# Travel Review Segmentation

**Abstract:**

Understanding the tastes of each user and the characteristics of each product is necessary to predict how a user will respond to a new product. This latent user and product dimensions can be discovered with the help of user feedback. A numeric rating and its accompanying text review is the most widely available form of user feedback. A measure which encapsulates the contents of such reviews is often necessary as they have been found to significantly influence the shopping behaviour of users. A fine-grained form of such measure that could act as perfect feedback about the product is a star rating. The review rating prediction tries to predict a rating corresponding to the given review.

**Problem Statement:**

Given the google rating data, use a hierarchical clustering algorithm to cluster reviews.

**Dataset Information:**

This data set is populated by capturing user ratings from Google reviews. Reviews on attractions from 24 categories across Europe are considered. Google user rating ranges from 1 to 5 and the average user rating per category is calculated.

**Variable Description:**

| Column | Description |
| --- | --- |
| User | Unique user id |
| Attribute 1 | Average ratings on churches |
| Attribute 2 | Average ratings on resorts |

| Attribute 3 | Average ratings on beaches |
|---|---|
| Attribute 4 | Average ratings on parks |
| Attribute 5 | Average ratings on theatres |
| Attribute 6 | Average ratings on museums |
| Attribute 7 | Average ratings on malls |
| Attribute 8 | Average ratings on zoo |
| Attribute 9 | Average ratings on restaurants |
| Attribute 10 | Average ratings on pubs/bars |
| Attribute 11 | Average ratings on local services |
| Attribute 12 | Average ratings on burger/pizza shops |
| Attribute 13 | Average ratings on hotels/other lodgings |
| Attribute 14 | Average ratings on juice bars |
| Attribute 15 | Average ratings on art galleries |
| Attribute 16 | Average ratings on dance clubs |
| Attribute 17 | Average ratings on swimming pools |
| Attribute 18 | Average ratings on gyms |
| Attribute 19 | Average ratings on bakeries |
| Attribute 20 | Average ratings on beauty & spas |
| Attribute 21 | Average ratings on cafes |
| Attribute 22 | Average ratings on viewpoints |
| Attribute 23 | Average ratings on monuments |
| Attribute 24 | Average ratings on gardens |

## Scope:

- Analyzing the existing data and getting valuable insights about the review pattern
- Data pre-processing including missing value treatment
- Cluster the reviews based on the optimum number of clusters ('k') with the help of dendrogram

## Learning Outcome:

The students will get a better understanding of how the variables are linked to each other and will be able to apply hierarchical clustering to determine review types.