# OSM Amenities1 Data Quality Report

## Overview

This report will outline the initial data quality findings on amenities data obtained from Overpass Turbo API which can be found at https://overpass-turbo.eu/. This report will include an overview of the dataset, and a review of the continuous and categorical features, including histograms and bar charts. On initial review, this dataset contains a lot of missing data for most features. The data that is present appears to be reasonable and logical, however a number of columns will need to be dropped. 42 features contain missing values.

## Summary

This dataset consists of information on different Amenities in New York City, including bars, restaurants, and cafes. The dataset has 46 features and 15055 rows. 36 features will have to be dropped due to missing values. There are no duplicate rows. Distribution of the data is consistent with expectations.

## Review Logical Integrity

Test 1: No date in x is before Feb 2022 or after the upload date (22nd May).

* 0 instances.

## Review Continuous Features

There are 3 continuous features in this dataset:

* id
* This feature has a mean of 6328565742.082298, a min value of 349323821.0 and a max value of 11943712730.0.
* lat
* This feature has a mean of 40.72131367367652, a min value of 40.5097277 and a max value of 40.9067767.
* lon
* This feature has a mean of -73.95595388388575, a min value of -74.2513558 and a max value of -73.7002573.