Analysis on Jobs and Leetcode questions

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

Motivation and Objective

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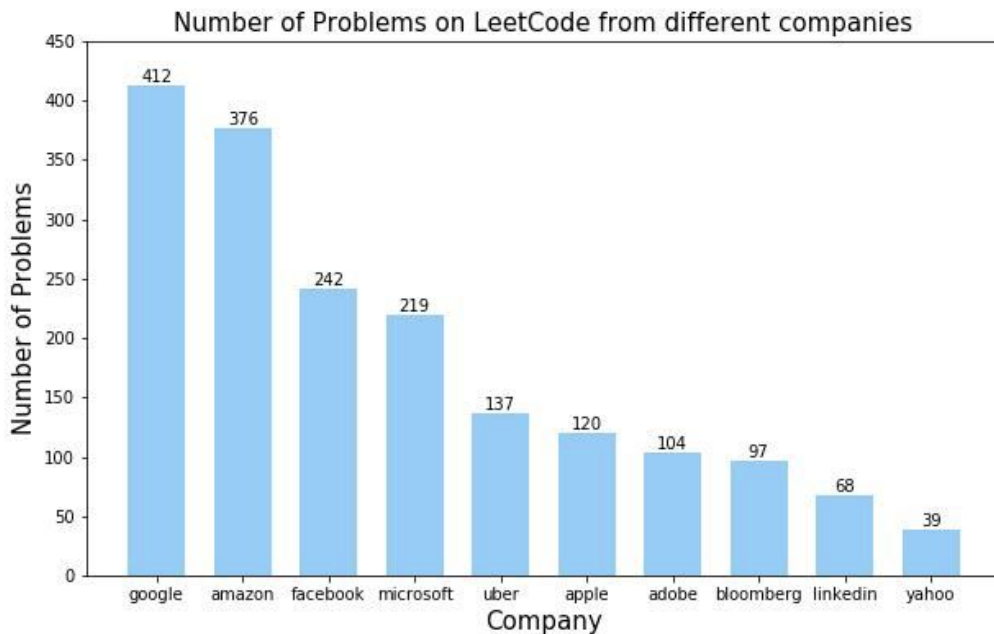
- Do I have enough knowledge for this company?

Jobs analysis

2017 US Jobs On Monster.com

Data source: monster.com

Most wanted titles

Development Lead Architect
Developer Network
Manager Security Application End
Administrator Analyst
Full QA Technician Technical
Front Consultant Software
Specialist
Engineer Net
Java DevOps C NET Business Stack
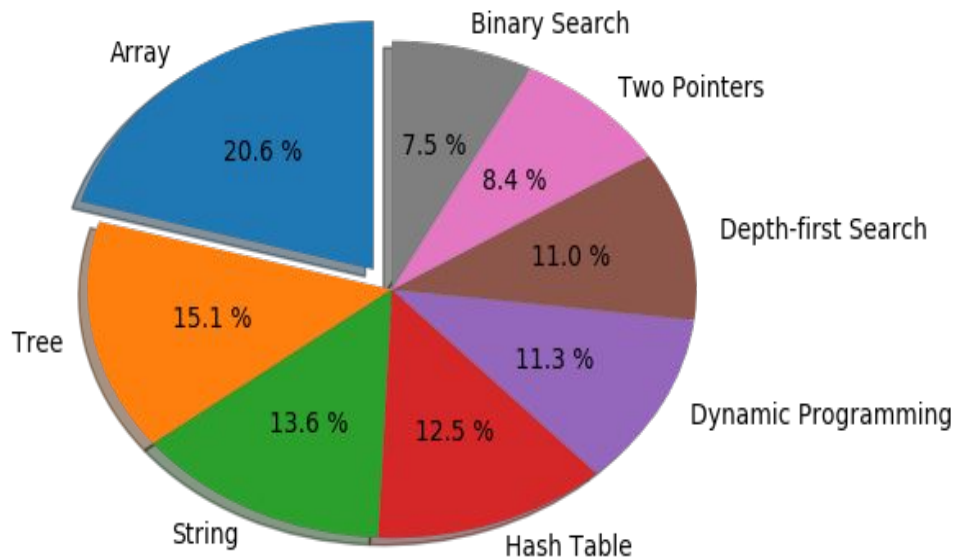IT Senior with Web
Support
Systems SAP and Data Project
Oracle

Data source: monster.com

Data set: why using leetcode?

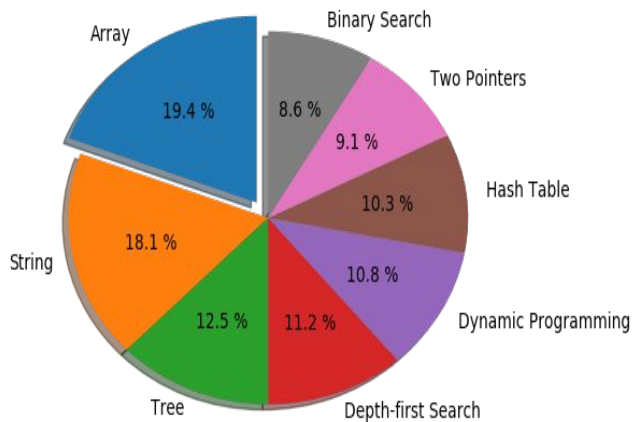# What questions are companies asking for?

Top 8 hot problems in Amazon

- Array — 20.6 %
- Binary Search — 7.5 %
- Two Pointers — 8.4 %
- Depth-first Search — 11.0 %
- Dynamic Programming — 11.3 %
- Hash Table — 12.5 %
- String — 13.6 %
- Tree — 15.1 %

Difficulties of The Problems In Amazon

- Easy — 27.9 %
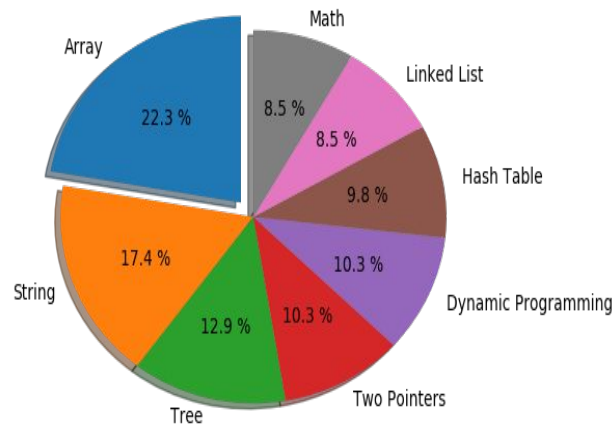- Hard — 21.2 %
- Medium — 50.9 %
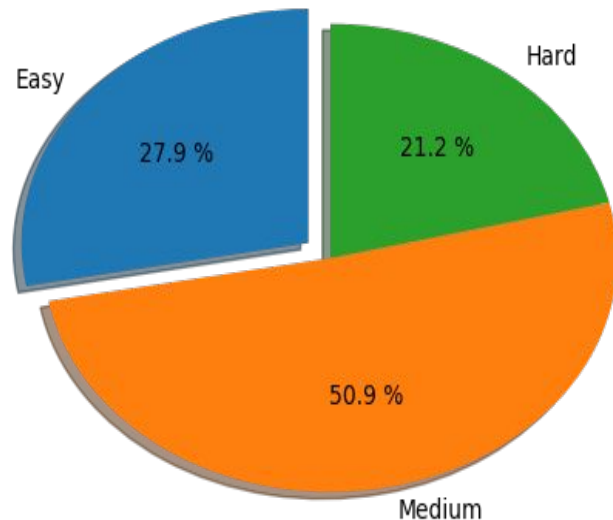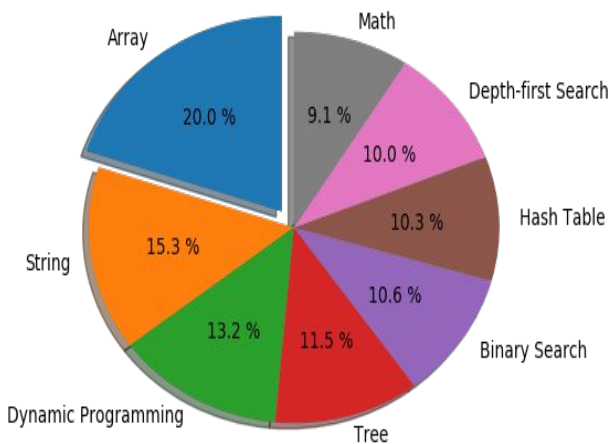
Average: 3.47 tags per question
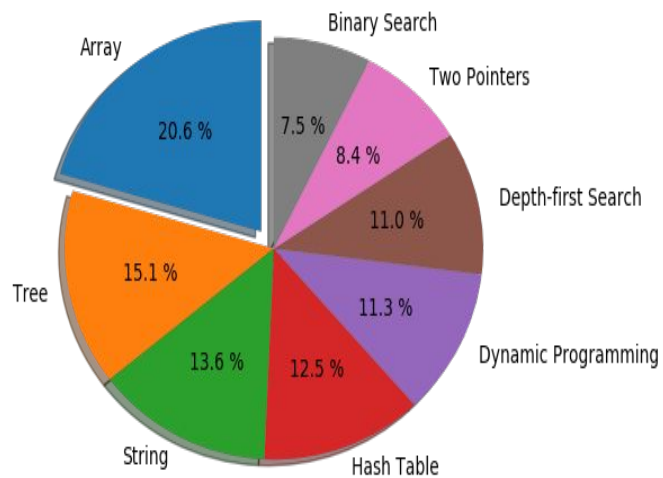
Top 8 hot problems in Facebook

Top 8 hot problems in Microsoft
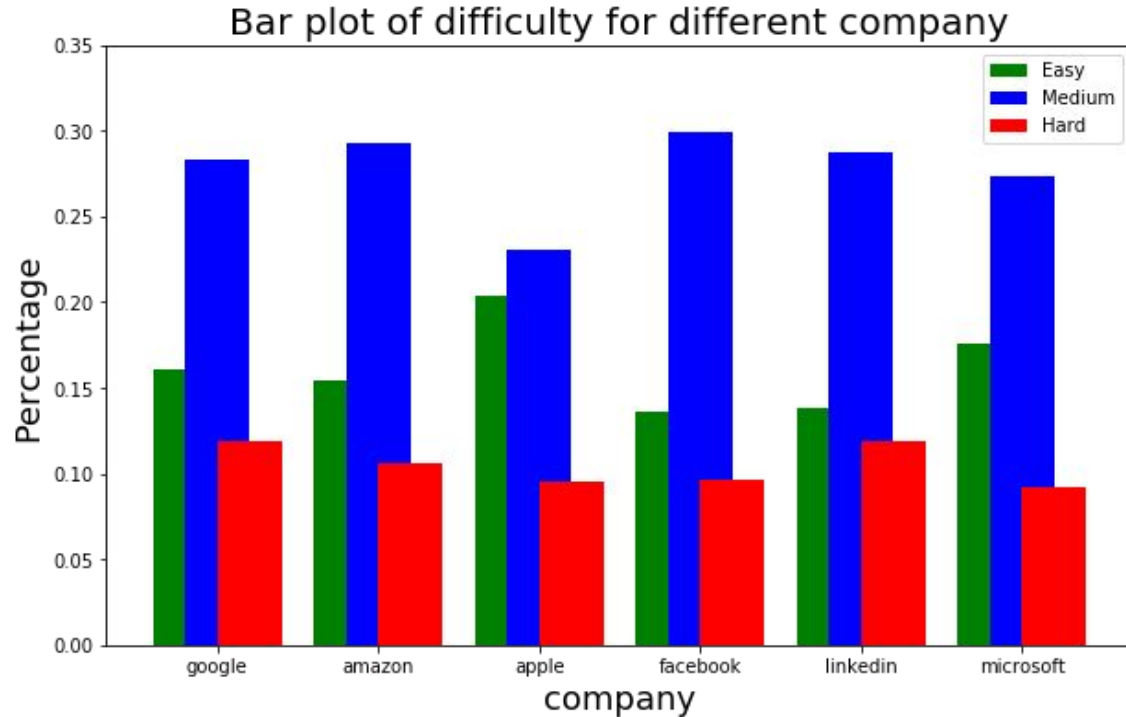
Top 8 hot problems in Google

Top 8 hot problems in Amazon

What's the difficulty companies are asking for?

Bar plot of difficulty for different company

Are YOU ready for the company?

Number of problems solved by user ['Bob']

Tag-wise classification of Accepted Solutions for User 'Bob'
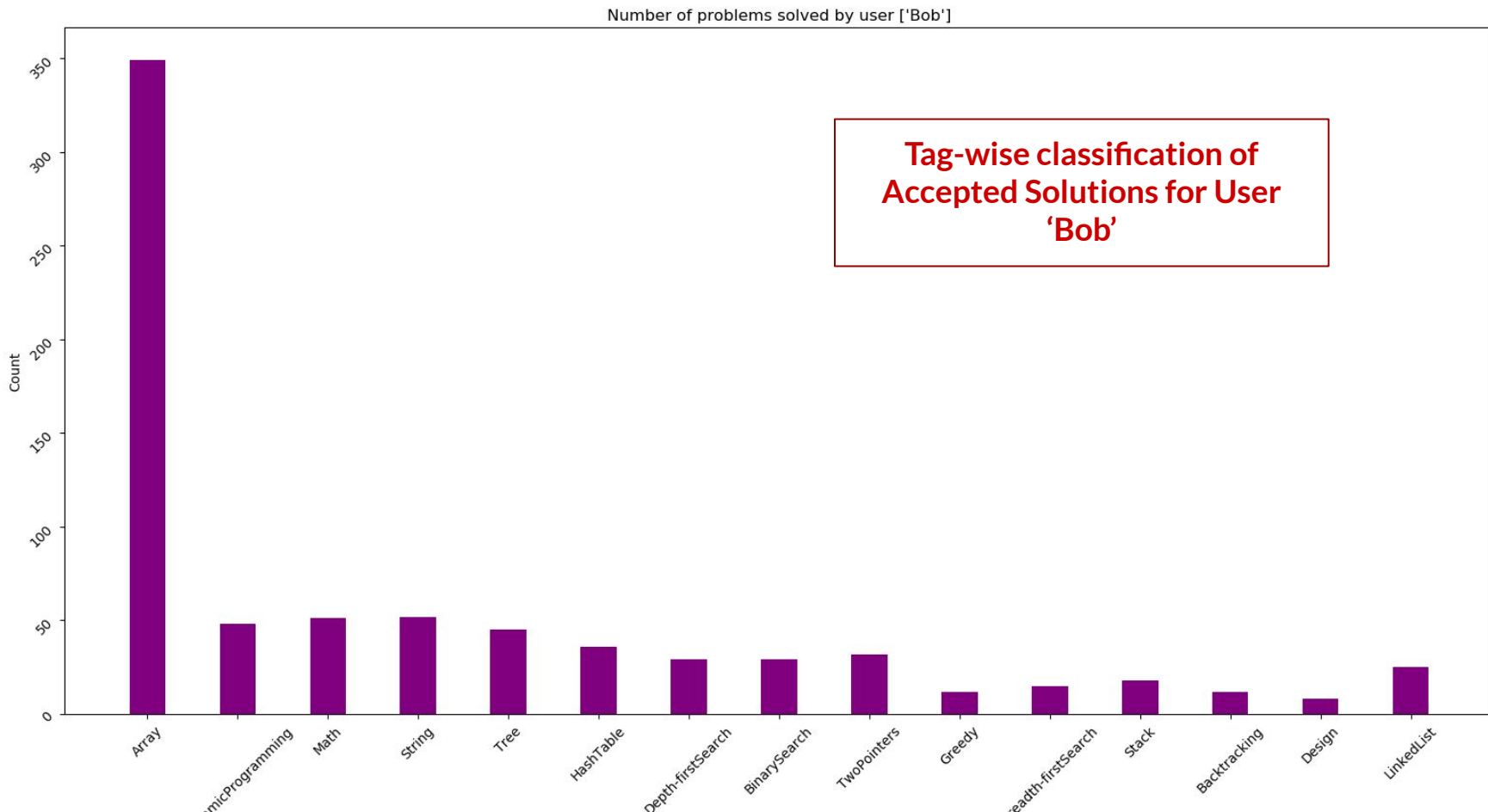
Are YOU ready for the company?

Number of questions solved per company by user

Are YOU ready for the company?

Number of problems solved by user ['Alice'] for company ['adobe']

Tag-wise classification of Accepted Solutions for User 'Alice' for Adobe

Number of problems solved by user ['Alice'] for company ['adobe']
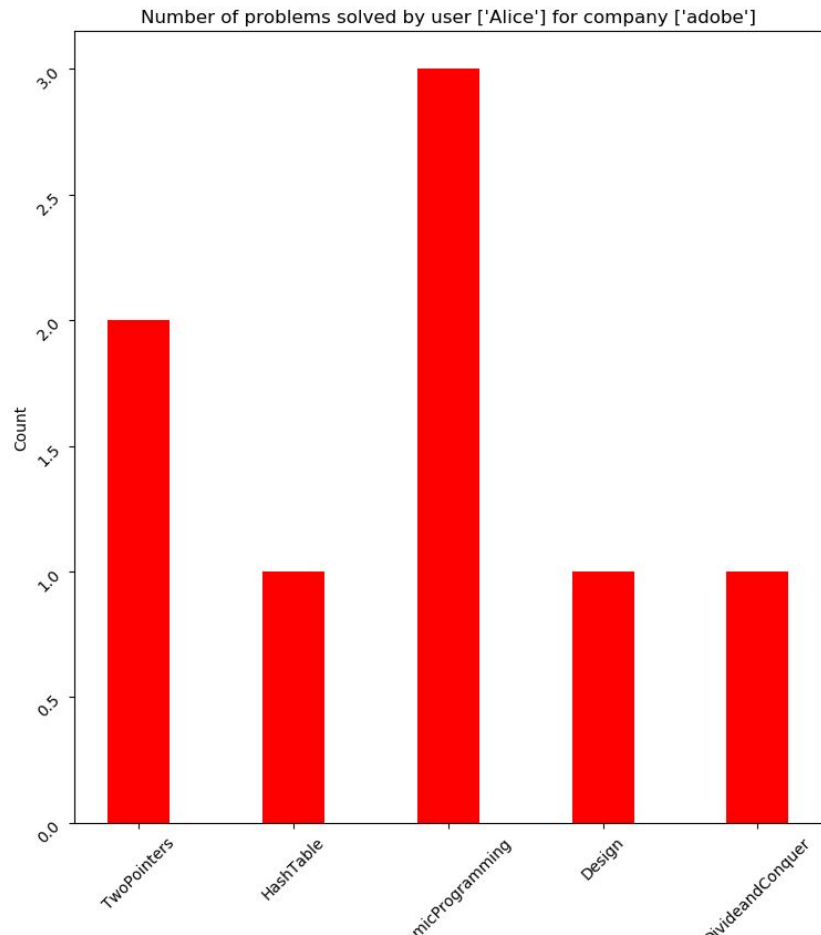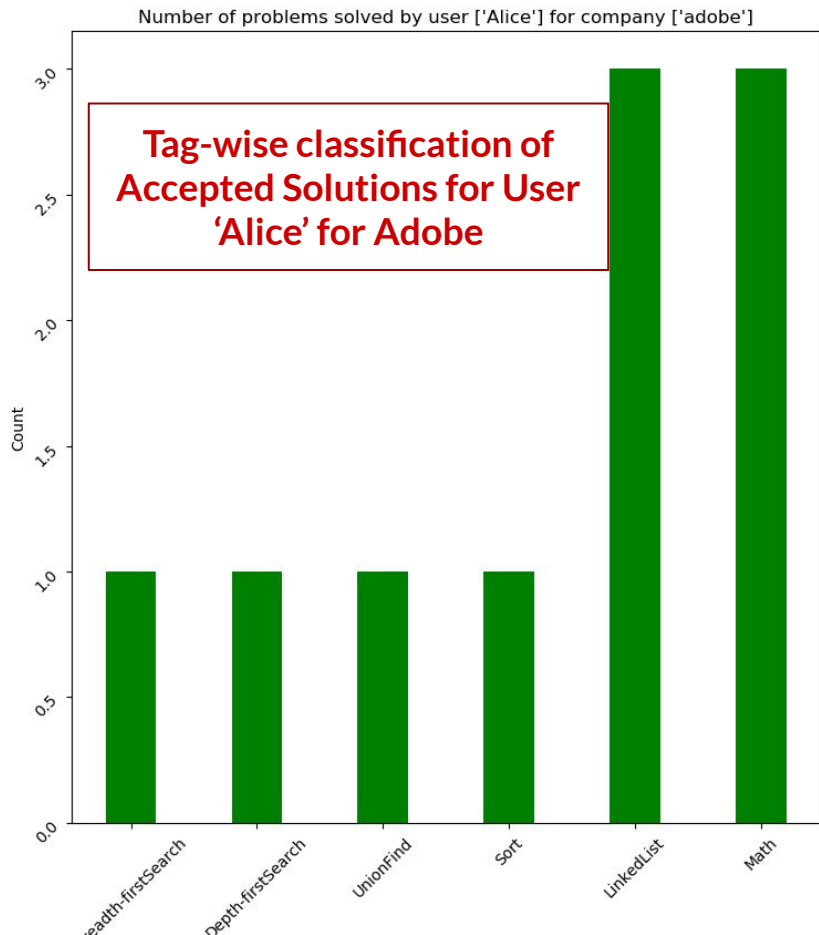
Are YOU ready for the company?

Number of questions of each tag solved for Facebook by user

Tag-wise comparison of number of questions relevant to Facebook across 5 users
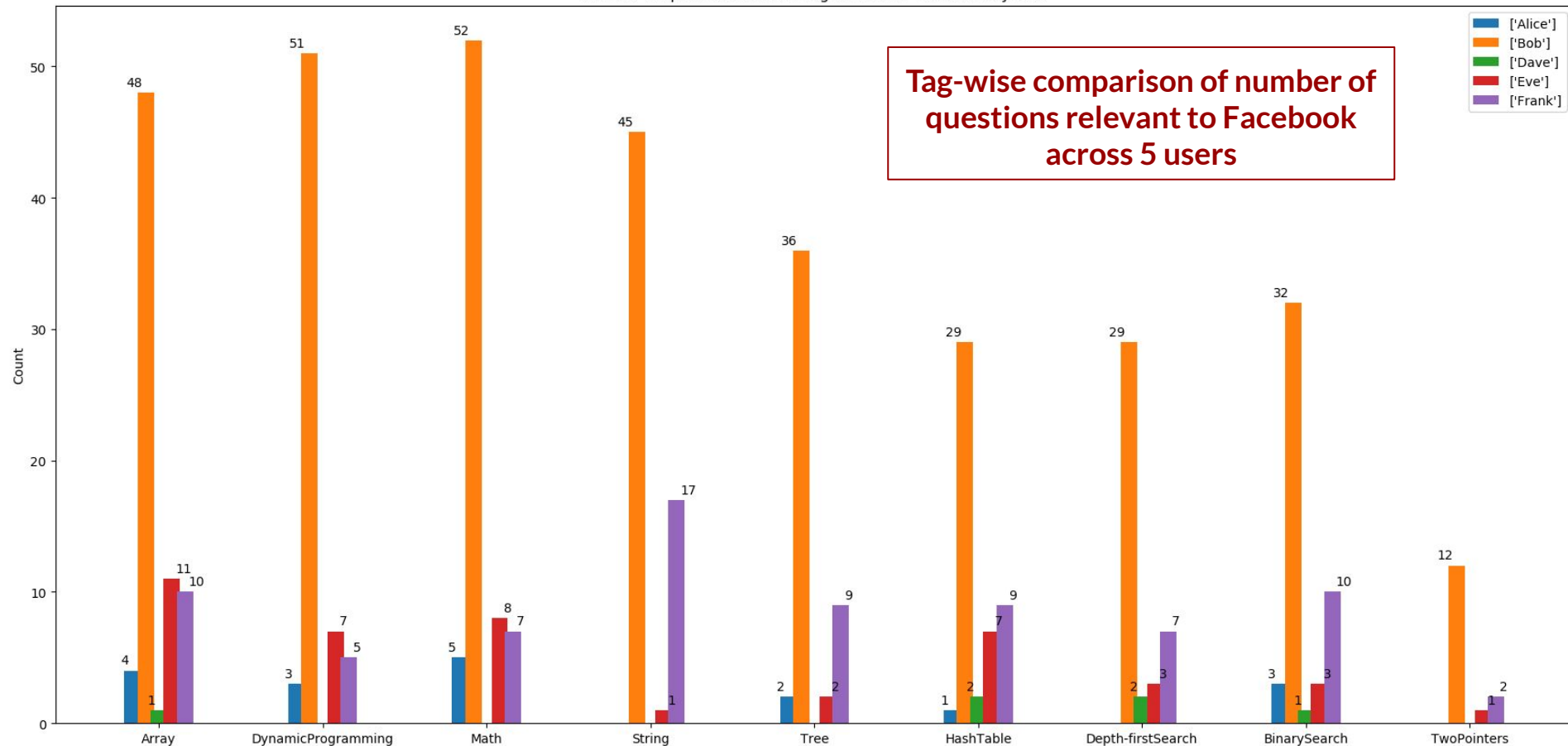
Data Preprocessing and Analysis

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

JSON

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

☐ Show problem tags

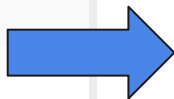| #	Title
1	Two Sum
4	Median of Two Sorted Arrays
11	Container With Most Water
15	3Sum
16	3Sum Closest
18	4Sum
26	Remove Duplicates from Sorted Array
27	Remove Element
31	Next Permutation
33	Search in Rotated Sorted Array

```
 1  {
 2    "two-sum": [
 3      "Array",
 4      "HashTable"
 5    ],
 6    "median-of-two-sorted-arrays": [
 7      "Array",
 8      "BinarySearch",
 9      "DivideandConquer"
10    ],
11    "container-with-most-water": [
12      "Array",
13      "TwoPointers"
14    ],
15    "3sum": [
16      "Array",
17      "TwoPointers"
18    ],
19    "3sum-closest": [
20      "Array",
21      "TwoPointers"
22    ],
23    "4sum": [
24      "Array",
```

Data flow:

leetcode.com/api/problems

← → C 🔒 https://leetcode.com/api/problems/all/

{"user_name": "ghostofchina", "num_solved": 0, "num_total": 1050, "ac_easy": 0, "ac_medium": ...
"question__article__live": null, "question__article__slug": null, "question__title": "Distant Barcod...
"total_acs": 3375, "total_submitted": 9091, "frontend_question_id": 1054, "is_new_question": tr...
false, "frequency": 0, "progress": 0}, {"stat": {"question_id": 1139, "question__article__live": null...
With One Swap", "question__title_slug": "previous-permutation-with-one-swap", "question__h...
"frontend_question_id": 1053, "is_new_question": true}, "status": null, "difficulty": {"level": 2}, "p...
{"question_id": 1138, "question__article__live": null, "question__article__slug": null, "question__t...
bookstore-owner", "question__hide": false, "total_acs": 4120, "total_submitted": 8063, "fronte...
{"level": 2}, "paid_only": false, "is_favor": false, "frequency": 0, "progress": 0}, {"stat": {"questi...
"question__title": "Height Checker", "question__title_slug": "height-checker", "question__hide"...
1051, "is_new_question": true}, "status": null, "difficulty": {"level": 1}, "paid_only": false, "is_favo...
"question__article__live": null, "question__article__slug": null, "question__title": "Actors and Dir...
"actors-and-directors-who-cooperated-at-least-three-times", "question__hide": false, "tota...
"is_new_question": true}, "status": null, "difficulty": {"level": 1}, "paid_only": true, "is_favor": fals...
"question__article__live": null, "question__article__slug": null, "question__title": "Customers Wh...
all-products", "question__hide": false, "total_acs": 606, "total_submitted": 918, "frontend_ques...
2}, "paid_only": true, "is_favor": false, "frequency": 0, "progress": 0}, {"stat": {"question_id": 113...
"question__title": "Last Stone Weight II", "question__title_slug": "last-stone-weight-ii", "questi...
"frontend_question_id": 1049, "is_new_question": false}, "status": null, "difficulty": {"level": 2}, "...

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:

CSV

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
  },
```

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

File

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].
```

```
 1  # [Two Sum][title]
 2
 3  ## Description
 4
 5  Given an array of integers, return **indices** of the two numbers such that
 6  they add up to a specific target.
 7
 8  You may assume that each input would have **_exactly_** one solution, and you
 9  may not use the _same_ element twice.
10
11  **Example:**
12
13      Given nums = [2, 7, 11, 15], target = 9,
14
15      Because nums[ **0** ] + nums[ **1** ] = 2 + 7 = 9,
16      return [ **0** , **1** ].
17
18
19
20
21
22  **Tags:** Array, Hash Table
23
24  **Difficulty:** Easy
25  |
26  ##
27
28  [title]: https://leetcode.com/problems/two-sum
29
```

Data flow:

Text file

```
1   # [Two Sum][title]
2
3   ## Description
4
5   Given an array of integers, return **indices** of the two numbers such that
6   they add up to a specific target.
7
8   You may assume that each input would have **_exactly_** one solution, and you
9   may not use the _same_ element twice.
10
11  **Example:**
12
13      Given nums = [2, 7, 11, 15], target = 9,
14
15      Because nums[ **0** ] + nums[ **1** ] = 2 + 7 = 9,
16      return [ **0** , **1** ].
17
18
19
20
21
22  **Tags:** Array, Hash Table
23
24  **Difficulty:** Easy
25  |
26  ##
27
28  [title]: https://leetcode.com/problems/two-sum
29
```

Python

```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 Madison	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 Princip	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	C	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 鉄?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
Combat Engineer - Construction and Engin		La Banta 46250

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()
```

Conclusion

- Arrays are the most popular topic that companies test candidates on

- **Medium** difficulty problems are the primary focus on a majority

- Practice a lot of problems but also check their relevance

- And most importantly,

Be like Bob

Thank you!

References:

Main data set:
Scraped from leetcode.com

Supporting dataset:

monster.com

Method:

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

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

Visualization:

Plotly, geopandas, pandas, wordcloud, shapefile, shapely