This data consists of review text and ratings given by the customers on TripAdvisor.

Use only ‘review\_full’ text column from the data to cluster the reviews to identify top themes of positive and negative reviews given by the customers.

This analysis will help business to identify what customers are liking and what they are disliking and improve the service accordingly.

For this problem statement you can use complete data and cluster it to find top themes.

Also think about how clustering model can be used to automatically cluster newly occurring reviews in future.

Provide your solution which consists of

* Text data cleaning
* Feature extraction from text data
* Clustering and number of clusters identified /suggested by you as Data Scientist
* Top positive and negative themes

Use the “New\_Delhi\_Reviews’’ table from the Database.db file for the project.