Title: Discover new places using Location Recommendation system

Name: Uma Nataraj
SJSU Id: 013727259

Description:
Social network applications like Facebook, Foursquare, Gowalla has helped their peers discover a new place because of the check in data updated on the website by their users. Check-in data is becoming popular these days which is defined as an update given by a user when he visits a new location such as park, night club, restaurant etc. User check-in data has helped significantly to ameliorate location recommendation system. This data can be used by other users to discover a new place or advertisers to provide specific advertisements for recommending a location. User check-in data has also helped data scientists to know more about user interests, pattern of preferences etc.

Goal of the project:

- Recommend new location to users using matrix factorization methods.
- Explore various techniques to improve the recommendation system and contrast different collaborative filtering methods.

Dataset:
Gowalla is a social networking website which includes details about user check in data. The dataset is obtained from the website below,
[1] https://snap.stanford.edu/data/loc-gowalla.html

References:

E. Cho, S. A. Myers, J. Leskovec. Friendship and Mobility: Friendship and Mobility: User Movement in Location-Based Social Networks ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011.

Data Analysis:
We start by exploring the properties of the dataset, the entire dataset is divided into four files namely, user check-in, user friendship, location info and user info files. This dataset consists of 407533 total users. The total number of locations recorded in the dataset is 2844145 and total number of check-in recorded is 36001959. We now use data preparation techniques for pruning the data. In order to remove inactive users, we only keep users who have done at least 25 check-ins. Similarly, locations which do not have check in count of more than 25 have been removed.

After initial data processing, the partial results are as shown below,

	No of users	No of locations	No of check-ins
Data before pruning	407533	2844145	36001959
Data after pruning	145823	277043	21453376