

U.S. Consumer: Finance Complaints

Analysing Data with SQL

Bruno M. Serafim

01/12/2021

Table of Contents

- Brief Introduction
- Objectives
- Results
- Conclusions
- Appendix (Queries)

Brief Introduction

Data

- Data from Consumer Financial Protection Bureau (CFPB), U.S.
 - (Downloaded from *Kaggle**)
- Collection of complaints about consumer financial products and services, sent from CFPB to companies for response
- SQLite database, with 18 columns and 555,957 rows
- Date range: 01/12/2011 - 25/04/2016

*<https://www.kaggle.com/kaggle/us-consumer-finance-complaints>

Context

- Solid economic recovery after 2008 sub-prime crisis*
- Heated credit market, consumer base expanding**
- Interest rates near 0%*
- U\$ 4 trillion boost from Federal Reserve (FED)*
- Expansionary fiscal policy*
 - (continuous budget deficit)

*<https://commonslibrary.parliament.uk/research-briefings/sn06901/>

**<https://www.mortgagecalculator.org/helpful-advice/mortgage-statistics.php>

Objectives

Objectives

- To analyse and bring insights about filed complaints, by category:
 - Products (and sub-products)
 - Companies
 - States
 - Consumer disputes
- To investigate complaints evolution throughout time
 - Analysing possible trends and seasonality

Results

Results

Complaints by product (2011 - 2016), ranked

Rank	Product	Number of Complaints	Percentage of Total
1	Mortgage	186,475	33.54%
2	Debt collection	101,052	18.18%
3	Credit reporting	91,854	16.52%
4	Credit card	66,468	11.96%
5	Bank account or service	62,563	11.25%
6	Consumer Loan	20,990	3.78%
7	Student loan	15,839	2.85%
8	Payday loan	3,877	0.70%
9	Money transfers	3,812	0.69%
10	Prepaid card	2,470	0.44%
11	Other financial service	557	0.10%

91.45%
of total

Results

Mortgage complains, ranked by issue (2011 - 2016)

Category: Mortgages			
Rank	Issue	Number of Complaints	Percentage of Total
1	Loan modification, collection, foreclosure	97,191	52.12%
2	Loan servicing, payments, escrow account	60,375	32.38%
3	Application, originator, mortgage broker	13,306	7.14%
4	Settlement process and costs	6,888	3.69%
5	Credit decision / Underwriting	4,418	2.37%
6	Other	4,297	2.30%

Most mortgage complaints
(84.50%) related to post-
aquisition issues



Access to mortgage loans
doesn't seem to be a major
problem

Results

Complaints by sub-product (2011 - 2016), ranked

Rank	Sub-product	Number of Complaints	Percentage of Total
1	Non-specified	158,322	28.48%
2	Other mortgage	74,319	13.37%
3	Conventional fixed mortgage	57,182	10.29%
4	Checking account	44,369	7.98%
5	Other (i.e. phone, health club, etc.)	29,617	5.33%
6	I do not know	21,634	3.89%
7	Conventional adjustable mortgage (ARM)	20,941	3.77%
8	Credit card	20,825	3.75%
9	FHA mortgage	19,152	3.44%
10	Non-federal student loan	17,969	3.23%
11	Medical	13,297	2.39%
12	Vehicle loan	12,168	2.19%
13	Other than previously specified	66,162	11.90%

Messy data
input



Unreliable
data

Results

Complaints by company (2011 - 2016), ranked - Top 10

Rank	Company	Number of Complaints	Percentage of Total
1	Bank of America	55,998	10.07%
2	Wells Fargo & Company	42,024	7.56%
3	JPMorgan Chase & Co.	33,881	6.09%
4	Equifax	31,828	5.72%
5	Experian	30,905	5.56%
6	Citibank	25,540	4.59%
7	TransUnion Intermediate Holdings, Inc.	25,534	4.59%
8	Ocwen	20,978	3.77%
9	Capital One	15,628	2.81%
10	Nationstar Mortgage	13,250	2.38%

2th largest bank by assets (2016)*

Largest bank by assets (2016)*

Bank of America: disproportionately high number of complaints, when compared to assets size

Top 10 (2011 - 2016):
6 Banks
4 Credit agencies

53.14% of total

*<https://www.relbanks.com/top-us-banks/assets-2016>

Results

Complaints of Bank of America (2011 - 2016), ranked by product - Top 10

Bank of America			
Rank	Product	Number of Complaints	Percentage of Total
1	Mortgage	36,228	64.70%
2	Bank account or service	10,440	18.64%
3	Credit card	7,224	12.90%
4	Debt collection	1,173	2.09%
5	Consumer Loan	541	0.97%
6	Money transfers	200	0.36%
7	Student loan	70	0.13%
8	Other financial service	44	0.08%
9	Credit reporting	37	0.07%
10	Prepaid card	33	0.06%

96.24%
of all complaints
of Bank of
America

Results

Complaints by state (2011 - 2016), ranked - Top 10

State	Number of Complaints	Percentage of Total	Rank (complaints)	Rank (population - 2019)*
California (CA)	81,700	14.70%	1	1
Florida (FL)	53,673	9.65%	2	3
Texas (TX)	41,352	7.44%	3	2
New York (NY)	38,266	6.88%	4	4
Georgia (GA)	24,548	4.42%	5	8
New Jersey (NJ)	22,408	4.03%	6	11
Pennsylvania (PA)	20,015	3.60%	7	5
Illinois (IL)	19,624	3.53%	8	6
Virginia (VA)	18,150	3.26%	9	12
Maryland (MD)	17,703	3.18%	10	19

Number of complaints in New Jersey and Maryland are disproportionately high (when compared to population size)



An analysis within these states is needed

*<https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html>

Results

Complaints comparison between states, ranked - Top 5

California			
Rank	Product	Number of Complaints	Percentage of Total
1	Mortgage	32,988	40.38%
2	Debt collection	13,568	16.61%
3	Credit reporting	11,965	14.65%
4	Credit card	9,078	11.11%
5	Bank account or service	8,519	10.43%

New Jersey			
Rank	Product	Number of Complaints	Percentage of Total
1	Mortgage	8,411	37.54%
2	Debt collection	3,304	14.74%
3	Bank account or service	3,046	13.59%
4	Credit reporting	2,881	12.86%
5	Credit card	2,870	12.81%

Maryland			
Rank	Product	Number of Complaints	Percentage of Total
1	Mortgage	7,000	39.54%
2	Debt collection	2,617	14.78%
3	Credit reporting	2,551	14.41%
4	Bank account or service	2,026	11.44%
5	Credit card	1,999	11.29%

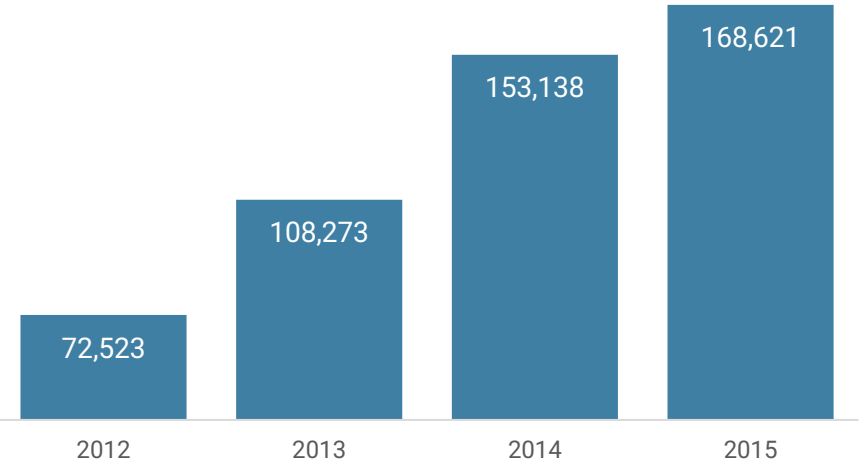
'Complaints distribution' nearly exactly the same



Further investigation is necessary to understand why New Jersey and Maryland had disproportional high number of complaints (compared to population size)

Results - Complaints over Time and Seasonality

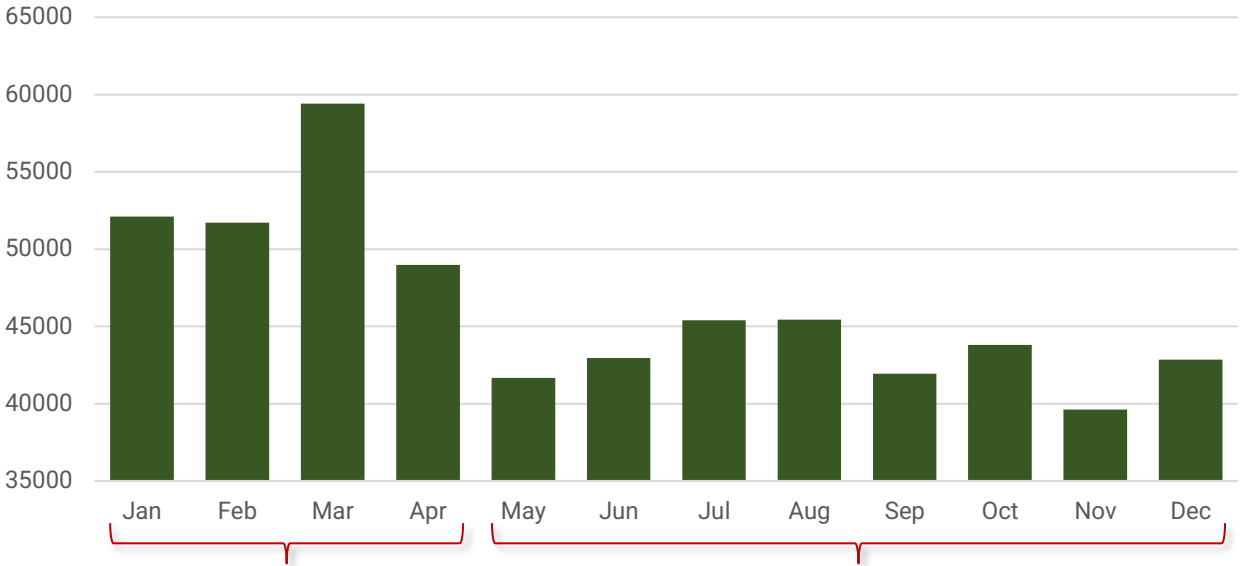
Complaints Throughout The Years



+132.5% (2012 to 2015)

Sharp increase in number of complaints

Complaints by Month (2012 - 2015) - Seasonality



Monthly av: 53,057
complaints
(14.5% above total average)

Monthly av: 42,966
complaints
(7.3% below total average)

First trimester (and quarter)
accrued more complains than
rest of the year

Relevant?

Results - Complaints over Time, by Product



Largest increase in
complaints number

Debt collection
complaints
increased...

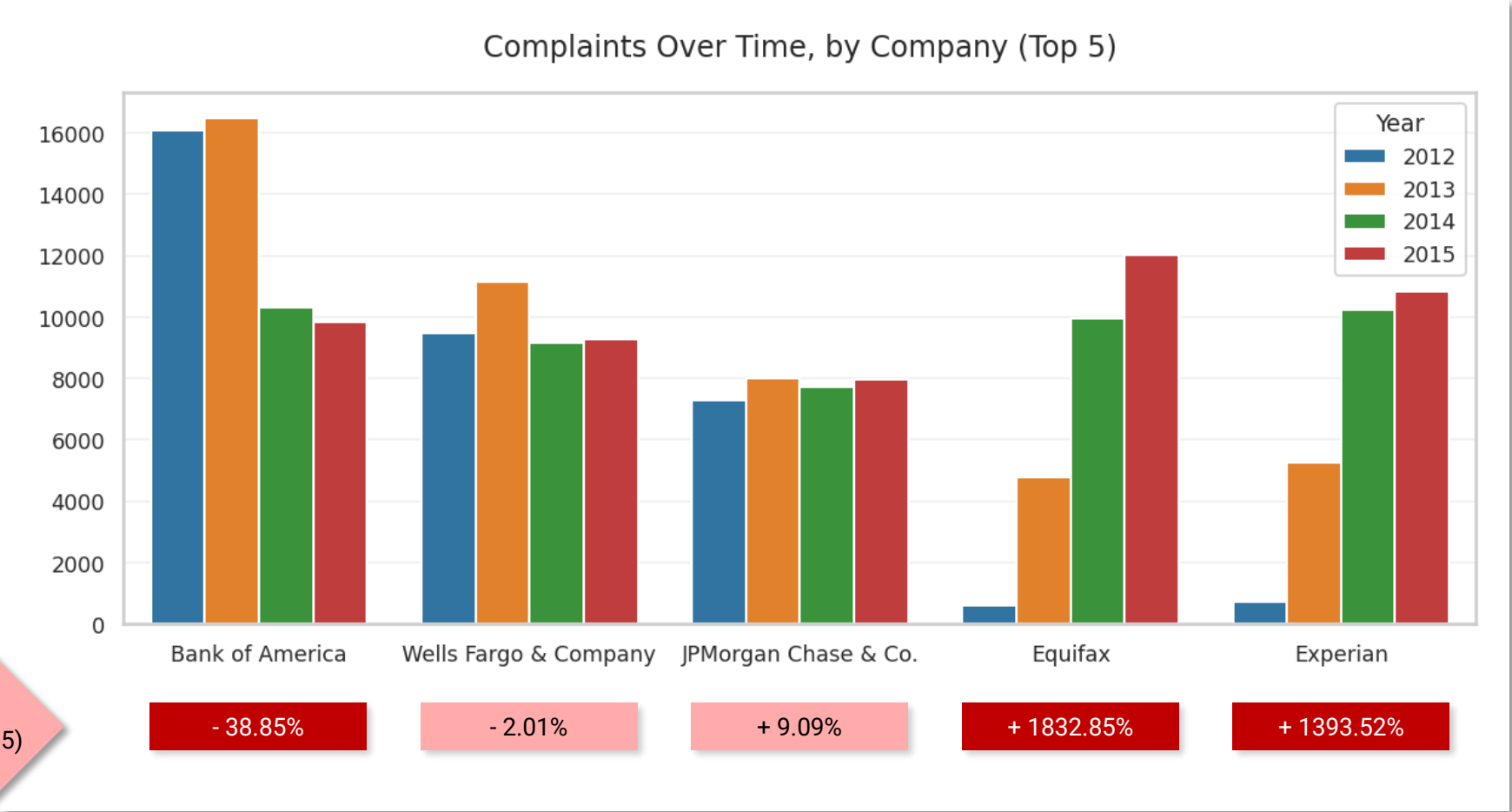


...but credit
delinquency
decreased*
(same period)



Further investigation
needed

Results - Complaints over Time, by Company



Results - Complaints Comparison between Equifax and Experian

Equifax			
Rank	Issue	Number of Complaints	Percentage of Total (from Equifax)
1	Incorrect information on credit report	22,538	70.81%
2	Credit reporting company's investigation	3,762	11.82%
3	Unable to get credit report/credit score	3,404	10.69%
4	Improper use of my credit report	1,050	3.30%
5	Credit monitoring or identity protection	940	2.95%

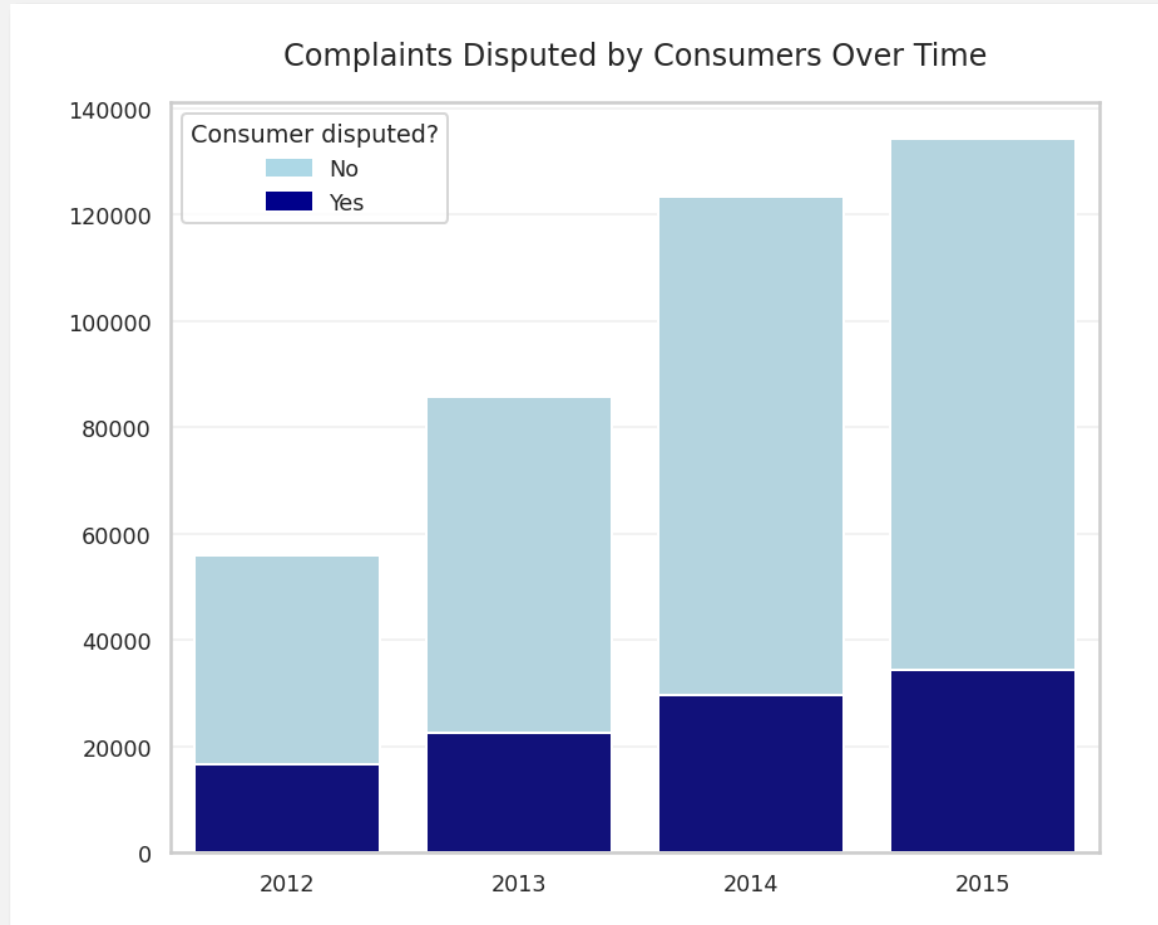
Experian			
Rank	Issue	Number of Complaints	Percentage of Total (from Experian)
1	Incorrect information on credit report	21,933	70.97%
2	Credit reporting company's investigation	3,913	12.66%
3	Unable to get credit report/credit score	2,609	8.44%
4	Credit monitoring or identity protection	1,204	3.90%
5	Improper use of my credit report	1,109	3.59%

'Issues distribution' nearly identical in both companies



It might suggest that the problem is within the sector (credit reporting), not necessarily with specific companies

Results - Complaints over Time



From 2012 to 2015:

Complaints



+ 132.5%

Consumer disputes



+ 107.5%

Further understanding of dispute process is necessary to investigate this difference



It might be related to the fact that number of complains of credit reporting agencies grew over time

Conclusions

Conclusions

Full data range (01/12/2011 – 25/04/2016)

- Mortgage was the product with more complaints (33.54%)
 - With loan modification (collection, foreclosure) and loan servicing (payments, escrow account) representing 84.50% of it
- Bank of America was the company with more complaints received (10.07%)
 - Mortgage was their most complained product (64.70%)
- California was the state with more complaints (14.70%), but also the one with largest population in U.S, though
 - New Jersey and Maryland, when compared to population sizes, were disproportionally high in number of complaints (4.03% and 3.18% respectively)
 - When compared to California, complaints distributions were nearly identical, though

Evolution throughout time (From 2012 to 2015)

- Number of complaints increased 132.50%
 - Monthly average between January and April was 14.5% larger than total monthly average
 - “Small” seasonality effect, causes not clear
 - Debt collection and Credit reporting had the largest increases
 - Led by Equifax (+1833%) and Experian (+1394%)
 - Both nearly identical in issues distribution

Appendix

Appendix - Queries

```
-- Complaints by product (2011 - 2016), ranked
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  product,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
group by product
order by 3 desc;

-- Mortgage complains, ranked by issue (2011 - 2016)
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  issue,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
where product = 'Mortgage'
group by issue
order by 3 desc
limit 10;

-- Complaints by sub-product (2011 - 2016), ranked
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  sub_product,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
group by sub_product
order by 3 desc;
```

(1)

```
-- Complaints by company (2011 - 2016), ranked - Top 10
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  company,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
group by company
order by 3 desc
limit 10;

-- Complaints of Bank of America (2011 - 2016), ranked by product - Top 10
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  product,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
where company = 'Bank of America'
group by product
order by 3 desc
limit 10;

-- Complaints by state (2011 - 2016), ranked - Top 10
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  state,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
group by state
order by 3 desc
limit 10;
```

(2)


```

-- Finding most common products, and how much of total each one represents, in California
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  product,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
where state = 'CA'
group by product
order by 3 desc
limit 10;

-- Finding most common products, and how much of total each one represents, in New Jersey
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  product,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
where state = 'NJ'
group by product
order by 3 desc
limit 10;

-- Finding most common products, and how much of total each one represents, in Maryland
select distinct RANK() over (
  order by count(*) desc
) as 'Rank',
  product,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
where state = 'MD'
group by product
order by 3 desc
limit 10;

```

(3)

```

-- Finding the complaints distribution over the years, between 2012 and 2015
select cast(substr(date_received, 7, 10) as integer) as 'Year',
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 1), 4) as 'percent. of total'
from consumer_complaints
where Year between 2012 and 2015
group by 1
order by 1;

-- Complaints by month (2012 - 2015) - Seasonality
select substr(date_received, 1, 2),
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total'
from consumer_complaints
group by 1
order by 1;

-- Complaints over time, by product (Top 5) from 2012 to 2015
select distinct row_number() over (
  partition by substr(date_received, 7, 10)
  order by count(*) desc
) as 'Rank',
  product,
  count(*) as 'number of complaints',
  round(1.0 * count(*) / sum(count(*)) over (
    order by 3), 4) as 'percent. of total',
  substr(date_received, 7, 10) as 'Year'
from consumer_complaints
group by Year, product
order by Year, 3 desc;

```

(4)

```
-- Complaints over time, by company (Top 5) from 2012 to 2015
select cast(substr(date_received, 7, 10) as integer) as 'Year',
       company,
       count(*) as 'number of complaints',
       1.0 * (LEAD(count(*)) over(
         partition by company
         order by company) - count(*) )/count(*) 'Change (in dec. points)'
from consumer_complaints
where (company = 'Equifax' or
       company = 'Experian' or
       company = 'Bank of America' or
       company = 'JPMorgan Chase & Co.' or
       company = 'Wells Fargo & Company') and
       (Year = 2012 or
        Year = 2015)
group by Year, company
order by 2;
```

```
-- Complains by issue (2011 - 2016), from 'Equifax'
select distinct issue,
       count(*) as 'number of complaints',
       round(1.0 * count(*) / sum(count(*)) over (
         order by 3), 4) as 'percent. of total'
from consumer_complaints
where company = 'Equifax'
group by issue
order by 3 desc
limit 10;
```

```
-- Complains by issue (2011 - 2016), from 'Experian'
select distinct issue,
       count(*) as 'number of complaints',
       round(1.0 * count(*) / sum(count(*)) over (
         order by 3), 4) as 'percent. of total'
from consumer_complaints
where company = 'Experian'
group by issue
order by 3 desc
limit 10;
```

```
-- Complaints disputed by consumers over time, between 2012 and 2015
select cast(substr(date_received, 7, 10) as integer) as 'Year',
       count(*) as 'number of complaints',
       "consumer_disputed?"
from consumer_complaints
where Year between 2012 and 2015
group by 1, 3
order by 1;
```

(5)