



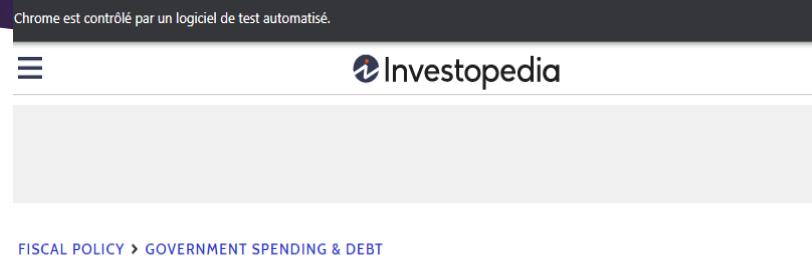
INFRASTRUCTURE AND INDUSTRY

THIS RESEARCH IS PRESENTED TO **DR.ZIAD ABDALLAH**
STUDENT: OMAR ABDELKADER
MAJOR: DATA SCIENCE
COURSE NAME: DATA SCRAPING

INFRASTRUCTURE



Infrastructure using Selenium webdriver (Implicit wait)



Infrastructure

By INVESTOPEDIA STAFF | Reviewed by MICHAEL J BOYLE | Updated Mar 25, 2021

What Is Infrastructure?

Infrastructure is the general term for the basic physical systems of a business, region, or nation. Examples of infrastructure include transportation systems, communication networks, sewage, water, and electric systems. These systems tend to be capital intensive and high-cost investments, and are vital to a country's economic development and prosperity.

Projects related to infrastructure improvements may be funded publicly, privately, or through [public-private partnerships](#). In economic terms, infrastructure often involves the production of [public goods](#) or production processes that support [natural monopolies](#).

KEY TAKEAWAYS

- Infrastructure are the basic systems that undergird the structure of the economy.

We use selenium to open a virtual browser

```
#Implicit Waits in Selenium Python
# import webdriver
from selenium import webdriver

driver = webdriver.Chrome()

# set implicit wait time
driver.implicitly_wait(10) # seconds

# get url
driver.get("https://www.investopedia.com/terms/i/infrastructure.asp")
```

Example of Regular Expression search

```
import re
from bs4 import BeautifulSoup
import requests
url = "https://www.investopedia.com/terms/i/infrastructure.asp"
r = requests.get(url)
html_content = r.text
x = re.findall("Infrastructure", html_content)
len(x)
```

28

We use regular expression to find word « infrastructure »

What's is Infrastructure using Beautiful Soup

```
from bs4 import BeautifulSoup
import requests
url="https://www.investopedia.com/terms/i/infrastructure.asp"
r = requests.get(url)
html_content=r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mntl-sc-block_1-0-1")
y = html_soup.find(id="mntl-sc-block_1-0-3")
print(x.text)
print(y.text)
```

Definition of
Infrastructure

Infrastructure is the general term for the basic physical systems of a business, region, or nation. Examples of infrastructure include transportation systems, communication networks, sewage, water, and electric systems. These systems tend to be capital intensive and high-cost investments, and are vital to a country's economic development and prosperity.

Projects related to infrastructure improvements may be funded publicly, privately, or through public-private partnerships. In economic terms, infrastructure often involves the production of public goods or production processes that support natural monopolies.

Understanding Infrastructure using BeautifulSoup

```
from bs4 import BeautifulSoup
import requests
url="https://www.investopedia.com/terms/i/infrastructure.asp"
r = requests.get(url)
html_content=r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mntl-sc-block_1-0-10")
y = html_soup.find(id="mntl-sc-block_1-0-12")
z = html_soup.find(id="mntl-sc-block_1-0-14")
print(x.text)
print(y.text)
print(z.text)
```

The term infrastructure first appeared in usage in the late 1880s. The word comes from French, with infra- meaning "below" and structure meaning "building." Infrastructure is the foundation upon which the structure of the economy is built, often times quite literally. In 1987, a panel of the U.S. National Research Council adopted the term "public works infrastructure" to refer to functional modes including highways, airports, telecommunications, and water supplies, as well as the combined systems that these elements comprise.

Applicable to large- and small-scale organizational frameworks, infrastructure can include a variety of systems and structures as long as there are physical components required. For example, the electrical grid across a city, state or country is infrastructure based on the equipment involved and the intent to provide a service to the areas it supports. Similarly, the physical cabling and components making up the data network of a company operating within a specific location are also the infrastructure for the business in question, as they are necessary to support business operations.

Because infrastructure very often involves the production of either public goods or goods that lend themselves to production by natural monopolies, it is very typical to see public financing, control, supervision, or regulation of infrastructure. This usually takes the form of direct government production or production by a closely regulated, legally sanctioned, and often subsidized monopoly. At smaller scales, infrastructure can also often take on the characteristics of club goods or goods most readily produced by localized monopolies, and can be provided within the context of a private firm producing infrastructure for use within the firm or provided by localized arrangements of formal or informal collective action.

Extracting a
paragraphe

Types of Infrastructure

using Beautiful Soup

```
from bs4 import BeautifulSoup
import requests
url = "https://www.investopedia.com/terms/i/infrastructure.asp"
r = requests.get(url)
html_content = r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mntl-sc-block_1-0-19")
y = html_soup.find(id="mntl-sc-block_1-0-22")
z = html_soup.find(id="mntl-sc-block_1-0-25")

def softinfrastructure():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-20")
    print(x.text)

def hardinfrastructure():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-23")
    print(x.text)

def criticalinfrasctructre1():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-26")
    print(x.text)

def criticalinfrasctructre2():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-28")
    print(x.text)

print(x.text)
softinfrastructure()
print(y.text)
hardinfrastructure()
print(z.text)
criticalinfrasctructre1()
criticalinfrasctructre2()
```

Soft Infrastructure

These types of infrastructure make up institutions that help maintain the economy. These usually require human capital and help deliver certain services to the population. Examples include the healthcare system, financial institutions, governmental systems, law enforcement, and education systems.

Hard Infrastructure

These make up the physical systems that make it necessary to run a modern, industrialized nation. Examples include roads, highways, bridges, as well as the capital/assets needed to make them operational (transit buses, vehicles, oil rigs/refineries).

Critical Infrastructure

These are assets defined by a government as being essential to the functioning of a society and economy, such as facilities for shelter and heating, telecommunication, public health, agriculture, etc. In the United States, there are agencies responsible for these critical infrastructures, such as Homeland Security (for the government and emergency services), the Department of Energy, and the Department of Transportation.

Along with the aforementioned sectors, infrastructure includes waste disposal services, such as garbage pickup and local dumps. Certain administrative functions, often covered by various government agencies, are also considered part of the infrastructure. Educational and healthcare facilities may also be included, along with specific research and development functions and necessary training facilities.

Special Considerations

using BeautifulSoup

```
from bs4 import BeautifulSoup
import requests

url = "https://www.investopedia.com/terms/i/infrastructure.asp"
r = requests.get(url)
html_content = r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mntl-sc-block_1-0-31")
y = html_soup.find(id="mntl-sc-block_1-0-34")
z = html_soup.find(id="mntl-sc-block_1-0-41")

def itinfrastructure():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-32")
    print(x.text)

def privateinvestmentinpublicinfrastructure1():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-35")
    print(x.text)

def privateinvestmentinpublicinfrastructure2():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-37")
    print(x.text)

def privateinvestmentinpublicinfrastructure3():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-39")
    print(x.text)

def infrastructureasanassetclass():
    url = "https://www.investopedia.com/terms/i/infrastructure.asp"
    r = requests.get(url)
    html_content = r.text
    html_soup = BeautifulSoup(html_content)
    x = html_soup.find(id="mntl-sc-block_1-0-42")
    print(x.text)

print(x.text)
itinfrastructure()
print(y.text)
privateinvestmentinpublicinfrastructure1()
privateinvestmentinpublicinfrastructure2()
privateinvestmentinpublicinfrastructure3()
print(z.text)
infrastructureasanassetclass()
```

IT Infrastructure

Many technical systems are often referred to as infrastructures, such as networking equipment and servers, due to the critical function they provide within specific business environments. Without the information technology (IT) infrastructure, many businesses struggle to share and move data in a way that promotes efficiency within the workplace. If IT infrastructure fails, many business functions cannot be performed.

Private Investment in Public Infrastructure

Sometimes private companies choose to invest in a country's infrastructure development as part of a business expansion effort. For example, an energy company may build pipelines and railways in a country where it wants to refine petroleum. This investment can benefit both the company and the country.

In 2005, Skyway Concession Company entered into a 99-year lease with the City of Chicago to operate and maintain the Chicago Skyway Bridge. As part of the agreement, Skyway receives all toll and concession revenue generated by the bridge, while the city benefited from a \$1.83 billion cash infusion and is no longer responsible for maintaining the bridge.

Individuals may also choose to fund improvements to certain pieces of public infrastructure. For example, an individual may fund improvements to hospitals, schools, or local law enforcement efforts.

Infrastructure as an Asset Class

Infrastructure is also an asset class that tends to be less volatile than equities over the long term and provides a higher yield. As a result, some companies and individuals like to invest in infrastructure funds for their defensive characteristics, such as funds involved in transportation or water infrastructure.



INDUSRTY



Industry using Selenium webdriver (Explicit Wait)

Chrome est contrôlé par un logiciel de test automatisé.



BUSINESS > BUSINESS ESSENTIALS

Industry

By WILL KENTON | Reviewed by JULIUS MANSA | Updated Apr 1, 2021

What Is an Industry?

An [industry](#) is a group of companies that are related based on their primary [business activities](#). In modern economies, there are dozens of industry classifications. Industry classifications are typically grouped into larger categories called [sectors](#).

Individual companies are generally classified into an industry based on their largest sources of revenue. For example, while an automobile manufacturer might have a financing division that contributes 10% to the firm's overall [revenues](#), the company would be classified in the automaker industry by most classification systems.

KEY TAKEAWAYS

- Similar companies are grouped together into industries, and there are a number of different industries, such as department stores and shoemakers.
- Industry grouping is based on the primary product that a

We use selenium to open a virtual browser

```
# import necessary classes
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC

# create driver object
driver = webdriver.Chrome()

# A URL that delays loading
driver.get("https://www.investopedia.com/terms/i/industry.asp")

try:
    # wait 10 seconds before looking for element
    element = WebDriverWait(driver, 10).until(
        EC.presence_of_element_located((By.ID, "myDynamicElement"))
)
finally:
    # else quit
    driver.quit()
```

Example of Regular Expression search

```
import re
from bs4 import BeautifulSoup
import requests
url = "https://www.investopedia.com/terms/i/industry.asp"
r = requests.get(url)
html_content = r.text
x = re.search("industry", html_content)
print(x)

<re.Match object; span=(977, 985), match='industry'>
```

We use regular expression to find word « industry »

What is an Industry using Beautiful Soup

```
from bs4 import BeautifulSoup
import requests
url = "https://www.investopedia.com/terms/i/industry.asp"
r = requests.get(url)
html_content = r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mntl-sc-block_1-0-1")
y = html_soup.find(id="mntl-sc-block_1-0-3")

print(x.text)
print(y.text)
```

An industry is a group of companies that are related based on their primary business activities. In modern economies, there are dozens of industry classifications. Industry classifications are typically grouped into larger categories called sectors.

Individual companies are generally classified into an industry based on their largest sources of revenue. For example, while an automobile manufacturer might have a financing division that contributes 10% to the firm's overall revenues, the company would be classified in the automaker industry by most classification systems.

We use regular expression to find word « industry »

Understanding an Industry

using Beautiful Soup

```
from bs4 import BeautifulSoup
import requests
url = "https://www.investopedia.com/terms/i/industry.asp"
r = requests.get(url)
html_content = r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mntl-sc-block_1-0-10")
print(x.text)
```

Similar businesses are grouped into industries based on the primary product produced or sold. This effectively creates industry groups, which can then be used to isolate businesses from those who participate in different activities. Investors and economists often study industries to better understand the factors and limitations of corporate profit growth. Companies operating in the same industry can also be compared to each other to evaluate the relative attractiveness of a company within that industry.

Special Considerations using BeautifulSoup

```
from bs4 import BeautifulSoup
import requests
url = "https://www.investopedia.com/terms/i/industry.asp"
r = requests.get(url)
html_content = r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mntl-sc-block_1-0-15")
y = html_soup.find(id="mntl-sc-block_1-0-17")

print(x.text)
print(y.text)
```

Stocks within the same industry often rise and fall as a group because the same macroeconomic factors impact all members of an industry. These macroeconomic factors can include changes in market sentiment on the part of investors—such as those based on a response to a particular event or piece of news—as well as changes directed specifically towards the specific industry, such as new regulations or increased raw material costs.

However, events relating to just one particular business can cause the associated stock to rise or fall separately from others within the same industry. This can be the result of certain events, including a differentiating product release, a corporate scandal in the news, or a change in leadership structures.

Industries vs. Sectors using Beautiful Soup

```
from bs4 import BeautifulSoup
import requests
url = "https://www.investopedia.com/terms/i/industry.asp"
r = requests.get(url)
html_content = r.text
html_soup = BeautifulSoup(html_content)
x = html_soup.find(id="mnl-sc-block_1-0-20")
y = html_soup.find(id="mnl-sc-block_1-0-22")
z = html_soup.find(id="mnl-sc-block_1-0-24")

print(x.text)
print(y.text)
print(z.text)
```

While both sectors and industries are classification systems used to group similar types of business operations, sectors are broader than industries.

For example, retail trade is a sector within the North American Industry Classification System (NAICS), and within that sector are industries, such as health and personal care stores, clothing stores, and shoe stores. Rite Aid Corporation and Gap, Inc. are members of the same consumer goods sector, but each would be listed in a different industry based on the specifics of the products they produce or sell. Rite Aid Corporation is classified within the health and personal care stores (NAICS Code 4461), while Gap, Inc. is classified within the clothing stores industry (NAICS Code 4481).

The North American Industry Classification System (NAICS), developed by the United States, Canada, and Mexico, is the standard upon which government agencies classify businesses when compiling statistical data. In the NAICS hierarchy, companies that use similar production processes are categorized in the same industry.

Comparison
between
infrastructure
and Sectors

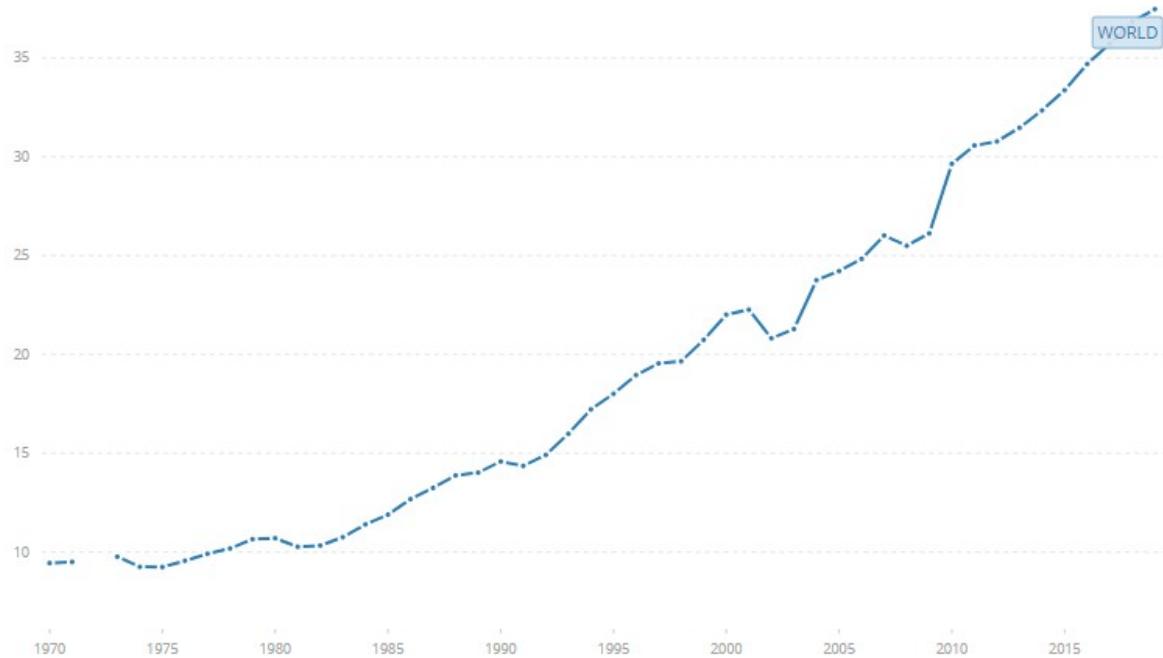
Industry Data Set using ParseHub

We use ParseHub to scrape all data and save it as csv file

Visualization **using Boket**

Data Set 1 Visualization

using Boket



Data Set 2 Visualization using Boket

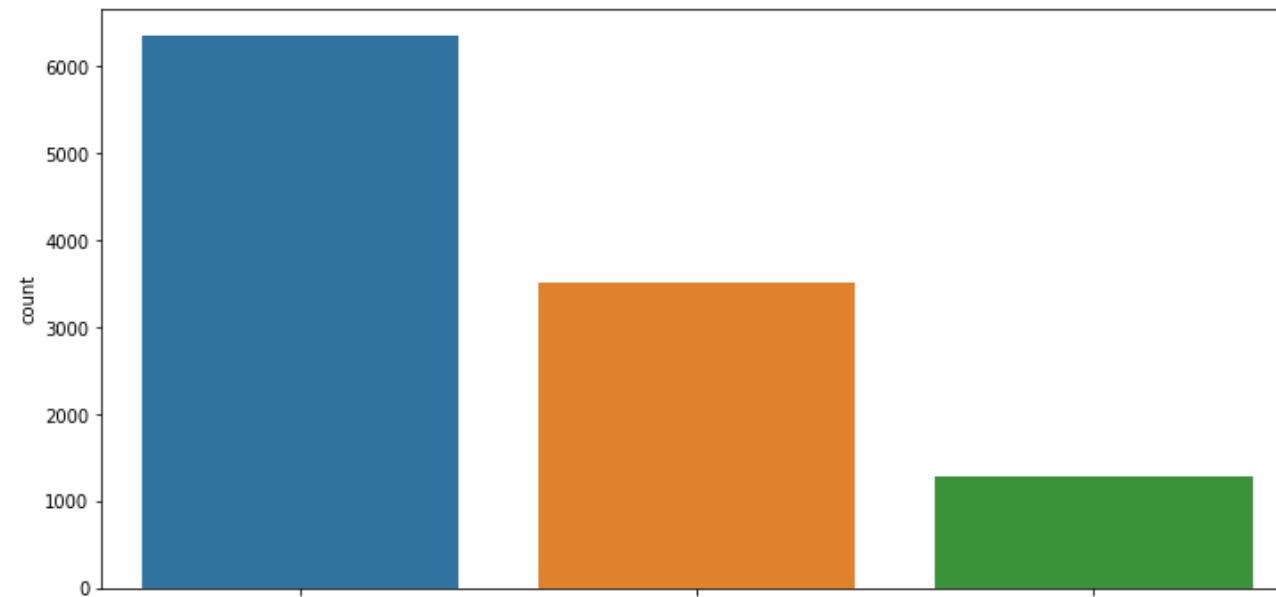
```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.gridspec import GridSpec
import seaborn as sns
import pandas as pd
base = pd.read_csv('data.csv')
base.head(10)
```

0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0	unknown	yes
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0	unknown	yes
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0	unknown	yes
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0	unknown	yes
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0	unknown	yes
5	42	management	single	tertiary	no	0	yes	yes	unknown	5	may	562	2	-1	0	unknown	yes
6	56	management	married	tertiary	no	830	yes	yes	unknown	6	may	1201	1	-1	0	unknown	yes
7	60	retired	divorced	secondary	no	545	yes	no	unknown	6	may	1030	1	-1	0	unknown	yes
8	37	technician	married	secondary	no	1	yes	no	unknown	6	may	608	1	-1	0	unknown	yes
9	28	services	single	secondary	no	5090	yes	no	unknown	6	may	1297	3	-1	0	unknown	yes

Data Set 2 Visualization

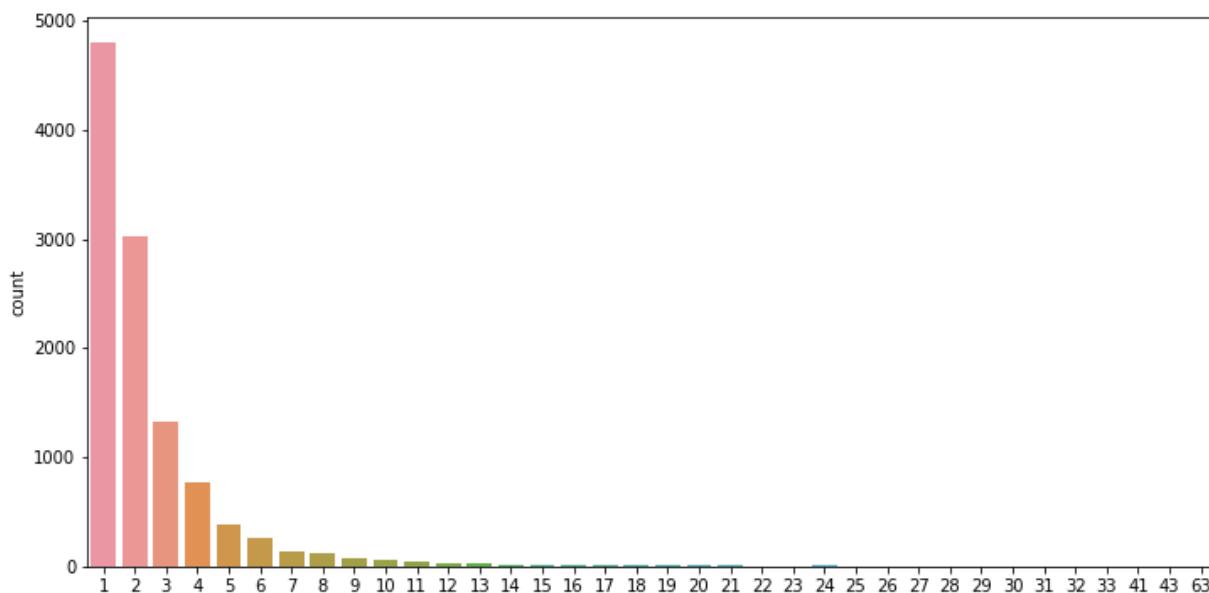
using Boket

```
plt.figure(figsize=(12,6))
sns.countplot(x="x",data=base)
plt.show()
```



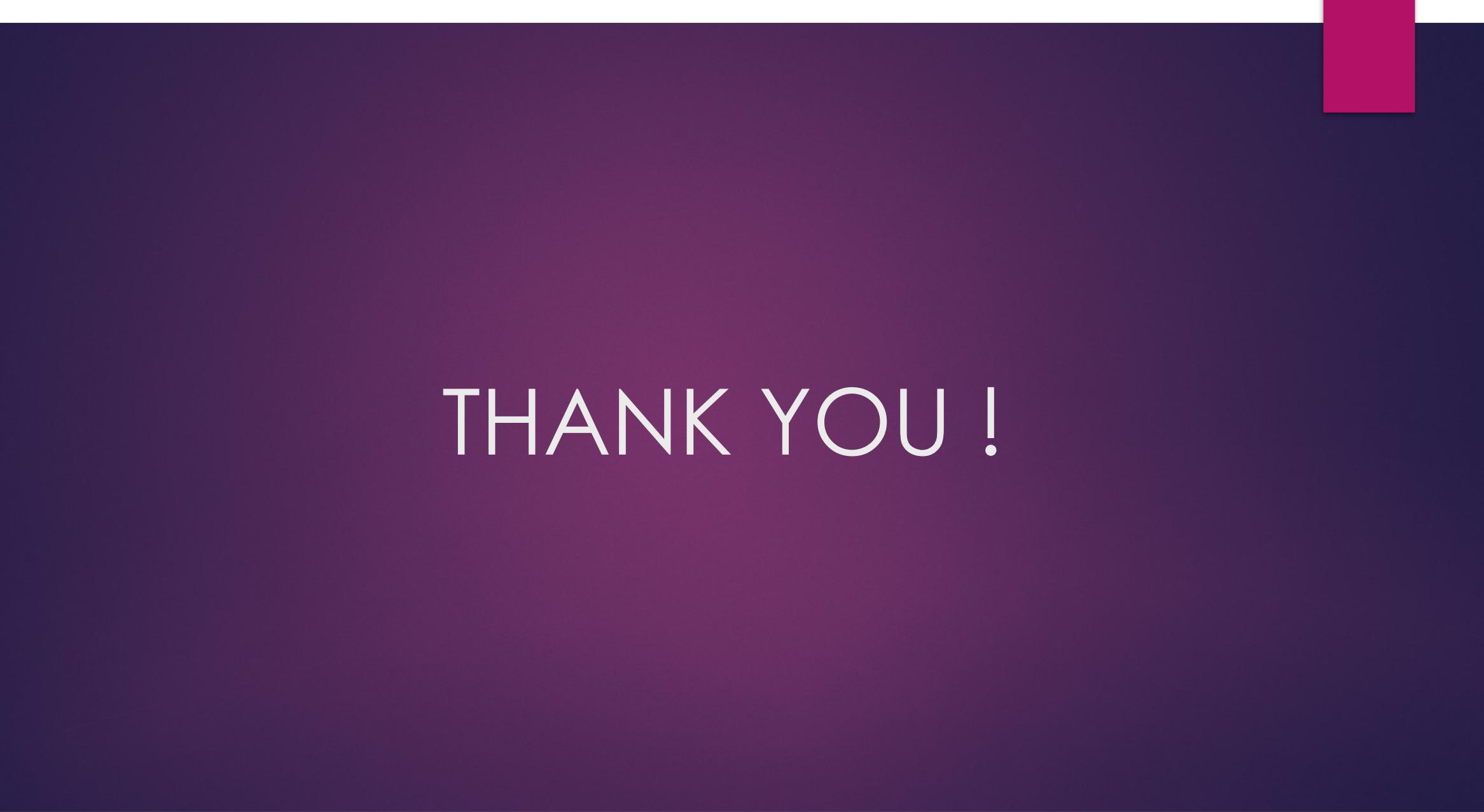
Data Set 2 Visualization using Boket

```
plt.figure(figsize=(12,6))
sns.countplot(x="campaign",data=base)
plt.show()
```



Conclusion

- ▶ I use selenium to open a virtual browser and extract data from it. Then we use beautiful soup to extract data about infrastructure and industry. And I use regular expression to search for specific patterns. Then I extract tabular data from ParseHub and I display this data using Bokeh.



THANK YOU !