**Data Science Capstone Project Report**

Holly Tibble

**Background**

Over the limited warm summer months in the Scottish city of Edinburgh, it is popular for locals to spend much of their day in the public parks – socialising, drinks and barbequing. Having recently procured a license for a mobile coffee shop, I am trying to identify the best spots in the city as a vendor.

As I will be operating out of a vehicle, I will not be able to provide customer toilets. As such, I would like to find spots with nearby public (or publicly accessible) toilets (maximum range one kilometre). My license allows me to park my cart on the edge of public parks, but I prefer to be nearby to park play areas as I think my target audience is probably parents and families, rather than young adults who are perhaps more likely to be consuming alcohol.

**Data**

The data I will be using to find suggested locations for my coffee cart will be the locations of public play areas, public toilets, and other local shops.

The play areas data was extracted from data.edinburghopendata.info, and was most recently updated in August 2017. The data contains the following variables:

* Site: the name of the play area (string)
* Play facilities: a description of the play area (string)
* Address: the street address of the play area (string)
* Postcode: the postcode of the play area (string)
* Telephone: the telephone number of the local council office responsible for the management of the play area (numeric)
* Location: the coordinates of the play area (numeric tuple)

The public toilets data is also from data.edinburghopendata.info, and was updated in July 2017. The data contains the following variables:

* Toilet: the name of the toilet (string)
* Type of facility: either public toilets, or the type of establishment if publicly accessible, such as ‘Library’ (string)
* Refurbished in 2013: a note of whether or not this toilet (public only) was refurbished in 2013 (string)
* Charge: the cost of using the toilet, where applicable (string)
* Refurbishment status: unknown, missing for all entries
* Facilities: description of the available facilities, such as baby changing and disabled access toilets (string)
* Opening times: the days and times for which the toilet is available for use (string)
* Telephone: the telephone number of the council office or establishment responsible for the upkeep of the facilities (numeric)
* Email: an email address for the council office or establishment responsible for the upkeep of the facilities (string)
* Website: a website for the establishment responsible for the upkeep of the facilities (string)
* Address: the street address of the toilet (string)
* Location: the coordinates of the toilet (numeric tuple)

Finally, foursquare data was used to identify nearby establishments, within a radius of 1km of the search query coordinates. A limit of 50 responses was applied, and the name, coordinates, and category of each venue was extracted from the json file returned by the query.