**1. Objectives of the project**

The Bureau of Consumer Financial Protection (BCFP) maintains an excellent data set of the complaints it receives from consumers in regards to their experience with financial products. This data set is consistently updated and offers many details about the complaints filed by consumers as well as the way companies respond to those complaints.

**customer**

* When customers deciding which company’s products to choose, they need objective information about which company may provide the best experience. Which, our database has that information that will help customer make better decisions.

**company**

* When companies need to find some insights from their business performance or their competitors, they will have easy access to our database and utilize the information to improve their service.

**2. Project Progressing**

**2.1 Sample data entry and database creation**

See appendix I

**2.2 ER model and related table**

See appendix II and III

**2.3 Data Flow Diagram**

See appendix IV

**2.4 Query example questions**

See appendix V

* *Appendix I/ Sample data entry and database creation*

CREATE TABLE products(ProductName VARCHAR(40),

ProductID VARCHAR(5),

PRIMARY KEY (ProductID));

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/product.csv'

INTO TABLE products

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

CREATE TABLE subproducts(SubProductName VARCHAR(40),

SubProductID VARCHAR(5),

PRIMARY KEY (SubProductID));

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/subproduct.csv'

INTO TABLE subproducts

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

CREATE TABLE company(CompanyName VARCHAR(60) PRIMARY KEY,

comphone VARCHAR(16),

comstreet VARCHAR(50),

comcity VARCHAR(40),

comstate VARCHAR(2),

comzip VARCHAR(5));

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/company.csv'

INTO TABLE company

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 lines

(CompanyName,comphone,comstreet,comcity,comstate,@comzip)

SET comzip = nullif(@comzip,00000);

CREATE TABLE States(StateName VARCHAR(20),

State VARCHAR(2) PRIMARY KEY,

StateArea DECIMAL(9,2),

Pop25over INT,

GraduateDeg DECIMAL (3,1),

HSorHigher DECIMAL(3,1),

BachelorsorHigher DECIMAL (3,1),

MedianEarnings INT,

Pop2010 INT,

Pop2011 INT,

Pop2012 INT,

Pop2013 INT,

Pop2014 INT,

Pop2015 INT);

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/states.csv'

INTO TABLE states

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

CREATE TABLE zips(ZipCode VARCHAR(5) PRIMARY KEY,

ZipArea DECIMAL(7,3),

Pop2010 INT,

Pop2011 INT,

Pop2012 INT,

Pop2013 INT,

Pop2014 INT,

Pop25over INT,

GraduateDeg DECIMAL (4,1),

HSorHigher DECIMAL (4,1),

BachelorsorHigher DECIMAL (4,1),

MedianEarnings INT);

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/zips.csv'

INTO TABLE zips

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

CREATE TABLE consumer(State VARCHAR(2) REFERENCES states,

ZipCode VARCHAR(5) REFERENCES zips,

ConsumerID VARCHAR(6) PRIMARY KEY);

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/consumers.csv'

INTO TABLE consumer

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

CREATE TABLE complaints(Rec\_Date DATE,

ProductID VARCHAR(5) REFERENCES products,

SubProductID VARCHAR(5) REFERENCES subproducts,

Issue VARCHAR(60),

Subissue VARCHAR(60),

CompanyName VARCHAR(60) REFERENCES company,

How\_submitted VARCHAR(15),

Sent\_to\_company DATE,

ComplaintID VARCHAR(7) PRIMARY KEY,

ConsumerID VARCHAR(6) REFERENCES consumer);

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/complaints.csv'

INTO TABLE complaints

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES

(@Rec\_Date, ProductID, SubProductID, Issue, Subissue, CompanyName, How\_submitted, @Sent\_to\_company, ComplaintID, ConsumerID)

SET Rec\_Date = str\_to\_date(@Rec\_Date,'%c/%e/%Y'),

Sent\_to\_company = str\_to\_date(@Sent\_to\_company,'%c/%e/%Y');

CREATE TABLE responses(CompanyName VARCHAR(60) REFERENCES company,

PublicResponse VARCHAR(130),

Response\_to\_consumer VARCHAR(40),

Timely VARCHAR(3),

ConsumerDisputed VARCHAR(4),

ComplaintID VARCHAR(7) PRIMARY KEY REFERENCES complaints);

LOAD DATA LOCAL INFILE 'C:/Users/Steven Spangler/OneDrive - Georgia State University/CIS 8040/Project/finaltables/responses.csv'

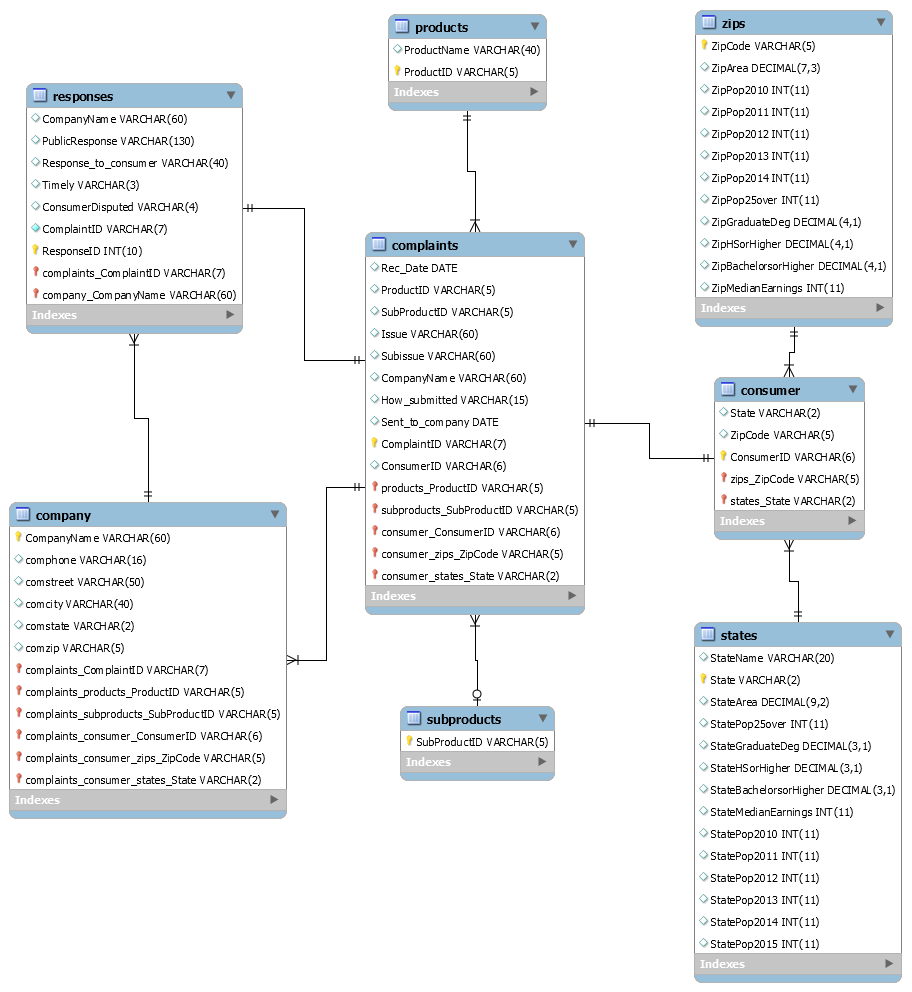
INTO TABLE responses

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

* *Appendix II/ ER model and related table*



* *Appendix V/ Query question examples*

**1, which state has the highest rate of complaints? Complaints per person.**

*SQL queries:*

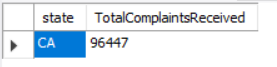
select co.state,count(co.state) as TotalComplaintsReceived

from complaints.consumer co

group by co.state

order by TotalComplaintsReceived desc limit 1;

*Output:*

**

**2 Are there more complaints in areas with higher education? Or in areas with lower education?**

*SQL queries:*

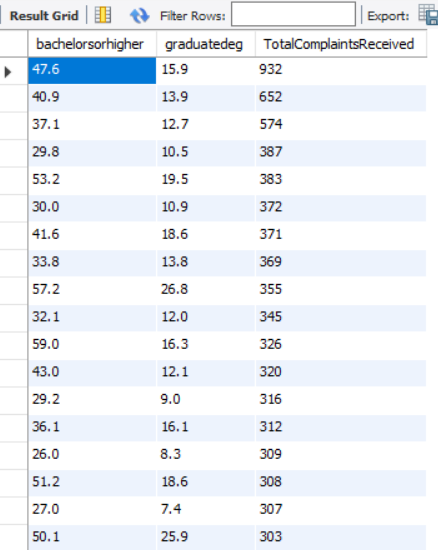
select zi.bachelorsorhigher,zi.graduatedeg,count(c.complaintid) as TotalComplaintsReceived

from complaints.complaints c,complaints.consumer co,complaints.zips zi

where c.consumerid=co.consumerid and co.zipcode=zi.zipcode group by co.zipcode

order by count(c.complaintid) desc ;

*Output*

*:* ****

**3 Are there more complaints in areas with higher income? Or in areas with lower income?**

*SQL queries:*

select zi.medianearnings,count(c.complaintid) as TotalComplaintsReceived

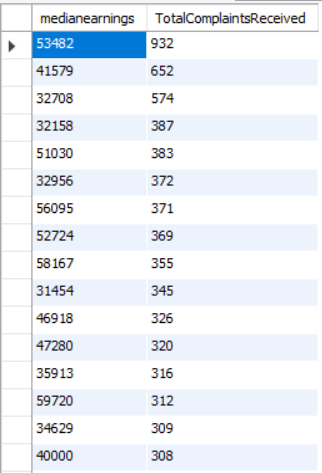
from complaints.complaints c,complaints.consumer co,complaints.zips zi

where c.consumerid=co.consumerid and co.zipcode=zi.zipcode

group by co.zipcode

order by count(c.complaintid) desc ;

*Output:*



**4 Does certain company have less complaints per population density.**

*SQL queries:*

select c.companyname,zi.pop25over,count(c.complaintid) as TotalComplaintsReceived

from complaints.complaints c,complaints.consumer co,complaints.zips zi

where c.consumerid=co.consumerid and co.zipcode=zi.zipcode

group by co.zipcode

order by count(c.complaintid) asc limit 20;

*Output:*



**5 Does certain product have more complaints than other product?**

*SQL queries:*

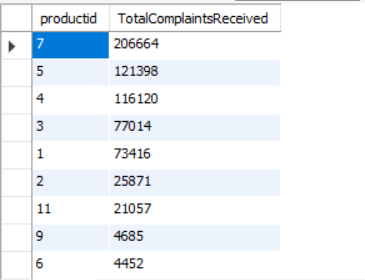
select productid,count(complaintid) as TotalComplaintsReceived

from complaints.complaints

group by productid

order by TotalComplaintsReceived desc;

*Output:*



**6 which company has the higher rate to timely response?**

*SQL queries:*

select companyname,count(Timely="Yes") as TimelyResponse from complaints.responses

group by companyname

order by TimelyResponse desc;

*Output:*



**7 which company has the lower rate of the response disputed by consumer?**

*SQL queries:*

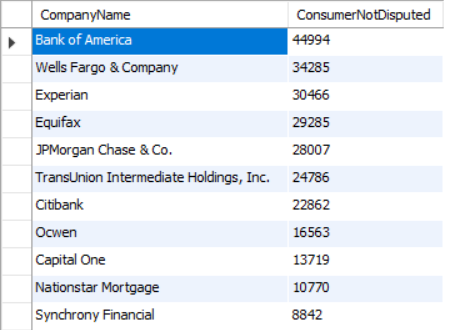
select CompanyName,count(consumerdisputed) as ConsumerNotDisputed

from complaints.responses where consumerdisputed="No"

group by CompanyName

order by ConsumerNotDisputed desc;

*Output:*



* *Appendix IV/ Data Flow Diagram*

