Run the following command:

*sudo vi /etc/my.cnf*

add following into the files:

*[mysqld]*

*local-infile*

*secure-file-priv=""*

*[mysql]*

*local-infile*

To run mysql

*mysql -u root -p*

CREATE DATABASE yelp;

CREATE TABLE yelp\_business

(

business\_id VARCHAR(100)

, name VARCHAR(100) DEFAULT NULL

, neighborhood VARCHAR(100) DEFAULT NULL

, address VARCHAR(200) DEFAULT NULL

, city VARCHAR(100) DEFAULT NULL

, state VARCHAR(100) DEFAULT NULL

, postal\_code VARCHAR(100) DEFAULT NULL

, latitude VARCHAR(100) DEFAULT NULL

, longitude VARCHAR(100) DEFAULT NULL

, stars DECIMAL(2, 1) DEFAULT NULL

, review\_count INT DEFAULT NULL

, is\_open INT DEFAULT NULL

, categories VARCHAR(500) DEFAULT NULL

, PRIMARY KEY (business\_id)

);

LOAD DATA LOCAL INFILE '/Users/geng/Workspace/ZTGG/project/project1/yelp\_business.csv'

INTO TABLE yelp.yelp\_business

FIELDS TERMINATED BY ','

OPTIONALLY ENCLOSED BY '"'

IGNORE 1 LINES

;

wc -l yelp\_business.csv

have you peeked the data before setting ddl?

what are the correct data types given our file?

have you checked the number of line is the same?

have you used the correct delimiter for fields and lines (which also depends on your OS)

have you correctly handled delimiters in special situations (wrapped within quotes)?

how did you handle empty or invalid values?

have you truncated any data during loading process?

are there any other warnings that you should be aware of?

CREATE TABLE yelp\_business\_hours

(

business\_id VARCHAR(100)

, monday VARCHAR(100) DEFAULT NULL

, tuesday VARCHAR(100) DEFAULT NULL

, wednesday VARCHAR(100) DEFAULT NULL

, thursday VARCHAR(100) DEFAULT NULL

, friday VARCHAR(100) DEFAULT NULL

, saturday VARCHAR(100) DEFAULT NULL

, sunday VARCHAR(100) DEFAULT NULL

, PRIMARY KEY (business\_id)

);

LOAD DATA LOCAL INFILE '/Users/geng/Workspace/ZTGG/project/project1/yelp\_business\_hours.csv'

INTO TABLE yelp.yelp\_business\_hours

FIELDS TERMINATED BY ','

IGNORE 1 LINES

;

CREATE TABLE yelp\_checkin

(

business\_id VARCHAR(100)

, weekday VARCHAR(100) DEFAULT NULL

, hour VARCHAR(100) DEFAULT NULL

, checkins VARCHAR(100) DEFAULT NULL

, PRIMARY KEY (business\_id)

);

LOAD DATA LOCAL INFILE '/Users/geng/Workspace/ZTGG/project/project1/yelp\_checkin.csv'

INTO TABLE yelp.yelp\_checkin

FIELDS TERMINATED BY ','

IGNORE 1 LINES

;

CREATE TABLE yelp\_tip

(

text VARCHAR(1000) DEFAULT NULL

, date DATE DEFAULT NULL

, likes INT DEFAULT NULL

, business\_id VARCHAR(100) DEFAULT NULL

, user\_id VARCHAR(100) DEFAULT NULL

);

LOAD DATA LOCAL INFILE '/Users/geng/Workspace/ZTGG/project/project1/yelp\_tip.csv'

INTO TABLE yelp.yelp\_tip

FIELDS TERMINATED BY ','

OPTIONALLY ENCLOSED BY '"'

ESCAPED BY ''

LINES TERMINATED BY '\r\n'

IGNORE 1 LINES

;

SELECT \*

FROM yelp.yelp\_tip

LIMIT 10200, 1

;

sed '10203q;d' yelp\_tip.csv

sed -n '10200,10300p' < yelp\_tip.csv

CREATE TABLE yelp\_user

(

user\_id VARCHAR(100)

, name VARCHAR(100) DEFAULT NULL

, review\_count INT DEFAULT NULL

, yelping\_since VARCHAR(100) DEFAULT NULL

, friends MEDIUMTEXT DEFAULT NULL

, useful INT DEFAULT NULL

, funny INT DEFAULT NULL

, cool INT DEFAULT NULL

, fans INT DEFAULT NULL

, elite VARCHAR(100) DEFAULT NULL

, average\_stars DECIMAL(3, 2) DEFAULT NULL

, compliment\_hot INT DEFAULT NULL

, compliment\_more INT DEFAULT NULL

, compliment\_profile INT DEFAULT NULL

, compliment\_cute INT DEFAULT NULL

, compliment\_list INT DEFAULT NULL

, compliment\_note INT DEFAULT NULL

, compliment\_plain INT DEFAULT NULL

, compliment\_cool INT DEFAULT NULL

, compliment\_funny INT DEFAULT NULL

, compliment\_writer INT DEFAULT NULL

, compliment\_photos INT DEFAULT NULL

, PRIMARY KEY (user\_id)

);

LOAD DATA LOCAL INFILE '/Users/geng/Workspace/ZTGG/project/project1/yelp\_user.csv'

INTO TABLE yelp.yelp\_user

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\r\n'

IGNORE 1 LINES

;

SELECT \*

FROM yelp.yelp\_user

LIMIT 1500, 1

;

CREATE TABLE yelp\_review

(

review\_id VARCHAR(100) DEFAULT NULL

, user\_id VARCHAR(100) DEFAULT NULL

, business\_id VARCHAR(100) DEFAULT NULL

, stars INT DEFAULT NULL

, date DATE DEFAULT NULL

, text LONGTEXT DEFAULT NULL

, useful INT DEFAULT NULL

, funny INT DEFAULT NULL

, cool INT DEFAULT NULL

);

LOAD DATA LOCAL INFILE '/Users/geng/Workspace/ZTGG/project/project1/yelp\_review.csv'

INTO TABLE yelp.yelp\_review

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

ESCAPED BY ''

LINES TERMINATED BY '\r\n'

IGNORE 1 LINES

;