test

November 3, 2024

1 Analyse de données d'un VSM

creation d'environement virtuel et activation de l'environement

[2]:
```
!python -m venv env
```

[3]:
```
!env\Scripts\activate
```

1.0.1 Installation des bibliothèques

[ ]:
```
!pip install pandas
!pip install seaborn
!pip install matplotlib.pyplot
```

[5]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import csv
```

2 Netoyage de données

[7]:
```
import pandas as pd

# Charger le fichier CSV
df = pd.read_csv('data.csv', sep=',', encoding='ISO-8859-1')

df.replace(['-', '', 'NaN'], pd.NA, inplace=True)

# Vérifier le nombre de valeurs manquantes après le remplacement
print(df.isnull().sum())

# Supprimer les lignes avec des NaN dans des colonnes importantes
df_cleaned = df.dropna(subset=['Country', 'Continent'])

# Vérifier le nombre de lignes après le nettoyage
print("Nombre de lignes après le nettoyage :", len(df_cleaned))
```

```
df_cleaned.to_csv('data_cleaned.csv', index=False)
```

```
Paper ID                                                                0
Manuscript Title                                                        0
Method used for optimization of the supply chain                        0
VSM  used alone or in combination with other optimization methods       0
Main benefits                                                           0
                                                                       ..
Unnamed: 58                                                           154
Unnamed: 59                                                           154
Unnamed: 60                                                           154
Unnamed: 61                                                           154
 .5                                                                   154
Length: 63, dtype: int64
Nombre de lignes après le nettoyage : 105
```

[8]:
```
df = pd.read_csv('data_cleaned.csv', sep=',',encoding='ISO-8859-1')

number_paper_by_continent=df['Continent'].value_counts()
number_paper_by_continent_df = number_paper_by_continent.reset_index()
number_paper_by_continent_df.columns = ['Continent', 'Count']
print(number_paper_by_continent_df)
america_count = number_paper_by_continent_df.
 ↪loc[number_paper_by_continent_df['Continent'].isin(['North America', 'South␣
 ↪America']), 'Count'].sum()

number_paper_by_continent_df = number_paper_by_continent_df[
    ~number_paper_by_continent_df['Continent'].isin(['North America', 'South␣
 ↪America'])
]

# Créer une nouvelle ligne pour "America"
new_row = pd.DataFrame({'Continent': ['America'], 'Count': [america_count]})

# Utiliser concat pour ajouter la nouvelle ligne
number_paper_by_continent_df = pd.concat([number_paper_by_continent_df,␣
 ↪new_row], ignore_index=True)
print(number_paper_by_continent_df)

plt.figure(figsize=(8, 6))
sns.barplot(x='Continent', y='Count', data=number_paper_by_continent_df,␣
 ↪palette='viridis')
plt.title('Nombre de papiers par continent')
plt.xlabel('')
plt.ylabel('Valeur')
plt.show()
```

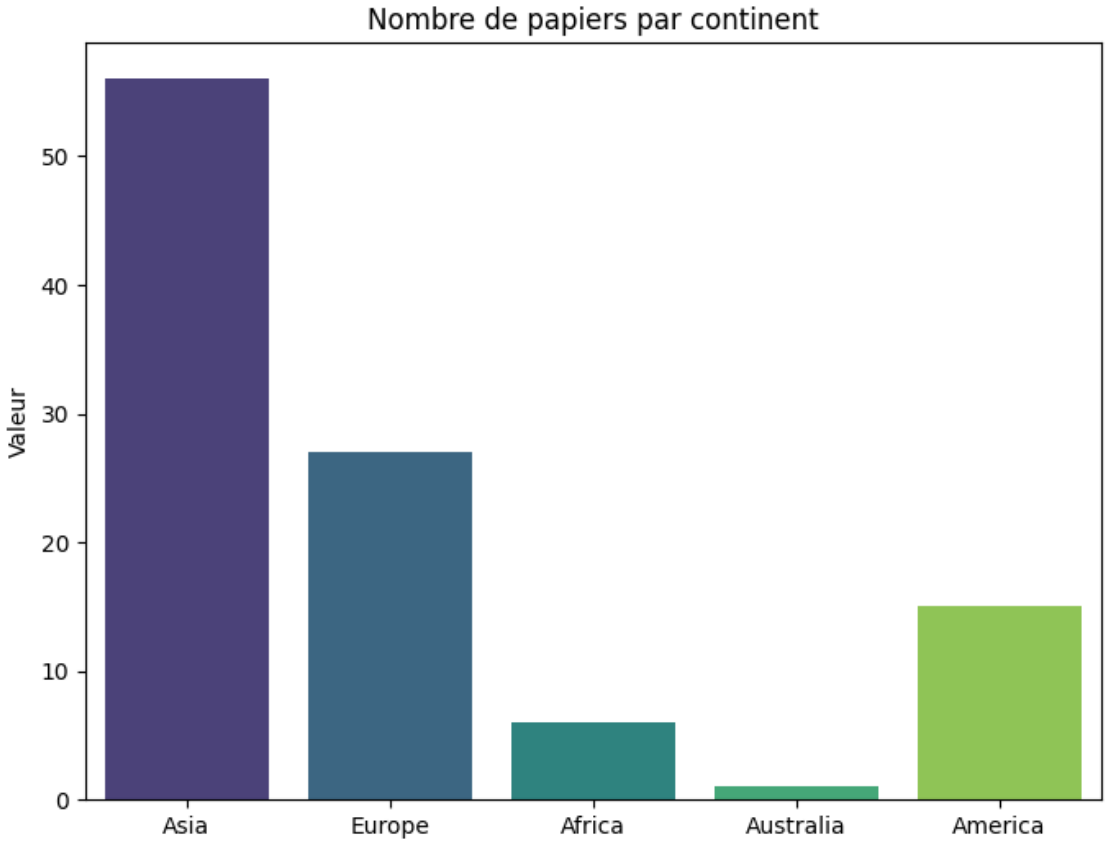
```
       Continent  Count
```

```
0           Asia     56
1         Europe     27
2  South America     12
3         Africa      6
4  North America      3
5      Australia      1
   Continent  Count
0       Asia     56
1     Europe     27
2     Africa      6
3  Australia      1
4    America     15
```

```
C:\Users\KELI\AppData\Local\Temp\ipykernel_10156\3014333849.py:21:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(x='Continent', y='Count', data=number_paper_by_continent_df,
palette='viridis')
```



3 Par Pays

[28]:
```
df = pd.read_csv('data_cleaned.csv', sep=',',encoding='ISO-8859-1')

number_paper_by_continent=df['Country'].value_counts()
number_paper_by_continent_df = number_paper_by_continent.reset_index()
number_paper_by_continent_df.columns = ['Country', 'Count']
print(number_paper_by_continent_df)


plt.figure(figsize=(14, 12))
sns.barplot(y='Country', x='Count', data=number_paper_by_continent_df,␣
 ↪palette='viridis')
plt.title('Nombre de papiers par continent')
plt.xlabel('Catégorie')
plt.ylabel('Valeur')
plt.show()
```

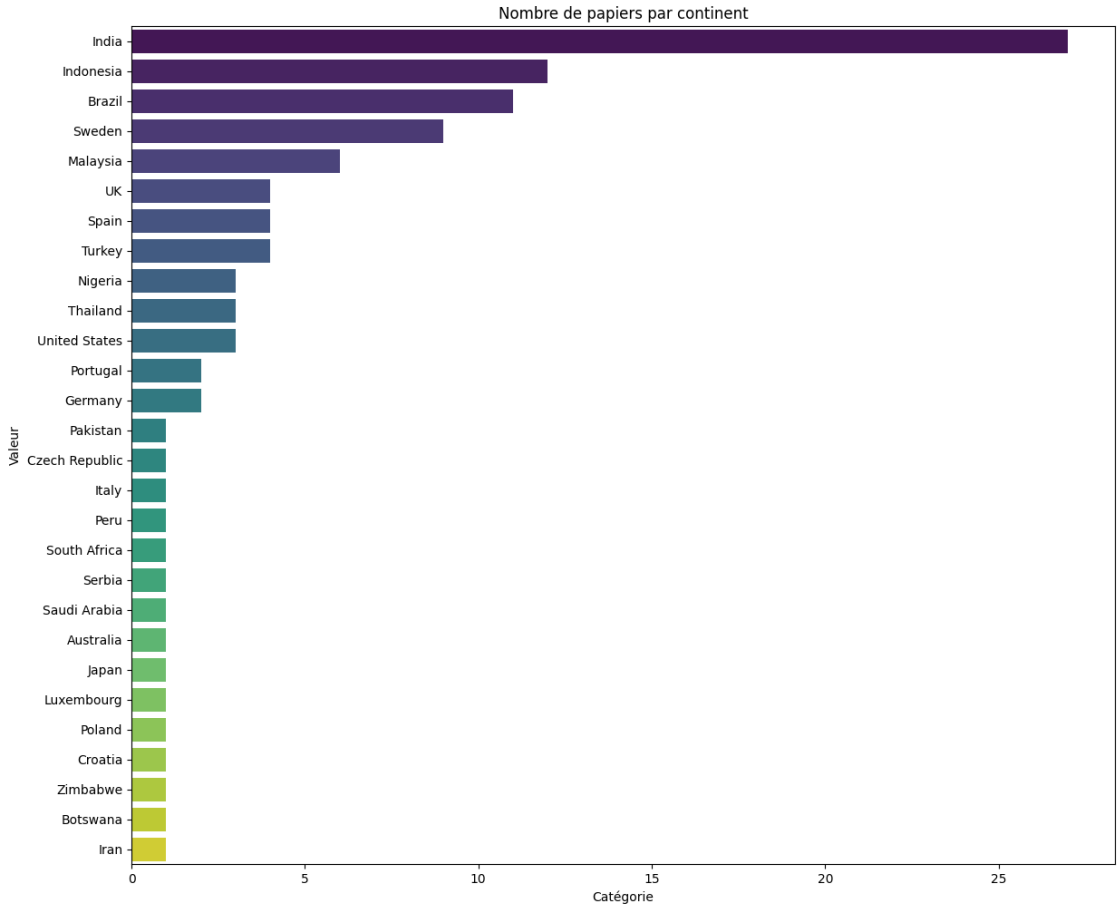
```
           Country  Count
0            India     27
1        Indonesia     12
2           Brazil     11
3           Sweden      9
4         Malaysia      6
5               UK      4
6            Spain      4
7           Turkey      4
8          Nigeria      3
9         Thailand      3
10   United States      3
11        Portugal      2
12         Germany      2
13        Pakistan      1
14  Czech Republic      1
15           Italy      1
16            Peru      1
17    South Africa      1
18          Serbia      1
19    Saudi Arabia      1
20       Australia      1
21           Japan      1
22      Luxembourg      1
23          Poland      1
24         Croatia      1
25        Zimbabwe      1
```

```
26        Botswana      1
27            Iran      1
```

```
C:\Users\KELI\AppData\Local\Temp\ipykernel_10156\3926689835.py:10:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(y='Country', x='Count', data=number_paper_by_continent_df,
palette='viridis')
```



4 Compagny Sector

[19]:
```
df = pd.read_csv('data_cleaned.csv', sep=',',encoding='ISO-8859-1')
```

```
print(df.columns)

number_paper_by_continent=df[' Company sector '].value_counts()
number_paper_by_continent_df = number_paper_by_continent.reset_index()
number_paper_by_continent_df.columns = [' Company sector ', 'Count']
print(number_paper_by_continent_df)


plt.figure(figsize=(20, 25))
sns.barplot(y=' Company sector ', x='Count', data=number_paper_by_continent_df,␣
 ↪palette='viridis')
plt.title('Nombre de papiers par continent')
plt.xlabel('Catégorie')
plt.ylabel('Valeur')
plt.tight_layout()
plt.show()
```

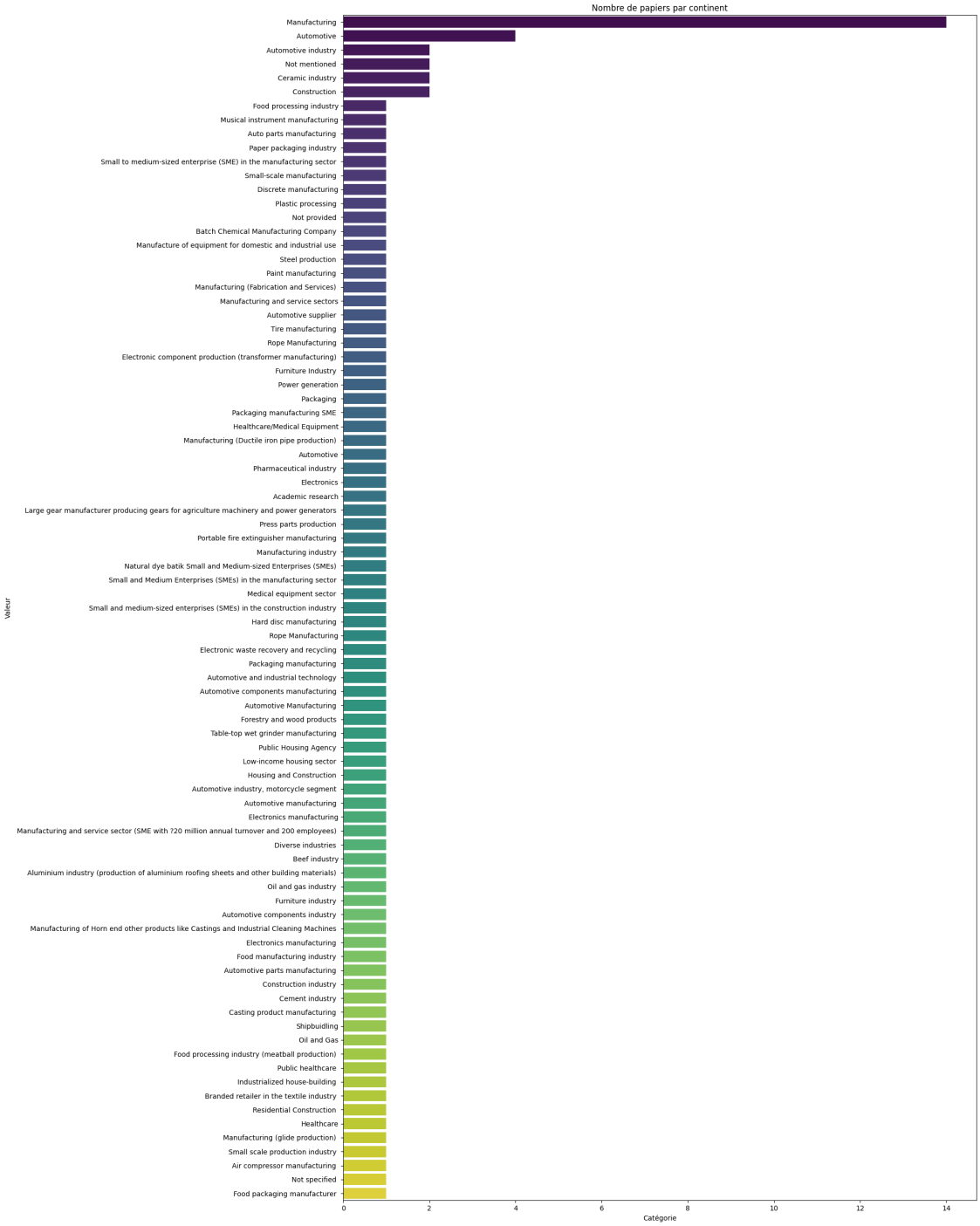
```
Index([' Paper ID ', ' Manuscript Title ',
       ' Method used for optimization of the supply chain ',
       ' VSM  used alone or in combination with other optimization methods ',
       ' Main benefits ', ' Place where the study was implemented ', 'Country',
       'Continent', ' Company sector ', 'Sector ', ' Year of publication ',
       ' Improvement rate for the KPIs ', ' ', 'Unnamed: 13', 'Unnamed: 14',
       'Unnamed: 15', 'Unnamed: 16', 'Unnamed: 17', 'Unnamed: 18',
       'Unnamed: 19', 'Unnamed: 20', 'Unnamed: 21', ' .1', 'Unnamed: 23',
       'Unnamed: 24', 'Unnamed: 25', 'Unnamed: 26', 'Unnamed: 27',
       'Unnamed: 28', 'Unnamed: 29', 'Unnamed: 30', 'Unnamed: 31', ' .2',
       'Unnamed: 33', 'Unnamed: 34', 'Unnamed: 35', 'Unnamed: 36',
       'Unnamed: 37', 'Unnamed: 38', 'Unnamed: 39', 'Unnamed: 40',
       'Unnamed: 41', ' .3', 'Unnamed: 43', 'Unnamed: 44', 'Unnamed: 45',
       'Unnamed: 46', 'Unnamed: 47', 'Unnamed: 48', 'Unnamed: 49',
       'Unnamed: 50', 'Unnamed: 51', ' .4', 'Unnamed: 53', 'Unnamed: 54',
       'Unnamed: 55', 'Unnamed: 56', 'Unnamed: 57', 'Unnamed: 58',
       'Unnamed: 59', 'Unnamed: 60', 'Unnamed: 61', ' .5'],
      dtype='object')
                        Company sector   Count
0                        Manufacturing      14
1                           Automotive       4
2                  Automotive industry       2
3                        Not mentioned       2
4                     Ceramic industry       2
..                                  …    …
80   Manufacturing (glide production)        1
81    Small scale production industry        1
82       Air compressor manufacturing        1
83                       Not specified       1
84        Food packaging manufacturer        1
```

```
[85 rows x 2 columns]
```

```
C:\Users\KELI\AppData\Local\Temp\ipykernel_10156\4008687539.py:13:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(y=' Company sector ', x='Count',
data=number_paper_by_continent_df, palette='viridis')
```


Nombre de papiers par continent
Manufacturing
Automotive
Automotive industry
Not mentioned
Ceramic industry
Construction
Food processing industry
Musical instrument manufacturing
Auto parts manufacturing
Paper packaging industry
Small to medium-sized enterprise (SME) in the manufacturing sector
Small-scale manufacturing
Discrete manufacturing
Plastic processing
Not provided
Batch Chemical Manufacturing Company
Manufacture of equipment for domestic and industrial use
Steel production
Paint manufacturing
Manufacturing (Fabrication and Services)
Manufacturing and service sectors
Automotive supplier
Tire manufacturing
Rope Manufacturing
Electronic component production (transformer manufacturing)
Furniture Industry
Power generation
Packaging
Packaging manufacturing SME
Healthcare/Medical Equipment
Manufacturing (Ductile iron pipe production)
Automotive
Pharmaceutical industry
Electronics
Academic research
Large gear manufacturer producing gears for agriculture machinery and power generators
Press parts production
Portable fire extinguisher manufacturing
Manufacturing industry
Natural dye batik Small and Medium-sized Enterprises (SMEs)
Small and Medium Enterprises (SMEs) in the manufacturing sector
Medical equipment sector
Small and medium-sized enterprises (SMEs) in the construction industry
Hard disc manufacturing
Rope Manufacturing
Electronic waste recovery and recycling
Packaging manufacturing
Automotive and industrial technology
Automotive components manufacturing
Automotive Manufacturing
Forestry and wood products
Table-top wet grinder manufacturing
Public Housing Agency
Low-income housing sector
Housing and Construction
Automotive industry, motorcycle segment
Automotive manufacturing
Electronics manufacturing
Manufacturing and service sector (SME with ?20 million annual turnover and 200 employees)
Diverse industries
Beef industry
Aluminium industry (production of aluminium roofing sheets and other building materials)
Oil and gas industry
Furniture industry
Automotive components industry
Manufacturing of Horn end other products like Castings and Industrial Cleaning Machines
Electronics manufacturing
Food manufacturing industry
Automotive parts manufacturing
Construction industry
Cement industry
Casting product manufacturing
Shipbuidling
Oil and Gas
Food processing industry (meatball production)
Public healthcare
Industrialized house-building
Branded retailer in the textile industry
Residential Construction
Healthcare
Manufacturing (glide production)
Small scale production industry
Air compressor manufacturing
Not specified
Food packaging manufacturer
Valeur
0
2
4
6
8
10
12
14
Catégorie

5 par produit

[20]:
```
df = pd.read_csv('data_cleaned.csv', sep=',',encoding='ISO-8859-1')


print(df.columns)

number_paper_by_continent=df['Sector '].value_counts()
number_paper_by_continent_df = number_paper_by_continent.reset_index()
number_paper_by_continent_df.columns = ['Sector ', 'Count']
print(number_paper_by_continent_df)


plt.figure(figsize=(20, 29))
sns.barplot(y='Sector ', x='Count', data=number_paper_by_continent_df,␣
 ↪palette='viridis')
plt.title('Nombre de papiers par continent')
plt.xlabel('Catégorie')
plt.ylabel('Valeur')
plt.show()
```

```
Index([' Paper ID ', ' Manuscript Title ',
       ' Method used for optimization of the supply chain ',
       ' VSM  used alone or in combination with other optimization methods ',
       ' Main benefits ', ' Place where the study was implemented ', 'Country',
       'Continent', ' Company sector ', 'Sector ', ' Year of publication ',
       ' Improvement rate for the KPIs ', ' ', 'Unnamed: 13', 'Unnamed: 14',
       'Unnamed: 15', 'Unnamed: 16', 'Unnamed: 17', 'Unnamed: 18',
       'Unnamed: 19', 'Unnamed: 20', 'Unnamed: 21', ' .1', 'Unnamed: 23',
       'Unnamed: 24', 'Unnamed: 25', 'Unnamed: 26', 'Unnamed: 27',
       'Unnamed: 28', 'Unnamed: 29', 'Unnamed: 30', 'Unnamed: 31', ' .2',
       'Unnamed: 33', 'Unnamed: 34', 'Unnamed: 35', 'Unnamed: 36',
       'Unnamed: 37', 'Unnamed: 38', 'Unnamed: 39', 'Unnamed: 40',
       'Unnamed: 41', ' .3', 'Unnamed: 43', 'Unnamed: 44', 'Unnamed: 45',
       'Unnamed: 46', 'Unnamed: 47', 'Unnamed: 48', 'Unnamed: 49',
       'Unnamed: 50', 'Unnamed: 51', ' .4', 'Unnamed: 53', 'Unnamed: 54',
       'Unnamed: 55', 'Unnamed: 56', 'Unnamed: 57', 'Unnamed: 58',
       'Unnamed: 59', 'Unnamed: 60', 'Unnamed: 61', ' .5'],
      dtype='object')
                 Sector   Count
0             Automotive     11
1          Manufacturing      5
2            Automotive       4
3           Construction      4
4            Electronics      3
..                     …    …
66      Casting products      1
67                Cement      1
```

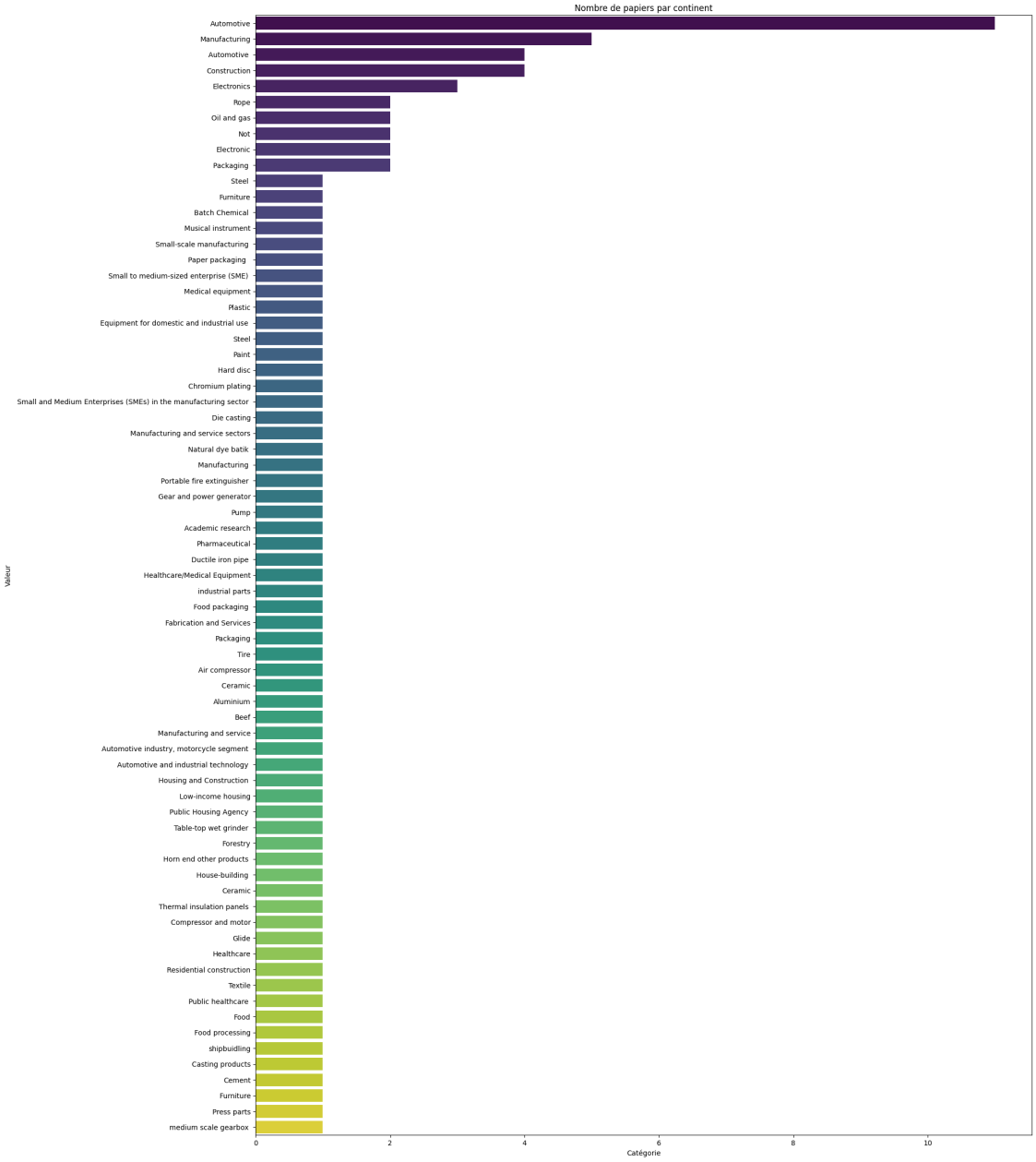
```
68              Furniture      1
69           Press parts       1
70  medium scale gearbox       1

[71 rows x 2 columns]
```

```
C:\Users\KELI\AppData\Local\Temp\ipykernel_10156\3971502843.py:13:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(y='Sector ', x='Count', data=number_paper_by_continent_df,
palette='viridis')
```



6 year publication

[25]:
```
df = pd.read_csv('data_cleaned.csv', sep=',',encoding='ISO-8859-1')


print(df.columns)

number_paper_by_year_publication=df[' Year of publication '].value_counts()
```

```
number_paper_by_year_publication_df = number_paper_by_year_publication.
 ↪reset_index()
number_paper_by_year_publication_df.columns = [' Year of publication ', 'Count']
print(number_paper_by_year_publication_df)


plt.figure(figsize=(18, 10))
sns.barplot(x=' Year of publication ', y='Count',␣
 ↪data=number_paper_by_year_publication_df, palette='viridis')
plt.title('Nombre de papiers par continent')
plt.xlabel('Catégorie')
plt.ylabel('Valeur')
plt.show()
```

```
Index([' Paper ID ', ' Manuscript Title ',
       ' Method used for optimization of the supply chain ',
       ' VSM  used alone or in combination with other optimization methods ',
       ' Main benefits ', ' Place where the study was implemented ', 'Country',
       'Continent', ' Company sector ', 'Sector ', ' Year of publication ',
       ' Improvement rate for the KPIs ', ' ', 'Unnamed: 13', 'Unnamed: 14',
       'Unnamed: 15', 'Unnamed: 16', 'Unnamed: 17', 'Unnamed: 18',
       'Unnamed: 19', 'Unnamed: 20', 'Unnamed: 21', ' .1', 'Unnamed: 23',
       'Unnamed: 24', 'Unnamed: 25', 'Unnamed: 26', 'Unnamed: 27',
       'Unnamed: 28', 'Unnamed: 29', 'Unnamed: 30', 'Unnamed: 31', ' .2',
       'Unnamed: 33', 'Unnamed: 34', 'Unnamed: 35', 'Unnamed: 36',
       'Unnamed: 37', 'Unnamed: 38', 'Unnamed: 39', 'Unnamed: 40',
       'Unnamed: 41', ' .3', 'Unnamed: 43', 'Unnamed: 44', 'Unnamed: 45',
       'Unnamed: 46', 'Unnamed: 47', 'Unnamed: 48', 'Unnamed: 49',
       'Unnamed: 50', 'Unnamed: 51', ' .4', 'Unnamed: 53', 'Unnamed: 54',
       'Unnamed: 55', 'Unnamed: 56', 'Unnamed: 57', 'Unnamed: 58',
       'Unnamed: 59', 'Unnamed: 60', 'Unnamed: 61', ' .5'],
      dtype='object')
    Year of publication   Count
0                  2017     19
1                  2015     15
2                  2018     10
3                  2016      9
4                  2019      8
5                  2014      7
6                  2021      5
7                  2022      4
8                  2023      4
9                  2008      4
10                 2011      4
11                 2007      3
12                 2012      3
13                 2013      3
```

```
14                    2020       3
15                    2006       2
16                    2005       1
17                    2009       1
```

```
C:\Users\KELI\AppData\Local\Temp\ipykernel_10156\103212372.py:13: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(x=' Year of publication ', y='Count',
data=number_paper_by_year_publication_df, palette='viridis')
```