```
-- created External Table
 drop table amazon_customer_reviews;
 CREATE External TABLE IF NOT EXISTS AMAZON_CUSTOMER_REVIEWS (id  STRING,
dateAdded STRING,
                                      dateUpdated STRING, name STRING,
                                      asins String,brand STRING,
                                      categories STRING, primaryCategories STRING,imageUrls
STRING,
                                      keys String,manufacturer STRING, manufacturerNumber
STRING,
                                      reviews_date STRING,reviews_dateSeen
STRING,reviews_didPurchase STRING,
                                      reviews_doRecommend STRING,reviews_id
STRING,reviews_numHelpful STRING,
                                      reviews_rating STRING,reviews_sourceURLs STRING,
                                      review_text STRING,
                                      reviews_title STRING,
                                      reviews_username STRING,
                                      sourceURLs STRING)
                                      ROW FORMAT SERDE
'org.apache.hadoop.hive.serde2.OpenCSVSerde'
                                      WITH SERDEPROPERTIES (  "separatorChar" = "\t",
                                         "quoteChar"  = """, "escapeChar"    = "\\")

                                      --ROW FORMAT DELIMITED FIELDS TERMINATED BY ' '
                                     STORED AS TEXTFILE;


--loading file into table
LOAD DATA LOCAL INPATH '/home/cloudera/Amazon_Consumer_Reviews.txt' OVERWRITE
INTO TABLE amazon_customer_reviews;


--checking the count of loaded records
select count(1) from amazon_customer_reviews;

--checking how data is loaded into table
describe amazon_customer_reviews;

--Top 10 products based on average ratings
select id,avg(reviews_rating) as Rating from amazon_customer_reviews group by id order by
```

```
Rating DESC limit 1000;

--Number of products per rating
select reviews_rating, count(id) as noOfProducts from amazon_customer_reviews group by
reviews_rating order by noOfProducts desc limit 5;

--number of unique products
SELECT name,COUNT (Distinct reviews_id) as review_count FROM
amazon_customer_reviews group by name sort by review_count DESC limit 5;

--most popular Products
select id, count(id) as mostordered from amazon_customer_reviews group by  id order by
mostordered desc limit 10;

--most valued customer.
select reviews_username,count(asins) as totalProducts from amazon_customer_reviews group
by reviews_username order by totalProducts desc limit 20;

-- Created Managed table
drop table amazon_customer_reviews_req_col;
CREATE TABLE IF NOT EXISTS AMAZON_CUSTOMER_REVIEWS_req_col(name STRING,
                                      reviews_rating STRING,reviews_text STRING,
                                     reviews_title STRING,reviews_username string)
                                      ROW FORMAT SERDE
'org.apache.hadoop.hive.serde2.OpenCSVSerde'
                                      WITH SERDEPROPERTIES (  "separatorChar" = "\t",
                                         "quoteChar"  = """, "escapeChar"    = "\\")
                                    STORED AS TEXTFILE  ;


-- formatted to managed table
describe formatted amazon_customer_reviews_req_col;

--loading data into the table from external table
insert overwrite table amazon_customer_reviews_req_col select name, reviews_rating,
review_text, reviews_title, reviews_username from amazon_customer_reviews ;


-- Removing null values in the column
select name, reviews_rating,reviews_text, reviews_title, reviews_username from
amazon_customer_reviews where name is not null AND review_text is not null AND
reviews_title is not null AND reviews_username is not null;
```

```
-- checking success of null value removal.
select count(name), count(reviews_rating), count(reviews_text), count(reviews_title) ,
count(reviews_username) from amazon_customer_reviews_req_col
where name is null AND reviews_text is null AND reviews_title is null AND reviews_username is
null AND reviews_rating is null;

--checking the table after processing.
select * from amazon_customer_reviews_req_col;
```