

# Banana Problem

Name: Kurumeti Mitesh

SRN: PES2UG24AM080

SECTION: 3<sup>rd</sup> sem AIML-B

Title: GitHub Trending Repositories

Subject:- UE24MA242A - Mathematics for Computer Science Engineers

## Python code:

```
[ ] pip install requests beautifulsoup4 pandas
```

```
Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-packages (2.32.4)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.12/dist-packages (4.13.5)
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (2.2.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests) (3.4.3)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests) (2.5.0)
Requirement already satisfied: certifi<=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests) (2025.8.3)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.12/dist-packages (from beautifulsoup4) (2.7)
Requirement already satisfied: typing-extensions>=4.0.0 in /usr/local/lib/python3.12/dist-packages (from beautifulsoup4) (4.15.0)
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

```
from bs4 import BeautifulSoup
import requests
import pandas as pd
from google.colab import files
import matplotlib.pyplot as plt
```

```
[ ] url = 'https://github.com/trending'

headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) '
    'AppleWebKit/537.36 (KHTML, like Gecko) '
    'Chrome/115.0.0.0 Safari/537.36'
}

response = requests.get(url, headers=headers)
print('Status code:', response.status_code)
```

```
Status Code: 200
```

```
[ ] soup = BeautifulSoup(response.content, "html.parser")
```

```
[ ] articles = soup.find_all("article", class_="Box-row")
```

```
[ ] repositories = []

for article in articles:
    repo_tag = article.find("h2", class_="h3 lh-condensed").find("a")
    repo_name = repo_tag.text.strip().replace("\n", "").replace(" ", "")

    stars_tag = article.find("a", href=lambda href: href and href.endswith("/stargazers"))
    if stars_tag:
        stars_text = stars_tag.text.strip().replace(",", "")
        try:
            stars = int(stars_text)
        except ValueError:
            stars = 0
    else:
        stars = 0

    lang_tag = article.find("span", itemprop="programmingLanguage")
    language = lang_tag.text.strip() if lang_tag else "N/A"

    repositories.append([repo_name, stars, language])
```

```
[ ] df = pd.DataFrame(repositories, columns=['Name', 'Stars', 'Language'])
df.to_csv('repositories.csv')
files.download('repositories.csv')
print(f'Data extracted and saved to repositories.csv')
```

```
Data extracted and saved to repositories.csv
```

```
filtered_df = df[df["Language"] != "N/A"]
lang_stars = filtered_df.groupby("Language")["Stars"].sum().reset_index()

x = range(len(lang_stars))
height = lang_stars["Stars"]
labels = lang_stars["Language"]

plt.bar(x, height, align="center", color="skyblue", edgecolor="black")
plt.xticks(x, labels, rotation=45, ha="right")

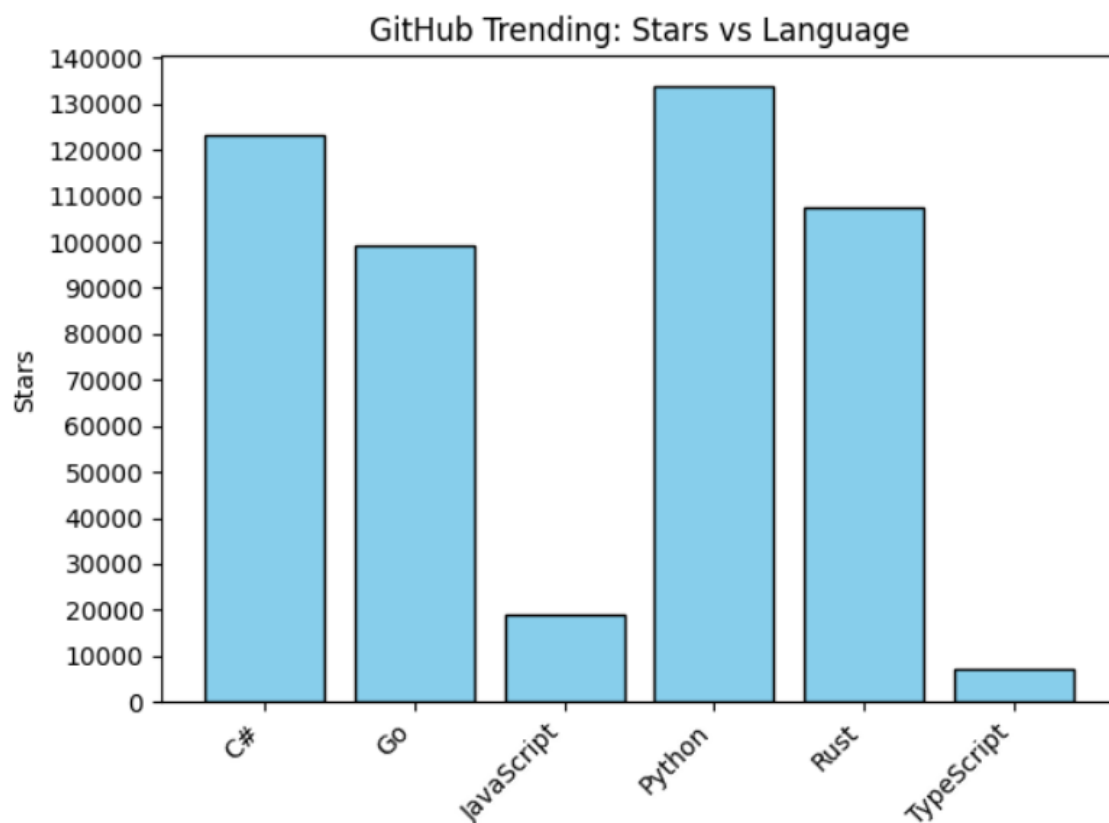
max_stars = height.max()
step = 10000
yticks = list(range(0, max_stars + step, step))
plt.yticks(yticks)

plt.title("Github Trending: Stars vs Language")
plt.ylabel("Stars")
plt.tight_layout()
plt.show()
```

Descriptive statistics results:

	A	B	C	D
1		Name	Stars	Language
2	0	bytebot-ai/bytebot	4374	TypeScript
3	1	LukeGus/Termix	2759	TypeScript
4	2	rustdesk/rustdesk	97462	Rust
5	3	crewAllnc/crewAI	37486	Python
6	4	aquasecurity/trivy	28308	Go
7	5	microsoft/PowerToys	123116	C#
8	6	pedroslopez/whatsapp-web.js	19088	JavaScript
9	7	trufflesecurity/trufflehog	20465	Go
10	8	eriklindernoren/ML-From-Scratch	27822	Python
11	9	pocketbase/pocketbase	50325	Go
12	10	weaviate/elysia	1266	Python
13	11	socfortress/Wazuh-Rules	935	Python
14	12	ansible/ansible	66194	Python
15	13	tensorzero/tensorzero	9954	Rust
16				

Visualization results:



Inferences drawn from the descriptive statistics and visualizations:

- 1) Among the trending repositories, Python stands out with the highest total stars, showing it's currently attracting a lot of attention.
- 2) TypeScript has the fewest stars, suggesting that projects in TypeScript are less at this moment.