When deciding on a location for a new business, we must consider how much of the surrounding market we can realistically expect to capture. While it is obviously beneficial to be located near a large population of people, this alone is not sufficient. We must also consider the presence of competitors, as a portion of the market will inevitably be drawn to existing businesses instead of ours. Given information about local population density and the locations of current competitors, we can attempt to define an optimal location or compare several candidate locations to determine which is most favourable.

To model this, consider a single person choosing between businesses of a particular type, such as car repair shops. We use a gravity-style model to describe, with some uncertainty, which business they are likely to choose. If the person is choosing between shop A located 1 km away and shop B located 3 km away, it is reasonable to assume they are more likely to visit the closer option. In this model, each business is assigned a "draw" that is inversely proportional to distance. Thus, the draw to shop A is $1/1 = 1$, while the draw to shop B is $1/3 \approx 0.33$. The probability that the person chooses a particular shop is then taken to be proportional to that shop's draw relative to the total draw from all available options.

More rigorously we have:

$$\begin{aligned}&A\ person\ with\ location\ \mathrm{P_1}\\&Two\ businesses\ with\ locations\ \mathrm{A_1, A_2}\\&\text{We have } \mathrm{P(P_1 \text{ chooses } A_1)} = \frac{\frac{1}{\|\mathrm{P_1\text{-}A_1}\|}}{\frac{1}{\|\mathrm{P_1\text{-}A_1}\|} + \frac{1}{\|\mathrm{P_1\text{-}A_2}\|}}\end{aligned}$$

For many existing locations and people considered we have:

$$\begin{aligned}&n\ people\ with\ locations\ \mathrm{P_1, P_2, \dots, P_n}\\&m\ businesses\ with\ locations\ \mathrm{A_1, A_2, \dots, A_m}\\&\mathrm{P(P_i \text{ chooses } A_j)} = \frac{\frac{1}{\|\mathrm{P_i\text{-}A_j}\|}}{\sum_{k=0}^{m} \frac{1}{\|\mathrm{P_i\text{-}A_k}\|}}\end{aligned}$$

We can define that last bit as $E_{i,j}$ as it tells us the expectation of customer i going to location j. If we want to estimate the total market size, then we need to take the sum of these expectations over all the people in the local area.

$$\text{Market size for location j} = \sum_{i=0}^{n} \mathrm{E_{i,j}} = \sum_{i=0}^{n} \left( \frac{\frac{1}{\|\mathrm{P_i\text{-}A_j}\|}}{\sum_{k=0}^{m} \frac{1}{\|\mathrm{P_i\text{-}A_k}\|}} \right)$$

In this program we can extract existing businesses from Overpass API and we can roughly estimate the location of where people live from ABS mesh-blocks data which gives us quite precise local population data (accurate to within roughly a city block). We can easily calculate the estimated market size and compare this value over a variety of locations to determine the best. Some examples are shown in location.ipynb.