LOCAL OPENSTREETMAP ANALYSIS SUMMARY

Objective	Analyse and Clean OpenStreetMap data from my local area and then perform a data query.
Description	Data Analyst Nanodegree assignment. I chose to work with MongoDB because of the flexibility it offers compared to SQL or SQLLite to evolve the database as new needs or new data emerges.
Data	Source: OpenStreetMap, Area: Cardiff newport bristol bath, England Link: mapzen.com/data/metro-extracts/#cardiff-newport-bristol-bath-england Size: 535MB Format: GeoJSON
Key Facts	2489219 **nodes** (single point in space i.e restaurant) 317867 **ways** (ordered list of nodes forming a line or closed polygon i.e park) 2255 **relations** (relationships between nodes or ways i.e bridge, street) 1992 users

New Data Structure

```
{"id": "660774542",
"created": {"user": "DaveF",
        "timestamp": "2014-07-10T12:47:37Z",
        "changeset": "24063039",
        "version": "4",
        "uid":"115894"},
"pos": [51.510174, -2.1964616], "type": "node", "id": "104779"}
"address": {"country": "GB",
        "housenumber": "13",
        "city": "Cardiff",
        "street": "Pethybridge Road"},
"amenity": "restaurant",
"type": "node",
"name": "Golf Clubhouse",
"node_refs": ["2946322217", "2946322224", "2946322207", "2946322206"]}
```

Data Cleaning Plan

Data Field	Issues	Fixes
Address	Normalising street names	"Crescent,": "Crescent","Cresent": "Crescent","HIll": "Hill", "Rd": "Road", "St": "Street", "avenue": "Avenue","hill":"Hill", "kingsway":"Kingsway","lane":"Lane","road":"Road","st":"Street"
Postcode	Capitalisation Data entry	"Bs": "BS", "Cf": "CF","bs": "BS", "cf": "CF", "ta": "TA" Manual change to osm file: - found postcode online
Amenity	Unrecognisable Capitalisation Spacing	Change entry based on name field Change all to lower case Change all spaces to underscores
City	Capitals City names Misspellings Hyphens	Change all to sentence case, consistency "Weston-Super-Mare" Remove all ', Bristol' and ', Cardiff' for consistency "Cardiff, Wales":"Cardiff", Remove all
FIXME	What is it?	Notes questioning entries. Delete 965 entries until confirmed

Data Query

Couple, aged 40, recently married, christian, baby due, dog owners. Moving to the area. Is it suitable?

veterinary count in the BA area : 0
school count in the BA area : 3
community_centre count in the BA area : 0
doctors count in the BA area : 1
Total ameneties in the BA area : 824

It's clear that amenity data is incomplete and the query can't be answered in a meaningful way.

Next Steps

Ordnance Survey Points of Interest? There is a fee and may be T&C's issues as well as teething problems converting the OS geocode to the Open Street Map node_refs or long/lat positional data.

Bath Hacked - local council data released and analysed by a local group. Relevant data could be imported into the current database to provided a meaningful answer the data query.