Web scraping using Selenium

In [1]:
```
pip install selenium
```

```
Requirement already satisfied: selenium in c:\users\frederick masangkay\anaconda3\lib\site-pack
ages (4.31.0)
Requirement already satisfied: urllib3<3,>=1.26 in c:\users\frederick masangkay\anaconda3\lib\s
ite-packages (from urllib3[socks]<3,>=1.26->selenium) (2.2.3)
Requirement already satisfied: trio~=0.17 in c:\users\frederick masangkay\anaconda3\lib\site-pa
ckages (from selenium) (0.29.0)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\frederick masangkay\anaconda3\li
b\site-packages (from selenium) (0.12.2)
Requirement already satisfied: certifi>=2021.10.8 in c:\users\frederick masangkay\anaconda3\lib
\site-packages (from selenium) (2024.8.30)
Requirement already satisfied: typing_extensions~=4.9 in c:\users\frederick masangkay\anaconda3
\lib\site-packages (from selenium) (4.11.0)
Requirement already satisfied: websocket-client~=1.8 in c:\users\frederick masangkay\anaconda3
\lib\site-packages (from selenium) (1.8.0)
Requirement already satisfied: attrs>=23.2.0 in c:\users\frederick masangkay\anaconda3\lib\site
-packages (from trio~=0.17->selenium) (25.3.0)
Requirement already satisfied: sortedcontainers in c:\users\frederick masangkay\anaconda3\lib\s
ite-packages (from trio~=0.17->selenium) (2.4.0)
Requirement already satisfied: idna in c:\users\frederick masangkay\anaconda3\lib\site-packages
(from trio~=0.17->selenium) (3.7)
Requirement already satisfied: outcome in c:\users\frederick masangkay\anaconda3\lib\site-packa
ges (from trio~=0.17->selenium) (1.3.0.post0)
Requirement already satisfied: sniffio>=1.3.0 in c:\users\frederick masangkay\anaconda3\lib\sit
e-packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: cffi>=1.14 in c:\users\frederick masangkay\anaconda3\lib\site-pa
ckages (from trio~=0.17->selenium) (1.17.1)
Requirement already satisfied: wsproto>=0.14 in c:\users\frederick masangkay\anaconda3\lib\site
-packages (from trio-websocket~=0.9->selenium) (1.2.0)
Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6 in c:\users\frederick masangkay\anac
onda3\lib\site-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: pycparser in c:\users\frederick masangkay\anaconda3\lib\site-pac
kages (from cffi>=1.14->trio~=0.17->selenium) (2.21)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\frederick masangkay\anaconda3\lib\site
-packages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.14.0)
Note: you may need to restart the kernel to use updated packages.
```

In [169…
```
from selenium.webdriver.common.by import By
import pandas as pd
import matplotlib.pyplot as plt
```

```
Matplotlib is building the font cache; this may take a moment.
```

In [104…
```
driver = webdriver.Edge()
driver.get("https://op.gg/champions")
```

In [105…
```
from selenium.webdriver.common.keys import Keys # for scrolled data
# Scroll down the page
driver.find_element(By.TAG_NAME, 'body').send_keys(Keys.END)
# Wait for content to load

import time
time.sleep(2) # Adjust the sleep time based on the website's loading speed
# Scrape the newly loaded content
new_content = driver.find_elements(By.TAG_NAME, 'tr')
```

In [115…
```
rank = []
champion = []
winrate = []
pickrate = []
banrate = []

for item in new_content:
    tds = item.find_elements(By.TAG_NAME, 'td')
    if len(tds) < 7:
        continue

    rank.append(tds[0].text)
    champ_name = tds[1].find_element(By.XPATH, './/a//strong').text
    champion.append(champ_name)
    winrate.append(tds[4].text)
    pickrate.append(tds[5].text)
    banrate.append(tds[6].text)
```

In [253…
```
df = pd.DataFrame({'Rank': rank, 'Champion': champion, 'Win Rate': winrate, 'Pick Rate': pickr
```

In [223…
```
df.tail()
```

Out[223…

	Rank	Champion	Win Rate	Pick Rate	Ban Rate
240	241	Corki	43.22%	0.86%	0.73%
241	242	Maokai	43.21%	0.59%	0.16%
242	243	Kalista	42.45%	2%	0.74%
243	244	Kayle	41.42%	2.27%	1.96%
244	245	Darius	40.54%	1.08%	12.93%

In [260…
```
highestWinrate = df.sort_values(by = ['Win Rate'], ascending = True)
```

In [261…
```
highestPickrate = df.sort_values(by = ['Pick Rate'], ascending = True)
```

In [262…
```
highestBanrate = df.sort_values(by = ['Ban Rate'], ascending = True)
```

In [188…
```
highestWinrate
```

Out[188...

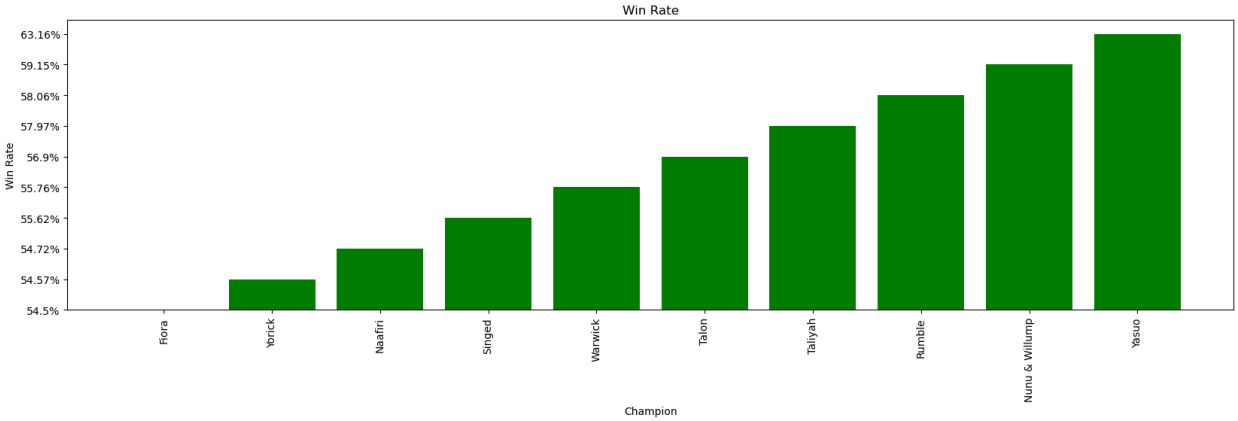
	Rank	Champion	Win Rate	Pick Rate	Ban Rate
19	20	Renekton	52.95%	6.1%	4.28%
18	19	Irelia	53.25%	4.85%	18.51%
17	18	Gnar	53.33%	3.85%	1.65%
16	17	Elise	53.35%	2.72%	11.49%
15	16	Mordekaiser	53.61%	3.56%	3.44%
14	15	Malphite	53.77%	4.38%	5.83%
13	14	Elise	53.85%	3.44%	11.58%
12	13	Nunu & Willump	53.97%	2.67%	0.48%
11	12	Taliyah	54.25%	3.08%	1%
10	11	Sett	54.45%	6.02%	1.76%
9	10	Fiora	54.5%	4.64%	5.57%
8	9	Yorick	54.57%	3.12%	24.8%
7	8	Naafiri	54.72%	1.16%	55.22%
6	7	Singed	55.62%	2.61%	0.87%
5	6	Warwick	55.76%	2.42%	3.06%
4	5	Talon	56.9%	2.53%	4.87%
3	4	Taliyah	57.97%	1.51%	0.92%
2	3	Rumble	58.06%	0.68%	1.51%
1	2	Nunu & Willump	59.15%	0.52%	0.32%
0	1	Yasuo	63.16%	0.96%	15.21%

In [268...

```
# Create the line plot
fig, ax = plt.subplots(figsize=(20, 5))
plt.bar(highestWinrate.Champion, highestWinrate['Win Rate'], color = 'green')

# Add titles and labels
plt.title('Win Rate')
plt.xlabel('Champion')
plt.xticks(rotation=90)
plt.ylabel('Win Rate')


# Show the plot
plt.show()
```

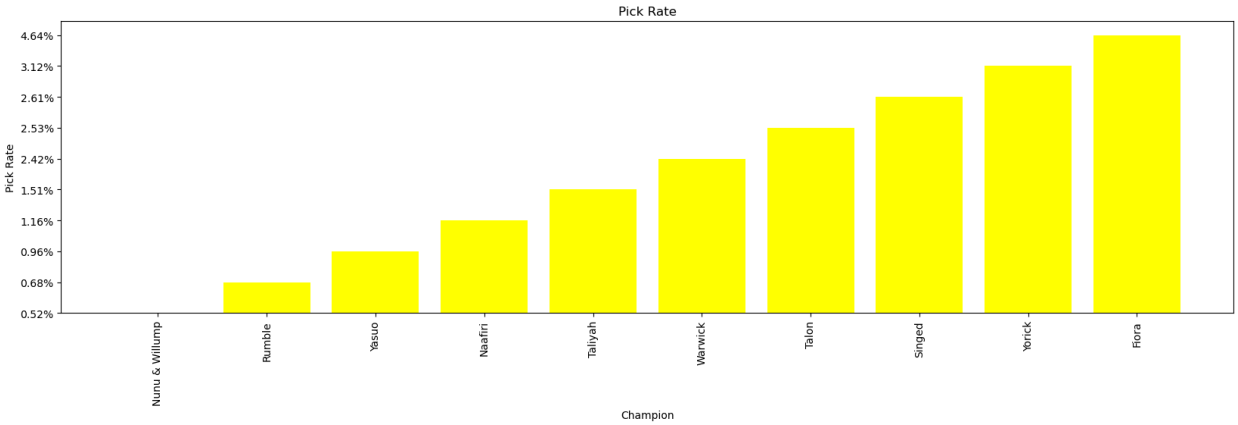


In [267…

```
fig, ax = plt.subplots(figsize=(20, 5))
plt.bar(highestPickrate.Champion, highestPickrate['Pick Rate'], color = 'yellow')

# Add titles and labels
plt.title('Pick Rate')
plt.xlabel('Champion')
plt.xticks(rotation=90)
plt.ylabel('Pick Rate')

# Show the plot
plt.show()
```

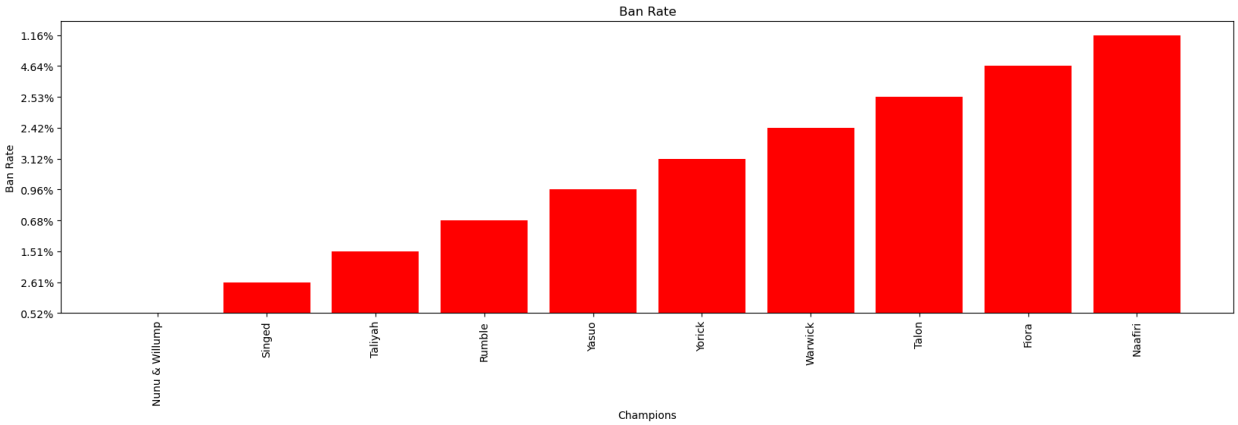


In [266…

```
fig, ax = plt.subplots(figsize=(20, 5))
plt.bar(highestBanrate.Champion, highestBanrate['Pick Rate'], color = 'red')

# Add titles and labels
plt.title('Ban Rate')
plt.xlabel('Champions')
plt.xticks(rotation=90)
plt.ylabel('Ban Rate')

# Show the plot
plt.show()
```


Ban Rate
1.16%
4.64%
2.53%
2.42%
3.12%
0.96%
0.68%
1.51%
2.61%
0.52%
Ban Rate
Nunu & Willump
Singed
Taliyah
Rumble
Yasuo
Yorick
Warwick
Talon
Fiora
Naafiri
Champions