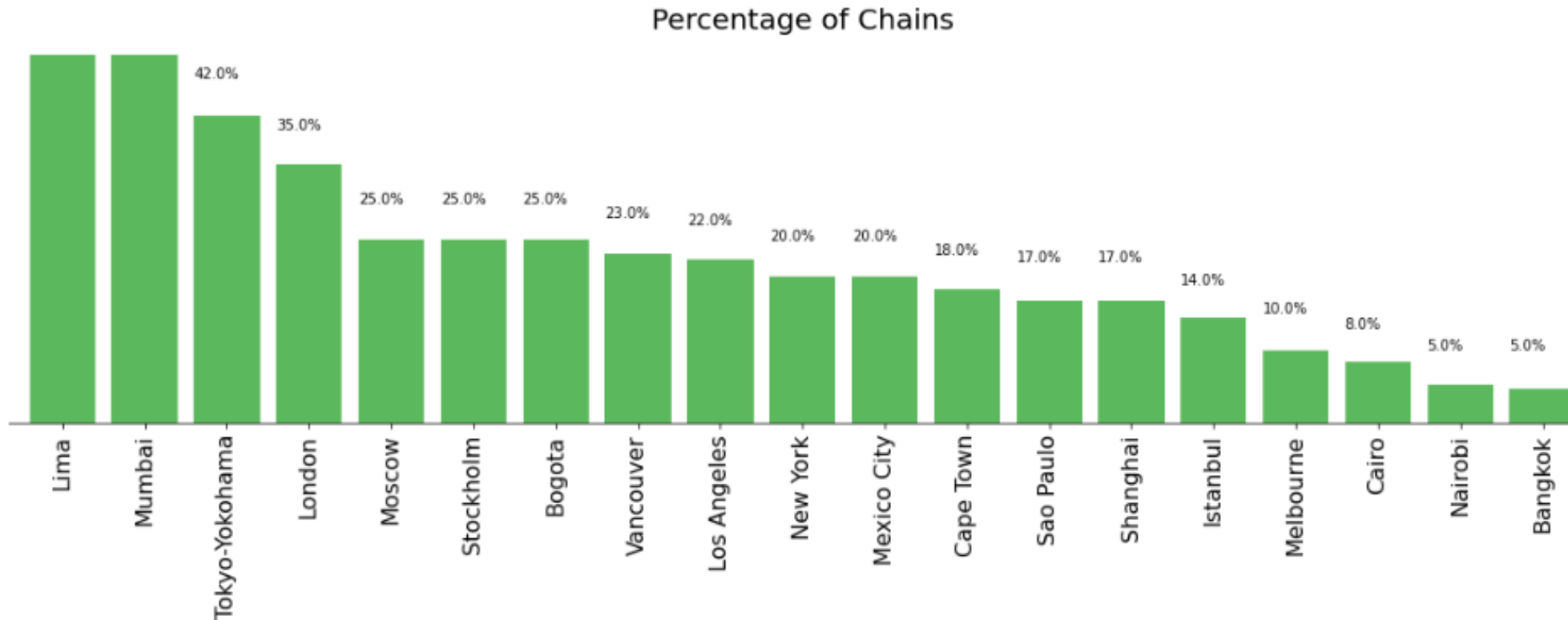

Number of coffee houses

METHODOLOGY: Frequency

o According to histogram, many cities have less than ten coffee shops in their city centres. The imperfections of the data however, call for more detailed analysis

• New York has artificially high number of venues as 'corner carts' are included in FourSquare, while a search for 'café' would yield more options in many cities

METHODOLOGY: Rankings

o Excluding Mumbai and Lima, which has very few coffee houses, Tokyo and London come at the top when it comes to coffee chain penetration

Percentage of Chains

RESULTS: Melbourne, chains, and Indians

o As expected, main financial centres such as New York and London have a large number of coffee shops in their centres

o Melbourne's top ranking was rather surprising given its 5m population. It seems to be home to a vibrant independent coffee culture

o Tokyo and London has the highest penetration of chains with 42% and 35%, respectively. London is dominated by Costa Coffee at its centre. Tokyo is served by a larger number of coffee chains. Moscow, Stockholm and Bogota also have relatively strong chain presence although the latter's figures are skewed by a very limited number of coffee venues

o Indian cities has a very limited number of coffee shops in their centres, especially considering their populations. I think this requires additional investigation into data

DISCUSSION: A fragmented market

o Based on this project, coffee market as defined by coffee shops is a very fragmented market. Although most city centres are penetrated by coffee houses, independents dominate, and there is further room for expansion by chains

o However, there are exceptions like Tokyo and London where chain share in centres have reached its limits

o FourSquare data proved practical for analysis of chains, and the project can be extended on this direction

o On the other hand, data was insufficient to compare cities on the basis of coffee cultures which we tried to identify by the number of coffee shops

CONCLUSION: Future directions

o In the West, we are living in a coffee culture where cafes have become an integral part of our daily lives. In the beginning of this project, my goal was to find differences across cultures and continents as well as explore the dominance of coffee chains globally

o Given the type of FourSquare data, it is hard to draw solid conclusions regarding consumer preferences in different cities

o However, there is room for further analysis on coffee chains. The project can be expanded to include more cities in the regions. It is also possible to branch out the project to analyse brand dominance in different cities