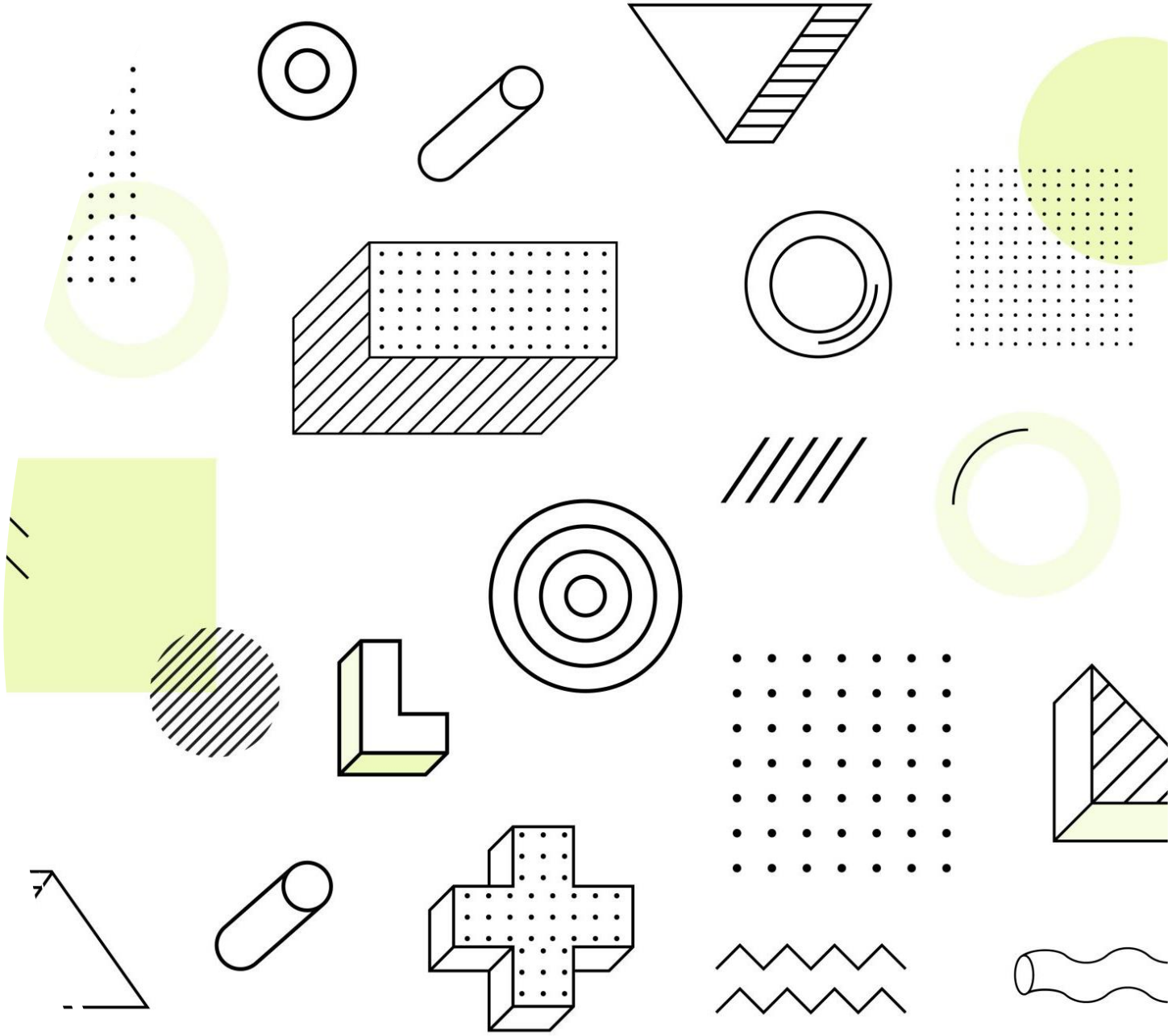
Data Science Challenge Library



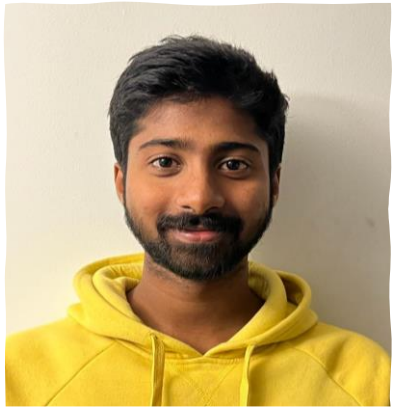
Project Summary

- The ***Data Science Challenge Library*** is a small-scale project aimed at creating a website to host and track data science challenges.
- The project will consist of several functionalities that will enhance the learning experience for students in the field of data science.
- The main objective is to provide a platform where students can practice and improve their skills in Artificial intelligence and Machine learning.
- To increase user engagement the website will consist of gamified elements to track their progress, to track the leaderboard.

Our team

Jahirabbas Joynal Khan
BSc in Computer Engineering

Web developer
- JavaScript, React JS
- MySQL



Kavuri Mahesh
BSc in Electronics & Communication Engineering

System Engineer
- Python, R, React
- Cloud Technologies (AWS, Azure)
- Tableau, SQL

Rushda Najeeb
BSc(Hons) in Software Engineering

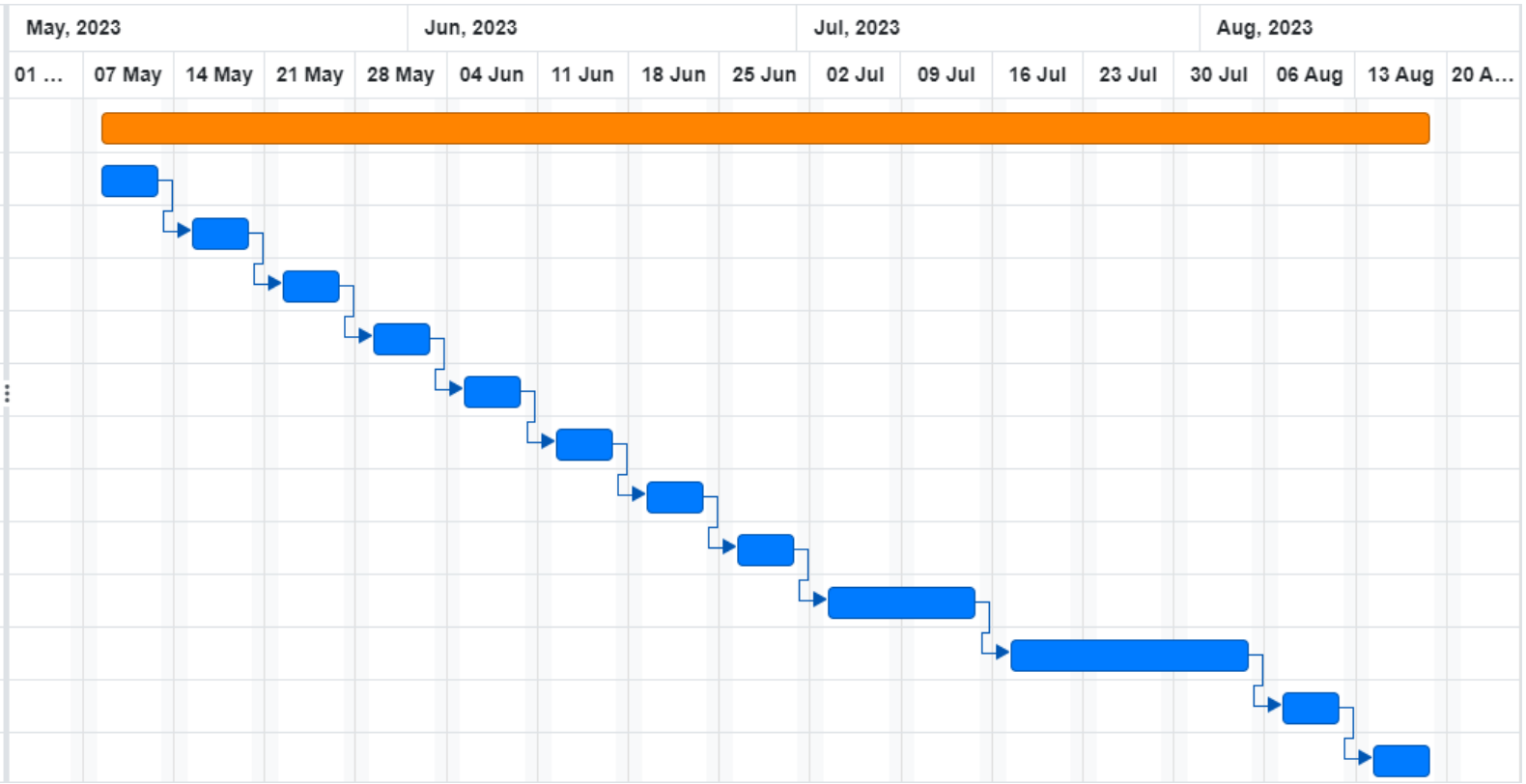
Software engineer
- Java, Javascript, Angular, Python
- PLSQL, SQL
- Mobile App Development

Pranavi Satheesan
BSc(Hons) in Data Science

Assistant Lecturer in Computing Faculty
-Python, R
-Tableau, PowerBI
-Java, Javascript
-SQL

Gantt Chart

ID	Name
1	Data Science Challenge Library
2	Selecting Project
3	Researching and Gathering requirements for t…
4	Defining project scope, budget, and preparing f…
5	Create a user-friendly website layout and interf…
6	Implement a user registration form to allow ne…
7	Design a database schema to store challenge …
8	Design a user-friendly form for users to submit …
9	Implement the "Case Studies" category as the …
10	Create a search bar to allow users to search fo…
11	Develop an algorithm or rules to verify user-su…
12	Testing all the functionalities
13	Making Final report and presentation



Project Progress

Targeted Audience (Students)

Demographics

Age: Students from high school to university level (typically aged 15-24).

Education Level: Students pursuing various academic disciplines.

Location: Nationwide or specific regions/countries depending on the scope of your project.

Preferences

- Tech-savvy: Students comfortable with using technology, including computers and the internet.
- Online Interactions: Students who actively engage in online platforms and social media.
- Learning Styles: Students who prefer interactive and hands-on learning experiences.

Goals

Skill Development: Students aim to enhance their programming skills and expand their knowledge in coding languages and concepts.

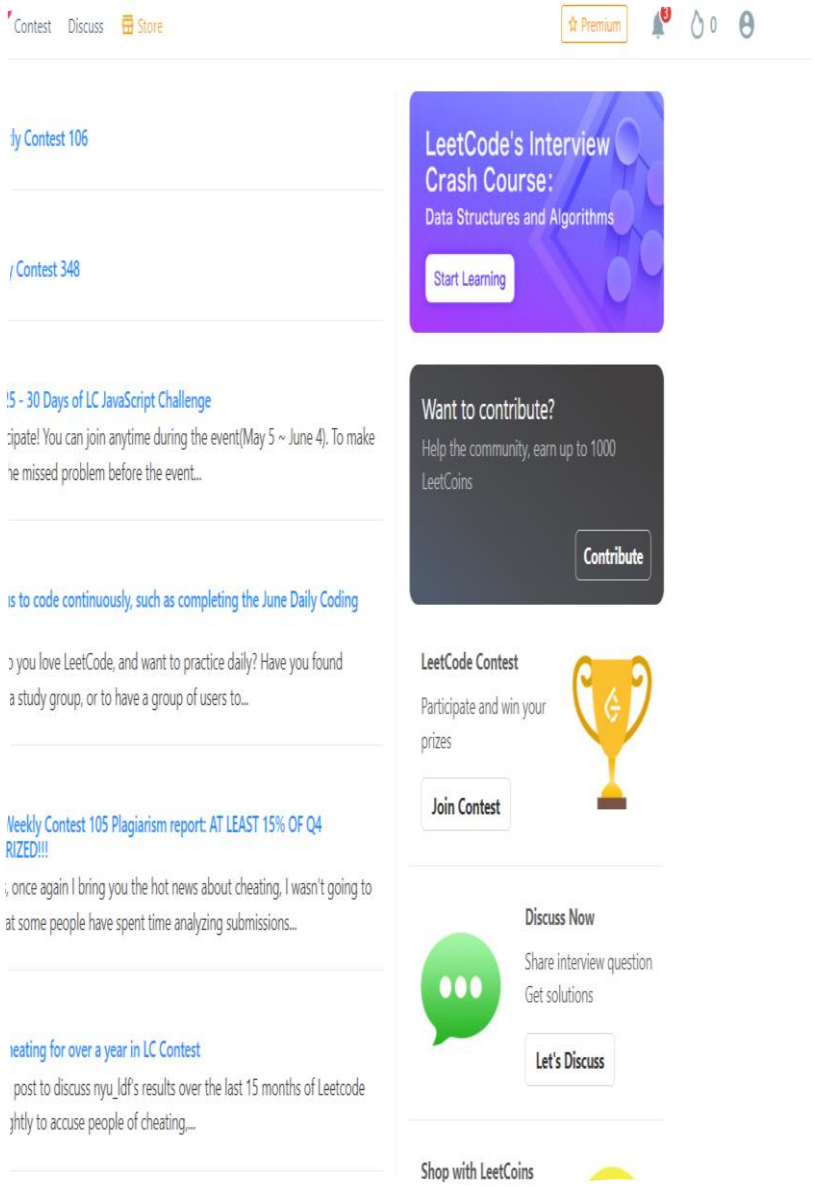
Academic Success: Students strive to excel in their academic courses that involve coding or programming components.

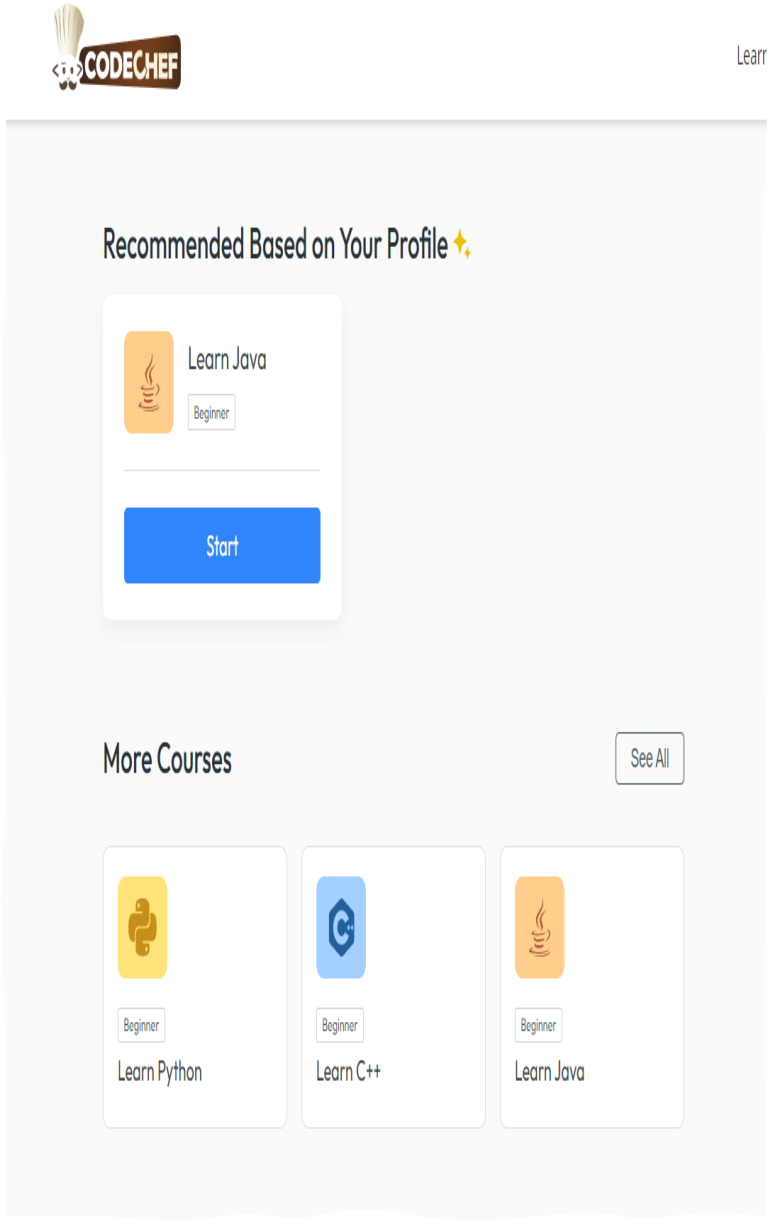
Career Advancement: Students aspire to secure internships, co-op positions, or job opportunities in the tech industry.

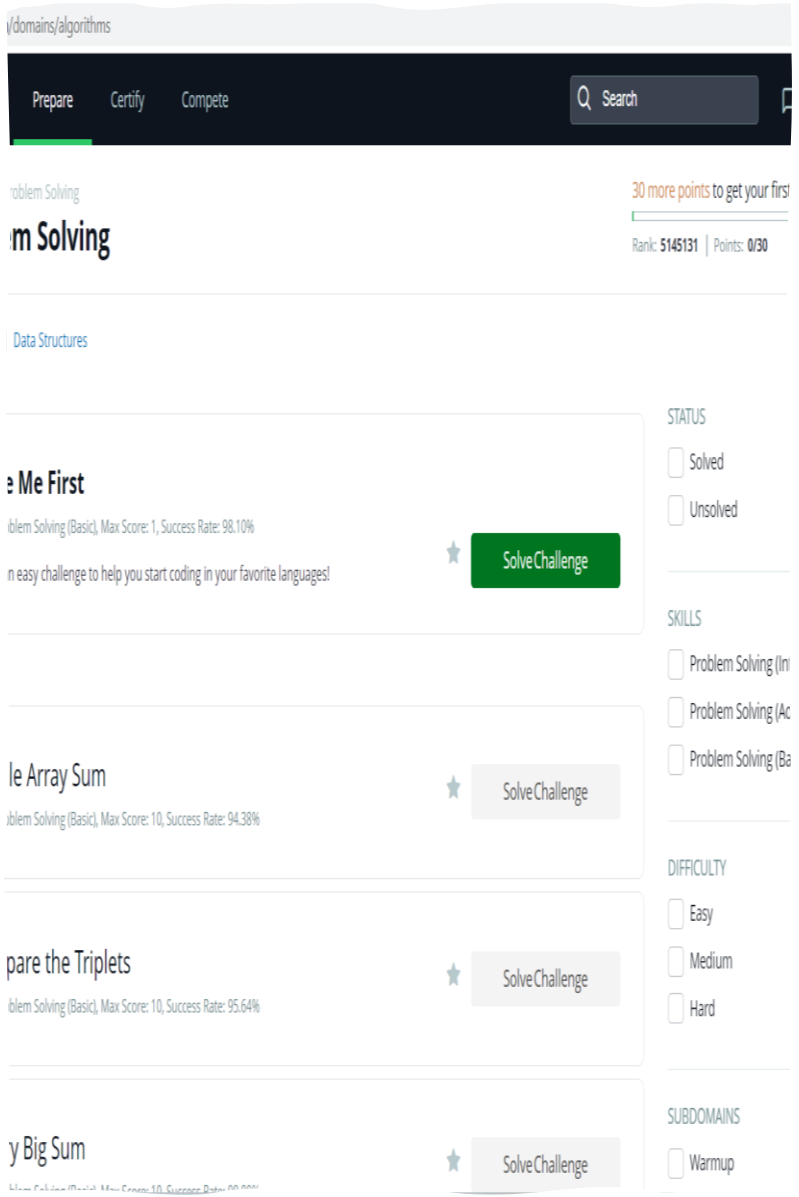
Challenges

- Time Constraints: Students may have limited time due to their academic commitments, so the platform should offer flexibility in completing coding challenges.
- Motivation and Engagement: Keeping students motivated and engaged throughout the coding challenges can be a challenge. Consider incorporating gamification elements or rewards to encourage participation.

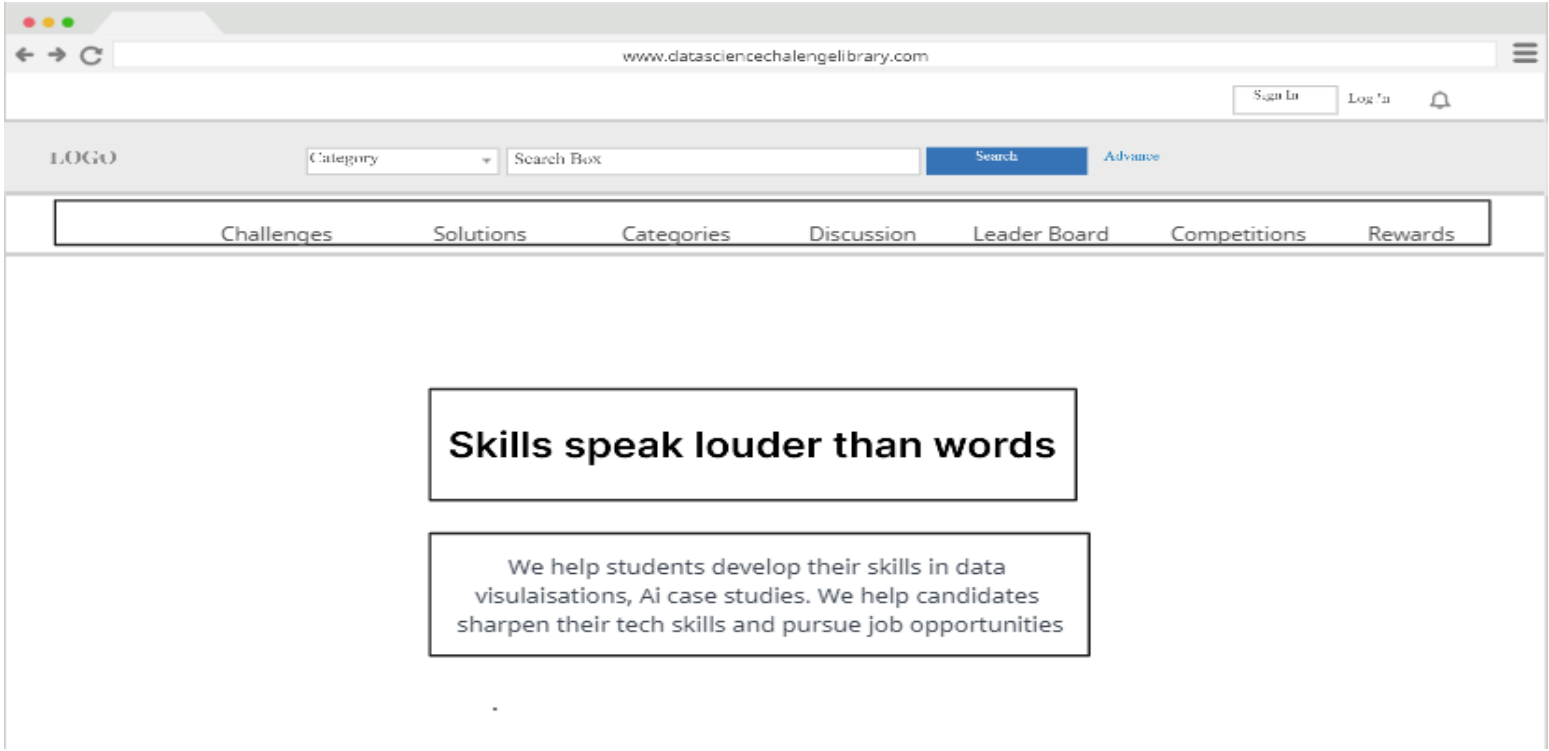
Researching Existing Website


Contest
Discuss
Store
Premium
ly Contest 106
Contest 348
5 - 30 Days of LC JavaScript Challenge
cipate! You can join anytime during the event(May 5 ~ June 4). To make
he missed problem before the event...
is to code continuously, such as completing the June Daily Coding
you love LeetCode, and want to practice daily? Have you found
a study group, or to have a group of users to...
Weekly Contest 105 Plagiarism report: AT LEAST 15% OF Q4
RIZED!!!
, once again I bring you the hot news about cheating, I wasn't going to
at some people have spent time analyzing submissions...
eating for over a year in LC Contest
post to discuss nyu_ldf's results over the last 15 months of Leetcode
ghtly to accuse people of cheating,...
LeetCode's Interview
Crash Course:
Data Structures and Algorithms
Start Learning
Want to contribute?
Help the community, earn up to 1000
LeetCoins
Contribute
LeetCode Contest
Participate and win your
prizes
Join Contest
Discuss Now
Share interview question
Get solutions
Let's Discuss
Shop with LeetCoins

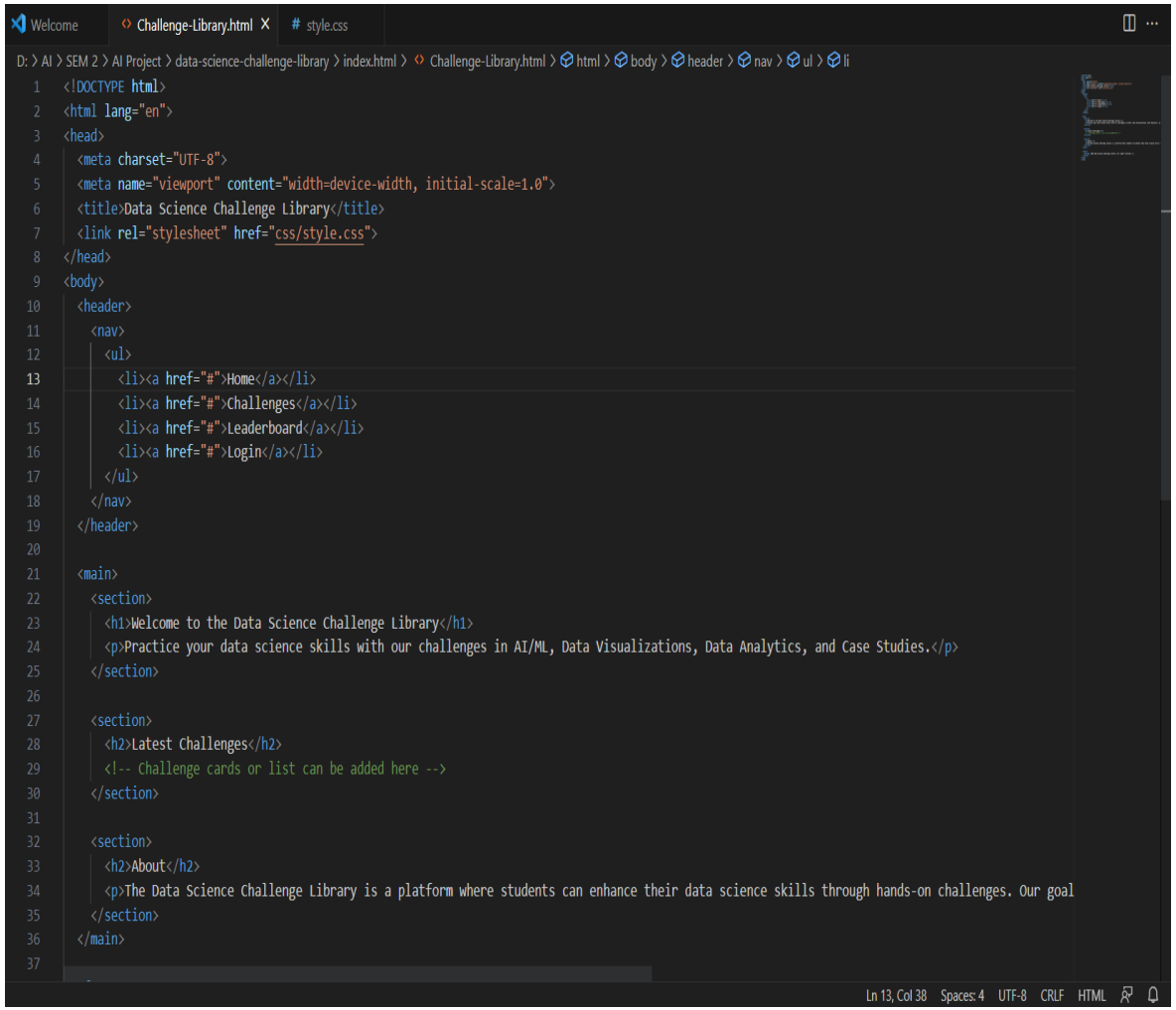

CODECHEF
Learr
Recommended Based on Your Profile
Learn Java
Beginner
Start
More Courses
See All
Beginner
Learn Python
Beginner
Learn C++
Beginner
Learn Java

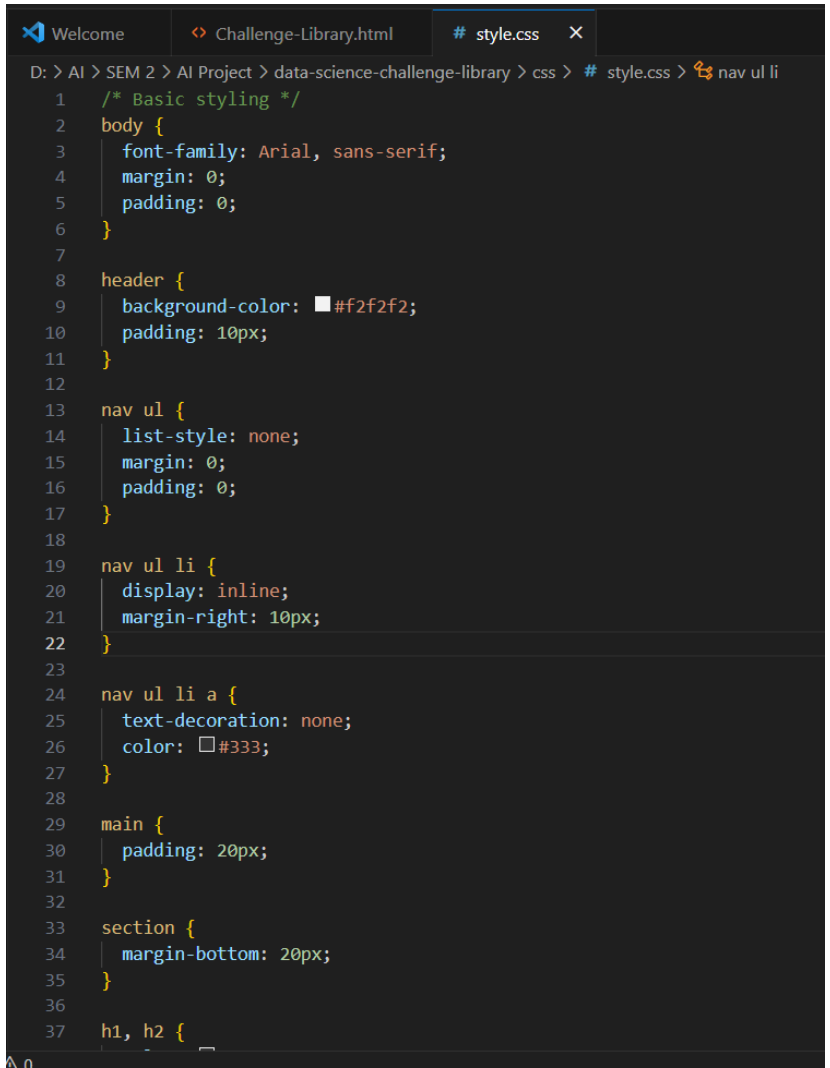

/domains/algorithms
Prepare
Certify
Compete
Search
oblem Solving
m Solving
30 more points to get your first
Rank: 5145131 | Points: 0/30
Data Structures
e Me First
blem Solving (Basic), Max Score: 1, Success Rate: 98.10%
n easy challenge to help you start coding in your favorite languages!
Solve Challenge
le Array Sum
blem Solving (Basic), Max Score: 10, Success Rate: 94.38%
Solve Challenge
pare the Triplets
blem Solving (Basic), Max Score: 10, Success Rate: 95.64%
Solve Challenge
y Big Sum
Solve Challenge
STATUS
Solved
Unsolved
SKILLS
Problem Solving (Int
Problem Solving (Ac
Problem Solving (Ba
DIFFICULTY
Easy
Medium
Hard
SUBDOMAINS
Warmup

Website - Wireframe



Home page initial layout code





Data Cleansing

```
[2]  import matplotlib.pyplot as plt
     import numpy as np
     import pandas as pd
     import seaborn as sns

     import warnings
     from sklearn.exceptions import ConvergenceWarning
     warnings.filterwarnings('ignore', category=FutureWarning)
     warnings.filterwarnings('ignore', category=DeprecationWarning)
     warnings.filterwarnings('ignore', category=ConvergenceWarning)

     %matplotlib inline
```

```
[3]  q_data = pd.read_excel('data.xlsx')
     q_data.head()
```

	ID	Type	Questions	Answers
0	1	case study	You are given a train data set having 1000 col...	Processing a high dimensional data on a limite...
1	2	case study	Is rotation necessary in PCA? If yes, Why? Wha...	Yes, rotation (orthogonal) is necessary becaus...
2	3	case study	You are given a data set. The data set has mi...	This question has enough hints for you to star...
3	4	case study	You are given a data set on cancer detection. ...	If you have worked on enough data sets, you sh...
4	5	case study	You are working on a time series data set. You...	Time series data is known to posses linearity....

```
[4]  #Drop ID Column
     q_data = q_data.drop('ID', axis=1)
```

```
[5]  #Dsiplay the first 5 rows
     q_data.head()
```

	Type	Questions	Answers
0	case study	You are given a train data set having 1000 col...	Processing a high dimensional data on a limite...
1	case study	Is rotation necessary in PCA? If yes, Why? Wha...	Yes, rotation (orthogonal) is necessary becaus...
2	case study	You are given a data set. The data set has mi...	This question has enough hints for you to star...

✓ 0s completed at 8:45PM

Data Visualization

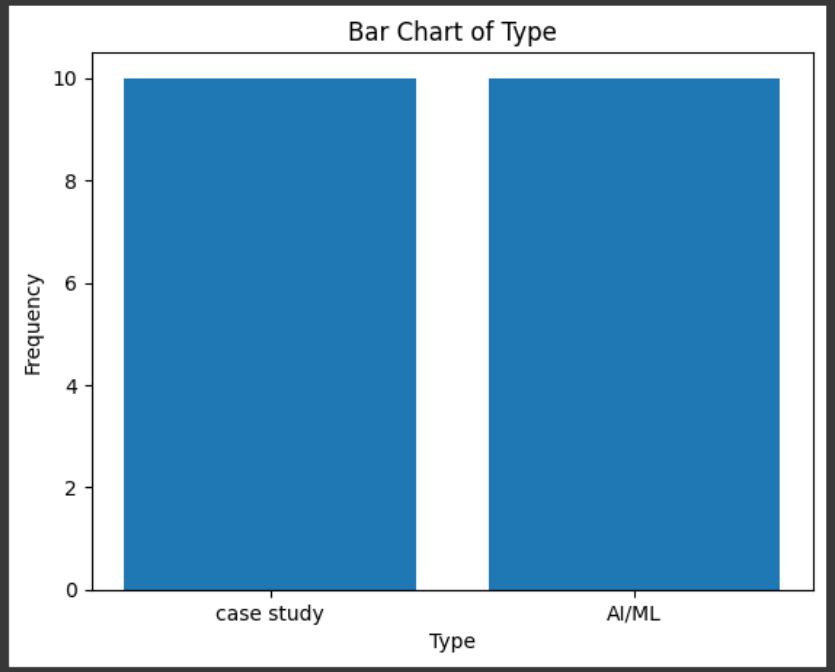
```
[9] #Display the shape of the data
    q_data.shape

    print('Number of questions: ', q_data['Questions'].size)
```

```
Number of questions:  20
```

```
    import pandas as pd

    # Assuming your data is stored in a DataFrame named "df"
    unique_values = q_data['Type'].unique()

    print('Questions categories:')

    # Print the unique values
    for value in unique_values:
        print(value)
```

```
Questions categories:
case study
AI/ML
```

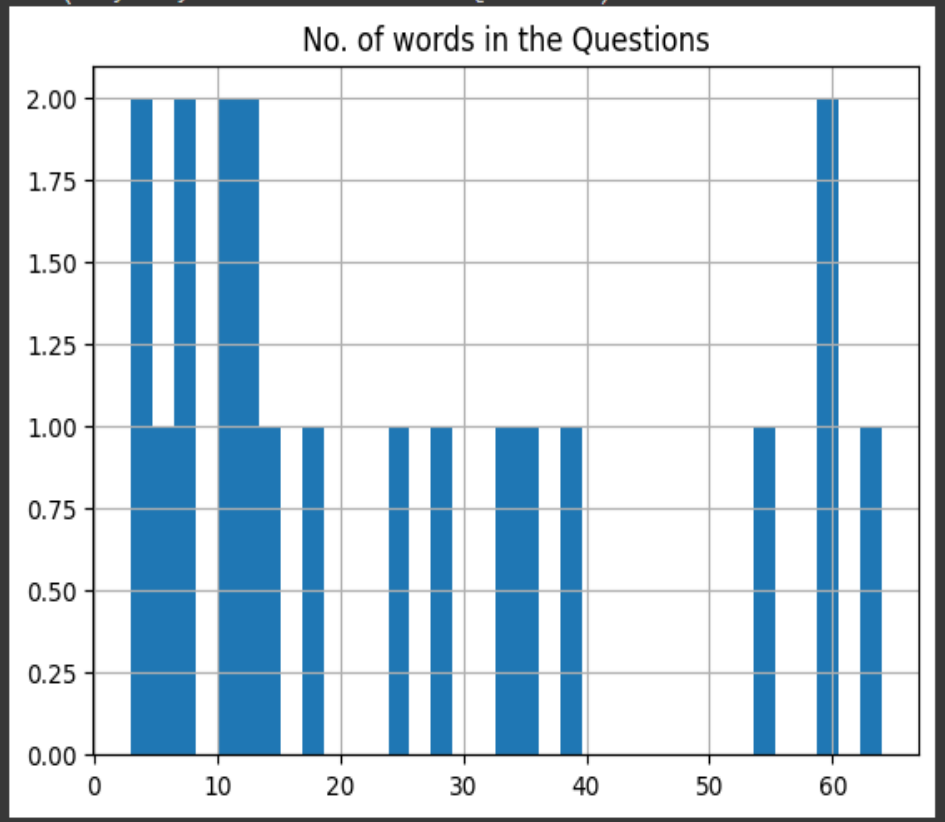
```
[ ] # Count the frequency of each unique value in the "Type" column
    value_counts = q_data['Type'].value_counts()

    # Create a bar chart
    plt.bar(value_counts.index, value_counts.values)

    # Add labels and title
    plt.xlabel('Type')
    plt.ylabel('Frequency')
    plt.title('Bar Chart of Type')

    # Display the chart
    plt.show()
```



```
[ ] q_data['qlength'] = q_data['Questions'].apply(lambda x : len(x.split(" ")))
    q_data['alength'] = q_data['Answers'].apply(lambda x : len(x.split(" ")))
```

```
[ ] q_data['qlength'].hist(bins=35)
    plt.title("No. of words in the Questions")
```

```
Text(0.5, 1.0, 'No. of words in the Questions')
```

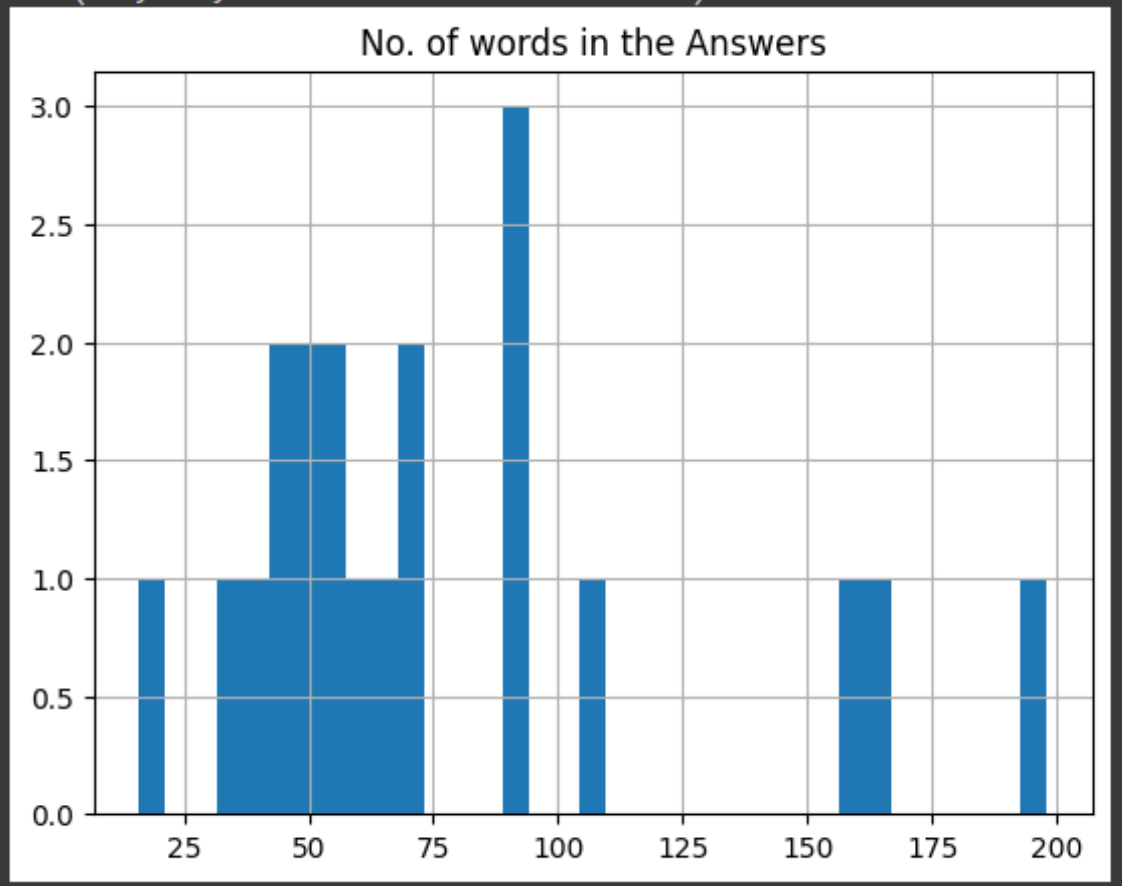



```
[ ] print('Mean word count of questions is %s' % np.mean(q_data['qlength']))
```

```
Mean word count of questions is 25.1
```

```
[ ] q_data['alength'].hist(bins=35)
    plt.title("No. of words in the Answers")
```

```
Text(0.5, 1.0, 'No. of words in the Answers')
```




```
[ ] print('Mean word count of answers is %s' % np.mean(q_data['alength']))
```

```
Mean word count of answers is 78.0
```

Consumption Habits – Empowering Students

Online Content Consumption:

- Our platform acknowledges that students spend a significant amount of time-consuming online content.
- We provide a rich collection of educational resources, interactive content, and learning materials to cater to their diverse academic needs.

Social Media Engagement:

- We understand the importance of social media for students' connectivity and sharing experiences.
- Our platform integrates social media features, enabling students to connect with peers, share achievements, and seek recommendations within our community.

Continued…

Peer Recommendations and Reviews:

- We leverage the power of peer recommendations and online reviews to enhance students' decision-making process.
- Our platform integrates community features where students can share their experiences, offer insights, and provide feedback on resources and challenges.

Budget Considerations:

- We understand the financial constraints faced by students and aim to provide affordable solutions.
- Our platform offers flexible pricing options, including student discounts and free tiers, ensuring that quality resources are accessible to all students.

Status report

- Researching about the existing websites **– All the members**
- Collecting question banks **– All the members**
- Reviewing technologies used by existing websites **– All the members**
- Preparing the status report for the current week **– All the members**

- Collecting datasets and performing necessary preprocessing techniques and data visualizations – **Rushda, Pranavi**
- Designing wireframe - **Jahir, Mahesh**

Upcoming week project plan

- Building a Login page for our website
- Unit testing for login functionality.
- Designing a database
- Finding more AI related case studies

Meeting Plans

- Tuesday-5hrs (including class time)
- Saturday 3hrs
- Sunday 3hrs

- Offline- In class Communication
- Online-Communication through Teams Channel & Group chats

Question & Answers

- **What is the primary goal of the website?** The primary goal of the website is to provide a platform for practicing data science challenges, allowing users to apply their skills and knowledge in real-world scenarios.
- **What database management system should be employed to store user information and challenge data?** Challenge data and user information can be stored in a relational database management system (e.g., PostgreSQL or MySQL) for efficient retrieval and management.
- **What programming languages, frameworks, or libraries should be used for building the website?** The website can be built using a combination of HTML, CSS, and JavaScript for the frontend, and a backend framework such as Django or Flask, utilizing Python for server-side logic.
- **What is the purpose of a discussion forum?** The discussion forum is given as functionality so that users can share their knowledge on any difficulties faced solving the challenge.
- **Do users receive feedback?** After submitting the answers for the challenges the user will be given a reward and automatically given feedback according to their answer.

Thank you