

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	

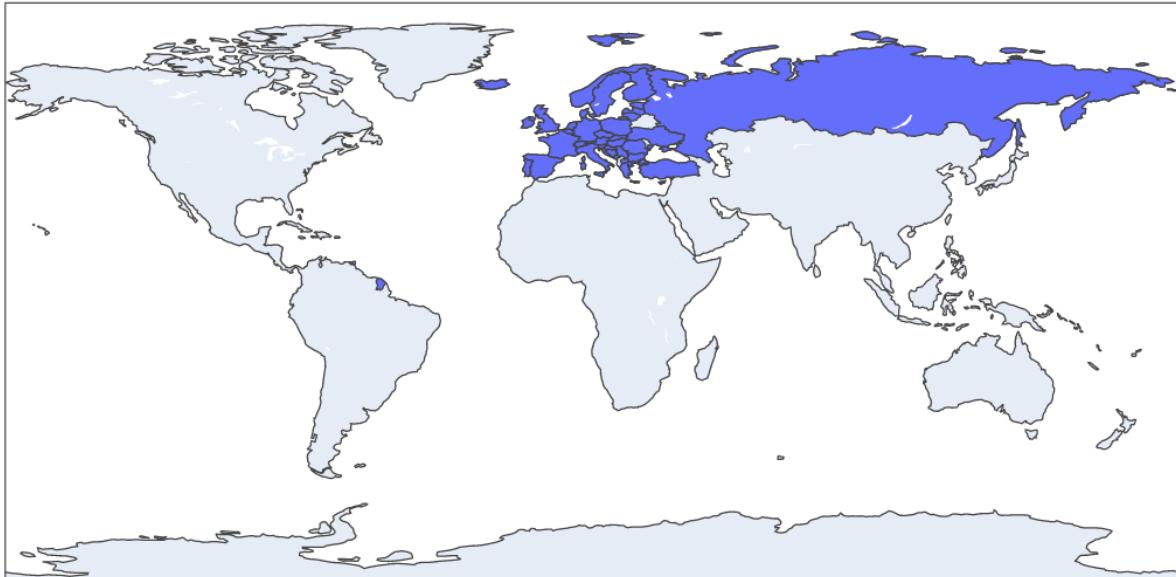
Country.ISO.Code

`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

Country.ISO.Code	AT	BA	BE	BG	CH	CY	CZ	DE	DK	EE	...	PL	\
HGF											...		
0	19	97	190	626	2	3	1985	28	1	2489	...	446	
1	0	14	11	31	0	0	60	0	0	114	...	46	

Country.ISO.Code	PT	RO	RS	RU	SE	SI	SK	TR	UA
HGF									
0	6535	15500	361	11731	2512	1693	3497	5	374
1	265	758	20	192	76	92	106	3	6

[2 rows x 38 columns]

```
Chi2 = 4.79e+02
p     = 3.4e-78
degrees of freedom = 37
```

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

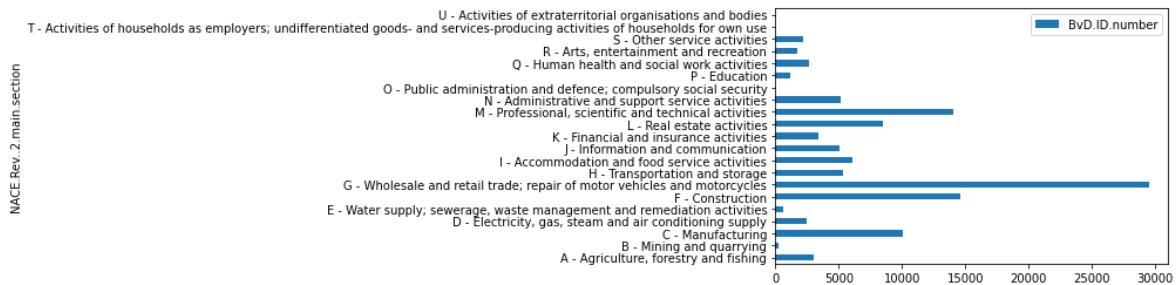
```

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	A	B	C	D	E	F	G	H	I	J	...	L
HGF											...	\
0	2938	213	9621	2429	593	14212	28334	5038	5942	4779	...	8340
1	99	8	444	33	43	385	1192	294	154	266	...	146
NACE	M	N	O	P	Q	R	S	T	U			
HGF												
0	13611	4899	16	1116	2504	1646	2151	2	1			
1	460	244	0	66	129	75	64	0	0			

[2 rows x 21 columns]

Chi² = 3.96e+02
p = 1.6e-71
degrees of freedom = 20

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	100	110	111	112	113	115	119	120	121	122	...	9529	\
HGF											...		
0	18	60	571	8	227	1	64	17	117	1	...	34	
1	2	1	20	0	5	0	1	1	7	0	...	1	
NACE	9600	9601	9602	9603	9604	9609	9700	9810	9900				
HGF													
0	4	131	957	129	186	351	1	1	1				
1	0	3	20	4	6	16	0	0	0				

[2 rows x 729 columns]

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

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

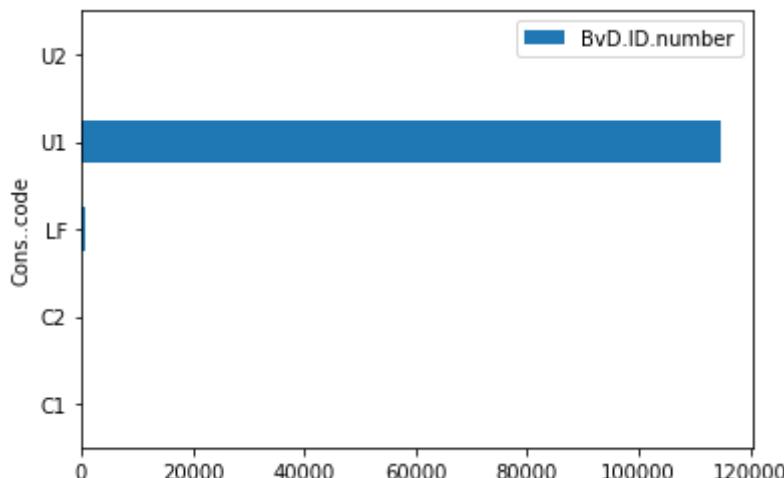
```
BvD.ID.number
RU65230449      4633
N0995818558      4759
SE5568024276      7311
N0995182165      3312
R027703886       4613
Name: NACE.Rev..2.Primary.code.s., dtype: category
Categories (729, int64): [100, 110, 111, 112, ..., 9609, 9700, 9810, 9900]
```

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 a 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]
```



CONS	C1	C2	LF	U1	U2
HGF					
0	116	103	784	110504	140
1	7	3	7	4170	6

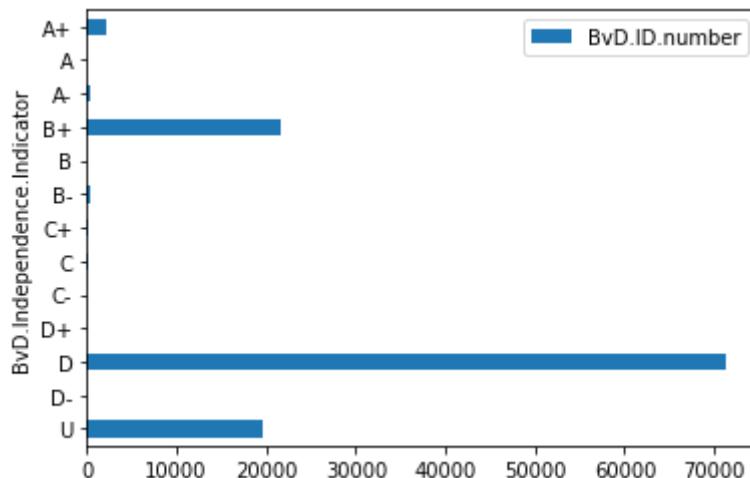
Chi² = 18.9
 p = 0.00084
 degrees of freedom = 4

BvD.Independence.Indicator

Characterize the degree of independence of a company with regard to its shareholders. It has been mapped to an 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+]
```



INDEP	-	A	A+	A-	B	B+	B-	C	C+	D	U
HGF											
0	3	20	2170	257	1	20777	346	104	197	68757	19015
1	0	1	63	21	0	837	16	13	8	2588	646

Chi² = 46.4
 p = 1.2e-06
 degrees of freedom = 10

BvD.major.sector

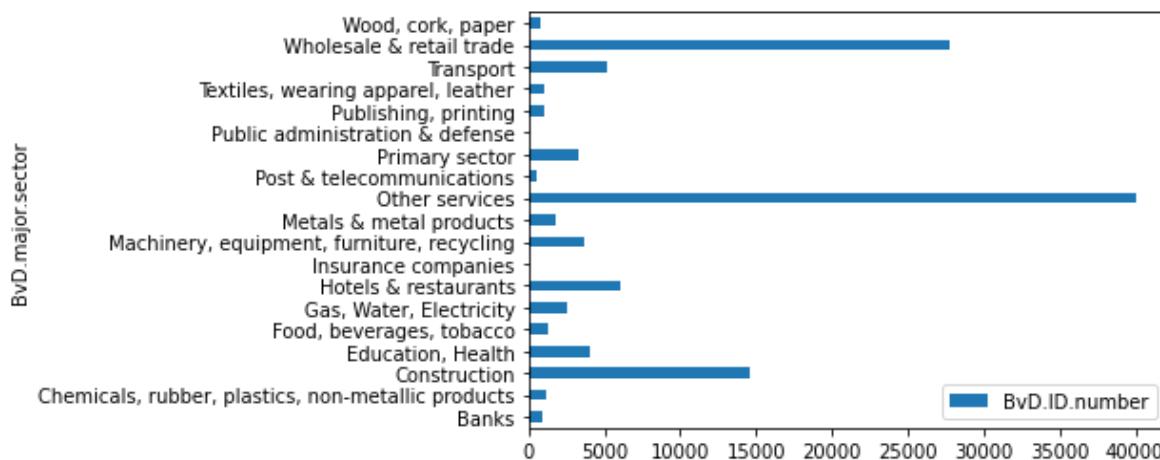
A different sector encoding from BvD. It encompasses 19 categories.

```

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.

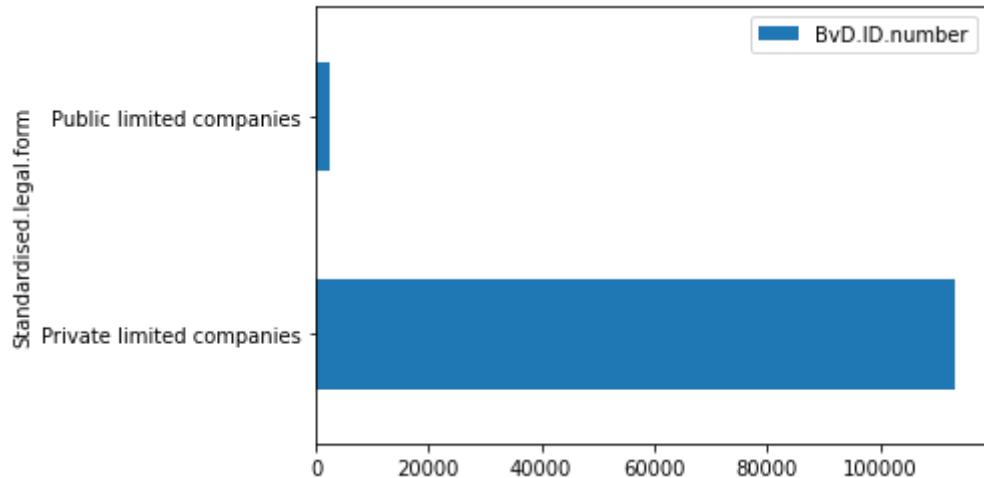


SECT	Banks	Chemicals, rubber, plastics, non-metallic products	Construction	Education, Health	Food, beverages, tobacco	Gas, Water, Electricity	Wholesale & retail trade	Wood, cork, paper
HGF								\
0	947	1156	14212	3809	1252	2494		
1	22	58	385	211	74	36		
SECT	Hotels & restaurants	Insurance companies	Machinery, equipment, furniture, recycling	Metals & metal products	Other services	Post & telecommunications	Primary sector	\
HGF								
0	5942	19	3466	1712	38654	447		
1	154	2	149	64	1340	39		
SECT	Primary sector	Public administration & defense	Publishing, printing	Textiles, wearing apparel, leather	Transport	Wholesale & retail trade	Wood, cork, paper	\
HGF								
0	3148	17	1042	1000	4919	26666		
1	107	0	43	52	286	1137		
SECT	Wood, cork, paper							
HGF								
0	745							
1	34							
Chi ²	=	2.7e+02						
p	=	7.4e-47						
degrees of freedom	=	18						

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]
```



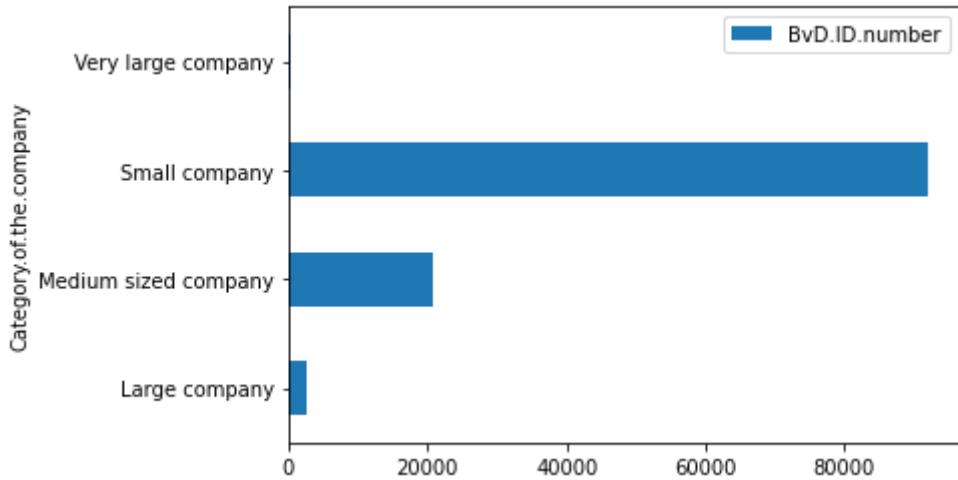
```
FORM  Private  Public
HGF
0      109204    2443
1       4056     137

Chi^2 =      21.6
p      =  3.3e-06
degrees of freedom = 1
```

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]
```

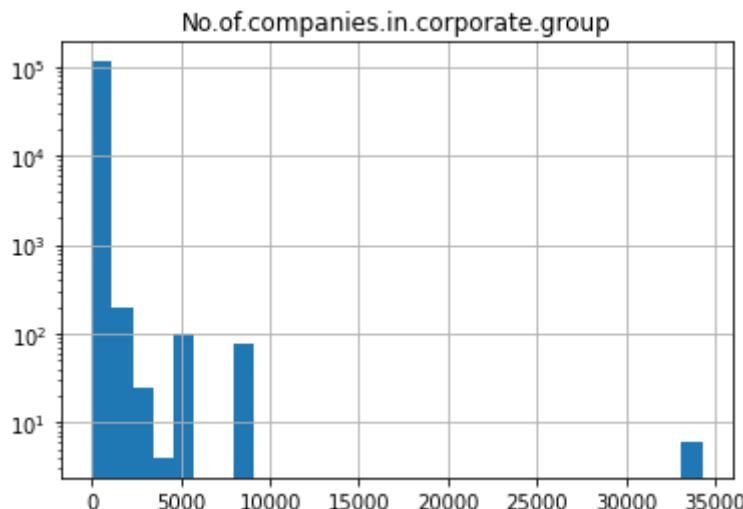


CAT	Large company	Medium sized	Small company	Very large c
HGF				
0	2485	19276	89542	344
1	267	1522	2373	31

$\text{Chi}^2 = 1.41\text{e+03}$
 $p = 2.3\text{e-306}$
 degrees of freedom = 3

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	Company.name
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.

```
HGF vs non-HGF for No.of.companies.in.corporate.group
Welch's t-test statistic = -0.7536
p-value = 0.4511
```

Optimization terminated successfully.

Current function value: 0.155660

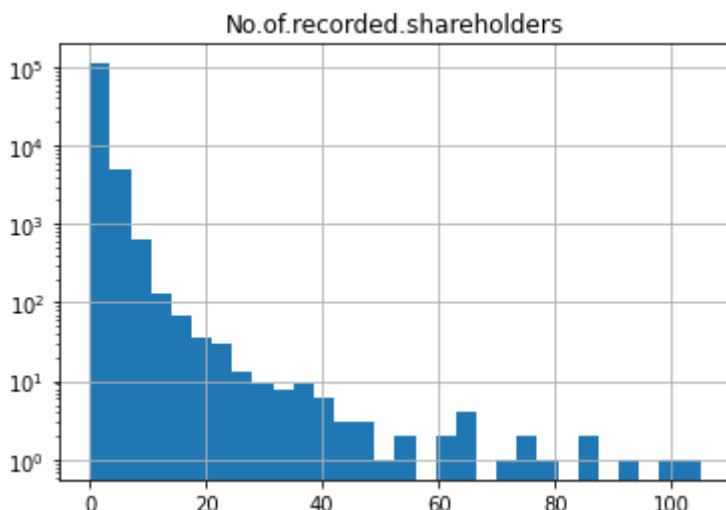
Iterations 7

Logit Regression Results

```
=====
Dep. Variable: HGF   No. Observations: 115840
Model: Logit    Df Residuals: 115838
Method: MLE     Df Model: 1
Date: Mon, 29 Jun 2020 Pseudo R-squ.: 1.205e-05
Time: 15:23:49 Log-Likelihood: -18032.
converged: True   LL-Null: -18032.
Covariance Type: nonrobust LLR p-value: 0.5098
=====
            coef      std err      z      P>|z|      [0.025      0.975]
-----
Intercept  -3.2825    0.016  -208.368      0.000     -3.313     -3.252
CORPGRP    2.408e-05  3.39e-05      0.710      0.478    -4.24e-05  9.06e-05
=====
```

No.of.recorded.shareholders

Numbers of recorded stakeholders.



```
HGF vs non-HGF for No.of.recorded.shareholders
Welch's t-test statistic = -4.809
p-value = 1.571e-06
```

Optimization terminated successfully.

Current function value: 0.155556

Iterations 7

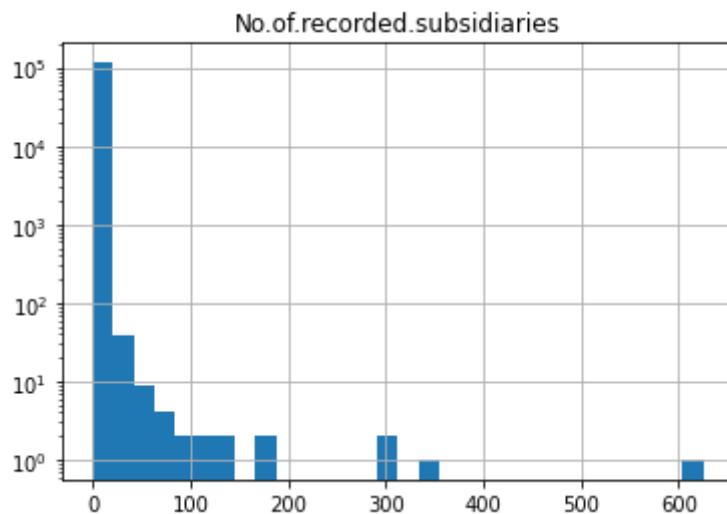
Logit Regression Results

```
=====
Dep. Variable: HGF   No. Observations: 115840
Model: Logit    Df Residuals: 115838
=====
```

```

Method: MLE Df Model: 1
Date: Mon, 29 Jun 2020 Pseudo R-squ.: 0.0006832
Time: 15:25:05 Log-Likelihood: -18020.
converged: True LL-Null: -18032.
Covariance Type: nonrobust LLR p-value: 6.913e-07
=====
          coef    std err      z     P>|z|    [0.025    0.975]
-----
Intercept -3.3319    0.018 -183.049    0.000   -3.368   -3.296
SHA        0.0315    0.006    5.698    0.000    0.021    0.042
=====
```

No.of.recorded.subsidiaries



outlier:

```

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

```

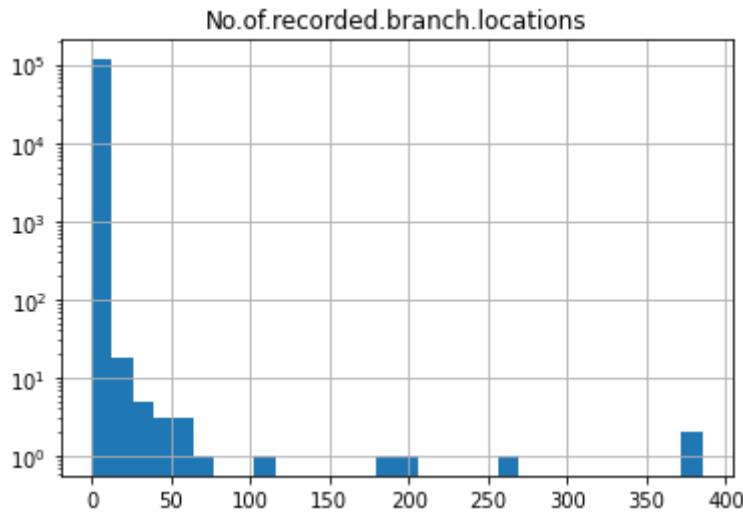
HGF vs non-HGF for No.of.recorded.subsidiaries
Welch's t-test statistic = -3.294
p-value = 0.0009946
```

```

Logit Regression Results
=====
Dep. Variable: HGF   No. Observations: 115840
Model: Logit   Df Residuals: 115838
Method: MLE   Df Model: 1
Date: Mon, 29 Jun 2020 Pseudo R-squ.: 8.499e-05
Time: 15:25:58 Log-Likelihood: -18030.
converged: True LL-Null: -18032.
Covariance Type: nonrobust LLR p-value: 0.08000
=====
          coef    std err      z     P>|z|    [0.025    0.975]
-----
Intercept -3.2831    0.016 -208.501    0.000   -3.314   -3.252
=====
```

SUB	0.0052	0.003	2.045	0.041	0.000	0.010
-----	--------	-------	-------	-------	-------	-------

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
```

```
HGF vs non-HGF for No.of.recorded.branch.locations
Welch's t-test statistic = -3.193
p-value = 0.001417
```

Optimization terminated successfully.

Current function value: 0.155622

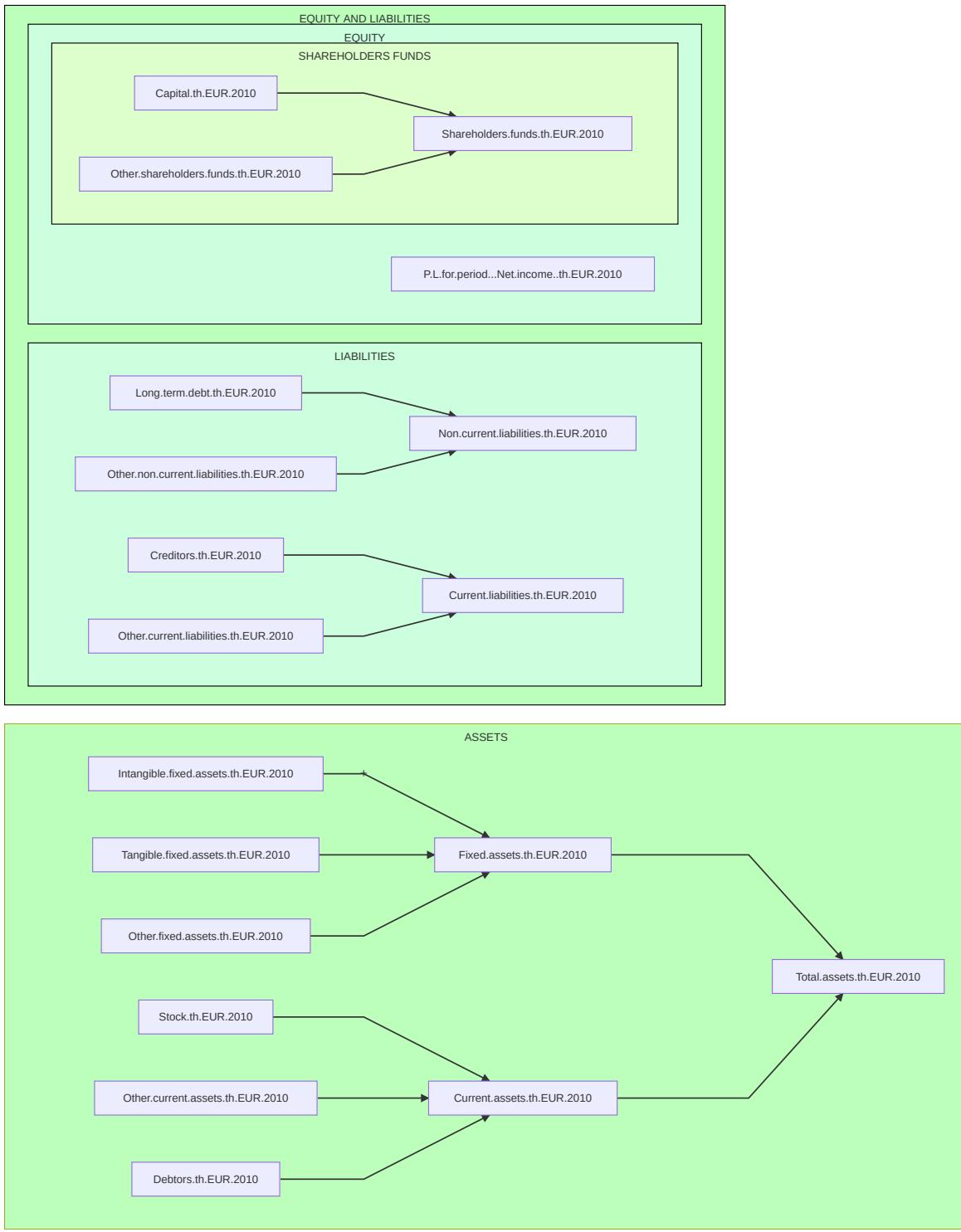
Iterations 7

Logit Regression Results

```
=====
Dep. Variable:           HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:          115838
Method:                MLE    Df Model:                  1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:     0.0002588
Time:          15:26:33    Log-Likelihood:   -18027.
converged:            True    LL-Null:        -18032.
Covariance Type:    nonrobust   LLR p-value:   0.002250
=====
```

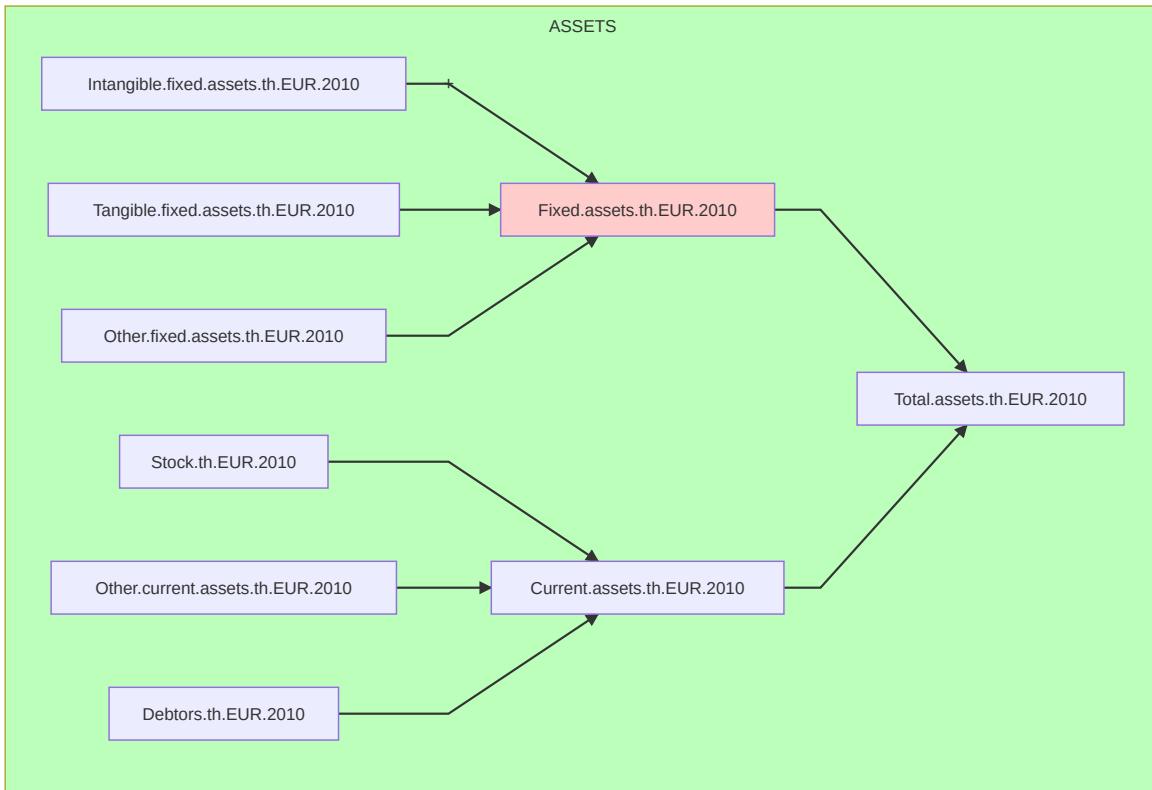
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2833	0.016	-208.582	0.000	-3.314	-3.252
BRA	0.0107	0.003	3.303	0.001	0.004	0.017

Balance sheet

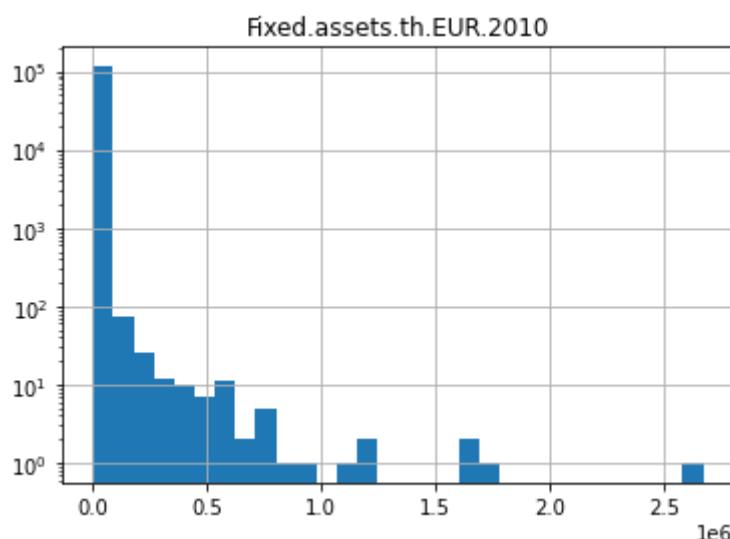


Assets

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
  
```

```

HGF vs non-HGF for Fixed.assets.th.EUR.2010
Welch's t-test statistic = 3.679
p-value = 0.0002362
  
```

Optimization terminated successfully.

Current function value: 0.155638

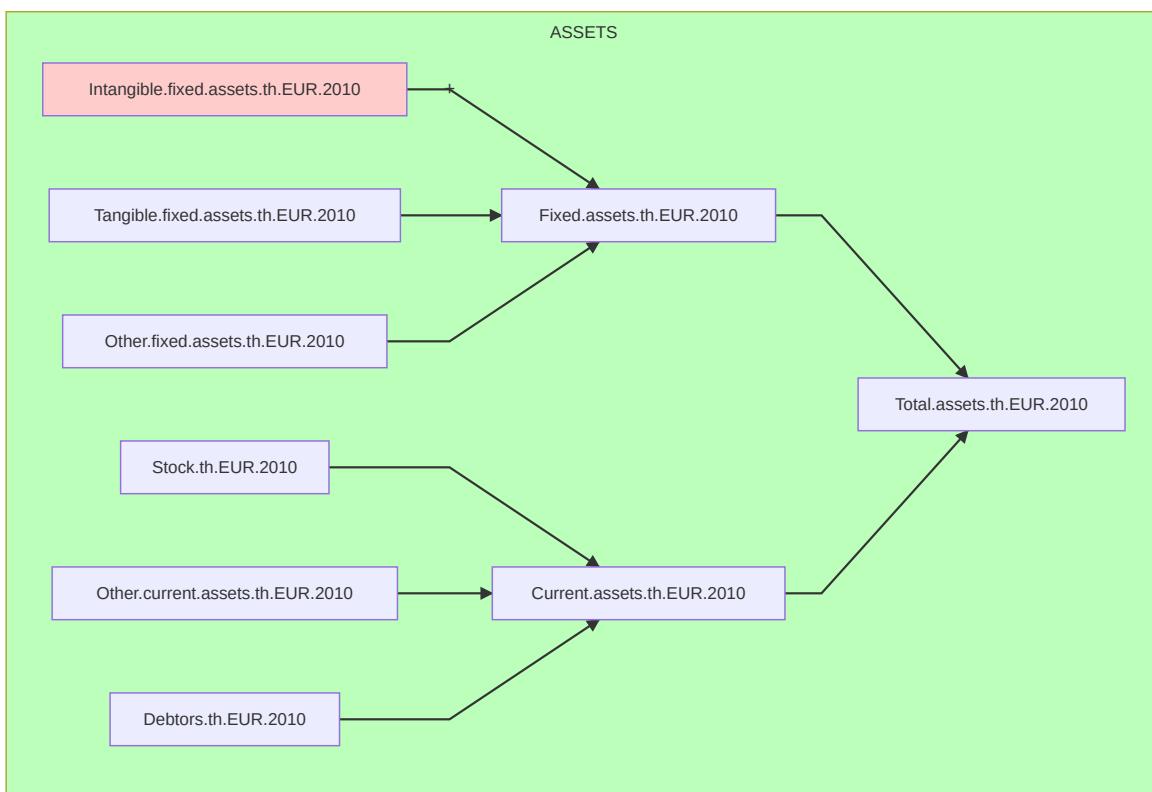
Iterations 9

Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.0001559
Time:	15:27:08	Log-Likelihood:	-18029.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	0.01773

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2793	0.016	-207.847	0.000	-3.310	-3.248
FX	-6.137e-06	3.82e-06	-1.607	0.108	-1.36e-05	1.35e-06

Intangible.fixed.assets.th.EUR.2010

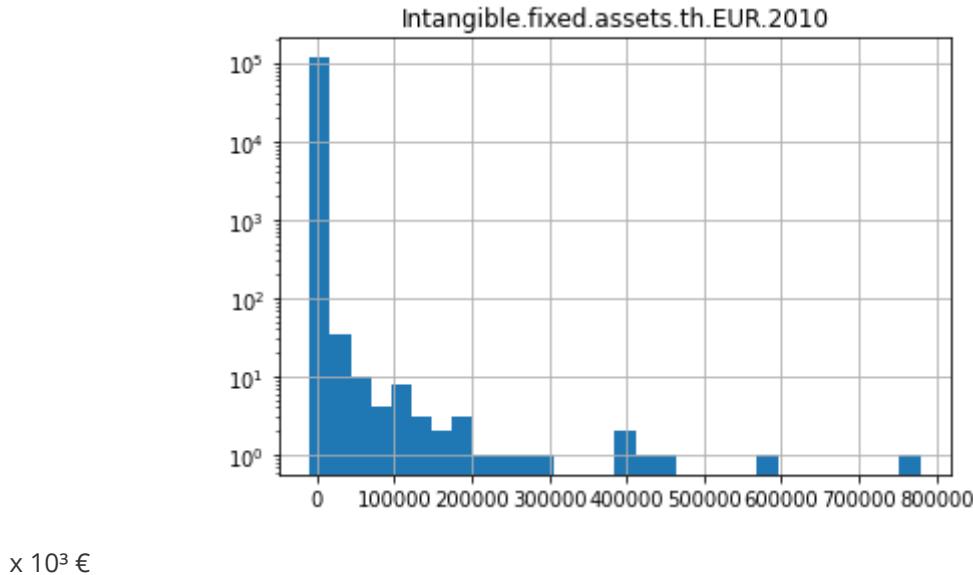


Intangible assets include operational assets that lack physical substance. For example, goodwill is a fixed asset, as are patents, copyrights, trademarks and franchises. A company's intangible assets are often not reported on a company's financial statements, or they may be reported at significantly 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.



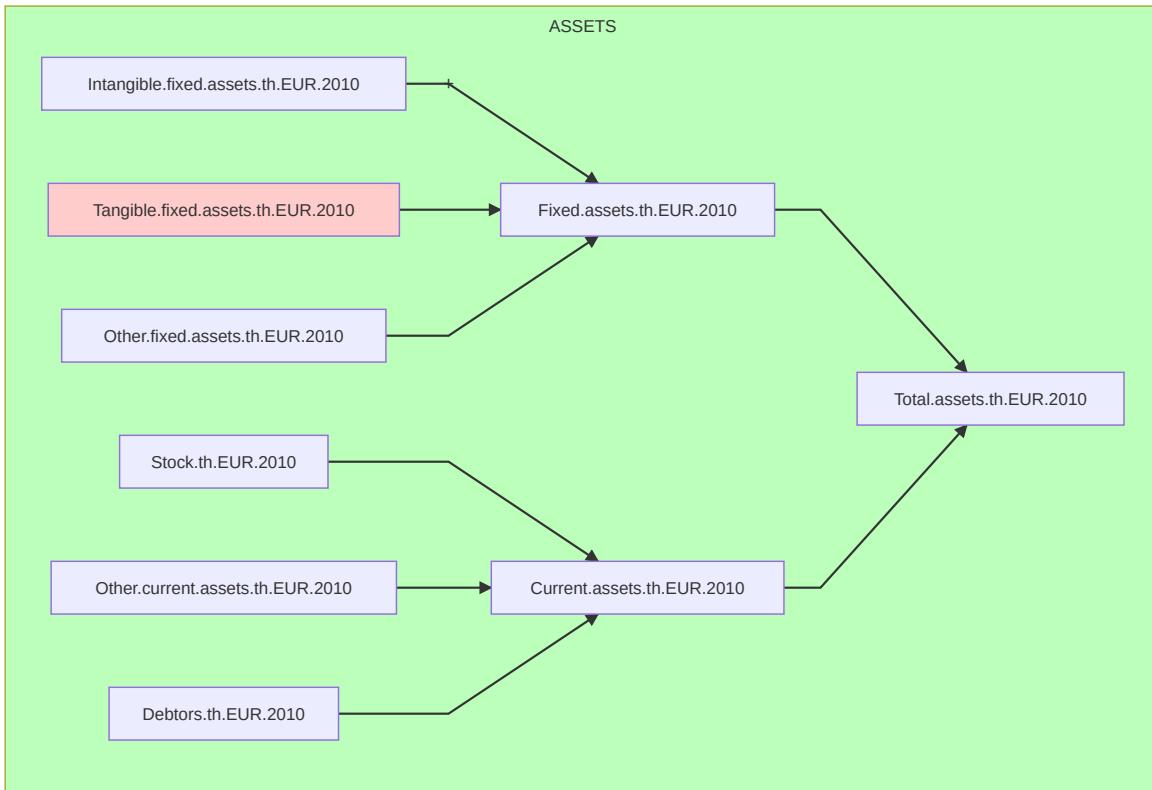
outlier:

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

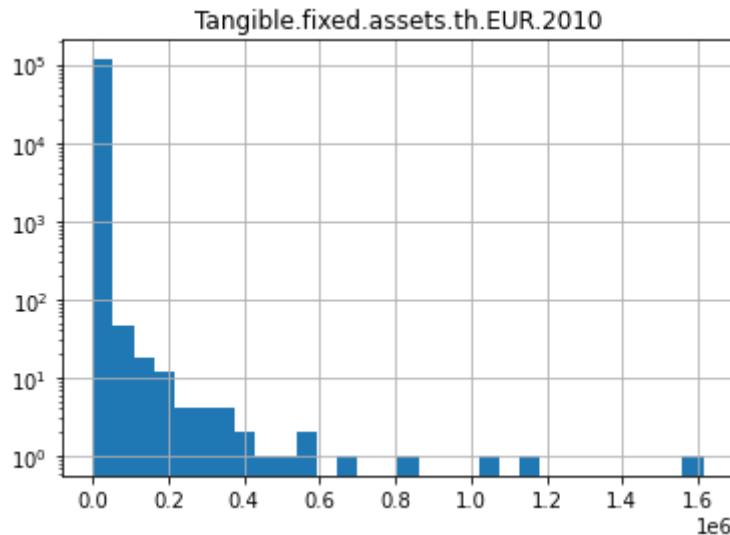
```
HGF vs non-HGF for Intangible.fixed.assets.th.EUR.2010
Welch's t-test statistic = 1.822
p-value = 0.06857
```

```
Optimization terminated successfully.
      Current function value: 0.155657
      Iterations 9
      Logit Regression Results
=====
Dep. Variable:                      HGF      No. Observations:             115840
Model:                            Logit      Df Residuals:                  115838
Method:                           MLE      Df Model:                      1
Date: Mon, 29 Jun 2020      Pseudo R-squ.:        3.442e-05
Time: 15:28:39      Log-Likelihood:     -18031.
converged:                         True      LL-Null:                  -18032.
Covariance Type:            nonrobust      LLR p-value:           0.2652
=====
              coef        std err         z      P>|z|      [0.025      0.975]
-----
Intercept   -3.2814       0.016     -208.487      0.000      -3.312     -3.251
INT        -9.435e-06  1.24e-05      -0.760      0.447     -3.38e-05  1.49e-05
=====
```

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
  
```

```

HGF vs non-HGF for Tangible.fixed.assets.th.EUR.2010
Welch's t-test statistic = 1.394
p-value = 0.1635
  
```

Optimization terminated successfully.

Current function value: 0.155655

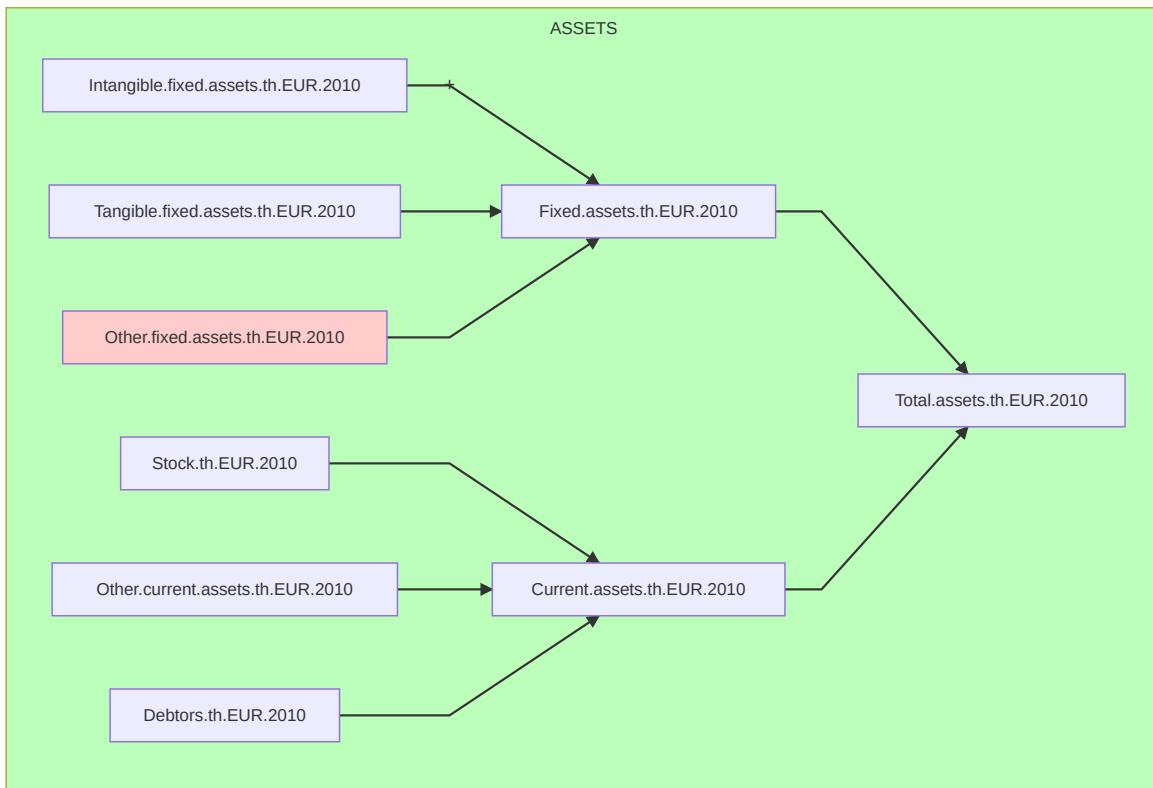
Iterations 8

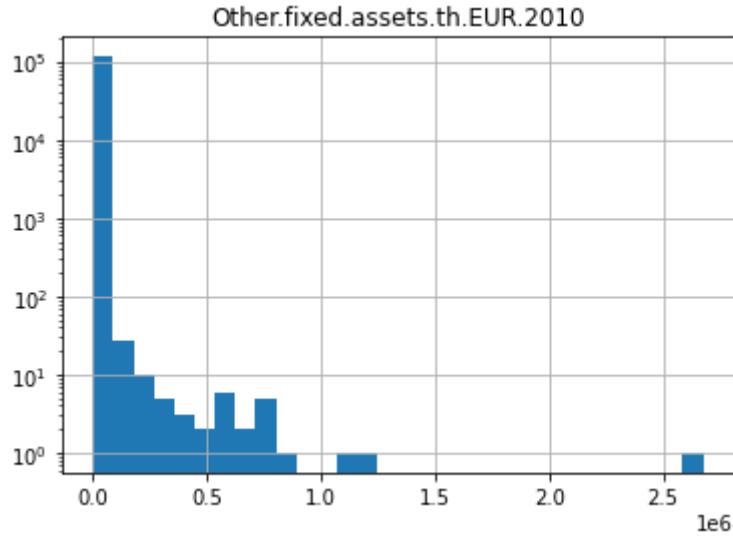
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	4.600e-05
Time:	15:29:04	Log-Likelihood:	-18031.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	0.1977

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2808	0.016	-208.167	0.000	-3.312	-3.250
T	-4.673e-06	4.82e-06	-0.969	0.332	-1.41e-05	4.78e-06

Other.fixed.assets.th.EUR.2010





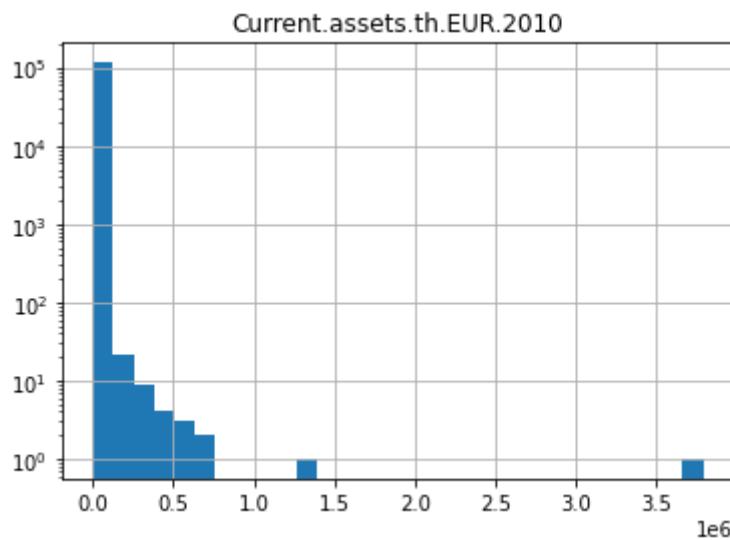
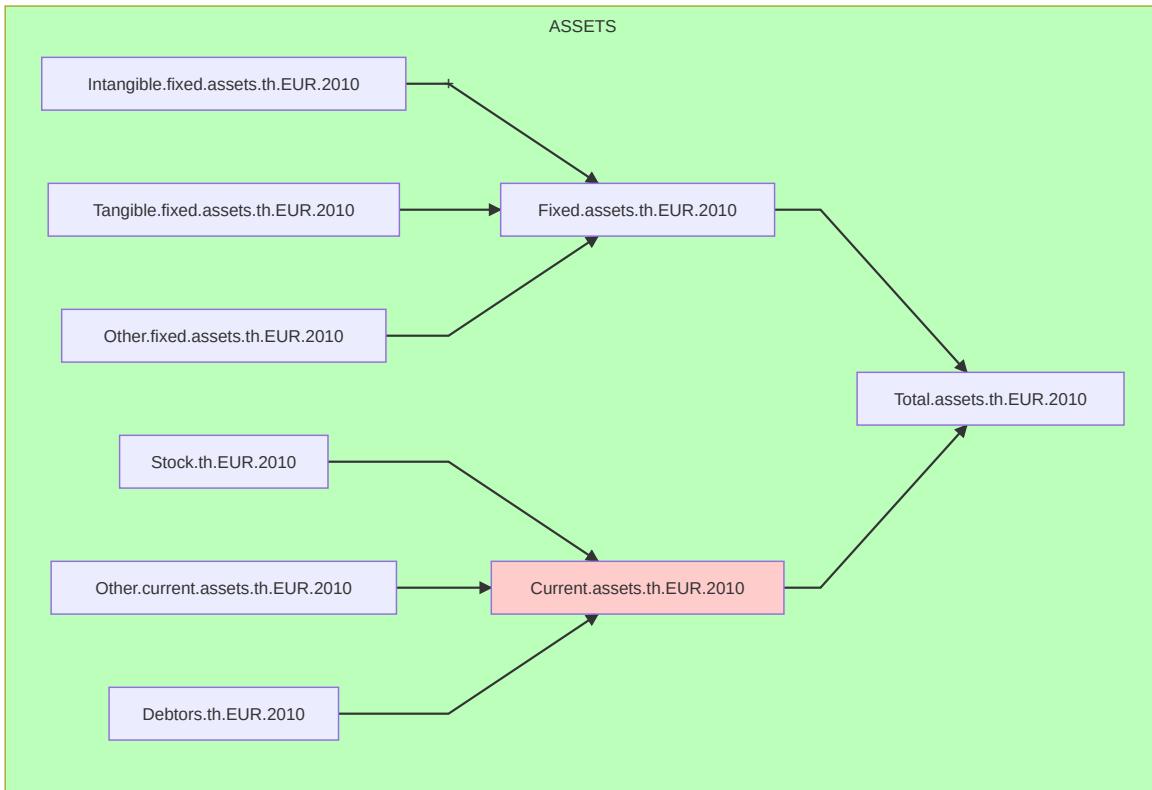
outlier:

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

```
HGF vs non-HGF for Other.fixed.assets.th.EUR.2010
Welch's t-test statistic = 4.793
p-value = 1.65e-06

Optimization terminated successfully.
    Current function value: 0.155642
    Iterations 10
                    Logit Regression Results
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit    Df Residuals:        115838
Method:                MLE     Df Model:             1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:  0.0001271
Time:          15:29:34    Log-Likelihood: -18030.
converged:         True    LL-Null:            -18032.
Covariance Type:  nonrobust   LLR p-value:  0.03230
=====
              coef    std err        z     P>|z|      [0.025      0.975]
-----
Intercept    -3.2804     0.016   -208.259      0.000     -3.311     -3.249
OF       -1.366e-05  1.07e-05    -1.272      0.203    -3.47e-05  7.38e-06
=====
```

Current.assets.th.EUR.2010



```

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

```

HGF vs non-HGF for Current.assets.th.EUR.2010
Welch's t-test statistic = 6.732
p-value = 1.687e-11
  
```

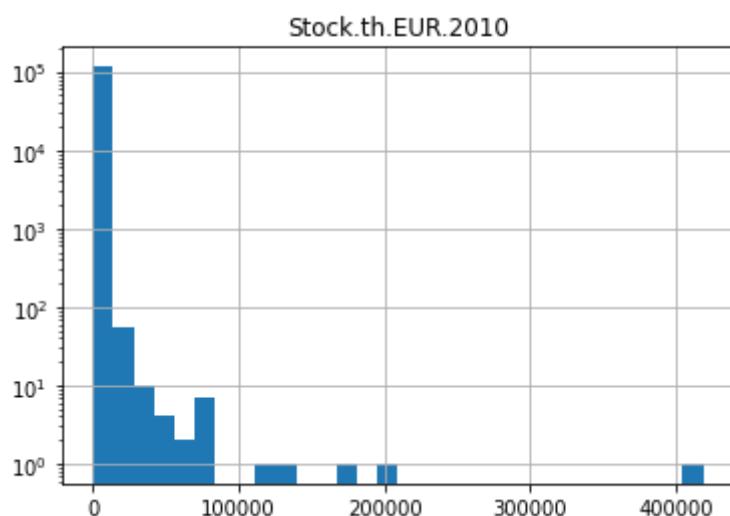
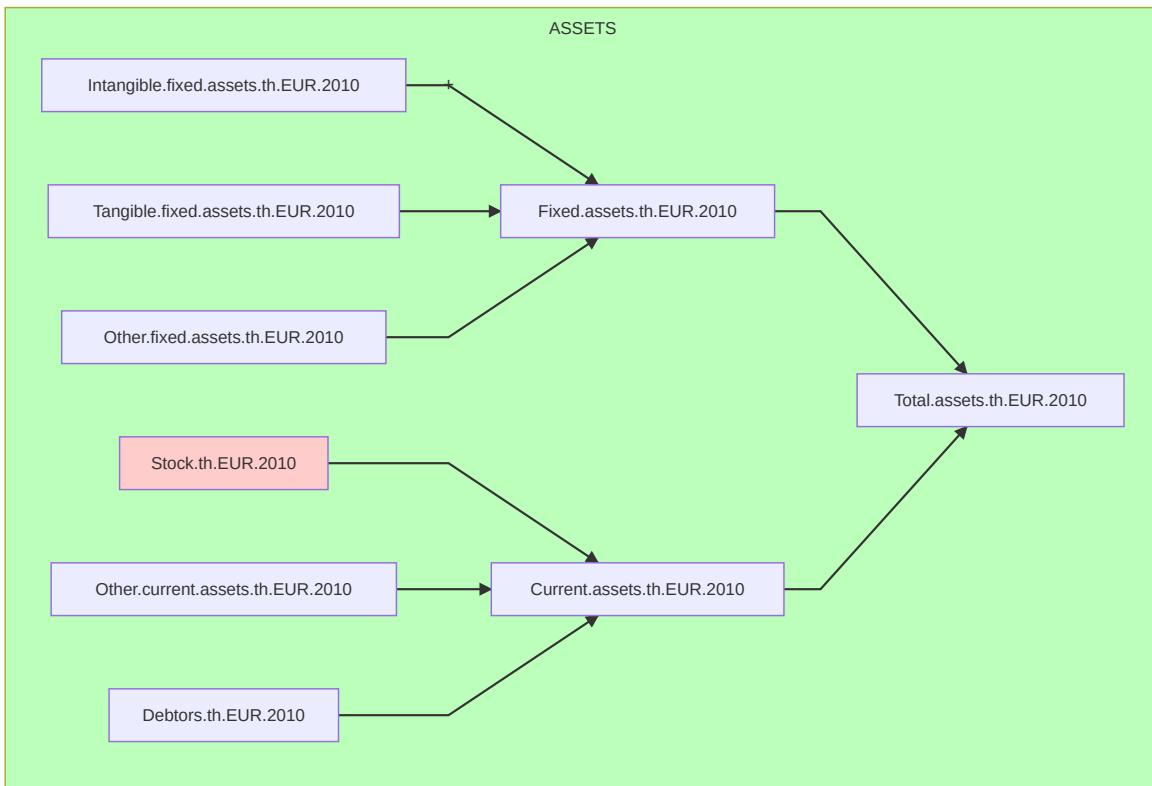
```

Optimization terminated successfully.
    Current function value: 0.155487
    Iterations 11
    Logit Regression Results
=====
  
```

```

Dep. Variable: HGF No. Observations: 115840
Model: Logit Df Residuals: 115838
Method: MLE Df Model: 1
Date: Mon, 29 Jun 2020 Pseudo R-squ.: 0.001126
Time: 15:30:01 Log-Likelihood: -18012.
converged: True LL-Null: -18032.
Covariance Type: nonrobust LLR p-value: 1.856e-10
=====
          coef    std err          z      P>|z|      [0.025      0.975]
-----
Intercept -3.2613     0.016   -201.717      0.000     -3.293     -3.230
CA        -0.0001  2.7e-05    -4.235      0.000     -0.000    -6.15e-05
=====
```

Stock.th.EUR.2010



outlier:

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

HGF vs non-HGF for Stock.th.EUR.2010
Welch's t-test statistic = 10.69
p-value = 1.155e-26

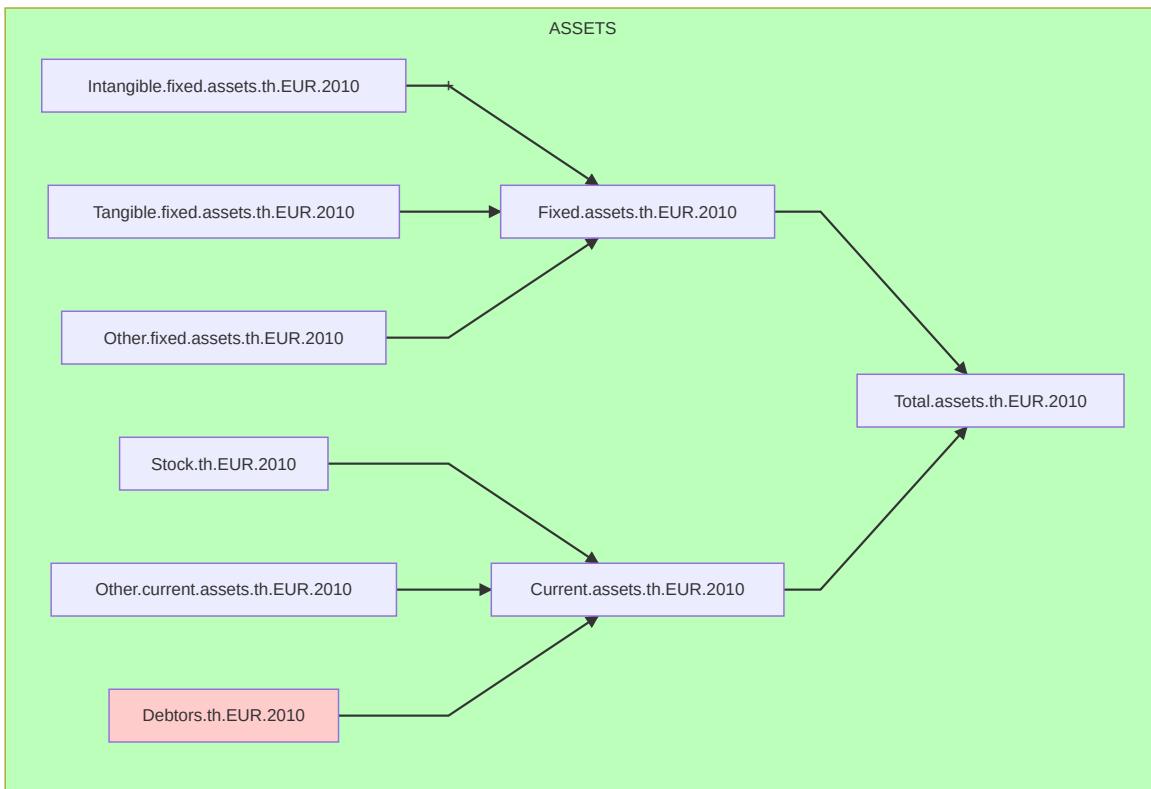
Optimization terminated successfully.
Current function value: 0.155221
Iterations 11

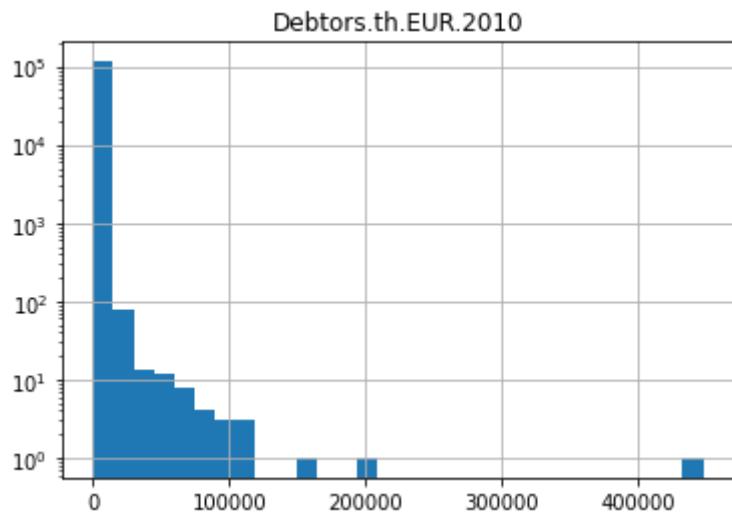
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840			
Model:	Logit	Df Residuals:	115838			
Method:	MLE	Df Model:	1			
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.002831			
Time:	15:30:21	Log-Likelihood:	-17981.			
converged:	True	LL-Null:	-18032.			
Covariance Type:	nonrobust	LLR p-value:	5.302e-24			
=====						
	coef	std err	z	P> z	[0.025	0.975]

Intercept	-3.2474	0.016	-201.604	0.000	-3.279	-3.216
S	-0.0015	0.000	-6.485	0.000	-0.002	-0.001
=====						

Debtors.th.EUR.2010





outlier:

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

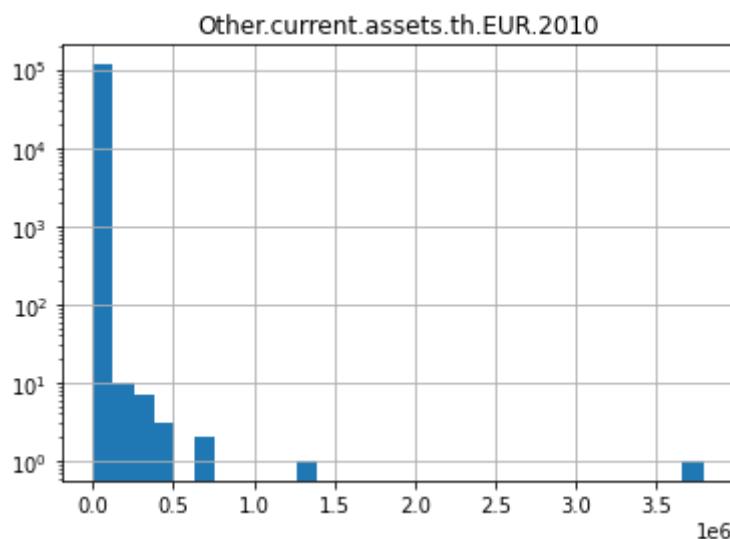
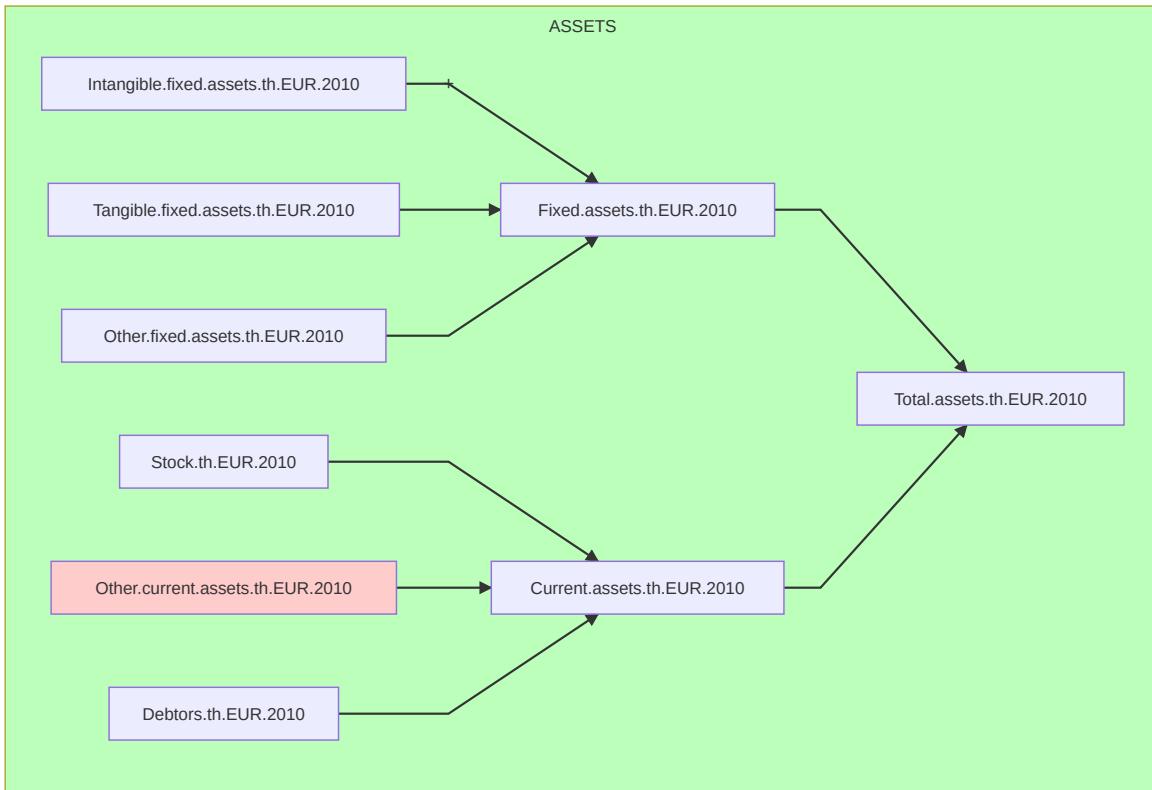
```
HGF vs non-HGF for Debtors.th.EUR.2010
Welch's t-test statistic = 7.135
p-value = 1.014e-12
```

```
Optimization terminated successfully.
    Current function value: 0.155553
    Iterations 10
```

Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.0007008
Time:	15:30:49	Log-Likelihood:	-18019.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	4.972e-07
<hr/>			
	coef	std err	z
<hr/>			
Intercept	-3.2697	0.016	-204.720
DEB	-0.0002	7.3e-05	-3.356
<hr/>			
	[0.025	0.975]	
<hr/>			
	-3.301	-3.238	
	-0.000	-0.000	
<hr/>			

Other.current.assets.th.EUR.2010



outlier:

```

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

```

HGF vs non-HGF for Other.current.assets.th.EUR.2010
Welch's t-test statistic = 4.135
p-value = 3.56e-05
  
```

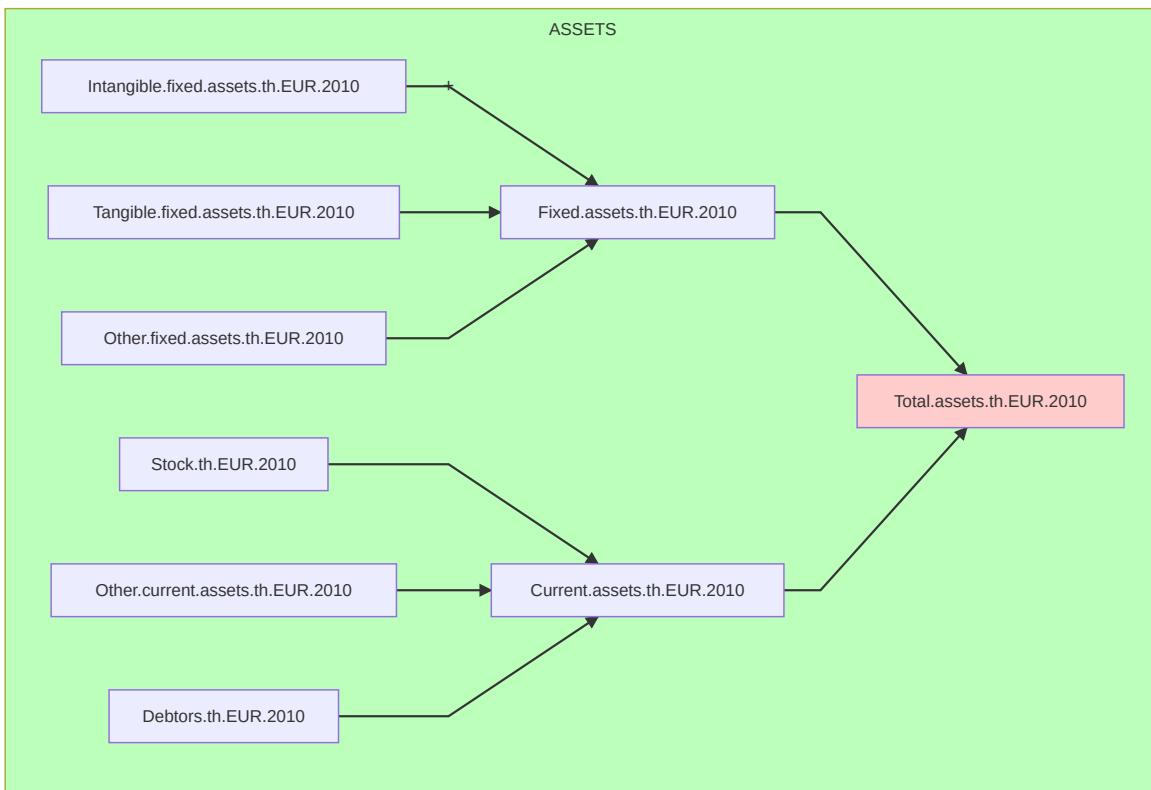
```

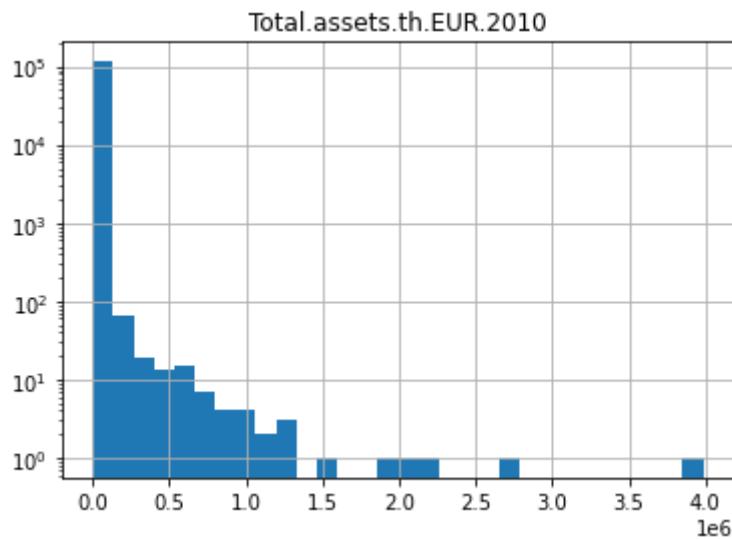
Optimization terminated successfully.
    Current function value: 0.155609
    Iterations 10
    Logit Regression Results
=====
```

Dep. Variable:	HGF	No. Observations:	115840		
Model:	Logit	Df Residuals:	115838		
Method:	MLE	Df Model:	1		
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.0003421		
Time:	15:31:18	Log-Likelihood:	-18026.		
converged:	True	LL-Null:	-18032.		
Covariance Type:	nonrobust	LLR p-value:	0.0004442		
=====					
	coef	std err	z		
			P> z	[0.025	0.975]

Intercept	-3.2750	0.016	-206.020	0.000	-3.306
OCA	-6.435e-05	2.72e-05	-2.366	0.018	-0.000
=====					

Total.assets.th.EUR.2010





outlier:

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

```
HGF vs non-HGF for Total.assets.th.EUR.2010
Welch's t-test statistic = 5.512
p-value = 3.645e-08
```

```
Optimization terminated successfully.
      Current function value: 0.155604
      Iterations 10
```

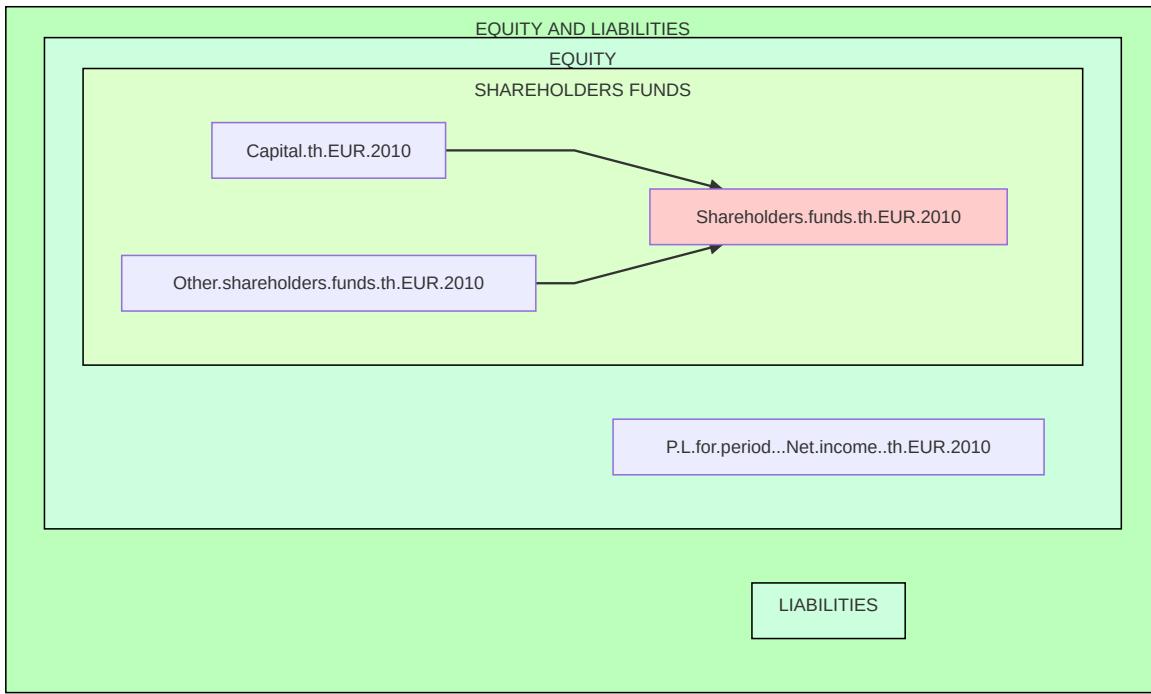
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840			
Model:	Logit	Df Residuals:	115838			
Method:	MLE	Df Model:	1			
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.0003744			
Time:	15:31:50	Log-Likelihood:	-18025.			
converged:	True	LL-Null:	-18032.			
Covariance Type:	nonrobust	LLR p-value:	0.0002385			
=====						
	coef	std err	z	P> z	[0.025	0.975]

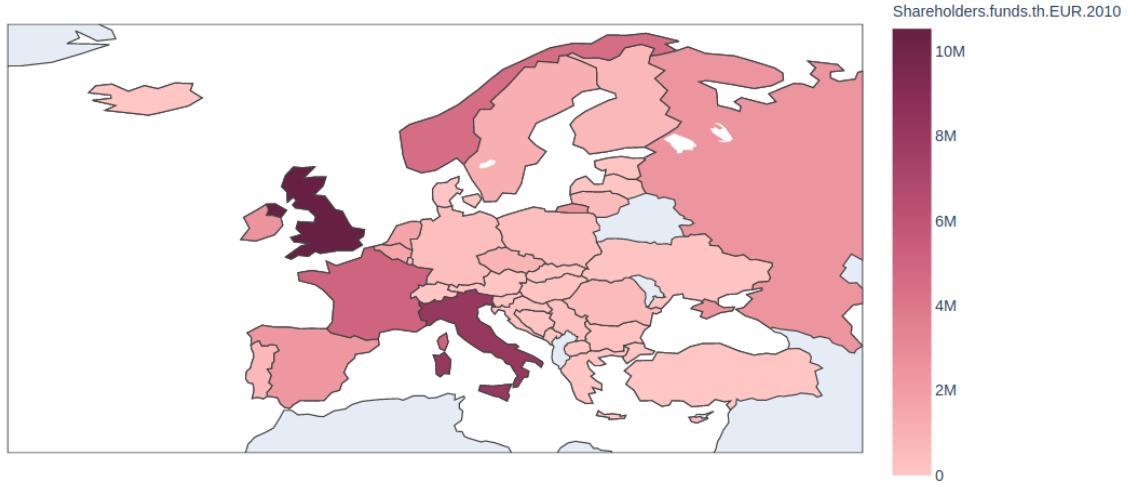
Intercept	-3.2751	0.016	-206.277	0.000	-3.306	-3.244
TA	-1.137e-05	4.72e-06	-2.406	0.016	-2.06e-05	-2.11e-06
=====						

Equity

Shareholders.funds.th.EUR.2010



Shareholders.funds.th.EUR.2010



outlier:

```
BvD.ID.number  
GB07450219 LONG ISLAND ASSETS LIMITED  
Name: Company.name, dtype: string  
  
HGF vs non-HGF for Shareholders.funds.th.EUR.2010  
Welch's t-test statistic = 1.909  
p-value = 0.05634
```

Optimization terminated successfully.

Current function value: 0.155657

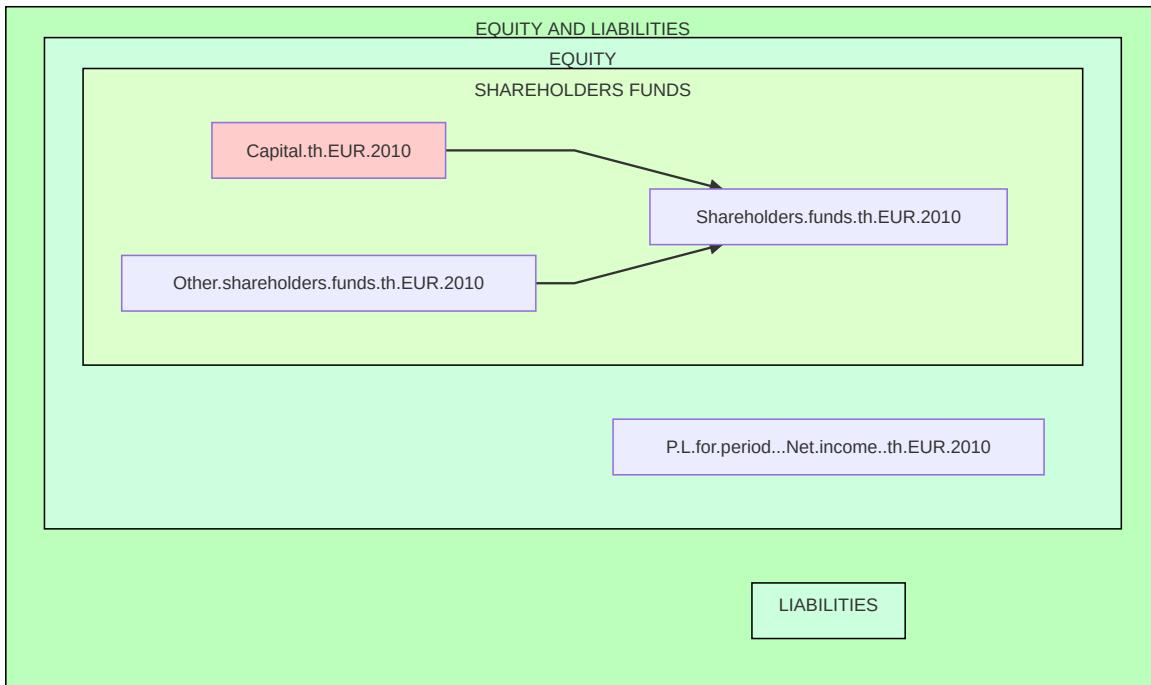
Iterations 8

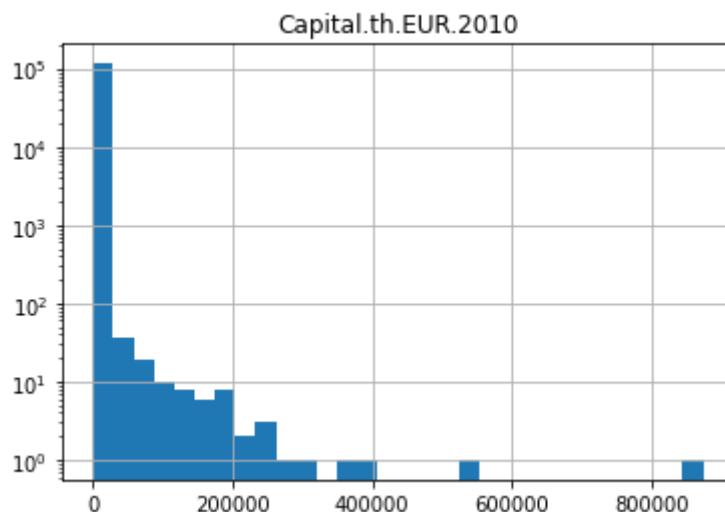
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	3.438e-05
Time:	15:32:19	Log-Likelihood:	-18031.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	0.2655

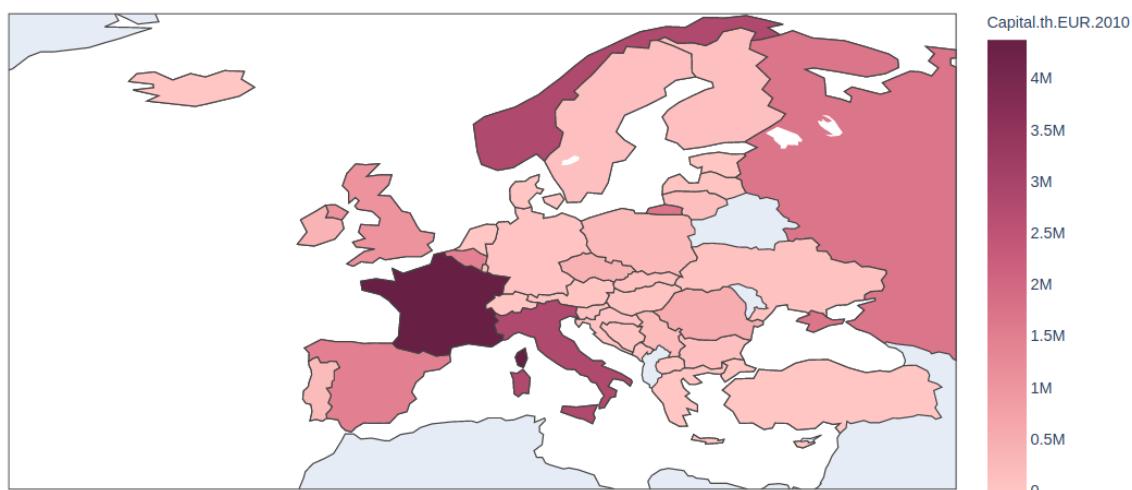
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2812	0.016	-208.351	0.000	-3.312	-3.250
SHA	-2.844e-06	3.41e-06	-0.834	0.404	-9.53e-06	3.84e-06

Capital.th.EUR.2010





Total capital per country



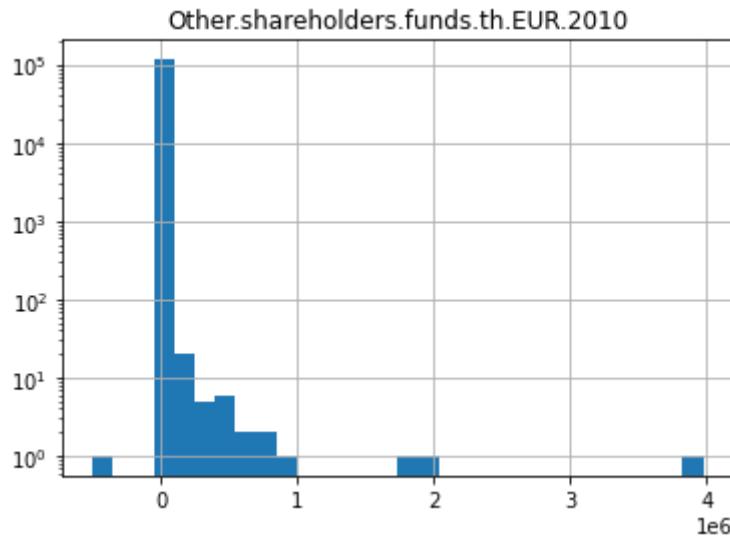
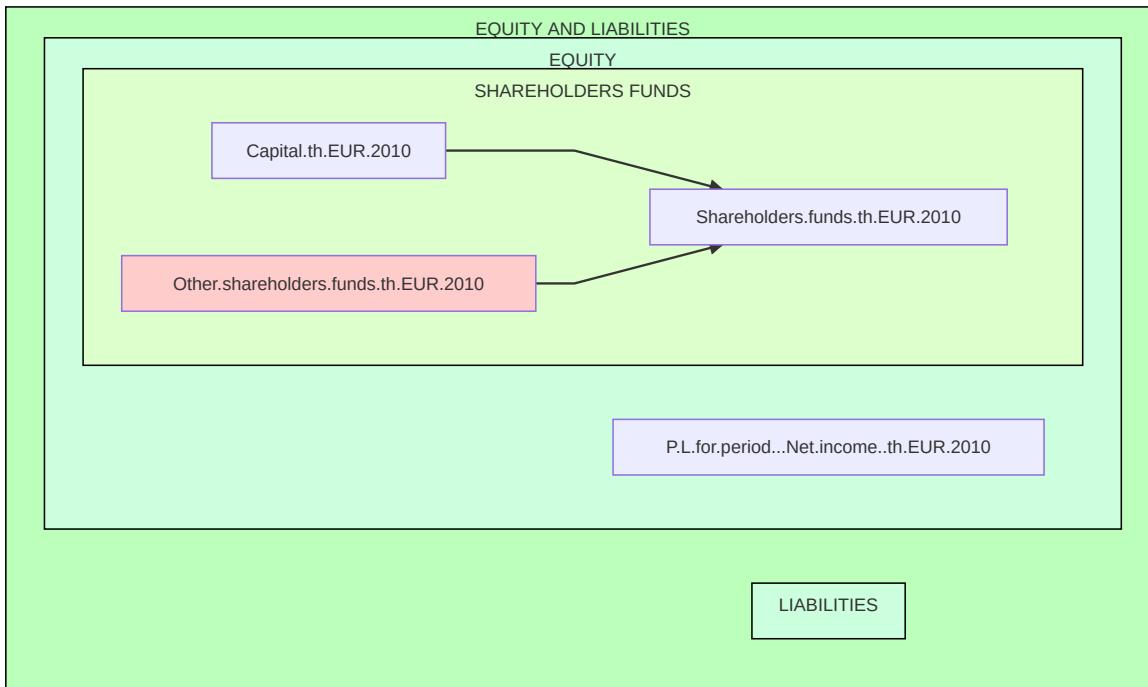
outlier:

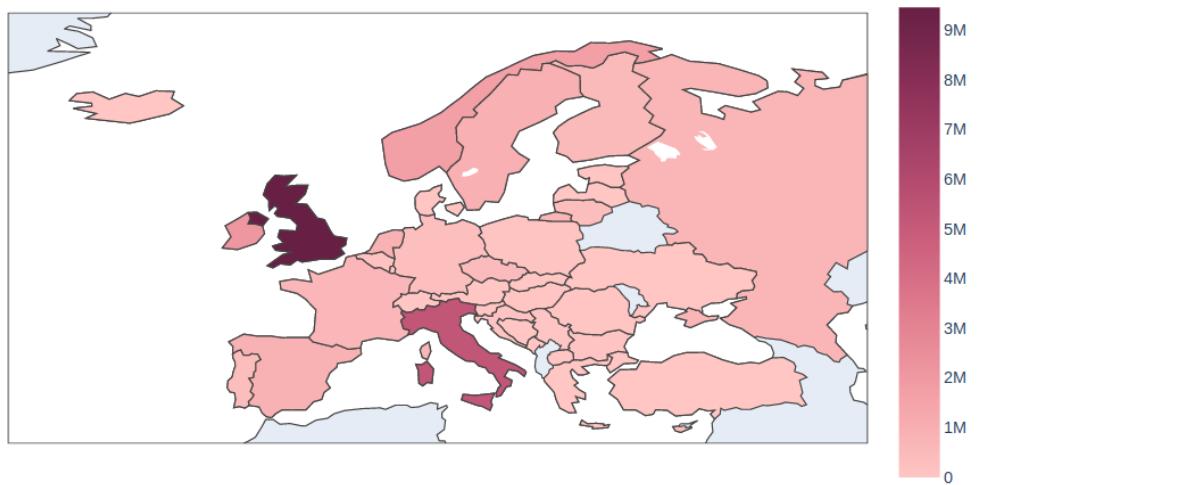
BvD.ID.number
FR519720643 IRIDIUM FRANCE
Name: Company.name, dtype: string

HGF vs non-HGF for Capital.th.EUR.2010
Welch's t-test statistic = 4.127
p-value = 3.699e-05

Intercept	-3.2801	0.016	-208.097	0.000	-3.311	-3.249
CAP	-1.822e-05	1.36e-05	-1.337	0.181	-4.49e-05	8.49e-06
<hr/>						

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
```

HGF vs non-HGF for Other.shareholders.funds.th.EUR.2010
Welch's t-test statistic = 0.8908
p-value = 0.3731

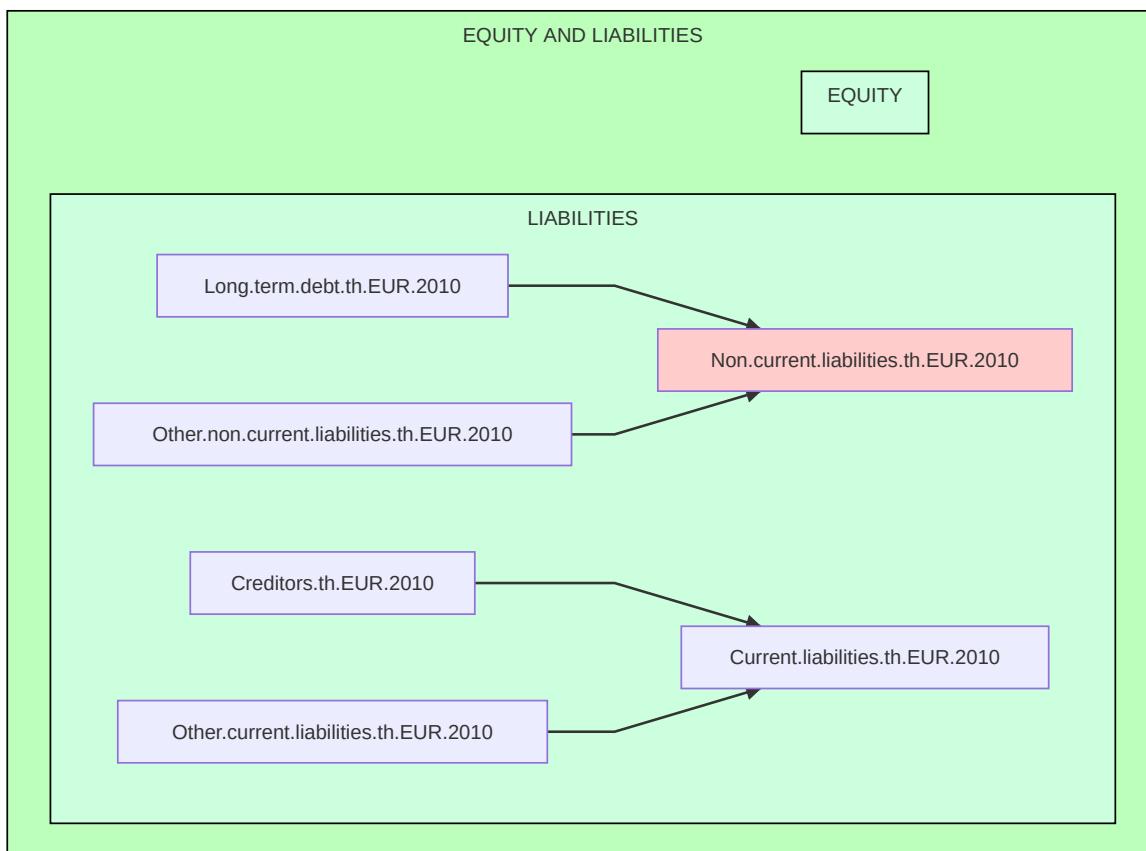
```

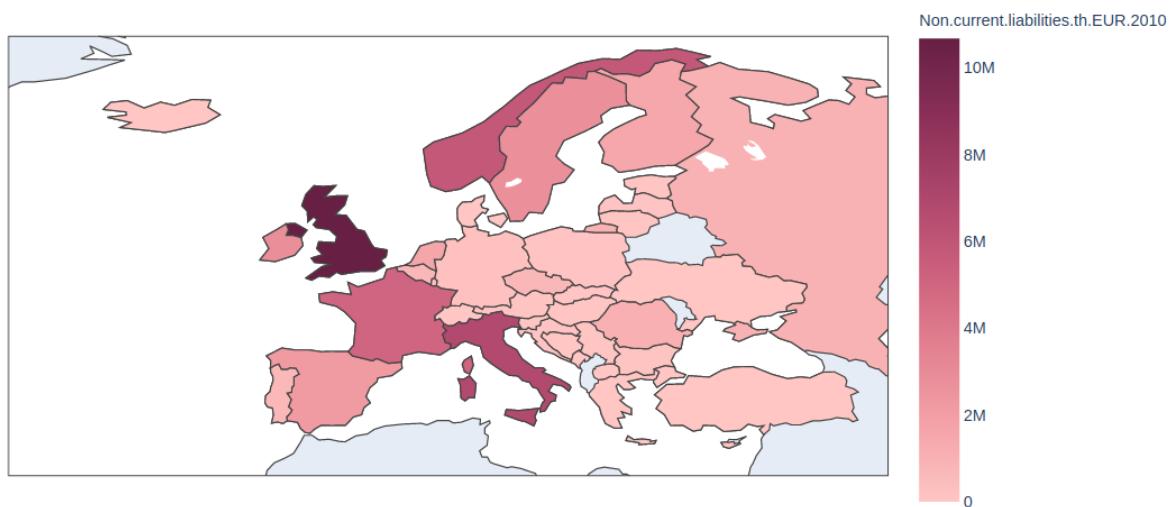
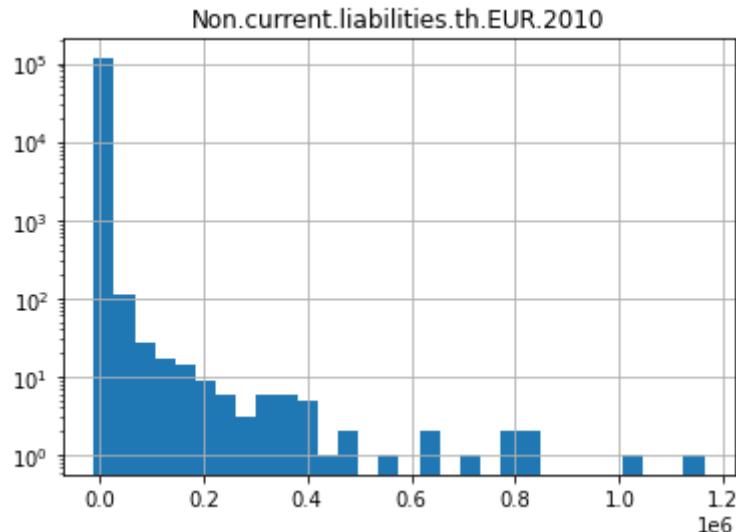
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:        115838
Method:                MLE     Df Model:             1
Date: Mon, 29 Jun 2020   Pseudo R-squ.:       4.778e-06
Time: 15:33:13           Log-Likelihood:      -18032.
converged:              True    LL-Null:            -18032.
Covariance Type:        nonrobust LLR p-value:     0.6781
=====
            coef      std err      z      P>|z|      [0.025      0.975]
-----
Intercept    -3.2818      0.016   -208.591      0.000      -3.313      -3.251
OSF         -6.672e-07  2e-06    -0.334      0.739      -4.59e-06   3.25e-06
-----

```

Liabilities

Non.current.liabilities.th.EUR.2010





HGF vs non-HGF for Non.current.liabilities.th.EUR.2010

Welch's t-test statistic = 7.868

p-value = 3.731e-15

Optimization terminated successfully.

Current function value: 0.155577

Iterations 10

Logit Regression Results

Dep. Variable: HGE No. Observations: 115840

Model: Logit Df Residuals: 115838

Method: MLE Pf Model:

Date: Mon, 29 Jun 2020 Pseudo R-sq.: 0.0005448

Time: 15:34:00 Log-Likelihood: -18022.

converged: True LL-Null: -18032.

Covariance Type: nonrobust LLR p-value: 9.307e-06

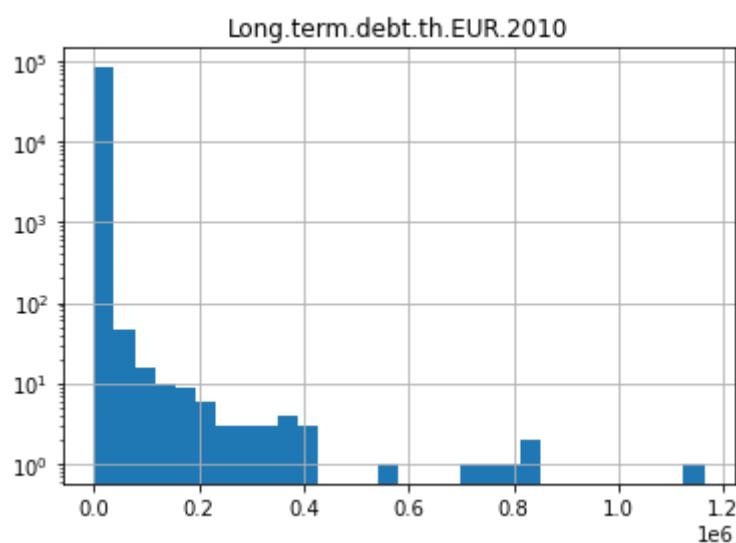
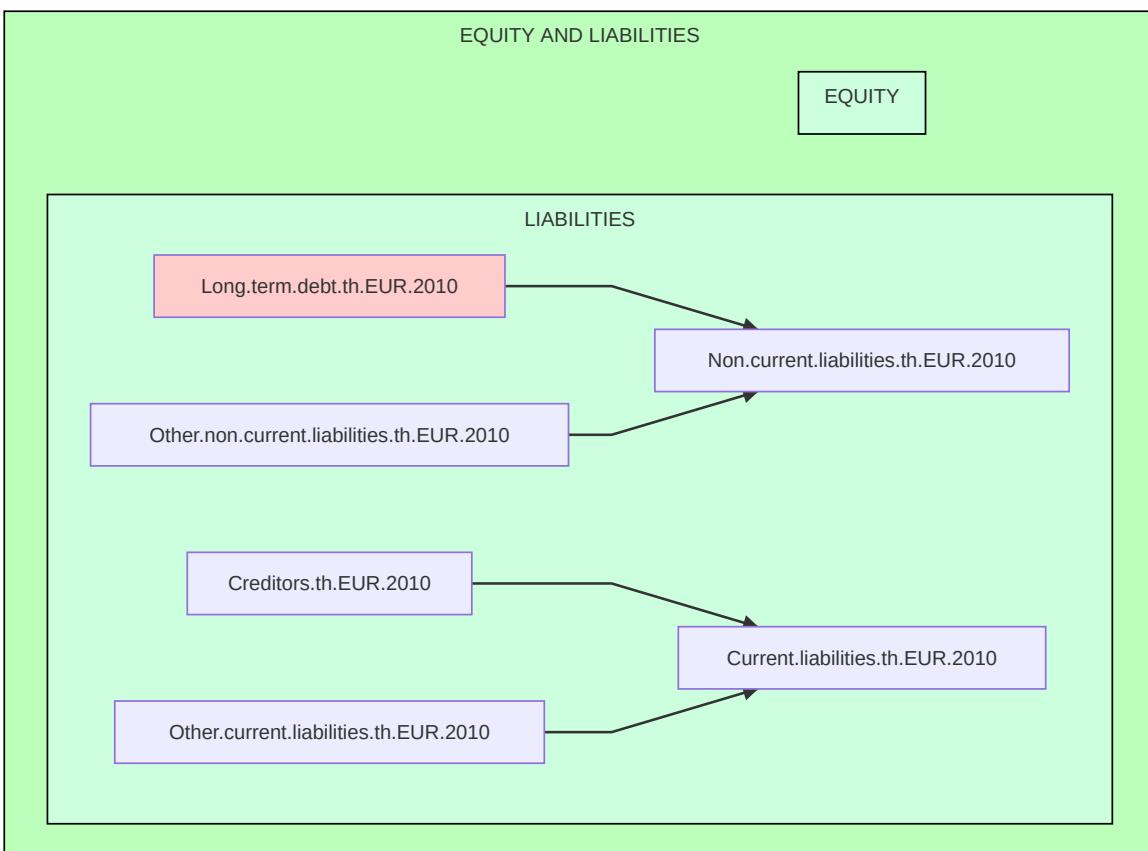
=====
=====

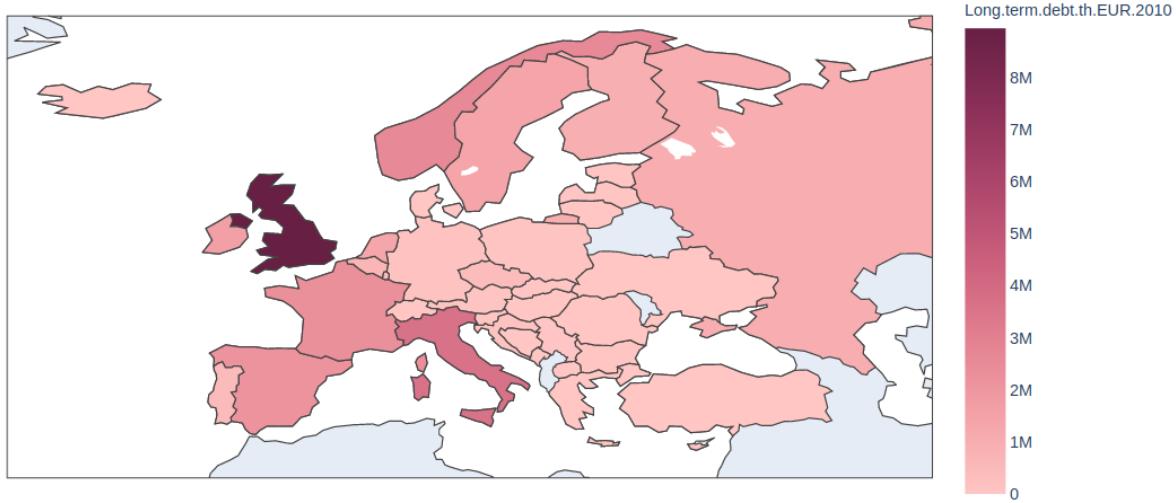
coef std_err z P>|z| [0.025 0.975]

Intercept -3.2743 0.016 -206.612 0.000 -3.305 -3.243

NCL -5.339e-05 1.92e-05 -2.774 0.006 -9.11e-05 -1.57e-05

Long.term.debt.th.EUR.2010



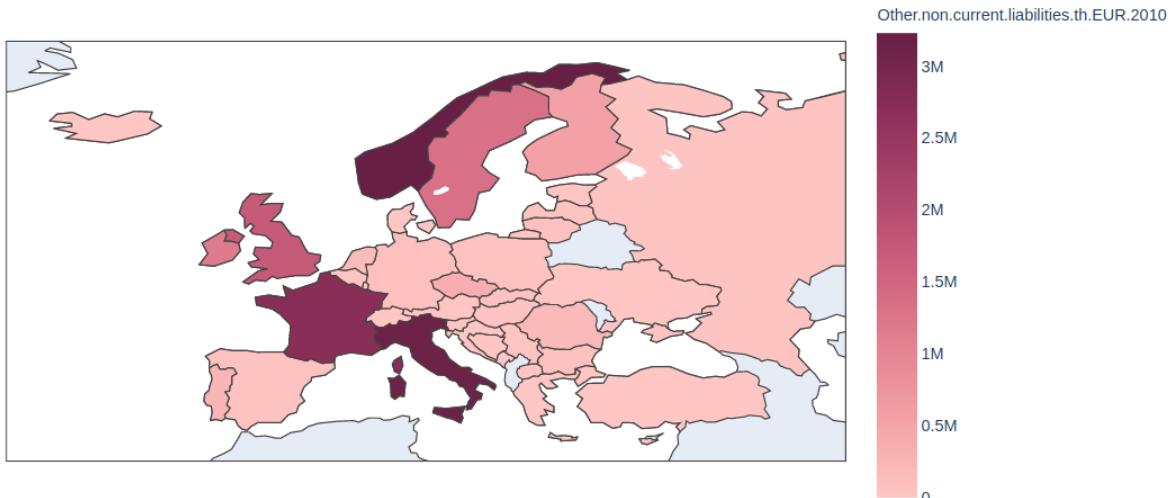
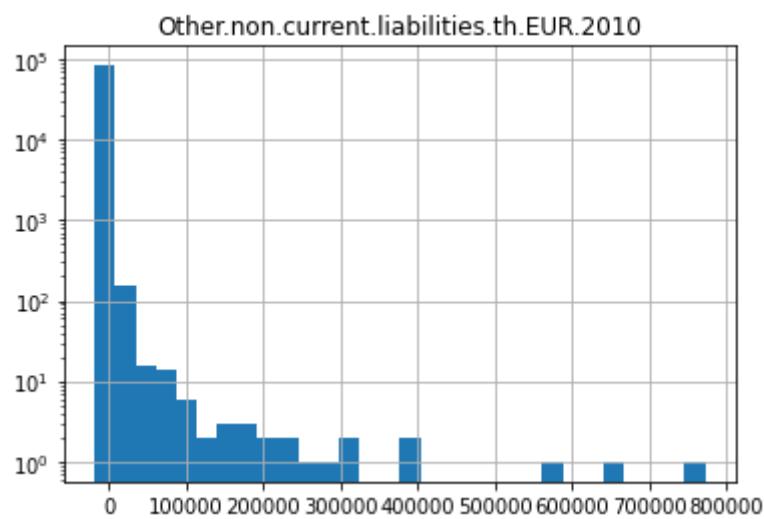
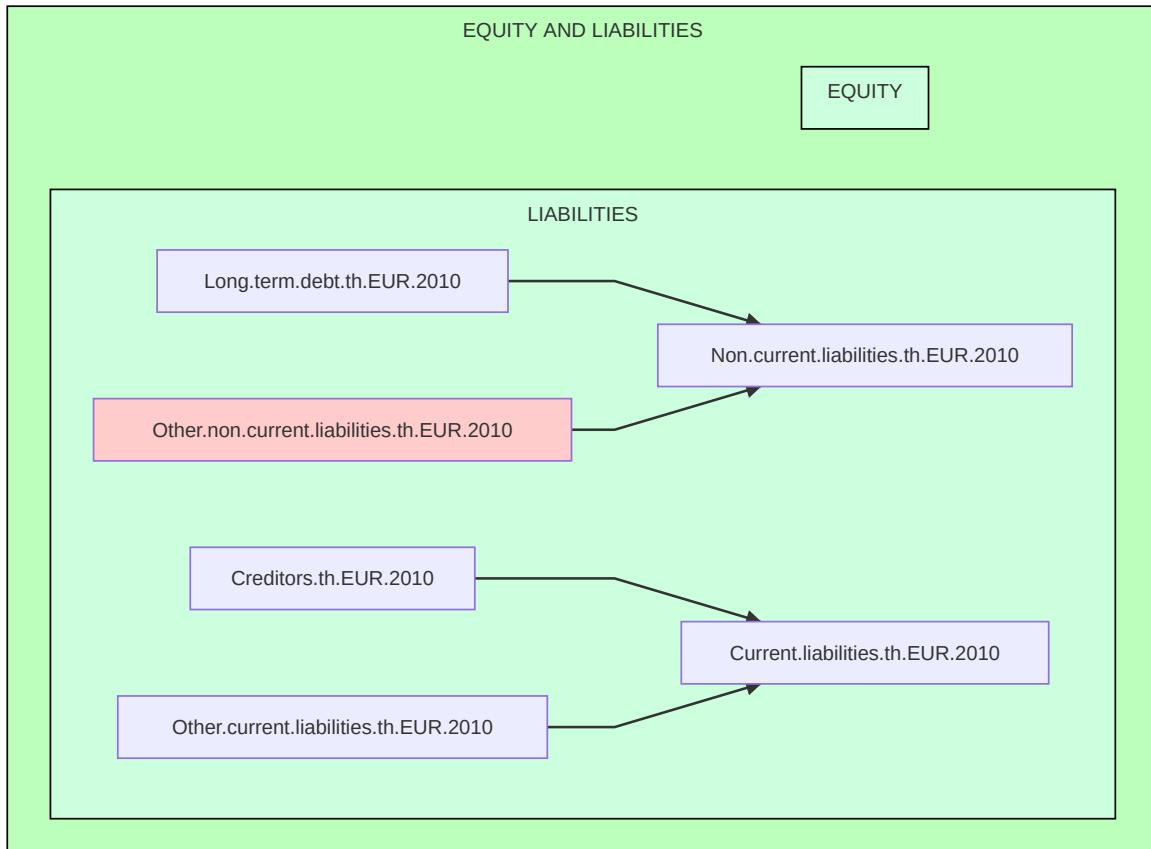


outlier:

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

```
Optimization terminated successfully.
    Current function value: 0.143145
    Iterations 10
                    Logit Regression Results
=====
Dep. Variable:          HGF    No. Observations:      83074
Model:                 Logit    Df Residuals:        83072
Method:                MLE     Df Model:             1
Date: Mon, 29 Jun 2020   Pseudo R-squ.: 0.0004459
Time: 15:34:25           Log-Likelihood: -11892.
converged:            True    LL-Null:        -11897.
Covariance Type:    nonrobust    LLR p-value: 0.001125
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept    -3.3885     0.020    -172.120      0.000     -3.427     -3.350
LTD       -4.824e-05  2.41e-05     -1.998      0.046    -9.56e-05   -9.12e-07
=====
```

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

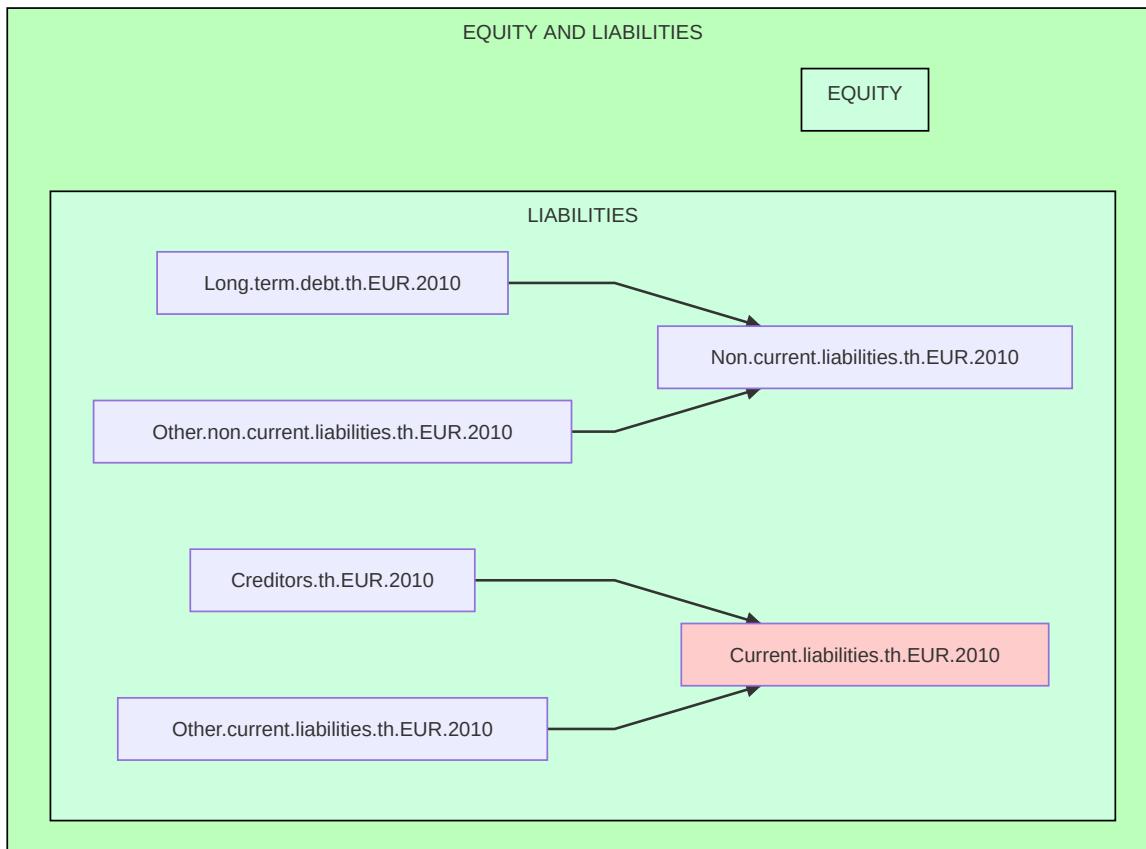


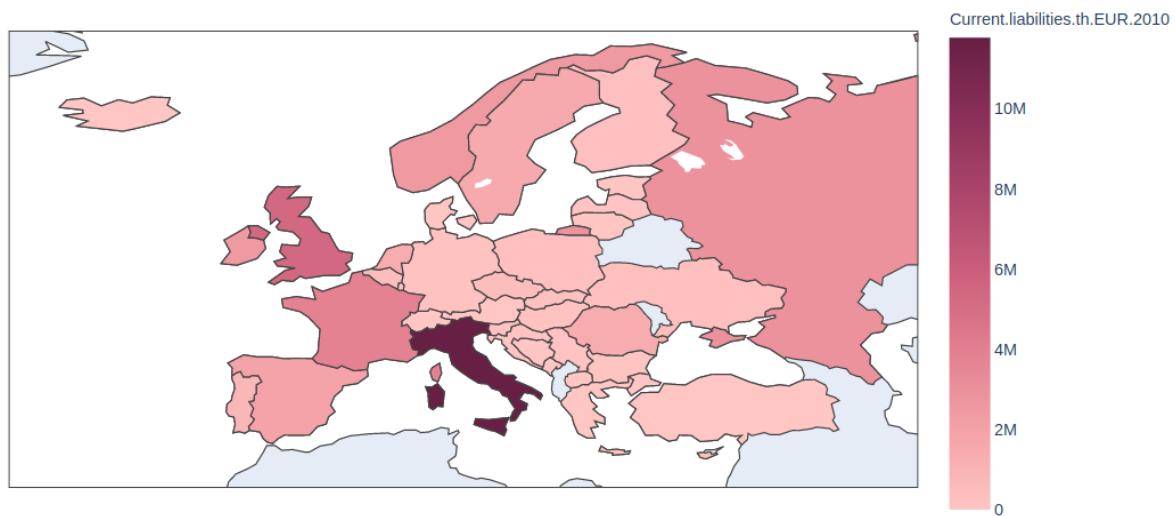
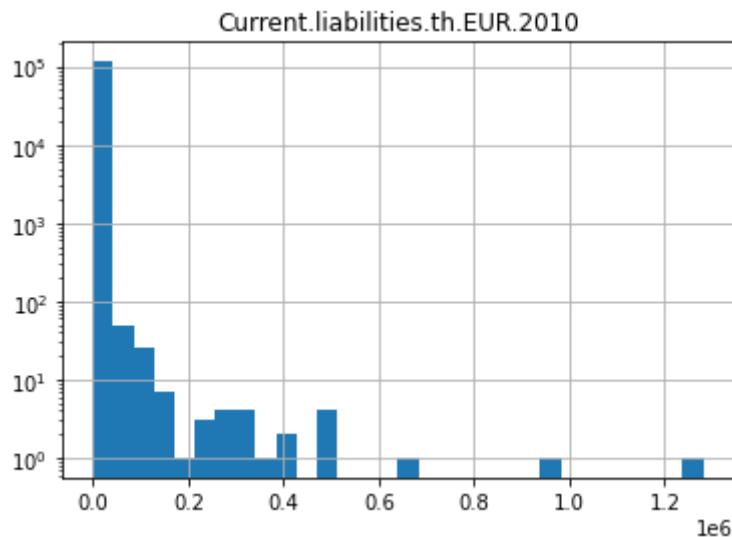
outliers:

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

Optimization terminated successfully.
  Current function value: 0.143132
  Iterations 10
      Logit Regression Results
=====
Dep. Variable:          HGF   No. Observations:      83074
Model:                 Logit   Df Residuals:        83072
Method:                MLE    Df Model:             1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:     0.0005412
Time:          15:34:51    Log-Likelihood:   -11891.
converged:            True    LL-Null:           -11897.
Covariance Type:    nonrobust   LLR p-value:    0.0003325
=====
              coef      std err       z   P>|z|      [0.025      0.975]
-----
Intercept    -3.3869      0.020    -171.972      0.000     -3.426     -3.348
ONCL        -0.0001  4.63e-05     -2.467      0.014     -0.000  -2.35e-05
=====
```

Current.liabilities.th.EUR.2010





outliers:

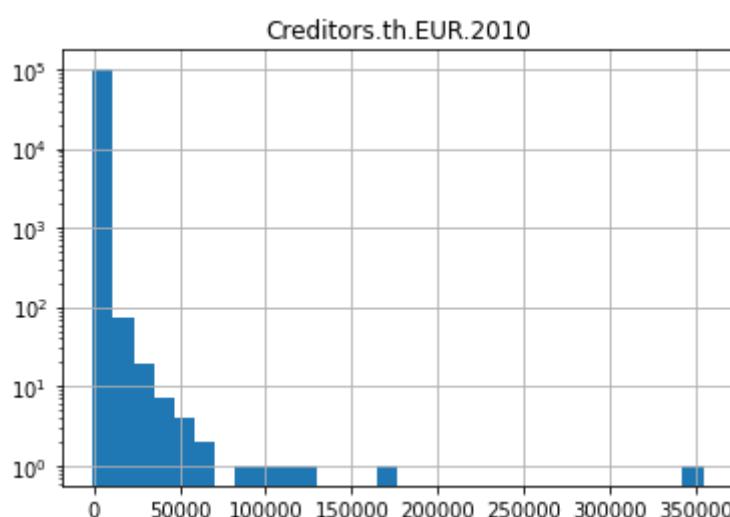
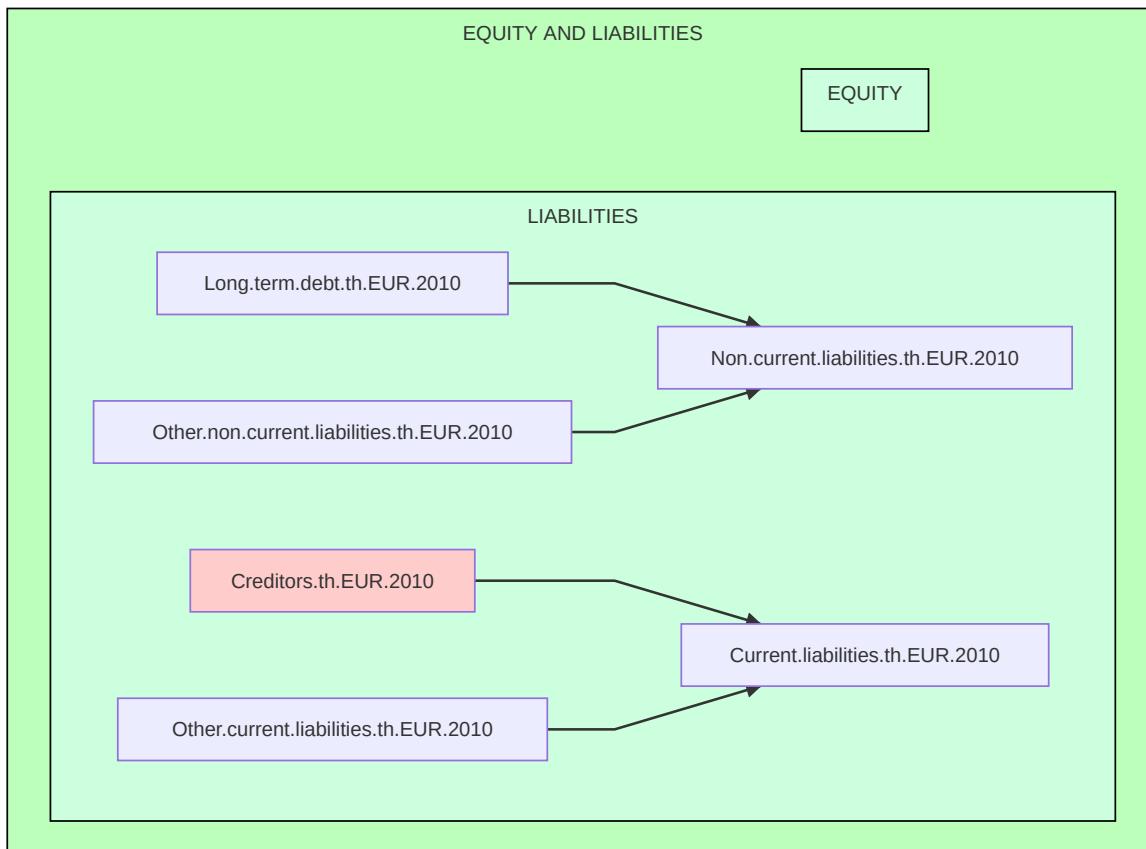
```
BvD.ID.number
IE480184      ESB FINANCE DESIGNATED ACTIVITY COMPANY
IE486122      ICG EOS LOAN FUND I LIMITED
GB07193500    PREMIER LOTTERIES CAPITAL UK LIMITED
GB07202475    PREMIER LOTTERIES INVESTMENTS UK LIMITED
IT10319310016 INFRASTRASPORTI.TO S.R.L.
Name: Company.name, dtype: string
```

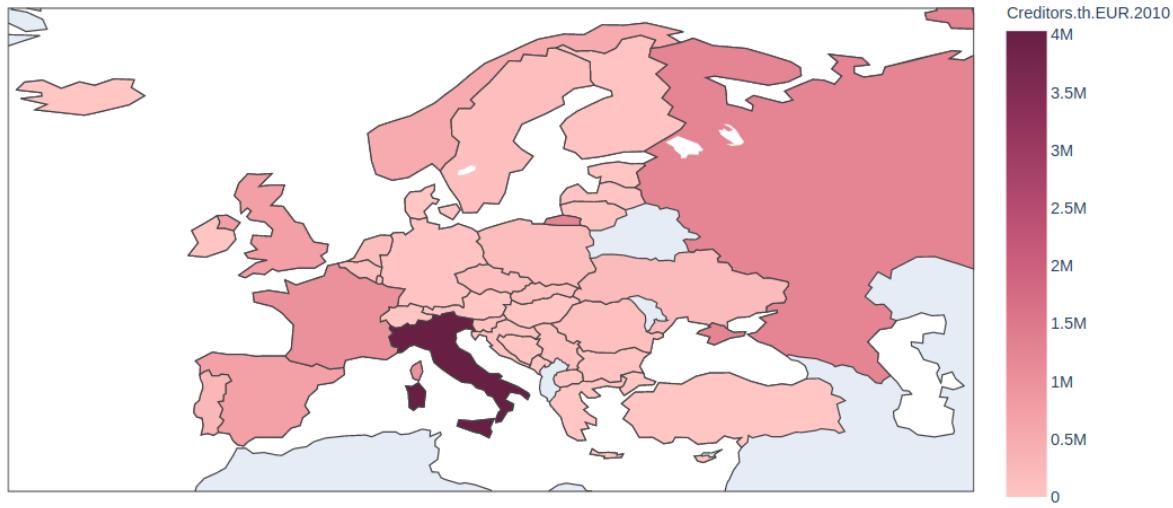
```
HGF vs non-HGF for Current.liabilities.th.EUR.2010
Welch's t-test statistic = 7.677
p-value = 1.691e-14
```

```
Optimization terminated successfully.
      Current function value: 0.155527
      Iterations 10
      Logit Regression Results
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:        115838
Method:                MLE    Df Model:             1
Date: Mon, 29 Jun 2020   Pseudo R-squ.:     0.0008644
Time: 15:35:15           Log-Likelihood:   -18016.
converged:            True    LL-Null:            -18032.
Covariance Type:    nonrobust  LLR p-value:  2.359e-08
```

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2648	0.016	-202.670	0.000	-3.296	-3.233
CL	-9.74e-05	2.52e-05	-3.868	0.000	-0.000	-4.8e-05

Creditors.th.EUR.2010





outliers:

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

Optimization terminated successfully.
    Current function value: 0.149458
    Iterations 10
        Logit Regression Results
=====
Dep. Variable:          HGF    No. Observations:      98541
Model:                 Logit   Df Residuals:         98539
Method:                MLE    Df Model:             1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:     0.001558
Time:          15:35:36      Log-Likelihood:   -14728.
converged:            True    LL-Null:           -14751.
Covariance Type:    nonrobust  LLR p-value:  1.211e-11
=====
              coef    std err        z     P>|z|      [0.025    0.975]
-----
Intercept    -3.3077    0.018   -183.576      0.000    -3.343    -3.272
CR          -0.0005    0.000     -4.689      0.000    -0.001    -0.000
=====
```

Other.current.liabilities.th.EUR.2010

EQUITY AND LIABILITIES

EQUITY

LIABILITIES

Long.term.debt.th.EUR.2010

Non.current.liabilities.th.EUR.2010

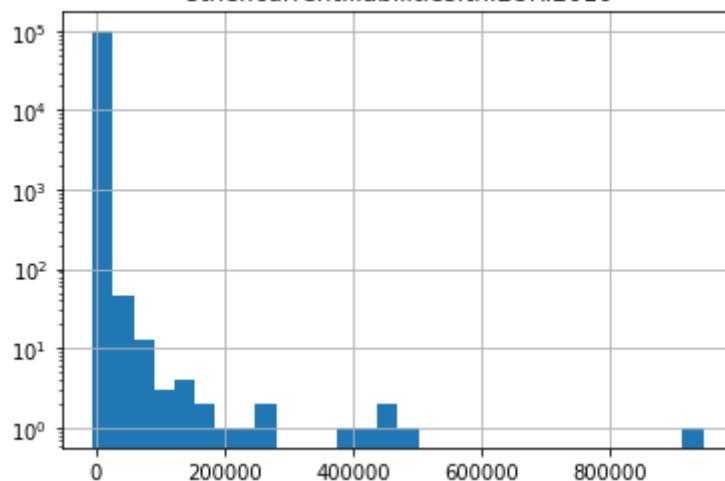
Other.non.current.liabilities.th.EUR.2010

Creditors.th.EUR.2010

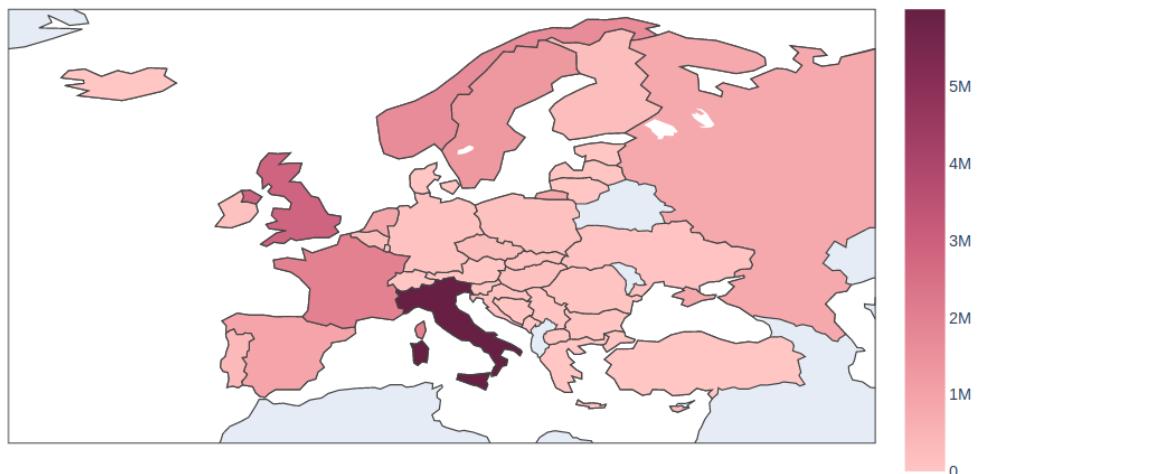
Current.liabilities.th.EUR.2010

Other.current.liabilities.th.EUR.2010

Other.current.liabilities.th.EUR.2010



Other.current.liabilities.th.EUR.2010

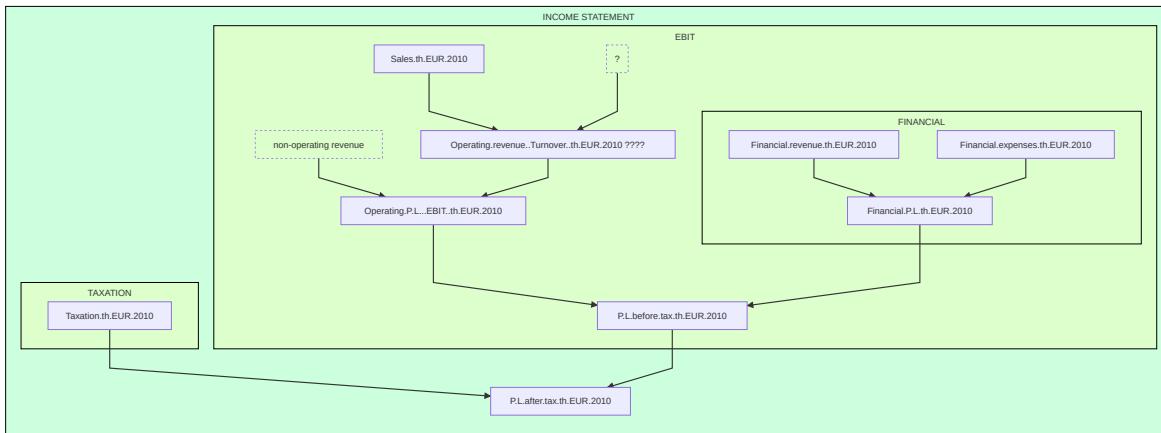


outlier:

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

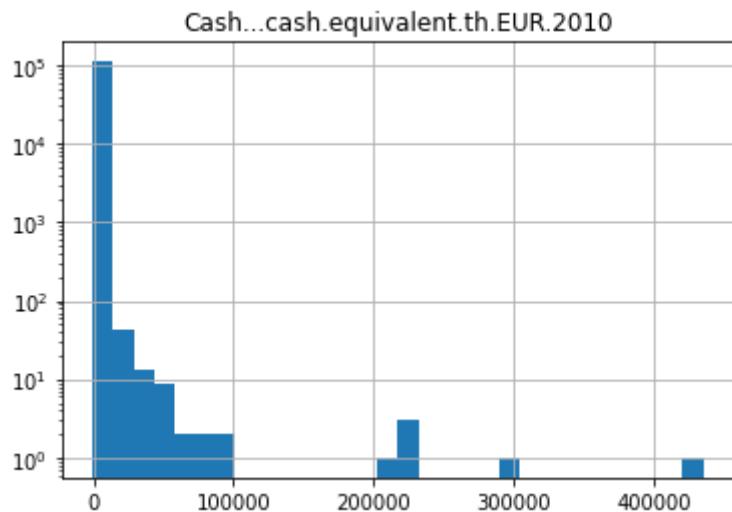
Optimization terminated successfully.
    Current function value: 0.148949
    Iterations 10
        Logit Regression Results
=====
Dep. Variable: HGF   No. Observations: 97463
Model: Logit   Df Residuals: 97461
Method: MLE    Df Model: 1
Date: Mon, 29 Jun 2020 Pseudo R-squ.: 0.0006357
Time: 15:36:15 Log-Likelihood: -14517.
converged: True   LL-Null: -14526.
Covariance Type: nonrobust LLR p-value: 1.727e-05
=====
            coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept  -3.3277    0.018  -185.296      0.000     -3.363     -3.292
OCL        -0.0001  4.46e-05     -3.012      0.003     -0.000    -4.69e-05
=====
```

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
```

Optimization terminated successfully.

Current function value: 0.157093

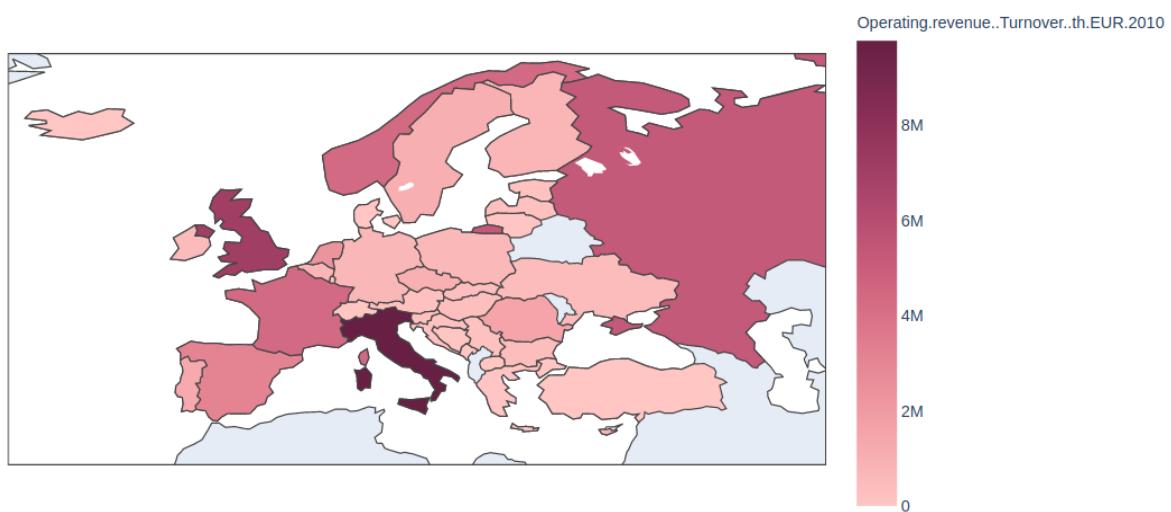
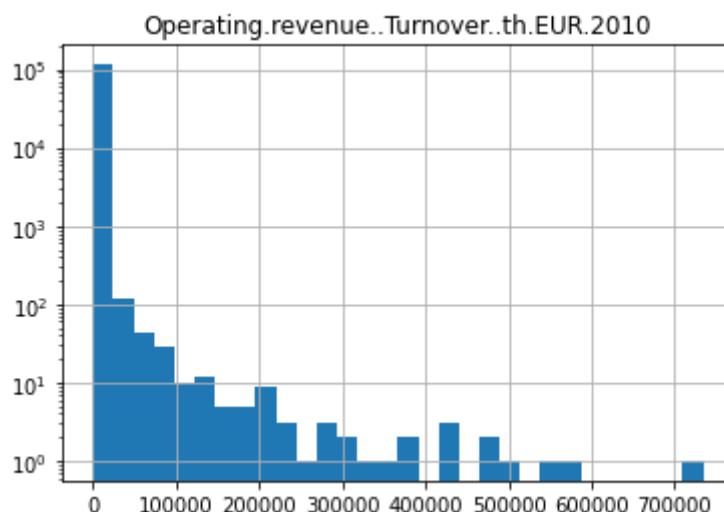
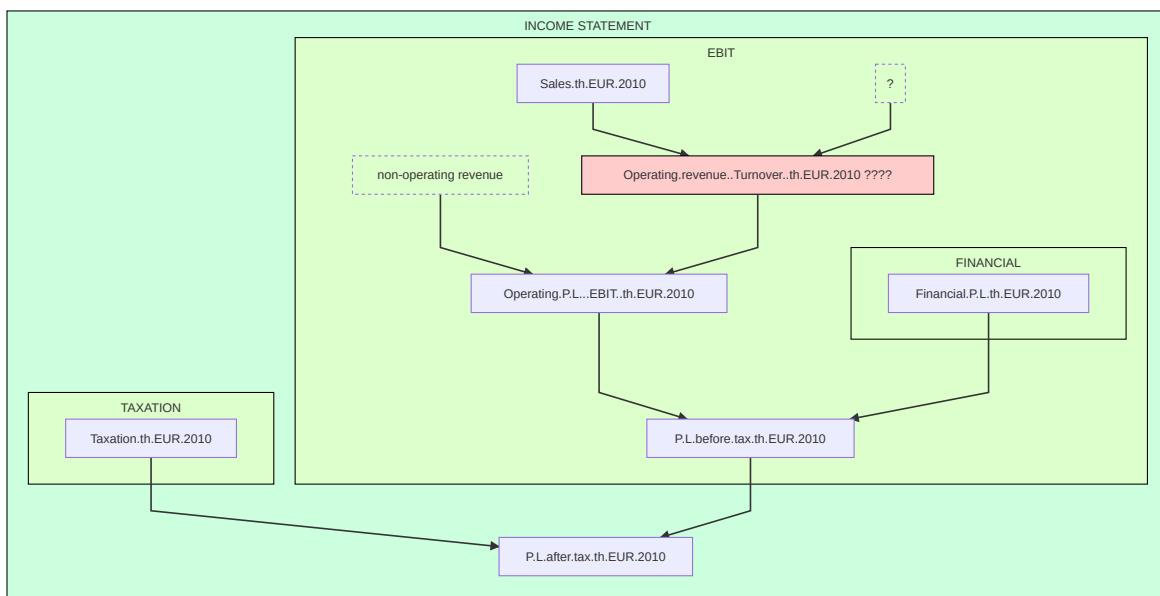
Iterations 9

Logit Regression Results

```
=====
Dep. Variable:          HGF    No. Observations:      110446
Model:                 Logit   Df Residuals:           110444
Method:                MLE    Df Model:                  1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:     0.0001870
Time:          15:36:36   Log-Likelihood:   -17350.
converged:            True    LL-Null:        -17354.
Covariance Type:    nonrobust   LLR p-value:     0.01086
=====
```

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2645	0.016	-201.981	0.000	-3.296	-3.233
OCL	-9.186e-05	5.11e-05	-1.797	0.072	-0.000	8.33e-06

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



HGF vs non-HGF for Operating.revenue..Turnover..th.EUR.2010
 Welch's t-test statistic = 9.251

```
p-value = 2.661e-20
```

```
Optimization terminated successfully.
```

```
    Current function value: 0.155219
```

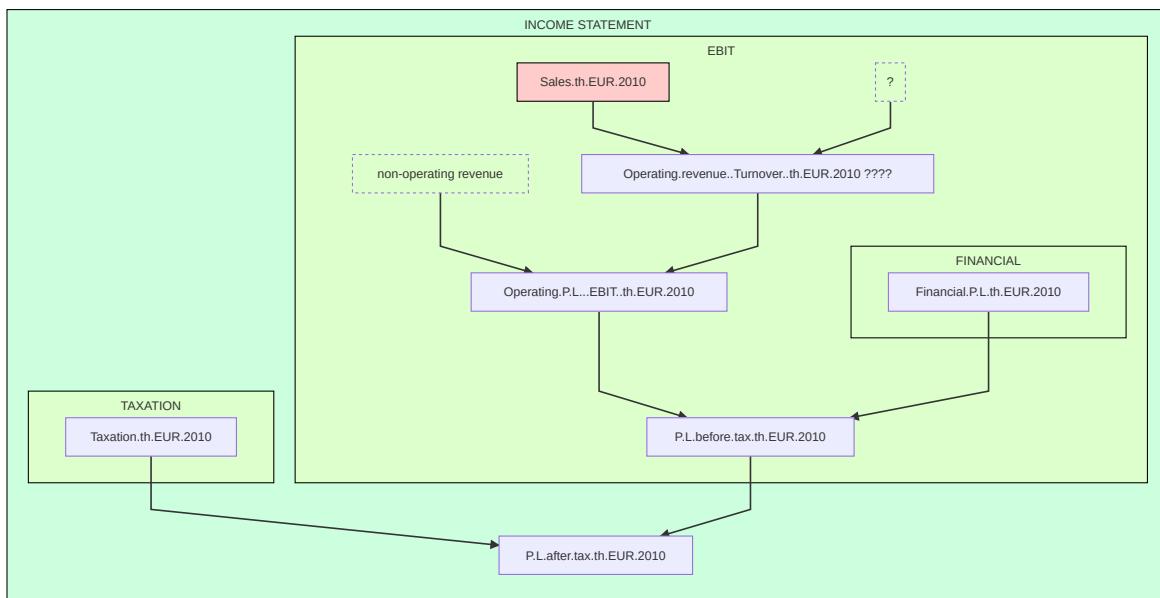
```
    Iterations 11
```

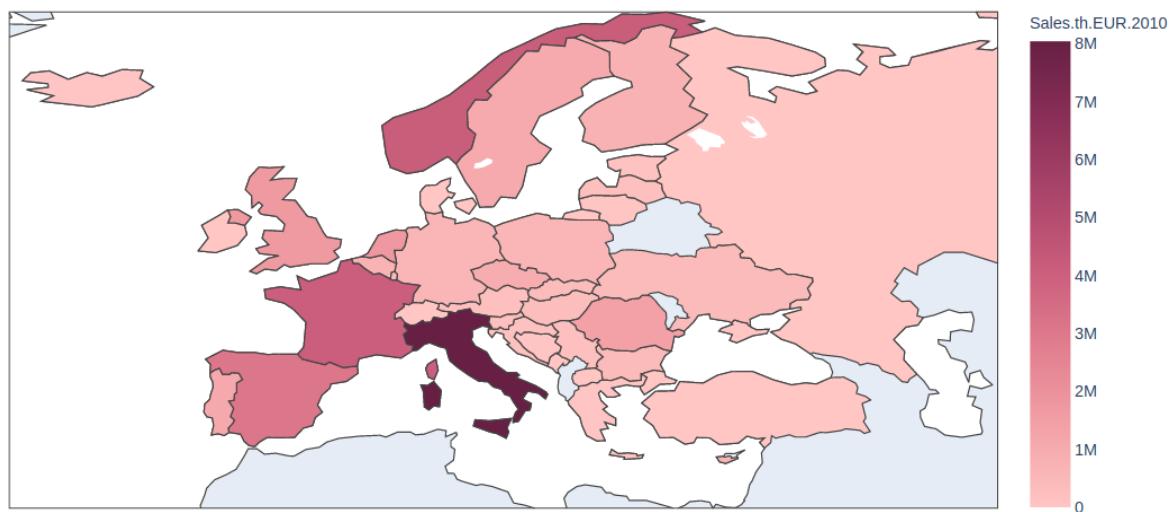
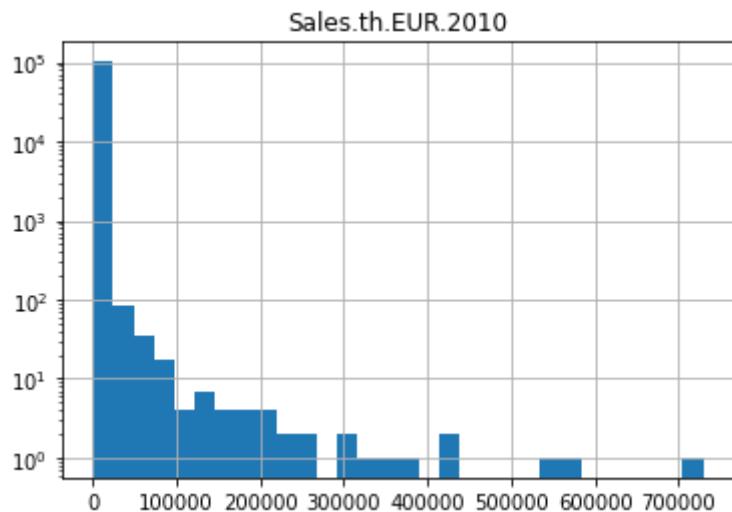
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.002844
Time:	15:37:05	Log-Likelihood:	-17981.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	4.183e-24

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2318	0.017	-193.739	0.000	-3.265	-3.199
ORT	-0.0003	4.57e-05	-6.834	0.000	-0.000	-0.000

Sales.th.EUR.2010





outliers:

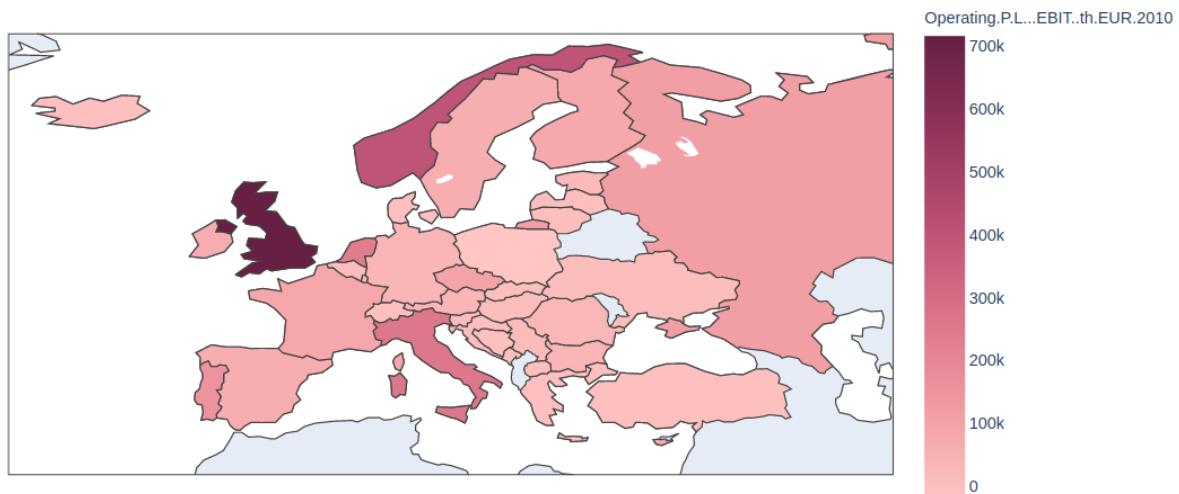
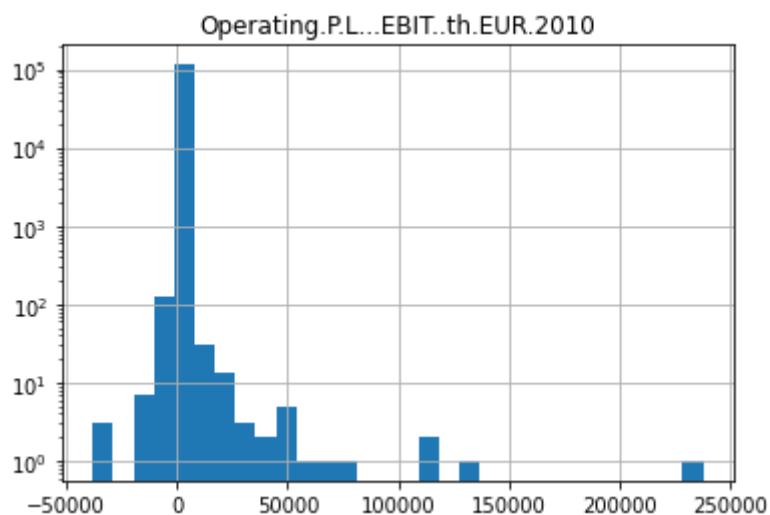
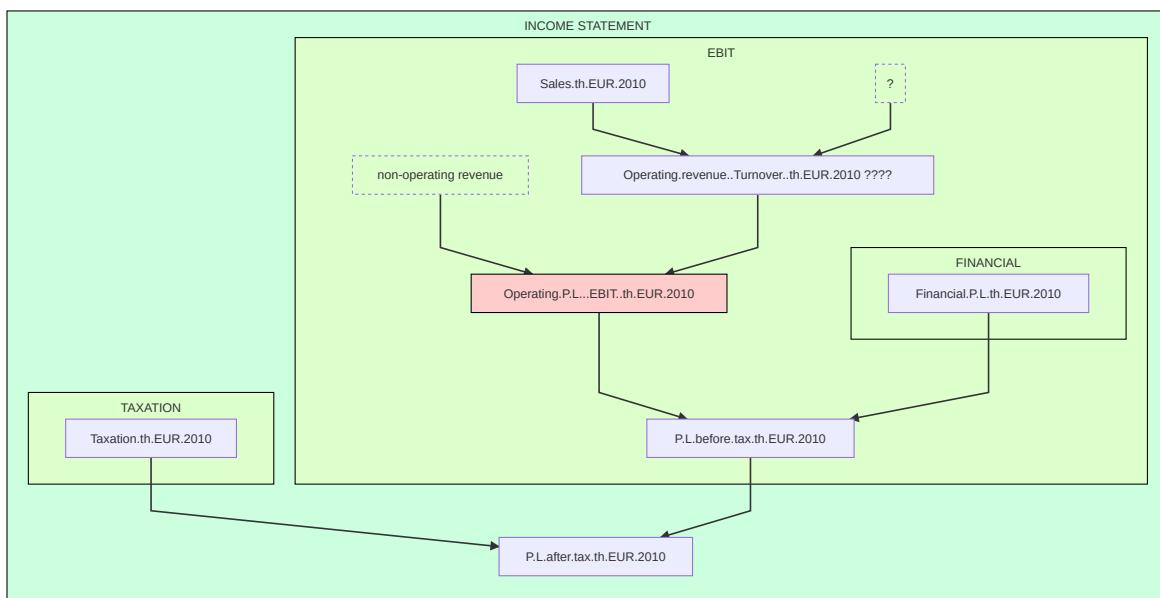
```
BvD.ID.number
CYC266578      HMS HYDRAULIC MACHINES & SYSTEMS GROUP PLC
GB07123187      ACACIA MINING PLC
IT10813301008    EOS S.R.L.

Name: Company.name, dtype: string

Optimization terminated successfully.
    Current function value: 0.160957
    Iterations 11
    Logit Regression Results
=====
Dep. Variable:          HGF    No. Observations:      102550
Model:                 Logit   Df Residuals:        102548
Method:                MLE    Df Model:             1
Date: Mon, 29 Jun 2020   Pseudo R-squ.:     0.005149
Time: 15:37:31           Log-Likelihood: -16506.
converged:            True    LL-Null:            -16592.
Covariance Type:    nonrobust   LLR p-value: 4.804e-39
=====

            coef      std err       z     P>|z|      [0.025      0.975]
-----
Intercept   -3.1452    0.018  -177.808    0.000    -3.180    -3.111
SAL         -0.0007  8.11e-05   -9.127    0.000    -0.001    -0.001
=====
```

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

HGF vs non-HGF for Operating.P.L...EBIT..th.EUR.2010
 Welch's t-test statistic = 4.539
 p-value = 5.771e-06

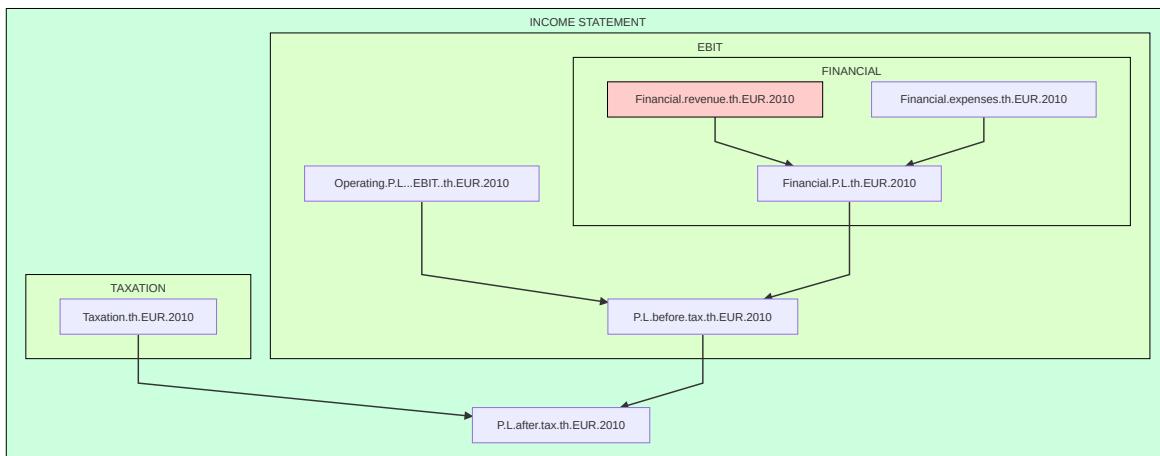
Optimization terminated successfully.

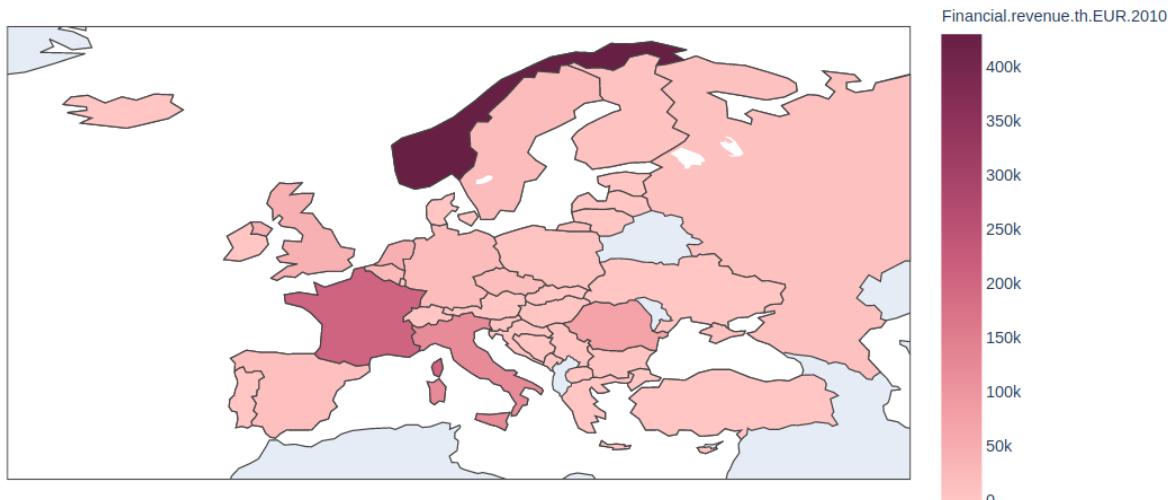
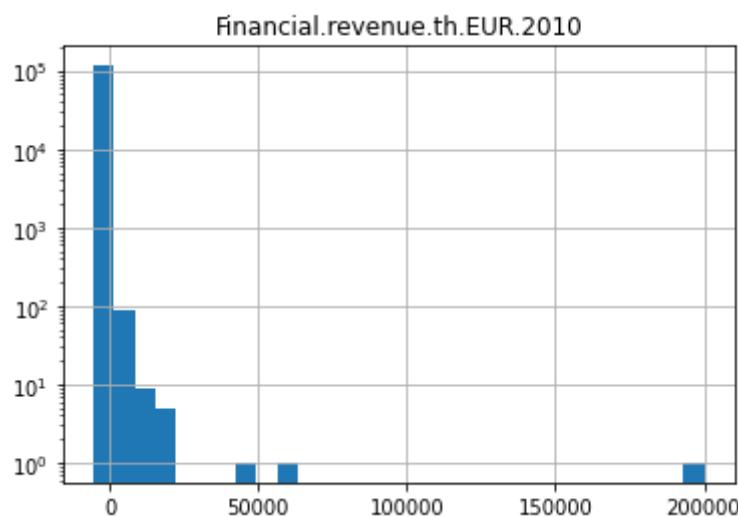
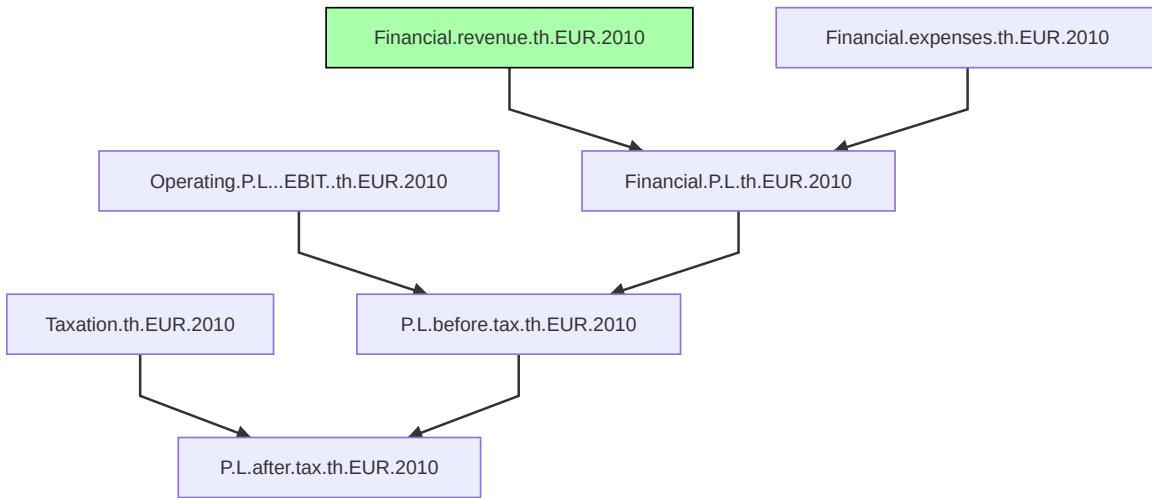
Current function value: 0.155585
 Iterations 7

Logit Regression Results

```
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:        115838
Method:                MLE    Df Model:             1
Date:      Mon, 29 Jun 2020 Pseudo R-squ.:     0.0004971
Time:          15:37:58  Log-Likelihood:   -18023.
converged:         True   LL-Null:            -18032.
Covariance Type:  nonrobust LLR p-value:  2.296e-05
=====
            coef      std err       z     P>|z|      [0.025      0.975]
-----
Intercept   -3.2814      0.016   -208.555      0.000     -3.312     -3.251
EBIT       -0.0001  3.08e-05     -4.079      0.000     -0.000   -6.54e-05
=====
```

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
```

Optimization terminated successfully.

Current function value: 0.155644

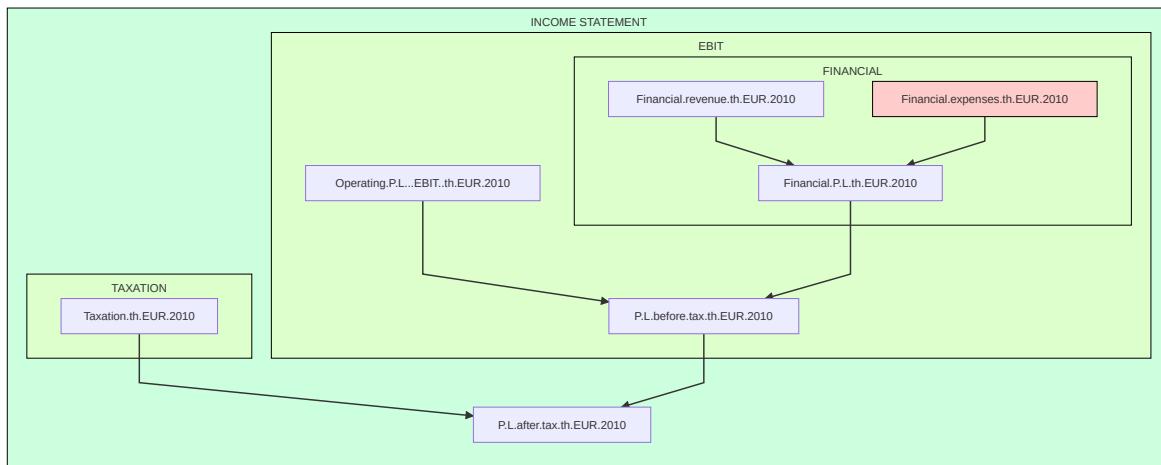
Iterations 9

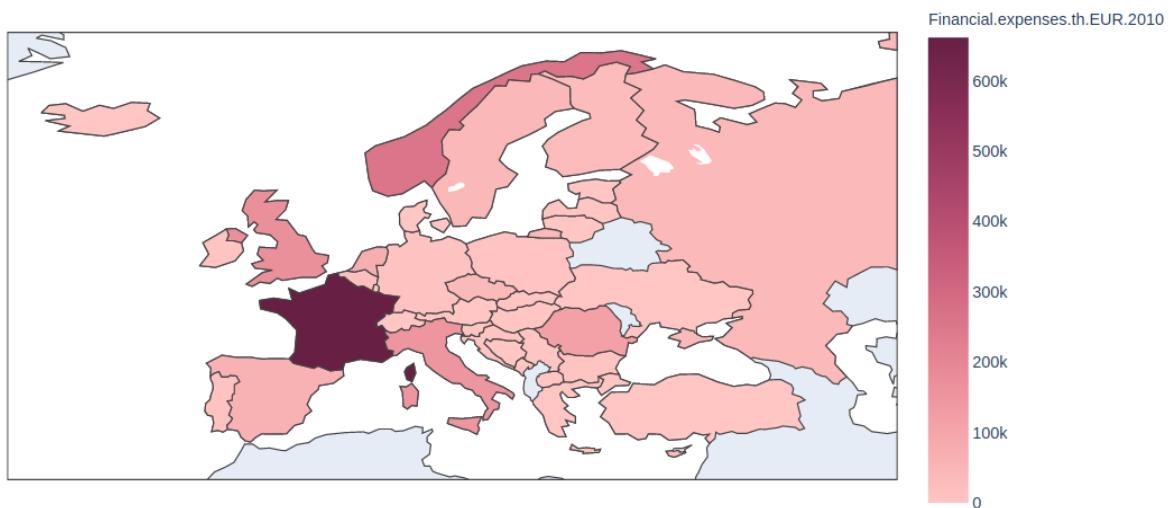
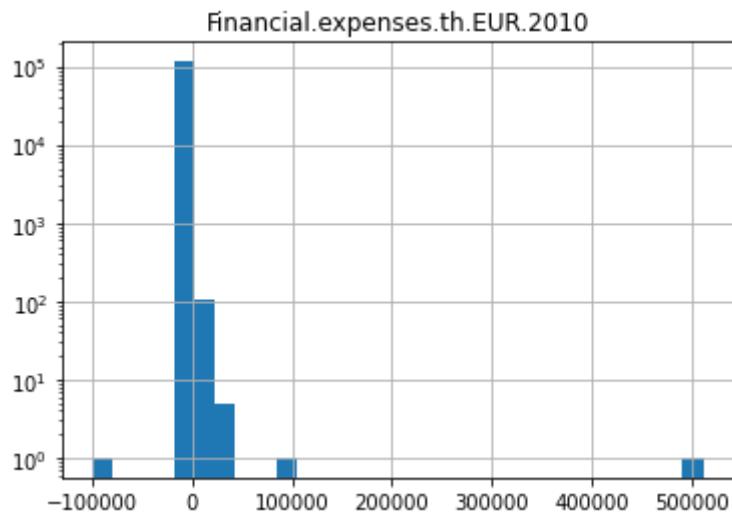
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.0001129
Time:	15:38:23	Log-Likelihood:	-18030.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	0.04361

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2807	0.016	-208.461	0.000	-3.312	-3.250
FR	-0.0003	0.000	-1.646	0.100	-0.001	5.98e-05

Financial.expenses.th.EUR.2010





positive outlier:

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

negative outlier:

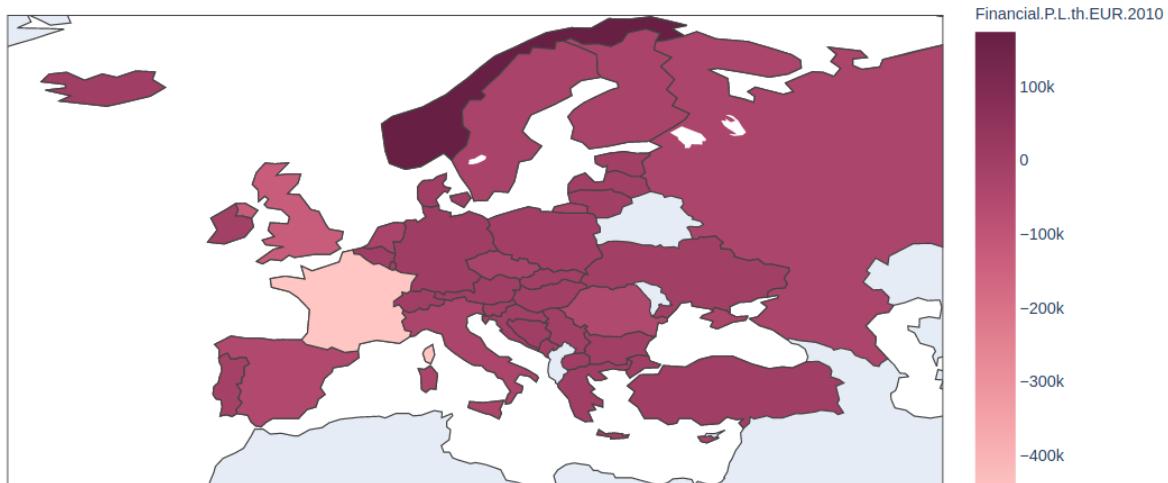
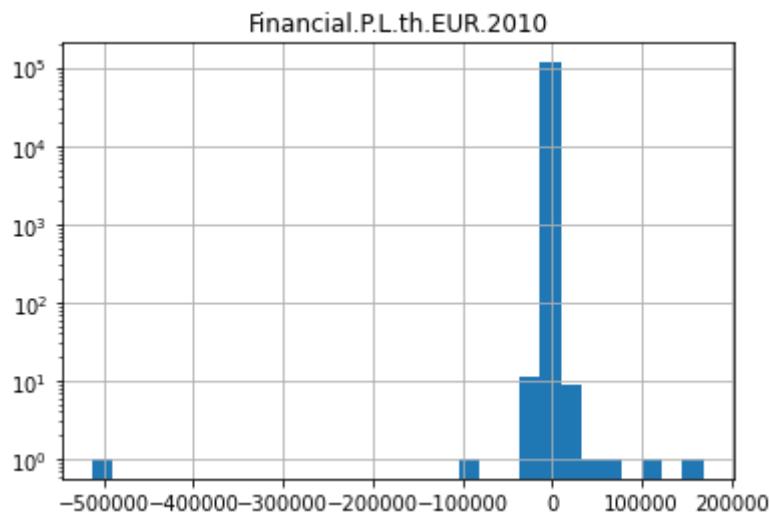
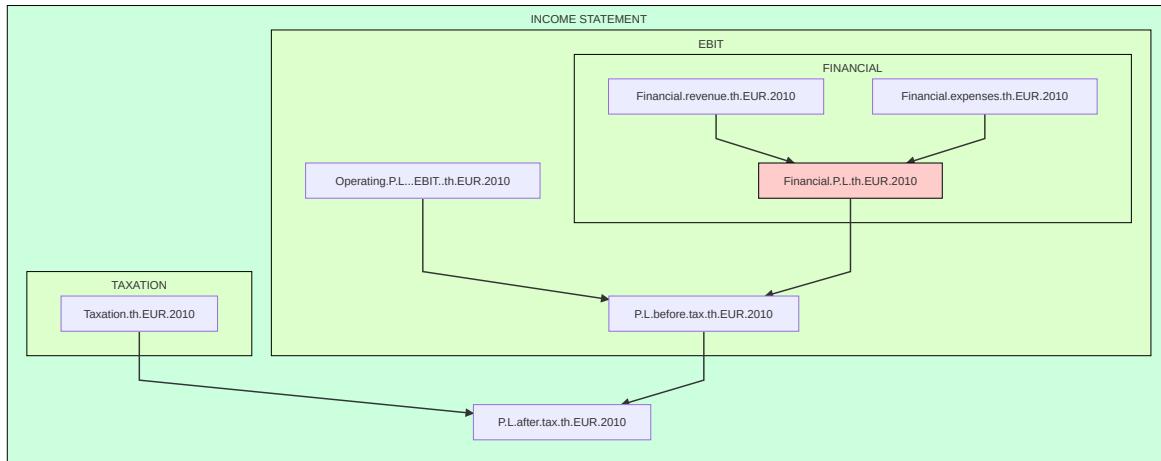
```
BvD.ID.number
GB07145051      CAPITAL & COUNTIES PROPERTIES PLC
Name: Company.name, dtype: string
```

HGF vs non-HGF for Financial.expenses.th.EUR.2010
 Welch's t-test statistic = 2.789
 p-value = 0.005292

Optimization terminated successfully.
 Current function value: 0.155655
 Iterations 7
 Logit Regression Results
 ======
 Dep. Variable: HGF No. Observations: 115840
 Model: Logit Df Residuals: 115838
 Method: MLE Df Model: 1
 Date: Mon, 29 Jun 2020 Pseudo R-squ.: 4.376e-05
 Time: 15:38:43 Log-Likelihood: -18031.

converged:	True	LL-Null:	-18032.			
Covariance Type:	nonrobust	LLR p-value:	0.2090			
<hr/>						
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2817	0.016	-208.616	0.000	-3.313	-3.251
FE	-2.787e-05	1.96e-05	-1.425	0.154	-6.62e-05	1.05e-05
<hr/>						

Financial.P.L.th.EUR.2010



Outlier:

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

```
HGF vs non-HGF for Financial.P.L.th.EUR.2010  
Welch's t-test statistic = -1.5  
p-value = 0.1337
```

Optimization terminated successfully.

Current function value: 0.155661

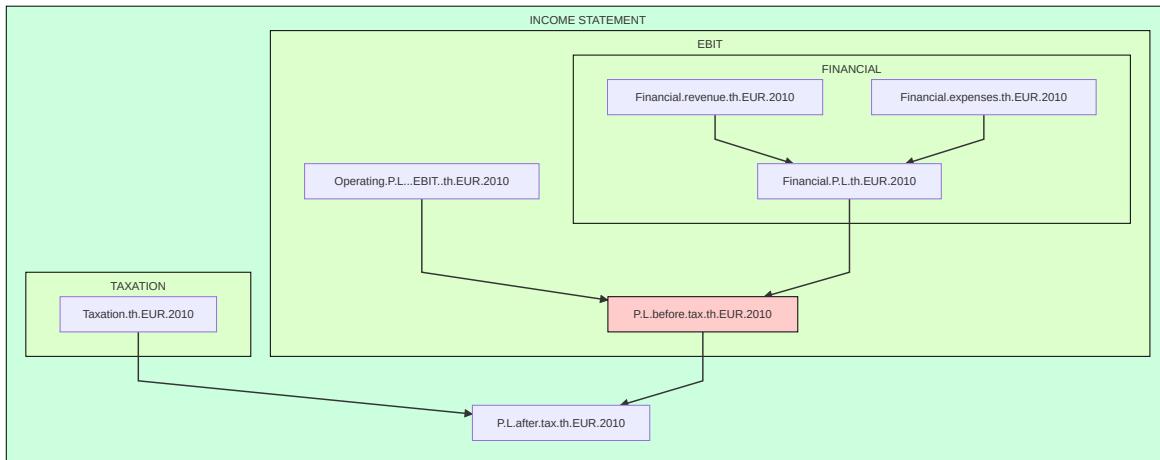
Iterations 7

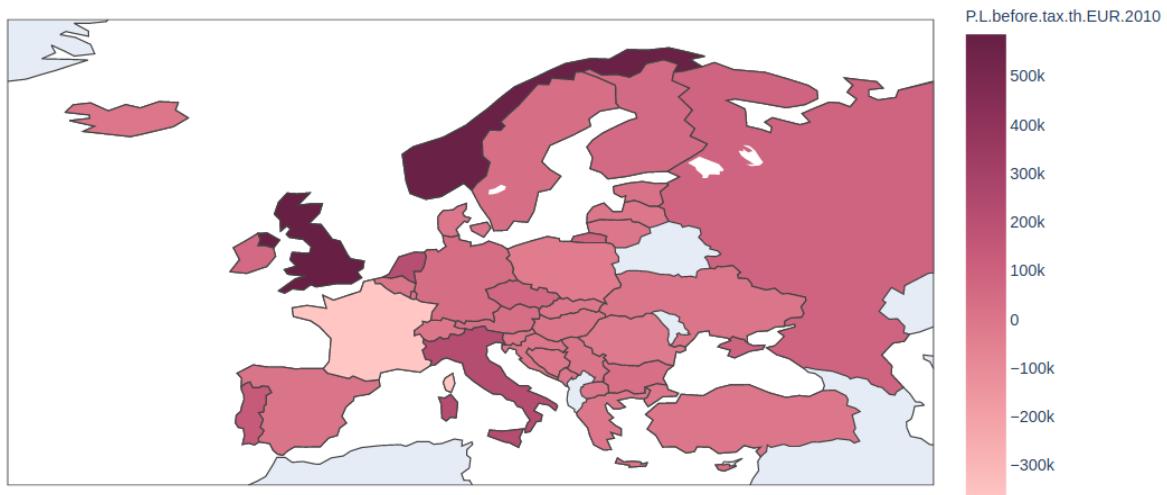
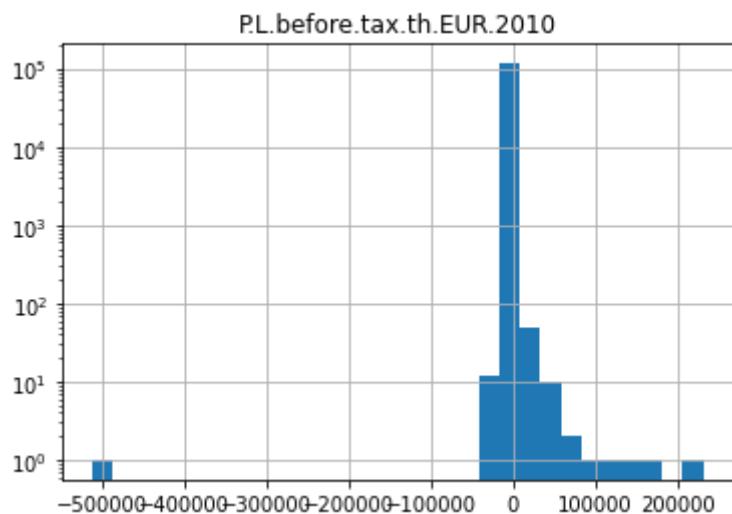
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	5.126e-06
Time:	15:39:06	Log-Likelihood:	-18032.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	0.6672

	coef	std err	z	P> z	[0.025	0.975]
<hr/>						
Intercept	-3.2819	0.016	-208.633	0.000	-3.313	-3.251
FPL	6.226e-06	1.53e-05	0.408	0.684	-2.37e-05	3.62e-05

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





outlier:

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

HGF vs non-HGF for P.L.before.tax.th.EUR.2010
Welch's t-test statistic = 3.793
p-value = 0.000149

Optimization terminated successfully.

Current function value: 0.155659

Iterations 7

Logit Regression Results

```

-----  

Dep. Variable: HGF No. Observations: 115840  

Model: Logit Df Residuals: 115838  

Method: MLE Df Model: 1  

Date: Mon, 29 Jun 2020 Pseudo R-squ.: 1.779e-05  

Time: 15:39:22 Log-Likelihood: -18032.  

converged: True LL-Null: -18032.  

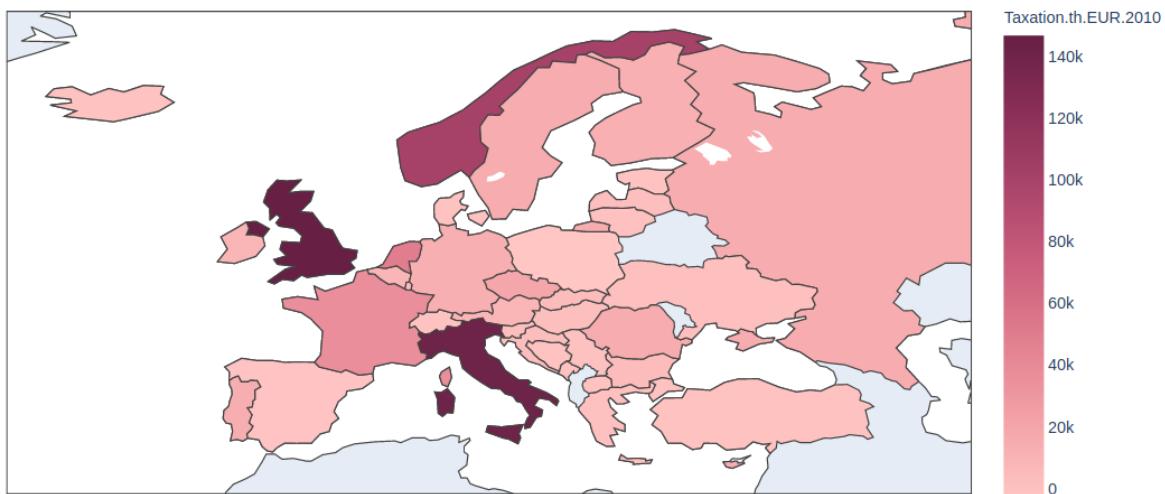
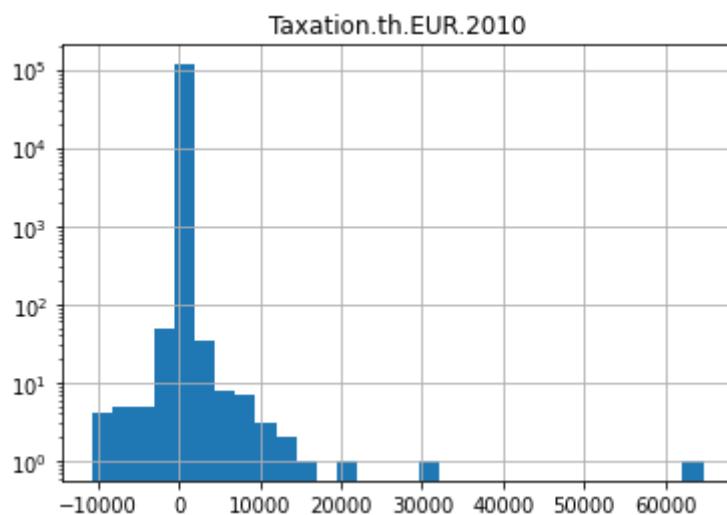
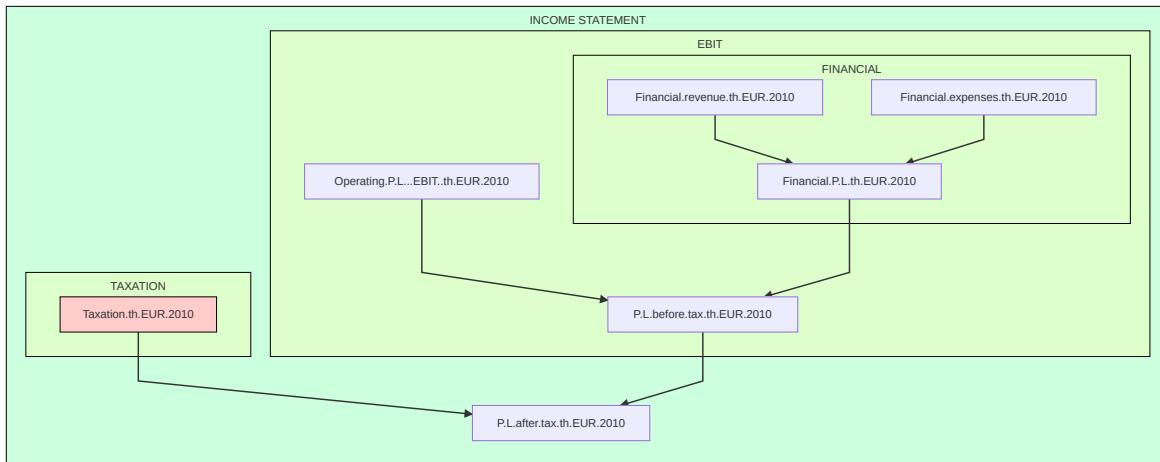
Covariance Type: nonrobust LLR p-value: 0.4232

```

	coef	std err	z	P> z	[0.025	0.975]
<hr/>						
Intercept	-3.2819	0.016	-208.631	0.000	-3.313	-3.251

FPL	-4.184e-06	4.37e-06	-0.957	0.339	-1.28e-05	4.39e-06
-----	------------	----------	--------	-------	-----------	----------

Taxation.th.EUR.2010



outlier:

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

HGF vs non-HGF for Taxation.th.EUR.2010

Welch's t-test statistic = 5.486

p-value = 4.128e-08

Optimization terminated successfully.

Current function value: 0.155649

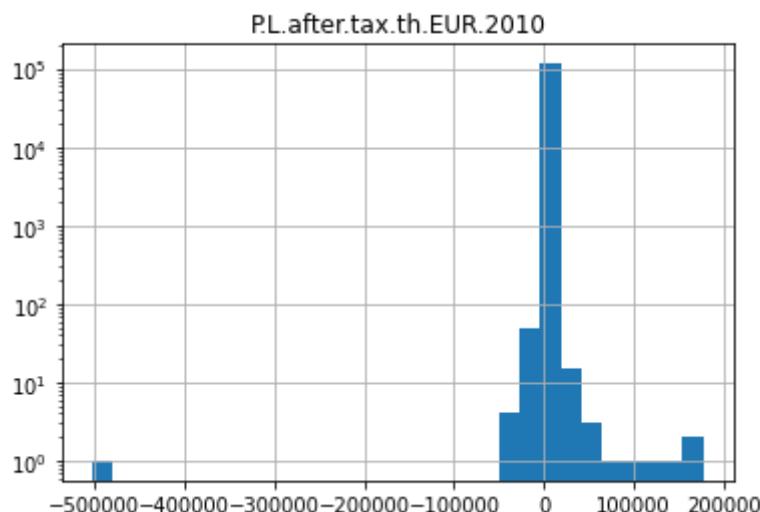
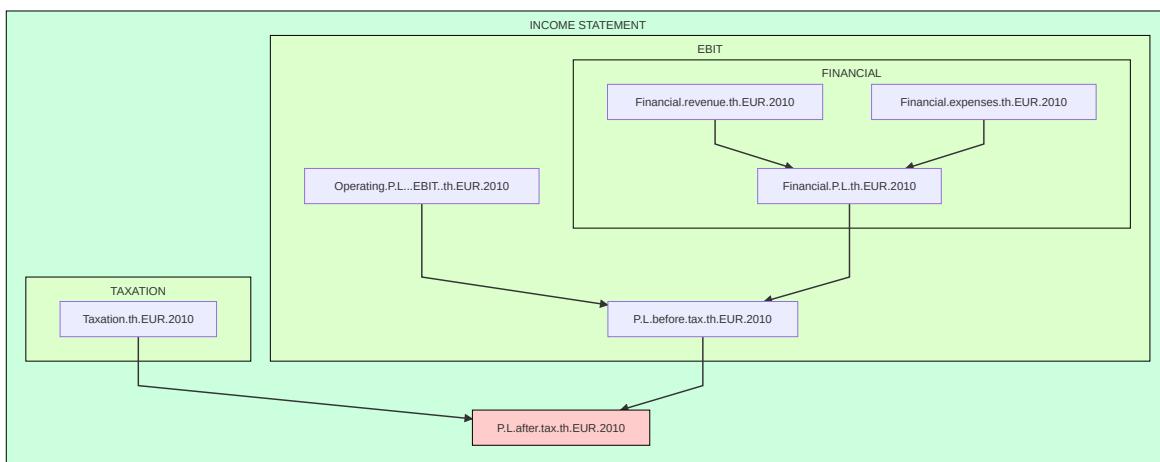
Iterations 7

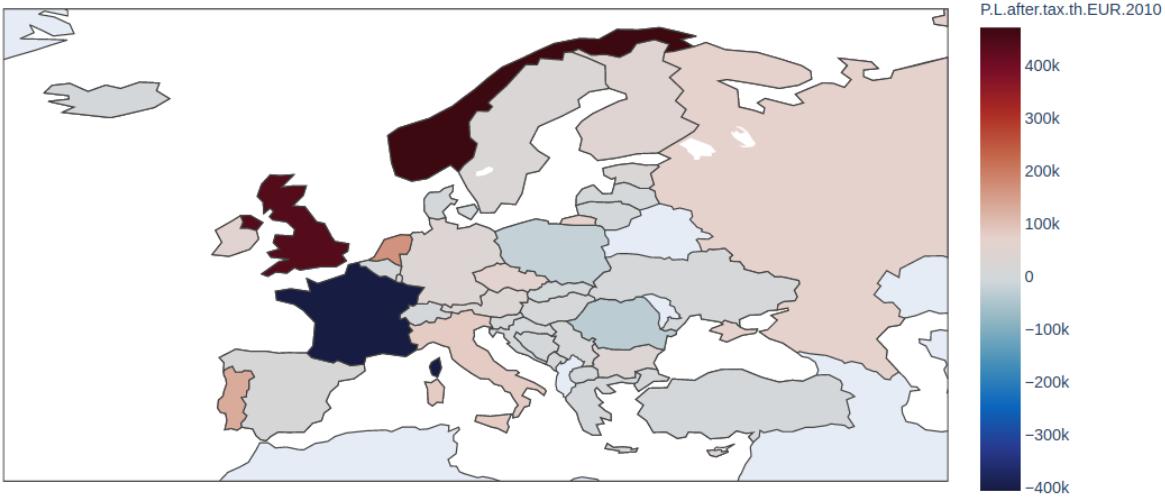
Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	8.533e-05
Time:	15:39:38	Log-Likelihood:	-18030.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	0.07940

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2814	0.016	-208.591	0.000	-3.312	-3.251
FPL	-0.0002	0.000	-1.794	0.073	-0.000	1.68e-05

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





outlier:

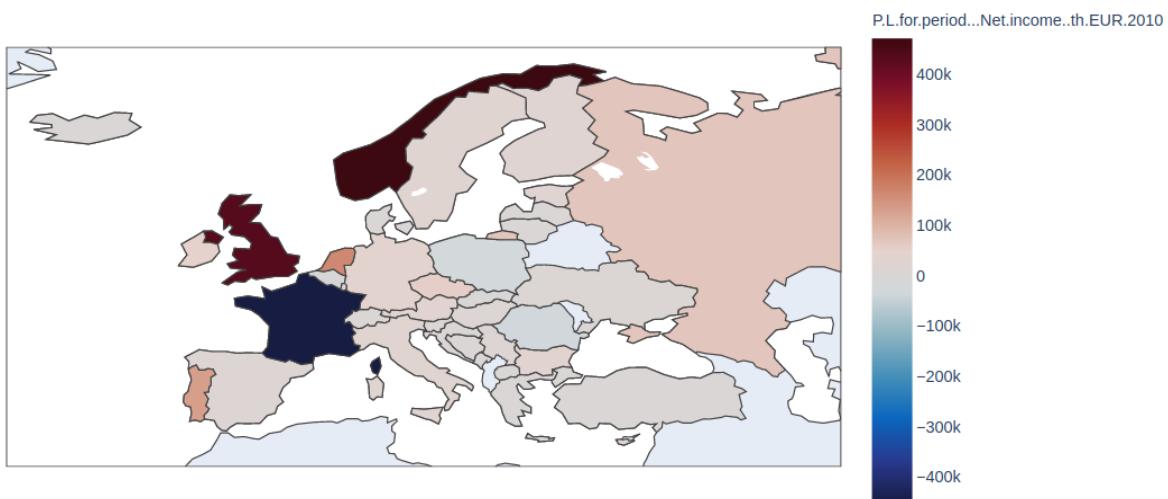
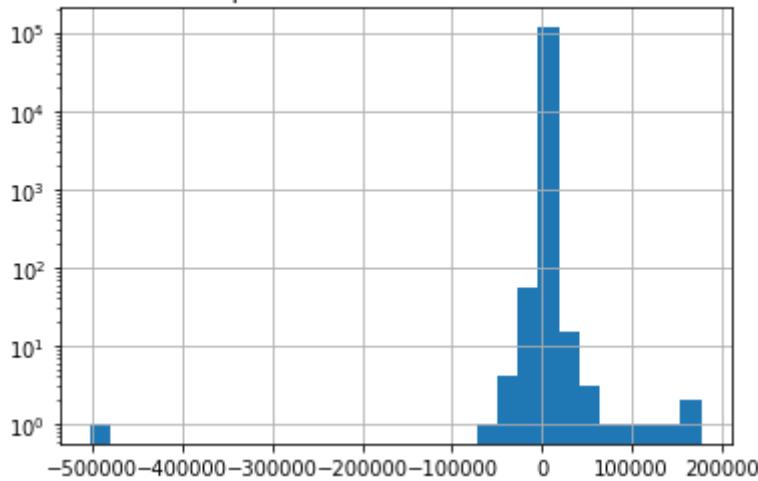
```
BvD.ID.number
FR519720643    IRIDIUM FRANCE
Name: Company.name, dtype: string
```

```
HGF vs non-HGF for P.L.after.tax.th.EUR.2010
Welch's t-test statistic = 3.352
p-value = 0.0008042
```

```
Optimization terminated successfully.
      Current function value: 0.155660
      Iterations 7
      Logit Regression Results
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:        115838
Method:                MLE    Df Model:             1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:  1.429e-05
Time:          15:40:03     Log-Likelihood: -18032.
converged:         True    LL-Null:            -18032.
Covariance Type:  nonrobust   LLR p-value:       0.4728
=====
            coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept    -3.2819      0.016     -208.632      0.000     -3.313     -3.251
PL          -3.991e-06  4.65e-06      -0.858      0.391    -1.31e-05  5.13e-06
=====
```

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
```

```
HGF vs non-HGF for P.L.for.period...Net.income..th.EUR.2010
Welch's t-test statistic = 3.123
p-value = 0.001791
```

```
Optimization terminated successfully.
      Current function value: 0.155660
      Iterations 7
```

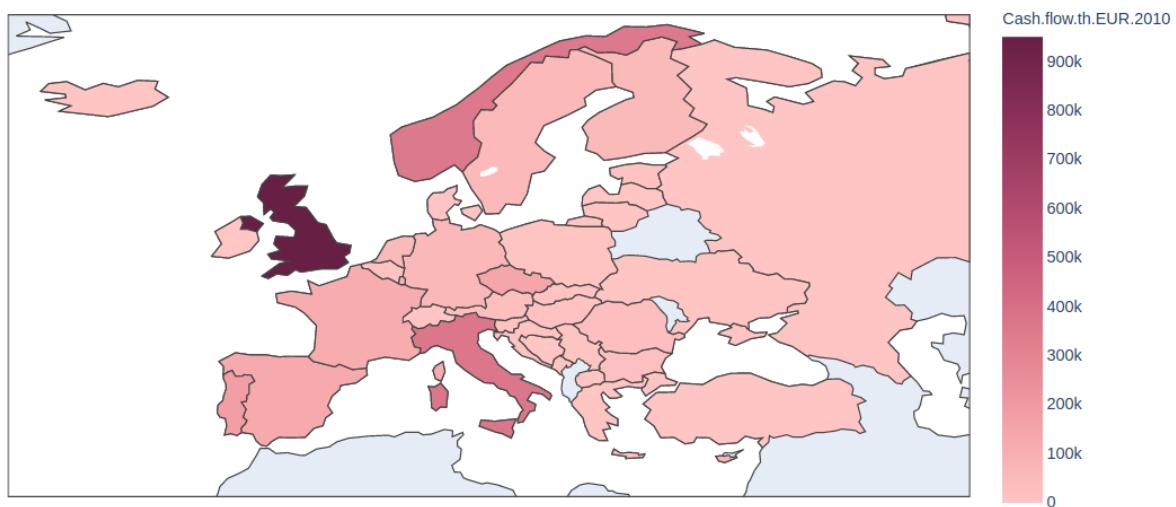
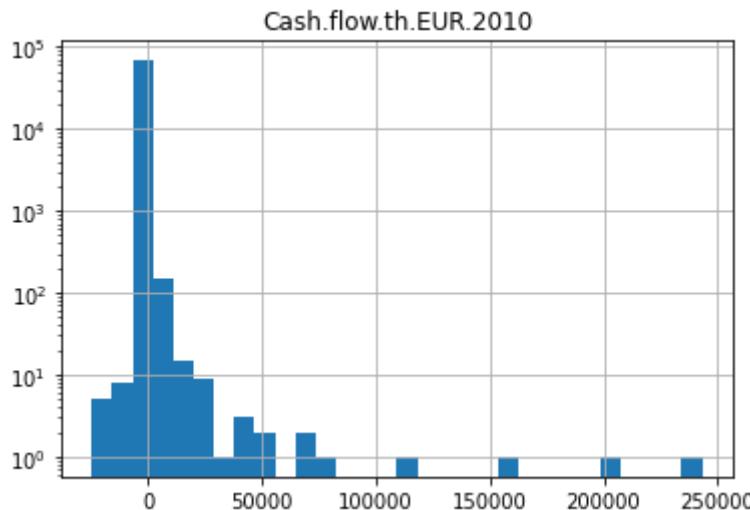
Logit Regression Results

```
=====
Dep. Variable:                  HGF      No. Observations:             115840
Model:                          Logit     Df Residuals:                  115838
Method:                         MLE      Df Model:                      1
Date:                Mon, 29 Jun 2020   Pseudo R-squ.:        1.267e-05
Time:                15:40:20       Log-Likelihood:   -18032.
converged:                     True      LL-Null:          -18032.
Covariance Type:            nonrobust    LLR p-value:        0.4991
=====
```

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2819	0.016	-208.632	0.000	-3.313	-3.251

PL	-3.82e-06	4.75e-06	-0.804	0.421	-1.31e-05	5.49e-06
----	-----------	----------	--------	-------	-----------	----------

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
```

Optimization terminated successfully.

Current function value: 0.162521

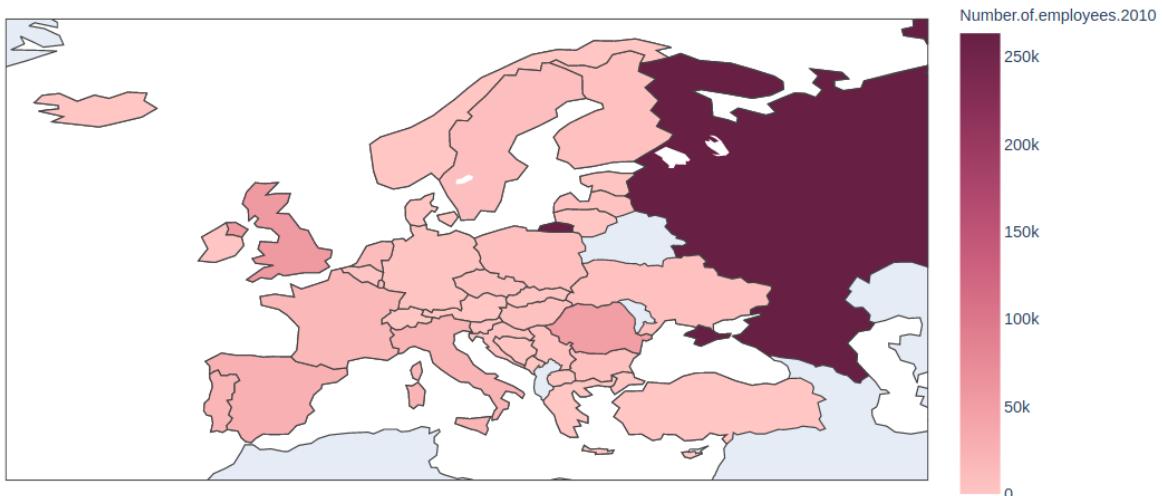
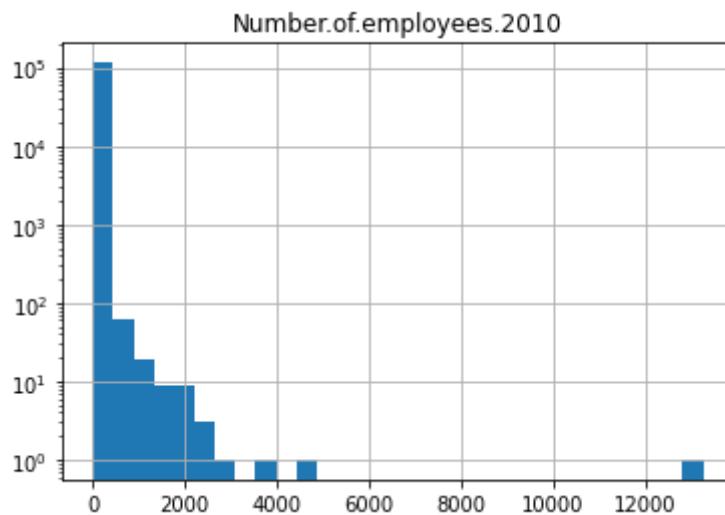
Iterations 7

Logit Regression Results

```
=====
Dep. Variable:           HGF    No. Observations:      68021
Model:                 Logit   Df Residuals:          68019
Method:                MLE    Df Model:                 1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:     0.0007085
Time:          15:40:47    Log-Likelihood:   -11055.
converged:            True    LL-Null:        -11063.
Covariance Type:    nonrobust  LLR p-value:    7.517e-05
=====
```

	coef	std err	z	P> z	[0.025	0.975]
<hr/>						
Intercept	-3.2206	0.020	-161.248	0.000	-3.260	-3.181
CF	-0.0001	3.46e-05	-4.110	0.000	-0.000	-7.44e-05
<hr/>						

Number.of.employees.2010



outlier:

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

```
HGF vs non-HGF for Number.of.employees.2010
Welch's t-test statistic =  8.77
p-value = 2.022e-18
```

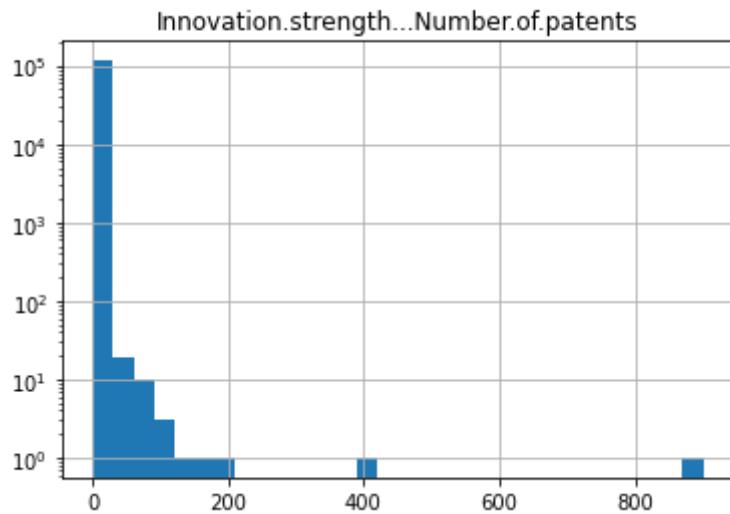
```
Optimization terminated successfully.
      Current function value: 0.155502
      Iterations 9
      Logit Regression Results
=====
```

```
Dep. Variable:          HGF    No. Observations:      115840
```

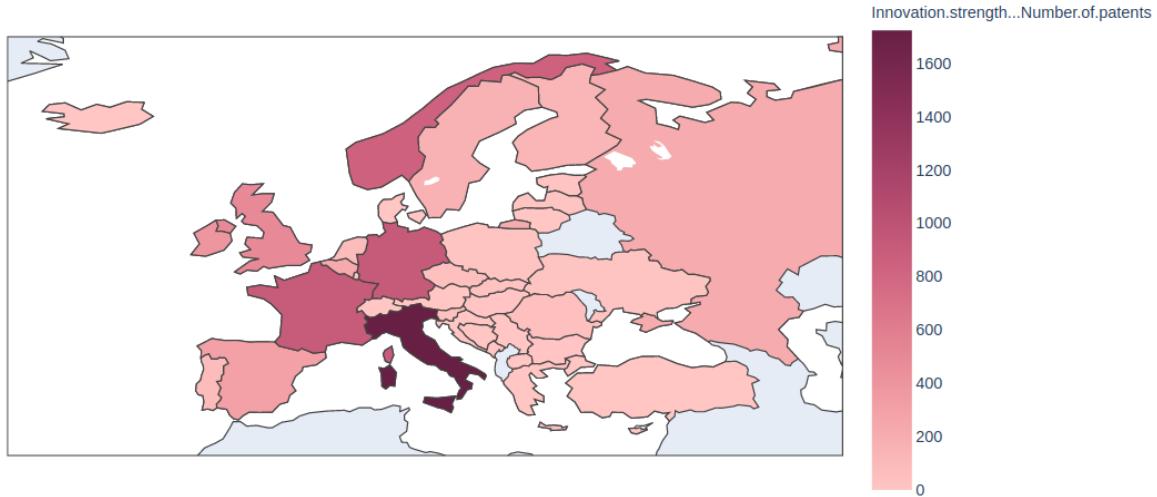
```

Model:                               Logit   Df Residuals:           115838
Method:                             MLE     Df Model:                  1
Date:      Mon, 29 Jun 2020          Pseudo R-squ.:        0.001025
Time:      15:41:13                Log-Likelihood:       -18013.
converged:                           True    LL-Null:            -18032.
Covariance Type: nonrobust         LLR p-value:      1.198e-09
=====
                                         coef    std err      z    P>|z|    [0.025    0.975]
-----
Intercept   -3.2594    0.016   -201.417    0.000   -3.291   -3.228
E          -0.0070    0.001    -4.796    0.000   -0.010   -0.004
=====
```

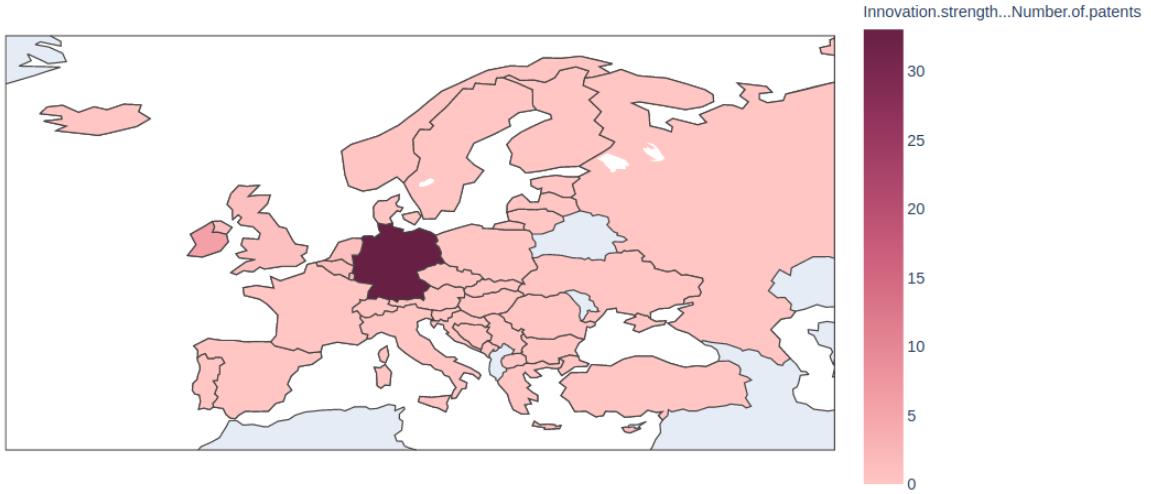
Innovation.strength...Number.of.patents



sum:



mean:



outliers:

```
BvD.ID.number
IE507678      HORIZON THERAPEUTICS PUBLIC LIMITED COMPANY
DE8190460728      WAVELIGHT GMBH
Name: Company.name, dtype: string
```

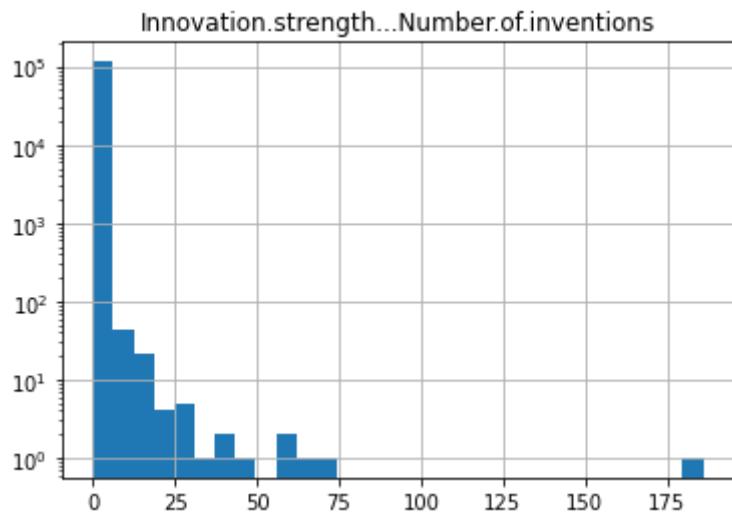
```
HGF vs non-HGF for Number.of.patents
Welch's t-test statistic = -2.068
p-value = 0.03868
```

```
Optimization terminated successfully.
    Current function value: 0.155654
    Iterations 7
```

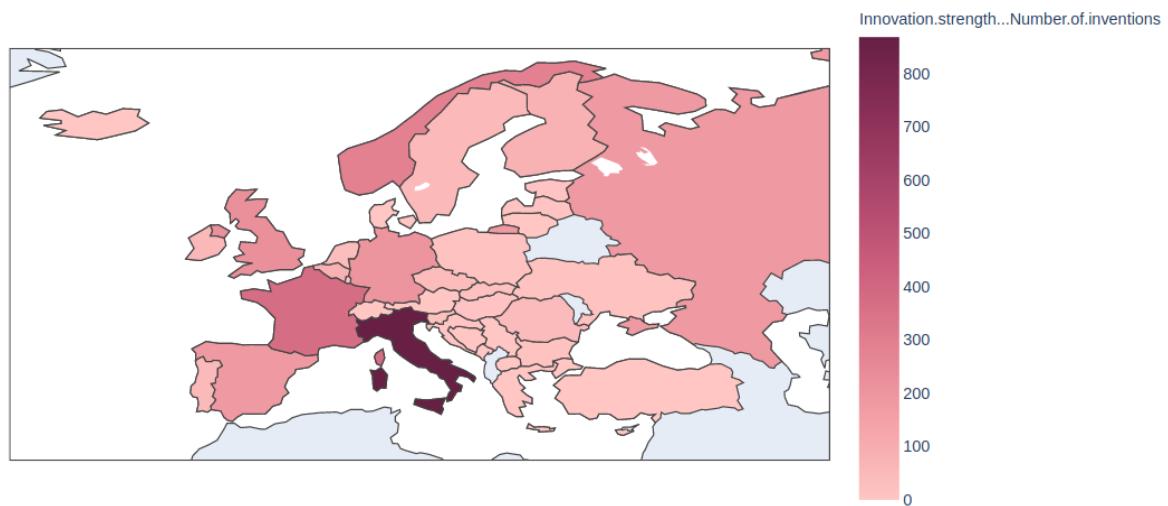
Logit Regression Results

```
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:           115838
Method:                MLE    Df Model:                  1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:     5.332e-05
Time:          15:41:34    Log-Likelihood:   -18031.
converged:            True    LL-Null:        -18032.
Covariance Type:    nonrobust   LLR p-value:     0.1655
=====
            coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept   -3.2822      0.016   -208.621      0.000     -3.313     -3.251
PAT         0.0034      0.002       1.641      0.101     -0.001      0.008
=====
```

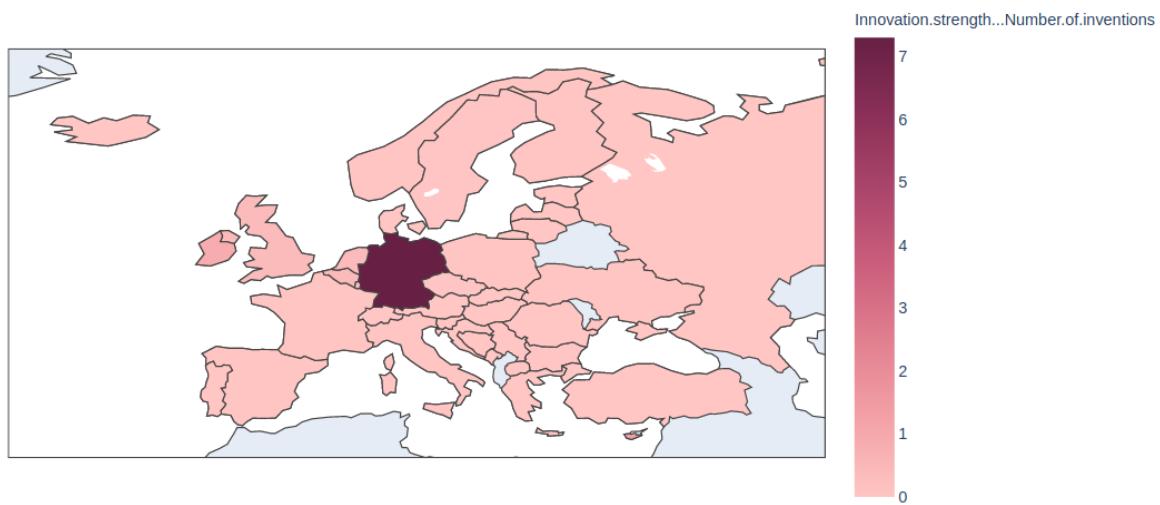
Innovation.strength...Number.of.inventions



sum:



mean:



outlier:

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

```
HGF vs non-HGF for Innovation.strength...Number.of.inventions
Welch's t-test statistic = -1.361
```

```

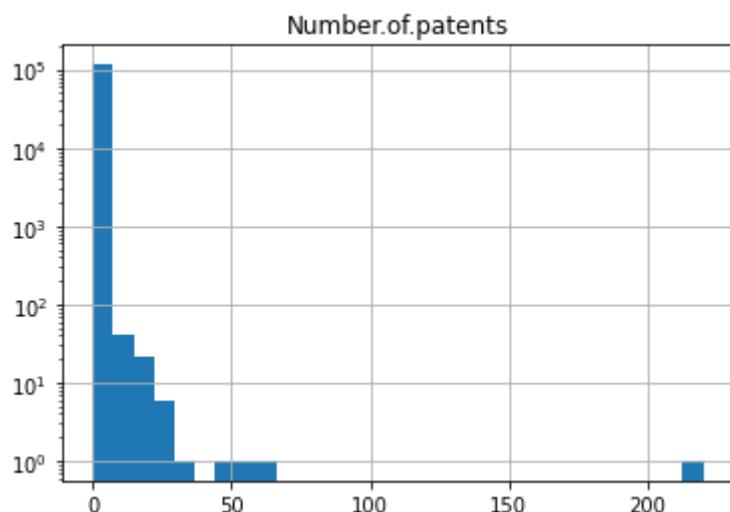
p-value = 0.1737

Optimization terminated successfully.
      Current function value: 0.155656
      Iterations 7

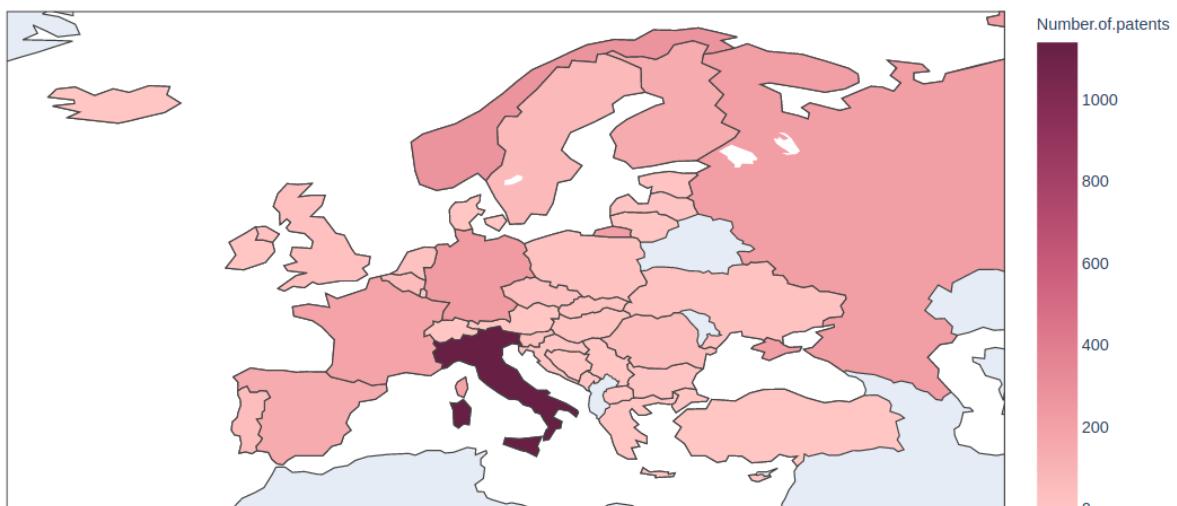
      Logit Regression Results
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:           115838
Method:                MLE    Df Model:                  1
Date: Mon, 29 Jun 2020  Pseudo R-squ.:       3.980e-05
Time: 15:41:56          Log-Likelihood:     -18031.
converged:            True   LL-Null:        -18032.
Covariance Type:    nonrobust  LLR p-value:      0.2309
=====

            coef      std err      z      P>|z|      [0.025      0.975]
-----
Intercept    -3.2824      0.016   -208.594      0.000     -3.313     -3.252
INV         0.0141      0.010      1.425      0.154     -0.005      0.034
=====
```

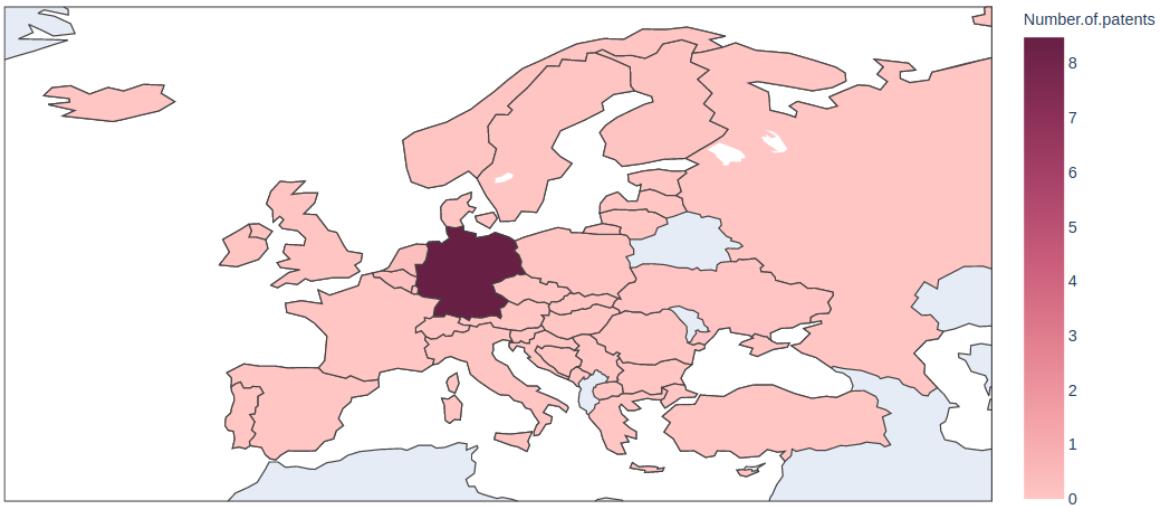
Number.of.patents



sum:



mean:



outlier:

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

```
HGF vs non-HGF for Number.of.patents
Welch's t-test statistic = -2.068
p-value = 0.03868
```

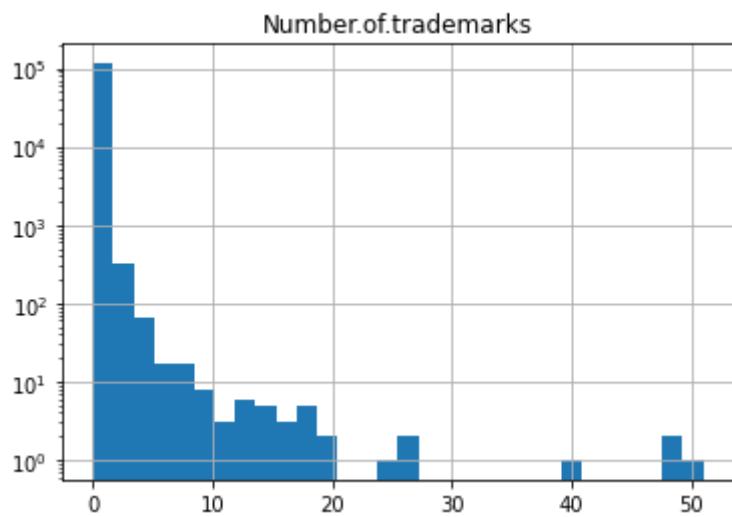
```
Optimization terminated successfully.
    Current function value: 0.155657
    Iterations 7
```

Logit Regression Results

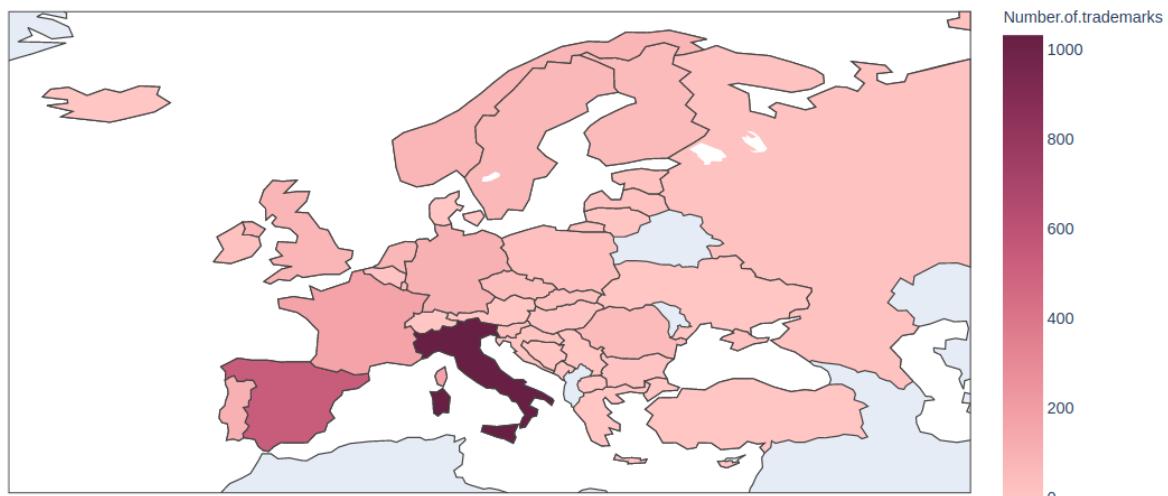
```
=====
Dep. Variable:          HGF    No. Observations:      115840
Model:                 Logit   Df Residuals:        115838
Method:                MLE    Df Model:             1
Date:      Mon, 29 Jun 2020   Pseudo R-squ.:     3.073e-05
Time:          15:42:17      Log-Likelihood:   -18031.
converged:            True    LL-Null:           -18032.
Covariance Type:    nonrobust   LLR p-value:     0.2925
=====
```

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2823	0.016	-208.605	0.000	-3.313	-3.251
P	0.0116	0.009	1.267	0.205	-0.006	0.029

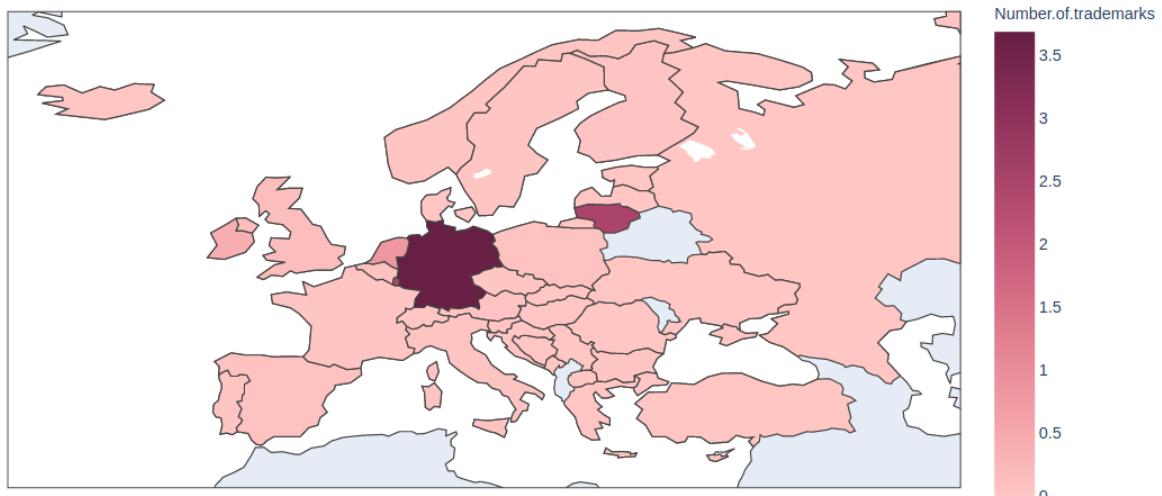
Number.of.trademarks



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	

```
HGF vs non-HGF for Number.of.trademarks
Welch's t-test statistic = -5.38
p-value = 7.824e-08
```

Optimization terminated successfully.

Current function value: 0.155560

Iterations 7

Logit Regression Results

Dep. Variable:	HGF	No. Observations:	115840
Model:	Logit	Df Residuals:	115838
Method:	MLE	Df Model:	1
Date:	Mon, 29 Jun 2020	Pseudo R-squ.:	0.0006523
Time:	15:42:41	Log-Likelihood:	-18020.
converged:	True	LL-Null:	-18032.
Covariance Type:	nonrobust	LLR p-value:	1.234e-06

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-3.2852	0.016	-208.489	0.000	-3.316	-3.254
T	0.0918	0.018	5.037	0.000	0.056	0.128

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
Total	115840

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:

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

Gazele

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

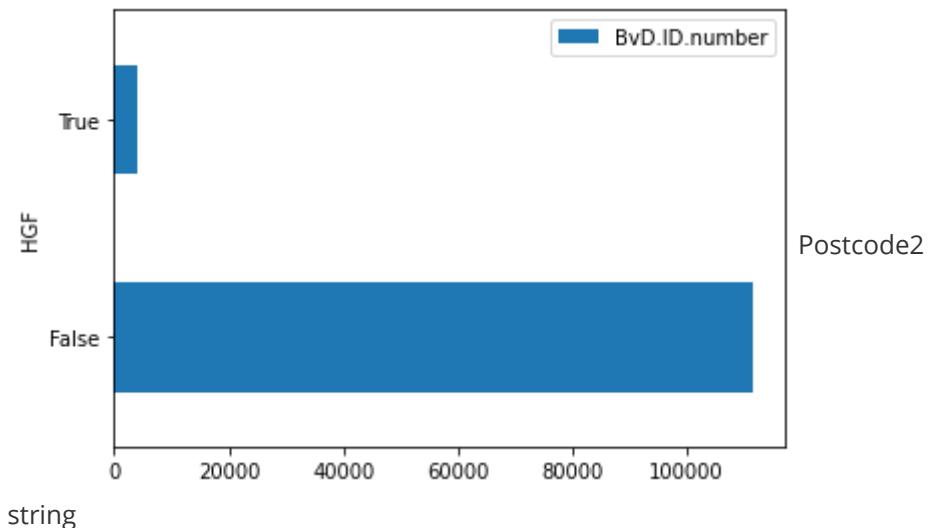
$$Gazele = \text{all} \left(\frac{turnover_t}{turnover_{t-1}} \geq 20\% \right), \text{for } t = 2010, \dots, 2014$$

Eurostat

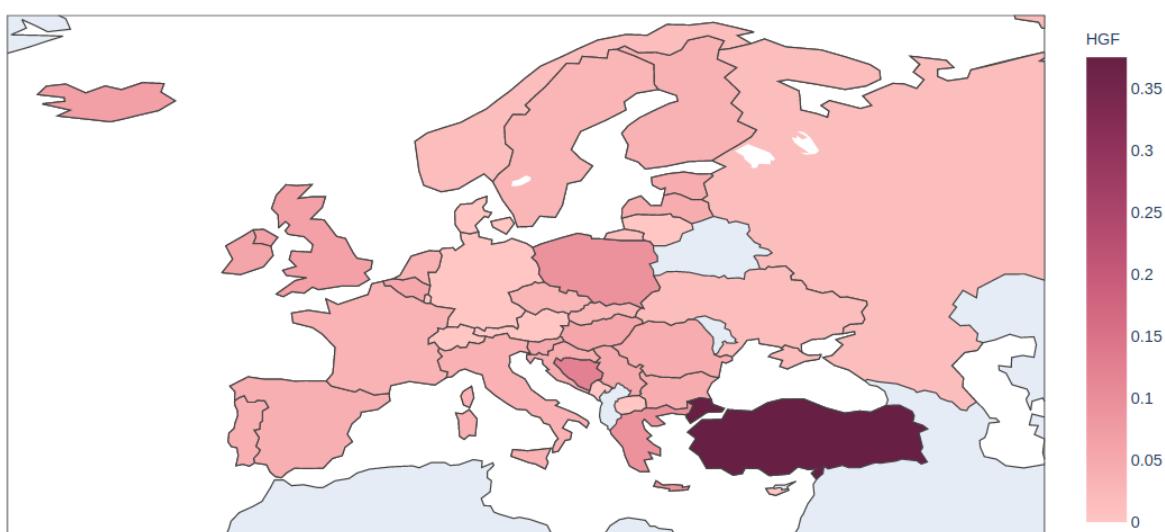
By Eurostat, 2007, being HGF means having a growth rate $\geq 20\%$ for 3 consecutive years.

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

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

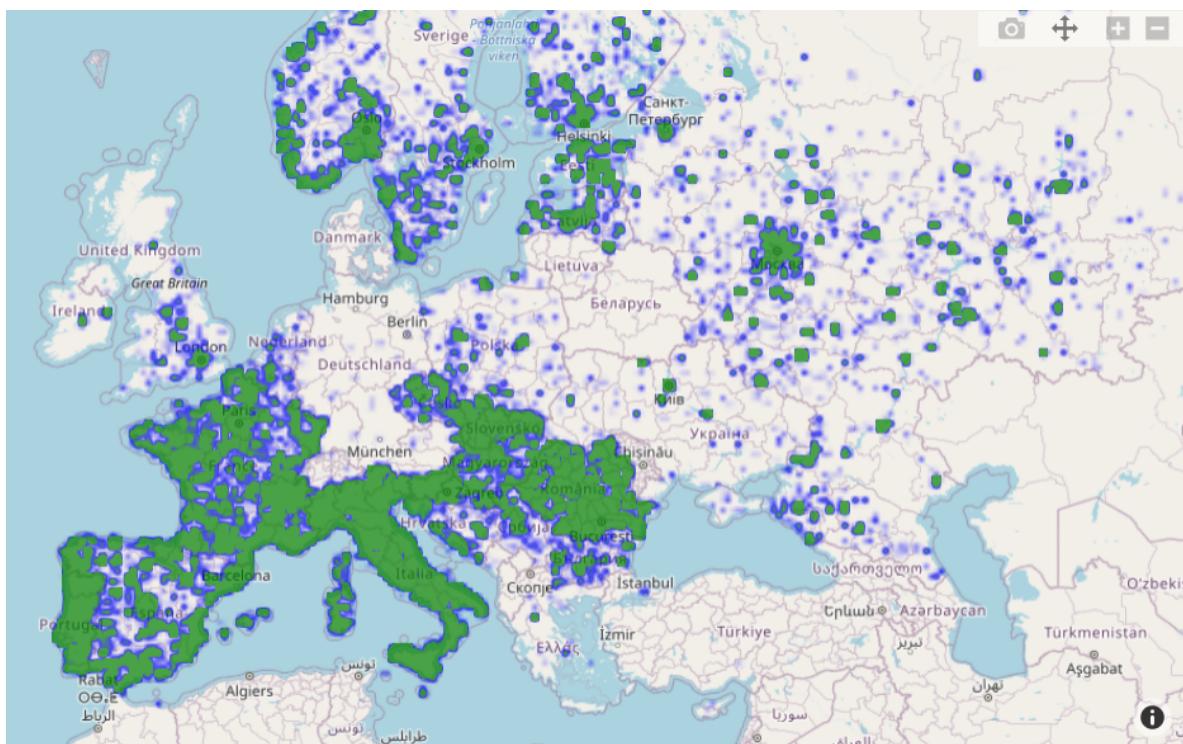
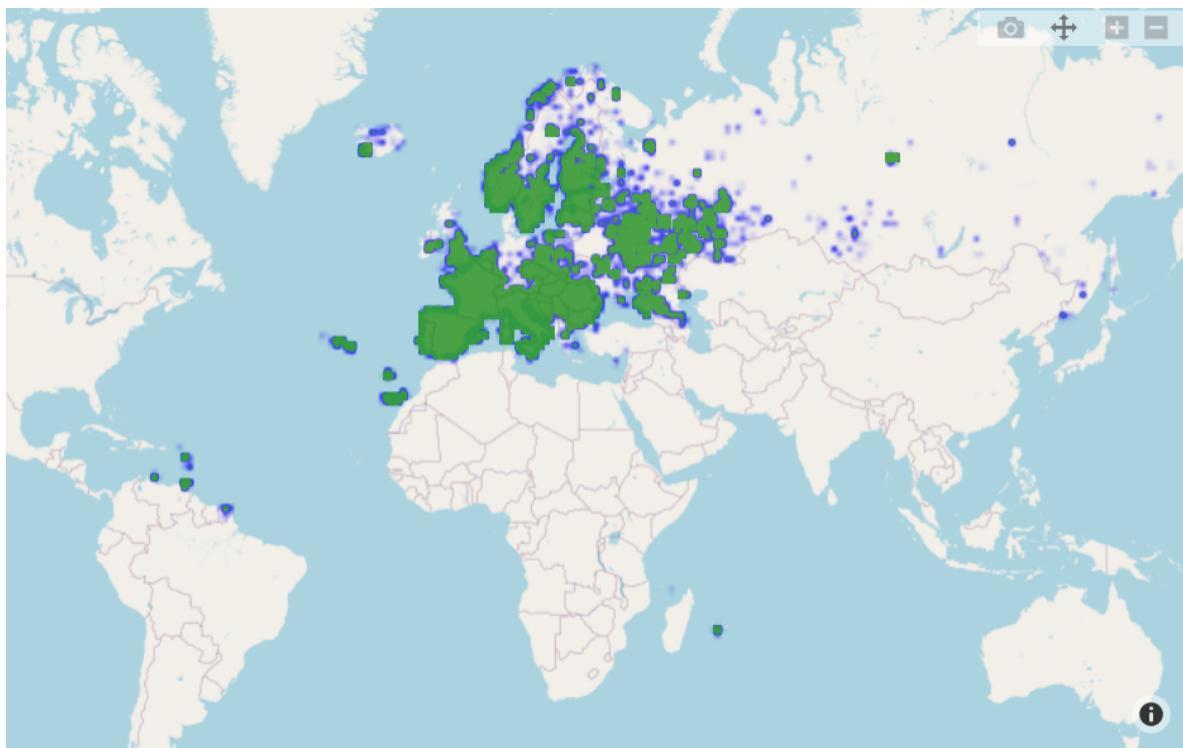


$p(HGF|country)$:



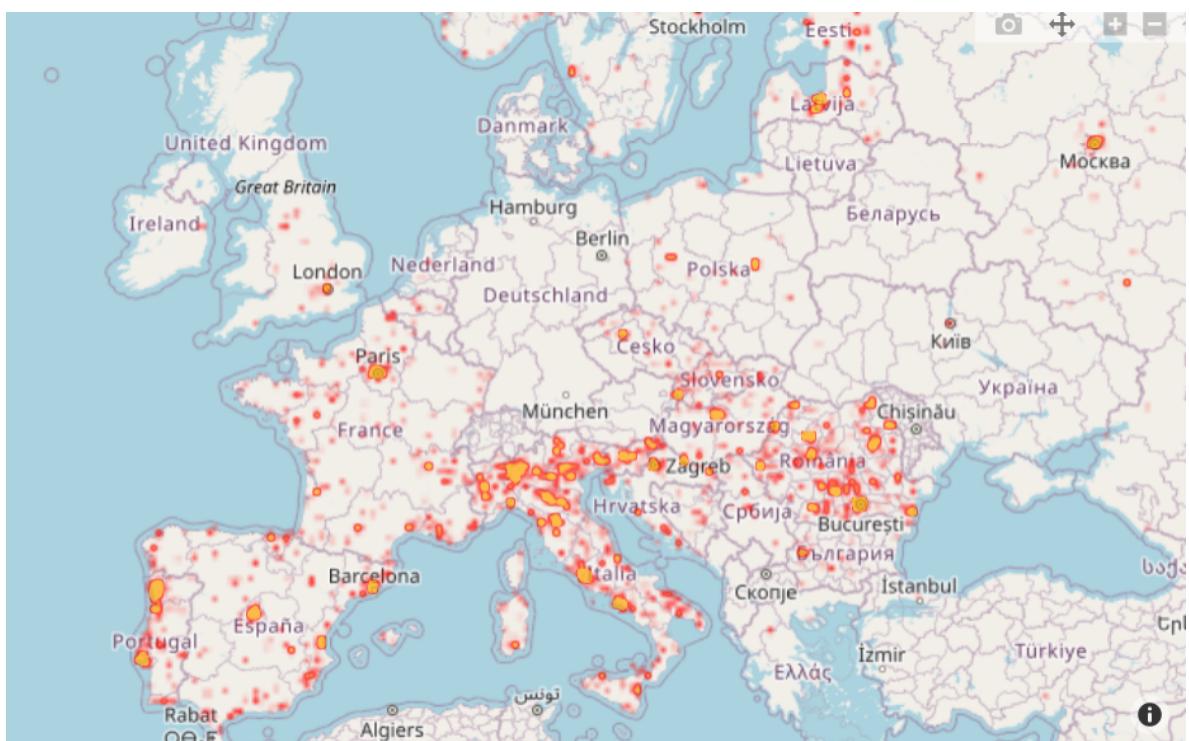
lat / lon

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.

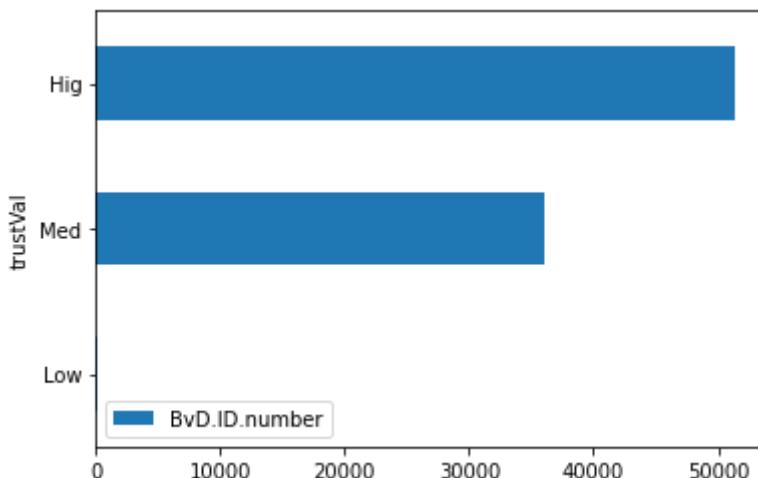
- A class {Low, Medium, High}
- A firm name, maybe the evaluator?

```
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
```

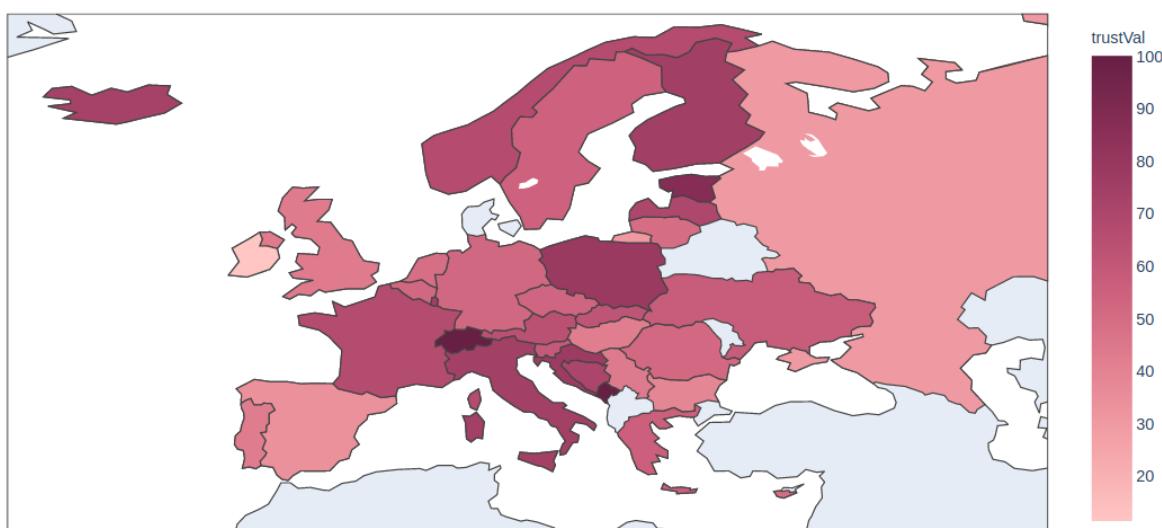
trustVal

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

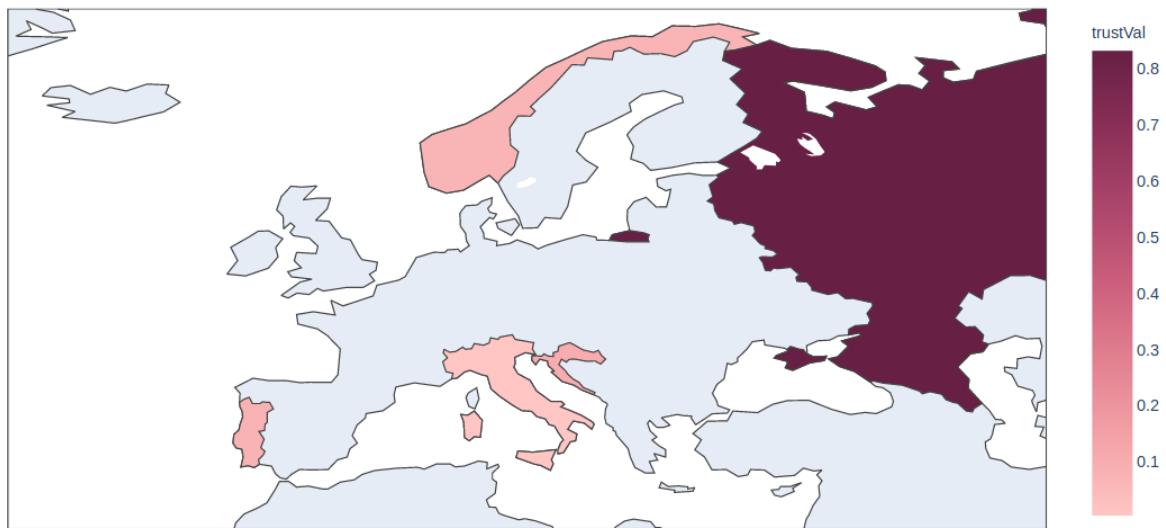
```
BvD.ID.number
IT04844800658 Hig
SE5568281116 Med
IT07016280963 Med
GB07234353 NaN
FR519497507 Hig
Name: trustVal, dtype: category
Categories (3, object): [Low < Med < Hig]
```



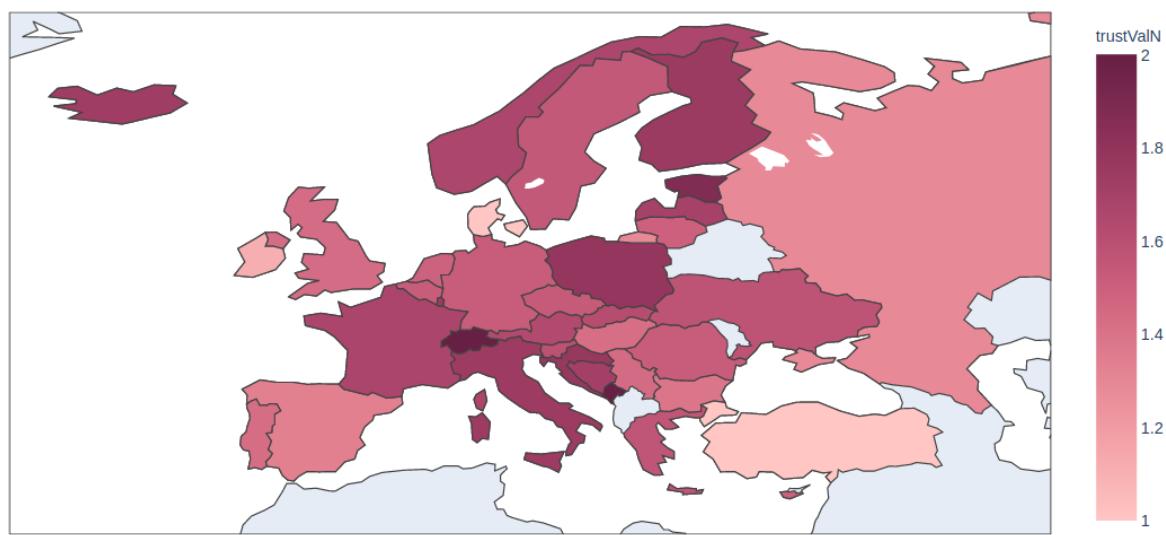
Percentage of companies with "High" rating:



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



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



trustVal	Low	Med	Hig
HGF			
False	77	34741	49523
True	1	1330	1873
 Chi ² = 1.36			
p = 0.51			
degrees of freedom = 2			