Introduction

The dataset in data folder is a preprocessed subset of the Yelp Challenge dataset. The whole Yelp Challenge and its documentation is available in: https://www.yelp.com/dataset/challenge.
In time of writing this description the most recent version of the dataset were for round 13 of Yelp Challenge (we used the most recent version of the dataset at that time of writing this description file).

Description of the dataset

There are three compressed csv files in the our dataset.

File name	Columns in the file		
yelp_businesses.csv.gz (row delimiter is TAB for this file)	**Column name**	**Description**	**Column's data type**
	business_id	The unique string depicting the business's id	string
	name	The business's name	string
	address	The full address of a business	string
	city	The city name of the business	string
	state	The state name of the business	string
	postal_code	The postal code of the business	string
	latitude	The latitude for where the business if located	float
	longitude	The longitude for where the business if located	float
	stars	The average number of stars the business has received from users	float
	review_count	The number of the reviews a business has got	integer
	categories	A list of categories that the business is associated with them; every item in the list is separated by ',' from other items	string
yelp_top_reviewers_with_reviews.csv.gz (row delimiter is TAB for this file)	**Column name**	**Description**	**Column's data type**

	review_id	The unique string depicting the review's id	string
	user_id	The string depicting the id of the user who wrote the review	string
	business_id	The unique string depicting the business's id	string
	review_text	The text of the review	string; base64 encoded
	review_date	The date and time when the review has been written	Unix timestamp
yelp_top_users_friendship_graph.csv.gz (row delimiter is "," for this file)	**Column name**	**Description**	**Column's data type**
	src_user_id	The user id of the source node	string
	dst_user_id	The user id of the destination node	string

Additional information:

- The column review_text is [base64](#) encoded so you may need to decode it before being able to use its contents
- The column review_date is in Unix timestamp format, so again you may need to convert it to a [date-time format](#) to be able to be use it
- Yelp's friendship graph is directed graph