



Feynn Labs AI Company

CodeMate

AI-Enabled Coding Tutor and Learning Platform

Muskan Sharma

Date: 04/07/2023

Abstract:

The AI-Enabled Coding Tutor and Learning Platform is an innovative online learning solution that harnesses the power of artificial intelligence and machine learning to provide personalized coding instruction and guidance. The platform offers a comprehensive set of features designed to enhance the learning experience and foster coding proficiency. Learners are guided through personalized learning paths, receive real-time code analysis and feedback, engage in interactive coding challenges, and collaborate with a community of peers. The platform aims to empower learners to develop practical coding skills, gain confidence, and excel in their coding journey.

Problem Statement:

The traditional approach to coding education faces several challenges that hinder effective learning experiences and skill development. These challenges include:

1. Limited Personalization:

Traditional coding education often follows a one-size-fits-all approach, neglecting the diverse learning styles, backgrounds, and skill levels of individual learners. This lack of personalization leads to disengagement, frustration, and suboptimal learning outcomes.

2. Insufficient Real-time Guidance:

Learners often encounter roadblocks and coding errors without immediate guidance or support. This can result in prolonged problem-solving times, reduced motivation, and limited opportunities for iterative learning.

3. Ineffective Feedback Mechanisms:

Feedback on coding exercises and projects is typically delayed or inadequate. Learners may not receive detailed explanations of their mistakes, preventing them from understanding and correcting their errors in a timely manner.

4. Limited Interactive Learning Opportunities:

Static learning materials, such as textbooks and video tutorials, often fail to provide hands-on coding experiences and interactive practice environments. This hampers the development of practical coding skills and stifles creativity.

5. Lack of Peer Interaction and Collaboration:

Traditional coding education methods often lack opportunities for learners to engage with their peers, share ideas, collaborate on projects, and receive feedback from a supportive community. This limits social learning and peer-to-peer knowledge exchange.

6. Inadequate Career Guidance:

Many coding learners struggle to navigate the complexities of the coding industry and find suitable career paths. The lack of career guidance, interview preparation, and job placement support inhibits learners' ability to effectively transition into the professional coding world.

Addressing these challenges requires an AI-enabled coding tutor and learning platform that offers personalized learning paths, real-time assistance, interactive exercises, comprehensive feedback mechanisms, collaborative opportunities, and career guidance. By leveraging AI/ML technologies, the platform can revolutionize coding education by providing learners with tailored experiences, timely support, engaging practice environments, and a pathway to successful coding careers.

Market/Customer/Business Need Assessment:

The market for an AI-enabled coding tutor and learning platform is driven by several factors that indicate a strong need and demand for such a product. The assessment includes:

1. Growing Demand for Coding Skills:

The demand for coding skills is rapidly increasing in various industries. As technology continues to advance, there is a growing need for professionals who can develop software, build

websites, analyze data, and automate processes. This creates a significant market for coding education and training platforms.

2. Ineffective Traditional Coding Education:

Traditional coding education methods often struggle to meet the demands of the modern learner. The one-size-fits-all approach and limited personalization hinder the effectiveness of teaching coding skills. There is a need for more adaptive and personalized learning experiences that cater to individual learners' needs and learning styles.

3. Increased Accessibility to Online Learning:

The availability and accessibility of online learning platforms have witnessed significant growth in recent years. Learners are increasingly turning to online resources for self-paced learning, convenience, and flexibility. An AI-enabled coding tutor and learning platform can tap into this trend and provide a comprehensive and engaging online learning experience.

4. Demand for Real-time Guidance and Support:

Learners often encounter challenges and roadblocks while learning to code. They require immediate guidance and support to overcome obstacles and gain a deeper understanding of coding concepts. An AI-powered tutor that offers real-time assistance and feedback addresses this need and enhances the learning experience.

5. Rising Importance of Personalized Learning:

Personalized learning is gaining traction in the education sector. Learners expect tailored experiences that cater to their specific needs and preferences. An AI-enabled coding tutor can provide personalized learning paths, adaptive content delivery, and customized feedback, ensuring learners have a more engaging and effective learning journey.

6. Strong Coding Community and Collaboration:

The coding community values collaboration, peer learning, and knowledge sharing. Learners seek opportunities to interact with their peers, work on coding projects together, and receive feedback from a supportive community. An AI-enabled learning platform can facilitate collaboration, discussion forums, and coding challenges, fostering a vibrant and engaged coding community.

7. Career Advancement and Job Placement:

Many learners pursue coding skills to enhance their career prospects. They seek guidance on job placement, resume building, and interview preparation. A comprehensive learning platform that includes career guidance and support can provide learners with the necessary skills and resources to excel in their coding careers.

Considering these factors, there is a strong market need for an AI-enabled coding tutor and learning platform. The demand for personalized, interactive, and comprehensive coding education is growing, and learners are actively seeking effective solutions that provide real-time guidance, adaptive learning experiences, community engagement, and career support.

Target Specifications and Characterization:

The target specifications and customer characteristics for an AI-enabled coding tutor and learning platform can be defined as follows:

1. Learner Profile:

The primary target audience includes individuals who are interested in learning coding or improving their coding skills. This may include:

- Beginners with no prior coding experience.
- Students pursuing computer science or related fields.
- Professionals seeking to enhance their coding skills for career advancement.
- Self-taught coders looking to expand their knowledge and deepen their understanding.

2. Age Range:

The platform can cater to a wide age range, including:

- High school and college students.
- Young adults and working professionals.
- Coding enthusiasts of any age who want to learn or improve their coding abilities.

3. Skill Levels:

The platform should accommodate learners at different skill levels, including:

- Absolute beginners with no coding knowledge.
- Intermediate learners with some coding experience.
- Advanced learners looking to deepen their expertise in specific programming languages or domains.

4. Learning Styles and Preferences:

The platform should offer flexibility and adaptability to cater to different learning styles and preferences. This includes providing visual, auditory, and interactive learning materials, as well as options for self-paced learning or guided instruction.

5. Technical Requirements:

The platform should be accessible across multiple devices, including desktop computers, laptops, tablets, and smartphones. It should support popular web browsers and have a responsive user interface for seamless user experience on different screen sizes.

6. User Experience:

The platform should prioritize a user-friendly and intuitive interface, making it easy for learners to navigate through the learning materials, access exercises, and interact with the AI tutor. Clear instructions, well-organized content, and a visually appealing design contribute to an engaging and enjoyable user experience.

7. Supportive Community:

The platform should foster a supportive community where learners can interact with each other, share ideas, collaborate on projects, and receive feedback. Community features such as discussion forums, chat platforms, and peer code reviews enhance the learning experience and provide opportunities for social learning.

8. Scalability:

The platform should have the ability to scale its resources and handle a growing user base. It should be designed to accommodate a large number of concurrent users without compromising performance or user experience.

By considering these specifications and customer characteristics, the AI-enabled coding tutor and learning platform can effectively target a diverse range of learners, provide a personalized learning experience, and meet the needs of individuals at different skill levels and learning preferences.

Benchmarking:

To understand the competitive landscape and assess existing products/services in the market, it is important to benchmark the AI-enabled coding tutor and learning platform against alternative solutions. Here are some popular products/services in the market that can be considered for benchmarking:

1. Codecademy:

Codecademy is an online learning platform that offers interactive coding courses and tutorials. It provides a wide range of programming languages and follows a hands-on approach to learning. Users can practice coding exercises, receive immediate feedback, and track their progress. However, Codecademy may lack personalized learning paths and real-time assistance compared to an AI-enabled coding tutor.

2. Udacity:

Udacity is an online learning platform that offers nanodegree programs in various domains, including coding and programming. It provides structured curriculum, project-based learning, and access to industry mentors. While Udacity offers comprehensive learning experiences, it may not offer the same level of personalized guidance and adaptive learning as an AI-enabled coding tutor.

3. Coursera:

Coursera is a widely recognized online learning platform that offers a wide range of courses, including coding and computer science. It collaborates with renowned universities and institutions to deliver high-quality content. While Coursera provides a diverse learning experience, it may lack real-time assistance and personalized learning paths compared to an AI-enabled coding tutor.

4. Pluralsight:

Pluralsight is an online learning platform that focuses on technology and software development skills. It offers video-based courses, assessments, and learning paths. Pluralsight provides a vast library of content, but it may not offer the same level of interactive coding exercises, real-time assistance, and personalized guidance as an AI-enabled coding tutor.

5. GitHub Learning Lab:

GitHub Learning Lab is a platform that offers interactive coding tutorials and projects focused on Git and GitHub. It provides hands-on learning experiences for version control and collaboration. While GitHub Learning Lab is valuable for specific topics, it may not cover a wide range of coding concepts or offer the same level of personalized learning and assistance as an AI-enabled coding tutor.

When benchmarking against these alternative products/services, the **AI-enabled coding tutor and learning platform can differentiate itself by offering personalized learning paths, real-time assistance, AI-powered code autocompletion, interactive coding exercises, comprehensive feedback mechanisms, and a supportive community**. The focus on adaptive learning, immediate guidance, and AI-driven features will provide a unique value proposition to learners in the coding domain.

Applicable Patents

In the field of AI and machine learning, there are several patents related to specific algorithms, techniques, or applications. These patents can cover various aspects of AI-enabled coding education, including natural language processing, recommendation systems, personalized learning algorithms, and interactive coding environments. Some of the relevant areas that may have applicable patents include:

1. Natural Language Processing (NLP) Techniques:

Patents related to NLP algorithms and techniques used for analyzing and processing human language can be applicable. These patents may cover aspects such as language understanding, sentiment analysis, question answering, and dialogue systems.

2. Recommendation Systems:

Patents related to recommendation algorithms and techniques used to suggest personalized learning paths, coding exercises, and learning resources based on the learner's profile, preferences, and progress can be relevant. These patents may cover collaborative filtering, content-based filtering, or hybrid recommendation systems.

3. Interactive Coding Environments:

Patents related to interactive coding environments that provide real-time assistance, code autocompletion, and debugging support can be applicable. These patents may cover features such as intelligent code editors, code analysis, and code execution environments.

4. Machine Learning Algorithms:

Patents related to specific machine learning algorithms or models used in the AI-enabled coding tutor and learning platform can be relevant. These patents may cover areas such as supervised learning, unsupervised learning, reinforcement learning, or deep learning architectures.

5. User Interface and Experience:

Patents related to innovative user interface designs, interactive features, and gamification elements in the AI-enabled coding tutor and learning platform can be applicable. These patents may cover aspects such as intuitive user interfaces, visualization techniques, and adaptive learning interfaces.

It's important to note that the availability of patents and their relevance to a specific AI-enabled coding tutor and learning platform can vary. Conducting a thorough patent search and consulting with legal professionals specializing in intellectual property would be essential to determine the specific patents that may apply.

Applicable Regulations

The regulations imposed on an AI-enabled coding tutor and learning platform can vary depending on the country or region where the product is being developed and deployed. Here are some potential areas of regulations that may apply:

1. Data Privacy and Protection:

Many countries have data protection regulations in place to safeguard the privacy of individuals' personal information. For example, the General Data Protection Regulation (GDPR) in the European Union sets requirements for collecting, processing, and storing personal data. Compliance with these regulations is crucial when handling user data in the learning platform.

2. Intellectual Property:

Intellectual property laws protect original works and inventions. It's important to ensure that the AI-enabled coding tutor and learning platform respects copyright laws and intellectual property rights. This includes proper attribution of code snippets, adherence to licensing terms for open-source libraries, and avoiding infringement of third-party copyrights.

3. Accessibility:

Accessibility regulations ensure that digital products and services are accessible to individuals with disabilities. For example, the Web Content Accessibility Guidelines (WCAG) provide standards for creating accessible web content. Compliance with accessibility guidelines ensures that the learning platform is usable by people with visual, hearing, motor, or cognitive impairments.

4. Education Regulations:

Educational institutions and platforms often need to comply with specific regulations related to education. These may include accreditation requirements, licensing, or compliance with educational standards set by relevant authorities. It's important to understand and comply with any regulations specific to the education sector in the target market.

5. Ethical Considerations:

While not legally binding, ethical considerations play a significant role in AI development. It is important to ensure that the AI-enabled coding tutor and learning platform adheres to ethical guidelines, such as fairness, transparency, and accountability in algorithmic decision-making. This includes avoiding bias in content or recommendations and being transparent about data usage and algorithmic processes.

It's crucial to thoroughly research and understand the regulations specific to the countries or regions where the AI-enabled coding tutor and learning platform will be deployed. Consulting with legal experts who specialize in technology, data protection, and education regulations in the target market can help ensure compliance and mitigate any legal or regulatory risks.

Applicable Constraints

1. Space Constraints:

Space constraints can be a consideration when developing and deploying an AI-enabled coding tutor and learning platform. This primarily applies to physical infrastructure requirements, such as server space or data centres to host the platform. Cloud-based solutions can help mitigate space constraints by leveraging infrastructure provided by cloud service providers.

2. Budget Constraints:

Budget constraints are an important consideration in the development of any product or service. The costs associated with developing the AI-enabled coding tutor and learning platform include software development, infrastructure costs, data acquisition or licensing, maintenance, marketing, and operational expenses. Budget constraints may influence decisions related to technology choices, team size, and overall project scope.

3. Expertise Constraints:

Developing an AI-enabled coding tutor and learning platform requires a multidisciplinary team with expertise in various areas. This may include software development, machine learning, natural language processing, user experience design, and educational content creation. Constraints may arise if there is a lack of expertise within the team or if additional resources or partnerships are required to bridge the skill gaps.

4. Technology Constraints:

Technology constraints can arise due to limitations or dependencies on specific frameworks, software libraries, or programming languages. Compatibility issues or licensing restrictions may affect the selection and integration of AI and ML tools. It's important to consider the technology constraints to ensure smooth implementation and avoid potential conflicts with existing systems or platforms.

5. Time Constraints:

Time constraints are often a factor in product development. The timeline for creating an AI-enabled coding tutor and learning platform can be influenced by factors such as market demand, competition, and the complexity of the features and algorithms to be implemented. It's important to manage time constraints effectively and prioritize key functionalities and user requirements to meet project milestones.

6. Regulatory and Compliance Constraints:

Compliance with legal and regulatory requirements, such as data privacy and protection laws, accessibility standards, and educational regulations, can impose constraints on the development and deployment of the AI-enabled coding tutor and learning platform. These constraints need to be carefully addressed to ensure legal compliance and avoid potential penalties or reputational damage.

Understanding and managing these constraints is essential for successful development and deployment of an AI-enabled coding tutor and learning platform. By identifying and addressing these constraints proactively, you can make informed decisions, allocate resources effectively, and mitigate any potential challenges or limitations during the development process.

Business Model

There are several potential monetization strategies for an AI-enabled coding tutor and learning platform. Here are some common business models to consider:

1. Subscription Model:

Offer a subscription-based pricing model where users pay a recurring fee to access the platform's features and content. Subscriptions can be offered on a monthly, quarterly, or annual basis, providing users with continuous access to the learning resources, personalized guidance, and additional premium features.

2. Freemium Model:

Implement a freemium model where the basic features and content of the learning platform are offered for free, while advanced features, personalized learning paths, or premium content are available for a fee. This model allows users to experience the platform before deciding to upgrade to the paid version for enhanced learning experiences.

3. Pay-per-Course Model:

Provide a pay-per-course model where users can select and pay for individual coding courses or learning modules. Each course or module is priced separately, allowing learners to choose specific topics or skills they want to focus on without committing to a full subscription.

4. Corporate Licensing:

Target corporate clients by offering licensing options for the AI-enabled coding tutor and learning platform. This allows organizations to provide their employees with access to the platform as part of their professional development initiatives. Licensing agreements can be customized based on the size of the organization and the number of users.

5. B2B Partnerships:

Collaborate with educational institutions, coding bootcamps, or other edtech companies to offer the AI-enabled coding tutor and learning platform as an integrated solution within their existing programs. This partnership model can involve revenue sharing or licensing agreements based on the number of users or usage metrics.

6. Sponsorships and Advertisements:

Explore opportunities for sponsorships or partnerships with relevant technology companies or coding-related brands. This can involve displaying targeted advertisements or sponsored content within the learning platform. Care should be taken to ensure that the advertisements do not interfere with the user experience or compromise the integrity of the learning environment.

7. Data Licensing and Analytics:

Leverage the data collected from user interactions and learning patterns to provide insights and analytics to educational institutions, corporate clients, or researchers. This can involve anonymized data analysis, benchmarking reports, or customized learning analytics solutions.

Concept Development :

1. Identify the Target Audience:

Start by identifying the target audience for the AI-enabled coding tutor and learning platform. Consider whether you want to cater to beginners, intermediate learners, or advanced programmers. Determine if you want to focus on a specific programming language, technology stack, or coding domain.

2. Assess Existing Solutions:

Research and assess existing coding tutor and learning platforms. Identify their strengths, weaknesses, and areas where they may be lacking. Look for gaps in the market or opportunities to offer unique features or a better learning experience.

3. Understand User Pain Points:

Conduct surveys, interviews, or user research to understand the pain points and challenges faced by learners in the coding domain. Identify common difficulties, areas where learners struggle, and the specific needs they have when it comes to learning coding.

4. Brainstorm Innovative Features:

Brainstorm and generate ideas for innovative features that can enhance the learning experience. Consider leveraging AI and ML technologies to provide personalized learning paths, intelligent code autocompletion, real-time code analysis, automated code review, interactive coding challenges, or virtual coding mentors.

5. Gamification and Engagement:

Explore the concept of gamification to make learning coding more engaging and enjoyable. Consider incorporating elements such as badges, points, leaderboards, coding competitions, or virtual rewards to incentivize progress and motivate learners.

6. Collaborative Learning:

Consider incorporating features that facilitate collaboration and peer learning. This can include discussion forums, code-sharing platforms, mentorship programs, or group coding projects. Encourage learners to interact, share knowledge, and learn from each other.

7. Personalized Learning Paths:

Leverage AI algorithms to analyze learners' skills, progress, and interests. Generate personalized learning paths tailored to each individual's needs and goals. Provide recommendations for courses, coding exercises, and resources based on their proficiency level and areas of improvement.

8. Real-world Project Integration:

Provide opportunities for learners to work on real-world projects or simulations to apply their coding skills in practical scenarios. Incorporate industry-relevant projects, case studies, or coding challenges that reflect the challenges they might encounter in their future careers.

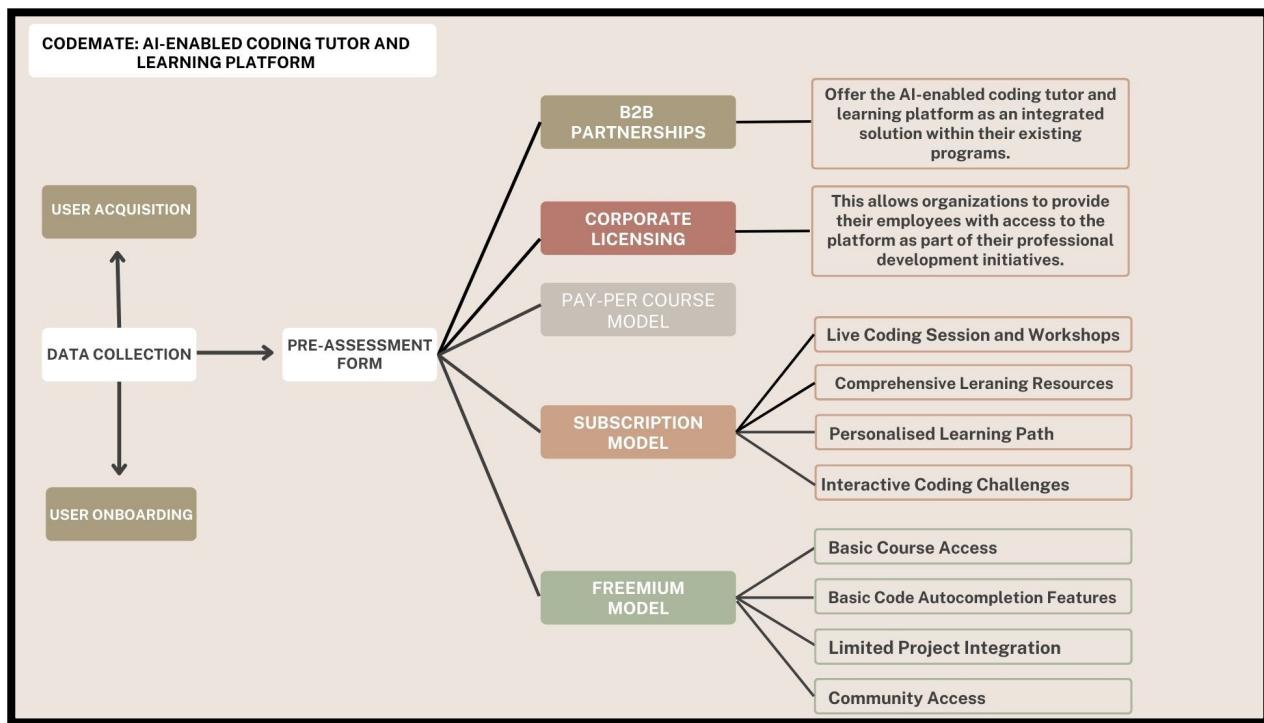
9. Seamless User Experience:

Emphasize a seamless and intuitive user experience in the design of the platform. Ensure that the interface is user-friendly, easy to navigate, and visually appealing. Incorporate interactive elements, visualizations, and progress tracking to enhance the learning journey.

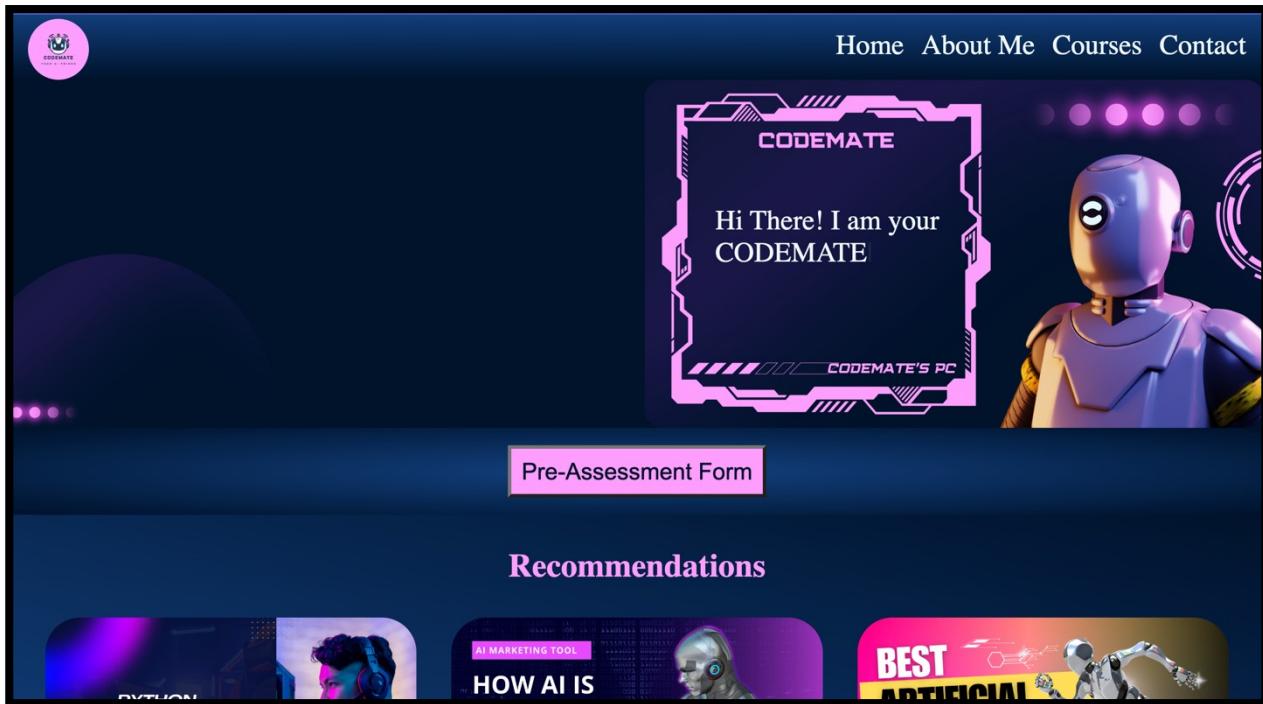
10. Continuous Improvement:

Plan for ongoing updates and improvements to the platform based on user feedback and emerging trends in the coding industry. Incorporate mechanisms for gathering user feedback, conducting user testing, and implementing iterative enhancements to keep the platform relevant and up-to-date.

Schematic Diagram:



User Interface:



Product Details:

1. How does it work?

The AI-Enabled Coding Tutor and Learning Platform leverages artificial intelligence and machine learning to enhance the learning experience for coding enthusiasts. Here's a high-level overview of how it works:

- a. User Registration and Profile Creation:** Learners create an account on the platform and set up their profiles, specifying their coding goals, proficiency level, and areas of interest.
- b. Personalized Learning Paths:** The platform analyzes user data and generates personalized learning paths tailored to each learner's needs. It recommends relevant courses, coding exercises, and resources based on the learner's proficiency level, goals, and areas for improvement.
- c. Interactive Learning Modules:** Learners engage in interactive coding challenges, hands-on exercises, and real-world projects to practice and apply their coding skills. They receive real-time code analysis, feedback, and suggestions for improvement.
- d. Community Interaction:** Learners can participate in discussion forums, code-sharing platforms, and mentorship programs to collaborate with peers, seek assistance, and gain insights from experienced coders.

e. Progress Tracking and Gamification: The platform tracks learners' progress, achievements, and milestones. Gamification elements such as badges, points, and leaderboards motivate learners and foster a sense of accomplishment.

f. Continuous Improvement and Updates: The platform evolves continuously, incorporating user feedback, industry trends, and emerging technologies. Updates, new coding challenges, and additional learning resources are regularly provided to keep the learning experience up-to-date.

2. Data Sources:

The AI-Enabled Coding Tutor and Learning Platform may utilize various data sources, including:

- **User Profile Data:** Information provided by learners during registration, such as coding experience, goals, and preferences.
- **Learning Progress Data:** Data on learners' performance, progress, and completion of coding challenges and exercises.
- **Code Repositories:** Access to code repositories for retrieving sample code, best practices, and real-world coding examples.
- **Community Data:** Data generated from learner interactions, discussions, and collaborations within the platform's community features.

3. Algorithms, Frameworks, Software, etc. Needed:

The AI-Enabled Coding Tutor and Learning Platform may require the following algorithms, frameworks, software, and technologies:

- **Machine Learning Algorithms:** Algorithms for personalized learning path generation, code analysis, and recommendation systems.
- **Natural Language Processing (NLP):** NLP techniques for code analysis, autocompletion, and generating feedback.
- **Programming Languages:** Support for various programming languages and frameworks to cater to different coding domains.
- **Web Development Technologies:** HTML, CSS, JavaScript, and frameworks like React or Angular for building the user interface.
- **Database Management System:** A database system to store user data, progress tracking, and content management.

4. Team Required to Develop:

The development team for the AI-Enabled Coding Tutor and Learning Platform may include the following roles:

- **Project Manager:** Oversees the overall development process, coordinates team efforts, and ensures timely delivery.
- **Software Engineers:** Responsible for developing the platform's backend, frontend, and database management components.
- **Data Scientists:** Work on algorithms, machine learning models, and data analysis to provide personalized learning experiences and code analysis.
- **UI/UX Designers:** Design the user interface, ensuring a seamless and intuitive user experience.
- **QA/Testers:** Conduct thorough testing to ensure the platform's functionality, performance, and security.

5. Cost:

The cost of developing an AI-Enabled Coding Tutor and Learning Platform can vary based on factors such as the complexity of features, the size of the development team, infrastructure requirements, and development time. It is recommended to conduct a detailed cost analysis considering the specific project scope and requirements to determine an accurate estimation.

External Search:

<https://unesdoc.unesco.org/ark:/48223/pf0000376709>

<https://hbr.org/2019/10/how-ai-and-data-could-personalize-higher-education>

<https://www.eklavyya.com/blog/ai-edtech-tools/>

DATASET:

Kaggle: <https://www.kaggle.com/datasets/kushagrathisside/student-skillset-analysis>

Github : <https://github.com/SharmaMuskan20/ML-CODEMATE-AI-Enabled-Coding-Tutor-and-Learning-Platform>

```
Importing Libraries
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: df=pd.read_csv("./Users/muskansharma/Desktop/ML internship task Feyynes lab/Studentdata (1).csv")

In [3]: df.head()

Out[3]:
      Unnamed: 0  2.Gender  3. Year of Study  4. Type of Schooling till 10th Grade:  2. Percentage in Class 10th?  3. Location of Class 10th School:  4. Type of Schooling for class 12th?:  5. Percentage in Class 12th?  6. Location of Class 12th School: ... 3. Exposure to Data structure and algorithms? 4. Exposure to GitHub ? 5. Are you associated with any developer Community? 6. Active in developer Communities 7. Participating in Hackathons and other competitions? 8. Have you made any Software based projects? 9. Have you made any Hardware based projects? 10. Have you ever pitched any idea? 11. Do you have any additional skills?  Achievements
0          0     Male       Third        CSE/IT    Normal Schooling >=80% and <90%   Town    Normal Schooling >=80% and <90%   Town   ...   Hard   Medium   Yes   Active Member   Participated   Yes   No   Yes   Cultural Activities   Scholarships,Hackathons,Research Papers,Project
1          1   Female      Second        CSIT    Schooling with Tuitions >= 90%   City    Schooling with Tuitions >=90%   City   ...   Medium   Medium   No   Moderately active   Not yet participated   Yes   No   No   No   Cultural Activities   Non
2          2     Male      Second        CSIT    Normal Schooling >= 90%   City    Coaching Based Schools >=90%   City   ...   Medium   Medium   No   Inactive Member   Not yet participated   Yes   No   No   Photography,Video Editing   Non
3          3     Male      Second        CSIT    Normal Schooling >=70% and <80%   City    Normal Schooling >=80% and <90%   City   ...   Medium   Easy   Yes   Moderately active   Not yet participated   Yes   Yes   No   None   Non
4          4     Male      Second        CSIT    Normal Schooling >=70% and <80%   City    Normal Schooling >=80% and <90%   City   ...   Medium   Hard   Yes   Active Member   Participated   Yes   No   Yes   None   Hackathons,Codin Competitions,Project
5 rows × 25 columns

In [4]: #renaming columns for simplification
df.columns

Out[4]: Index(['Unnamed: 0', '2.Gender', '3. Year of Study', '4. Branch', '1. Type of Schooling till 10th Grade:', '2. Percentage in Class 10th?', '3. Location of Class 10th School:', '4. Type of Schooling for class 12th?:', '5. Percentage in Class 12th?', '6. Location of Class 12th School:', '...', '3. Exposure to Data structure and algorithms?', '4. Exposure to GitHub ?', '5. Are you associated with any developer Community?', '6. Active in developer Communities', '7. Participating in Hackathons and other competitions?', '8. Have you made any Software based projects?', '9. Have you made any Hardware based projects? (Arduino, Raspberry Pi, Robots or any other)?', '10. Have you ever pitched any idea?', '11. Do you have any additional skills?', 'Achievements'], dtype='object')

In [5]: ...
```

```
In [1]: df=df.rename(columns={ '2.Gender': 'Gender', '3. Year of Study': 'Year of Study', '4. Branch': 'Branch',
 '5. Type of Schooling till 10th Grade': 'Type of Schooling 10th', '2. Location of Class 10th': 'Location of Class 10th',
 '4. Type of Schooling for class 12th': 'Type of Schooling 12th', '5. Percentage in Class 12th?': 'In Class 12th',
 '6. Location of Class 12th school': 'Location of Class 12th', '7. Dropped an year after School?': 'Dropped by after School',
 '8. Want to be first generation graduate': 'Want to be 1st generation graduate',
 '9. Involved in additional volunteer groups(NCC,NSS, Social Welfare Groups or any other)': 'Involved in additional volunteer',
 '1. How many programming languages do you know? ': 'The number of programming languages',
 '2. Actively involved in Specific technology?': 'Involved in specific technology?',
 '3. Exposure to Data structures and algorithms?': 'Data structures and algorithms?',
 '4. Any additional skills?': 'Any additional skills?'},
 '5. Are you associated with any developer Community?': 'Associated with any developer Community',
 '6. Active in developer Communities': 'Developer Communities',
 '7. Participating in Hackathons and other competitions?': 'Hackathons and other competitions',
 '8. Have you made any Software based project?': 'Software based projects',
 '9. Have you made any Hardware based project (Arduino, Raspberry Pi, Robots or any other)?': 'Hardware based project',
 '10. Have you ever pitched any idea?': 'Pitch',
 '11. Do you have any additional skills?': 'Any additional skills'})
```

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 256 entries, 0 to 255
Data columns (total 25 columns):
 #   Column          Non-Null Count  Dtype  
 --- 
 0   Unnamed: 0      256 non-null    int64  
 1   Gender         256 non-null    object  
 2   Year of Study  256 non-null    object  
 3   Branch         256 non-null    object  
 4   Type of Schooling <10th        256 non-null    object  
 5   % in Class 10th       256 non-null    object  
 6   Location of Class 10th       256 non-null    object  
 7   Type of Schooling 12th       256 non-null    object  
 8   % in Class 12th       256 non-null    object  
 9   Location of Class 12th       256 non-null    object  
 10  Droped by after School     256 non-null    object  
 11  Want to be 1st generation graduate 256 non-null    object  
 12  Involved in additional volunteer 256 non-null    object  
 13  The number of programming languages 256 non-null    object  
 14  Involved in specific technology? 256 non-null    object  
 15  Data structures and algorithms? 256 non-null    object  
 16  Github?          256 non-null    object  
 17  Associated with any developer Community 216 non-null    object  
 18  Developer Communities       216 non-null    object  
 19  Hackathons and other competitions 256 non-null    object  
 20  Software based projects      256 non-null    object  
 21  Hardware based project      256 non-null    object  
 22  Pitch              256 non-null    object  
 23  Any additional skills       256 non-null    object  
 24  Achievements          256 non-null    object  
 25  Dyperl int64           256 non-null    object  
dtypes: int64(1), object(24)
memory usage: 50.1+ KB
```

```
In [25]: df['Branch'].unique()
```

```
Out[25]: array(['CSE/IT', 'CSIT', 'CSE', 'IT', 'CS', 'ECE', 'CSAI', 'EN', 'CSEME',
 'ME', 'CE', 'MCA'], dtype=object)
```

```
Renaming all Computer Science Branches to one
```

```
In [26]: df['Branch'].replace(to_replace=['CSE/IT', 'CSIT', 'CSE', 'IT', 'CS'], value='CSIT', inplace=True)
```

```
In [27]: df['Branch'].unique()
```

```
Out[27]: array(['CSIT', 'ECE', 'CSAI', 'EN', 'CSEME', 'ME', 'CE', 'MCA'],
 dtype=object)
```

Handling missing values

```
In [7]: df.isnull().sum()
```

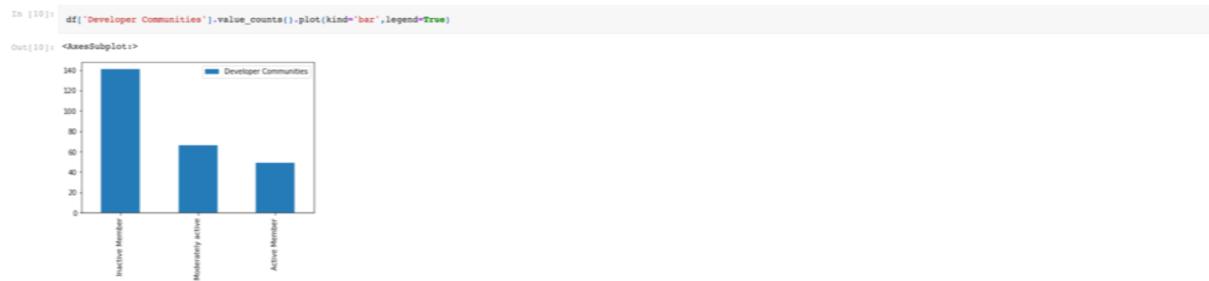
```
Out[7]: Unnamed: 0          0
Gender          0
Year of Study   0
Branch          0
Type of Schooling <10th  0
% in Class 10th       0
Location of Class 10th  0
Type of Schooling 12th  0
% in Class 12th       0
Location of Class 12th  0
Dropped by after School 0
Want to be 1st generation graduate 0
Involved in additional volunteer 0
The number of programming languages 0
Involved in specific technology? 0
Data structures and algorithms? 0
Github?          0
Associated with any developer Community 0
Developer Communities       40
Hackathons and other competitions 0
Software based projects      0
Hardware based project      0
Pitch              0
Any additional skills       0
Achievements          0
Dyperl int64           0
```

```
In [8]: #treating null values
df.fillna('Inactive Member', inplace=True)
```

```
In [9]: df['Associated with any developer Community'].value_counts().plot(kind='bar', legend=True)
```

```
Out[9]: <AxesSubplot: >
```

```
In [10]: df['Developer Communities'].value_counts().plot(kind='bar', legend=True)
```



INFERENCE:

1. Maximum Students are not associated with any developer community.
2. Among the associated members it is evident that active members are very less that means students are unaware about the importance of a community and need to be more participative.

```
In [11]: def achieve(i):
    if i == 'None':
        return 0
    else:
        return 1
```

```
In [12]: df['Achievements2']=df['Achievements'].apply(achieve)
```

```
In [13]: df.head()
```

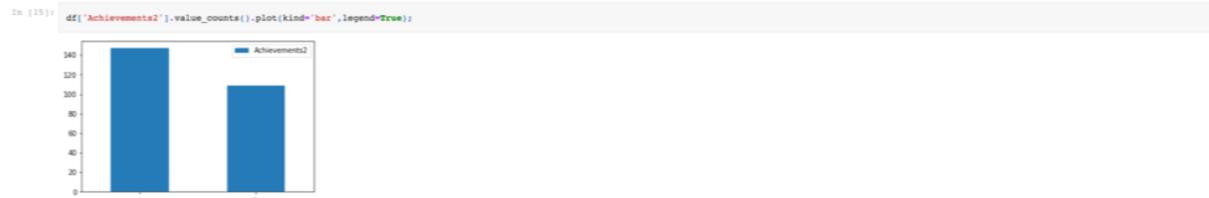
	Unnamed: 0	Gender	Year of Study	Branch	Type of Schooling <10th	% in Class 10th	Location of Class 10th	Type of Schooling 12th	% in Class 12th	Location of Class 12th	Github?	Associated with any developer Community	Developer Communities	Hackathons and other competitions	Software based projects	Hardware based project	Pitch	Any additional skills	Achievements	Achievements2
0	0	Male	Third	CSE/IT	Normal Schooling >=80% and <90%	Town	Normal Schooling	>=80% and <90%	Town	Medium	Yes	Active Member	Participated	Yes	No	Yes	Cultural Activities	Scholarships,Hackathons,Research Papers,Projects	1	
1	1	Female	Second	CSIT	Schooling with Tuitions >= 90%	City	Schooling with Tuitions	>=90%	City	Medium	No	Moderately active	Not yet participated	Yes	No	No	Cultural Activities	None	0	
2	2	Male	Second	CSIT	Normal Schooling >= 90%	City	Coaching Based Schools	>=90%	City	Medium	No	Inactive Member	Not yet participated	Yes	No	No	Photography/Video Editing	None	0	
3	3	Male	Second	CSIT	Normal Schooling >=70% and <80%	City	Normal Schooling	>=80% and <90%	City	Easy	Yes	Moderately active	Not yet participated	Yes	Yes	No	None	None	0	
4	4	Male	Second	CSIT	Normal Schooling >=70% and <80%	City	Normal Schooling	>=80% and <90%	City	Hard	Yes	Active Member	Participated	Yes	No	Yes	None	Hackathons,Coding Competitions,Projects	1	

	Unnamed: 0	Gender	Year of Study	Branch	Type of Schooling <10th	% in Class 10th	Location of Class 10th	Type of Schooling 12th	% in Class 12th	Location of Class 12th	Github?	Associated with any developer Community	Developer Communities	Hackathons and other competitions	Software based projects	Hardware based project	Pitch	Any additional skills	Achievements	Achievements2
0	0	Male	Third	CSE/IT	Normal Schooling >=80% and <90%	Town	Normal Schooling	>=80% and <90%	Town	Medium	Yes	Active Member	Participated	Yes	No	Yes	Cultural Activities	Scholarships,Hackathons,Research Papers,Projects	1	
1	1	Female	Second	CSIT	Schooling with Tuitions >= 90%	City	Schooling with Tuitions	>=90%	City	Medium	No	Moderately active	Not yet participated	Yes	No	No	Cultural Activities	None	0	
2	2	Male	Second	CSIT	Normal Schooling >= 90%	City	Coaching Based Schools	>=90%	City	Medium	No	Inactive Member	Not yet participated	Yes	No	No	Photography/Video Editing	None	0	
3	3	Male	Second	CSIT	Normal Schooling >=70% and <80%	City	Normal Schooling	>=80% and <90%	City	Easy	Yes	Moderately active	Not yet participated	Yes	Yes	No	None	None	0	
4	4	Male	Second	CSIT	Normal Schooling >=70% and <80%	City	Normal Schooling	>=80% and <90%	City	Hard	Yes	Active Member	Participated	Yes	No	Yes	None	Hackathons,Coding Competitions,Projects	1	

5 rows x 26 columns

```
In [14]: cat=['Gender', 'Year of Study', 'Branch', 'Type of Schooling <10th', ' % in Class 10th', ' Location of Class 10th', ' Type of Schooling 12th', ' % in Class 12th', ' Location of Class 12th', ' Dropped ly after School', 'Want to be 1st generation graduate', 'Want to be 2nd generation graduate', 'The number of programming languages', 'Involved in specific technology?', 'Data structure and algorithms?', 'Github?', 'Associated with any developer Community', 'Developer Communities', 'Hackathons and other competitions', 'Software based projects', 'Hardware based project', 'Pitch', 'Any additional skills']
```

1) Non achievement and achievement



INFERENCE:

It is evident and positive observation that there are more number of students who have achieved something in coding domain. Though there is not huge difference between the number of students who have achievements and others who do not.

2) Histplot by Non achievement and achievement

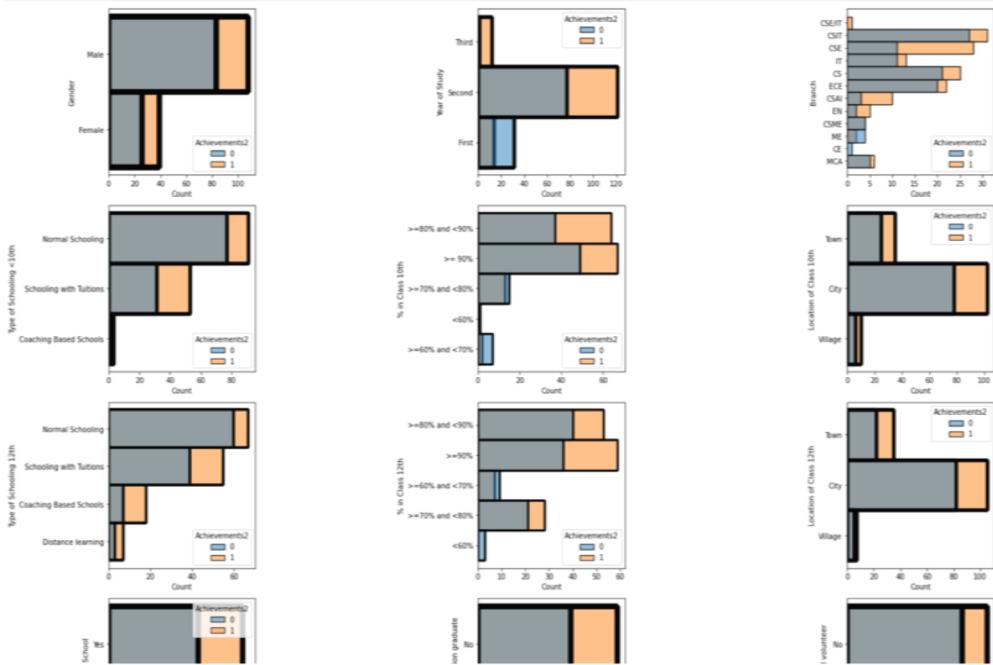
```
In [16]: fig = plt.figure(figsize=(20,30))
```

2) Histplot by Non achievement and achievement

```
In [16]: fig = plt.figure(figsize=(20,30))

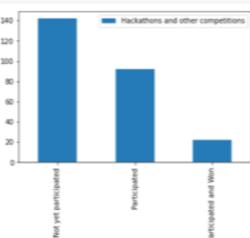
for i in range(len(cat)):
    plt.subplot(8,3,i+1)
    sns.histplot(data=df,y=df[cat[i]],hue=df['Achievements2'])

plt.tight_layout()
plt.show()
```





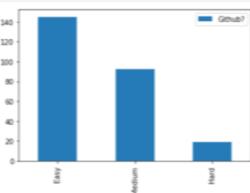
```
In [17]: df['Hackathons and other competitions'].value_counts().plot(kind='bar', legend=True);
```



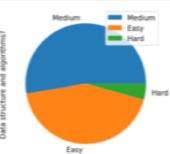
INFERENCE:

There is a need to inculcate coding competitions among students in order to increase their skills and make them aware about where they stand when compared to people of same domain.

```
In [21]: df['Github?'].value_counts().plot(kind='bar', legend=True);
```



```
In [22]: df['Data structure and algorithms?'].value_counts().plot(kind='pie', legend=True);
```



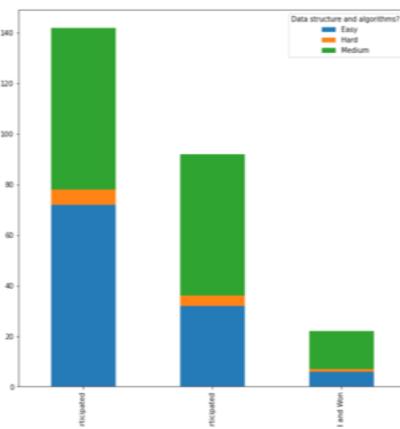
INFERENCE:

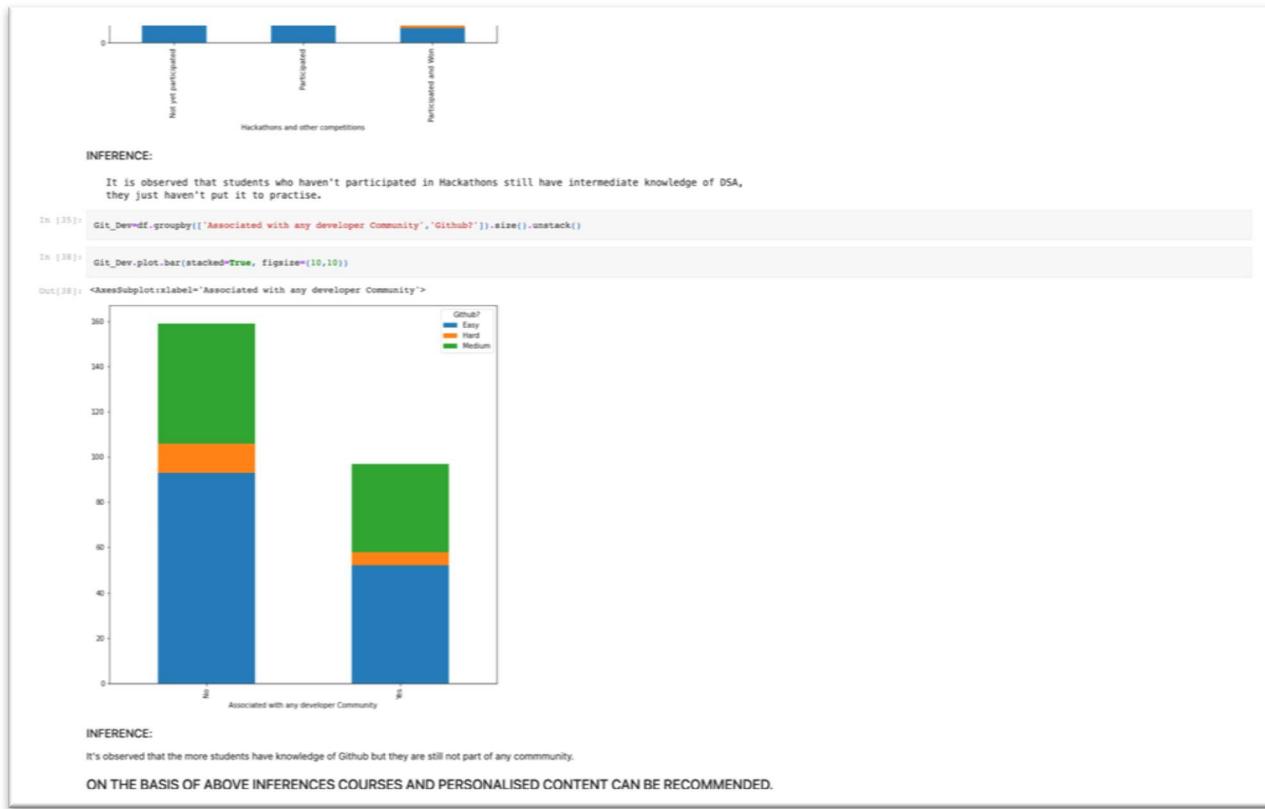
It is evident that students already know about DSA just their expertise differ.

```
In [31]: Dsa_Hackathon=df.groupby(['Hackathons and other competitions','Data structure and algorithms?']).size().unstack()
```

```
In [32]: Dsa_Hackathon.plot.bar(stacked=True, figsize=(10,10))
```

```
Out[32]: <AxesSubplot:xlabel='Hackathons and other competitions'>
```





CONCLUSION

In conclusion, the AI-Enabled Coding Tutor and Learning Platform presents a transformative solution for enhancing coding education through the integration of artificial intelligence and machine learning. By leveraging advanced technologies, personalized learning paths, real-time code analysis, interactive coding challenges, and a collaborative community, the platform aims to revolutionize the way individuals learn and improve their coding skills.

The platform addresses the growing demand for effective and accessible coding education, catering to learners of different proficiency levels and coding backgrounds. It offers a user-friendly interface, personalized learning paths, and real-time feedback to guide learners through their coding journey and help them build practical coding skills.

Through benchmarking existing products, analyzing the market, and considering customer needs, the AI-Enabled Coding Tutor and Learning Platform ensures a competitive edge by offering unique features, customization, and continuous improvement based on user feedback and emerging industry trends.

The development of the platform requires a dedicated team comprising project managers, software engineers, data scientists, and UI/UX designers. Collaborative efforts, adherence to best practices, and meticulous testing are essential for creating a robust and reliable learning environment.

From a business perspective, the platform's monetization model can include subscription plans, freemium options, partnerships with educational institutions, and potentially licensing the platform to other organizations. A well-defined business model ensures sustainable revenue generation and growth opportunities.

In conclusion, the AI-Enabled Coding Tutor and Learning Platform holds great potential to empower learners, provide personalized guidance, foster collaboration, and ultimately contribute to the advancement of coding education. By equipping learners with essential coding skills, the platform paves the way for future opportunities in the tech industry and beyond.

Market Analysis

Introduction:

The landscape of education and skill development has evolved significantly with the integration of artificial intelligence and technology. An AI-enabled coding tutor and learning platform leverages advanced algorithms to provide personalized, efficient, and interactive coding education. However, not all users within this expansive domain share the same requirements, learning styles, and objectives. Market segmentation analysis becomes a vital tool in identifying and catering to diverse user groups, enhancing the overall user experience, and fostering better learning outcomes.

Primary Dimension: User characteristics and preferences:

These attributes can potentially be used to create segments within the market of users using the AI-enabled coding tutor platform:

1. **Engagement patterns and behavior:** The first trial appointment date, first payment date, homework done, desktop enter, add homework done, call date, first visit date, nps score and average score could provide insights into how actively users are engaging with the platform, whether they complete assignments, their satisfaction level (NPS score), and their learning progress.

Based on these attributes, we can create market segments like:

- **Engaged Learners:** with high average score and homework done values indicating active participation and completion of assignments.
- **Paid Subscribers:** Users who have made payments as indicated by paywall paid.
- **Trial Users:** Users who have a first trial appointment date but haven't made a payment yet.

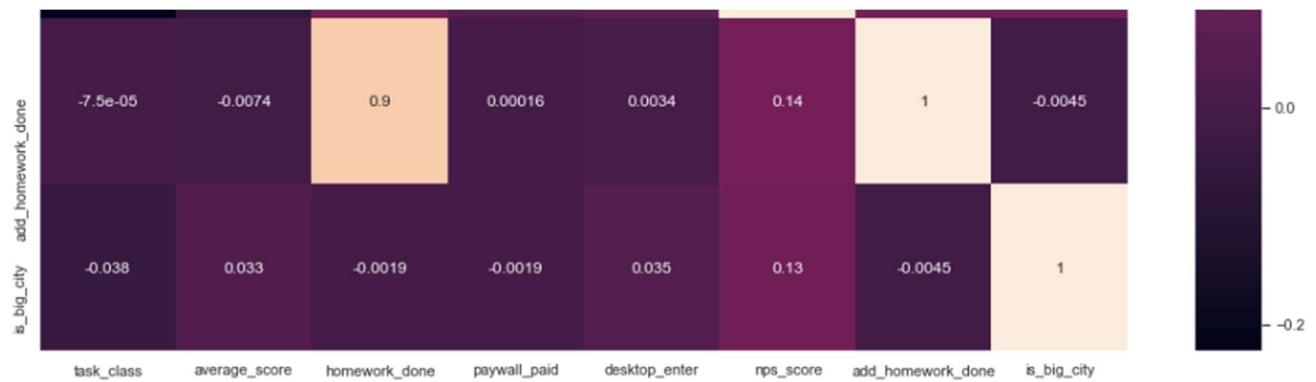
- **Operating system preferences** : Segments based on ‘os’ preference, which could influence platform optimization for different operating system.
- **Geographic segments**: Based on ‘region’ and ‘is big city’ catering to users from different geographical locations.

```
[ ] df_[df_['user_id'] == 'ffad160245d011e860b8d2576d06a7a1']
```

	user_id	first_trial_appointment_date	first_payment_date	os	tutor	job	task_class	average_score	homework_done	paywall_paid
118420	ffad160245d011e860b8d2576d06a7a1	2020-06-26 11:57:00		NaT iOS	I am doing now	Finance	6.0	83.333333	3	0

```
▶ df_corr = data1.corr()
# plot the heatmap and annotation on it
fig, ax = plt.subplots(figsize=(18,18))
sns.heatmap(df_corr, xticklabels=df_corr.columns, yticklabels=df_corr.columns, annot=True, ax=ax);
```





```

data1.groupby('os').agg({'user_id': ['count'],
    'homework_done': [np.mean], 'add_homework_done': [np.mean],
    'task_class': [np.mean], 'average_score': [np.mean],
    'nps_score': [np.mean], 'paywall_paid': [np.mean],
    'desktop_enter': [np.mean],
})

```

os	user_id	homework_done	add_homework_done	task_class	average_score	nps_score	paywall_paid	desktop_enter
	count	mean	mean	mean	mean	mean	mean	mean
Android	42476	3.647307	0.114935	6.664719	76.725451	6.794393	0.000024	0.802124
Linux	44	3.590909	0.159091	6.558140	72.910613	NaN	0.000000	1.000000
Mac OS	1035	3.477295	0.261836	6.646199	81.151649	1.000000	0.000000	0.998068
Solaris	1	2.000000	0.000000	5.000000	95.500000	NaN	0.000000	1.000000
Ubuntu	12	2.250000	0.000000	7.166667	87.136364	NaN	0.000000	1.000000
Unknown	15	1.133333	0.000000	7.000000	78.490909	NaN	0.000000	0.200000
Windows	22244	3.211698	0.088428	6.767658	78.909539	6.764706	0.000135	0.999011
iOS	52737	3.454558	0.110776	6.855201	76.834572	6.860000	0.000076	0.773290

Financial Modeling

EdTech = "education" + "technology" EdTech refers to hardware and software designed to enhance teacher-led learning in classrooms and improve students' education outcomes. This includes a wide range of tools, such as online learning platforms, virtual classrooms, educational software, and interactive whiteboards. Edtech can be used to facilitate teaching and learning, assess student progress, and provide personalized learning experiences.

Accordingly, there are several discrete aspects to describing the intellectual and technical development of educational technology:

Educational technology as the theory and practice of educational approaches to learning.
 Educational technology as technological tools and media, for instance massive online courses, that assist in the communication of knowledge, and its development and exchange. This is usually what people are referring to when they use the term "edtech". Educational technology for learning management systems (LMS), such as tools for student and curriculum management, and education

management information systems (EMIS). Educational technology as back-office management, such as training management systems for logistics and budget management, and Learning Record Store (LRS) for learning data storage and analysis. Educational technology itself as an educational subject; such courses may be called "computer studies" or "information and communications technology (ICT)".

```
import gc

from scipy.stats import norm
import matplotlib.pyplot as plt
import seaborn as sns

[ ] data1 = pd.read_csv('/Users/muskansharma/Documents/edtech_data.csv')
data1.head()

      client_id          user_id first_trial_appointment_date first_payment_date   os tutor    job task_class average_score h
0  2989400374e8abf2c4ac8921531cd1c0  92cc220d5171deb66e947d5b4ffce8bd        NaN        NaN  Android  NaN  NaN     6.0  80.222222
1  c8bce51779d3f746baaf2f59ddef69cd  80cf7575a4dc3a3a8cb30e3262d23e19        NaN        NaN  Windows  I am
                                            doing now  IT     11.0  98.250000
2  a6bad682ac10999865570bee0e11b9f6  3b671290e690217ce17052c9c52afa31        NaN        NaN  Android  No  NaN     5.0  68.000000
3  b85c31b416b405dd9b9498bdde703f4e  03e4d2186bc5231cd7f302ac76e1de76        NaN        NaN    iOS  NaN  NaN     7.0  78.571429
4  93c44d3b5d850f60ada1faf9db83b7f7  ec8a79f01156b3c82a27ca947f49d319        NaN        NaN  Android  No  Finance  6.0  92.000000

[ ] df_ = data1.groupby('user_id').agg('count').reset_index()[['user_id', 'client_id']]
#df_['client_id'] = df_df['client_id']
df_[df_['client_id'] > 1].head()

      user_id  client_id
15  000796759b25145e0dc8a0fad33d801        2
20  000c159a37f9e587091fed0d69fa4b2c        2
48   001b9da631dbc7ac11a525a48a61aa6        2
122  004626f806f715350ea69fd9b50787bf        2
182  00674f08b030c10c0230181f0ab31322        2
```

It appears that each row represents information about a specific user and their interactions with the platform.

- **Engagement Analysis:** We have analyzed user engagement by looking at the number of completed homework assignments, average scores, and additional homework done. Identified users who actively participate and those who might need more encouragement.
- **Payment Conversion:** Explored the relationship between trial appointments and first payments to understand how many users convert from trial to paying customers.

- **Platform Usage Trends:** Analyze trends related to first visit dates, desktop usage, and call dates to understand when and how users are interacting with the platform.

Business Modeling

Introduction:

Business modeling involves creating a structured representation of how a business operates, generates revenue, and interacts with its stakeholders. In the context of programming languages, business modeling might involve analyzing how these languages impact various aspects of the software development industry.

Here's how could you approach it:

a. Market Analysis:

- Identify the demand for developers skilled in each programming language.
- Analyze job posting, salary trends, and the overall market size for each language.
- Determine which industries or domains have a high demand for developers in specific languages.

b. Developer Tools and Ecosystem:

- Explore the tools, libraries, frameworks, and IDEs associated with each language.
- Analyze how developer preferences for certain languages impact the tooling market.

For visualization I'm including only those programming languages whose mean value is greater than 2.5. And I'm doing this using the result as boolean mask.

```
[ ] mask = data.mean() > 2.5

data = data.loc[:, mask]

clms = data.columns.tolist()
clms

['C/C++',
 'C#',
 'Java',
 'Javascript',
 'Matlab',
 'Objective-C',
 'Perl',
 'PHP',
 'Python',
 'Visual Basic']
```

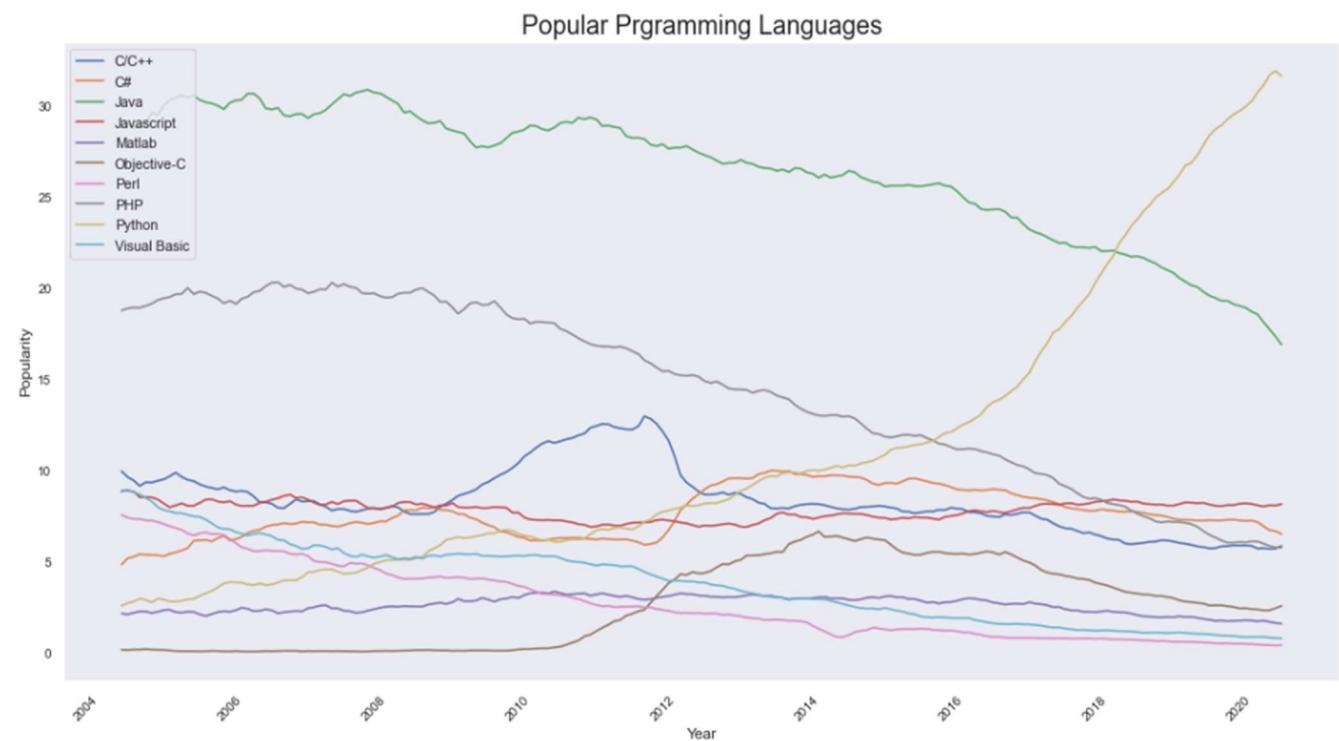
The above clms list showing our top 10 programming languages.

► Visualization

```
plt.figure(figsize = (15, 8))
sns.set(style = 'dark')

for language in clms:
    sns.lineplot(x = data.index, y = data[language], label = language)

plt.ylabel('Popularity', fontsize = 12)
plt.xlabel('Year', fontsize = 12)
plt.title('Popular Programming Languages', fontsize = 20)
plt.legend(loc = 2)
plt.yticks(fontsize = 10)
plt.xticks(rotation = 45, fontsize = 10)
plt.tight_layout()
plt.show()
```



This graph is showing importance of programming language has been increasing and developers with high demand might have better job prospects and potentially higher salaries. Staying updated with language trends can help developers align their skills with market needs. Also, startups or business that align with the trends of popular languages might find more opportunities for growth and innovation.

Future Outlook

An AI-Enabled coding tutor and learning platform has the potential to evolve significantly in the future, driven by advancements in technology, changes in education paradigms, and the ever-evolving needs of learners and developers. Here's how such a platform could be enhanced in the future:

1. **Personalized Learning:** Future versions of the platform could utilize more advanced AI algorithms to offer even more personalized learning experiences. This might involve tailoring content, exercises, and challenging to each individual's learning style, pace, and skill level.
2. **Adaptive Assessment:** The platform could incorporate adaptive assessment methods, identifying a learner's strengths and weaknesses in real-time. This would allow the platform to provide targeted practice exercises to address specific areas of improvement.
3. **Advanced Code Analysis:** Improved AI capabilities could lead to more sophisticated code analysis. The platform might be able to identify subtle bugs, suggest optimizations, and provide insights into best practices.
4. **Interactive coding challenges:** Future platforms could offer interactive coding challenges that simulate real-world scenarios. This would provide learners with practical experience in problem-solving and collaboration.
5. **Natural Language Interaction:** Enhanced natural language processing could enable learners to interact with the platform using natural language. This might include asking questions, seeking explanations, and receiving code-related guidance in conversational form.