ABH Internship report

Data Analytics: Month 2

Shapefiles of OSM data for Arizona is imported in postgis using QGIS. Only shape files of Arizona POIs are used.

**ST\_GeomFromText**: creates geometry out of WKT represented object

**ST\_Buffer**: creates a geometry/geography that represents all points whose distance from this Geometry/geography is less than or equal to distance (if SRID is 4326, unit is degrees)

**ST\_Transform**: Returns a new geometry with its coordinates transformed to a different spatial reference system

**ST\_MakePoint**: makes point out two variables, for example lat and long

**ST\_Intersection:** returns intersection (point, polygon…) from two geometries

**ST\_IsValid:** Test if an ST\_Geometry value is well formed. For geometries that are invalid, the PostgreSQL NOTICE will provide details of why it is not valid

**Data cleaning:** In OSM data, benches (3700), public toilets (500), waste-baskets (2100), fountains(364), atm (136) are deleted

-Outlier: poi in Las Vegas (but state is AZ)

Intersection of pois\_points and yelp buffered (11 m) points and osm name is not null: 3217 records

Intersection of pois\_points and yelp buffered (111 m) points and osm name is not null: 57272 records

Intersection of pois\_points and yelp buffered (11 m) points and osm name is not null and osm name= yelp name: 588 records

Intersection of pois\_points and buffered (111 m) yelp points and lower(name) are equal: 1460 records

-problem with categories in business Arizona table: 32458 distinct categories, cannot be manually standardized