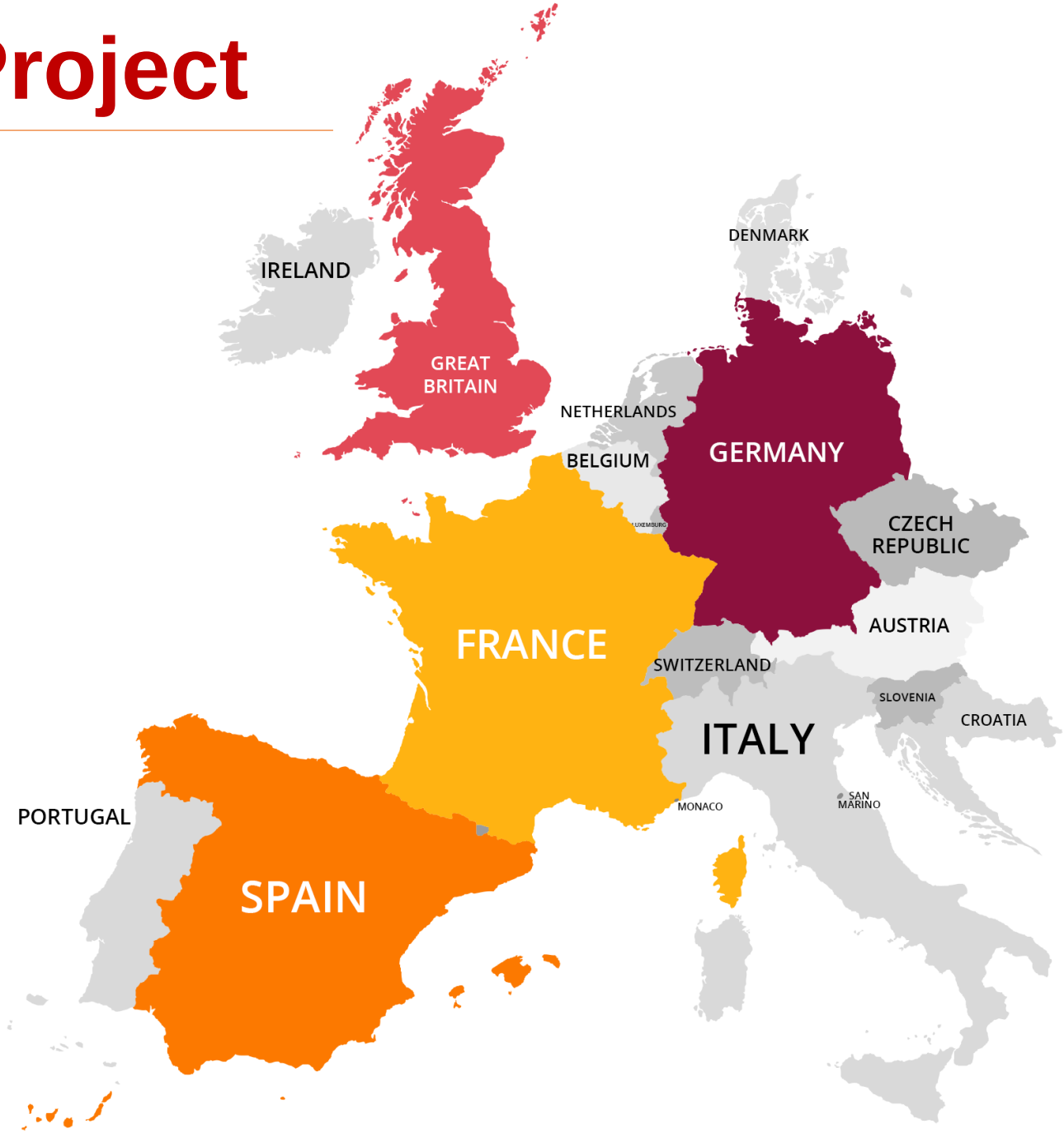
Coursera Capstone Project

IBM Data Science



Alin Pirvan

April 2020

Introduction

- Coffee is the most popular drink worldwide with around two billion cups consumed every day.
- In the UK, we now drink approximately 95 million cups of coffee per day. The coffee industry creates over 210,000 UK jobs. The Gross Value-Added contribution from the UK coffee industry to the economy is estimated to be £9.1 billion.
- In this project, we will atempt to use FourSquare and K-Means clustering to find the optimal location for opening a new cafe.

Areas of Glasgow

- Glasgow has a population density of 3,400 people per square kilometer, which makes it the most densely populated city in Scotland. The larger Greater Glasgow area has an estimated population of 1.2 million.
- This represents about 42% of the population of Scotland.

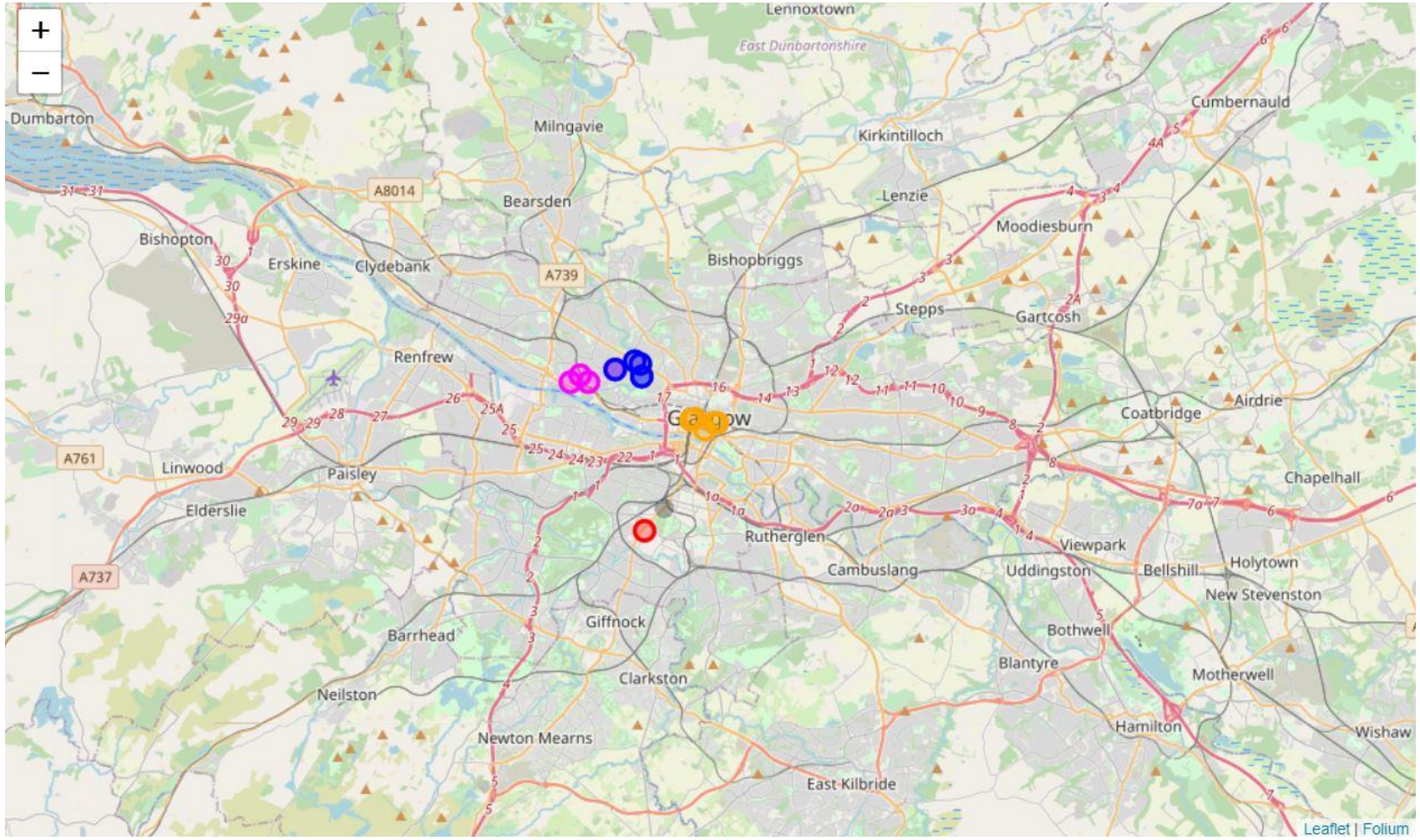
	District	Population	Area (km²)	Density (/km²)	District_Coord	Latitude	Longitude
0	Govanhill	9,725	0.86	11,308	(55.8363741, -4.2581531)	55.836374	-4.258153
1	Pollokshields	9,738	1.59	6,125	(55.8422663, -4.2849973)	55.842266	-4.284997
2	Partick	8,884	0.85	10,452	(55.8699211, -4.3094365)	55.869921	-4.309437
3	Hillhead	6,275	0.96	6,536	(55.8752091, -4.293281)	55.875209	-4.293281
4	Govan	5,860	1.63	3,595	(55.860879, -4.3185273)	55.860879	-4.318527
5	Gorbals	6,030	0.83	7,265	(55.851813, -4.2531625)	55.851813	-4.253163
6	Shawlands	7,015	0.52	13,490	(55.8292301, -4.2924584)	55.829230	-4.292458

Data Analysis

- FourSquare API will be used to acquire information on the most popular venues of Glasgow, the most frequented venues and use all this in combination with K-Means clustering of neighborhoods to determine the optimal location for a coffee shop.

Map of Glasgow

- We used FourSquare API to create a map of Glasgow and superimpose the venues as markers, using folium.
- We can then create a dataframe of all venues in Glasgow and find the most frequented ones.
- Finally, we can use this data to establish which district would have our required customer basis.



Finding the best location

Gorbals

	Venue	Freq
0	Pub	0.07
1	Coffee Shop	0.07
2	Seafood Restaurant	0.05
3	Bar	0.05
4	Restaurant	0.04

Govanhill

	Venue	Freq
0	Bar	0.15
1	Indian Restaurant	0.11
2	Fast Food Restaurant	0.07
3	Train Station	0.07
4	Supermarket	0.04

Langside

	Venue	Freq
0	Grocery Store	0.12
1	Italian Restaurant	0.12
2	Café	0.12
3	Restaurant	0.08
4	Pizza Place	0.08

Pollokshields

	Venue	Freq
0	Park	0.14
1	Hotel Bar	0.07
2	Roller Rink	0.07
3	Café	0.07
4	Supermarket	0.07

Govan

	Venue	Freq
0	Discount Store	0.15
1	Museum	0.15
2	Gas Station	0.15
3	Grocery Store	0.08
4	Metro Station	0.08

Hillhead

	Venue	Freq
0	Café	0.13
1	Bar	0.08
2	Pub	0.07
3	Coffee Shop	0.05
4	Indian Restaurant	0.05

Partick

	Venue	Freq
0	Café	0.14
1	Pub	0.07
2	Bar	0.07
3	Coffee Shop	0.05
4	Restaurant	0.04

Shawlands

	Venue	Freq
0	Café	0.20
1	Pub	0.10
2	Supermarket	0.10
3	Italian Restaurant	0.07
4	Grocery Store	0.07

- We can see that cafes are the most popular venues in Govan, Partick and Hillhead.
- Using this information, we can move forward and use K-Mean clustering algorithm to find our preferred location.

K-Means Clustering

- We use the K-Means algorithm to determine the best location for opening a new coffee shop.
- Partick and Hillhead are both good districts with great student traffic, many local attractions and a good rent value.