

# ORBIS Analytics

---

---

*Andrea Lorenzon*

## **Table of Contents:**

### **ORBIS Analytics**

#### **Dataset**

#### **Data import**

#### **ORBIS: Variable description**

BvD.ID.number

Company.name

Country.ISO.Code

Postcode

City

NACE codes

    NACE.Rev..2.main.section

    NACE.Rev..2.Core.code..4.digits.

    NACE.Rev..2.Primary.code.s.

    Cons..code

BvD.Independence.Indicator

BvD.major.sector

Standardised.legal.form

Category.of.the.company

No.of.companies.in.corporate.group

No.of.recorded.shareholders

No.of.recorded.subsidiaries

No.of.recorded.branch.locations

Balance sheet

    Assets

        Fixed.assets.th.EUR.2010

        Intangible.fixed.assets.th.EUR.2010

        Tangible.fixed.assets.th.EUR.2010

        Other.fixed.assets.th.EUR.2010

        Current.assets.th.EUR.2010

        Stock.th.EUR.2010

        Debtors.th.EUR.2010

        Other.current.assets.th.EUR.2010

        Total.assets.th.EUR.2010

    Equity

        Shareholders.funds.th.EUR.2010

        Capital.th.EUR.2010

        Other.shareholders.funds.th.EUR.2010

    Liabilities

        Non.current.liabilities.th.EUR.2010

        Long.term.debt.th.EUR.2010

        Other.non.current.liabilities.th.EUR.2010

        Current.liabilities.th.EUR.2010

        Creditors.th.EUR.2010

Other.current.liabilities.th.EUR.2010  
Income statement  
Cash...cash.equivalent.th.EUR.2010  
Operating.revenue..Turnover..th.EUR.2010  
Sales.th.EUR.2010  
Operating.P.L...EBIT..th.EUR.2010  
Financial.revenue.th.EUR.2010  
Financial.expenses.th.EUR.2010  
Financial.P.L.th.EUR.2010  
P.L.before.tax.th.EUR.2010  
Taxation.th.EUR.2010  
P.L.after.tax.th.EUR.2010  
P.L.for.period...Net.income..th.EUR.2010  
Cash.flow.th.EUR.2010  
Number.of.employees.2010  
Innovation.strength...Number.of.patents  
Innovation.strength...Number.of.inventions  
Number.of.patents  
Number.of.trademarks  
Trademarks...Type  
HGF  
CAGR  
Gazele  
Eurostat  
lat / lon  
trust  
trustVal

## Dataset

```
HGFFinal.merge.csv
Size: 44,7 MB (44745032 bytes)
MD5 checksum: 420d345c68dc3998b8403ab07d0fecf8
```

Note. All maps represents the SUM of the variable grouped by country.

```
n = 115840
```

## Data import

```
import pandas as pd
import geopandas
import descartes
import matplotlib.pyplot as plt
from shapely.geometry import Point, Polygon
import plotly.express as px
```

```

def makePt(x,y):
    if pd.isnull(x):return None
    if pd.isnull(y):return None
    return Point(x,y)

data = pd.read_csv('./HGF-Andrea/HGFFinal.merge.csv', low_memory=False)

# set data types

for var in data.columns:
    try:
        data[var]=data[var].astype(float)
    except:
        pass

# parse to int
for var in ['NACE.Rev..2.Core.code..4.digits.',
            'NACE.Rev..2.Primary.code.s.',
            'No.of.companies.in.corporate.group',
            'No.of.recorded.shareholders',
            'No.of.recorded.branch.locations',
            ]:
    data[var]=data[var].astype(int)

for var in ['Cons..code','NACE.Rev..2.main.section',
            'NACE.Rev..2.Core.code..4.digits.',
            'NACE.Rev..2.Primary.code.s.',
            'BvD.ID.number',
            'BvD.major.sector',
            'Standardised.legal.form',
            'Trademarks...Type',
            'Postcode', 'Postcode2', 'category.of.the.company',
            ]:
    data[var]=data[var].astype('category')

# redundant unique index
data=data.drop('idx', axis=1)
data.set_index('BvD.ID.number')

# order categories for independence indicator

cat_dtype = pd.api.types.CategoricalDtype(
    categories=['U',
                'D-', 'D', 'D+',
                'C-', 'C', 'C+',
                'B-', 'B', 'B+',
                'A-', 'A', 'A+'], ordered=True)

data['BvD.Independence.Indicator']=data['BvD.Independence.Indicator'].astype(cat_dtype)

# null "nan" postcodes
data['Postcode'][data['Postcode']=="nan"]=None

# create Trust value as ordered categorical
data['trustVal']=data['trust'].astype(str).apply(lambda x:
x[0:3]).astype(pd.api.types.CategoricalDtype(
    categories=['Low', 'Med', 'Hig'], ordered=True))

```

```

    ''
for var in ['Company.name', 'Country.ISO.Code',
            'City', 'Postcode2',
            'trust'
            ]:
    data[var]=data[var].astype(pd.StringDtype())

data['HGF']=data['HGF'].astype(bool)

# geolocation
data['point'] = data.apply(lambda row: makePt(row['lon'],row['lat']), axis=1)
geo_data = geopandas.GeoDataFrame(data, geometry='point')
geo_data.crs = {'init' : 'epsg:4326'}

```

# ORBIS: Variable description

## BvD.ID.number

Primary key. Unique (cardinality = n). 2 letters and 8÷12 digits. The two letters appear to be the `Country.ISO.Code`.

```

CategoricalIndex(['IS4203100990', 'GR997722505', 'BG201066368', 'BG201251947',
                  'BG201331418', 'IS4102101180', 'BG201222746', 'BG201124711',
                  'BG201005899', 'BA4281217330002',
                  ...
                  'SK45369747', 'SK45284300', 'SK45349304', 'SK45457824',
                  'SK45432112', 'SK45480371', 'SK45452245', 'SK45407851',
                  'SK45430268', 'SK45418527'],
                  categories=['AT9010104250', 'AT9030242392', 'AT9070278738',
                  'AT9070279036', 'AT9090150166', 'AT9110712698', 'AT9110713446', 'AT9110713447',
                  ...], ordered=False, name='BvD.ID.number', dtype='category', length=115840)

```

## Company.name

`dtype: string`

Contains the firm name. All caps, sometimes includes firm's < forma giuridica? >

BvD.ID.number	
BG201056516	AW TRONICS OOD
IT02637960606	HOME DESIGN - S.R.L.
FR519561369	FINANCIERE HL
SK45371296	EU MANAGEMENT, S.R.O.
FR529210692	TOTAL E&P WELL RESPONSE
Name: Company.name, dtype: string	

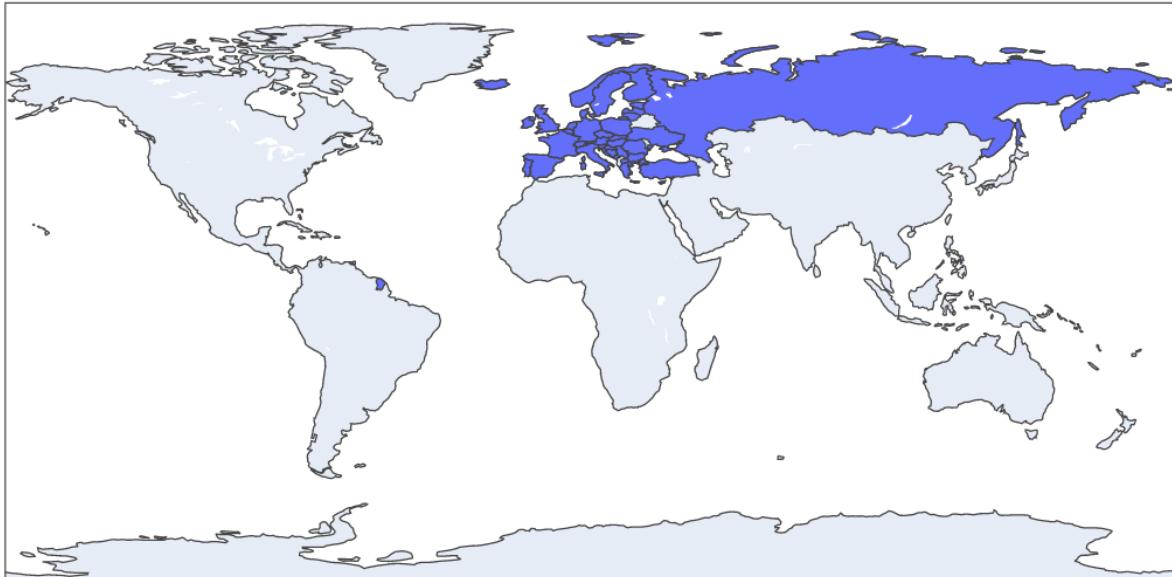
## Country.ISO.Code

`dtype: string`

Contains a two-letter [ISO 3166 alpha-2](#) code from 38 countries. Verify if it's included in the `BvD.ID.number` index.

```
data['Country.ISO.Code'].unique()
```

```
['IS', 'GR', 'BG', 'BA', 'BE', 'IE', 'CY', 'DE', 'AT', 'DK', 'GB', 'CH', 'CZ',
'EE', 'ES', 'FI', 'FR', 'HR', 'HU', 'IT', 'RS', 'PL', 'UA', 'ME', 'NL', 'LU',
'MT', 'MK', 'TR', 'LT', 'LV', 'NO', 'PT', 'RO', 'RU', 'SE', 'SI', 'SK']
Length: 38, dtype: string
```



*Involved countries*

## Postcode

Postcode of the firm. Refers to a different encoding for each country.

```
BvD.ID.number
FR522454743      81640
RU66322917       129085
EE11951827       10151
FR519806830       18700
R027831630        nan
Name: Postcode, dtype: category
Categories (29914, object): [00-024, 00-042, 00-066, 00-102, ..., Y019 6ED, Y026
4GB, Y041 5NS, nan]
```

Present only in 107506 rows. Missing in 8334 rows.

## City

All caps name of the firm's city. Missing in 88 entries.

# NACE codes

NACE uses four hierarchical levels:

- Level 1: 21 sections identified by alphabetical letters A to U;
- Level 2: 88 divisions identified by two-digit numerical codes (01 to 99);
- Level 3: 272 groups identified by three-digit numerical codes (01.1 to 99.0);
- Level 4: 615 classes identified by four-digit numerical codes (01.11 to 99.00).

The first four digits of the code, which is the first four levels of the classification system, are the same in all European countries. National implementations may introduce additional levels. The fifth digit might vary from country to country and further digits are sometimes placed by suppliers of databases.

links: [Reference to all NACE codes](#) , [Wikipedia: NACE codes](#).

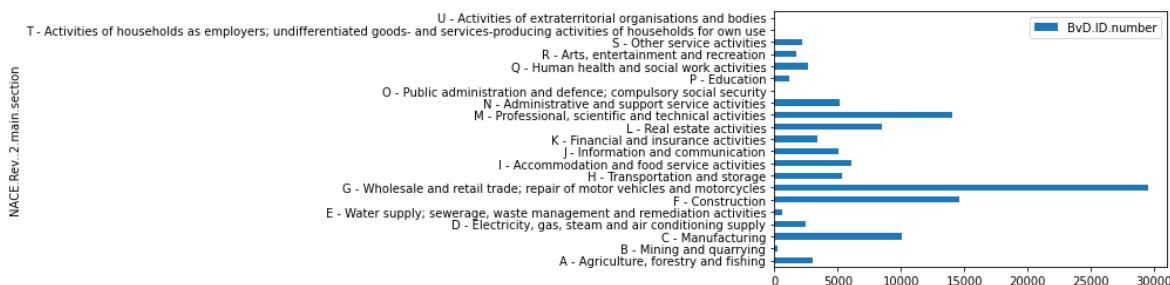
## NACE.Rev..2.main.section

Level 1 NACE code. A letter, and the sector description.

```
BvD.ID.number
FR521201111 F - Construction
FR523714624 G - Wholesale and retail trade; repair of moto...
ESB85891794 S - Other service activities
IT03972880235 I - Accommodation and food service activities
PT509412866 M - Professional, scientific and technical act...
Name: NACE.Rev..2.main.section, dtype: category
```

All 21 sections are being represented.

```
pd.DataFrame(data['NACE.Rev..2.Core.code..4.digits.']).reset_index().groupby('NA
CE.Rev..2.main.section').count().plot.barh()
```



number of entries per section

## NACE.Rev..2.Core.code..4.digits.

4 digit NACE code. 729 different categories.

```

BvD.ID.number
N0995138697      4110
PT509493599       4339
R027726219       4711
FR520957143       161
IT03102890831     4120
Name: NACE.Rev..2.Core.code..4.digits., dtype: category
Categories (729, int64): [100, 110, 111, 112, ..., 9609, 9700, 9810, 9900]

```

## NACE.Rev..2.Primary.code.s.

4 digit NACE code. Similar to the former? maybe duplicate? 729 different categories.

```

BvD.ID.number
RU65230449      4633
N0995818558      4759
SE5568024276      7311
N0995182165      3312
R027703886       4613
Name: NACE.Rev..2.Primary.code.s., dtype: category
Categories (729, int64): [100, 110, 111, 112, ..., 9609, 9700, 9810, 9900]

```

## Cons..code

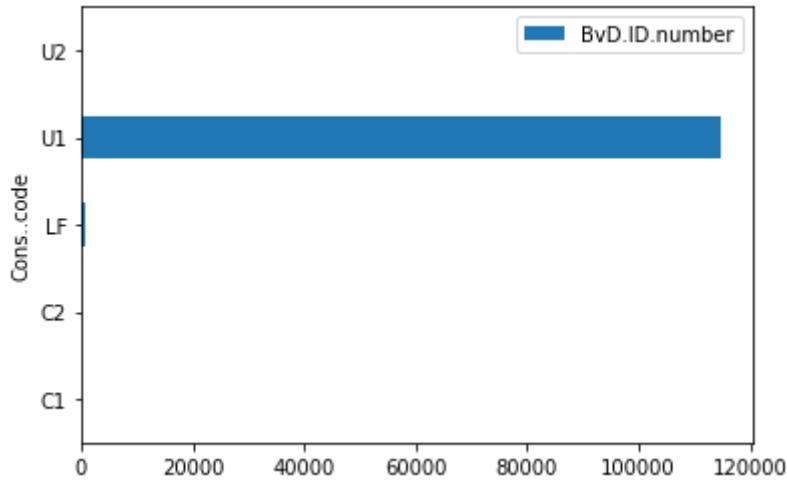
### Bankscope consolidation code

- **C1:** statement of a mother company integrating the statements of its controlled subsidiaries or branches with no unconsolidated companion,
- **C2:** statement of a mother company integrating the statements of its controlled subsidiaries or branches with an unconsolidated companion,
- **U1:** statement not integrating the statements of the possible controlled subsidiaries or branches of the concerned company with no consolidated companion.
- **U2:** statement not integrating the statements of the possible controlled subsidiaries or branches of the concerned company with an consolidated companion.
- **LF:** limited financials: information based on rounded figures officially available, sometimes collected from other directories or websites.

```

BvD.ID.number
R026440005      U1
IT03232890982     U1
IT11001531000     U1
RU67267304       LF
PT509588980      U1
Name: Cons..code, dtype: category
Categories (5, object): [C1, C2, LF, U1, U2]

```

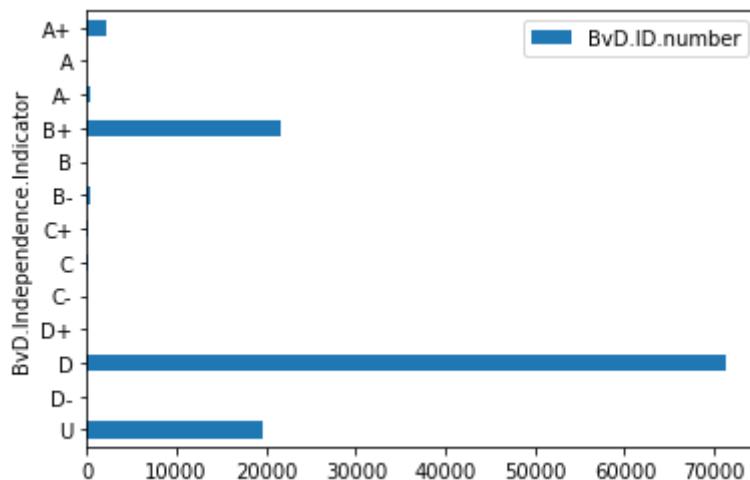


## BvD.Independence.Indicator

Characterize the degree of independence of a company with regard to its shareholders. It has been mapped to ad ordered category, with null value (U) being set at the lowest value. *Check difference between U e not rated.*

links: [Variable description](#)

```
BvD.ID.number
R026798381      D
RU67068071       D
FR527515381     U
RU64795818       D
IT06649101216   D
Name: BvD.Independence.Indicator, dtype: category
Categories (13, object): [U < D- < D < D+ ... B+ < A- < A < A+]
```



## BvD.major.sector

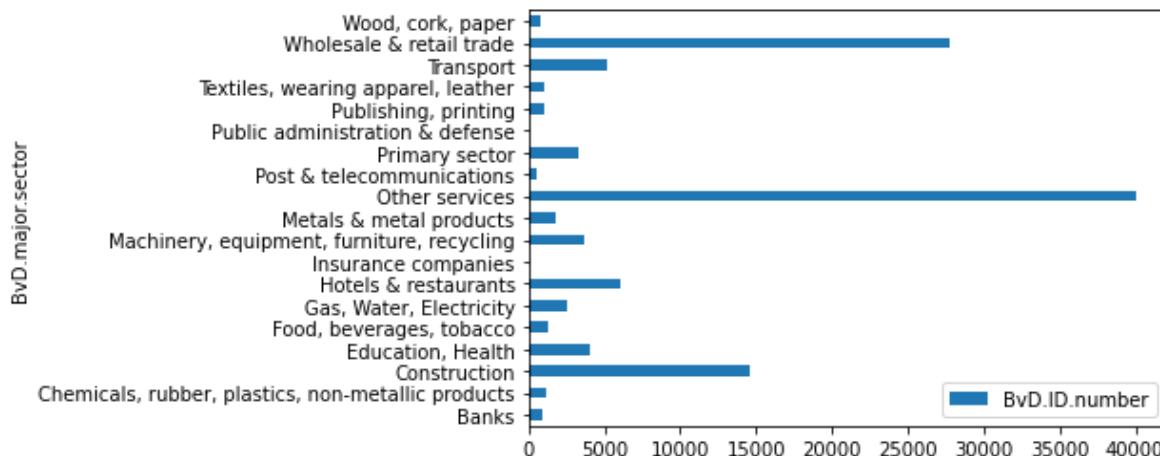
A different sector encoding from BvD. It encompass 19 categories.

```

BvD.ID.number
R026474245          Other services
HR76526891156      Chemicals, rubber, plastics, non-metallic prod...
RU68874348          Textiles, wearing apparel, leather
CZ28117956          Primary sector
R015587044         Other services
Name: BvD.major.sector, dtype: category
Categories (19, object): [Banks, Chemicals, rubber, plastics, non-metallic
prod..., Construction, Education, Health, ..., Textiles, wearing apparel,
leather, Transport, Wholesale & retail trade, Wood, cork, paper]

```

There are no missing values.



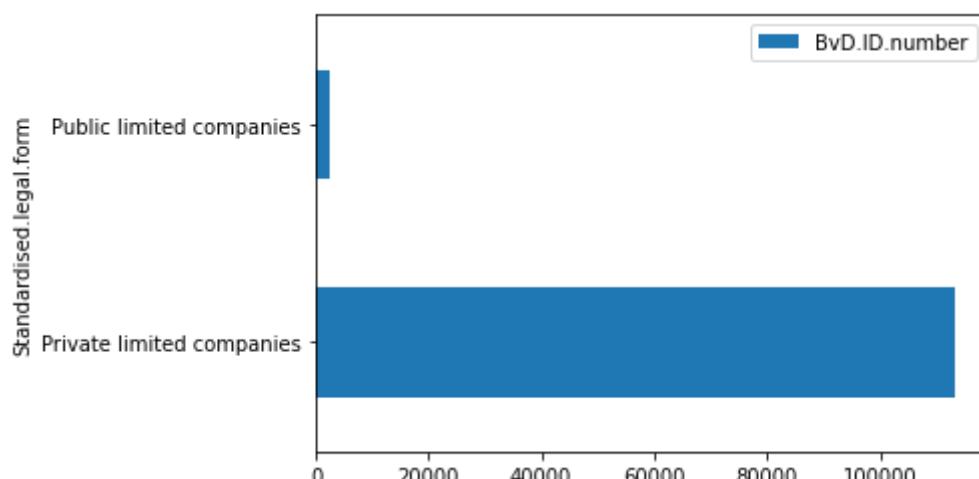
## Standardised.legal.form

Two level factor, stating if the company is public or private.

```

BvD.ID.number
ESB85955946    Private limited companies
FR519336481    Private limited companies
PT509284035    Public limited companies
EE11920653     Private limited companies
N0995237563    Private limited companies
Name: Standardised.legal.form, dtype: category
Categories (2, object): [Private limited companies, Public limited companies]

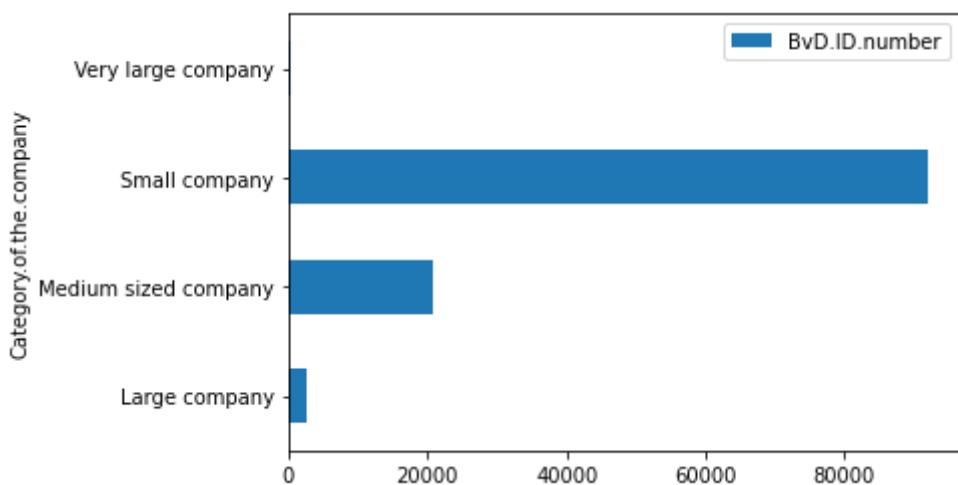
```



## Category.of.the.company

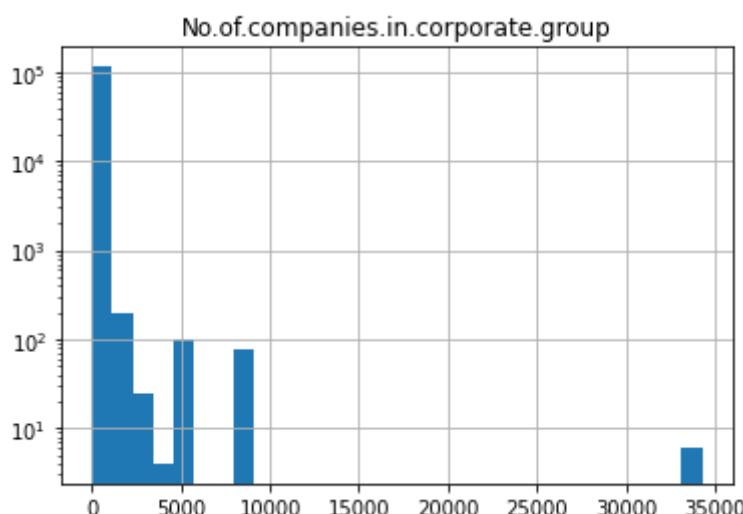
4 level factor stating the dimension of the company. Could not find objective inclusion criteria anywhere.

```
BvD.ID.number
R026541921           Small company
R026543973           Small company
R027249764          Medium sized company
PT509553923           Small company
SE5568008675          Small company
Name: Category.of.the.company, dtype: category
Categories (4, object): [Large company, Medium sized company, Small company,
Very large company]
```



## No.of.companies.in.corporate.group

Number of companies in the corporate group. The largest part of entries has 0 companies in the group (assuming: no group).



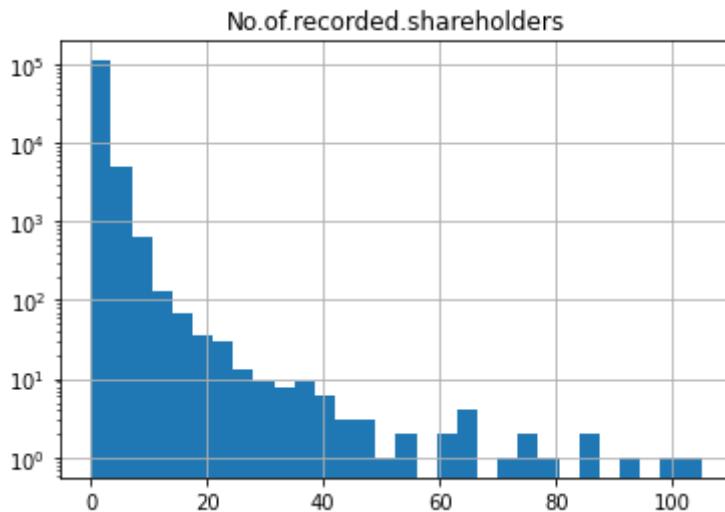
outliers:

BvD.ID.number	
IE488184	SKY HIGH III LEASING DESIGNATED ACTIVITY COMPANY
IT07063570969	HB SERVIZI S.R.L
IT07182390968	POLIAMBULATORIO BICOCCA S.R.L.
N0995590271	ELKEM RANA AS
R01590899	ADAMA AGRICULTURAL SOLUTIONS SRL
R025221180	EDPR ROMANIA SRL
Name: Company.name, dtype: string	

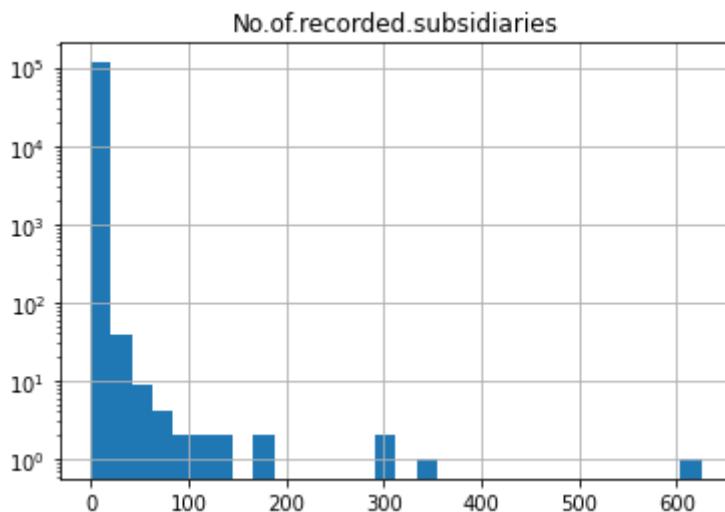
Note: log scale on x axis.

## No.of.recorded.shareholders

Numbers of recorded stakeholders.



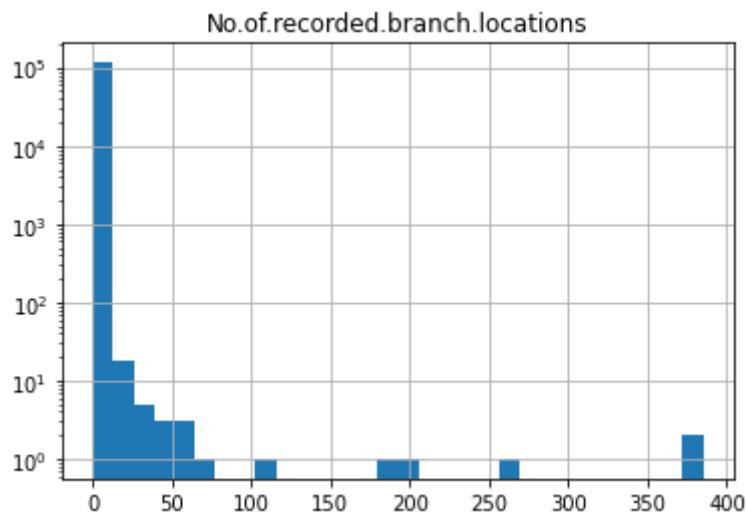
## No.of.recorded.subsidiaries



outlier:

BvD.ID.number	
RS20661283	AKCIONARSKI FOND
Name: Company.name, dtype: string	

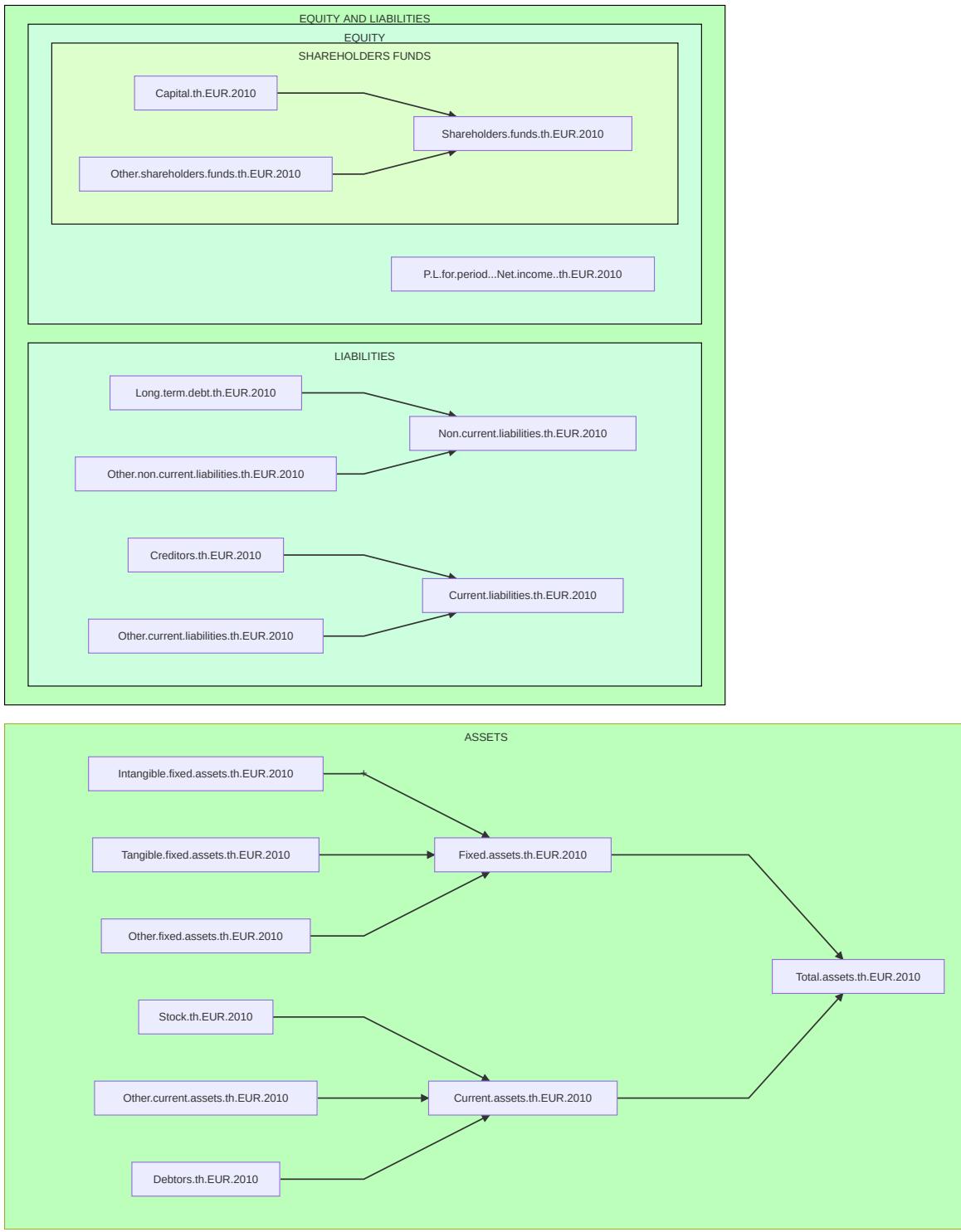
## No.of.recorded.branch.locations



outliers:

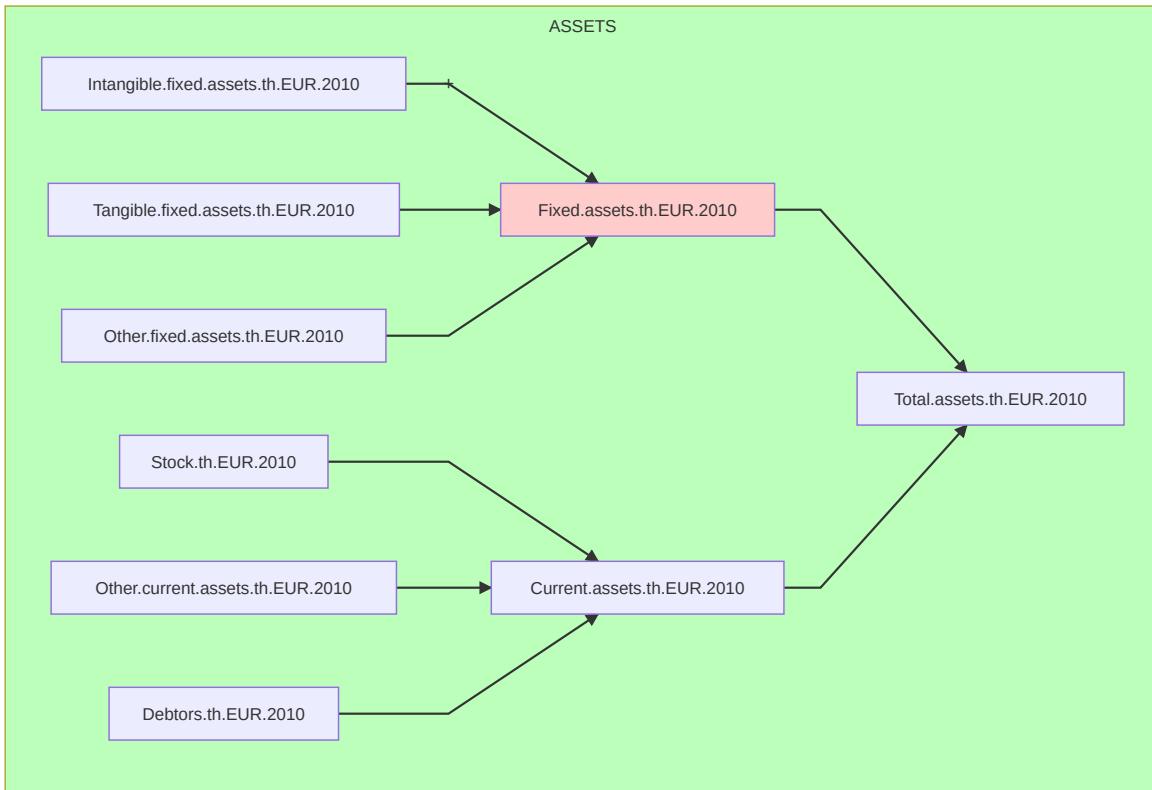
```
BvD.ID.number
FR528648892          CHAUSSON MATERIAUX
FR524237351      S A S LOT AGRICULTURE ET ENERGIE SOLAIRE
Name: Company.name, dtype: string
```

## Balance sheet

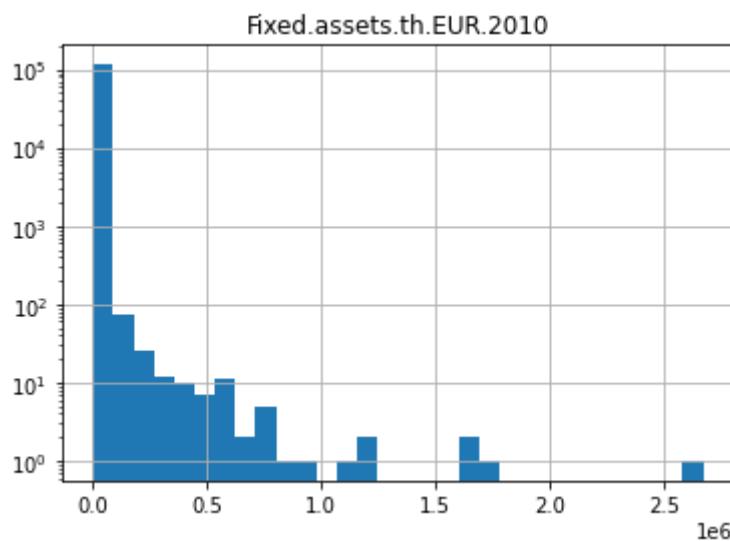


## Assets

### Fixed.assets.th.EUR.2010



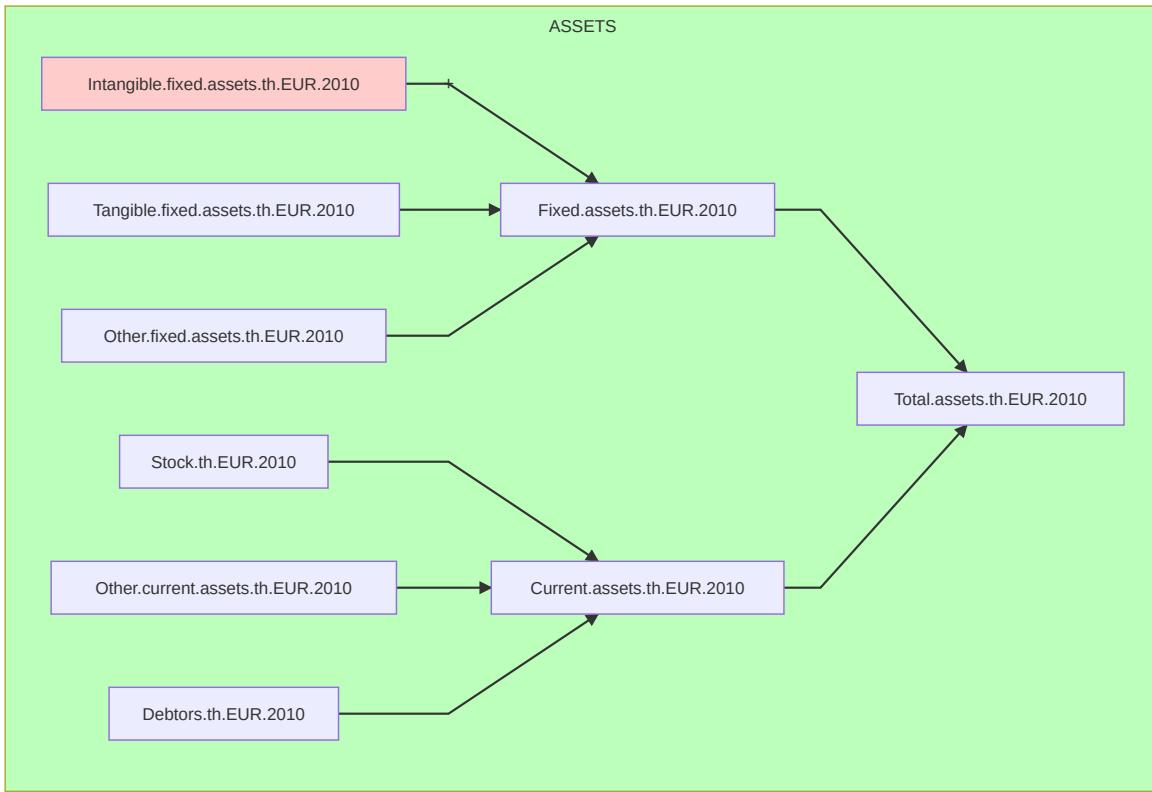
A fixed asset is a long-term tangible piece of property or equipment that a firm owns and uses in its operations to generate income. Fixed assets are not expected to be consumed or converted into cash within a year. Fixed assets most commonly appear on the balance sheet as [property, plant, and equipment](#) (PP&E).



```

outlier:
BvD.ID.number
IE486605      SAP IRELAND US-FINANCIAL SERVICES DESIGNATED A...
Name: Company.name, dtype: string
  
```

## Intangible.fixed.assets.th.EUR.2010



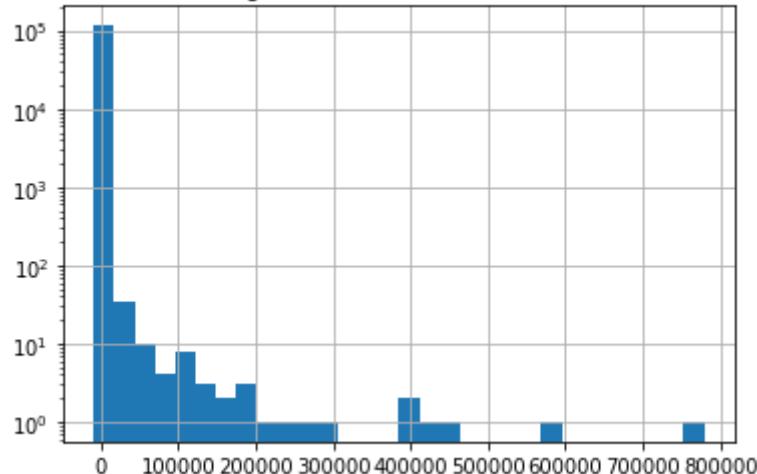
Intangible assets include operational assets that lack physical substance. For example, goodwill is a fixed asset, as are patents, copyrights, trademarks and franchises. A company's intangible assets are often not reported on a company's financial statements, or they may be reported at significantly less than their actual value. This is because assets are accounted for at their historical cost.

Unlike tangible fixed assets such as a building or machinery, intangibles are often developed internally without any direct, measurable cost that can be capitalized. When an intangible is purchased, however, or when costs can be directly traced to the development of the asset, the cost is recorded as an intangible asset on the balance sheet.

Intangible assets are valued at their cost of acquisition. A purchased intangible is valued based on the amount paid for the asset. Research and development costs associated with developing an intangible are expensed for the year in which they were incurred.

However, costs of registering patents or trademarks and legal fees incurred to defend a company's right of use are included in the cost of acquisition, which is reported as an intangible asset on the balance sheet.

### Intangible.fixed.assets.th.EUR.2010

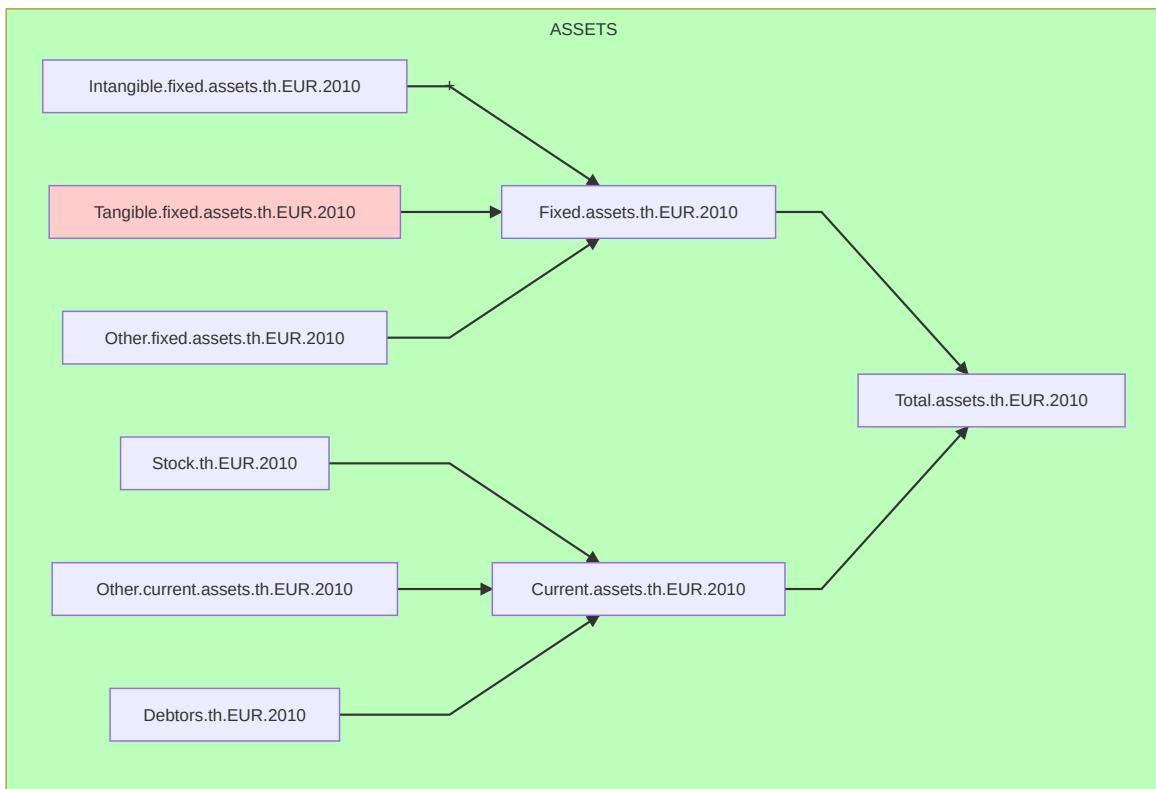


x 10<sup>3</sup> €

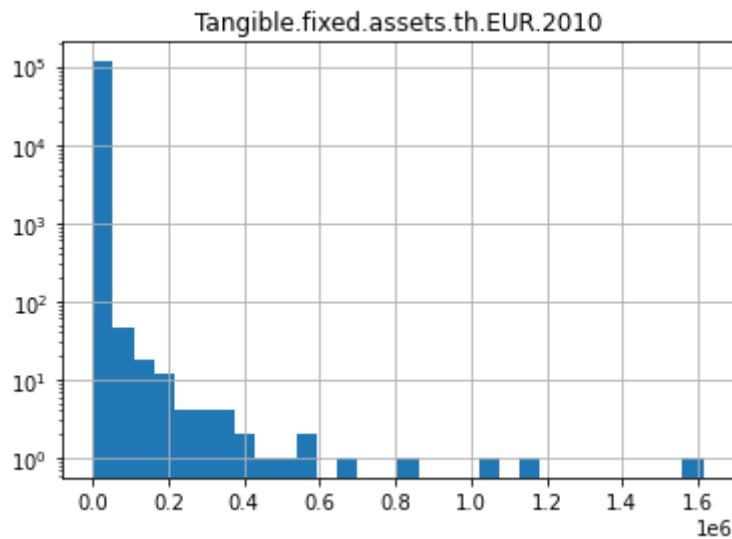
outlier:

```
BvD.ID.number  
IT10969001006      LOTTERIE NAZIONALI S.R.L.  
Name: Company.name, dtype: string
```

### Tangible.fixed.assets.th.EUR.2010



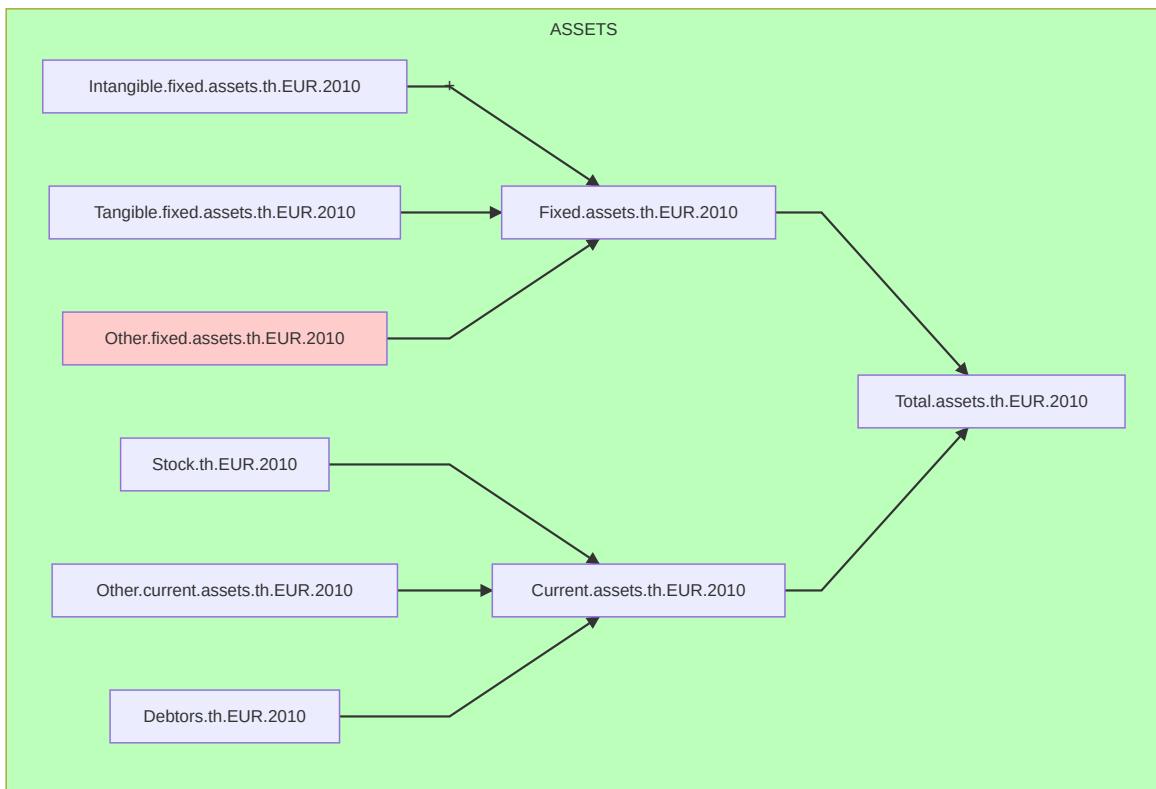
Tangible fixed assets generally refer to assets that have a physical value. Examples of this are your business premises, equipment, inventory and machinery. Tangible fixed assets have a market value that needs to be accounted for when you file your annual accounts. Some of these assets, for example computer equipment, will incur depreciation, which needs to be factored into your accounts.

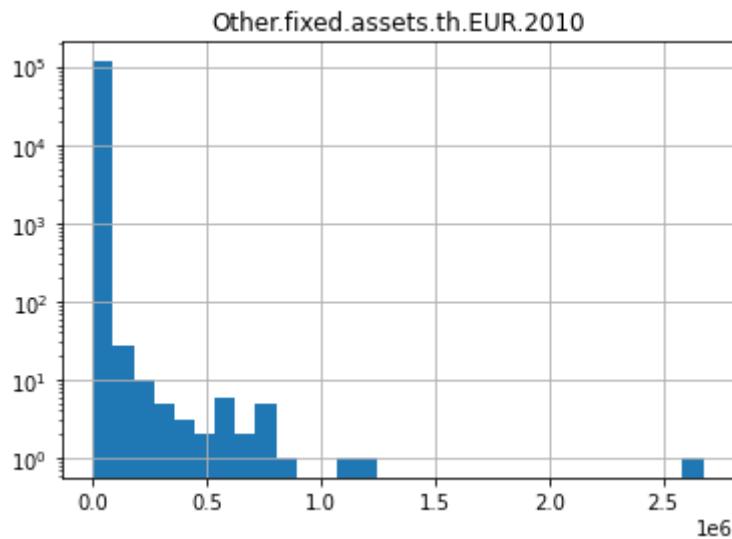


outlier:

```
BvD.ID.number
GB07145051    CAPITAL & COUNTIES PROPERTIES PLC
Name: Company.name, dtype: string
```

## Other.fixed.assets.th.EUR.2010

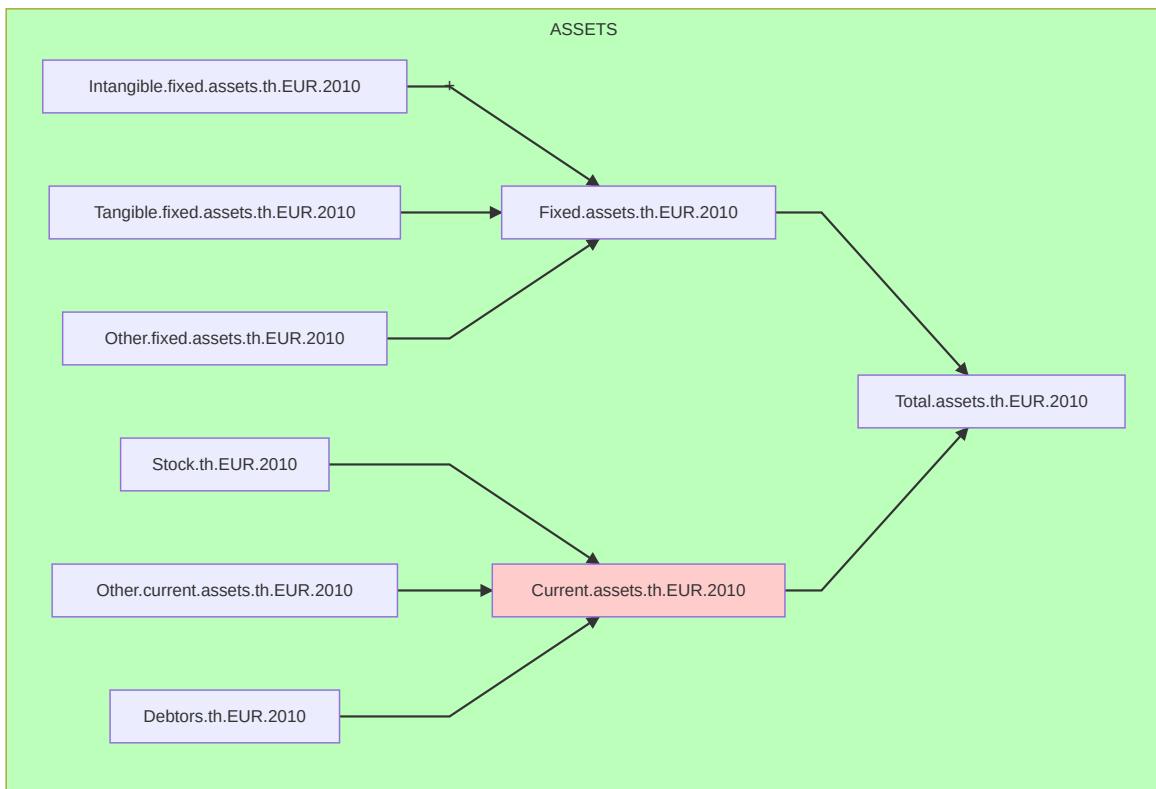


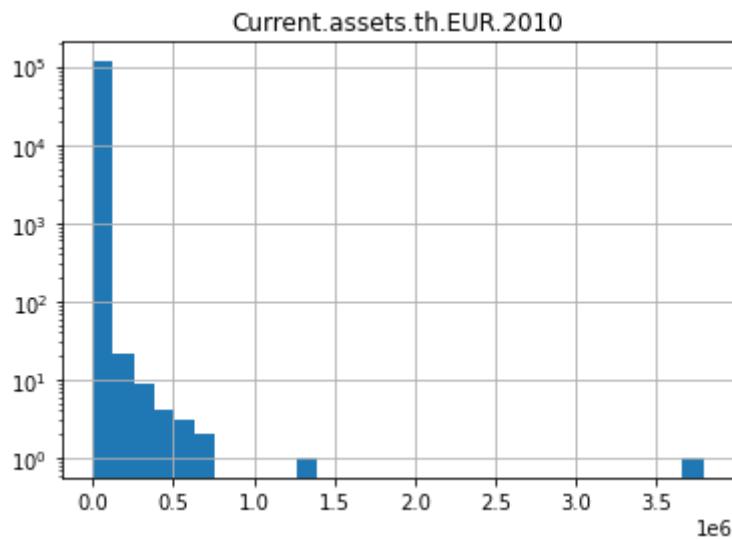


outlier:

```
BvD.ID.number
IE486605    SAP IRELAND US-FINANCIAL SERVICES DESIGNATED A...
Name: Company.name, dtype: string
```

## Current.assets.th.EUR.2010



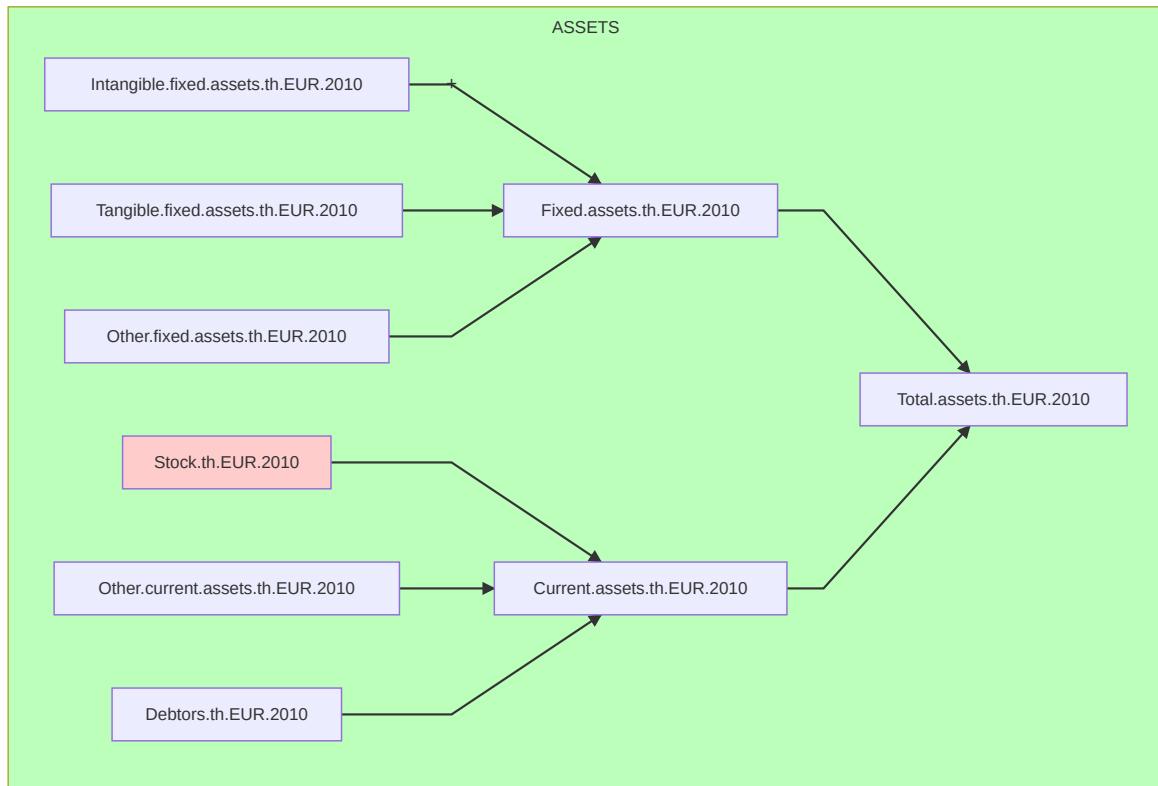


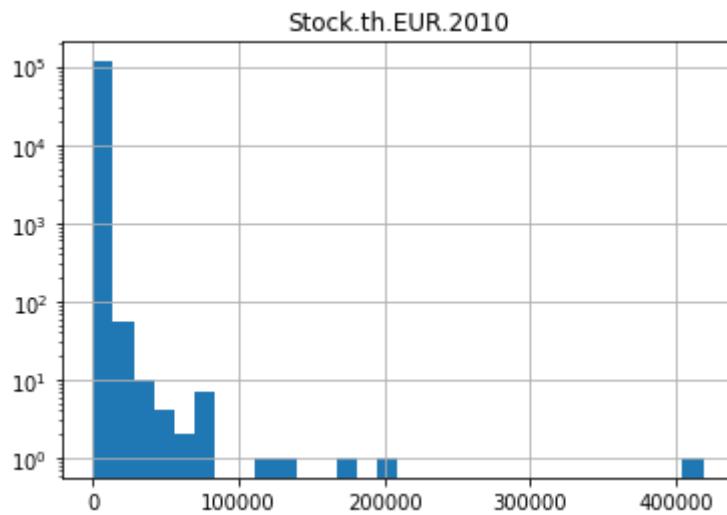
```

outlier:
BvD.ID.number
GB07450219      LONG ISLAND ASSETS LIMITED
Name: Company.name, dtype: string

```

## Stock.th.EUR.2010

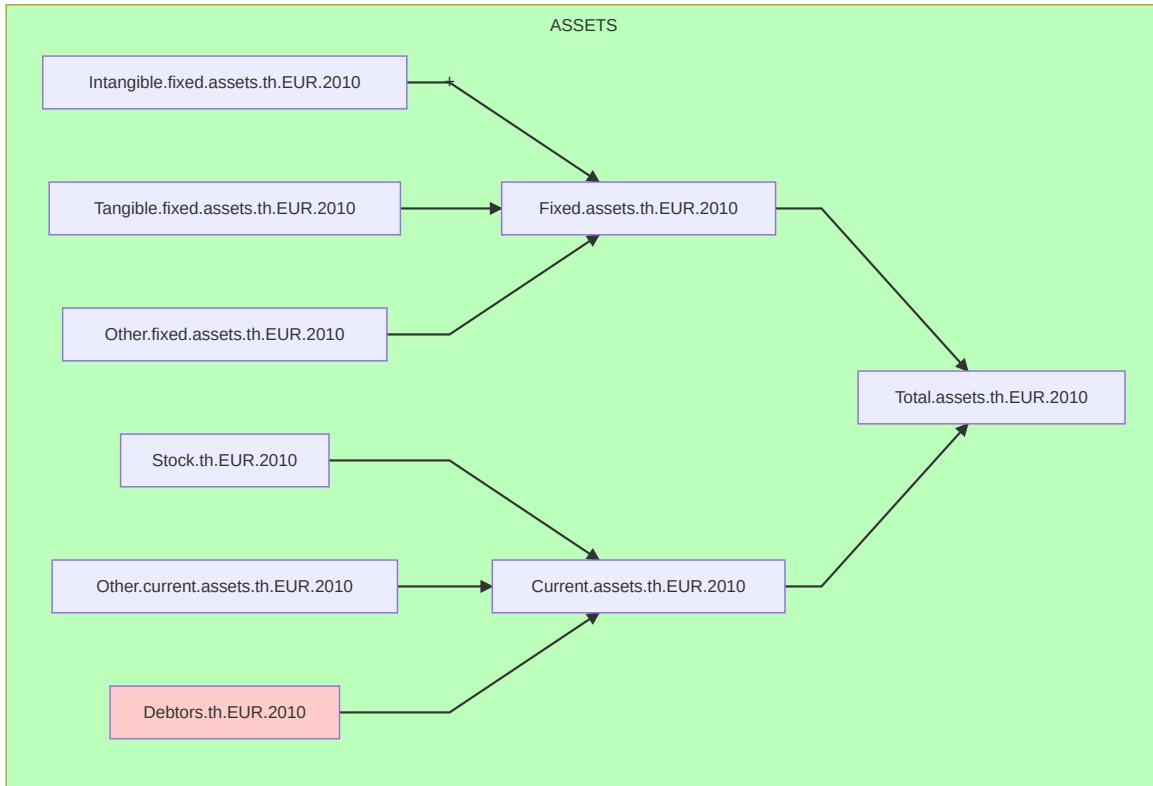


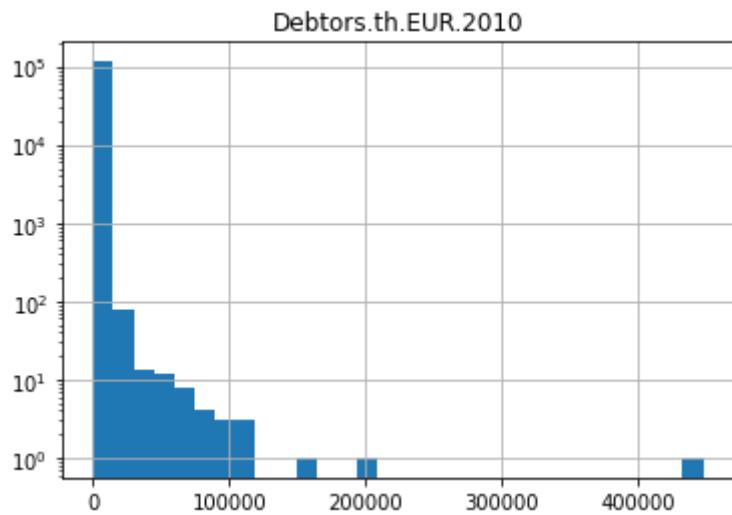


outlier:

```
BvD.ID.number
IT07099900966    MILANOESTO S.P.A.
Name: Company.name, dtype: string
```

## Debtors.th.EUR.2010

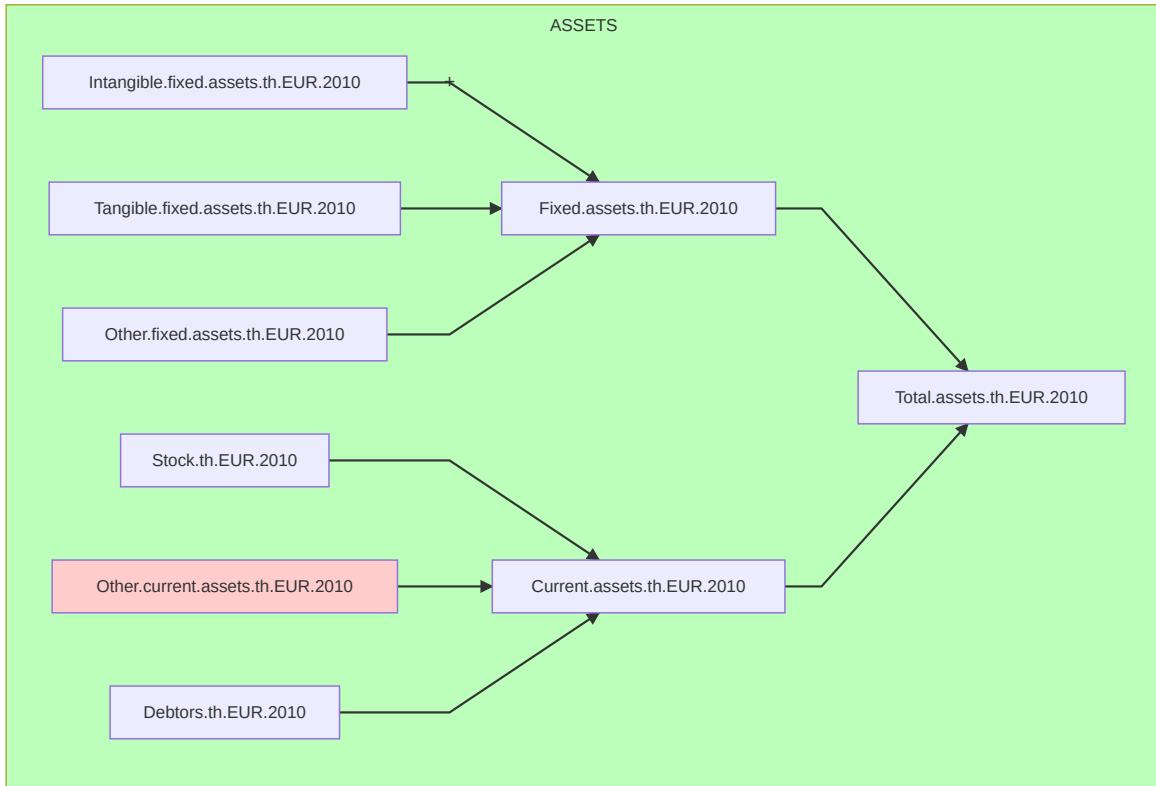


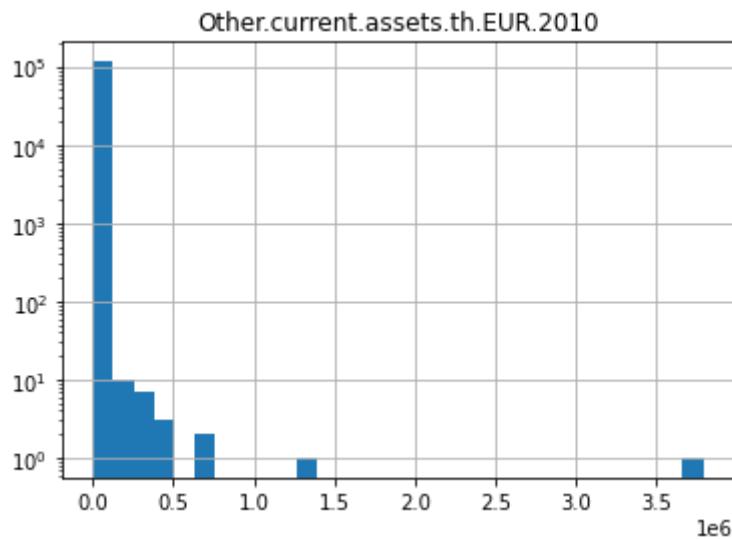


outlier:

```
BvD.ID.number
GB07246104      MACSCO 22 LIMITED
Name: Company.name, dtype: string
```

## Other.current.assets.th.EUR.2010

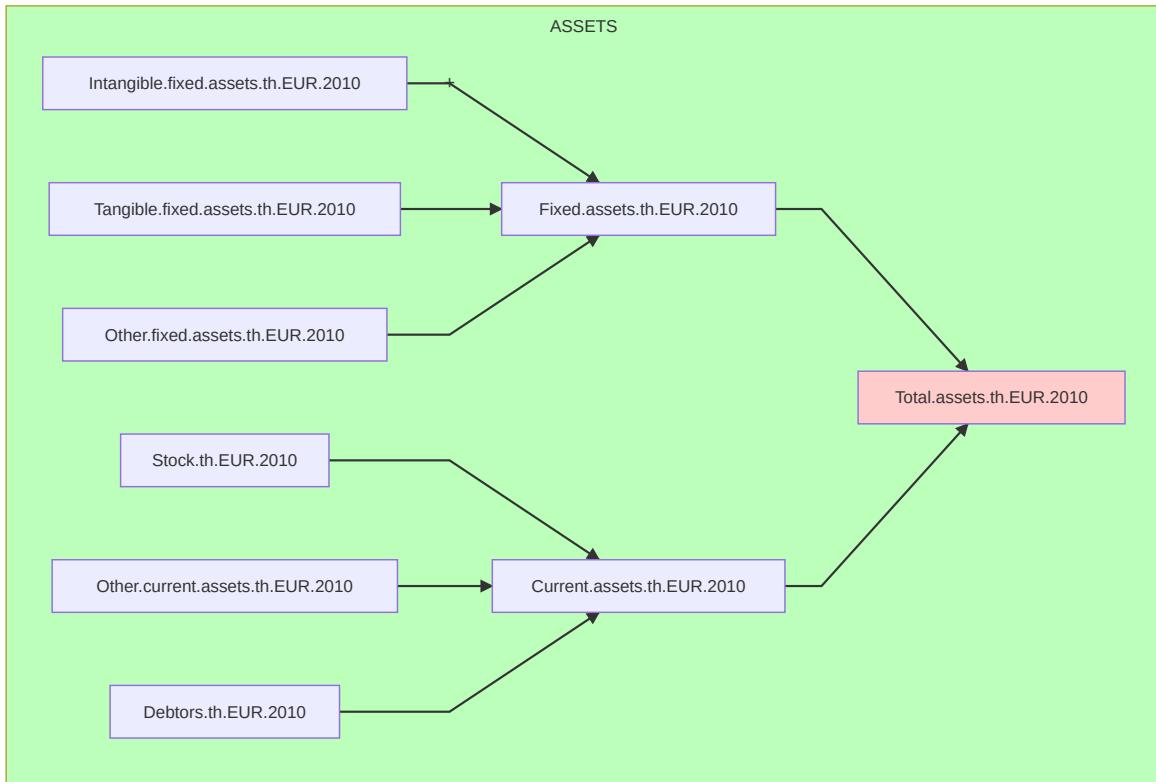


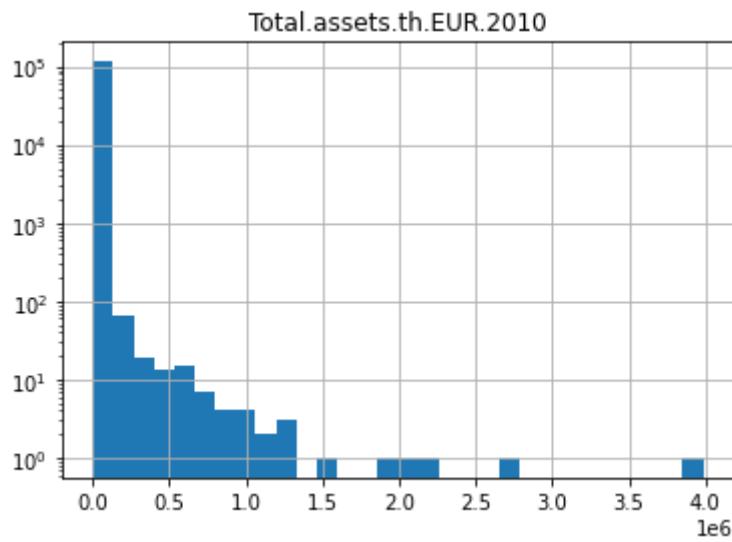


outlier:

```
BvD.ID.number
GB07450219    LONG ISLAND ASSETS LIMITED
Name: Company.name, dtype: string
```

## Total.assets.th.EUR.2010



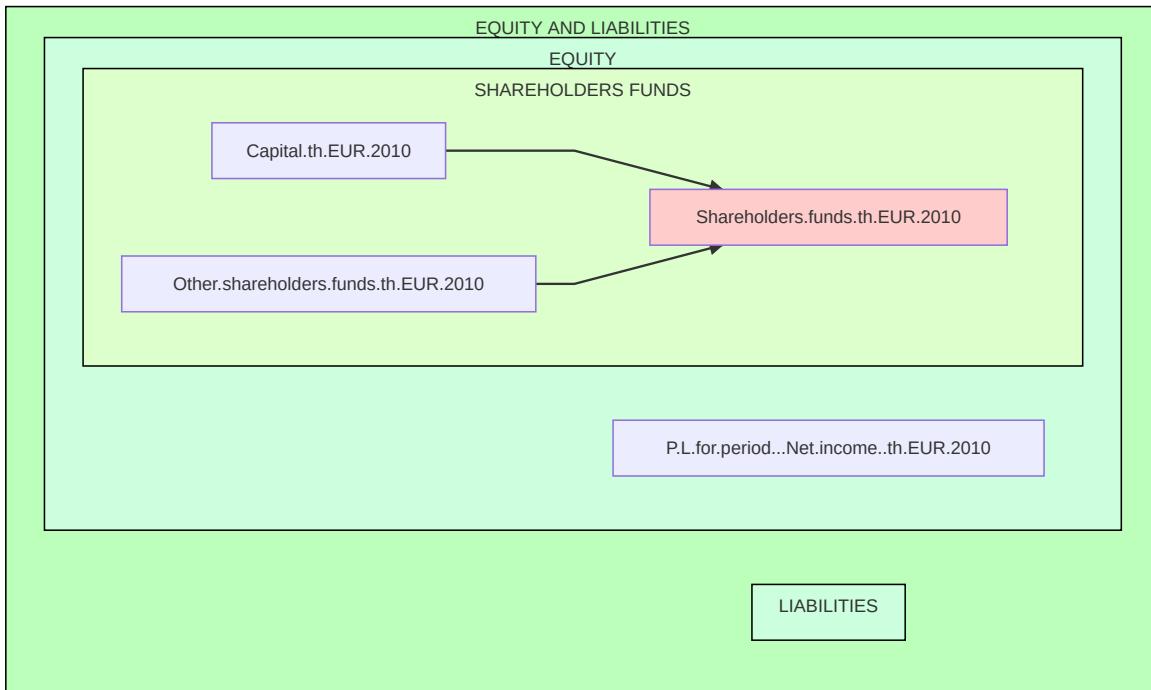


outlier:

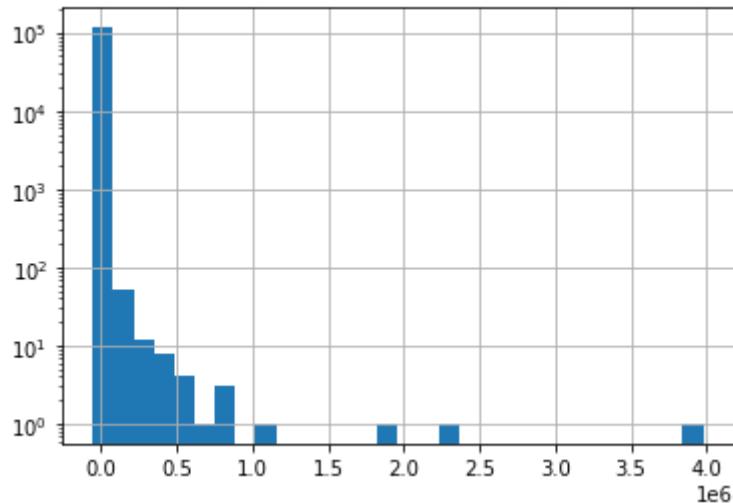
```
BvD.ID.number
GB07450219    LONG ISLAND ASSETS LIMITED
Name: Company.name, dtype: string
```

## Equity

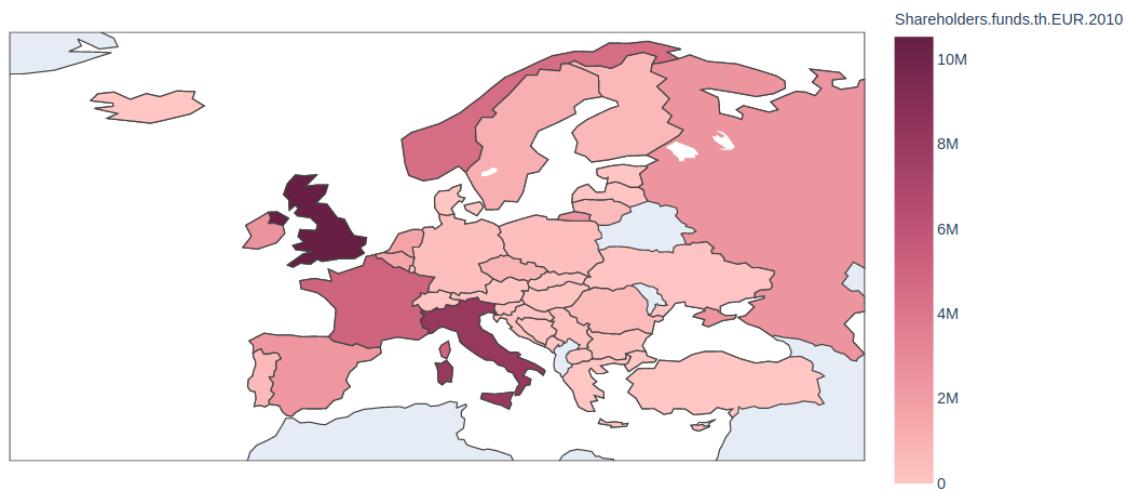
### Shareholders.funds.th.EUR.2010



Shareholders.funds.th.EUR.2010



Shareholders.funds.th.EUR.2010



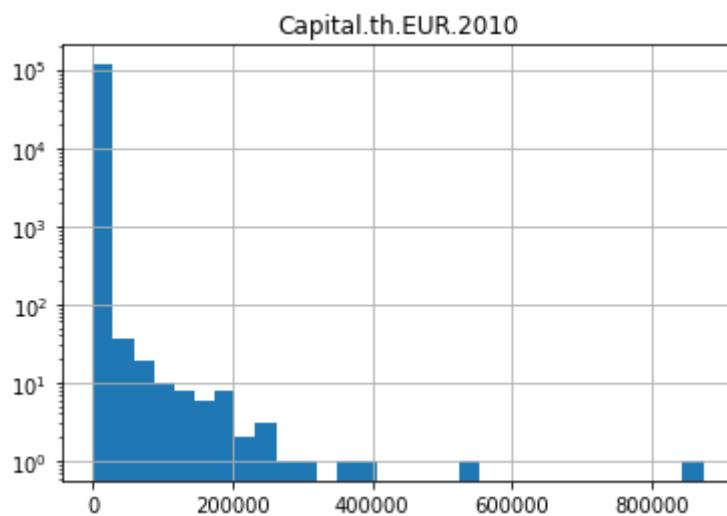
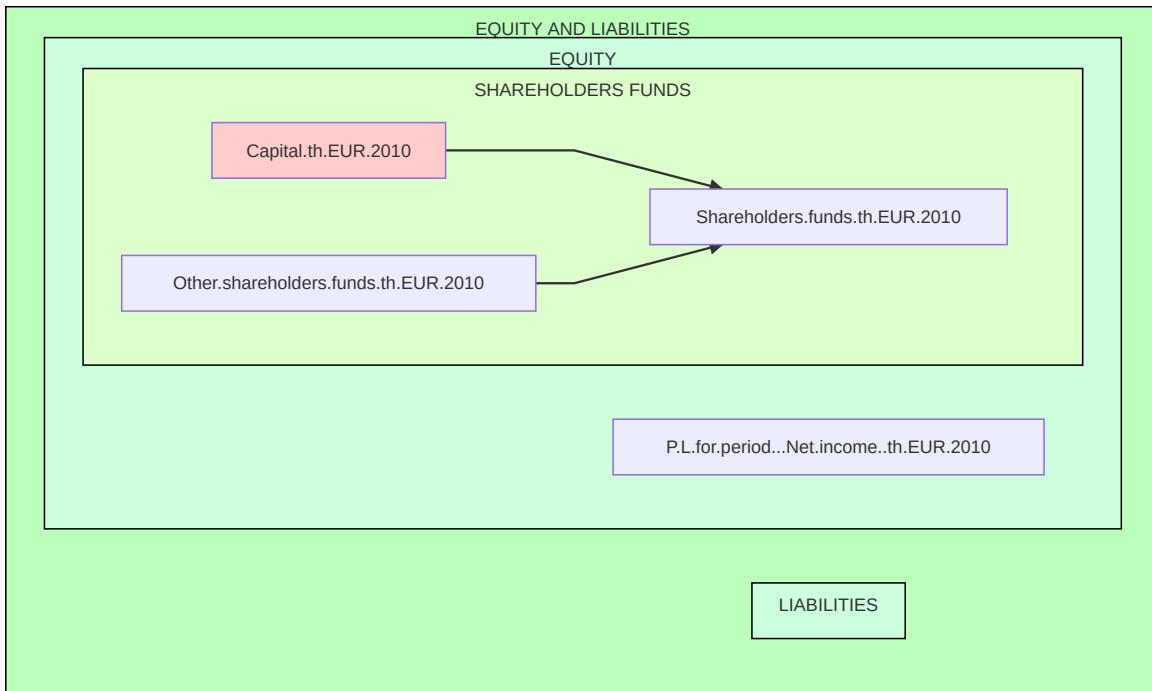
outlier:

BvD.ID.number

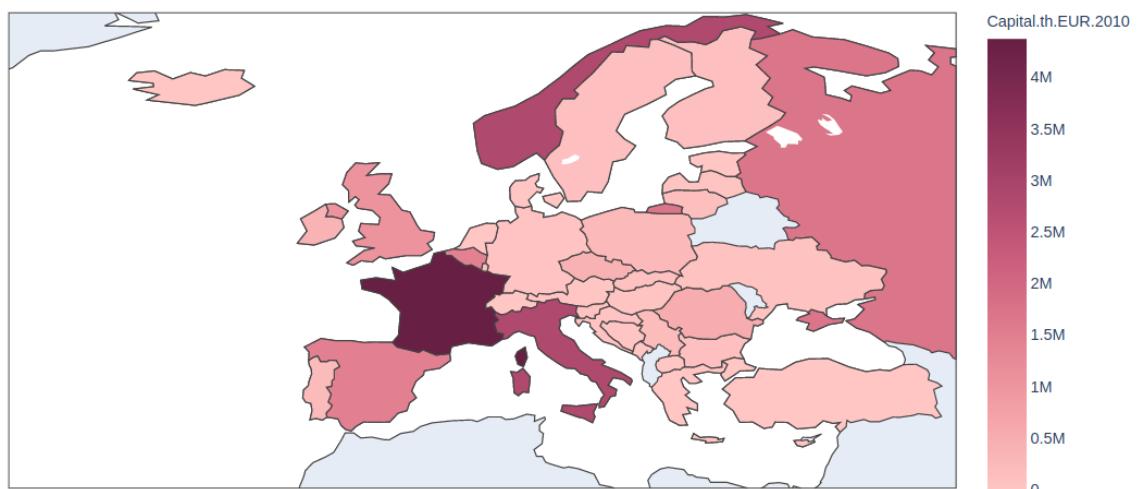
GB07450219 LONG ISLAND ASSETS LIMITED

Name: Company.name, dtype: string

## Capital.th.EUR.2010



Total capital per country



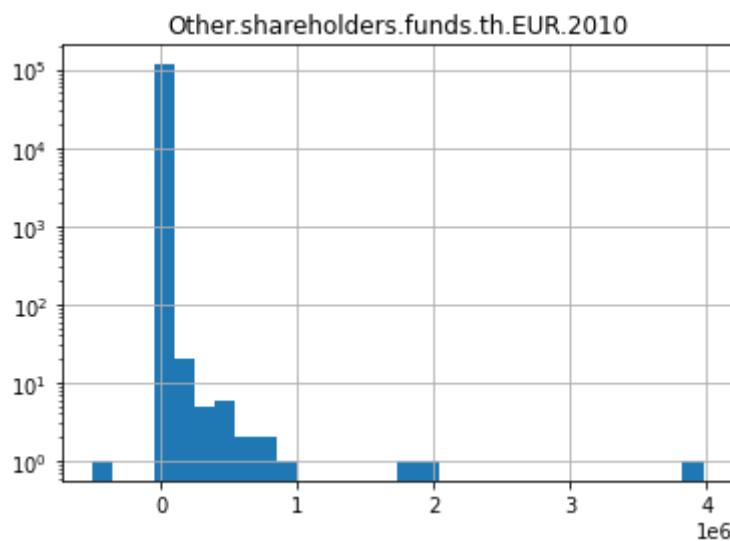
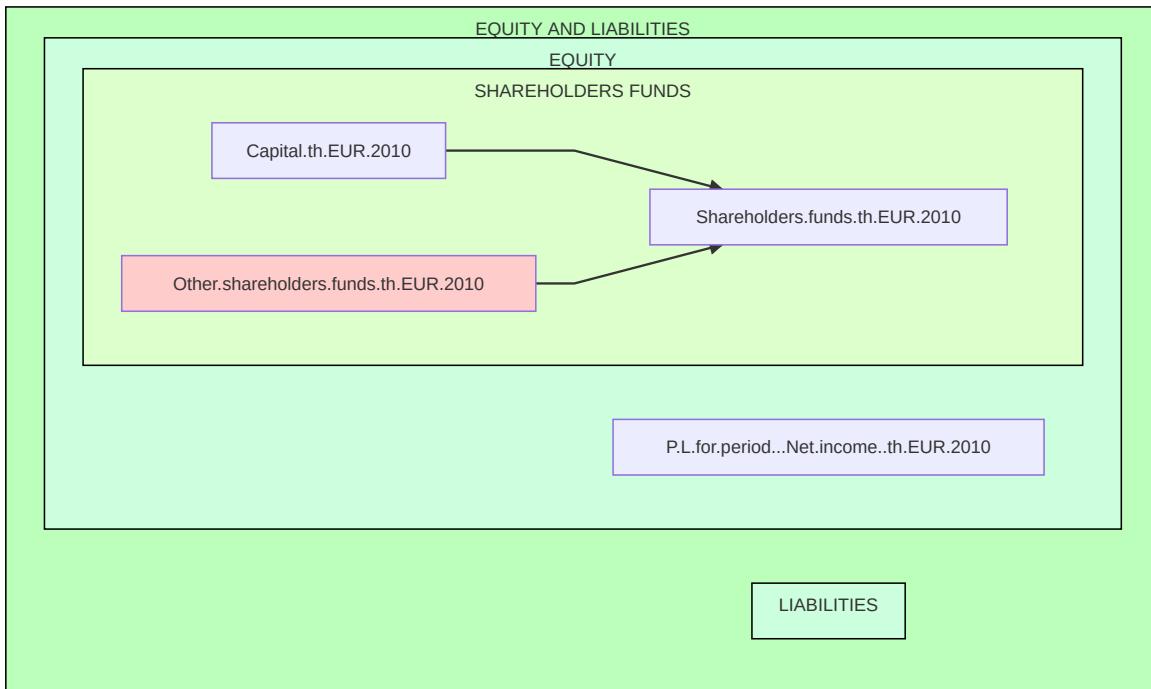
outlier:

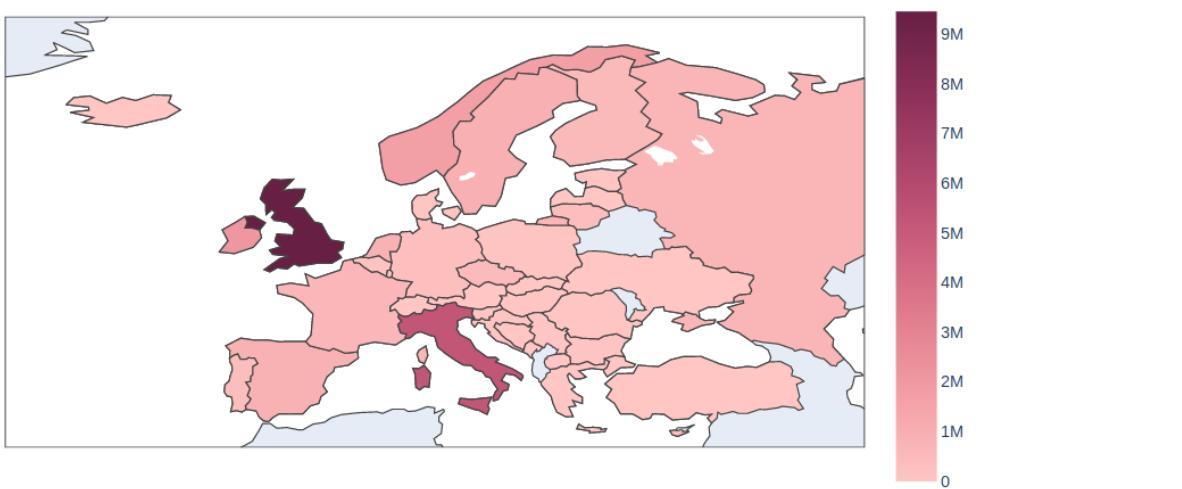
BvD.ID.number

FR519720643 IRIDIUM FRANCE

Name: Company.name, dtype: string

## Other.shareholders.funds.th.EUR.2010





42359 companies has negative values in the range (-)

outlier:

```
BvD.ID.number  
GB07450219 LONG ISLAND ASSETS LIMITED  
Name: Company.name, dtype: string
```

Distribution of negative values:



outlier:

```
vD.ID.number  
FR519720643 IRIDIUM FRANCE  
Name: Company.name, dtype: string BvD.ID.number  
FR519720643 -502140.0  
Name: Other.shareholders.funds.th.EUR.2010, dtype: float64
```

## Liabilities

### Non.current.liabilities.th.EUR.2010

## EQUITY AND LIABILITIES

EQUITY

### LIABILITIES

Long.term.debt.th.EUR.2010

Other.non.current.liabilities.th.EUR.2010

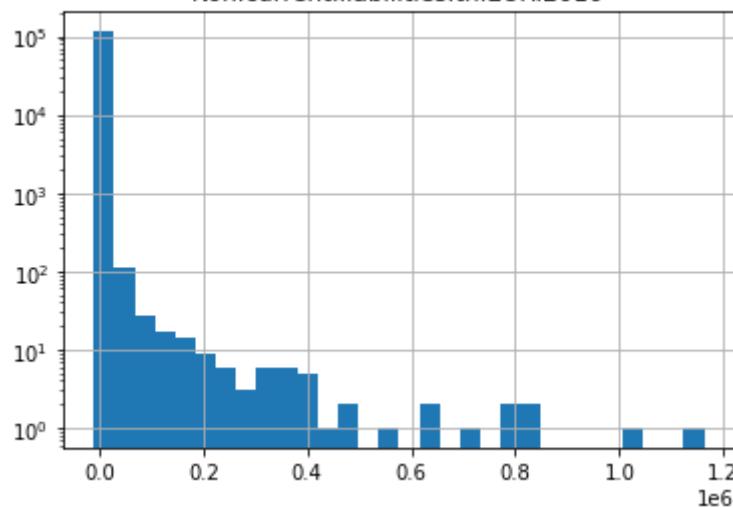
Creditors.th.EUR.2010

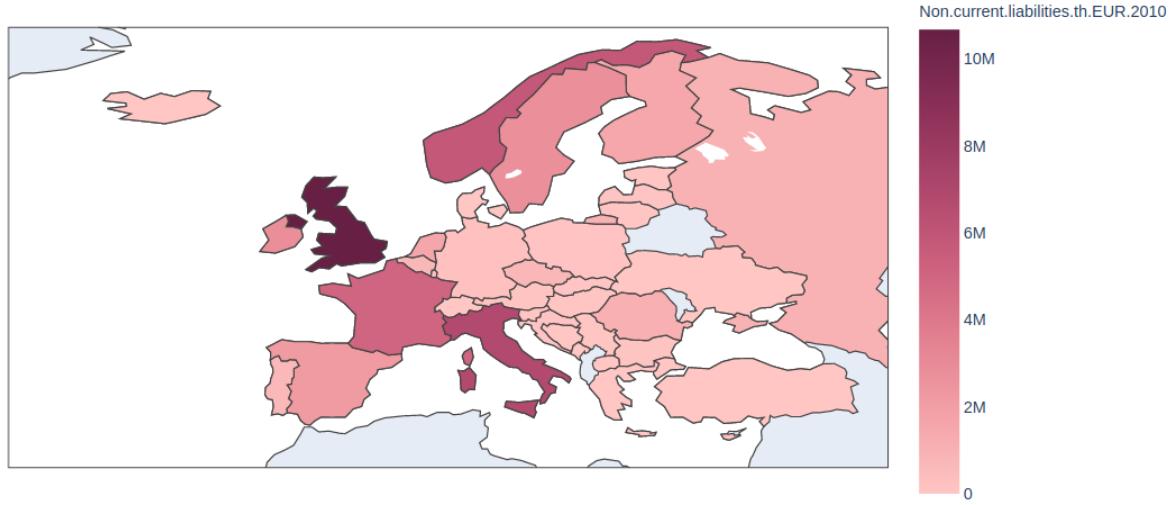
Other.current.liabilities.th.EUR.2010

Non.current.liabilities.th.EUR.2010

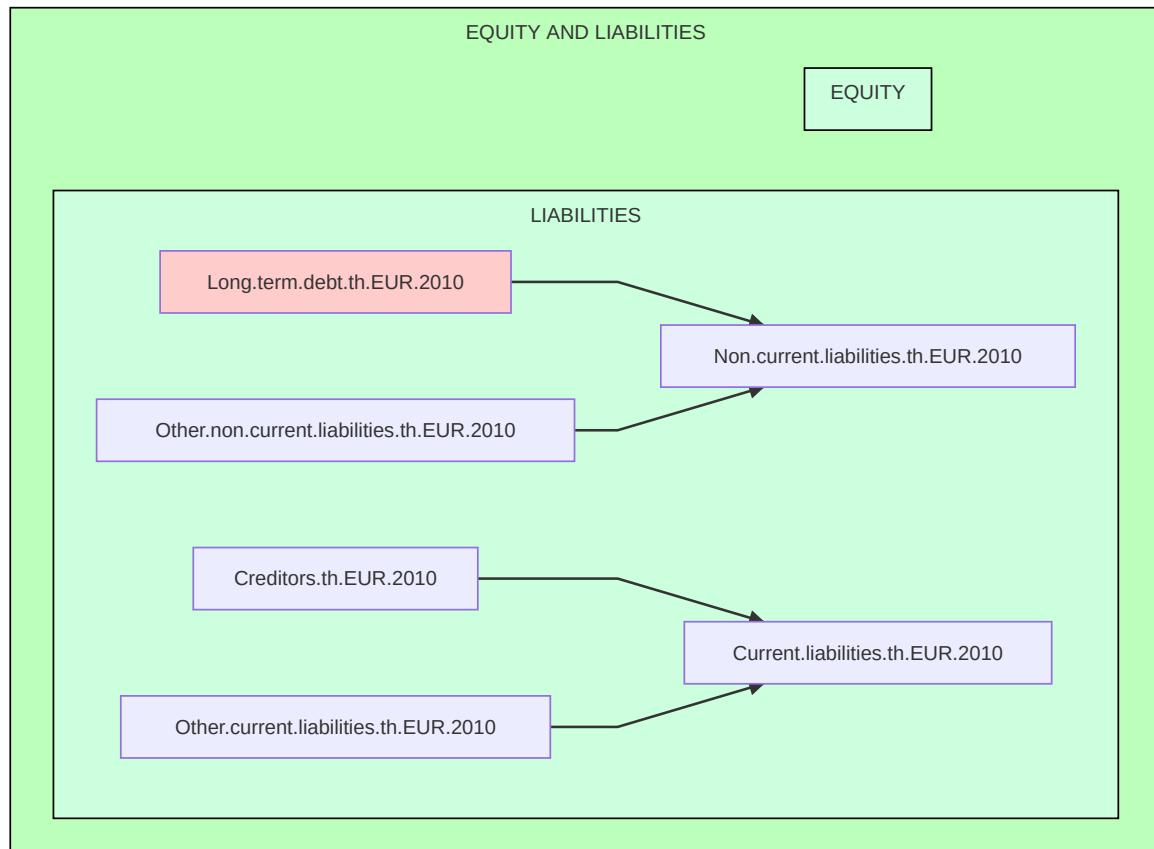
Current.liabilities.th.EUR.2010

Non.current.liabilities.th.EUR.2010

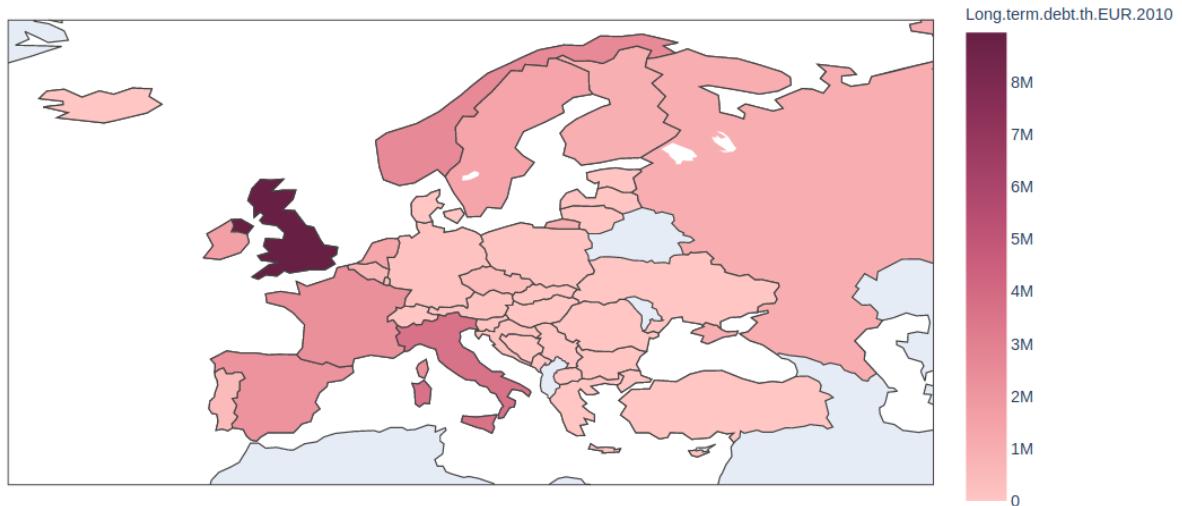
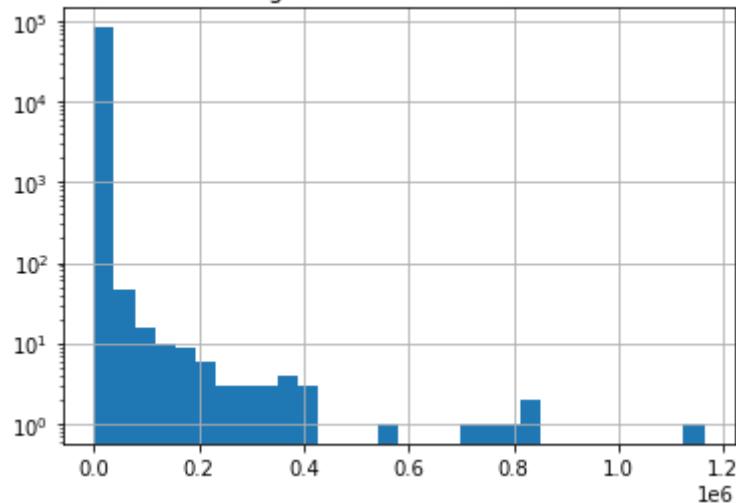




### Long.term.debt.th.EUR.2010



Long.term.debt.th.EUR.2010



outlier:

```
BvD.ID.number  
GB07251526      TESCO PROPERTY FINANCE 3 PLC  
Name: Company.name, dtype: string
```

### Other.non.current.liabilities.th.EUR.2010

## EQUITY AND LIABILITIES

EQUITY

### LIABILITIES

Long.term.debt.th.EUR.2010

Non.current.liabilities.th.EUR.2010

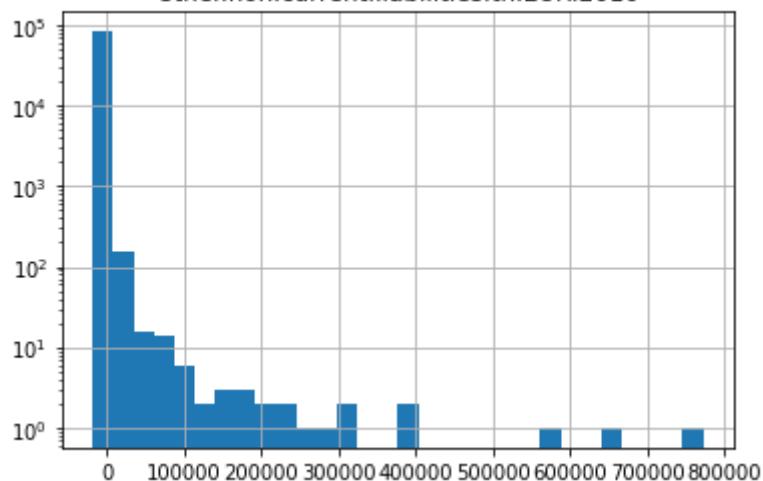
Other.non.current.liabilities.th.EUR.2010

Creditors.th.EUR.2010

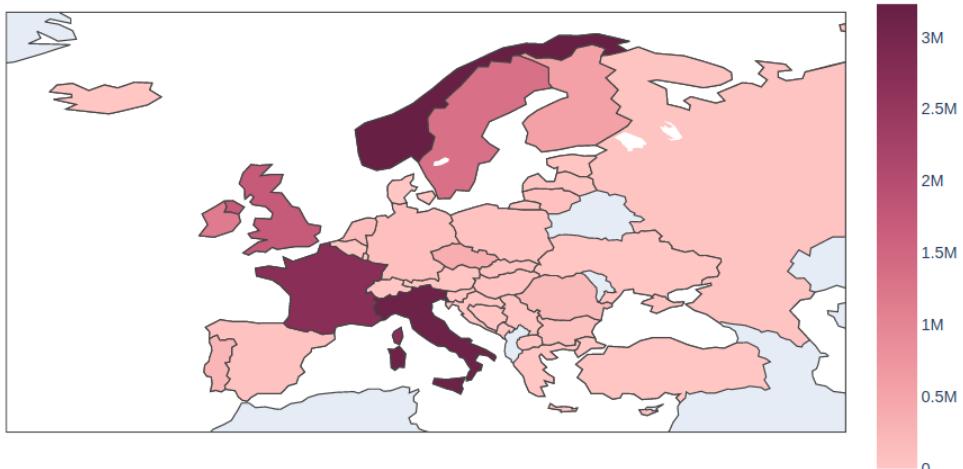
Current.liabilities.th.EUR.2010

Other.current.liabilities.th.EUR.2010

Other.non.current.liabilities.th.EUR.2010



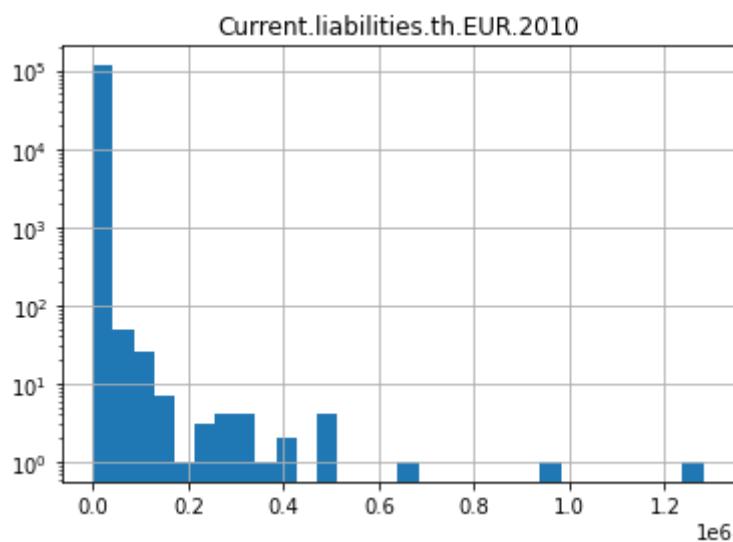
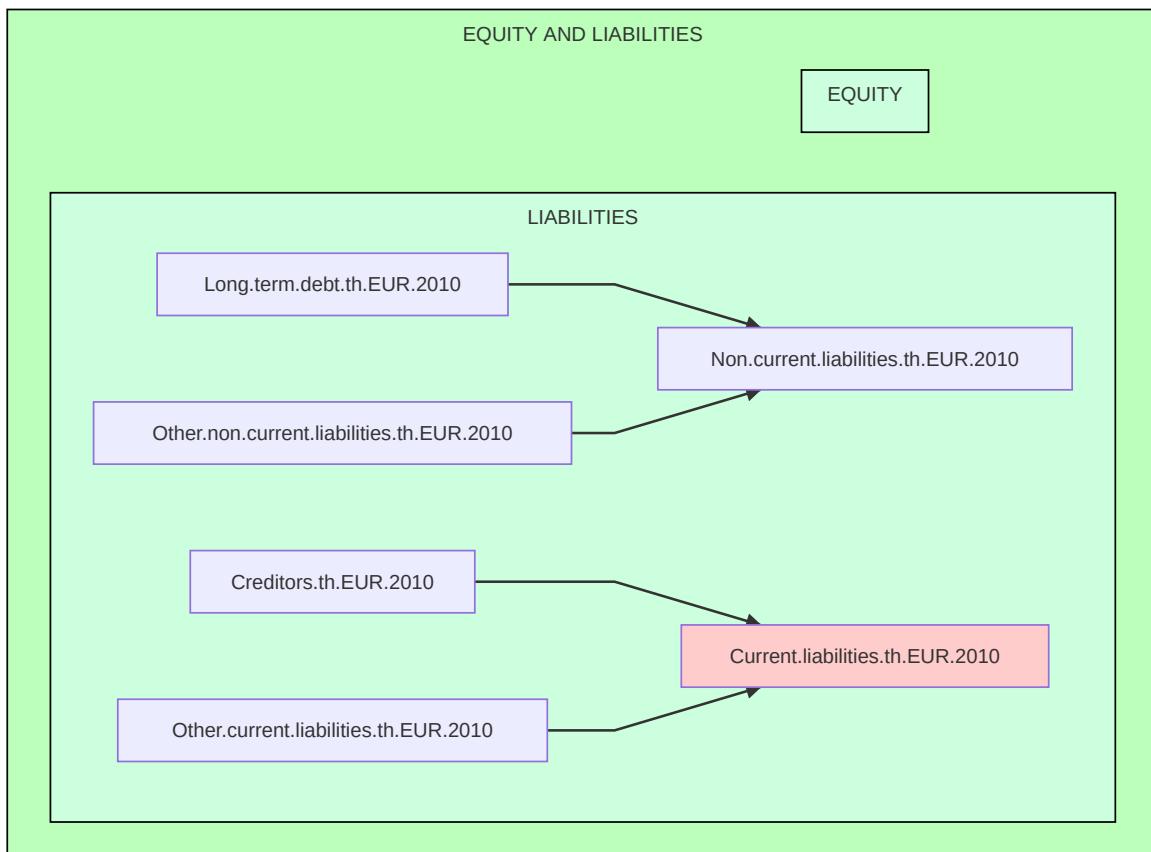
Other.non.current.liabilities.th.EUR.2010

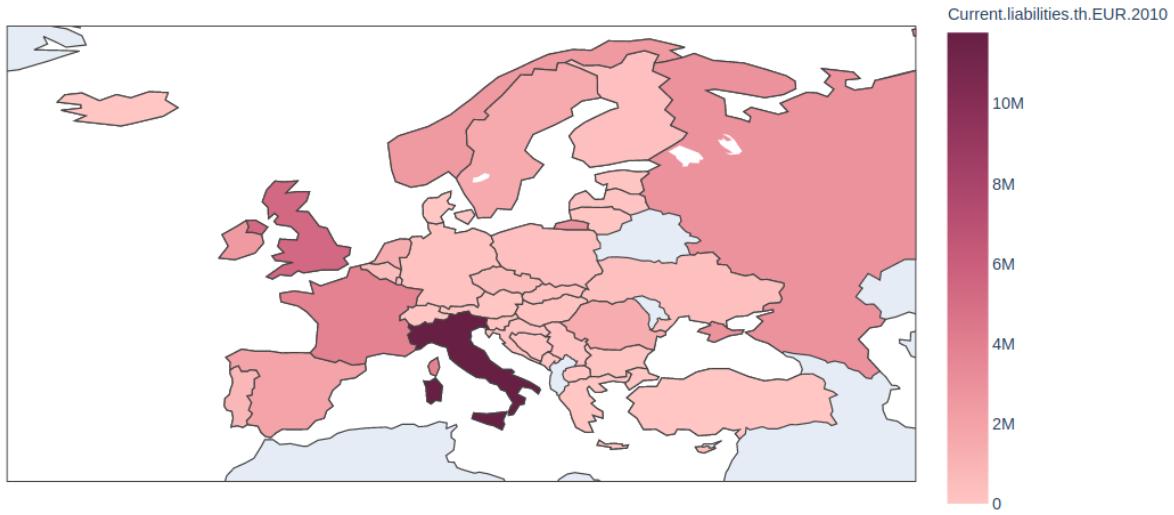


outliers:

vd.ID.number	SCF CAPITAL DESIGNATED ACTIVITY COMPANY
IE487769	DIACINE FRANCE
FR521029926	INDUSTRIINVESTERINGER AS
N0995633604	

## Current.liabilities.th.EUR.2010

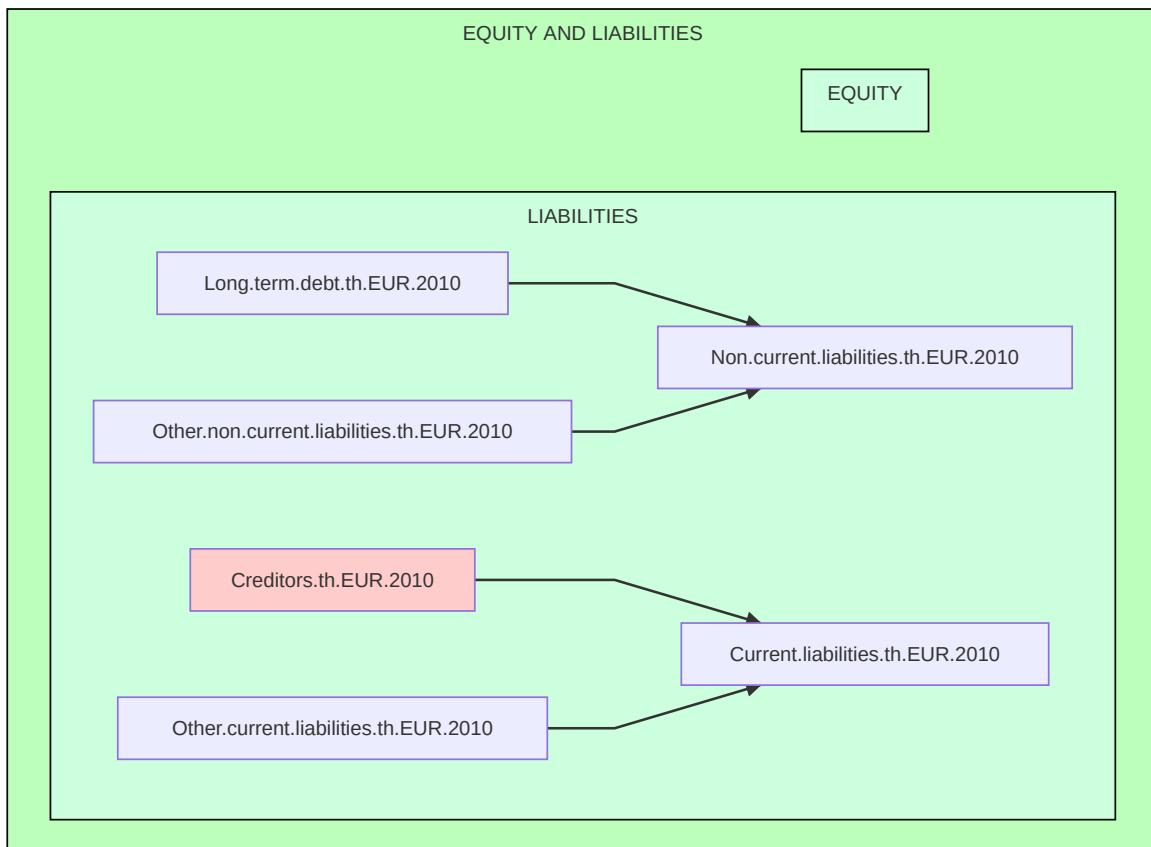




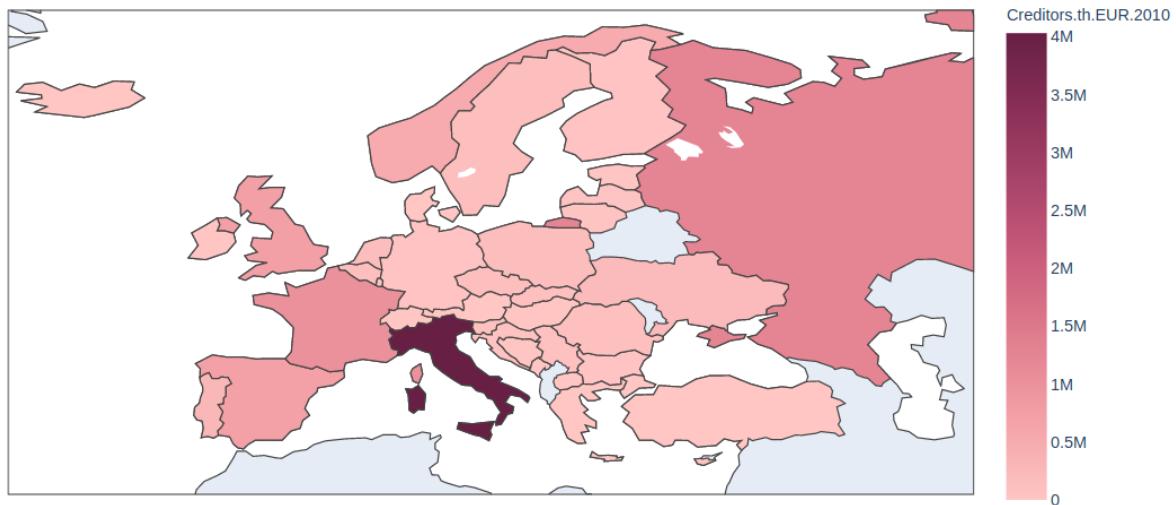
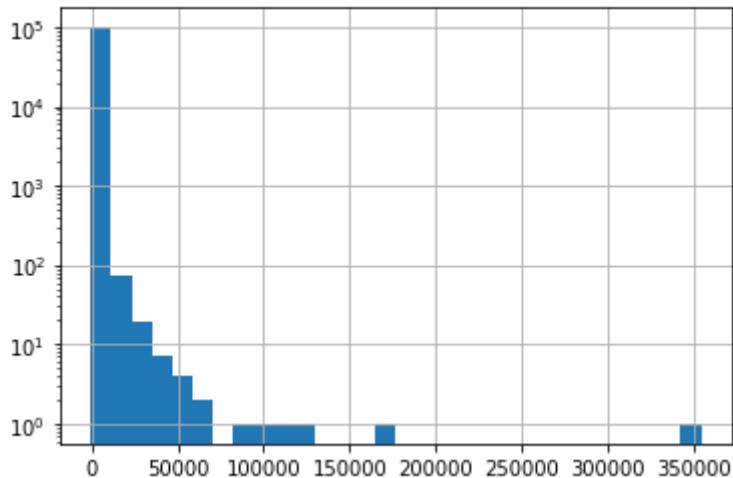
outliers:

BvD.ID.number	
IE480184	ESB FINANCE DESIGNATED ACTIVITY COMPANY
IE486122	ICG EOS LOAN FUND I LIMITED
GB07193500	PREMIER LOTTERIES CAPITAL UK LIMITED
GB07202475	PREMIER LOTTERIES INVESTMENTS UK LIMITED
IT10319310016	INFRATRASPORTI.TO S.R.L.
Name: Company.name, dtype: string	

## Creditors.th.EUR.2010



Creditors.th.EUR.2010



outliers:

```
BvD.ID.number
GB07254605      ED BROKING GROUP LIMITED
IT10969001006    LOTTERIE NAZIONALI S.R.L.
Name: Company.name, dtype: string
```

### Other.current.liabilities.th.EUR.2010

## EQUITY AND LIABILITIES

EQUITY

### LIABILITIES

Long.term.debt.th.EUR.2010

Non.current.liabilities.th.EUR.2010

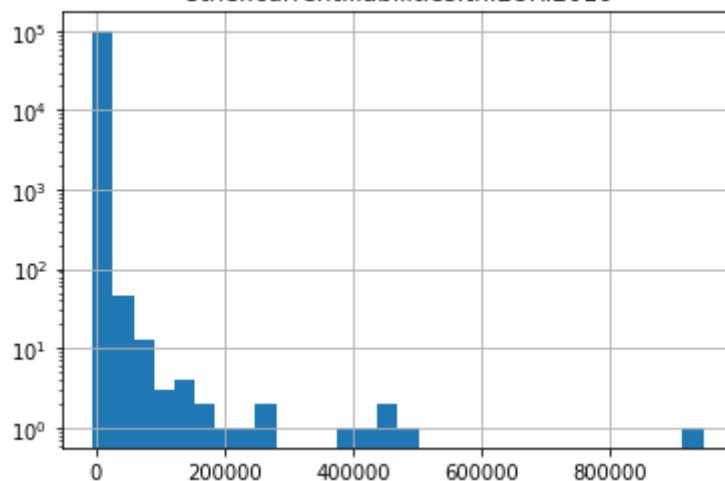
Other.non.current.liabilities.th.EUR.2010

Creditors.th.EUR.2010

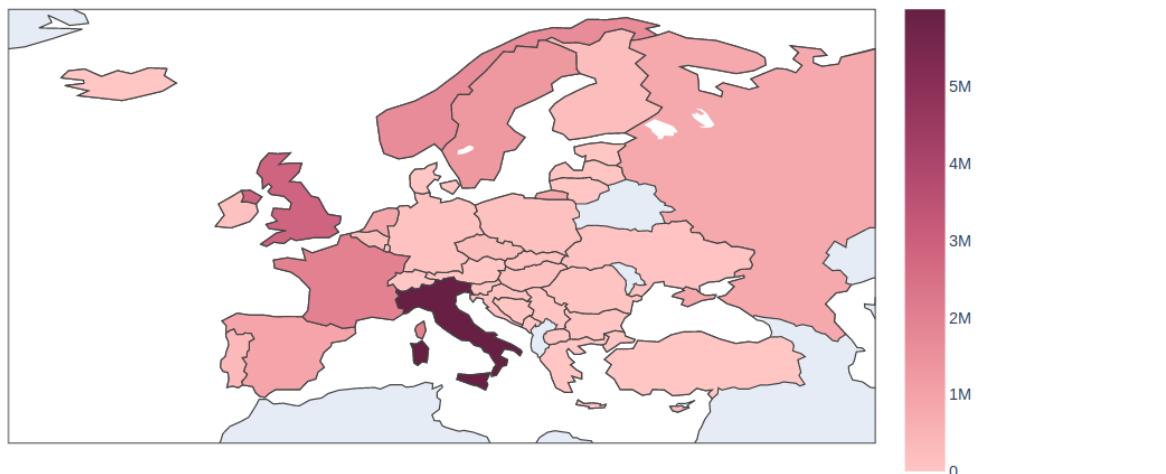
Current.liabilities.th.EUR.2010

Other.current.liabilities.th.EUR.2010

Other.current.liabilities.th.EUR.2010



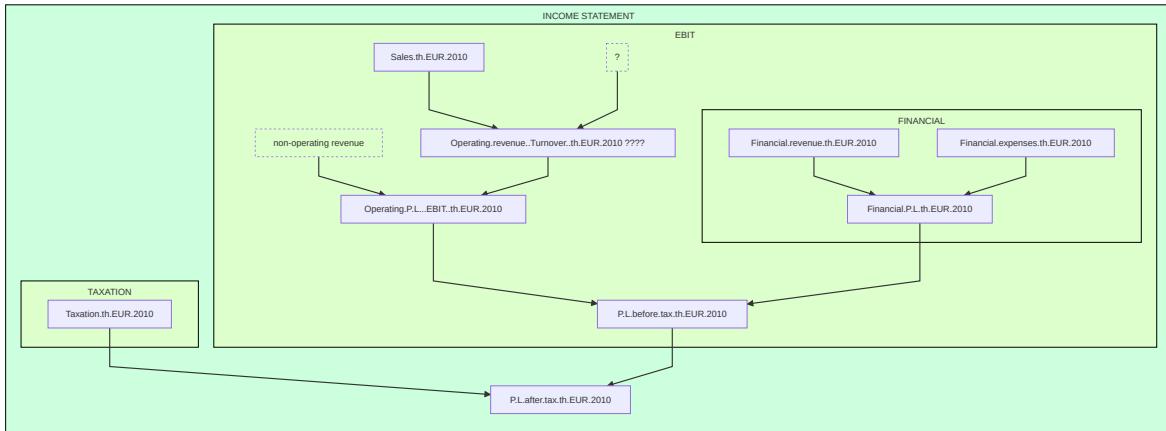
Other.current.liabilities.th.EUR.2010



outlier:

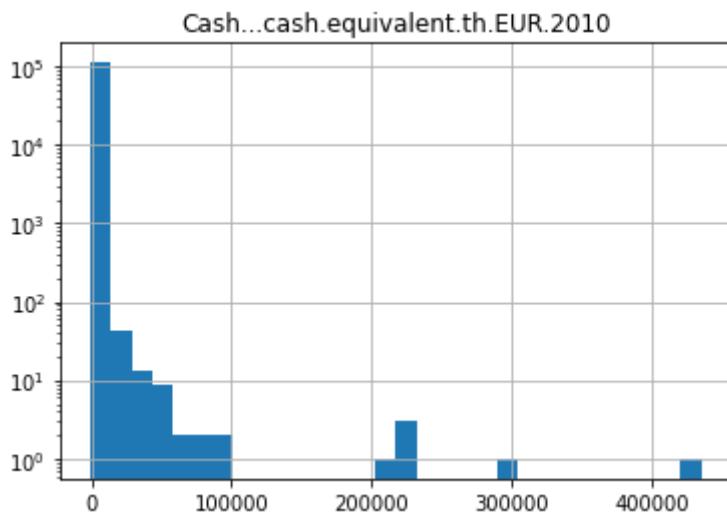
```
BvD.ID.number  
IT10319310016 INFRATRASPORTI.TO S.R.L.  
Name: Company.name, dtype: string
```

## Income statement



## Cash...cash.equivalent.th.EUR.2010

< to be placed in the dependancy graph >



outliers:

400+K bin:

```
BvD.ID.number  
N0995216604 WALLENIUS WILHELMSEN ASA  
Name: Company.name, dtype: string
```

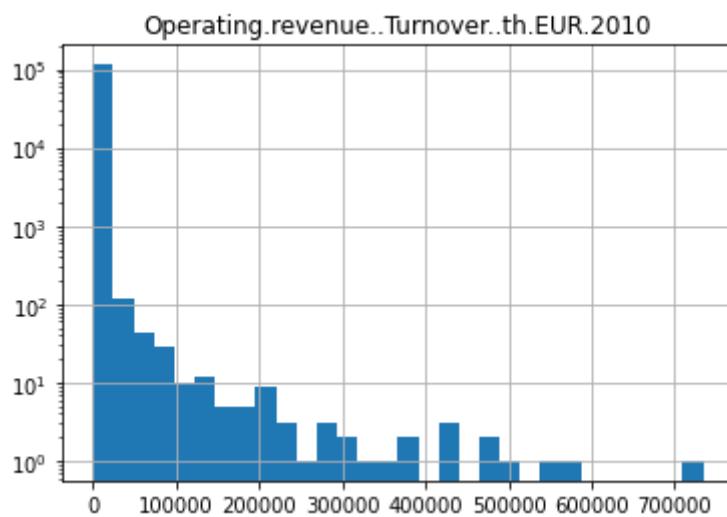
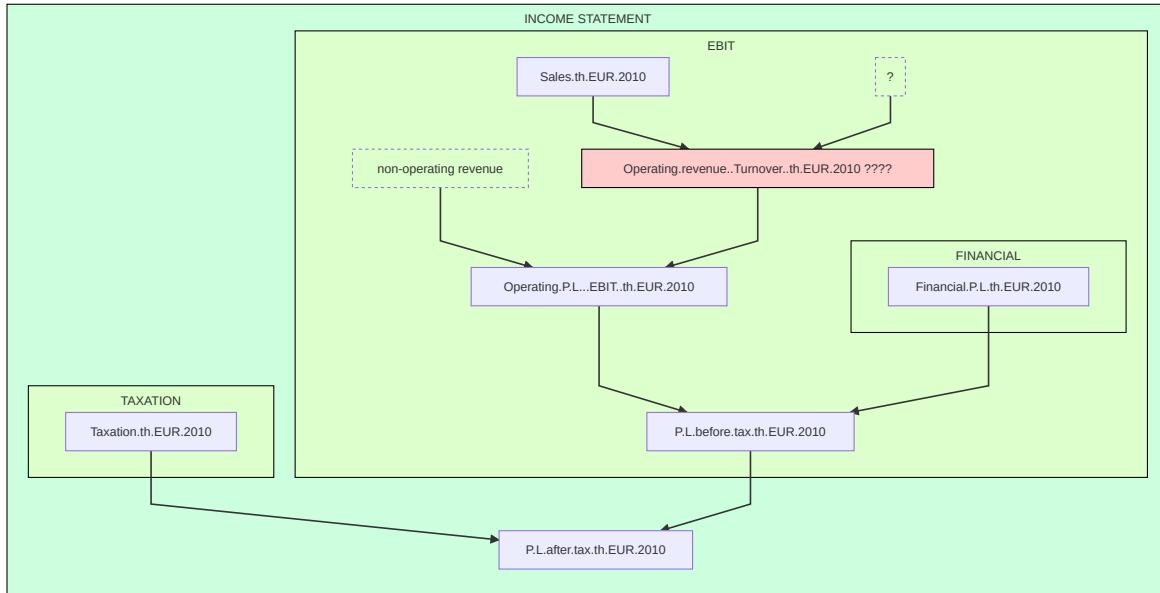
300K bin:

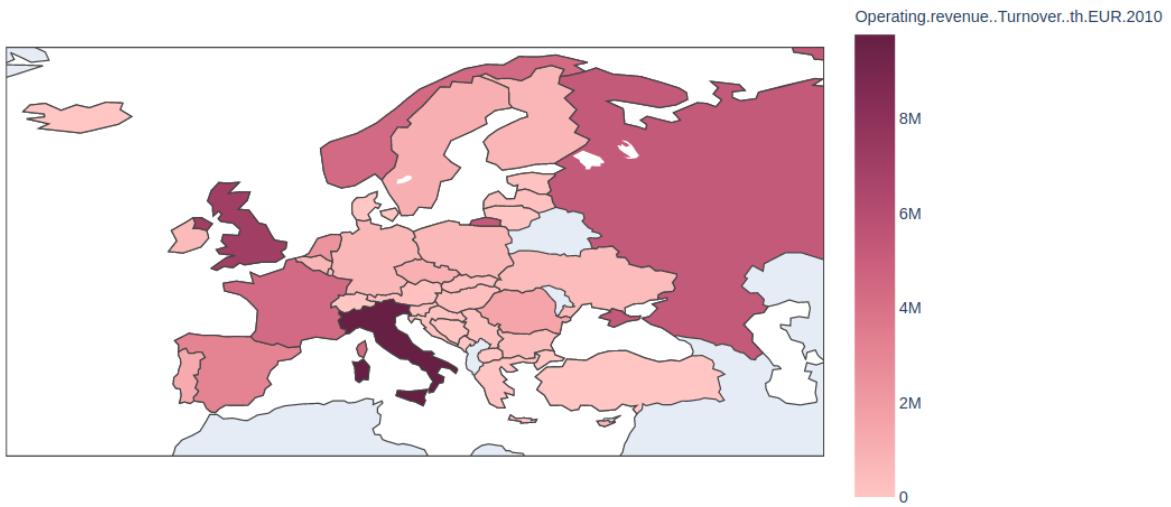
```
BvD.ID.number  
GB07123187 ACACIA MINING PLC  
Name: Company.name, dtype: string
```

200K+ bin:

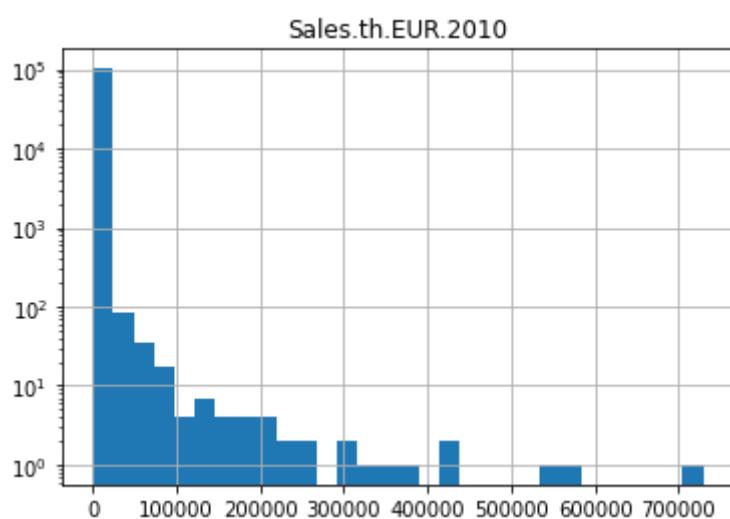
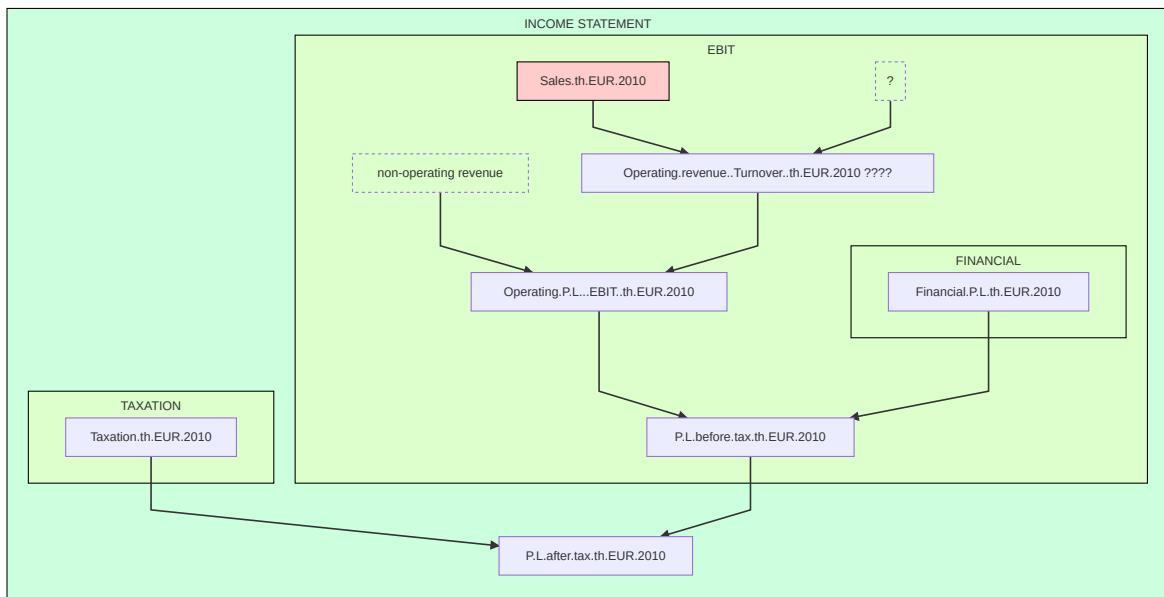
BvD.ID.number  
BE0831465984 XIX-INVEST  
GB07145051 CAPITAL & COUNTIES PROPERTIES PLC  
GB07254605 ED BROKING GROUP LIMITED  
GB07283266 HIGHBRIDGE COBALT LIMITED  
Name: Company.name, dtype: string

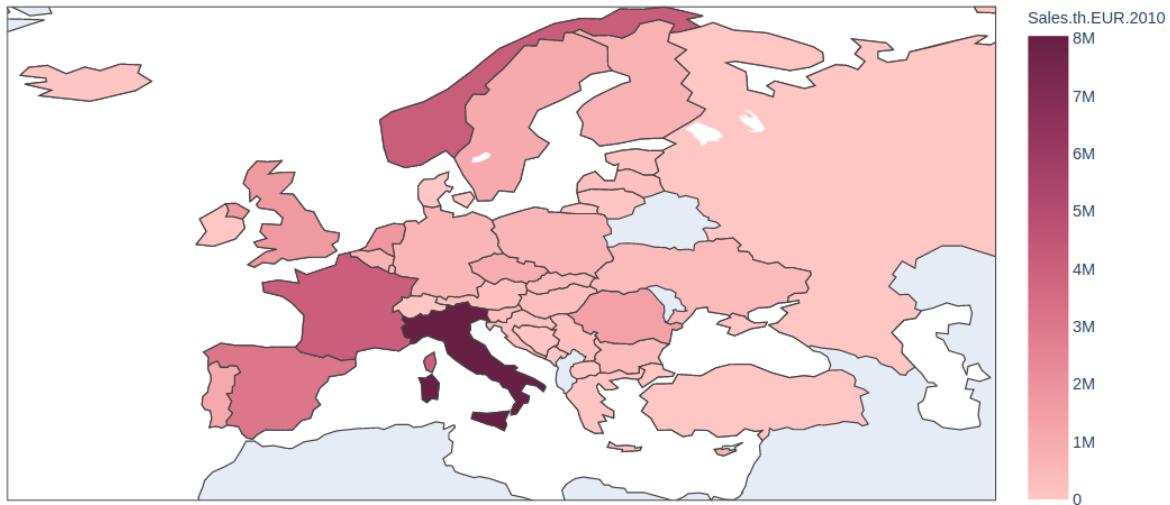
## Operating.revenue..Turnover..th.EUR.2010





## Sales.th.EUR.2010

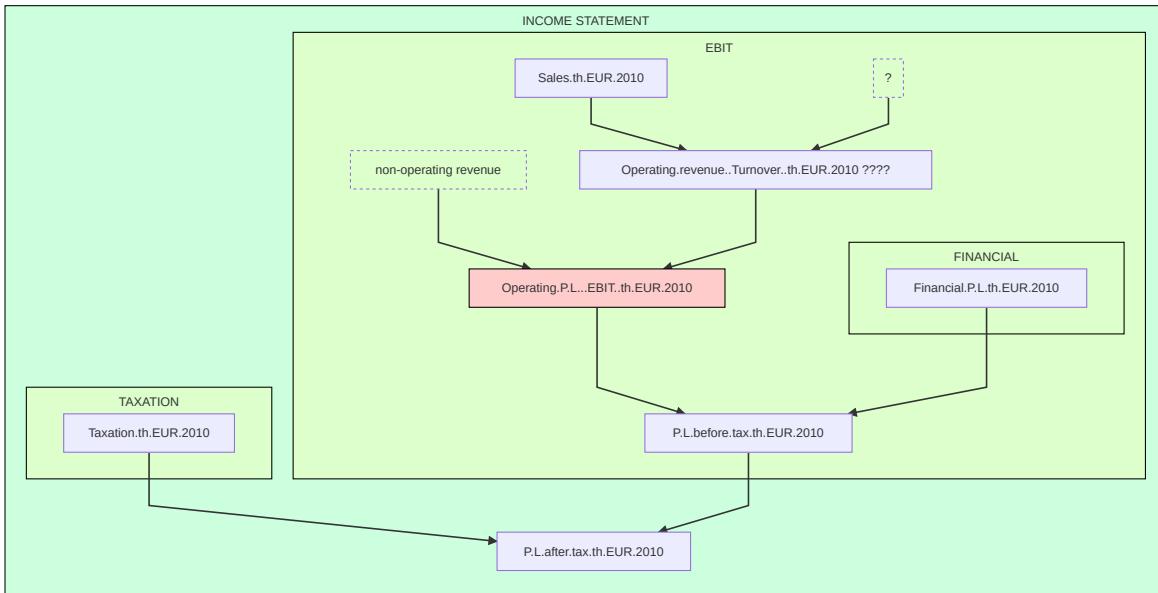




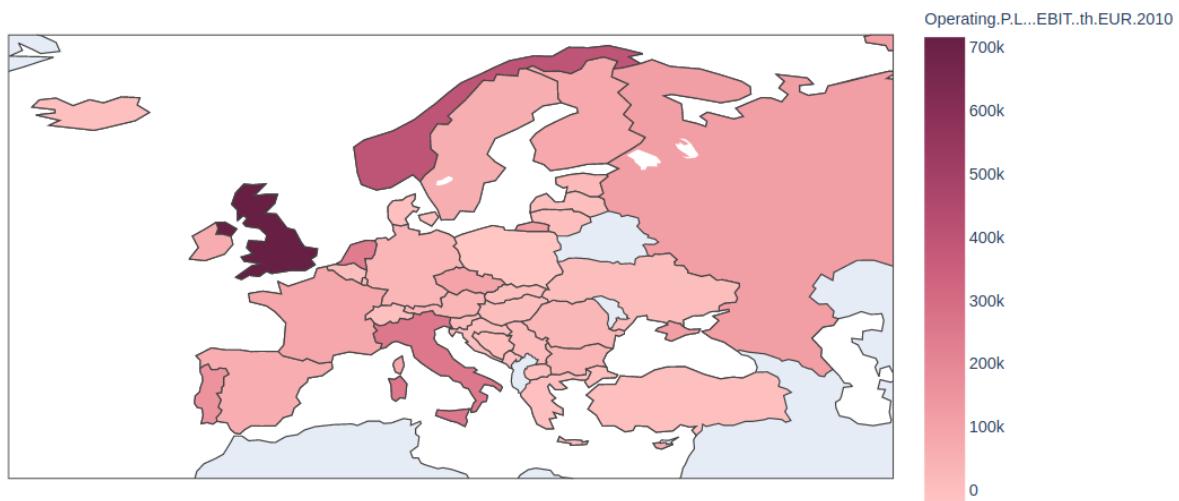
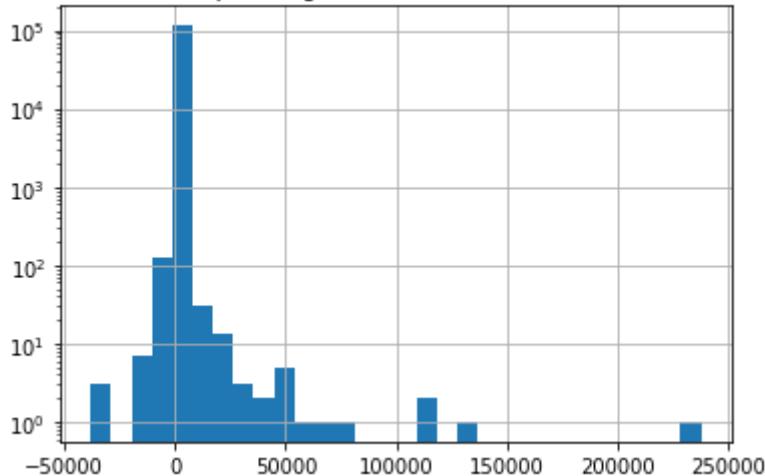
outliers:

```
BvD.ID.number
CYC266578      HMS HYDRAULIC MACHINES & SYSTEMS GROUP PLC
GB07123187      ACACIA MINING PLC
IT10813301008    EOS S.R.L.
Name: Company.name, dtype: string
```

## Operating.P.L...EBIT..th.EUR.2010



Operating.P.L...EBIT..th.EUR.2010



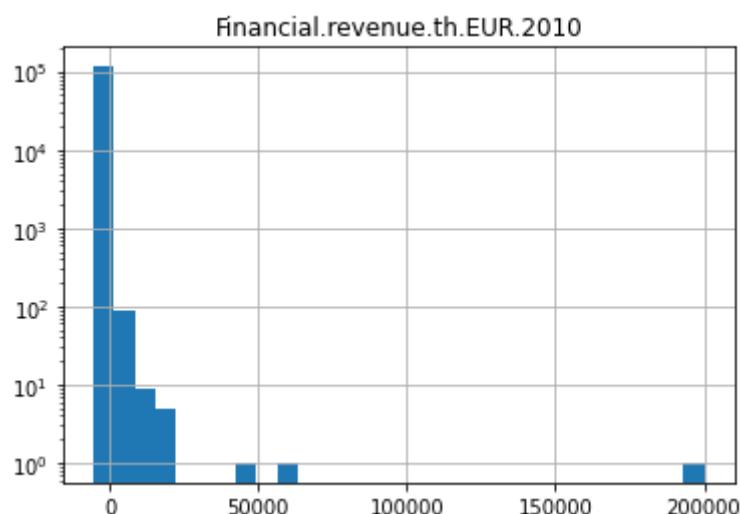
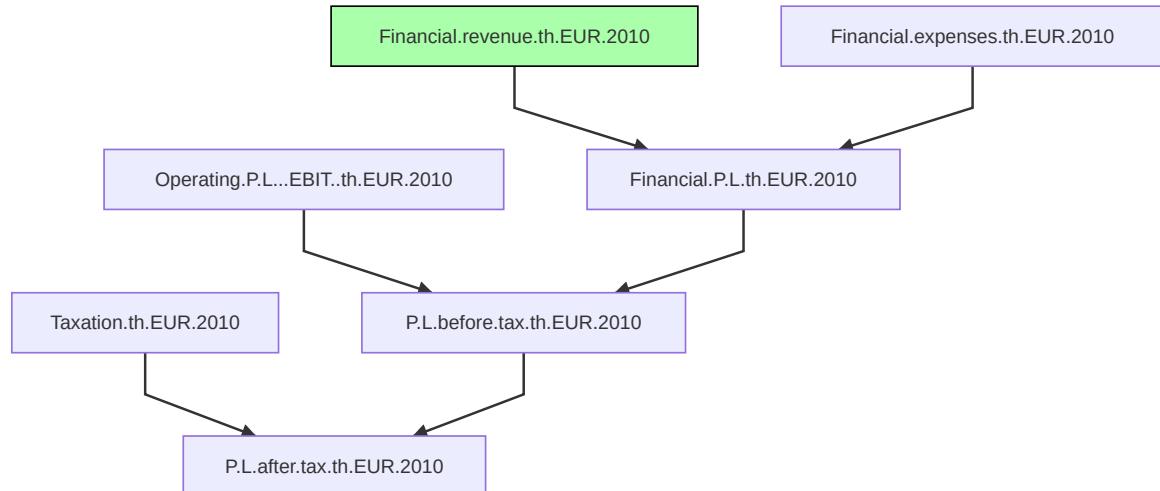
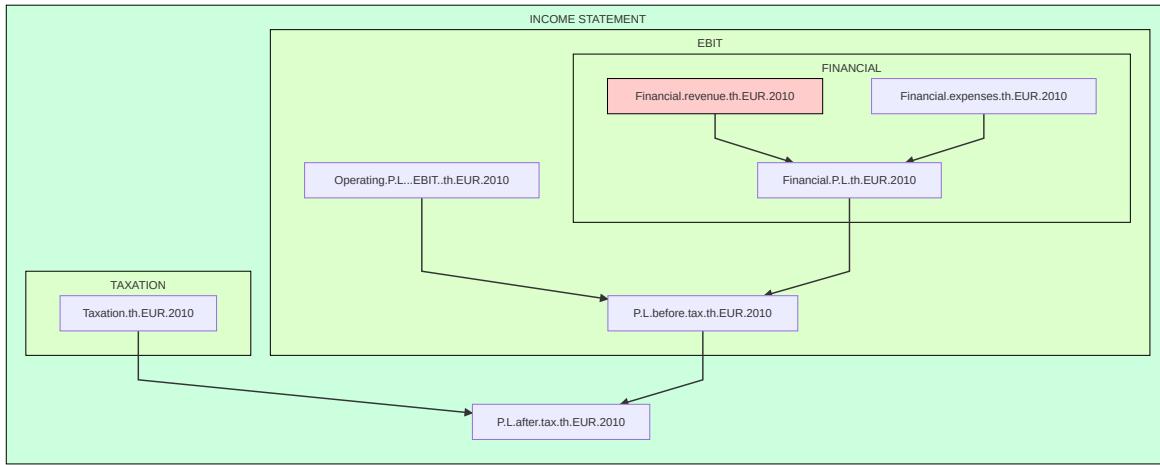
Positive outliers:

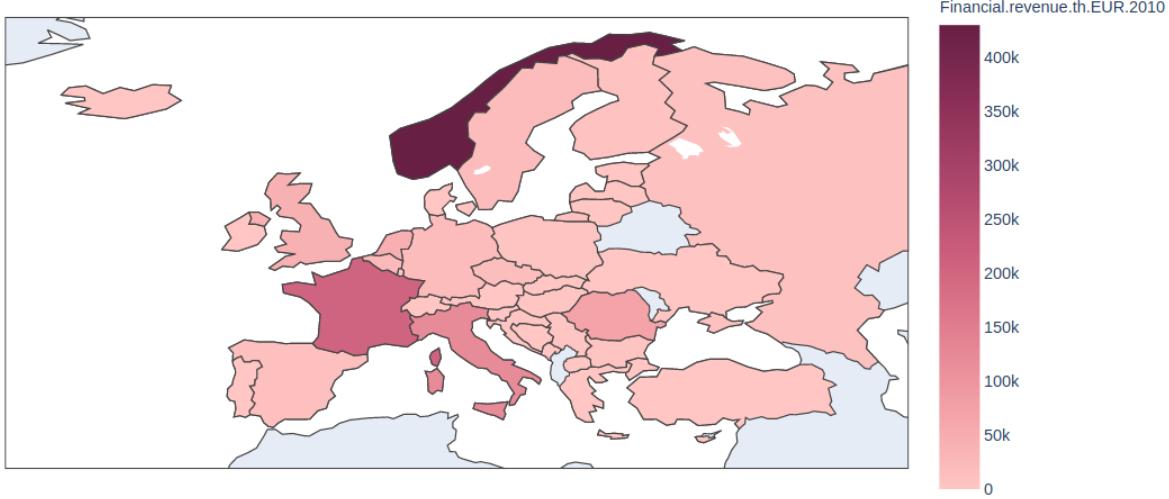
```
BvD.ID.number
GB07123187          ACACIA MINING PLC
NL50397931          ATLANTIC AURUM INVESTMENTS B.V.
N0995216604          WALLENIUS WILHELMSEN ASA
PT509444229         MOTA-ENGIL AFRICA - SGPS, S.A.
Name: Company.name, dtype: string
```

Negative outliers:

```
BvD.ID.number
IE507678          HORIZON THERAPEUTICS PUBLIC LIMITED COMPANY
PL301339040        SAMSUNG ELECTRONICS POLAND MANUFACTURING SP. Z...
RU65519055         AKTSIONERNOE OBSHCHESTVO TATTEPLOSBYT
Name: Company.name, dtype: string
```

## Financial.revenue.th.EUR.2010





outliers:

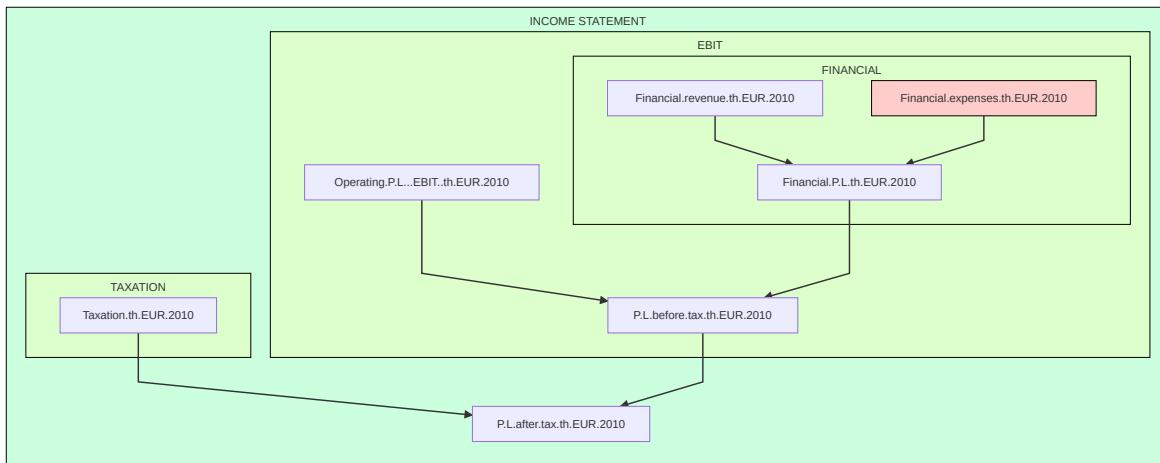
50K cluster:

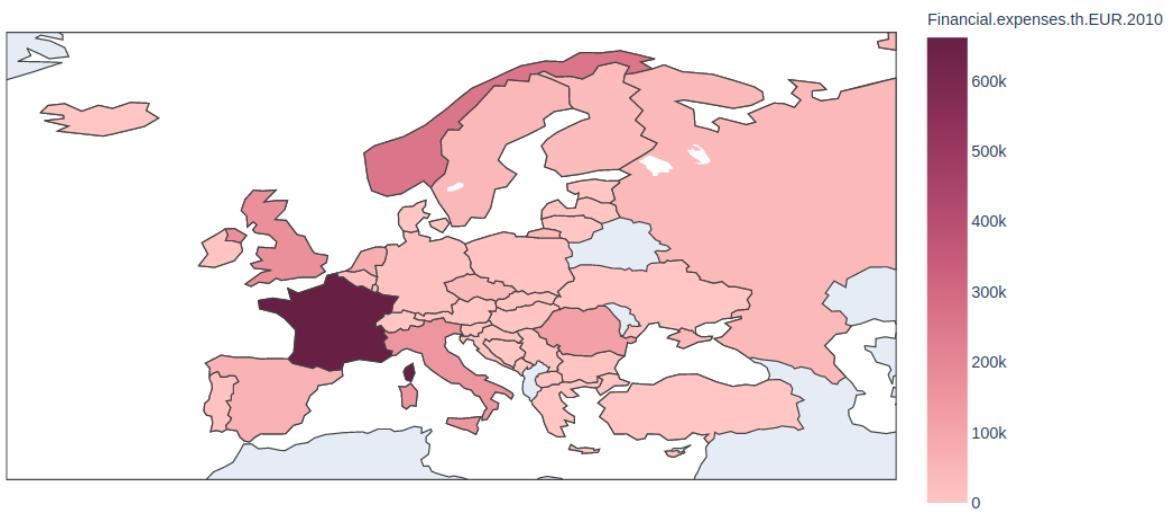
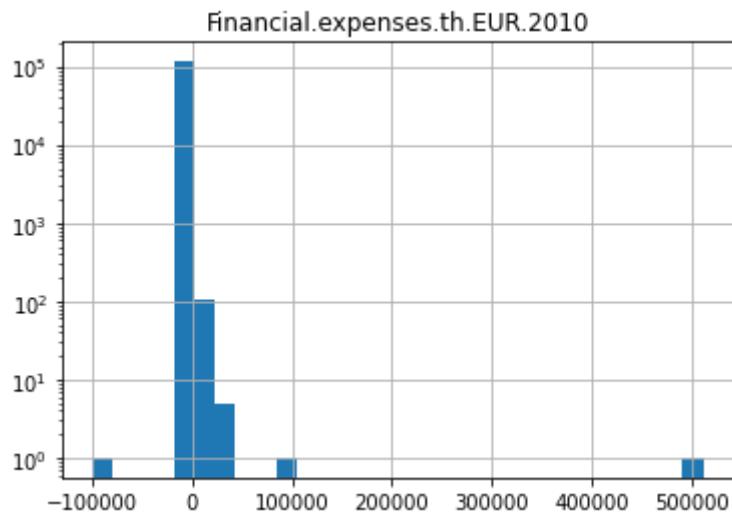
```
BvD.ID.number
FR527925143           SOFAQUE
N0996031454          DOLPHIN INVEST AS
Name: Company.name, dtype: string
```

200K cluster:

```
BvD.ID.number
N0995633604      INDUSTRIINVESTERINGER AS
Name: Company.name, dtype: string
```

## Financial.expenses.th.EUR.2010





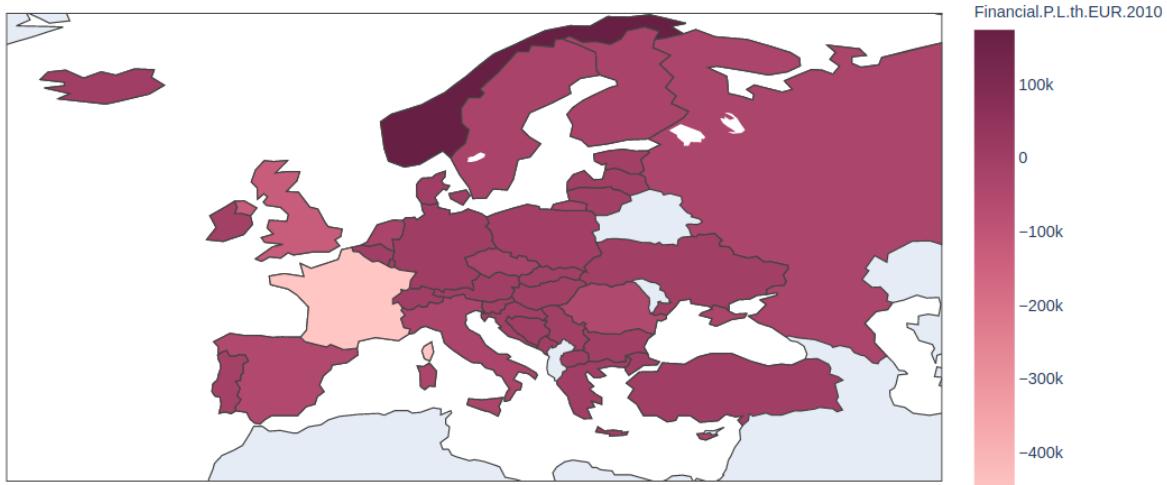
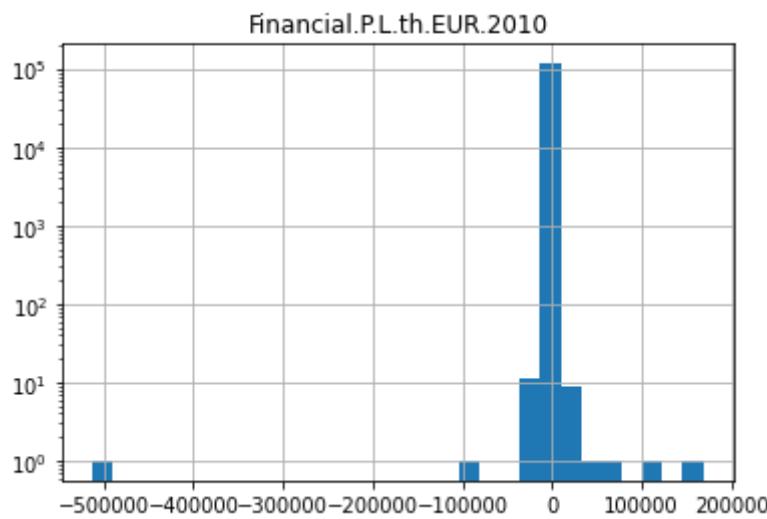
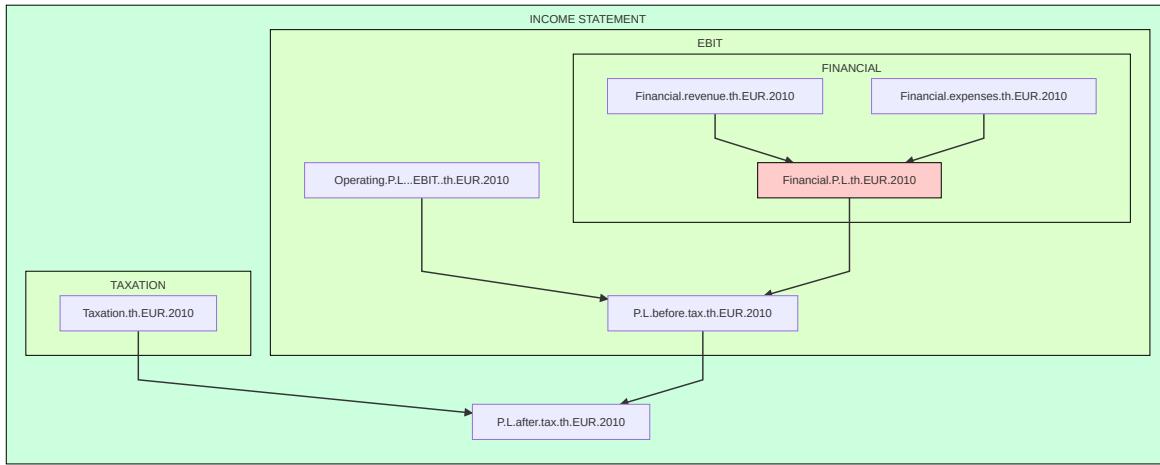
positive outlier:

```
BvD.ID.number
FR519720643    IRIDIUM FRANCE
Name: Company.name, dtype: string
```

negative outlier:

```
BvD.ID.number
GB07145051    CAPITAL & COUNTIES PROPERTIES PLC
Name: Company.name, dtype: string
```

## Financial.P.L.th.EUR.2010

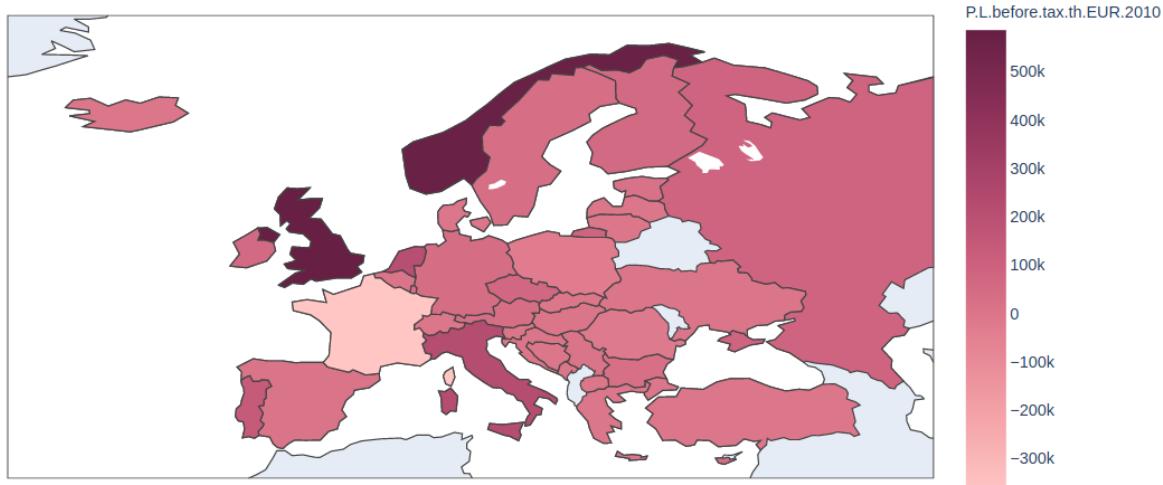
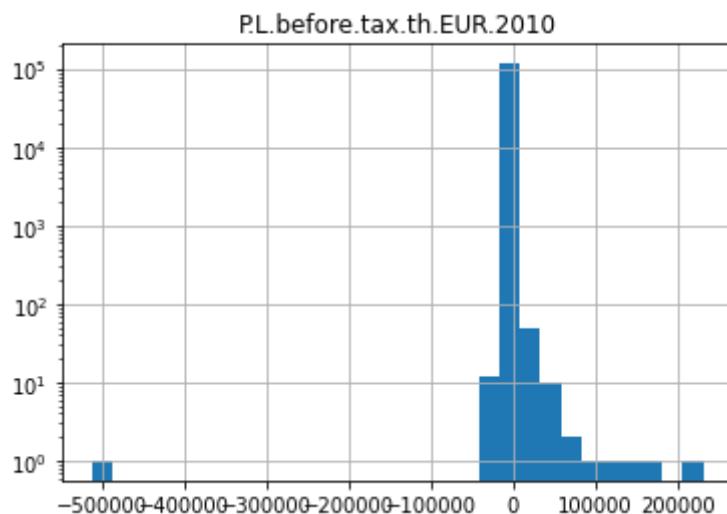
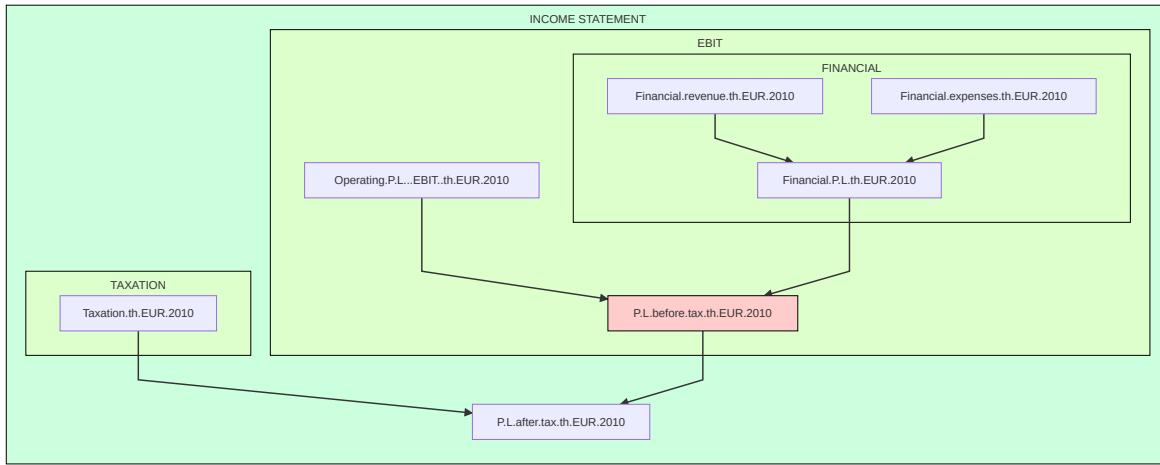


Outlier:

```

BvD.ID.number
FR519720643      IRIDIUM FRANCE
Name: Company.name, dtype: string
    
```

**P.L.before.tax.th.EUR.2010**

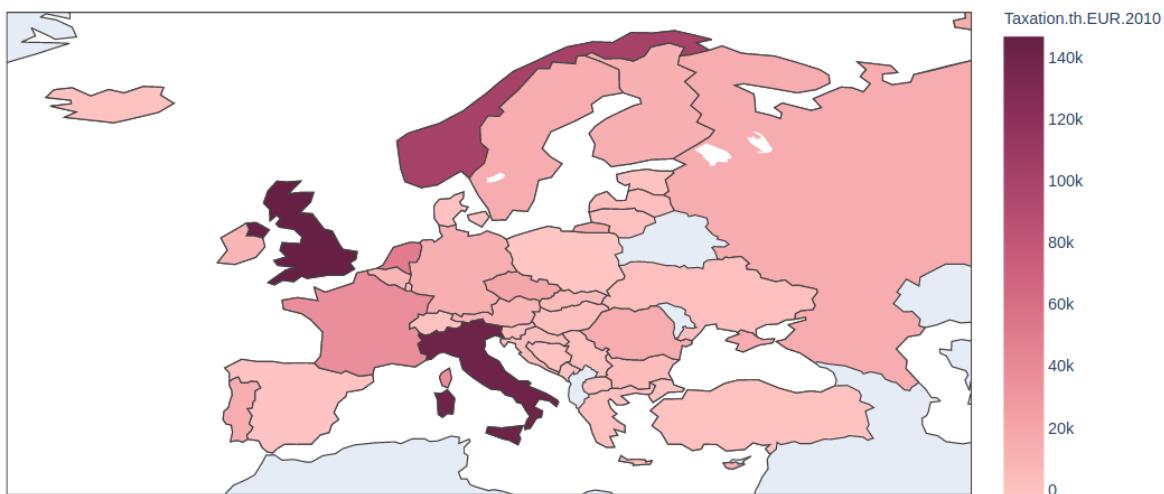
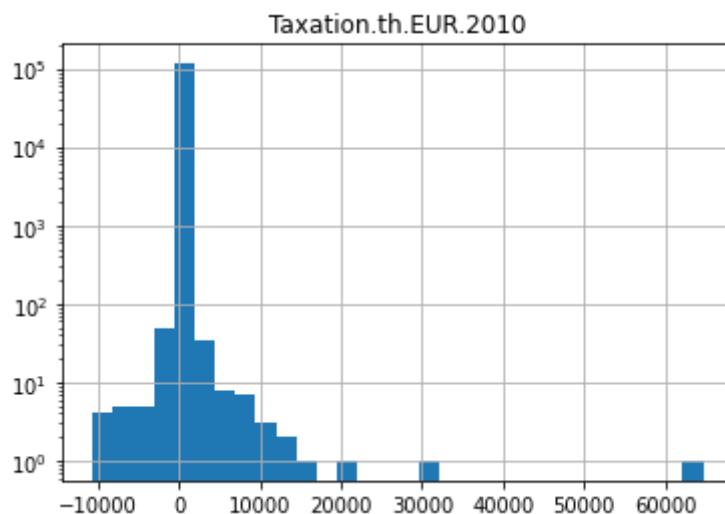
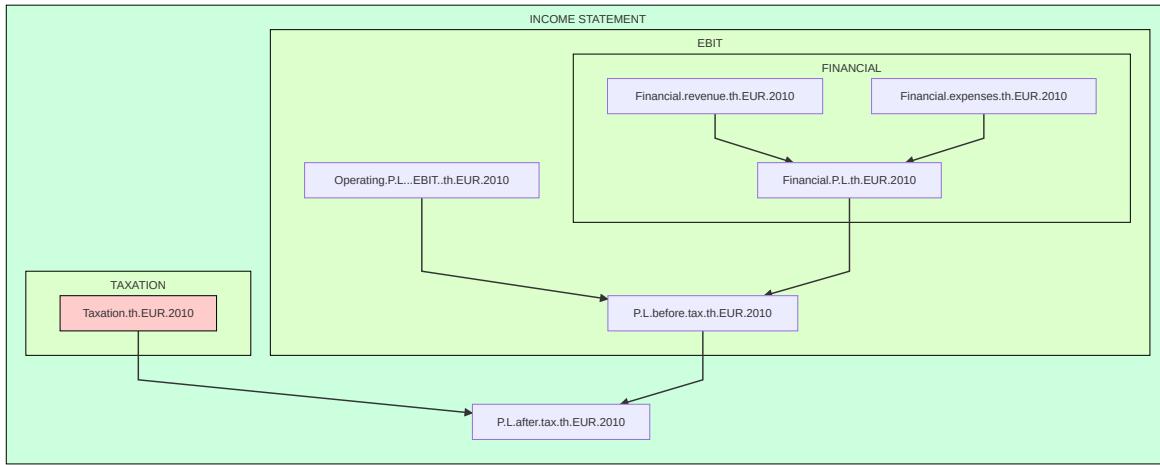


outlier:

```

BvD.ID.number
FR519720643      IRIDIUM FRANCE
Name: Company.name, dtype: string
  
```

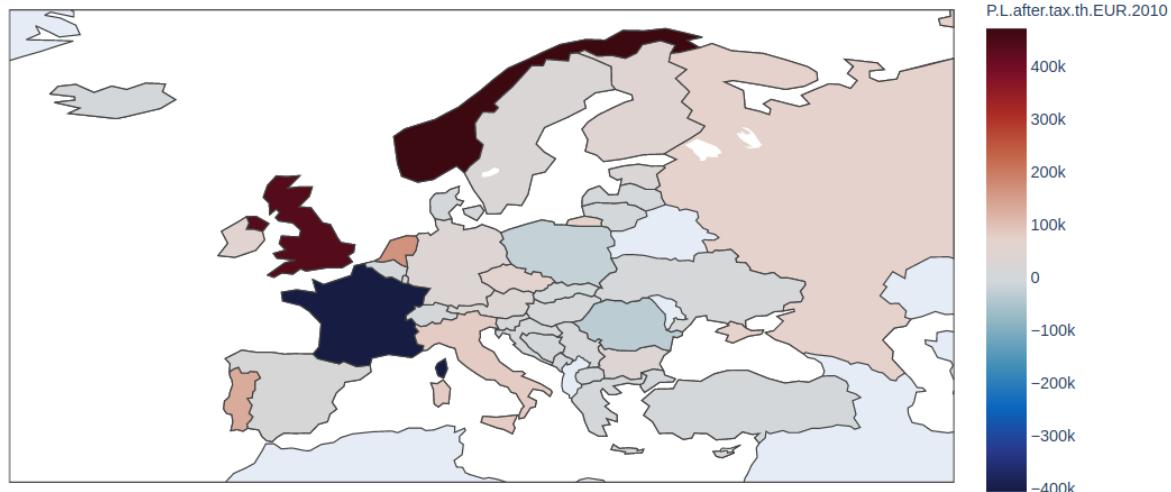
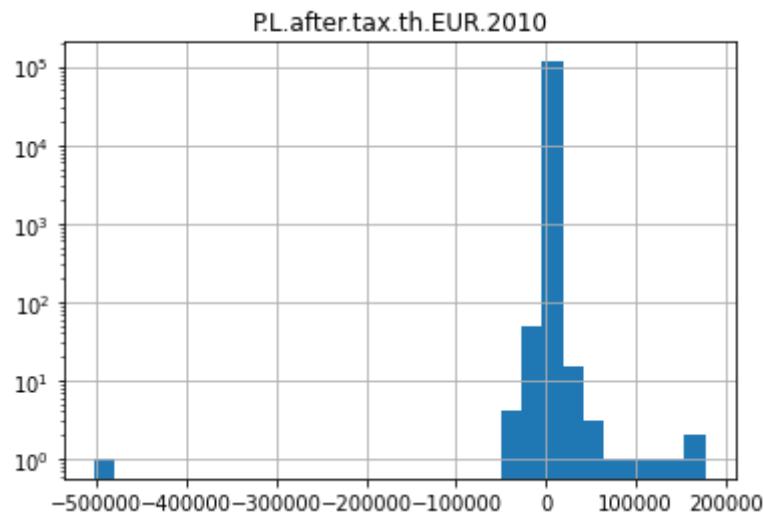
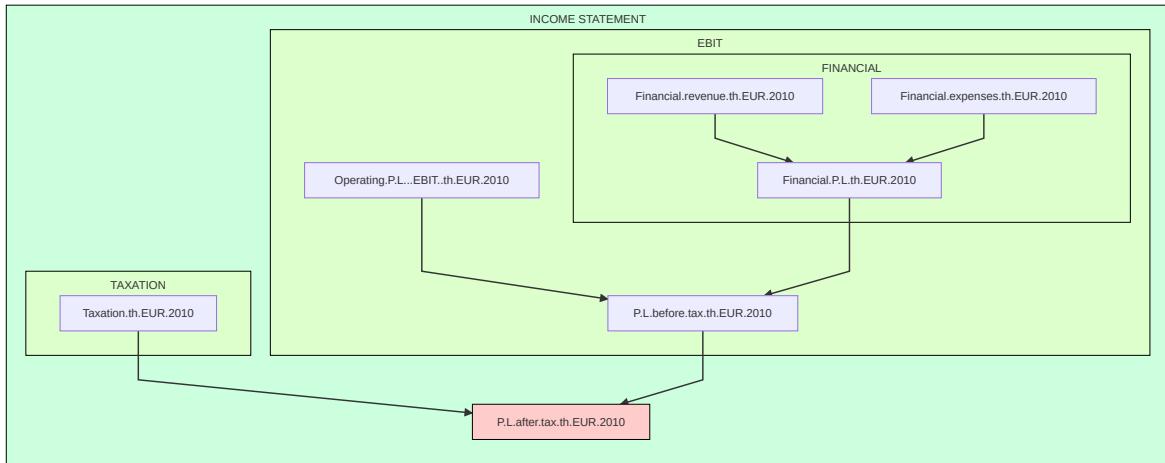
## Taxation.th.EUR.2010



outlier:

```
BvD.ID.number
GB07123187    ACACIA MINING PLC
Name: Company.name, dtype: string
```

**P.L.after.tax.th.EUR.2010**

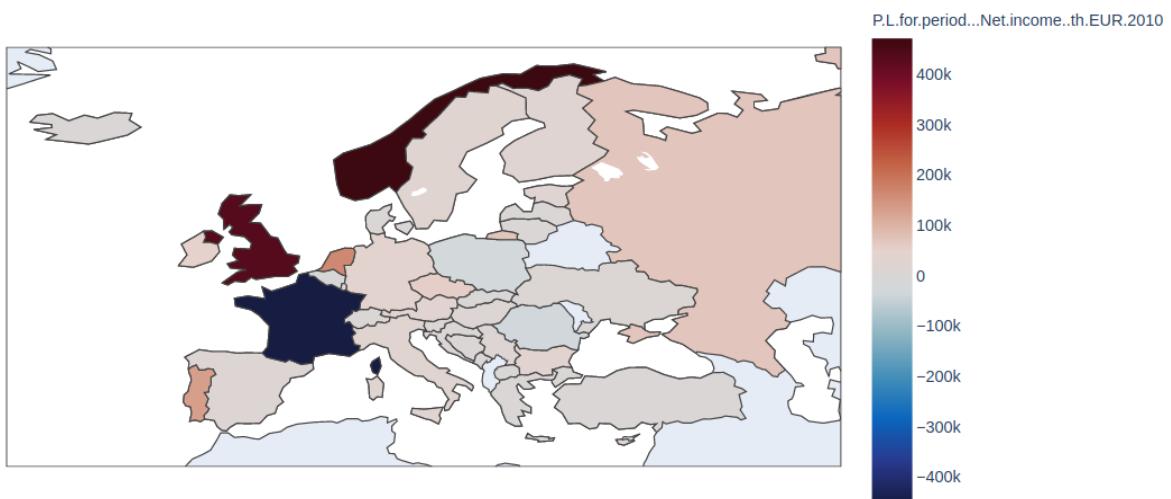
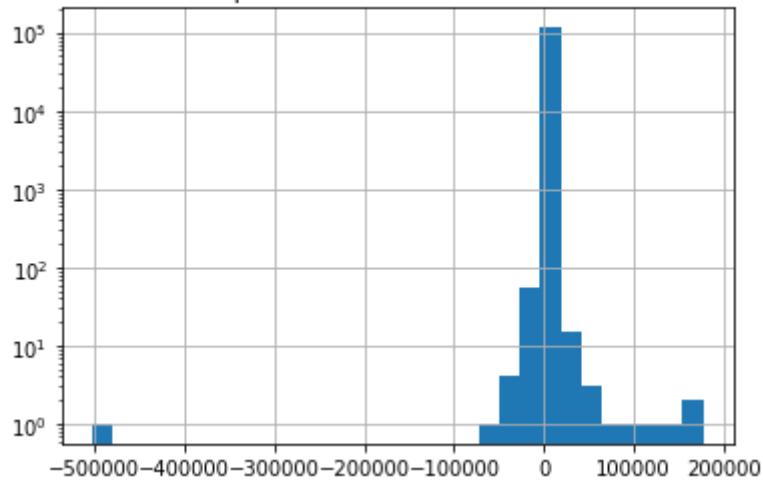


outlier:

BvD.ID.number  
FR519720643     IRIDIUM FRANCE  
Name: Company.name, dtype: string

**P.L.for.period...Net.income..th.EUR.2010**

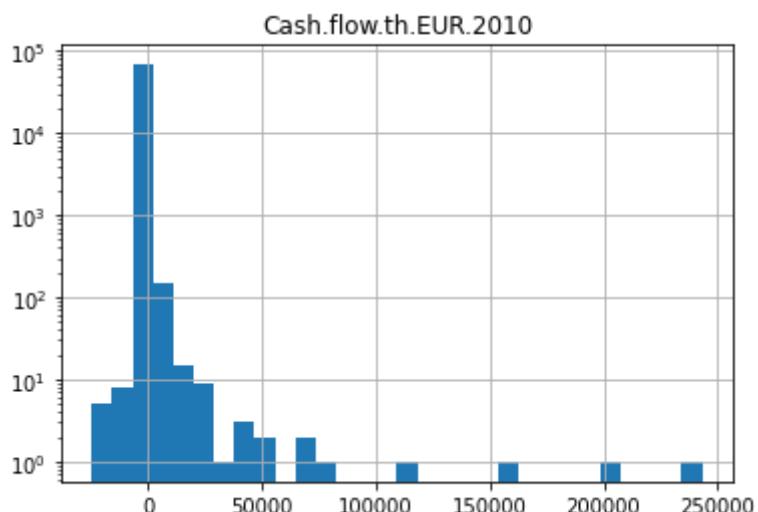
P.L.for.period...Net.income..th.EUR.2010

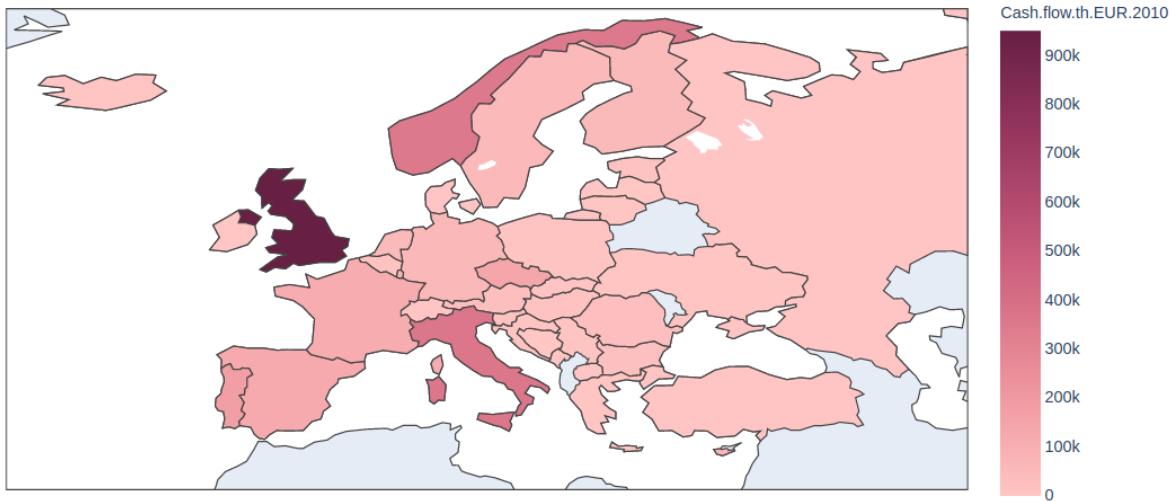


outlier:

```
BvD.ID.number  
FR519720643    IRIDIUM FRANCE  
Name: Company.name, dtype: string
```

## Cash.flow.th.EUR.2010

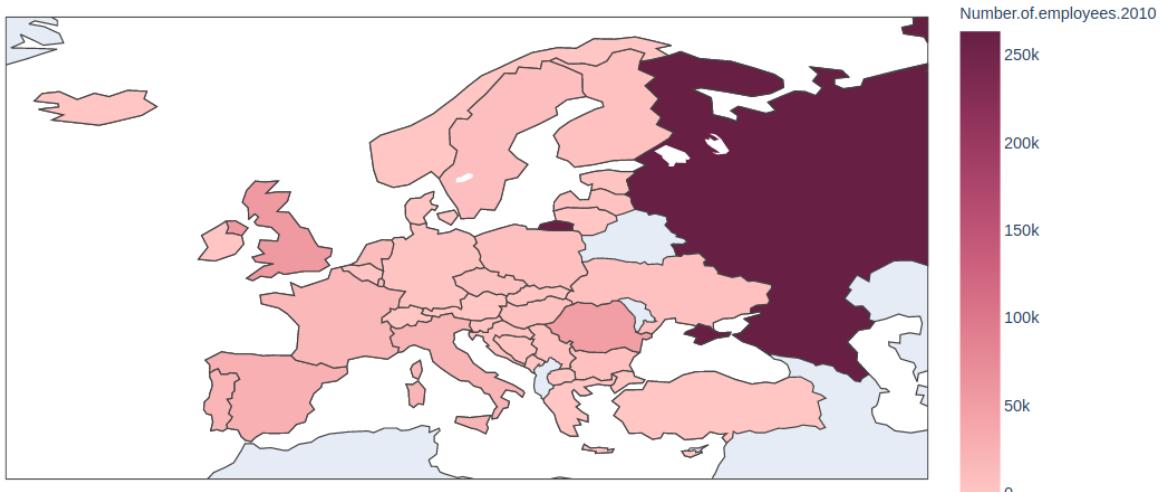
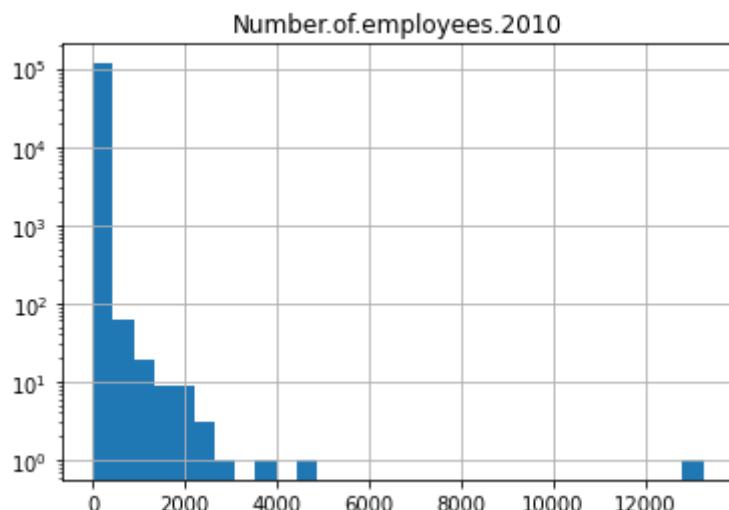




outliers:

BvD.ID.number	
GB07123187	ACACIA MINING PLC
GB07145051	CAPITAL & COUNTIES PROPERTIES PLC
GB07140891	ENQUEST PLC
PT509444229	MOTA-ENGIL AFRICA - SGPS, S.A.
Name: Company.name, dtype: string	

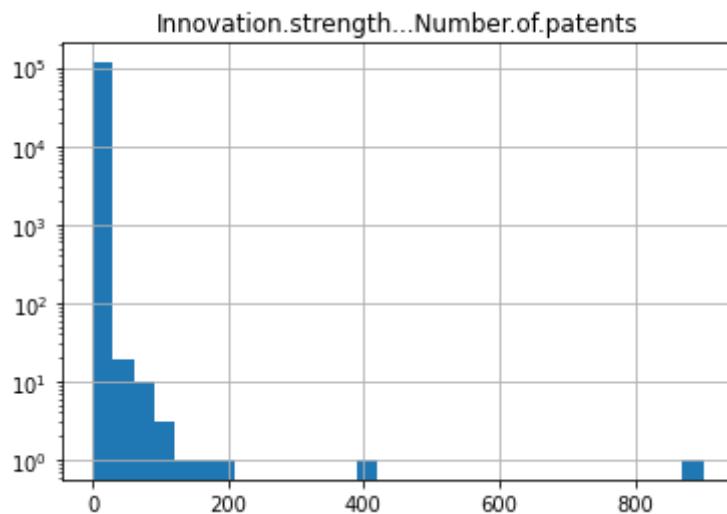
## Number.of.employees.2010



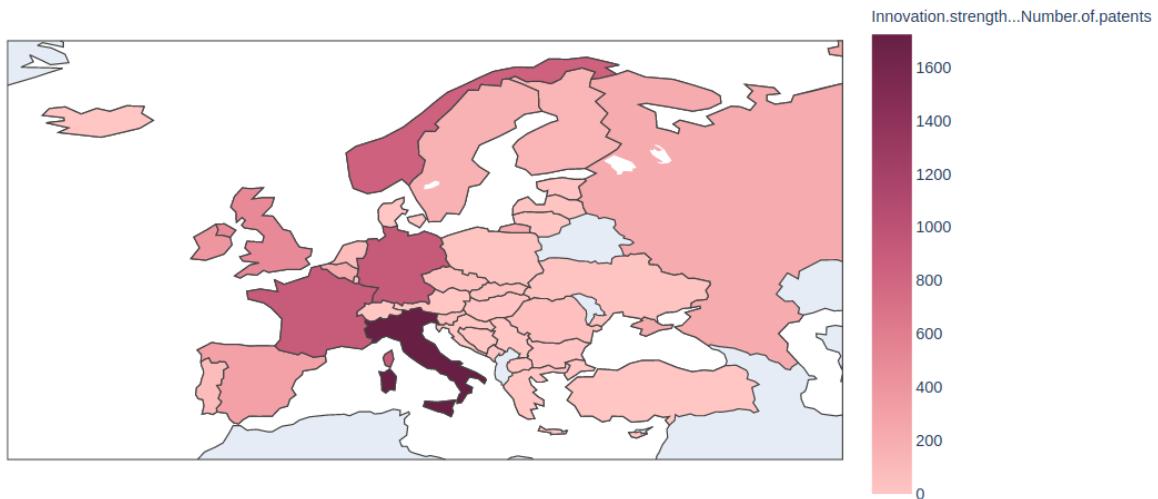
outlier:

```
BvD.ID.number  
GB07158140 CARE UK HEALTH & SOCIAL CARE INVESTMENTS LIMITED  
Name: Company.name, dtype: string
```

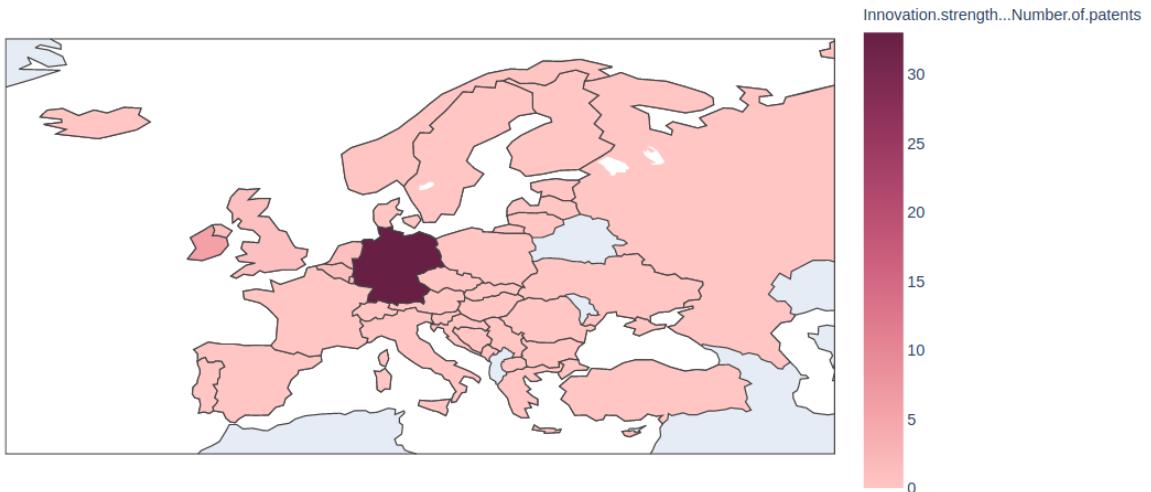
## Innovation.strength...Number.of.patents



sum:



mean:



outliers:

BvD.ID.number

IE507678

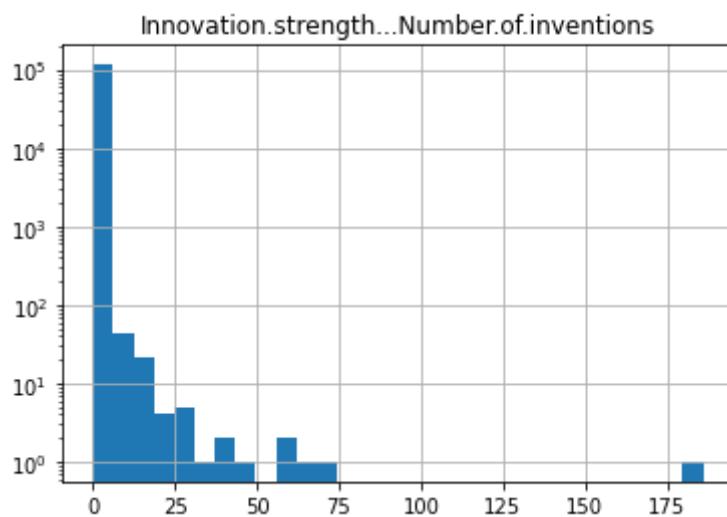
HORIZON THERAPEUTICS PUBLIC LIMITED COMPANY

DE8190460728

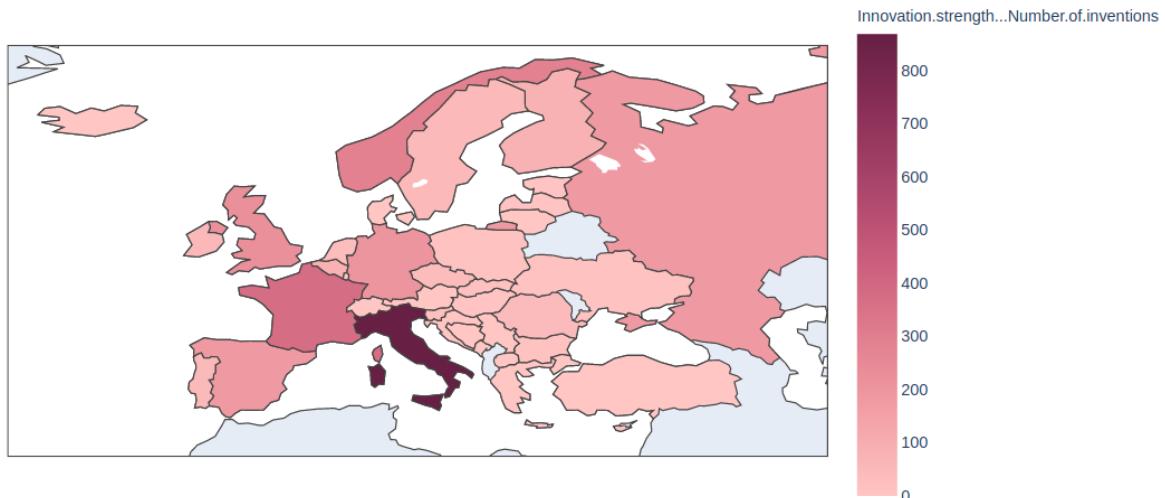
WAVELIGHT GMBH

Name: Company.name, dtype: string

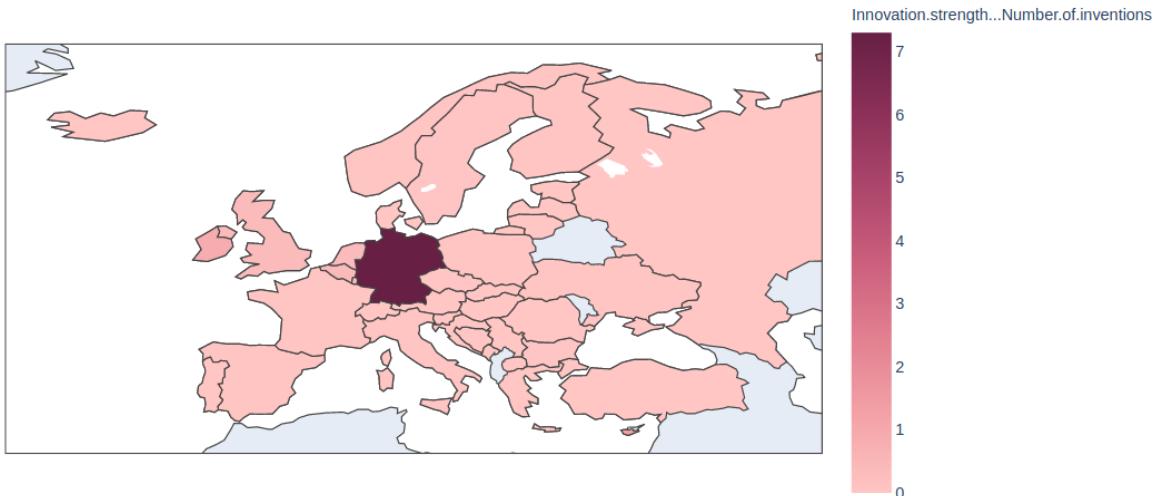
## Innovation.strength...Number.of.inventions



sum:



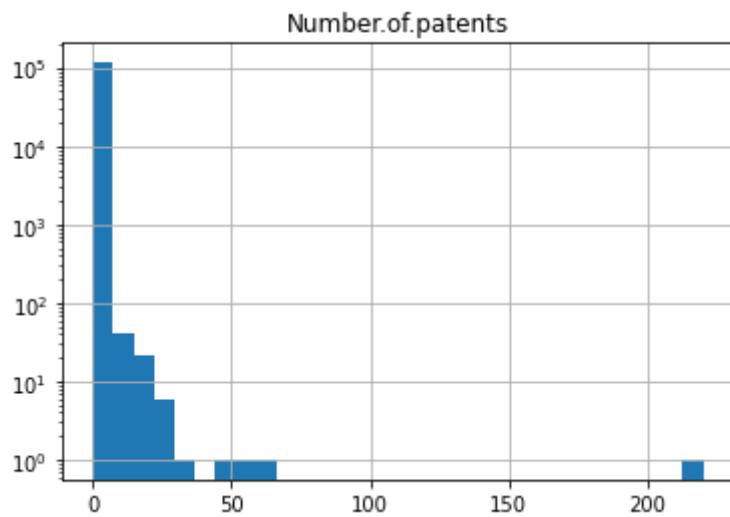
mean:



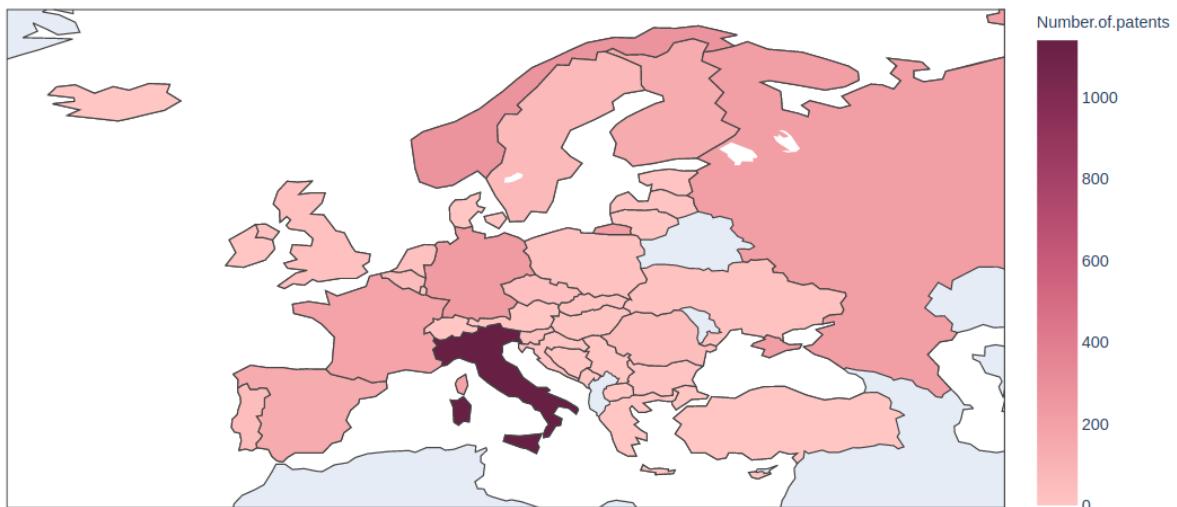
outlier:

BvD.ID.number  
DE8190460728 WAVELIGHT GMBH  
Name: Company.name, dtype: string

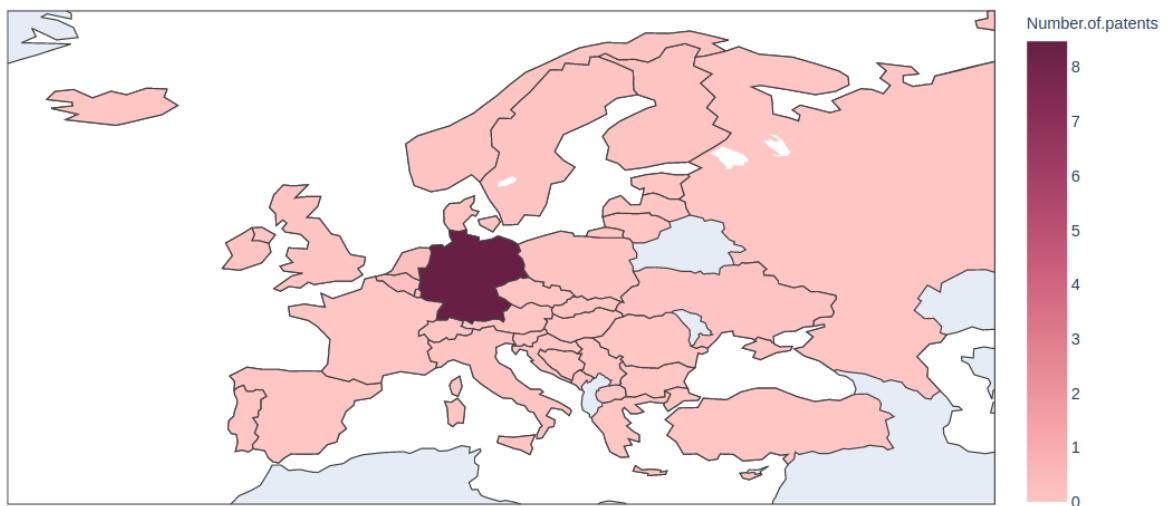
## Number.of.patents



sum:



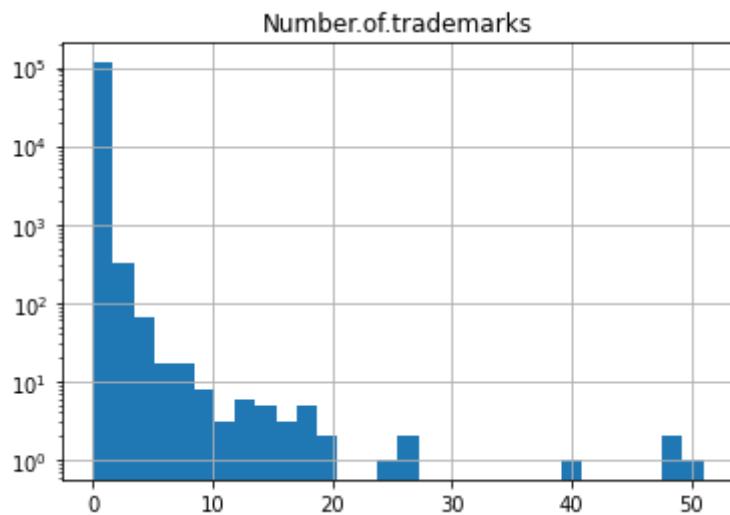
mean:



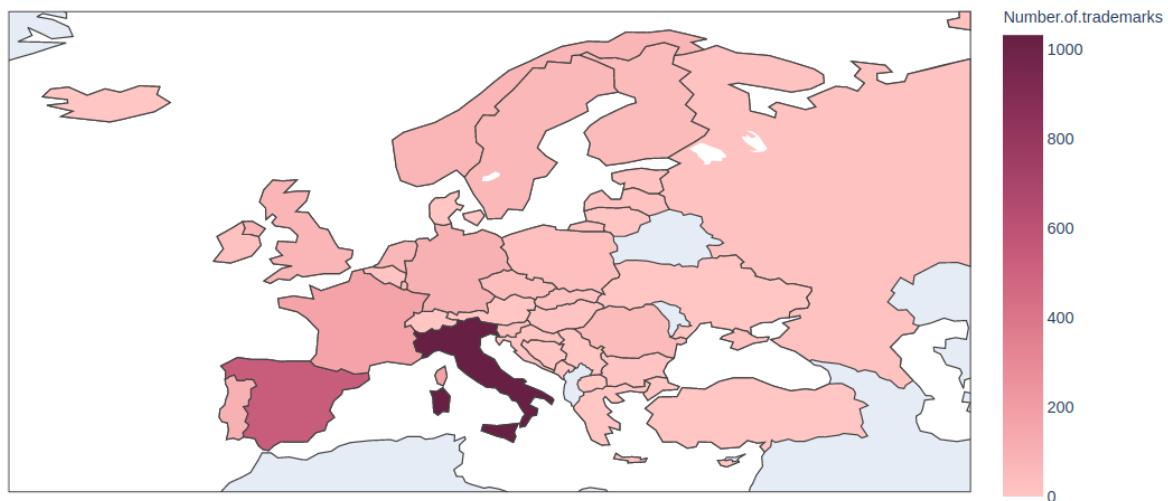
outlier:

BvD.ID.number  
DE8190460728 WAVELIGHT GMBH  
Name: Company.name, dtype: string

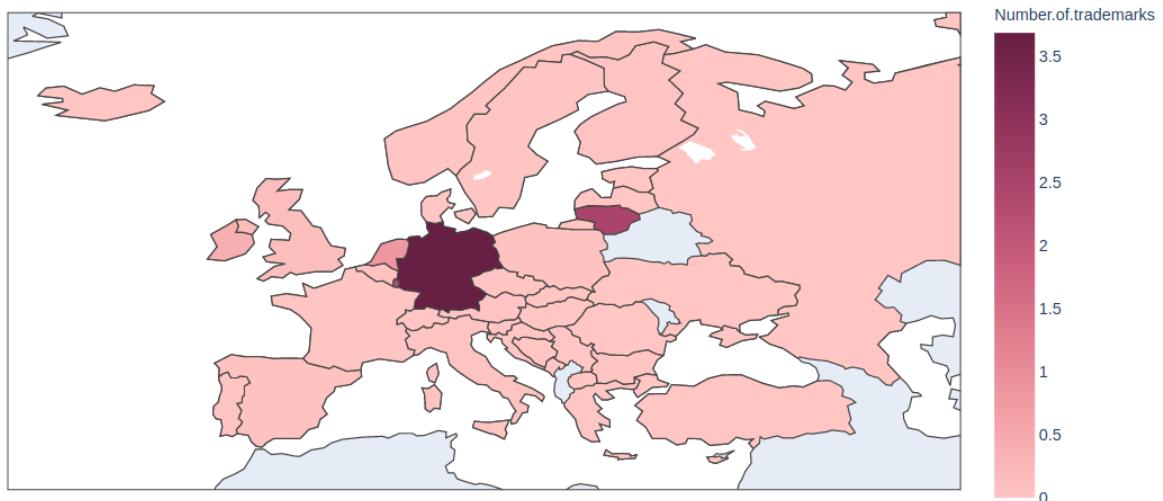
## Number.of.trademarks



sum:



mean (Lituania!):



outliers:

BvD.ID.number	
DE4250480683	VOLMARY GMBH
DE8190460728	WAVELIGHT GMBH
IT07237530964	SALROS S.R.L.
LULB157784	MEDA PHARMA SARL
Name: Company.name, dtype: string	

## Trademarks...Type

BvD.ID.number	
R026941545	No
PT509482171	No
IT06789671218	No
FR519321806	No
N0995113562	No
Name: Trademarks...Type, dtype: category	
Categories (4, object): [Figurative, No, Other, Word]	

Trademarks...Type	n
Figurative	779
Word	400
Other	3
No	114658
<b>Total</b>	<b>115840</b>

## HGF

High Growth Firm. Boolean. There are many definitions of HGF.

### CAGR

Companies with an average growth rate  $\geq 20\%$  for the first 5 years:

$$CAGR = \left( \frac{\text{turnover}_{2014}}{\text{turnover}_{2010}} \right)^{1/4} - 1 \geq 20\%$$

### Gazelle

From Birch, 1995, gazeles are firm with a growth rate that remains  $>20\%$  for the first 5 years

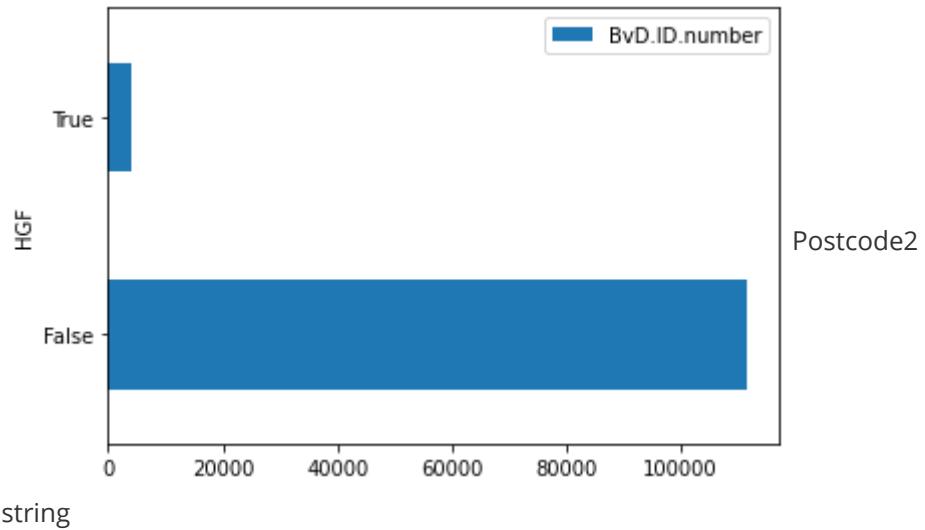
$$\text{Gazelle} = \text{all} \left( \frac{\text{turnover}_t}{\text{turnover}_{t-1}} \geq 20\% \right), \text{for } t = 2010, \dots, 2014$$

### Eurostat

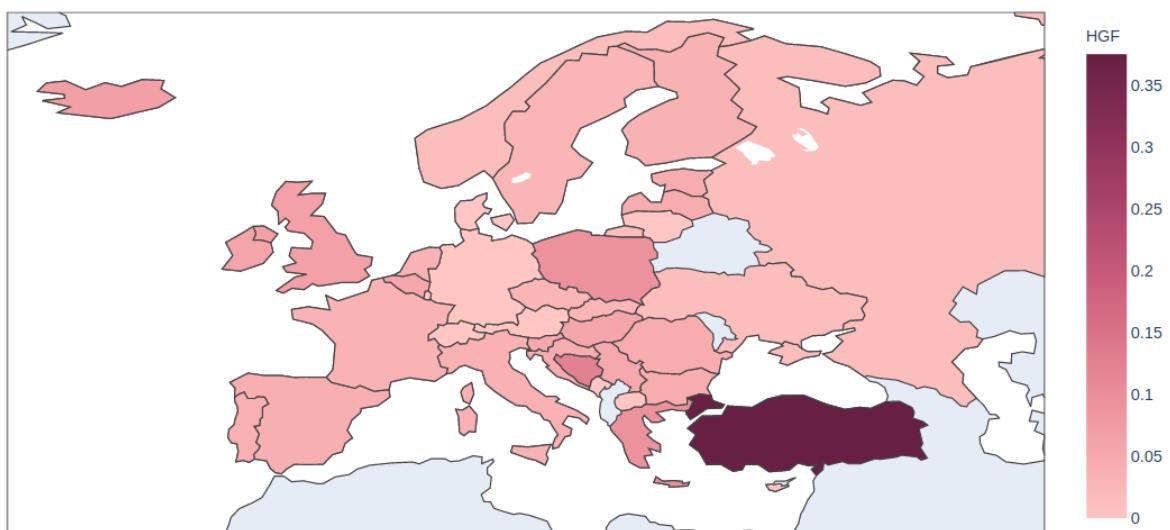
By Eurostat, 2007, being HGF means having a growth rate  $\geq 20\%$  for 3 consecutive years.

$$\text{Eurostat} = \exists t \in \{2010, 2011, 2012\} \left( \frac{\text{turnover}_t}{\text{turnover}_{t-1}} \geq 20\% \wedge \frac{\text{turnover}_{t+1}}{\text{turnover}_t} \geq 20\% \wedge \frac{\text{turnover}_{t+2}}{\text{turnover}_{t+1}} \geq 20\% \right)$$

In this dataset HGF is computed by the second option, Gazele.

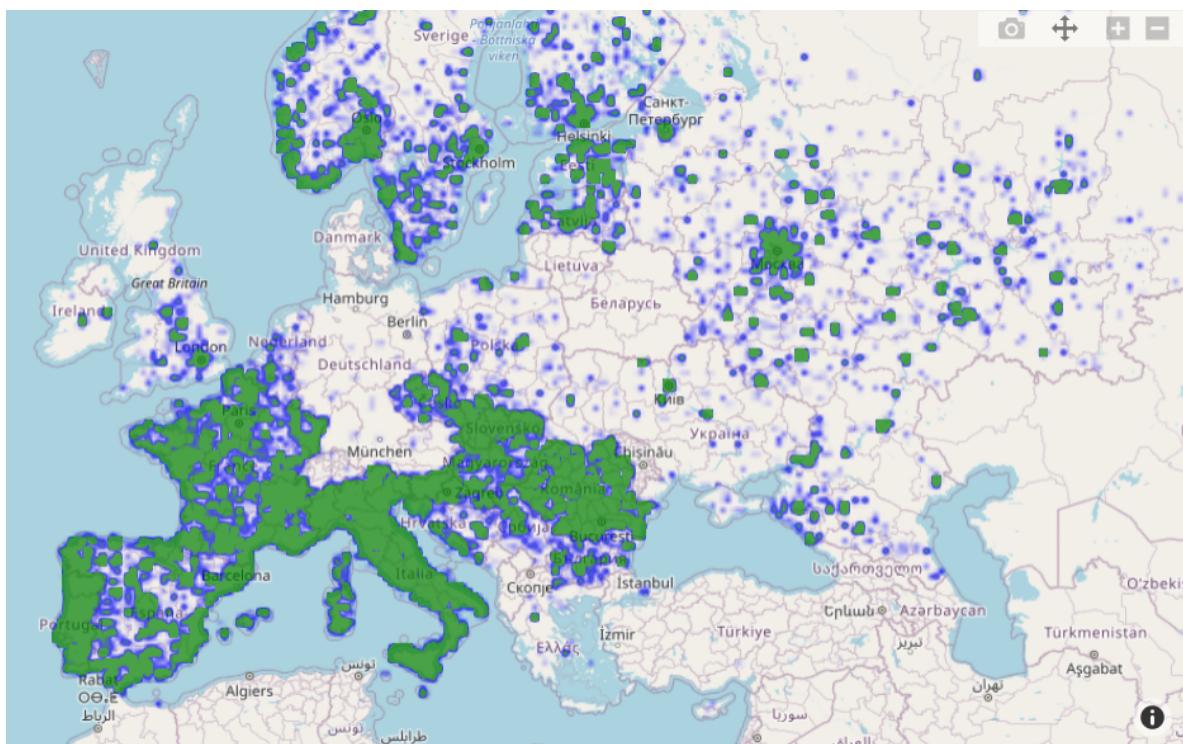
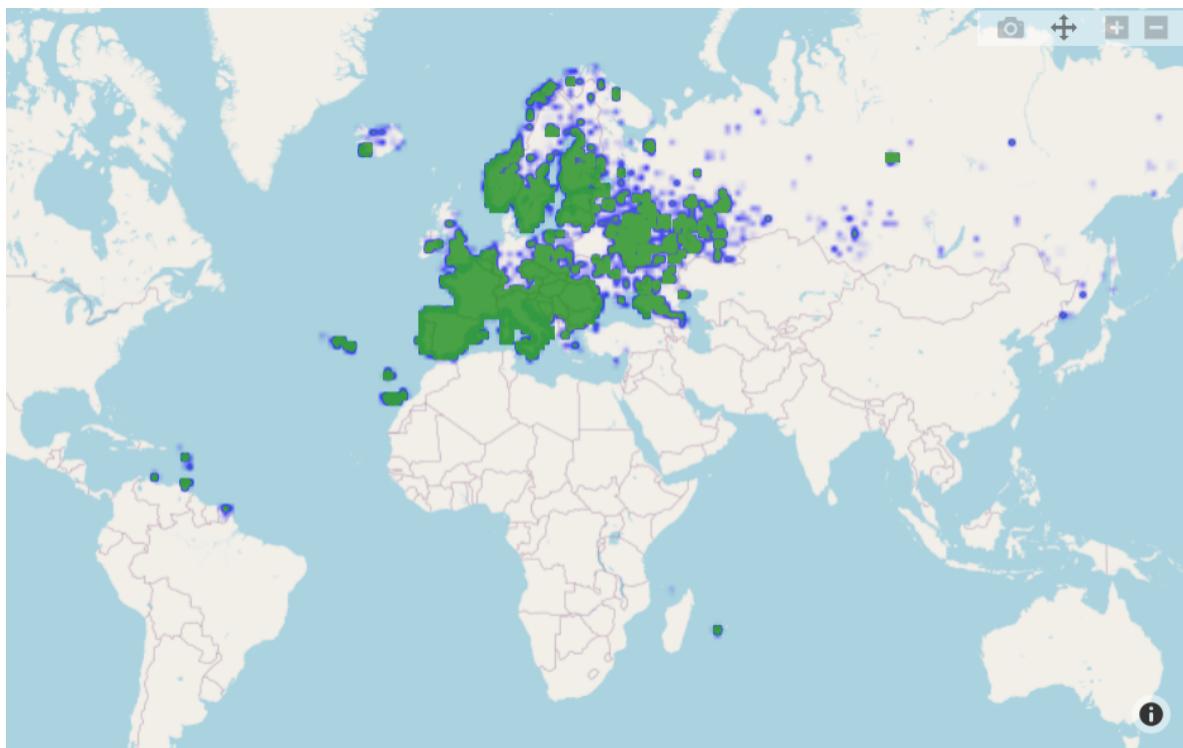


$p(HGF|country)$ :



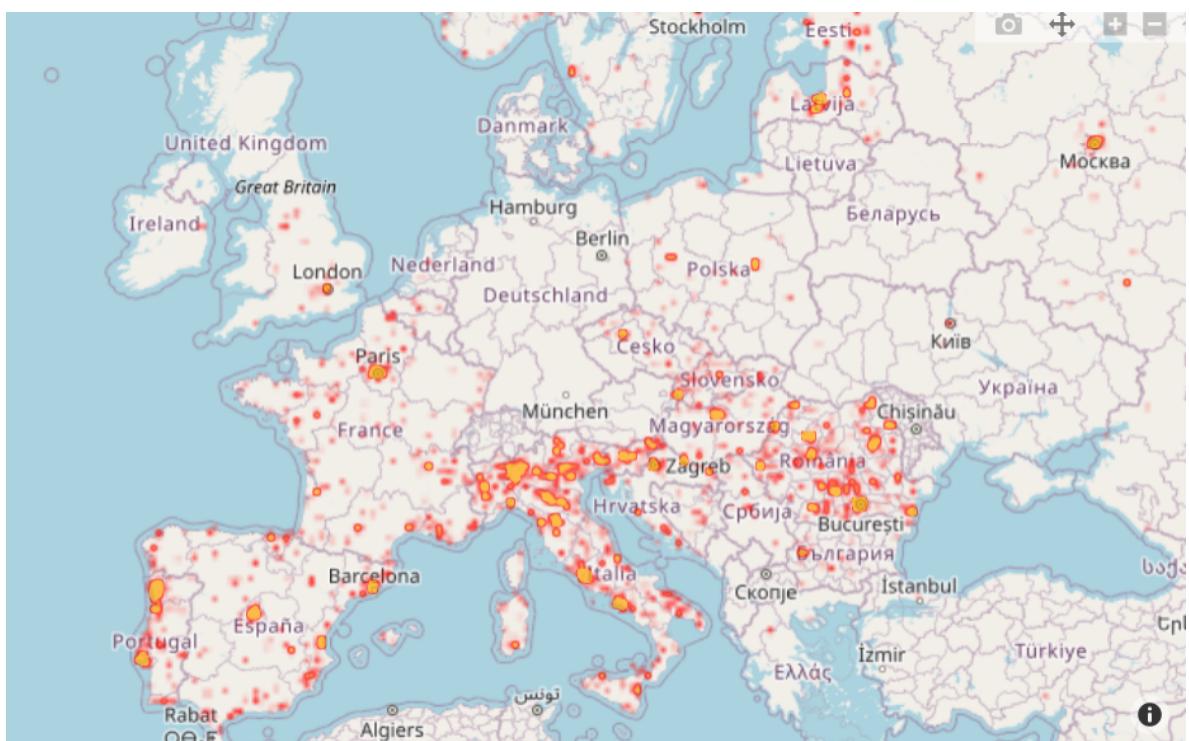
## lat / lon

Geographic coordinates of the firm. Not available for ~8000 samples.





HGF=True:



(These maps are navigable like Google Maps on my pc, I can provide an interactive version.)

## trust

Trust evaluation, as string made of two parts.

- A class {Low, Medium, High}
- A firm name, maybe the evaluator?

```

BvD.ID.number
R026547053 <NA>
HR05541138225 Medium: [DAKOVO, 31400, Croatia] instead of [C...
IT06738671210 High: [ESSECI ITALIA S.R.L., POGGIOMARINO, 800...
ESB72172646 High: [NUMENTI SL, PUERTO REAL, 11510, Spain]
IT02439920352 High: [SAN TOMMASO S.R.L., CANTU, 22063, Italy]
Name: trust, dtype: string

```

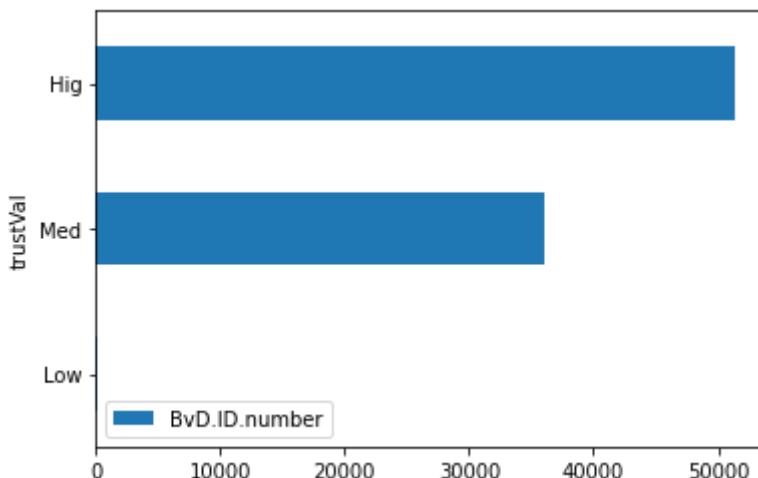
## trustVal

To give meaning to `trust`, I've created a variable with the ranked level of trust. Class labels are in `{'Low','Med','Hig'}`.

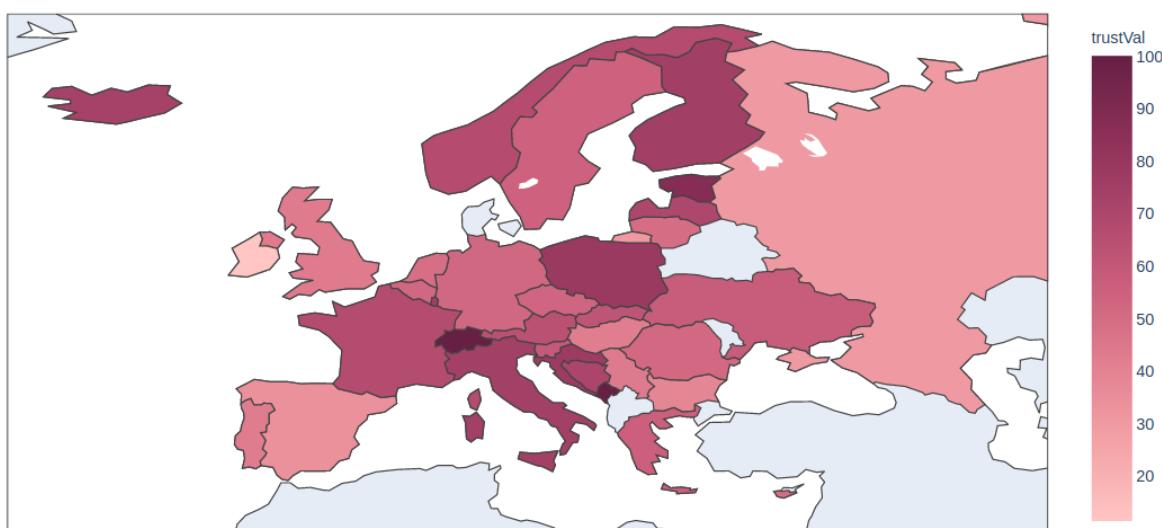
```

BvD.ID.number
IT04844800658 Hig
SE5568281116 Med
IT07016280963 Med
GB07234353 NaN
FR519497507 Hig
Name: trustVal, dtype: category
Categories (3, object): [Low < Med < Hig]

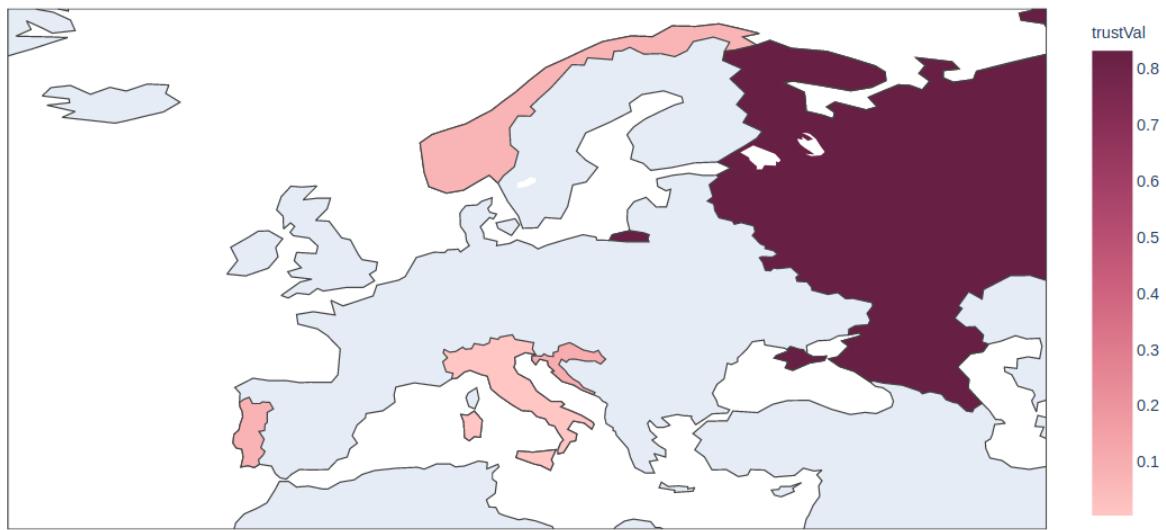
```



Percentage of companies with "High" rating:



Percentage of companies with "Low" rating (*note the different y axis*):



mean trust: (high = 2, low = 0)

