**Project Details**

**Group Members:**

Alishba Tahir L1F16BSCS0040

Arooj Fatima L1F16BSCS0354

Huba Waheed L1F16BSCS0175

Nimra Mubashir L1S16BSCS0035

**Description:**

We want to make a personalized hotel recommendation system. Normally we see a list of hotels that have good rating while booking a hotel but these recommendations are not personalized. These ratings are ignorant because they are according to the reviews given by previous users for example a hotel might have many bad reviews on its food service but it would still be recommended to a user who prefers good food because the reviews of its other services are good so it is rated as one of the best hotels. We will be implementing it in Apache Spark as it is becoming increasingly popular due to its processing power being much greater than Apache Hadoop MapReduce.

**Dataset:**

We will be using the same dataset provider by our advisor for our FYP from kaggle.

**Features of Dataset**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ratings | Title(User Reviews) | Username | Number of reviews | ID | Hotel name | Location | Date stayed | Offering\_id |
| 1 |  |  |  |  | Jewel of the Upper Wets Side | Canada |  |  |
| 2 |  |  |  |  | A Nice Stay for NYC | NYC |  |  |
| 3 |  |  |  |  | Historic Luxury Hotel | Boston |  |  |
| 4 |  |  |  |  |  | Norway |  |  |
| 5 |  |  |  |  |  | And many more locations |  |  |

The dataset consists of json arrays that contain three other json arrays within it. The three arrays are ratings, author and date stayed.

1. Service, cleanliness, sleep quality, location, rooms, and review are elements of array ratings.
2. Username, num\_reviews, num\_cities, id and location are elements of array author.
3. Offereing\_id, num\_helpful\_votes, date, id and via mobile are elements of date stayed.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ratings | Service | cleanliness | Sleep quality | location | rooms | review |  |
| author | username | num\_reviews | num\_cities | id | Location |  |  |
| Date stayed | Offering id | num\_helpful\_votes | Date | Id | Via mobile |  |  |