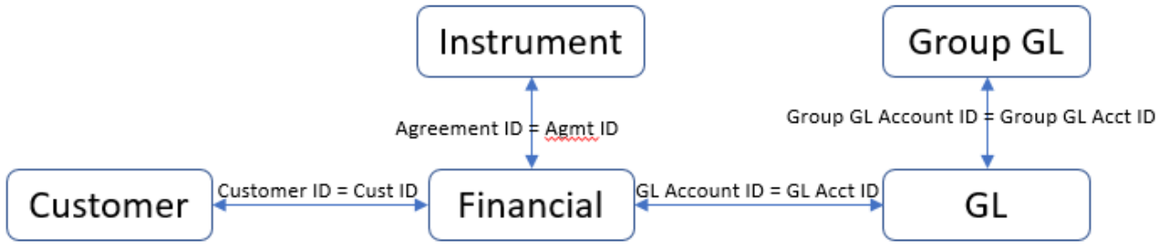
Recruitment case: Data Analyst
Jan 2023

Candidate: Siarhei Thor

Q1. Read and profile the data. Explain your understanding about the data and provide insights.

Data evaluation

Data model



Tables Info

Customer
Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	Customer_ID	6763 non-null	int64
1	Local_Customer_ID	6763 non-null	object
2	Customer_Name	5367 non-null	object
3	NACE_Code_L1	6544 non-null	object
4	NACE_Name_L1	6524 non-null	object
5	NACE_Code	6089 non-null	float64
6	Bankrupcy_Flag	6308 non-null	float64
7	Rating_Score	5454 non-null	object
8	Country	6308 non-null	object
9	Customer_Responsible_Unit	6293 non-null	float64
10	Sector_Code	6292 non-null	float64
11	Sector_Name	6292 non-null	object

Instrument
Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	Agreement_ID	23441 non-null	int64
1	Account_Number	23441 non-null	int64
2	Account_Number_Old	23441 non-null	int64
3	Performing_Non_Performing	23441 non-null	object
4	Effective_Date	23440 non-null	datetime64[ns]
5	Closing_Date	22967 non-null	datetime64[ns]
6	Maturity_Date	23159 non-null	datetime64[ns]
7	Registration_Date	19853 non-null	datetime64[ns]
8	Country	23441 non-null	object
9	Basel_FT_ID	23007 non-null	object
10	Last_Repricing_Date	14814 non-null	datetime64[ns]
11	Agreement_Purpose	23164 non-null	object
12	Amortization_Method	17210 non-null	object

Financial
Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	Agreement_ID	39356 non-null	int64
1	GL_Account_ID	39356 non-null	int64
2	Source	39356 non-null	object
3	Customer_ID	39356 non-null	int64
4	Amount	39356 non-null	float64

GL
Data columns (total 6 columns):

#	Column	Non-Null Count	Dtype
0	GL_Account_ID	17 non-null	int64
1	GL_Account	17 non-null	int64
2	GL_Name	17 non-null	object
3	Group_GL_Account_ID	17 non-null	int64
4	GL_Account_Type	17 non-null	int64
5	GL_Account_Name	17 non-null	object

Group_GL
Data columns (total 3 columns):

#	Column	Non-Null Count	Dtype
0	Group_GL_Account_ID	5 non-null	int64
1	Group_GL_Account	5 non-null	object
2	Group_GL_Name	5 non-null	object

Answer

Data model explains how different tables relate to one another.

Given extract of the following tables:

Customer
Data regarding customer information.

Instrument
Data regarding financial instrument information.

Financial
Data containing all transactions. This table can be used for aggregations.

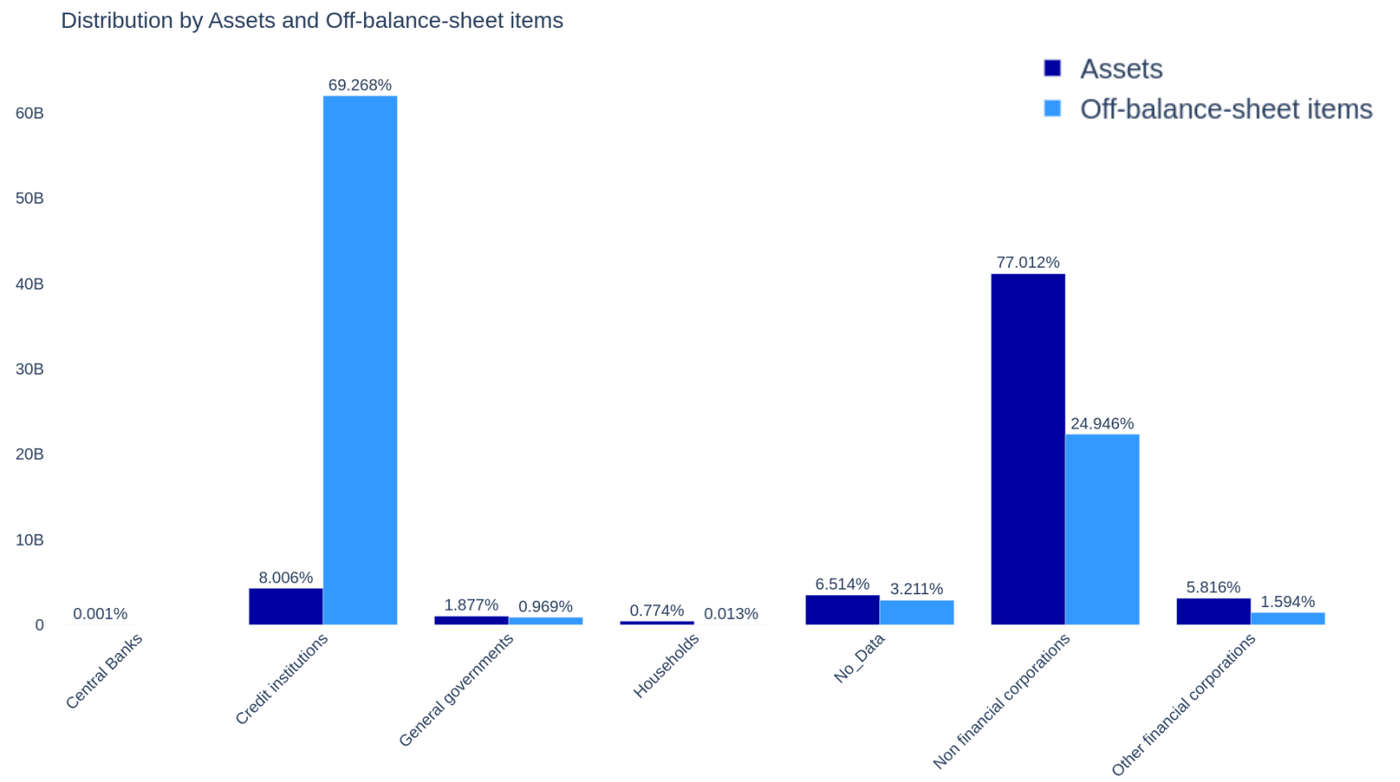
GL
Data regarding account type according to general ledger.

Group_GL
Data regarding account group according to general ledger.

Q2. Explain the assets and off balances for each customer sector category.

Total amounts per sector

Distribution of amount per sector and per assets/off-balance-sheet items



Answer

Most assets:
- Non financial corporation sector 77.01%

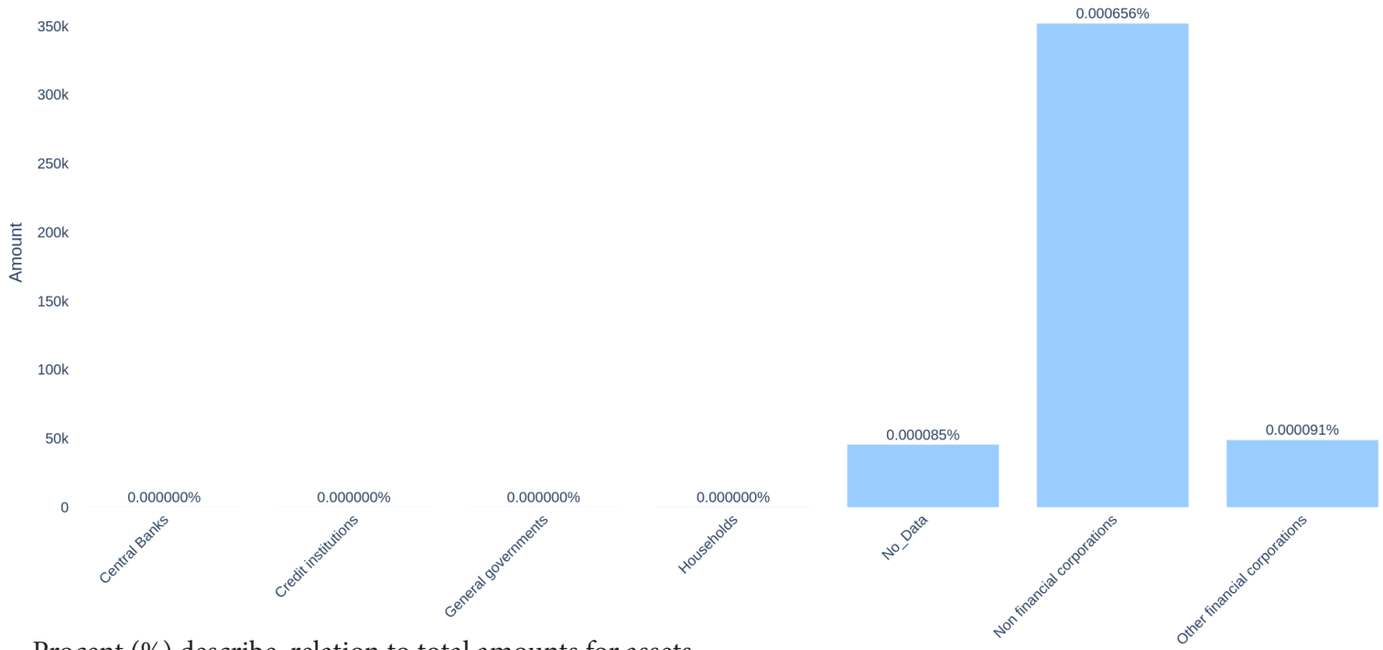
Most off-balance-sheet amounts:
- Credit institutions 69.27%
- Non financial corporation 24.95%

Worth noting that assets and off-balance-sheet items with missing information (No_Data) account for resp. 6.51% and 3.21%

Q3. Explain the total amount for each sector category before and after adjustment.

Adjustments per Sector

Distribution of adjustments per sector



Procent (%) describe relation to total amounts for assets.

Answer

- Adjustment are found in Assets category.
- These adjustments are very small amounts in relation to the total amounts.
- Most adjustments made in NON FINANCIAL CORPORATIONS sector.

Other observations and insights

In the category NON FINANCIAL CORPORATIONS:

- Most amount for positive adjustment (146 650) attributed to ‘accrued interest income’ GL account.
- Most amount for negative adjustment (-229 817) attributed to ‘EIR correction for loans non-demand’ GL account.

Q4a. Explain the amounts aggregated on countries.

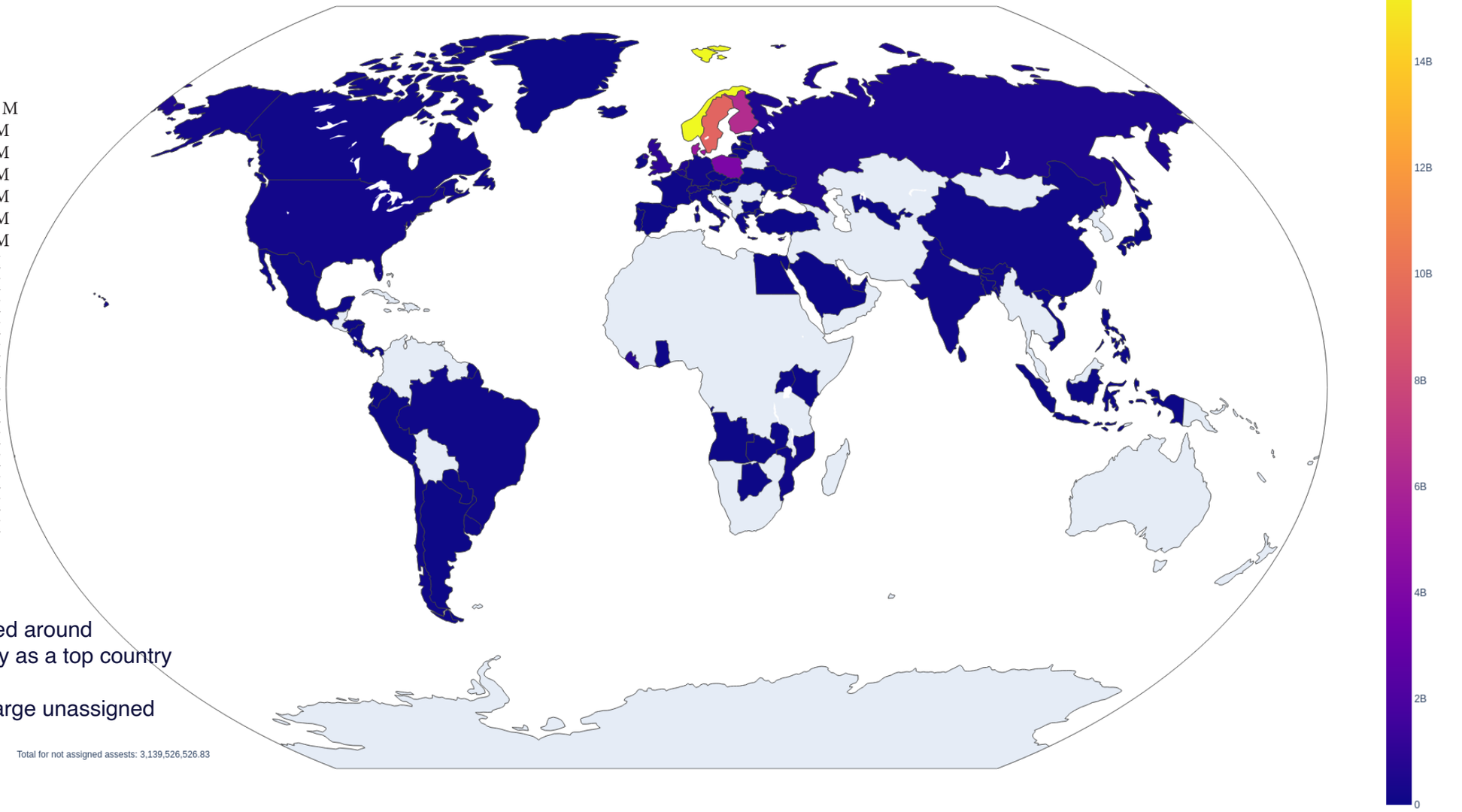
Assets per customers country

Top Countries

1	Norway	15643.92 M
2	Sweden	9410.55 M
3	Finland	6397.21 M
4	Denmark	5150.31 M
5	Poland	3905.86 M
6	No_Data	3139.53 M
7	Bermuda	1200.23 M
8	United Kingdom	968.10 M
9	Jersey	936.39 M
10	Liberia	773.13 M
11	Cyprus	736.66 M
12	Netherlands	560.27 M
13	Russia	498.62 M
14	Marshall Islands	394.48 M
15	Belgium	376.23 M
16	Luxembourg	330.31 M
17	Cayman Islands	329.73 M
18	Switzerland	238.71 M
19	Singapore	205.27 M
20	United States	181.22 M

Insights

- Most assets consentrated around scandinavia with Norway as a top country
- There is a reamrkable large unassigned amount (No_Data)

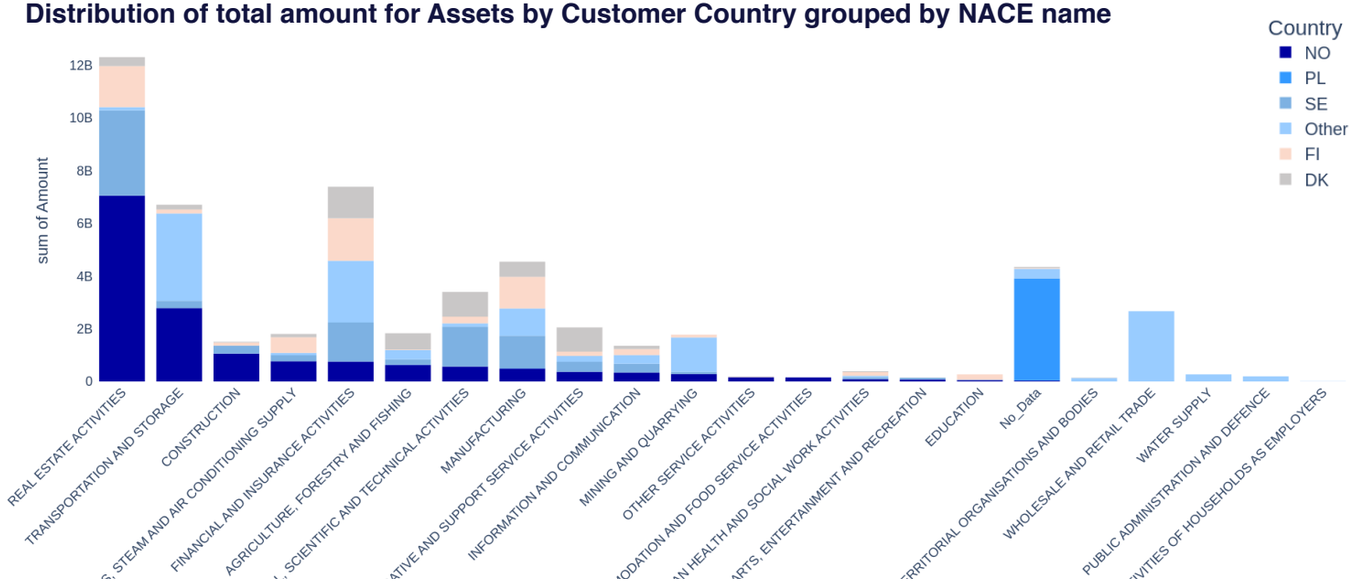


link for interactive map on github: https://htmlpreview.github.io/?https://raw.githubusercontent.com/SiarheiThor/example_projects/main/choropleth_map.html

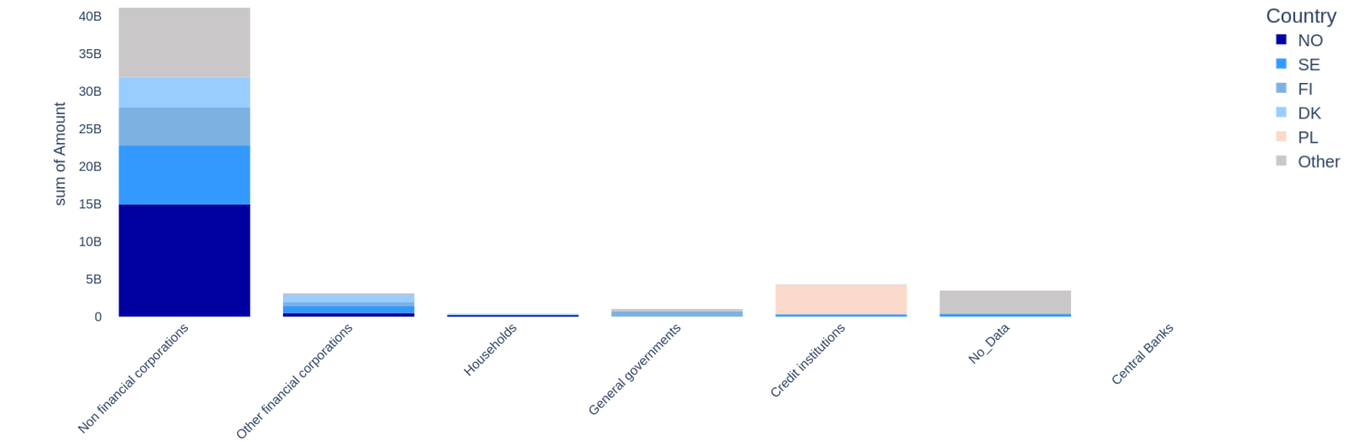
Q4b. Show which country has maximum assets per NACE code and sector category.

Amounts per NACE name

Distribution of total amount for Assets by Customer Country grouped by NACE name



Distribution of total amount for Assets by Customer Country grouped by Sector category



Answer

Maximum assets per NACE code:
Norway, REAL ESTATE ACTIVITIES: 7 055 171 578.3

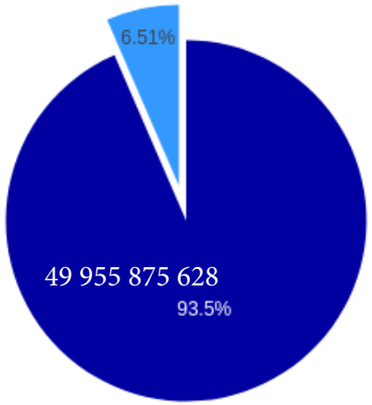
Maximum assets per Sector category:
Norway, Non financial corporations: 14 970 574 545.25

Q5. What is the total assets for missing Customer Responsible Unit? Can you find any trend from customer or instrument perspective?

Customer Responsible Unit (CRU)

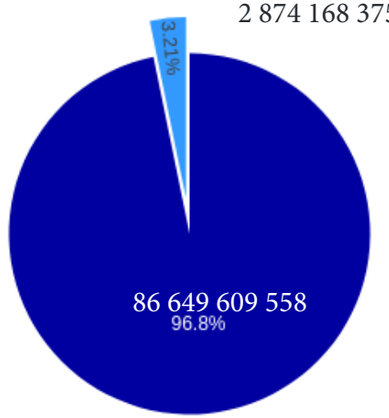
Total Amounts (Assets)

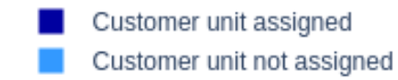
3 480 878 046



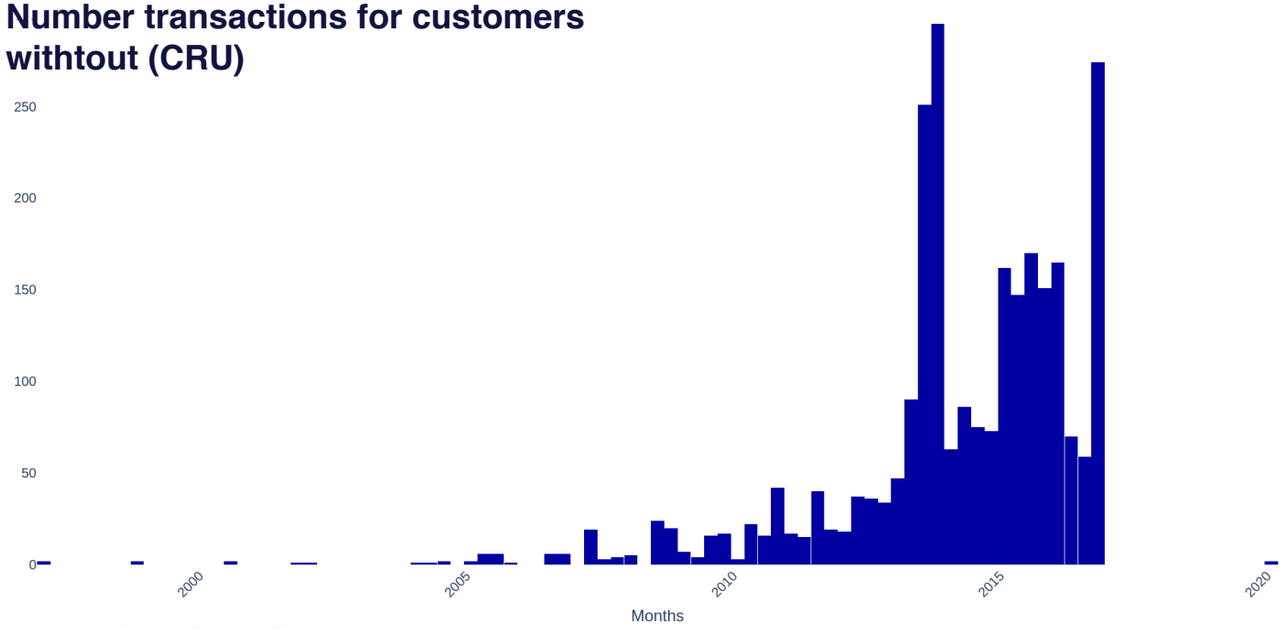
Total Amounts (Off-balance)

2 874 168 375





Number transactions for customers withtout (CRU)



Answer

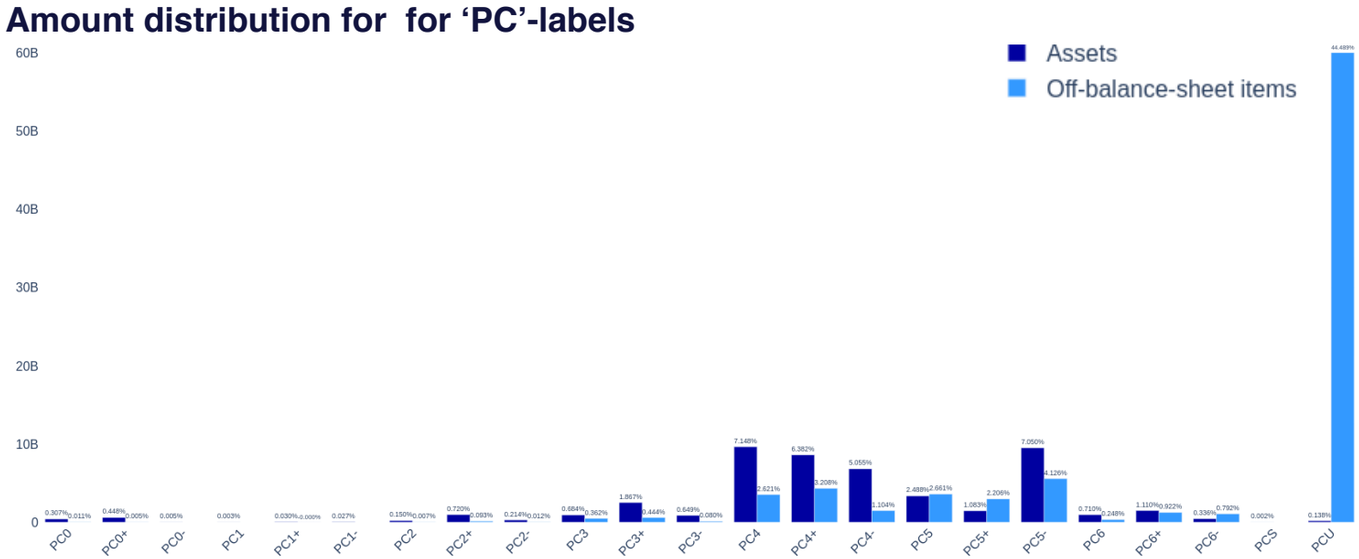
Total assets for missing Customer Responsible Unit:
3 480 878 046.

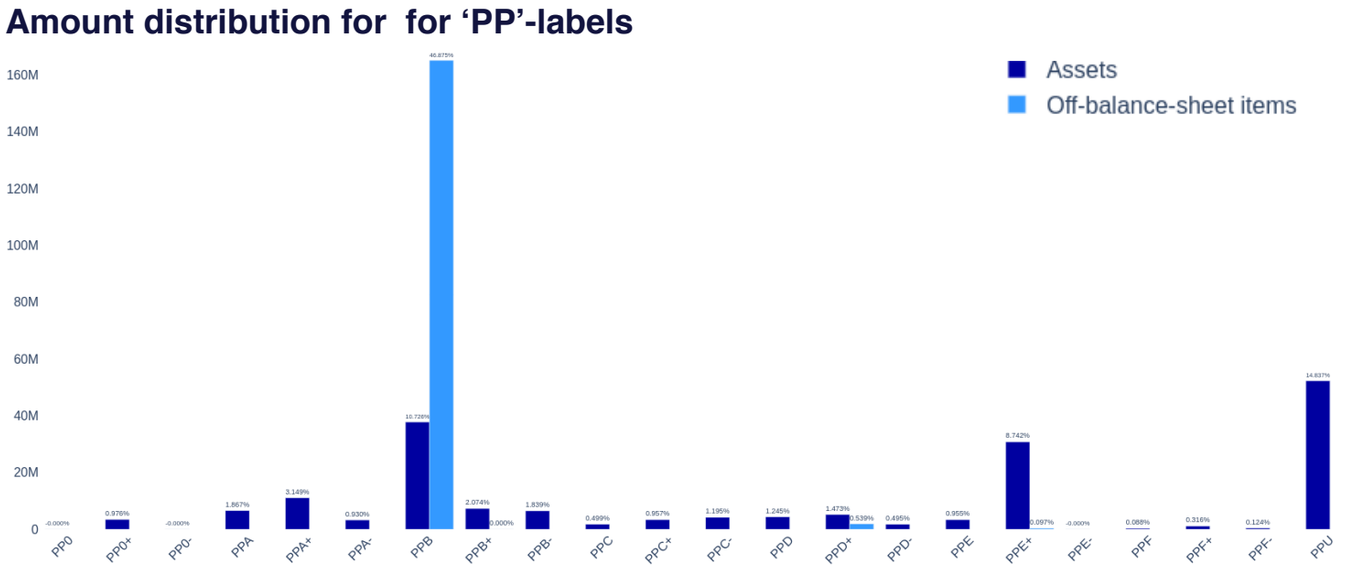
Possible trends

- Majority of clients without data on (CRU) are assigned to 'WHOLESALE AND RETAIL TRADE'
- All client without data on (CRU) are missing the Sector Name.
- Most transaction were made for FINANCING Agreement_ Purpose
- Transactions for customers without (CRU) started to appear around 2005 with a substantial increase around 2011.
- Most of the customers transactions appeared between 2014 and 2016.
- There is some data discrepancy within effective date column, 14 data point have dates around 1900

Q6. Considering the rating/scores <..> Explain identified anomalies with performance of instruments.

Rating Scores





Answer

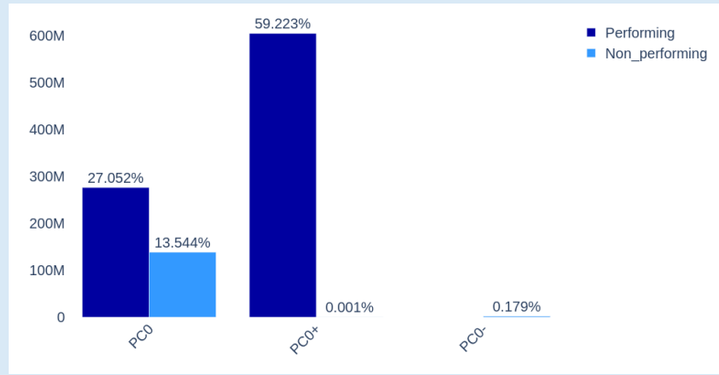
1. There are two statistical outliers (anomalies) within rating PCU and PPB from the off-balance-sheet items GL group.

2. All non-performing instrument are within Assets GL group.

3. Only 14% of bad rated (PC0, PC0+, PC0-) customers accounted as non performing.

4. Most of the assets with low ranking (86%) seem still to perform according to data.

Distribution of performing and Non-performing assets by ranking



Other observations and insights

Ratings across the data have two types of coding: 'PC[0-6,S,U(+)]' and 'PP[0,A-F,U(+)]'.

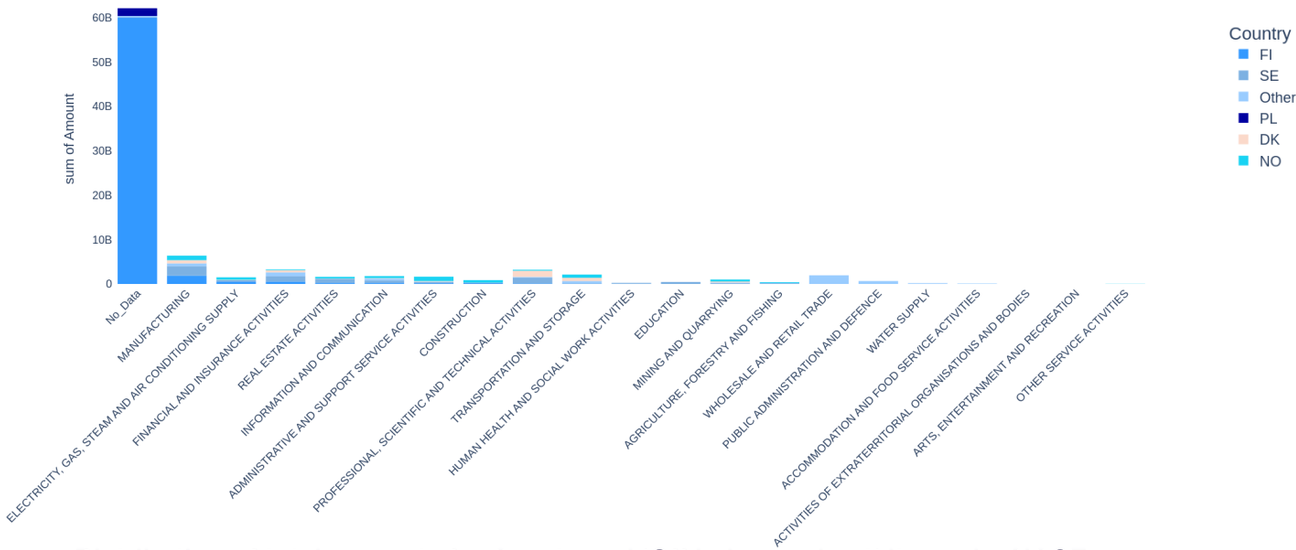
PC ratings a more common with the most amounts distributed between PC3 and PC6

There is a relatively large amount of assets without ranking in PP category - (PPU rating)

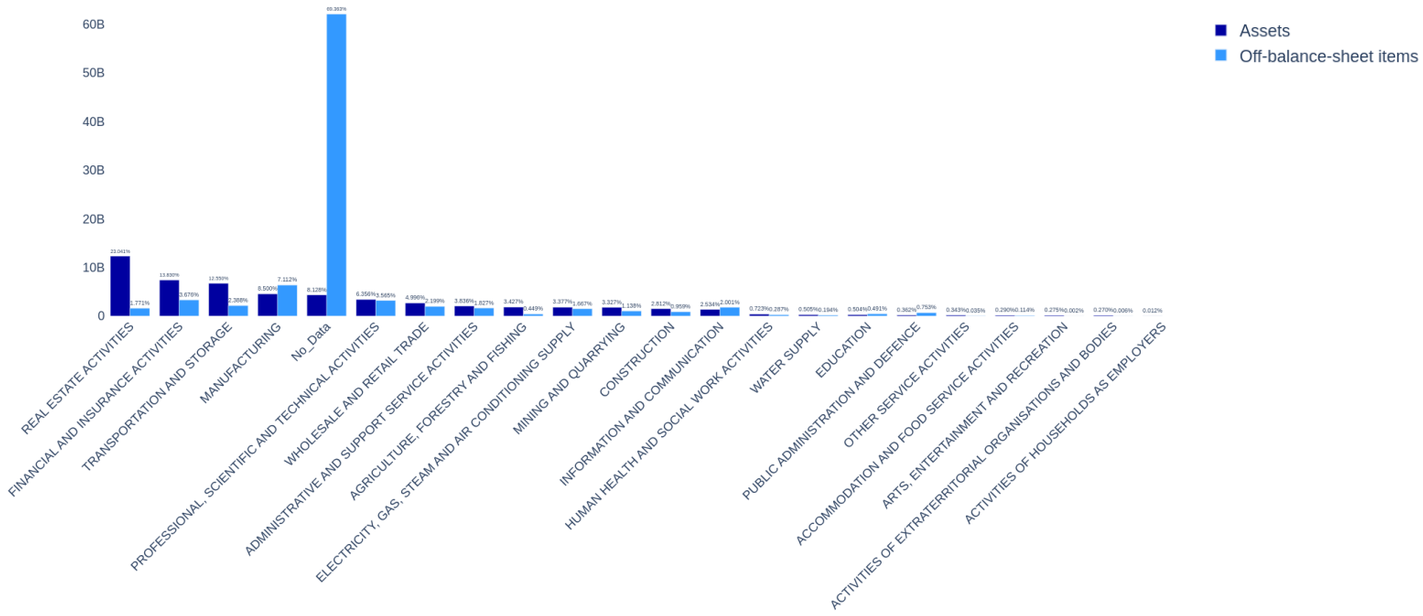
Q7. Explain any other observation about the data.

Amounts per NACE name

Distribution of total amount for ‘Off balance sheet‘ item by Customer Country grouped by NACE category



Distribution of total amounts for Assets and ‘Off balance sheet‘ items by NACE name



Other observations

Maximum off-balance-sheet per NACE code:
Finland, No Data: 60 007 105 838.35

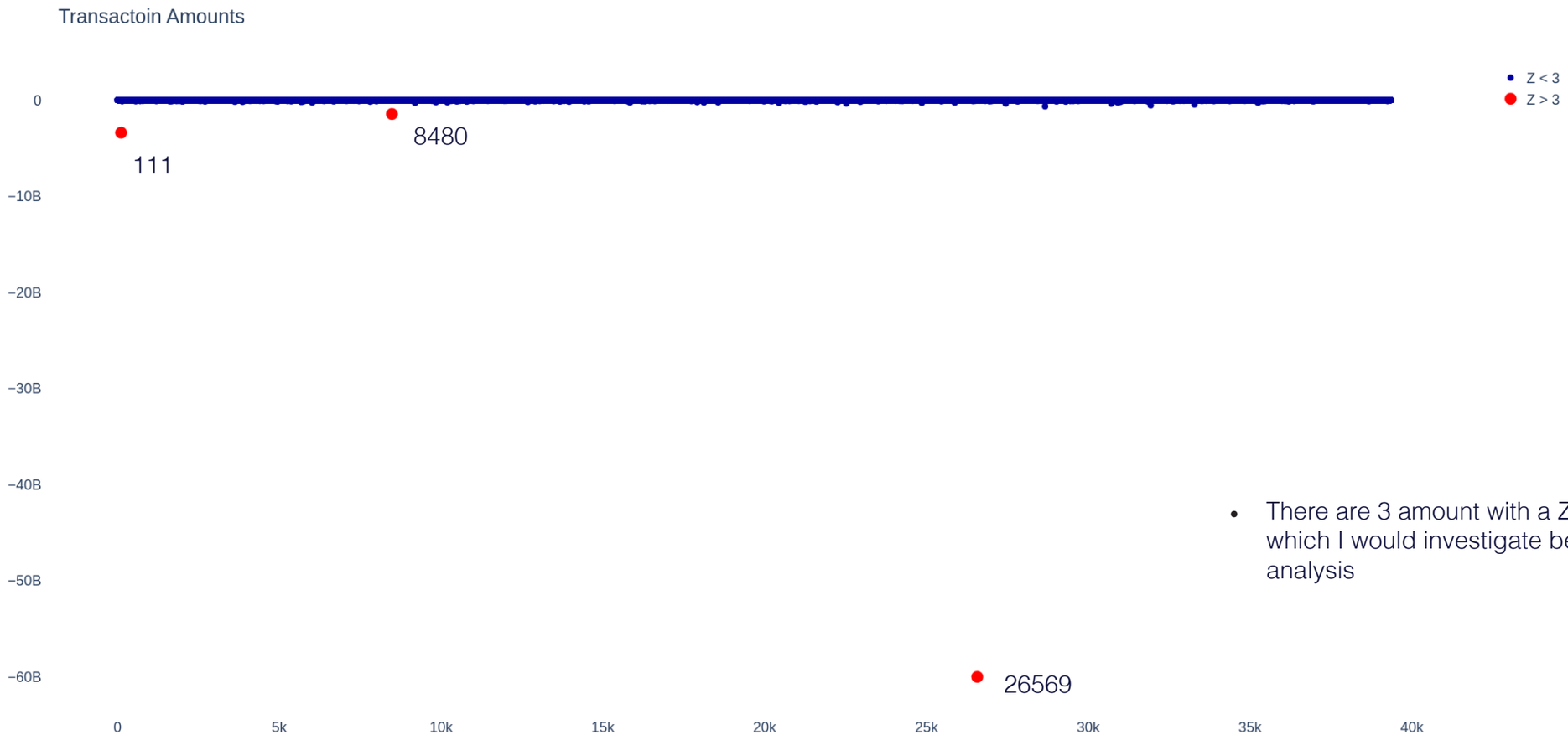
In relation to other NACE categories within both assets and off-balance this amount is far beyond average values.

Insights

Before analysing previous tasks, I would first investigate and deal with the statistical outliers as it creates skewed or disproportional results.

Q7. Explain any other observation about the data.

Outliers detection for Amounts



- There are 3 amount with a Z-score above 3 which I would investigate before doing further analysis

	Agreement_ID	GL_Account_ID	Source	Customer_ID	Amount	Sector_Name	GL_Account_Name
111	1000379280791	1000000008710	229	1000004430504	-3.394932e+09	Credit institutions	Assets
8480	1000174226255	1000000008590	229	1000004430504	-1.453529e+09	Credit institutions	Off-balance-sheet items
26569	1000112687365	1000000008590	229	1000029025316	-6.000000e+10	Credit institutions	Off-balance-sheet items

Thank you!