

PARTICIPANTS AND THEIR PARTS

Student Name: Ozan Can Demir

Student ID: 402533

Scraper: Scrapy + project documentation

Student Name: Gamze Topal

Student ID: 436823

Scraper: BeautifulSoup + project documentation

Student Name: Recep Arda Kaya

Student ID: 435060

Scraper: Selenium + project documentation

WEBSITE

In our project, we used the same website for all our three scraper which is <https://remoteok.io/>.

Remoteok is the platform where only remote jobs are listed and there are currently more than 30 thousand remote job postings and the postings are divided into many different categories and detailed. People see the most suitable position among the postings they can see with a simple listing such as daily, weekly, monthly, and they can make their applications.

The focus of the <https://remoteok.io/> project in a specific area is beneficial for job seekers as well as for companies trying to find men to work remotely. It should not be easy to reach people who aim to work remotely among millions of job postings. The number of people the project reaches is increasing day by day and we can say that the biggest reason for this is the API it provides. Thanks to the API, people can use all the postings currently on the system (according to the founder Pieter Levels, they publish almost 80 percent of the existing remote job postings) as they wish.

TOPIC

In our project, we aim to analyze one hundred fifty (150) job posts out of thirty (30) thousand posts published at Remoteok website based on their locations, titles, companies and skills distributions. After our analysis, we obtained the outputs which were shared with you on this project.

SCRAPER MECHANICS

In our project, each job post is an element and we extracted one hundred fifty elements along with their following details title, location, company and tags.

TECHNICAL DESCRIPTION AND DATA ANALYSIS

We obtained the outputs as below:

1- LOCATIONS:

```
Worldwide  24.863388
🌐         22.404372
America    3.825137
North      2.732240
United     2.732240
...
+12        0.273224
of          0.273224
Within     0.273224
MST        0.273224
US         0.273224
Length: 78, dtype: float64
```

Based on our outputs, we can see that the distributions of the job posts within 150 posts are mainly weighted as Worldwide with the percentage % 47.26. Please note that Worldwide and the Globe icon represent Worldwide.

2- TITLE

```
Engineer    12.500000
Developer   10.381356
Senior      9.957627
Software    6.991525
Backend     3.389831
...
Solutions   0.211864
Wordpress   0.211864
Azure       0.211864
Performance 0.211864
Scrum       0.211864
Length: 146, dtype: float64
```

Based on our outputs, we can see that Engineer was occurred by % 12.5 inside of the title within 150 job posts which we extracted. Engineer was followed by Developer by % 10.3. The least occurred one is Scrum by % 0.21.

It shows us that Engineer is the most popular job title posted within 150 job posts.

3- COMPANY

```
Toptal      4.761905
Group       1.731602
Digital     1.731602
AG          1.298701
ColorElephant 1.298701
...
Cornershop   0.432900
Deodorant    0.432900
HUGHES       0.432900
Fi           0.432900
PolicyFly    0.432900
Length: 192, dtype: float64
```

When we look at the COMPANY outputs, the data show us that Toptal is the company that posts the greatest number of jobs in the website by % 4.76 while Policyfly posts the least number of jobs by % 0.43.

4- TAGS

```
dev      11.073254
engineer  7.836457
senior    5.110733
react     4.088586
javascript 3.577513
...
science   0.170358
software  0.170358
fast       0.170358
graphql   0.170358
docker    0.170358
Length: 157, dtype: float64
```

The most interesting part for us in our analysis was the TAGS. It gave us the clear understanding that in the demand for developer is really high. We can claim it by looking the output that % 11.07 of job post within 150 has tags as dev. It is followed by Engineer by %7.83. And the least occurred tag is Docker.