

# Atlan | Data Science Intern Challenge

**Aim:** Identify commercial centers using Points of Interest (POI) data

There's a lot of open data available about the demographics and geography of the planet. But this information is not necessarily supervised in any particular structure from which insights can be drawn.

This task requires you to create clusters of distinct commercial centers or markets using points of interest data of a city (the city could be yours). Points of interest (POI) data provides location information of different places along with their defining tags like school, type of outlets, type of building, etc.

POI data refers to the coordinates of any physical entity with a tag describing its type like commercial buildings, schools, hospitals, restaurants, etc.

## Objective:

1. Get Points of Interest from open data sources like open street maps (OSM).
2. Understand how spatial location data works
  1. Understand spatial vector data types and how to manipulate it using your language of choice.
  2. Understand necessary GIS concepts like projections, spatial clustering, etc.
3. Figure out a way of clustering these points into commercial centers/markets. You can use standard size polygons also to cluster the points.
4. Find and label the most significant clusters, statistically and intuitively.
5. Visualize the resultant commercial centres/markets.

## Data Source:

POI (OSM) data- <https://www.openstreetmap.org/#map=11/28.6518/77.2219>

You can use overpy (a python frontend for overpass API of OSM) to get OSM data for the desired city.

(Tip: you can use [overpass-turbo](#) to frame queries)

## Submissions:

1. Point or polygon file with commercial centers/markets in ESRI shapefile.
2. Documentation around the methodology, analysis, and final results that you want to share. Do use graphs and charts to substantiate your analysis. (Bonus points if you use GitHub pages / RPosts / etc. to share your documentation)
3. Script(s) and their documentation. (Bonus points for using Jupyter Notebook or GitHub ReadMe.)

4. Visualizations. (Bonus points if you use interactive dashboards)