CODE

1. CREATE TABLE CONSUMER_COMPLAINT_DATABASE1(Date_Received string, Product string, Sub_Product string, Issue String,Sub_issue string,Consumer_complaint_narrative string,Company_public_response string, Company string, State string, Zip_code int, submitted_via string,Date_sent_to_company string,Company_response_to_consumer string,Timely_response string,Consumer_disputed string,Complaint_ID string)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
TBLPROPERTIES("skip.header.line.count=1");

2. LOAD DATA INPATH '\HdiSamples\Consumer_Complaints.CSV' OVERWRITE INTO TABLE CONSUMER_COMPLAINT_DATABASE1;

3. select product, count(product) AS INC_OCCURE from CONSUMER_COMPLAINT_DATABASE1 GROUP BY product ORDER BY INC_OCCURE desc;

4. select sub_product, count(sub_product) AS INC_OCCURE from CONSUMER_COMPLAINT_DATABASE1 GROUP BY sub_product ORDER BY INC_OCCURE desc;

5. select state, count(state) AS INC_OCCURE from CONSUMER_COMPLAINT_DATABASE1 GROUP BY state ORDER BY INC_OCCURE desc;

6. select zip_code,count(zip_code) AS INC_OCCURE from CONSUMER_COMPLAINT_DATABASE1 GROUP BY zip_code ORDER BY INC_OCCURE desc;

7. select company,count(company) AS INC_OCCURE from CONSUMER_COMPLAINT_DATABASE1 GROUP BY company ORDER BY INC_OCCURE desc;

8. select submitted_via,count(submitted_via) AS INC_OCCURE from CONSUMER_COMPLAINT_DATABASE1 GROUP BY submitted_via ORDER BY INC_OCCURE desc;

9. SELECT CONSUMER_COMPLAINT_DATABASE1.company_response_to_consumer as response, count(*) as occured from
(select case when company_response_to_consumer rlike 'Closed with explanation' then 'Closed with explanation'
when company_response_to_consumer rlike 'Closed with non-monetary relief' then 'Closed with monetary relief'
when company_response_to_consumer rlike 'web' then 'web'
when company_response_to_consumer rlike 'closed' then 'closed'
else 'others'
end as company_response_to_consumer from
CONSUMER_COMPLAINT_DATABASE1)CONSUMER_COMPLAINT_DATABASE1 group by CONSUMER_COMPLAINT_DATABASE1.company_response_to_consumer order by occured;

10. SELECT YEAR,COUNT(YEAR) from DB3 GROUP BY YEAR;