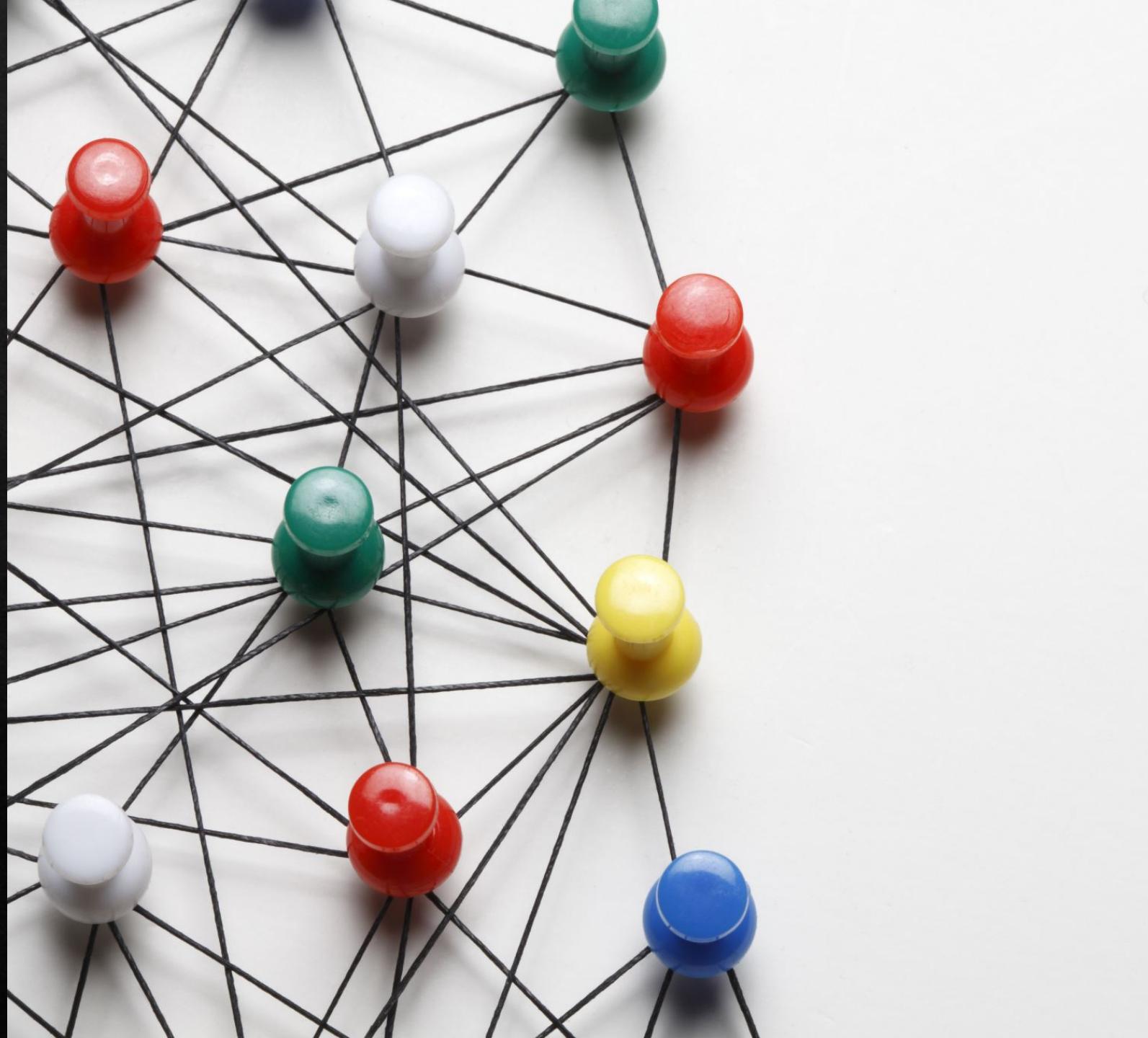


Capstone Project - The Battle of Neighborhoods

Location selection

Based on

Data Science



Objective

- ❖ Leverage the Foursquare location data to explore or compare neighbourhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.
- ❖ The submission should include:
 - ❖ Week 1:
 - ❖ A description of the problem and a discussion of the background. (**15 marks**)
 - ❖ A description of the data and how it will be used to solve the problem. (**15 marks**)
 - ❖ Week 2:
 - ❖ A link to your Notebook on your Github repository, showing your code. (**15 marks**)
 - ❖ A full report consisting of all of the following components (**15 marks**):
 - ❖ Introduction where you discuss the business problem and who would be interested in this project.
 - ❖ Data where you describe the data that will be used to solve the problem and the source of the data.
 - ❖ Methodology section which represents the main component of the report
 - ❖ Results section where you discuss the results.
 - ❖ Discussion section.
 - ❖ Conclusion section where you conclude the report.
 - ❖ Your choice of a presentation or blogpost. (**10 marks**)

Introduction

- ❖ For this assignment it is assumed that we advice a company in their decision process of choosing a location for their Europe headquarter. The company has pre-selected three cities in which a local office already exists:
- ❖ [London \(GBP\)](#)
- ❖ [Zurich \(CHE\)](#)
- ❖ [Barcelona \(SPA\)](#)

Each of these cities are attractive places to live and work. All are very close to international airports and are appreciated by Expats. The organisation wants to settle his new headquarter based on the following decision matrix:

Nb.	Attribute	Details	Weight
1.	Labour climate	labour cost, productivity, skills availability	6
2.	Political environment	Effectivness of goverment, policy consistency, corruption	7
3.	Access to Universities	Nb. of Univerities and reputation, ranking, size	5
4.	Quality of life	standard of living, recreation, health	4
5.	Labour and skill abundance	Skill levle, unemployment rate, labour unions, wage rate	1
6.	Cost of labour	Productivity, exchange rate,	2
7.	Tax structure	Corp. tax rate, social sec. cost, rgulatory barriers	3

Data Sources

The sources of data acquired for this assignment are

- ❖ well known
- ❖ regarded NGO institutions
- ❖ Government sites
- ❖ data collection platforms e.g. Foursquare

Name of organization	Data access / Report	Data attributes	File format
Foursquare	API	Id, name, contact, location, categories, verified, stats, url, hours, popular, menu, price, rating, hereNow, storeId, description	API (json)
OECD	Unemployment	Annual, harmonised unemployment rate in %, over 10 years	API (csv)
OECD	Unit labour cost	Annual, Unit labour costs and labour productivity (employment based), Total economy in %, over 10years	API (csv)
OECD	Corporate Tax	Statutory corporate income tax rate, Total economy in %, over 10years	API (csv)
OECD	Labour wage	Average annual wages, Total economy in EUR, over 10years	API (csv)
Transparency International	CPI Data Set	CPI rank, CPI score, CPI std. error	XLS
Opendata.swiss	Swiss city index	Administration, Construction and housing, Crime, criminal justice, Culture, media, information society, sport, National economy, Education and science, Energy, Finances, Geography, Mobility and Transport, Public order and security, Politics, Population, Prices, Social security, Health, Territory and environment, Tourism, Work and income	API (json)
Stadt Zurich	District information	Geospatial data as polygon lat, long coordinates	Json
Statista	Average price of residential property	Average prices of 120 square meter apartments located in the most important cities of 38 European countries	Web scraping

Methodology

Two step approach:

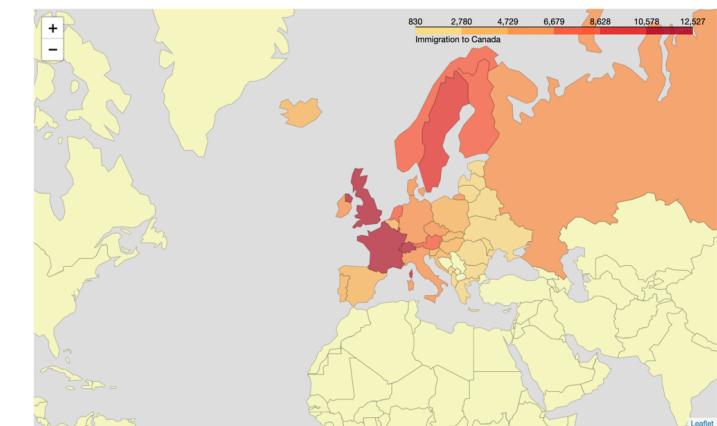
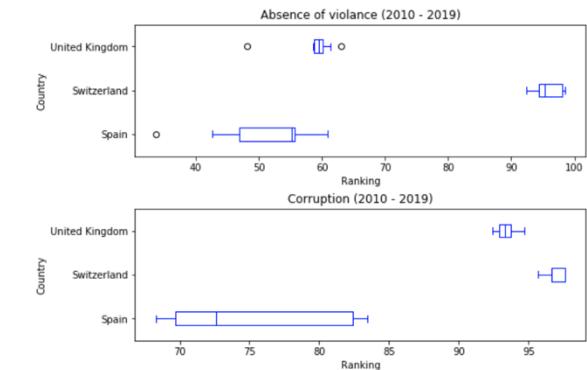
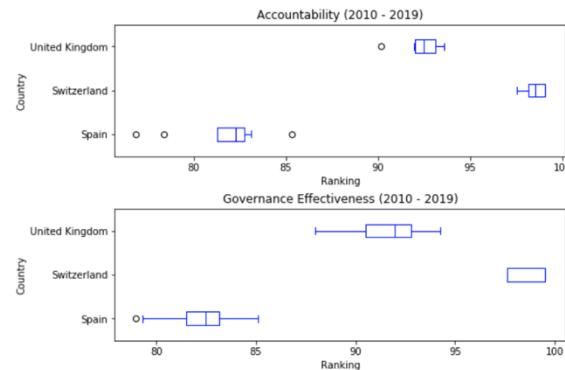
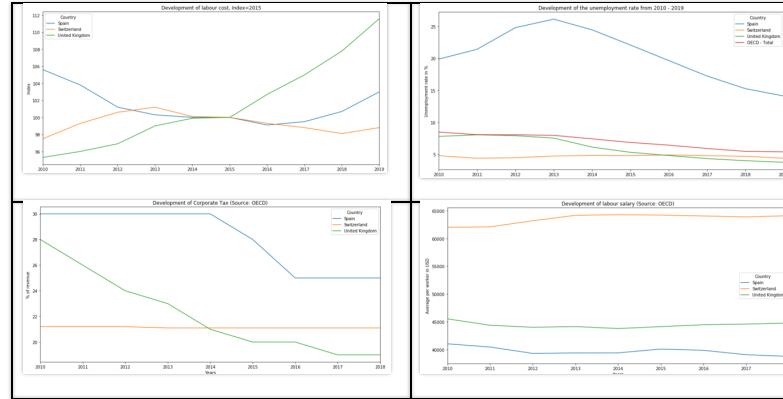
- ❖ Select a country among the three candidates (SPA, CHE, GBR) based on macro economical data as shown above and run a decision tree analysis on the subject to define the key selection points i.e. the high entropy gains.
- ❖ Based on the preselection, and with the data of mainly Foursquare, the different boroughs were analysed for the suit to the companies need and GenY compatibility. For this a clustering and prediction model was deployed on the data

Data and Observations

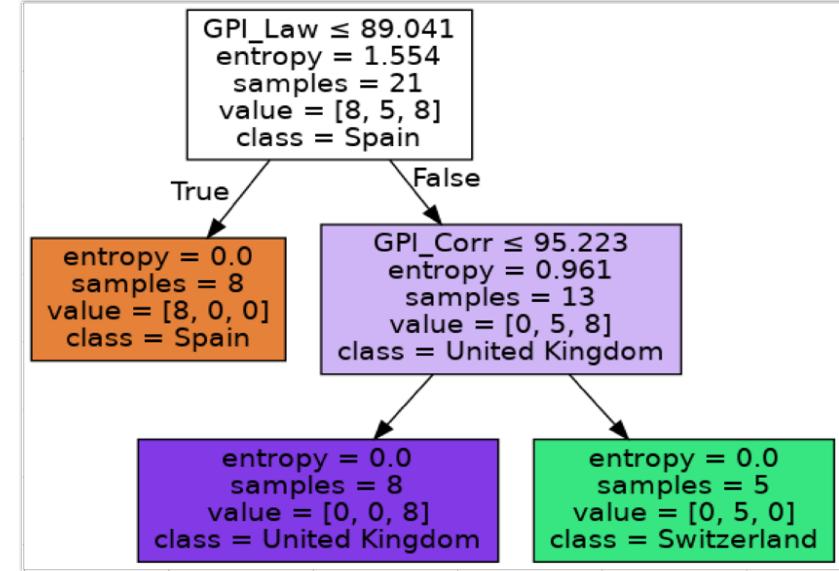
Cost view

Regulatory view

Macro Economic



Country Selection



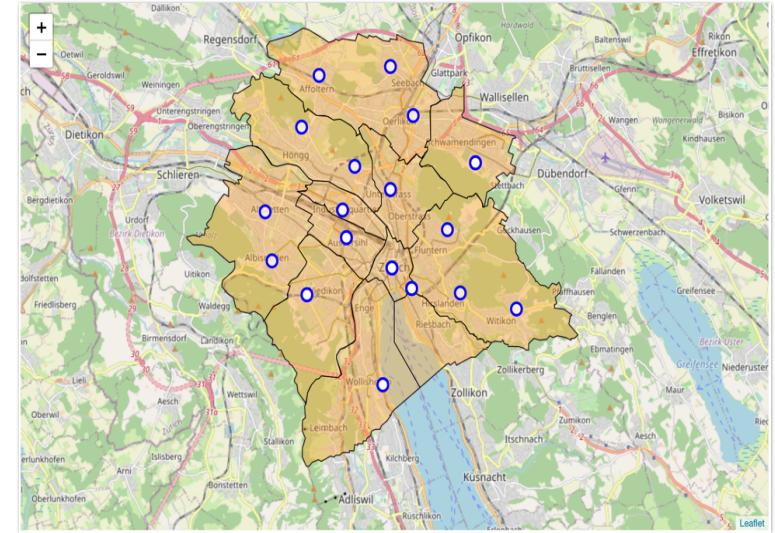
Purpose of the above data review was to identify the data sets providing the highest information gains on the nodes with the least variables to look at.

The model indicates that only two variables:

1. Index for “Rule of law”
2. Corruption reliance index

Finetuning the selection

Zurich as an international
City is subdivided into 11
boroughs



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
17	Fluntern	Tram Station	Hotel Bar	Fondue Restaurant	Hotel	Cupcake Shop	Deli / Bodega	Department Store	Design Studio	Dessert Shop	Food & Drink Shop
31	Saatlen	Tram Station	Swiss Restaurant	Pizza Place	Athletics & Sports	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Eastern European Restaurant
32	Schwamendingen Mitte	Tram Station	Swiss Restaurant	Pizza Place	Athletics & Sports	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Eastern European Restaurant
33	Hirzenbach	Tram Station	Swiss Restaurant	Pizza Place	Athletics & Sports	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Eastern European Restaurant

We'd recommend to locate the new Europe headquarter in
Zurich in the borough of Fluntern/ZH

Discussion of the results / method

- ❖ This work was highly hypothetical and hence might not hold all data points , a company would consider.
- ❖ It can however lead the process as a template.
- ❖ The challenge in this work was less based on the statistical validity of the analysis but primarily on the demonstration of the following skills:
 - ❖ Finding, collecting harmonizing data from multiple sources
 - ❖ Building a data framework which allows the deployment of machine learning techniques, despite the small sample of data
 - ❖ Application of multiple data visualization techniques, i.e. boxplot, superimposed maps and decision trees
 - ❖ Handling of at least two machine learning techniques
 - ❖ Formulation of report demonstrating the skills