**Capstone Project - Cafes in Noida**

**A description of the data**: the data used to solve this problem is geolocation data collected from [FourSquare](https://foursquare.com/). Adequate explanation and discussion, with examples, of the data is the following. Data is a single dataframe, containing at least a location of the café. **Explanation** of the location data is a standard tuple (lat, lng), where lat stands for latitude and lng for longitude. Some other metadata like name, postal code and so on is also collected, but let us **discuss** that they are not absolutely necessary for the analysis. **Example** of the data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Identifier | Name | Shortname | Address | Poastalcode | Latitude | Londitude |
| 4ebfa65377c885a64e5f9052 | Abhishe | Café | TGIP | 201301 | 28.570080 | 77.3237 |
| 510d017be4b0dc1da493a09e | kaffiiaa | Café | Sector-18 | 201301 | 28.568715 | 77.3242 |
| 4c13d982127f9521d8c02425 | teasta | Tea room | 16 complex | 201301 | 28.565530 | 77.3400 |
| 5b5203c4b9a5a8002ce00945 | starbuck | Coffee shop | DLF | 201301 | 28.567397 | 77.3207 |
| 4c6d43bf1585ef3bac000a9e | Café coffee day | Coffee shop | GIP-mall | 201301 | 28.567525 | 77.3252 |

**Data will be used** in the following way: by knowing the locations of already existing cafes, it's possible to apply unsupervised learning technique like kernel density estimation (KDE) to determine the area of influence of the existing cafes, and start up new café which is not in the area of influence.