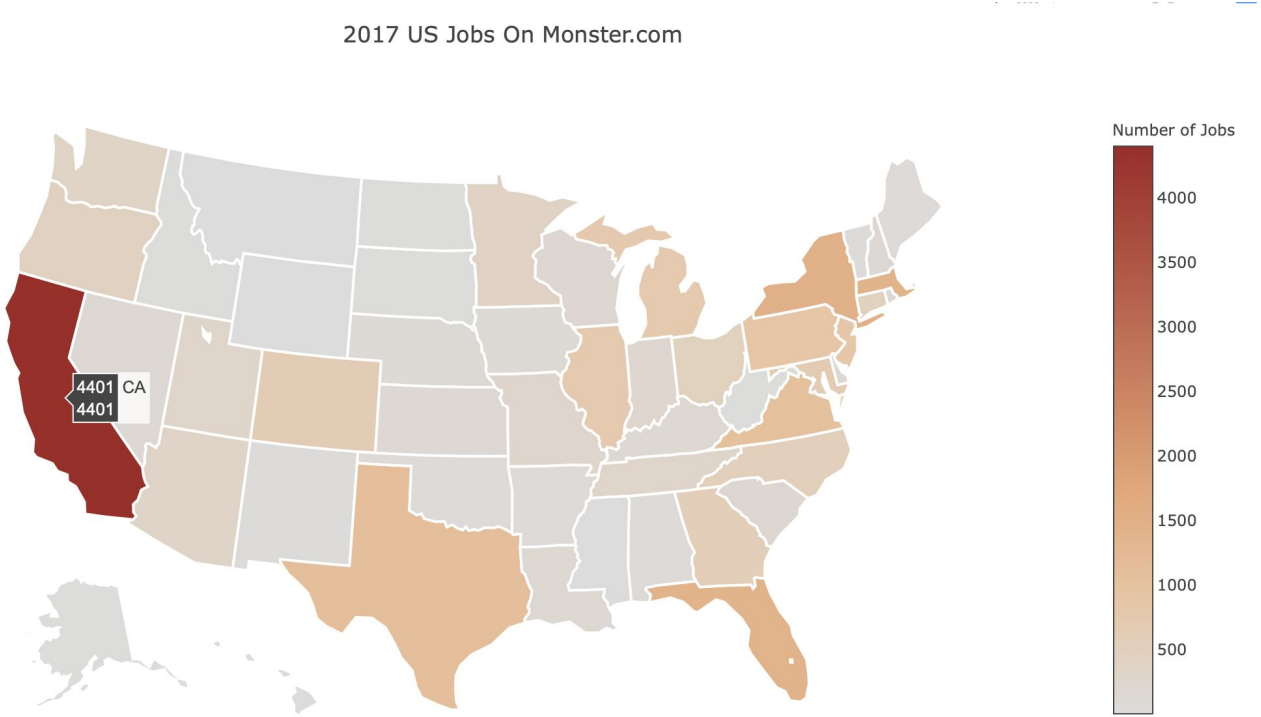
Analysis on Jobs and leetcode questions

Group 14: Shreya, Zeyuan Li, Meihua Su, Zhongke Ma

Motivation and Objective

- Which type of question is more likely to occur?
- Do I have enough knowledge for this company?

Jobs analysis

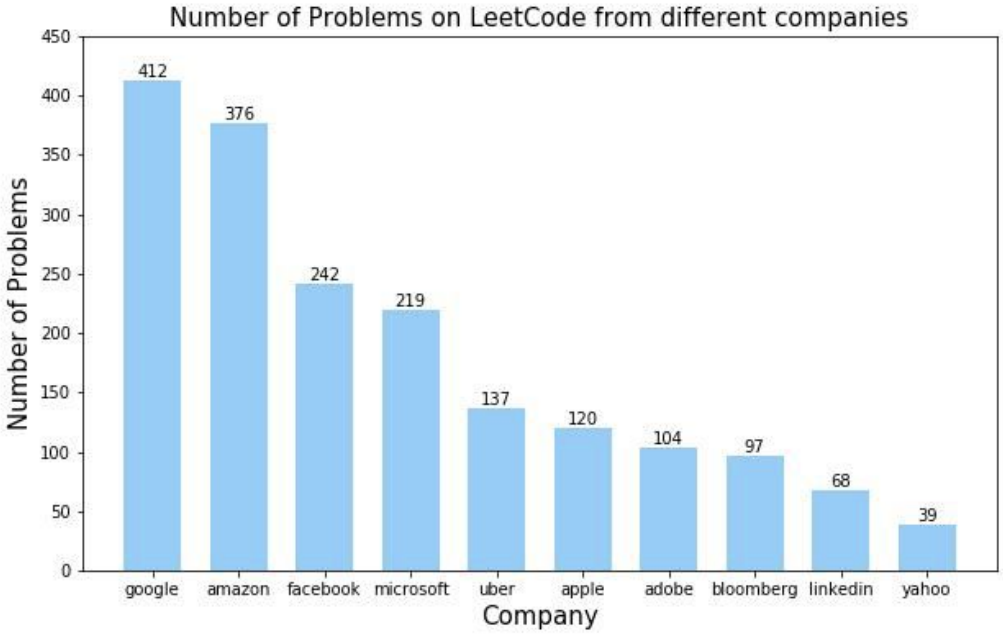


Data source: monster.com

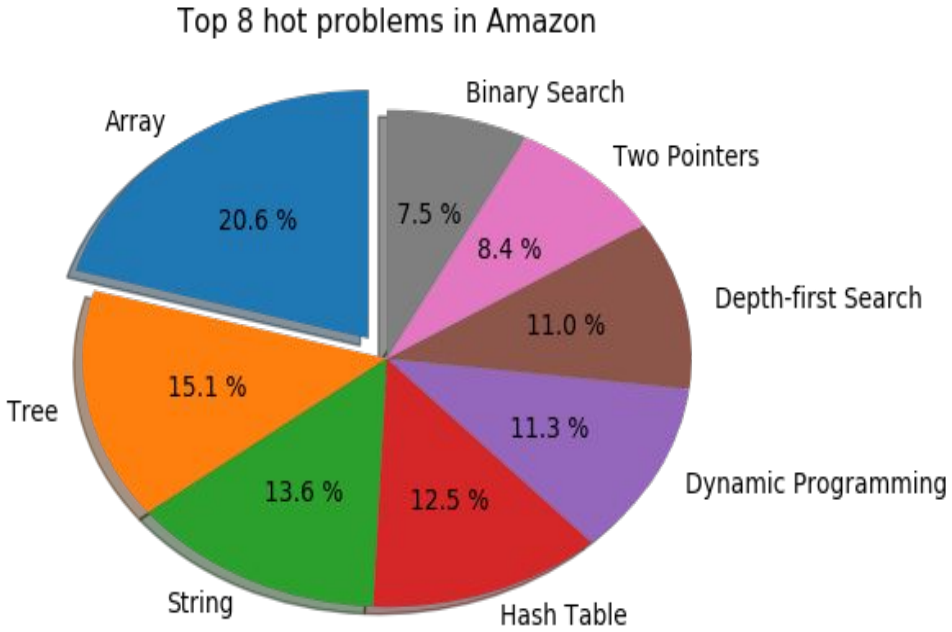
Most wanted titles

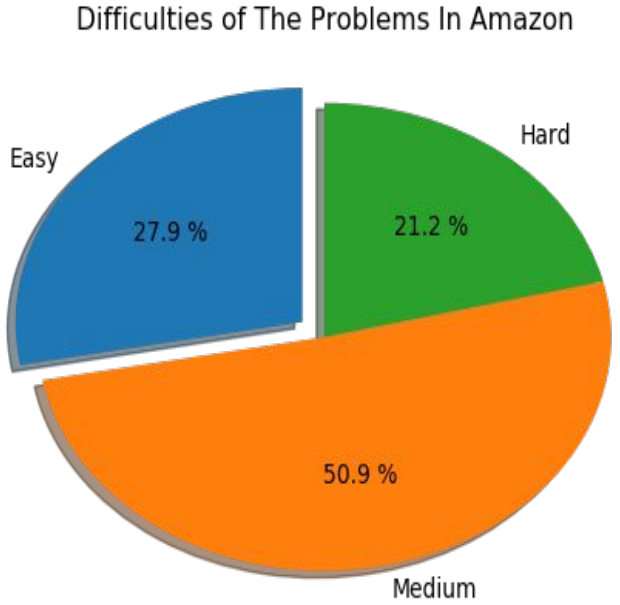
Data source: monster.com

Data set: why using leetcode?

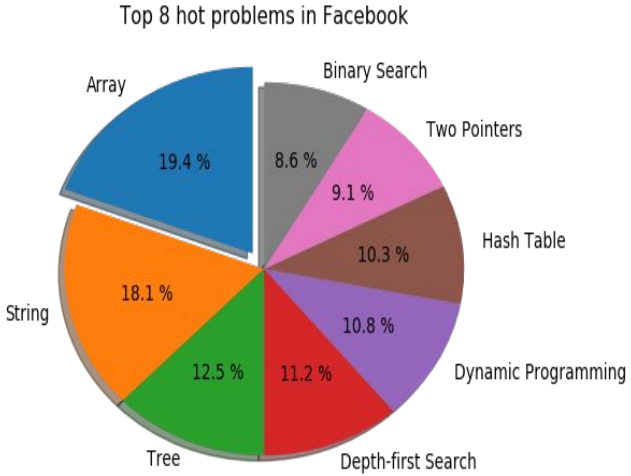


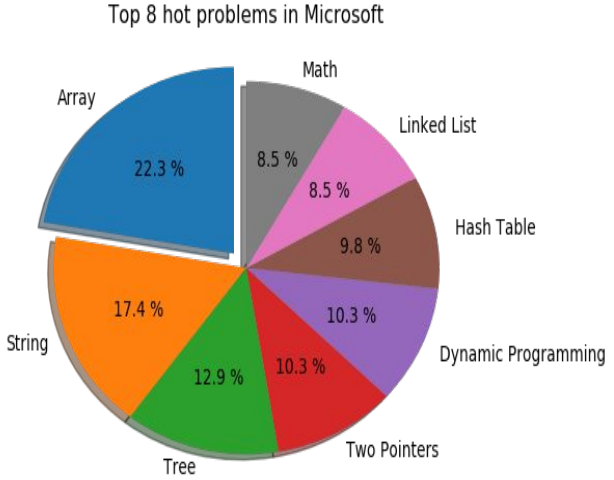
What questions are companies asking for?

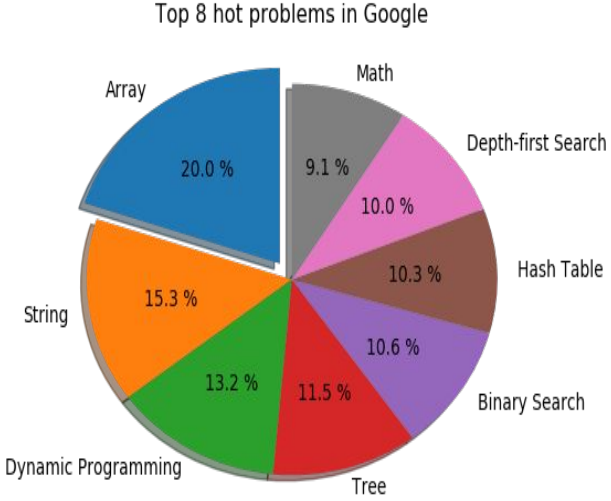


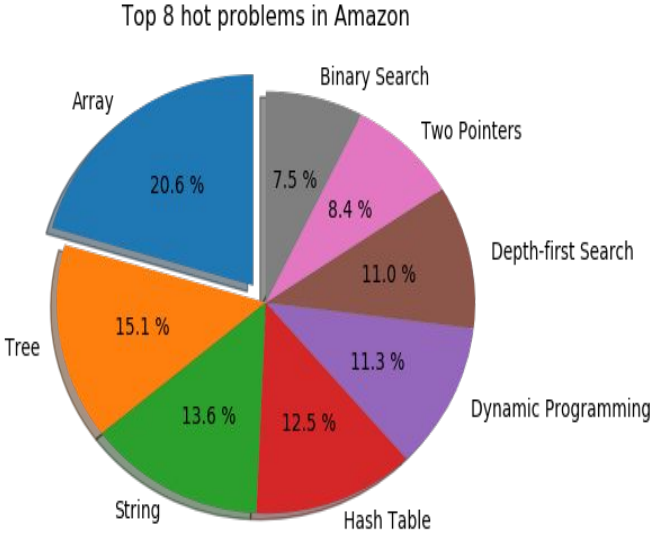


Average: 3.47 tags per question

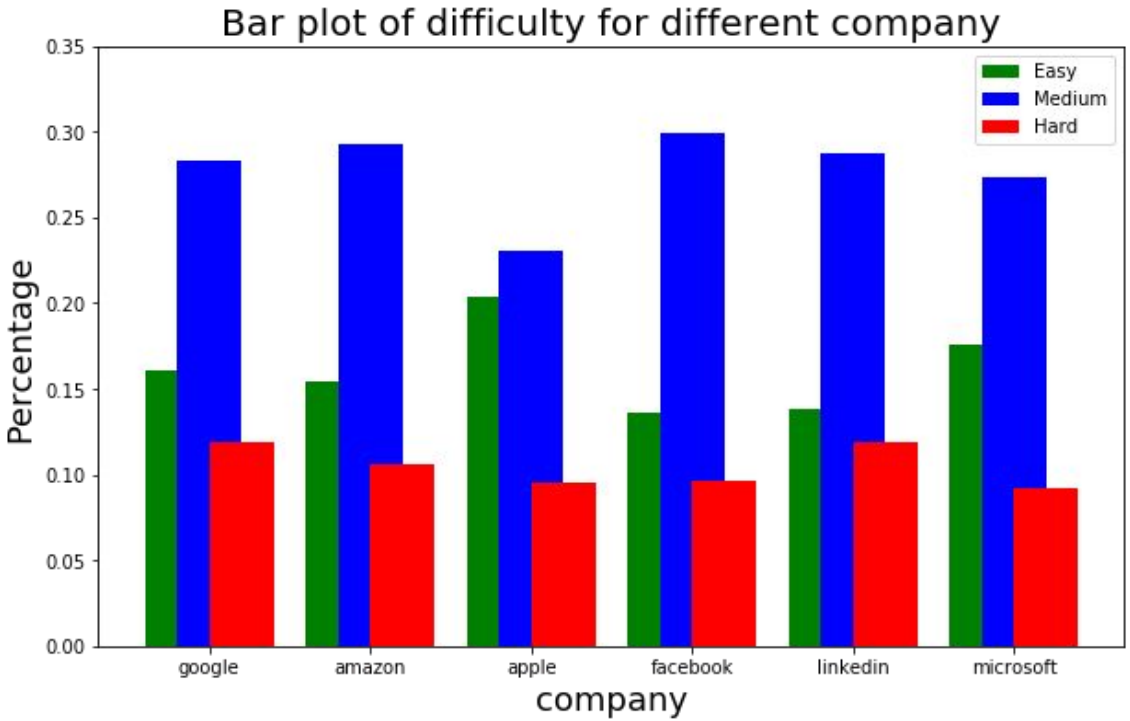

Top 8 hot problems in Facebook
Array
19.4 %
Binary Search
8.6 %
Two Pointers
9.1 %
Hash Table
10.3 %
String
18.1 %
Dynamic Programming
10.8 %
Tree
12.5 %
Depth-first Search
11.2 %


Top 8 hot problems in Microsoft
Array
22.3 %
Math
8.5 %
Linked List
8.5 %
Hash Table
9.8 %
String
17.4 %
Dynamic Programming
10.3 %
Tree
12.9 %
Two Pointers
10.3 %

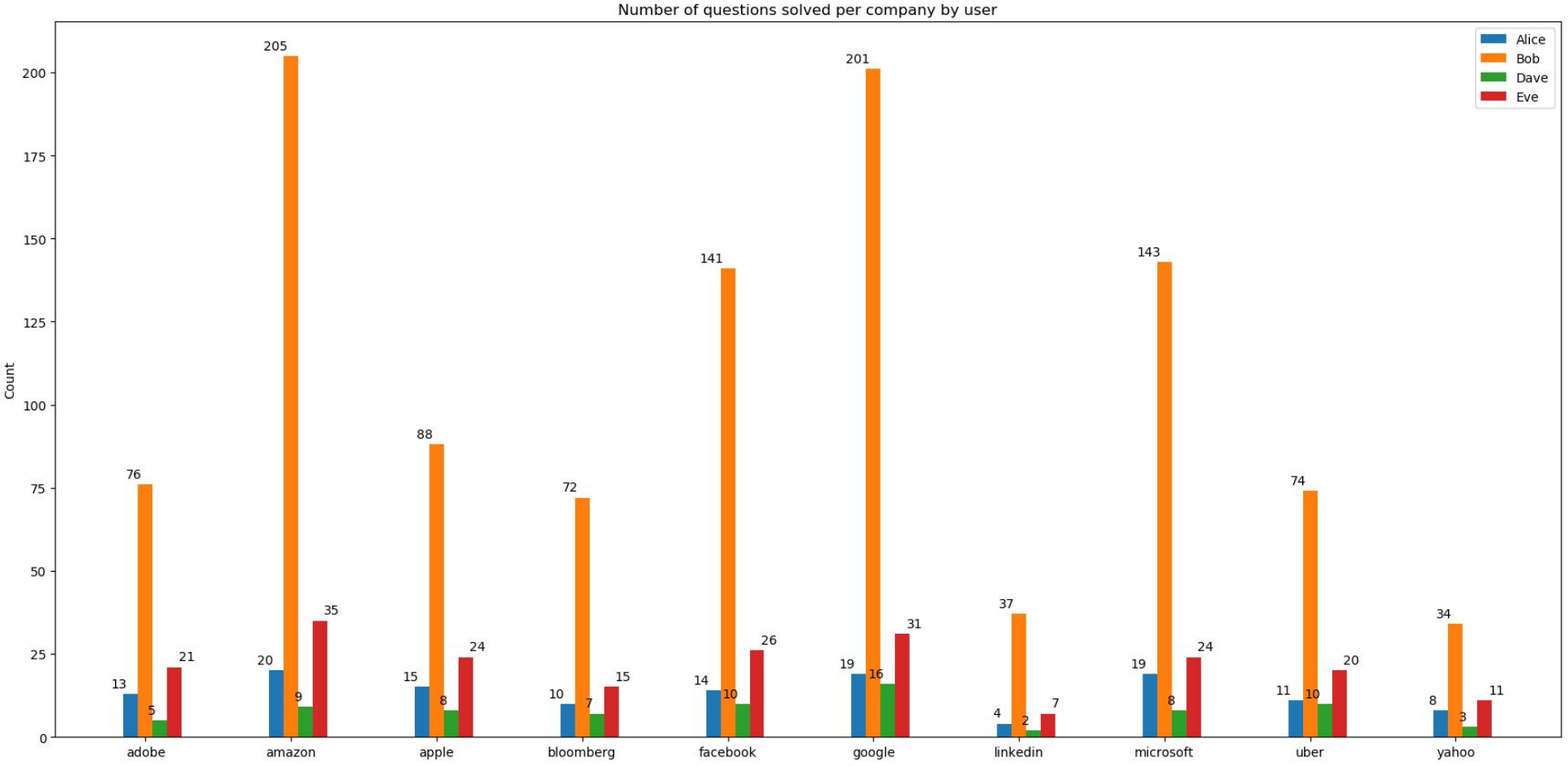

Top 8 hot problems in Google
Array
20.0 %
Math
9.1 %
Depth-first Search
10.0 %
Hash Table
10.3 %
String
15.3 %
Binary Search
10.6 %
Dynamic Programming
13.2 %
Tree
11.5 %


Top 8 hot problems in Amazon
Array
20.6 %
Binary Search
7.5 %
Two Pointers
8.4 %
Depth-first Search
11.0 %
Tree
15.1 %
Dynamic Programming
11.3 %
String
13.6 %
Hash Table
12.5 %

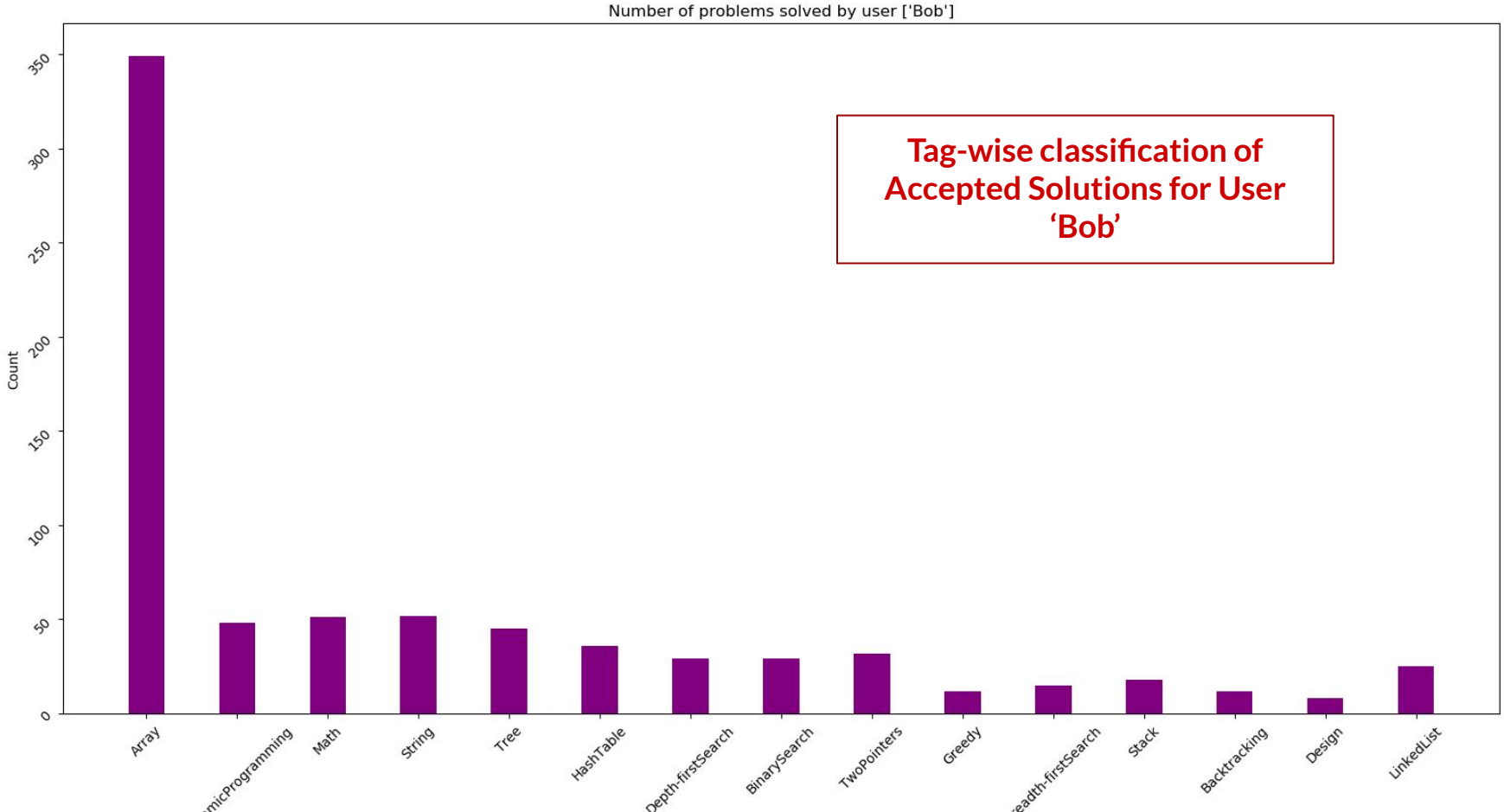
What's the difficulty companies asking for?



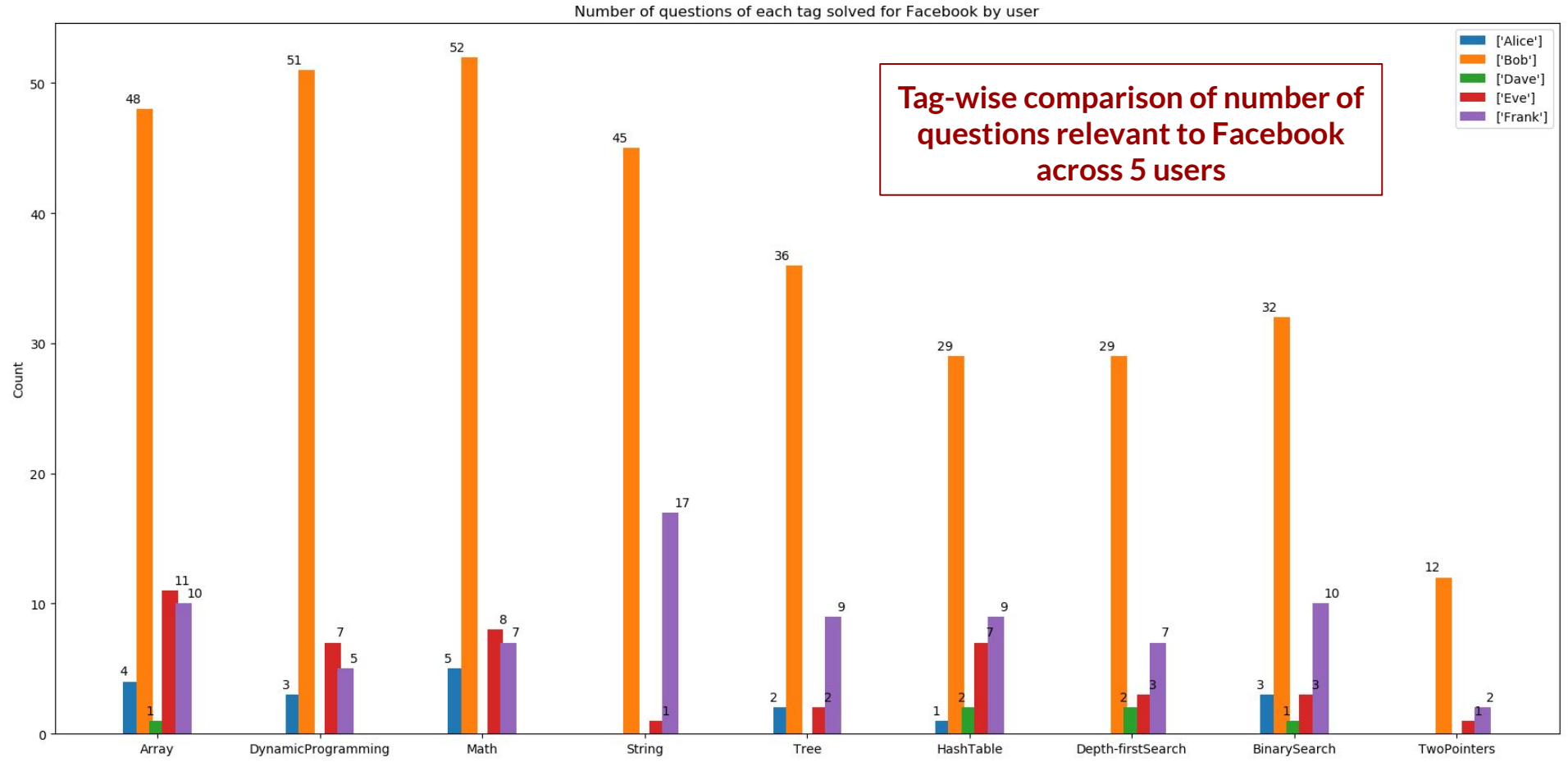
Are YOU ready for the company?



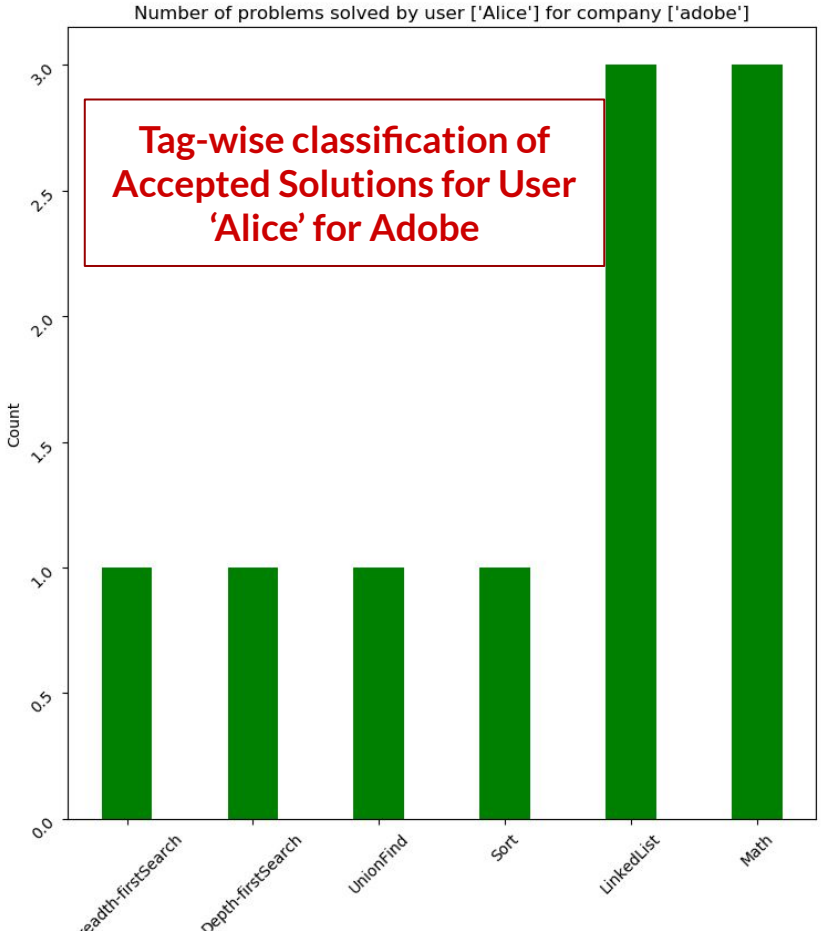
Are YOU ready for the company?

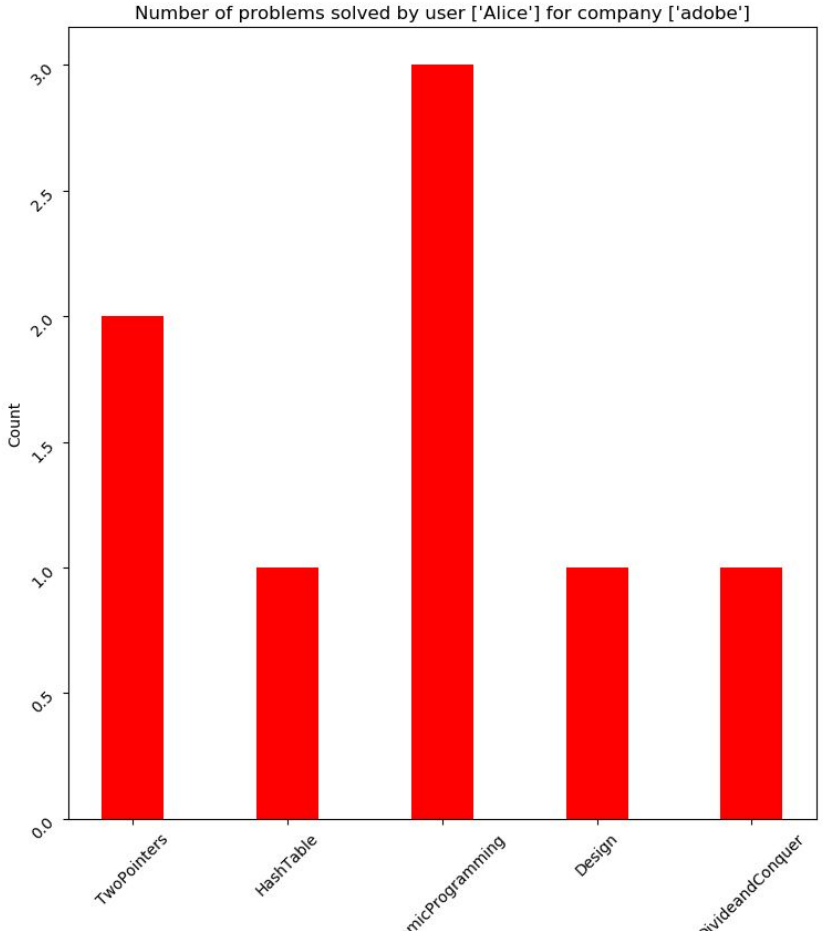


Are YOU ready for the company?



Are YOU ready for the company?





Data Preprocessing and Analysis

Datasets were obtained from -

- Monster.com job postings for 2017
- leetcode.com/company (using a premium Leetcode account)
- leetcode.com/api
  - Problems
  - Submissions
- leetcode.com/tags

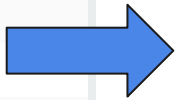
Data flow:

leetcode.com/tag

https://leetcode.com/tag/array/

Show problem tags

#	Title
1	Two Sum
4	Median of Two Sorted Arrays
11	Container With Most Water
15	3Sum
16	3Sum Closest
18	4Sum
26	Remove Duplicates from Sorted Array
27	Remove Element
31	Next Permutation
33	Search in Rotated Sorted Array



JSON


```
{
  "two-sum": [
    "Array",
    "HashTable"
  ],
  "median-of-two-sorted-arrays": [
    "Array",
    "BinarySearch",
    "DivideandConquer"
  ],
  "container-with-most-water": [
    "Array",
    "TwoPointers"
  ],
  "3sum": [
    "Array",
    "TwoPointers"
  ],
  "3sum-closest": [
    "Array",
    "TwoPointers"
  ],
  "4sum": [
    "Array",
```

Data flow:

leetcode.com/api/problems

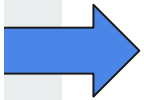
JSON


```
{
  "stat": {
    "question_id": 1014,
    "question__article__live": true,
    "question__article__slug": "k-closest-points-to-origin",
    "question__title": "K Closest Points to Origin",
    "question__title_slug": "k-closest-points-to-origin",
    "question__hide": false,
    "total_acs": 59215,
    "total_submitted": 94880,
    "frontend_question_id": 973,
    "is_new_question": false
  },
  "status": "ac",
  "difficulty": {
    "level": 2
  },
  "paid_only": false,
  "is_favor": false,
  "frequency": 0,
  "progress": 0
},
{
  "stat": {
    "question_id": 1013,
    "question__article__live": null,
    "question__article__slug": null,
    "question__title": "Fibonacci Number",
    "question__title_slug": "fibonacci-number",
    "question__hide": false,
    "total_acs": 55579,
    "total_submitted": 83098,
    "frontend_question_id": 509,
    "is_new_question": false
  },
```

Data flow:

JSON


```
    "is_new_question": false
  },
  "status": "ac",
  "difficulty": {
    "level": 2
  },
  "paid_only": false,
  "is_favor": false,
  "frequency": 0,
  "progress": 0
},
{
  "stat": {
    "question_id": 1013,
    "question__article__live": null,
    "question__article__slug": null,
    "question__title": "Fibonacci Number",
    "question__title_slug": "fibonacci-number",
    "question__hide": false,
    "total_acs": 55579,
    "total_submitted": 83098,
    "frontend_question_id": 509,
    "is_new_question": false
  },
```

CSV

user	company	tags	count
Alice	adobe	Array	21
Alice	adobe	DynamicProgramming	4
Alice	adobe	Math	3
Alice	adobe	String	5
Alice	adobe	Tree	0
Alice	adobe	HashTable	2
Alice	adobe	Depth-firstSearch	1
Alice	adobe	BinarySearch	0
Alice	adobe	TwoPointers	3
Alice	adobe	Greedy	0
Alice	adobe	Breadth-firstSearch	1
Alice	adobe	Stack	0
Alice	adobe	Backtracking	0
Alice	adobe	Design	1
Alice	adobe	LinkedList	3
Alice	adobe	Heap	1
Alice	adobe	BitManipulation	1
Alice	adobe	Sort	3
Alice	adobe	Graph	0
Alice	adobe	UnionFind	1
Alice	adobe	DivideandConquer	2

Data flow:

Website

1. Two Sum

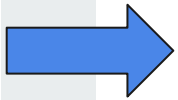
Easy 10790 361 Favorite Share

Given an array of integers, return **indices** of the two numbers such that they add up to a specific target.

You may assume that each input would have ***exactly*** one solution, and you may not use the *same* element twice.

Example:

```
Given nums = [2, 7, 11, 15], target = 9,

Because nums[0] + nums[1] = 2 + 7 = 9,
return [0, 1].
```



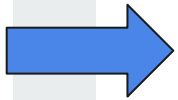
File

```
# [Two Sum][title]

## Description

Given an array of integers, return **indices** of the two numbers such that
they add up to a specific target.

You may assume that each input would have **_exactly_** one solution, and you
may not use the _same_ element twice.

**Example:**

    Given nums = [2, 7, 11, 15], target = 9,

    Because nums[ **0** ] + nums[ **1** ] = 2 + 7 = 9,
    return [ **0** , **1** ].

**Tags:** Array, Hash Table

**Difficulty:** Easy

##

[title]: https://leetcode.com/problems/two-sum
```

Data flow:

Text file

```
# [Two Sum][title]

## Description

Given an array of integers, return **indices** of the two numbers such that
they add up to a specific target.

You may assume that each input would have **_exactly_** one solution, and you
may not use the _same_ element twice.

**Example:**

    Given nums = [2, 7, 11, 15], target = 9,

    Because nums[ **0** ] + nums[ **1** ] = 2 + 7 = 9,
    return [ **0** , **1** ].




**Tags:** Array, Hash Table

**Difficulty:** Easy
##

[title]: https://leetcode.com/problems/two-sum
```



Python

```
file=open(file_name,'r')
for line in file:
    problem_name=line[30:-2]
    for term in folders_in_all_questions_folder:
        if problem_name in term:

            description_problem=open(all_questions_folde

            for line2 in description_problem:
                if '**Tags:**' in line2:
                    #print(line2)
                    line_count+=1
                    tag_list=line2[10:-1].split(', ')
                    for tag in tag_list:
                        if tag in tag_ratio:
                            tag_ratio[tag]=tag_ratio[tag
                            tag_count = tag_count+len(ta
```

Data flow:

CSV file

job_title	j	location
IT Support Technician Job in Madison	Fu	Madison, WI 53702
Business Reporter/Editor Job in Madisor	Fu	Madison, WI 53708
Johnson & Johnson Family of Companies J	Fu	DePuy Synthes Companies
Engineer - Quality Job in Dixon	Fu	Dixon, CA
Shift Supervisor - Part-Time Job in Can	Fu	Camphill, PA
Construction PM - Charlottesville Job i	Fu	Charlottesville, VA
CyberCoders Job Application for Princi	Fu	Contact name Tony Zeric
Mailroom Clerk Job in Austin	Fu	Austin, TX 73301
Housekeeper Job in Austin	Pa	Austin, TX 78746
Video Data Management /Transportation Te		Chesterfield, MO
Aflac Insurance Sales Agent Job in Berr	Fu	Berryville, VA 22611
Sales Associate Job in Columbus	Fu	Columbus, IN
Junior Proofreader Job in Boston	Fu	Boston, MA
Primrose Private Preschool Teacher Job	Fu	Houston, TX 77098
Superintendent Job in Houston	Fu	Houston, TX
Engineer II - Outside Plant Engineer J	Fu	Des Moines, IA 50317
Horizontal Construction Engineers Job in		Wadesboro 28170
Pest Control Technician 款?Field Servic	Fu	Denver, CO
Content Developer for MATLAB Code Gener	Fu	Natick, MA
Technician - Robot & Multi-Axis CNC Fie	Fu	Carter Lake, IA 51510
Business & Strategy Analyst Job in Hous	Fu	Houston, TX
Real Estate Underwriter -Jr Job in Hous	Fu	Houston, TX 77027
Carbet Engineer - Construction and Engin		La Porte 46350

Python

```
In [54]: wc = WordCloud(background_color='white',
                        width=4000,
                        height=3200,
                        ).generate_from_frequencies(most_freq)
         wc.to_file('jobcloud.png')
         plt.imshow(wc)
         plt.axis('off')
         plt.show()
```

References:

Main data set:
Scrapied from leetcode.com

Other data set:

monster.com

Method:

For the file extraction from Leetcode problem sets to txt files:

https://github.com/gcyml/leetcode-crawler