**Data Sources**

For identifying the travel locations in India, there are many data sources available. I have used following travel website to identify top travel destinations across India.

<https://www.holidify.com/collections/tourist-places-in-india>

From the source, I am planning to use following fields from above web source–

1. Tourist place
2. State of India where the tourist place located
3. Speciality of the place
4. Web source rating of the place
5. Best time to visit
6. Photo of the location from Web source

Apart from this, I would be utilizing [foursquare](https://foursquare.com/) for top 100 nearby venues (based on visitor’s rating) in neighbourhood of the identified tourist places. This data would help to identify similarities based on the venues. Following data would be extracted from source –

1. Venue
2. Venue Category

In addition, to search data for the location, we would utilize latitude and longitude data available with geocoder libraries.

**Assumptions:**

1. Ratings given by visitors for each venue in foursquare data represent true popularity of venue.
2. Tourist places chosen and ratings by web source data are reliable

**Limitations**:

1. More data sources and information would help developing more accurate model. For the project data is taken from 2 sources
2. For outliers (Very Unique locations) more calculations are required and higher levels of machine learning models could be beneficial.