1. CREATE DATABASE mmishra2;
2. use mmishra2;
3. DROP TABLE IF EXISTS dictionary\_us;
4. CREATE EXTERNAL TABLE if not exists dictionary\_us (

type string,

length int,

word string,

pos string,

stemmed string,

polarity string )

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE

LOCATION '/user/**mmishra2**/project/tables/dictionary\_us';

1. DROP TABLE IF EXISTS dictionary\_ge;
2. CREATE EXTERNAL TABLE if not exists dictionary\_ge (

word string,

word1 string,

misc string,

polarity string,

stemmed string,

misc1 string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE

LOCATION '/user/**mmishra2**/project/tables/dictionary\_ge';

1. DROP TABLE IF EXISTS dictionary\_fr;
2. CREATE EXTERNAL TABLE if not exists dictionary\_fr (

word string,

polarity string,

joy int,

fear int,

sadness int,

anger int,

surprise int,

disgust int )

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ';'

STORED AS TEXTFILE

LOCATION '/user/**mmishra2**/project/tables/dictionary\_fr';

1. DROP TABLE IF EXISTS review\_fr;
2. CREATE EXTERNAL TABLE if not exists review\_fr (

marketplace string,

customer\_id int,

review\_id string,

product\_id string,

product\_parent int,

product\_title string,

product\_category string,

star\_rating int,

helpful\_votes int,

total\_votes int,

vine string,

verified\_purchase string,

review\_headline string,

review\_body string,

review\_date timestamp )

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE

LOCATION '/user/**mmishra2**/project/tables/fr';

1. DROP TABLE IF EXISTS review\_us;
2. CREATE EXTERNAL TABLE if not exists review\_us (

marketplace string,

customer\_id int,

review\_id string,

product\_id string,

product\_parent int,

product\_title string,

product\_category string,

star\_rating int,

helpful\_votes int,

total\_votes int,

vine string,

verified\_purchase string,

review\_headline string,

review\_body string,

review\_date timestamp )

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE

LOCATION '/user/**mmishra2**/project/tables/us';

1. DROP TABLE IF EXISTS review\_uk;
2. CREATE EXTERNAL TABLE if not exists review\_uk (

marketplace string,

customer\_id int,

review\_id string,

product\_id string,

product\_parent int,

product\_title string,

product\_category string,

star\_rating int,

helpful\_votes int,

total\_votes int,

vine string,

verified\_purchase string,

review\_headline string,

review\_body string,

review\_date timestamp )

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE

LOCATION '/user/**mmishra2**/project/tables/uk';

1. DROP TABLE IF EXISTS review\_de;
2. CREATE EXTERNAL TABLE if not exists review\_de (

marketplace string,

customer\_id int,

review\_id string,

product\_id string,

product\_parent int,

product\_title string,

product\_category string,

star\_rating int,

helpful\_votes int,

total\_votes int,

vine string,

verified\_purchase string,

review\_headline string,

review\_body string,

review\_date timestamp )

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE

LOCATION '/user/**mmishra2**/project/tables/de';

1. alter table review\_de SET SERDEPROPERTIES ("timestamp.formats"="yyyy-MM-dd");
2. alter table review\_us SET SERDEPROPERTIES ("timestamp.formats"="yyyy-MM-dd");
3. alter table review\_uk SET SERDEPROPERTIES ("timestamp.formats"="yyyy-MM-dd");
4. alter table review\_fr SET SERDEPROPERTIES ("timestamp.formats"="yyyy-MM-dd");
5. DROP TABLE IF EXISTS review;
6. CREATE TABLE review AS

select \* from review\_de where review\_id is not null and star\_rating is not null

union

select \* from review\_fr where review\_id is not null and star\_rating is not null

union

select \* from review\_uk where review\_id is not null and star\_rating is not null

union

select \* from review\_us where review\_id is not null and star\_rating is not null;

1. **Create table based on review count group by star rating**

CREATE TABLE IF NOT EXISTS rating ROW FORMAT DELIMITED

FIELDS TERMINATED BY "," STORED AS TEXTFILE

LOCATION "/user/**mmishra2**/project/tables/one" AS

SELECT COUNT(review\_id)count ,star\_rating FROM review

GROUP BY star\_rating ORDER BY star\_rating;

1. **Create table based on top ten highest number of reviews given by unique users over ten years**

CREATE TABLE IF NOT EXISTS users ROW FORMAT DELIMITED

FIELDS TERMINATED BY "," STORED AS TEXTFILE

LOCATION "/user/**mmishra2**/project/tables/two" AS

SELECT COUNT(review\_id) count, customer\_id FROM review

GROUP BY customer\_id ORDER BY count DESC LIMIT 10;

1. **Create table based on top ten highest number of reviews grouped by year and month**

CREATE TABLE IF NOT EXISTS review\_date ROW FORMAT DELIMITED

FIELDS TERMINATED BY "," STORED AS TEXTFILE

LOCATION "/user/mmishra2/project/tables/three" AS

SELECT COUNT(review\_id) count, YEAR(review\_date)year, MONTH(review\_date) month

FROM review

GROUP BY YEAR(review\_date),MONTH(review\_date)

ORDER BY year, month;

1. **Create table based on review count by product category**

CREATE TABLE IF NOT EXISTS product ROW FORMAT DELIMITED

FIELDS TERMINATED BY "," STORED AS TEXTFILE

LOCATION "/user/**mmishra2**/project/tables/four" AS

SELECT COUNT(review\_id) count, product\_category FROM review

GROUP BY product\_category;

1. **create table for top ten popular products based on average rating of 5 and having maximum review count**

CREATE TABLE IF NOT EXISTS popular ROW FORMAT DELIMITED

FIELDS TERMINATED BY "," STORED AS TEXTFILE

LOCATION "/user/**mmishra2**/project/tables/five" AS

SELECT COUNT(review\_id)count, AVG(star\_rating) rating,

product\_title, product\_category, marketplace FROM review

GROUP BY product\_title, product\_category, marketplace

HAVING AVG(star\_rating) = 5 ORDER BY count DESC LIMIT 10;

1. **create table showing rating of product category “Baby” by country**

CREATE TABLE IF NOT EXISTS baby ROW FORMAT DELIMITED

FIELDS TERMINATED BY "," STORED AS TEXTFILE

LOCATION "/user/**mmishra2**/project/tables/seven" AS

SELECT review\_id, star\_rating, marketplace, review\_date

FROM review WHERE product\_category = 'Baby';

1. **create table eventually based on sentimental analysis of reviews by country (created using a series of steps)**
2. CREATE VIEW IF NOT EXISTS v1 AS

SELECT marketplace,review\_id, review\_body FROM review ;

1. CREATE VIEW IF NOT EXISTS v2 AS

SELECT marketplace,review\_id,words FROM v1

LATERAL VIEW EXPLODE(SENTENCES(LOWER(review\_body))) dummy as words;

1. CREATE VIEW IF NOT EXISTS v3 AS

SELECT marketplace,review\_id,word FROM v2

LATERAL VIEW EXPLODE( words ) dummy as word;

1. CREATE VIEW IF NOT EXISTS v4 AS

SELECT marketplace,review\_id,v3.word,

CASE d\_us.polarity

WHEN 'negative' THEN -1

WHEN 'positive' THEN 1

ELSE 0 END AS polarity

FROM v3 LEFT OUTER JOIN dictionary\_us d\_us on v3.word = d\_us.word

WHERE marketplace ='US'

UNION

CREATE VIEW IF NOT EXISTS v4 AS

SELECT marketplace,review\_id,v3.word,

CASE d\_us.polarity

WHEN 'negative' THEN -1

WHEN 'positive' THEN 1

ELSE 0 END AS polarity

FROM v3 LEFT OUTER JOIN dictionary\_us d\_us on v3.word = d\_us.word

WHERE marketplace ='UK'

UNION

SELECT marketplace,review\_id,v3.word,

CASE d\_fr.polarity

WHEN 'negative' THEN -1

WHEN 'positive' THEN 1

ELSE 0 END AS polarity

FROM v3 LEFT OUTER JOIN dictionary\_fr d\_fr on v3.word = d\_fr.word

WHERE marketplace = 'FR'

UNION

SELECT marketplace,review\_id,v3.word,

CASE d\_ge.polarity

WHEN 'negative' THEN -1

WHEN 'positive' THEN 1

ELSE 0 END AS polarity

FROM v3 LEFT OUTER JOIN dictionary\_ge d\_ge on v3.word = d\_ge.word

WHERE marketplace = 'DE';

1. CREATE view IF NOT EXISTS v5 AS

SELECT marketplace, review\_id,

CASE

WHEN SUM( polarity ) > 0 THEN 'positive'

WHEN SUM( polarity ) < 0 THEN 'negative'

ELSE 'neutral' END AS sentiment

FROM v4 GROUP BY marketplace, review\_id;

1. DROP TABLE IF EXISTS sentiment;
2. CREATE TABLE IF NOT EXISTS sentiment ROW FORMAT DELIMITED

FIELDS TERMINATED BY "," STORED AS TEXTFILE

LOCATION "/user/**mmishra2**/project/tables/All\_Countries" AS

SELECT \* FROM v5;

1. **Sentimental analysis using Ngram of the Least Rated product in US with review count greater than 50**
2. SELECT product\_id, product\_title, marketplace, product\_category, FORMAT\_NUMBER(AVG(star\_rating),2) AS avg\_rating, COUNT(\*) AS num

FROM review\_us

GROUP BY product\_id, product\_title, marketplace, product\_category

HAVING num > 50 ORDER BY avg\_rating limit 5;

1. SELECT product\_id, product\_title, marketplace, product\_category, FORMAT\_NUMBER(AVG(star\_rating),2) AS avg\_rating, COUNT(\*) AS num

FROM review\_us

WHERE product\_id = 'B00B5P37IG'

GROUP BY product\_id, product\_title, marketplace, product\_category;

1. SELECT EXPLODE(NGRAMS(SENTENCES(LOWER(review\_body)), 2, 5))

AS bigrams

FROM review\_us

WHERE product\_id= 'B00B5P37IG';

1. SELECT EXPLODE(NGRAMS(SENTENCES(LOWER(review\_body)), 3, 5))

AS trigrams

FROM review\_us

WHERE product\_id= 'B00B5P37IG';

1. SELECT review\_body

FROM review\_us

WHERE product\_id= 'B00B5P37IG'

AND review\_body LIKE '%waste of %'

LIMIT 3;