**Data**

The following factors were referenced in the Introduction – Business Problem section:

1. What places (cities) consume the most coffee?
2. What are the population, GDP, and GDP per capita in these cities?
3. How saturated is the coffee shop market in these cities?

I will now explain the motivation and data sources for these items.

The first – coffee consumption – is obviously something in which our stakeholders would take interest. Specifically, we will use the list provided in the following site for an idea of how popular coffee is (how much is drunk on a per capita basis) in different countries:

<https://www.caffeineinformer.com/caffeine-what-the-world-drinks>

This is reported at the country level. As we would like to limit our scope to capital cities, we will make the assumption that the per capita consumption provided at the country level is a good reflection of the per capita consumption in the capital city. So, for example, the data source tells us that **Serbs** consume 5.4kg of dry weight (before brewing process) coffee per capita per year. We will thus assume that coffee consumption is at the same level in the capital city of **Belgrade**.

The second factor – macroeconomic statistics – will provide us with a metric to help us gauge how wealthy the population in a given country is. Again, we will apply the GDP per capita findings for each country to its capital city. This data will be obtained from the World Bank and can be viewed at the following page:

<https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita>

It’s important to note that this is reported in PPP (Purchasing Power Parity) so it already accounts for the differences in value of a dollar in different countries. Population data will be obtained from:

<https://en.wikipedia.org/wiki/List_of_national_capitals_by_population>

Finally, I will use Foursquare to count the number of cafes within a 5 kilometer radius of the center of the capital city. This – in conjunction with population data – will provide us with a population-per-café (PPC) number. Then, we will take this number and multiply it by coffee-consumed-per-capita to understand how much coffee can be attributed to each café. Again, we will use **Belgrade** as an example: The per-capita-consumption is **5.4kg**. The population of Belgrade is about **1.2 million**. Once we find the number of cafes (Café Count) in a 5 kilometer radius from the center of Belgrade, we will compute:

Coffee Potential for new shop (CPNS) = (5.4 kg \* 1,200,000 / Café Count)

As costs per cup of coffee will certainly vary by country, we will use GDP per capita ratios between countries to adjust for this. And again, these CPNS numbers may be orders of magnitude too high or too low based on the accuracy of location reporting. So we will likely end up relying on the relative CPNS values to determine the city with the best opportunity.

Of particular importance to our business is to consider the relationship between GDP per capita and on-premise coffee consumption. How does the economic situation in a city influence the way in which the local population consumes coffee? Unfortunately we don’t have on vs off-premise coffee consumption data for all of the countries in our data. But we do have *beer* consumption data for the 50 US States:

<https://www.brewersassociation.org/insights/importance-on-premise-craft-brewers/>

<http://scottjanish.com/interactive-map-total-beer-consumption-state/>

And we do have the on/off premise consumption of coffee split in the US in the year 2004:

<https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?locations=US>

This will all come in handy. More on this in Methodology!