E-Learning vs. Traditional Coaching: An Exploratory Data Analysis and Predictive Modeling Approach

A project work made under the guidance of **Vigor Council**

Submitted by:

Kritika Mittal

Kriti Khurana

Data Analyst Internship @ Vigor Council

June, 2024

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to our supervisor, **Dr. B.P. Sharma**, for his invaluable guidance and support throughout this project. This project helped us to explore and undertake research work and gaining practical knowledge and expertise to perform tasks. We also extend our thanks to the **Vigor Council** for providing the necessary resources and facilities.

Thanking You,

Kriti Khurana

Kritika Mittal

DECLARATION

We, **Kritika Mittal** and **Kriti Khurana**, hereby declare that the project entitled " **E-Learning vs. Traditional Coaching: An Exploratory Data Analysis and Predictive Modeling Approach**" is a result of our original research work carried out under the guidance and supervision of **Dr. B.P. Sharma**, **President** at **Vigor Council**.

This project work is undertaken as part of **our internship as Data Analysts** and is submitted to **Vigor Council**. We affirm that the research and findings presented in this project are genuine. All sources of information and data have been acknowledged appropriately. We also declare that any help received in carrying out this project and preparing the report has been duly acknowledged.

ABSTRACT

Understanding student preferences for online versus offline learning environments is crucial for modern education. This project analyzes these preferences using a comprehensive survey and advanced data analysis techniques. We collected data from students on various aspects, including perceived effectiveness, convenience, accessibility, and overall satisfaction with each learning mode.

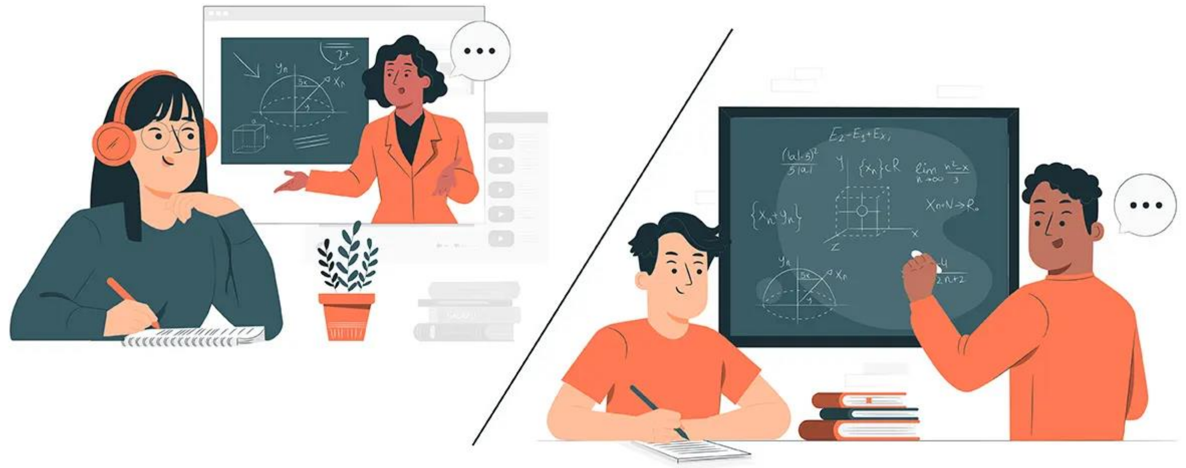
Exploratory data analysis revealed significant trends and correlations, forming the basis for predictive modeling. Machine learning algorithms—Logistic Regression, SVM, and Random Forest—were employed to classify student preferences. These models were evaluated using accuracy, precision, recall, and F1-score metrics, highlighting the impact of demographic and behavioral factors on learning preferences.

A Power BI dashboard was developed to visualize findings interactively, enabling stakeholders to explore the data through filters and slicers.

The study reveals a nuanced landscape where both online and offline education have distinct advantages, appealing to different student segments. These insights can guide educational institutions in designing hybrid learning models that cater to diverse needs, enhancing overall learning outcomes and satisfaction. This project offers a data-driven perspective on the future of education, emphasizing the importance of flexibility and adaptability in teaching methodologies.

INTRODUCTION

We live in an era where digital technology permeates every aspect of our lives, simplifying and enhancing our quality of life. One key area where digital technology has seen substantial growth is education. Over the years, educators have increasingly integrated digital tools to make teaching and learning more efficient and engaging. However, when the COVID-19 pandemic led to the closure of nearly all educational institutions worldwide in March 2020, digital technology became essential.



The pandemic caused an unprecedented disruption in education, affecting approximately 1.6 billion learners across over 200 countries and impacting more than 94% of the global student population (Pokhrel & Chhetri, 2021). To cope with this sudden shift, all educational activities transitioned to digital platforms. This rapid shift accelerated the growth of digital education in India, making online learning the new norm. As a result, technological proficiency became crucial for continuing education. With digital learning's rise, traditional teacher-centred lectures have been replaced by more student-centred approaches like group projects, discussions, and hands-on activities (Zhu & Liu, 2020).

Despite the omnipresence of digital devices among today's youth, not all are equally equipped for a tech-driven future due to existing digital divides within the community. To meet the needs of the younger generation and prepare them for a digital future, schools and education systems must undergo a significant digital transformation (Iivari et al., 2020).

ONLINE VS OFFLINE COACHING

In the evolving landscape of education, the debate between online and offline coaching has become more prominent. Each mode has its unique advantages and disadvantages, and understanding these can help learners make informed decisions.

ONLINE COACHING

Benefits:

1. **Accessibility and Flexibility:** One of the most significant advantages of online coaching is its accessibility. Learners can access educational materials and sessions from anywhere, eliminating geographical barriers. This flexibility allows students to learn at their own pace and on their own schedule, which is particularly beneficial for those balancing other commitments like work or family.
2. **Wide Range of Resources:** Online platforms often offer a wealth of resources, including videos, interactive simulations, and e-books. These resources can cater to various learning styles, making education more inclusive.

3. **Cost-Effective:** Online coaching can be more affordable than traditional offline methods. Costs associated with commuting, physical materials, and infrastructure are minimized, making education more accessible to a broader audience.
4. **Customized Learning Experience:** Online platforms often use algorithms to personalize learning experiences, adapting to the pace and style of individual students. This can enhance the effectiveness of the learning process.

Demerits:

1. **Lack of Personal Interaction:** Online coaching lacks face-to-face interaction, which can impact the learning experience. The absence of direct communication with instructors and peers may lead to feelings of isolation and reduced motivation.

2. **Technical Issues:** Dependence on technology means that technical problems, such as poor internet connectivity or software glitches, can disrupt learning. Not all students have access to high-quality digital devices or reliable internet connections.
3. **Self-Discipline Required:** Online learning demands a high level of self-discipline and time management skills. Without the structure of a physical classroom, some students may struggle to stay focused and complete their courses.
4. **Limited Practical Experience:** Subjects that require hands-on practice, such as laboratory sciences or certain vocational skills, may not be as effectively taught online due to the lack of physical presence and equipment.

OFFLINE COACHING

Benefits:

1. **Direct Interaction:** Offline coaching provides direct interaction with instructors and peers, facilitating immediate feedback and a more engaging learning environment. This interaction can enhance understanding and retention of information.
2. **Structured Learning Environment:** The structured environment of a physical classroom can help students stay disciplined and focused. Regular schedules and physical presence can contribute to better time management and accountability.
3. **Practical Experience:** For subjects requiring hands-on practice, offline coaching offers the necessary resources and supervision. This can be critical for fields that require practical skills and laboratory work.

4. **Social Skills Development:** Attending physical classes helps students develop social skills through interactions with peers and teachers. These skills are essential for personal and professional growth.

Demerits:

1. **Limited Flexibility:** Offline coaching often follows a rigid schedule, which can be challenging for students with other commitments. Commuting to a physical location can also be time-consuming and inconvenient.

2. **Higher Costs:** The costs associated with offline coaching, including transportation, physical materials, and facility maintenance, can be higher than those for online education. This may limit access for some students.
3. **Geographical Limitations:** Students are limited to the educational resources available in their vicinity. This can be a significant disadvantage for those living in remote areas without access to high-quality coaching centres.
4. **Pace of Learning:** In a physical classroom, the pace of instruction is often set to accommodate the average student, which may not cater to individual learning speeds. Some students may feel left behind, while others may find the pace too slow.

LITERATURE REVIEW

EFFICACY OF ONLINE AND OFFLINE COACHING

Several studies have explored the effectiveness of online versus offline coaching. According to Allen et al. (2011), online learning can be as effective as traditional face-to-face instruction when the design and implementation of the online courses are carefully planned. Similarly, Means et al. (2010) conducted a meta-analysis of online learning studies and found that students in online conditions performed modestly better than those receiving face-to-face instruction. However, the effectiveness of online learning is often contingent on the quality of the technological infrastructure and the digital literacy of both students and educators (Bates, 2015).

In contrast, Bernard et al. (2014) argue that offline coaching continues to be highly effective due to the direct interaction between students and instructors, which can enhance comprehension and retention of information. The structured environment and immediate feedback in traditional classroom settings are also highlighted as significant contributors to effective learning.

LEARNER ENGAGEMENT

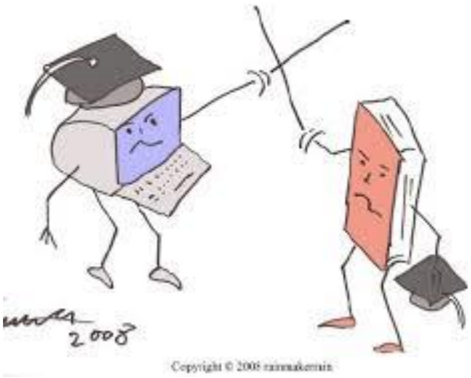
Engagement is a critical factor in the learning process, and various studies have examined how it differs between online and offline coaching. Kearsley and Shneiderman's (1999) Engagement

Theory posits that technology can enhance engagement by providing interactive and collaborative learning opportunities. Online platforms often incorporate multimedia resources, discussion forums, and interactive simulations that can engage learners in dynamic ways (Garrison & Kanuka, 2004).

However, some researchers argue that offline coaching may foster greater engagement through personal interactions and social presence. According to Tinto (1997), the sense of community and belonging in physical classrooms can significantly influence student engagement and retention. Moreover, face-to-face interactions can facilitate the development of social and communication skills, which are essential components of the learning experience (Wentzel, 1999).

EDUCATIONAL OUTCOME

Educational outcomes are a critical measure of the success of any instructional method. Research comparing online and offline coaching outcomes has yielded mixed results. A study by Xu and Jaggars (2013) found that students enrolled in online courses were more likely to struggle and had lower completion rates compared to those in traditional classrooms. This suggests that online education may pose challenges in terms of maintaining student motivation and persistence.



Conversely, other studies have shown positive outcomes for online learning. For instance, Al-Qahtani and Higgins (2013) reported that blended learning, which combines online and offline elements, resulted in higher student performance compared to purely offline methods. This hybrid approach leverages the strengths of both modalities, providing flexibility while maintaining some level of personal interaction.

TECHNOLOGICAL AND PEDAGOGICAL INNOVATIONS

The rapid development of educational technology has introduced new tools and methodologies that impact both online and offline coaching. The flipped classroom model, where students engage with lecture content online and participate in interactive activities in class, exemplifies how technology can enhance traditional teaching methods (Bergmann & Sams, 2012). This approach has been shown to improve student engagement and learning outcomes by promoting active learning and critical thinking (Chen Hsieh et al., 2017).

Furthermore, the integration of artificial intelligence and adaptive learning technologies in online platforms has the potential to personalize education and cater to individual learning needs (Holmes et al., 2019). These innovations can provide real-time feedback and adjust instructional content based on student performance, thereby enhancing the learning experience.

RESEARCH OBJECTIVE

The objective of this project is to analyze and compare the preferences and effectiveness of online versus offline coaching among users. By conducting a comprehensive survey and applying exploratory data analysis (EDA) alongside various classification algorithms, the project aims to:

1. Understand Demographics: Gather and analyze demographic data of users to identify trends and patterns in coaching preferences across different age groups, genders, occupations, and other relevant demographic factors.

2. Assess Perceptions: Evaluate user perceptions and thoughts regarding the general concept of coaching, including their motivations, expectations, and satisfaction levels with both online and offline coaching methods.

3. Compare Preferences: Determine the preferences for online versus offline coaching, identifying key factors that influence these choices, such as convenience, cost, interaction quality, and perceived effectiveness.

4. Evaluate Effectiveness: Use classification models to predict user preferences based on survey responses and assess the accuracy of these predictions. This includes comparing the performance of logistic regression, random forest, and SVM models.

5. Identify Best Practices: Highlight the strengths and weaknesses of online and offline coaching from the users' perspectives, providing actionable insights for educators, coaches, and policymakers to enhance the learning experience in both formats.

6. Model Performance: Evaluate and compare the performance of different classification models using confusion matrices and other relevant metrics to determine the most effective model for predicting coaching preferences.

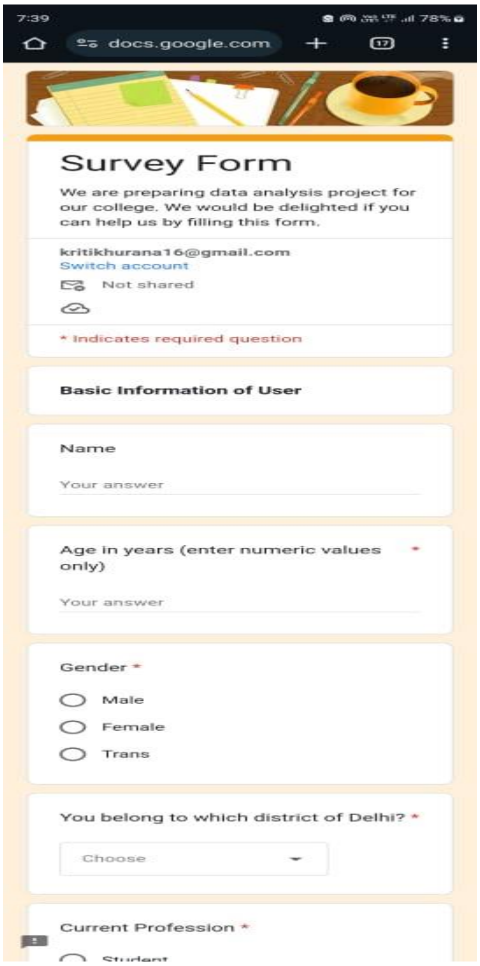
By achieving these objectives, the project seeks to provide a data-driven understanding of the current landscape of coaching, offering valuable insights into user preferences and the relative effectiveness of online versus offline coaching methods.

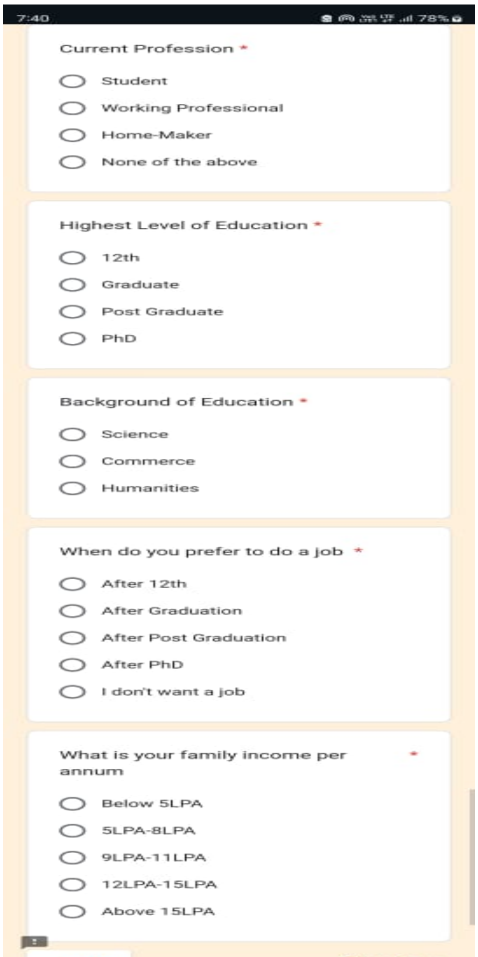
METHODOLOGY

SURVEY FORM

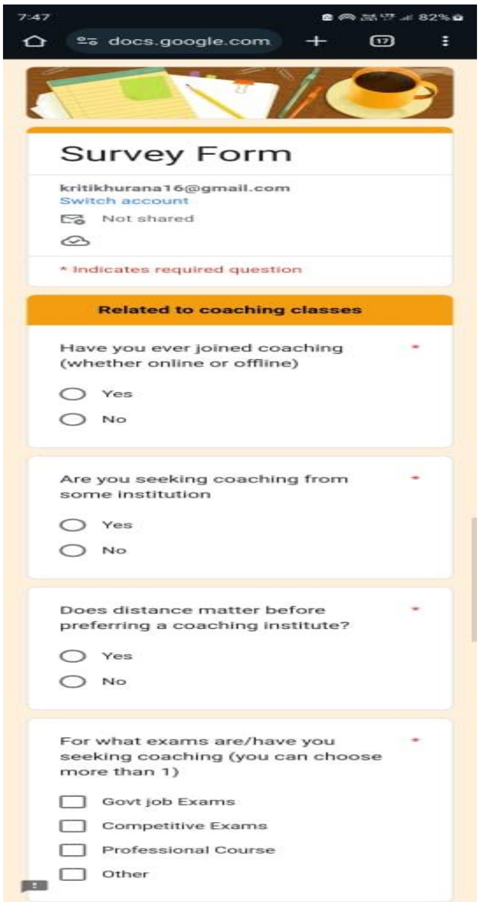
We have created a survey using Google Forms. We have also not collected the email Ids and Name section is set to be not mandatory to keep up with the privacy of the user. Furthermore, the survey form consists of 35 questions and is divided into 3 parts:

1) Personal Details of the users





2) Coaching Preferences and Choices



docs.google.com

Survey Form

kritikhurana16@gmail.com
Switch account
Not shared

* Indicates required question

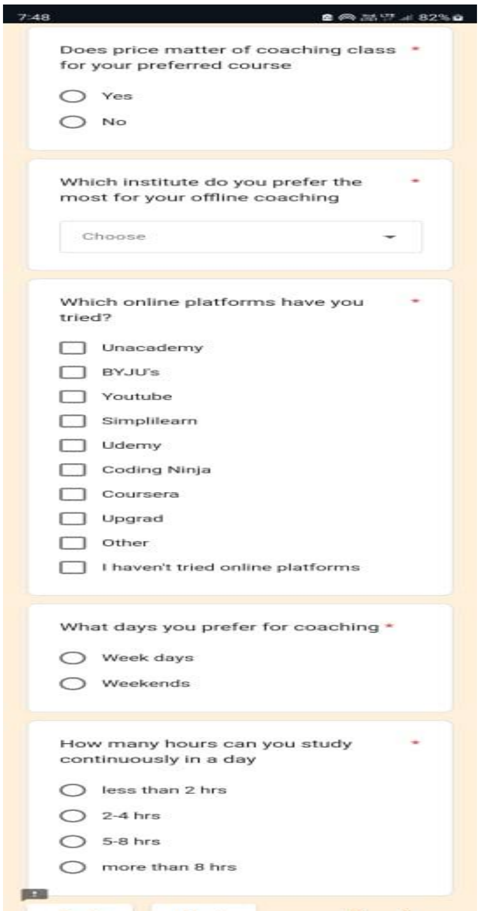
Related to coaching classes

Have you ever joined coaching (whether online or offline) *
- Yes
- No

Are you seeking coaching from some institution *
- Yes
- No

Does distance matter before preferring a coaching institute? *
- Yes
- No

For what exams are/have you seeking coaching (you can choose more than 1) *
- Govt job Exams
- Competitive Exams
- Professional Course
- Other



Does price matter of coaching class for your preferred course *
- Yes
- No

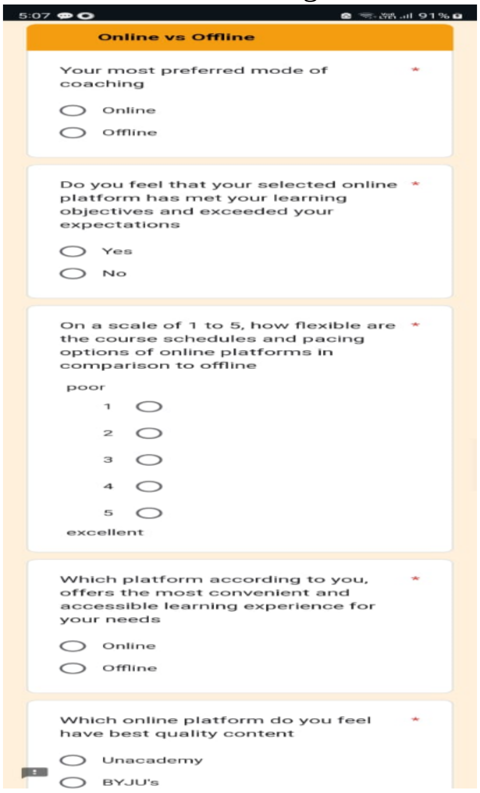
Which institute do you prefer the most for your offline coaching *
Choose

Which online platforms have you tried? *
- Unacademy
- BYJU's
- Youtube
- Simplilearn
- Udemy
- Coding Ninja
- Coursera
- Upgrad
- Other
- I haven't tried online platforms

What days you prefer for coaching *
- Week days
- Weekends

How many hours can you study continuously in a day *
- less than 2 hrs
- 2-4 hrs
- 5-8 hrs
- more than 8 hrs

3) Online vs Offline Thoughts



Online vs Offline

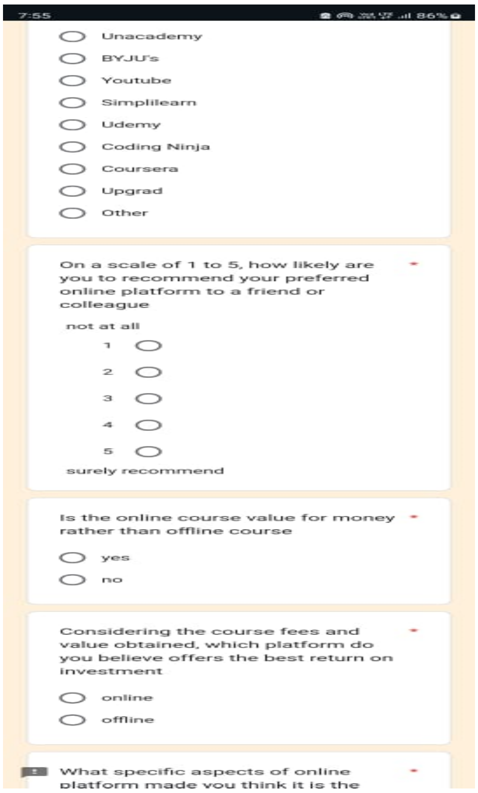
Your most preferred mode of coaching *
- Online
- Offline

Do you feel that your selected online platform has met your learning objectives and exceeded your expectations *
- Yes
- No

On a scale of 1 to 5, how flexible are the course schedules and pacing options of online platforms in comparison to offline *
poor
1 2 3 4 5
excellent

Which platform according to you, offers the most convenient and accessible learning experience for your needs *
- Online
- Offline

Which online platform do you feel have best quality content *
- Unacademy
- BYJU's



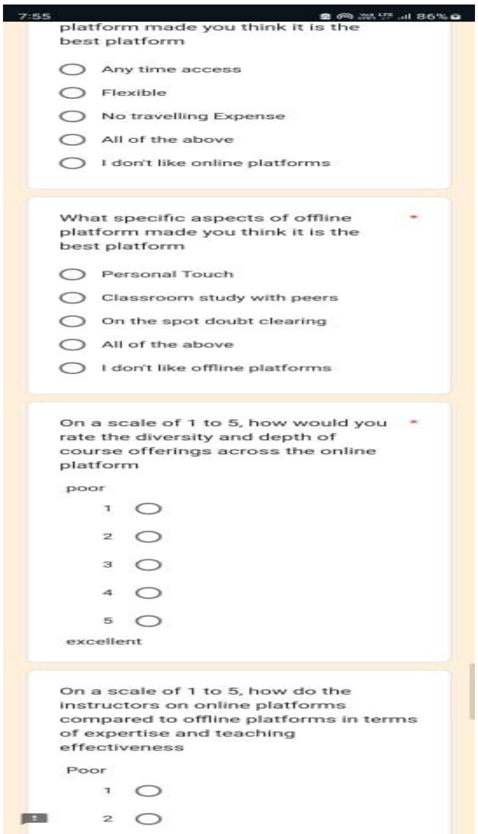
- Unacademy
- BYJU's
- Youtube
- Simplilearn
- Udemy
- Coding Ninja
- Coursera
- Upgrad
- Other

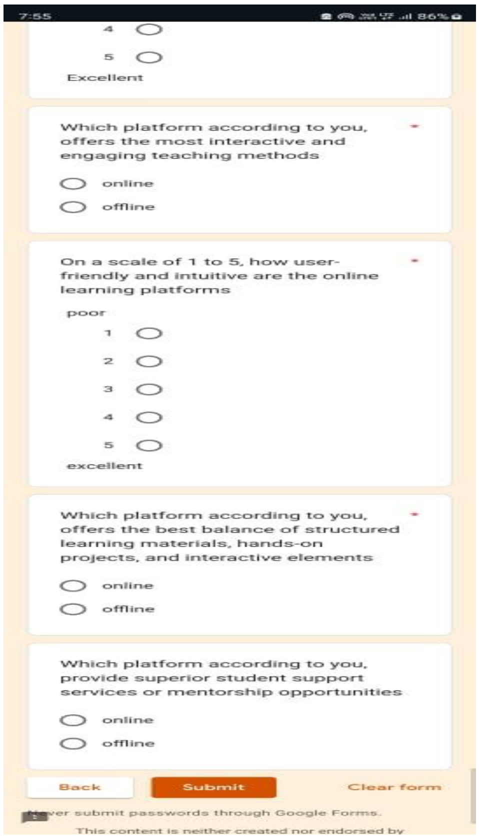
On a scale of 1 to 5, how likely are you to recommend your preferred online platform to a friend or colleague *
not at all
1 2 3 4 5
surely recommend

Is the online course value for money rather than offline course *
- yes
- no

Considering the course fees and value obtained, which platform do you believe offers the best return on investment *
- online
- offline

What specific aspects of online platform made you think it is the





DATASET

The responses were then downloaded from google sheet to .xlsx format. We received a total of 271 responses within a week which was enough to go ahead with our project.

	A	B	C	D	E	F	G	H
1	Timestamp	Name	Age in yeai	Gender	You belong to which dis	Current Professic	Highest Level of Ed	Background of Ec
2	5-23-2024 10:52:09	Kriti khurana	24	Female	North-West	Student	Graduate	Commerce
3	5-23-2024 11:00:12	Sweety	23	Female	New Delhi	Working Professi	Post Graduate	Commerce
4	5-23-2024 11:00:41	Naveen	23	Male	North-West	Student	Post Graduate	Commerce
5	5-23-2024 11:01:49	Rishub Sriva	24	Male	North-West	Working Professi	Graduate	Science
6	5-23-2024 11:04:30	Aditya	25	Male	New Delhi	Working Professi	Post Graduate	Science
7	5-23-2024 11:06:39	Manvi	27	Female	I don't belong in Delhi	Working Professi	Graduate	Science
8	5-23-2024 11:08:49	Yuvraj	27	Male	I don't belong in Delhi	Working Professi	Post Graduate	Science
9	5-23-2024 11:09:25	Rahul Khurai	30	Male	I don't belong in Delhi	Working Professi	Graduate	Commerce
10	5-23-2024 11:10:59	Deepali Paliv	39	Female	I don't belong in Delhi	Working Professi	Graduate	Humanities
11	5-23-2024 11:12:00		34	Male	Central	Working Professi	Graduate	Commerce
12	5-23-2024 11:15:35	Nisha Bhasir	39	Female	Central	Working Professi	Graduate	Humanities
13	5-23-2024 11:26:08	Rishabh garç	21	Male	Central	Student	Graduate	Commerce
14	5-23-2024 11:28:20	Prashansa	23	Female	North	Student	Graduate	Humanities
15	5-23-2024 11:30:30	Sonal Arora	33	Female	I don't belong in Delhi	Working Professi	Post Graduate	Science
16	5-23-2024 11:30:37	Samridhi	32	Female	I don't belong in Delhi	Working Professi	Post Graduate	Science
17	5-23-2024 11:31:20	RITU KHURA	38	Female	North	Working Professi	Post Graduate	Humanities
18	5-23-2024 11:42:56	Piyush	26	Male	South-West	Working Professi	Graduate	Science
19	5-23-2024 11:47:31	Pooja	39	Female	West	Working Professi	Post Graduate	Humanities
20	5-23-2024 11:52:04	Vaidehi	17	Female	North-West	Student	12th	Science
21	5-23-2024 11:52:20	Shailja	37	Female	South-East	Working Professi	Post Graduate	Commerce
22	5-23-2024 12:10:36	Shaloo	38	Female	West	Working Professi	Post Graduate	Humanities
23	5-23-2024 12:17:40	Sanjay kalra	39	Male	I don't belong in Delhi	Working Professi	Post Graduate	Science
24	5-23-2024 12:22:24	Komal Arora	39	Female	South-West	Working Professi	Post Graduate	Commerce
25	5-23-2024 12:33:16	Vaibhav Raw	24	Male	North-West	Working Professi	Post Graduate	Commerce
26	5-23-2024 12:45:13	Vandana Jair	38	Female	South-West	Working Professi	Post Graduate	Science
27	5-23-2024 13:06:43	Kritika	22	Female	Central	Working Professi	Graduate	Commerce
28	5-23-2024 13:17:32	Mohd Hamm	24	Male	South-East	Student	Post Graduate	Commerce
29	5-23-2024 13:56:54		22	Female	South-East	Student	Post Graduate	Humanities
30	5-23-2024 13:57:27	Kirti	23	Female	North-West	Student	Graduate	Humanities
31	5-23-2024 13:58:56		19	Male	West	Student	12th	Science
32	5-23-2024 14:04:24	radhika	30	Female	New Delhi	Working Professi	Graduate	Humanities
33	5-23-2024 14:06:31	rani	25	Female	Shahdara	Student	12th	Commerce

	I	J	K	L	M	N	O	P	
1	When do you prefer to do a j	What is your family inco	Have you ever joine	Are yo	Does di	For what exams are/have you see	Does pr	Which institu	Which on
2	After Graduation	Above 15LPA	Yes	No	No	Competitive Exams, Professional	No	TIME Institut	Unacader
3	After Graduation	9LPA-11LPA	No	No	Yes	Competitive Exams	Yes	Other	Unacader
4	After Post Graduation	9LPA-11LPA	Yes	Yes	Yes	Competitive Exams	Yes	TIME Institut	Unacader
5	After Graduation	Above 15LPA	Yes	No	Yes	Competitive Exams	Yes	MADE EASY	Udemy, C
6	After Post Graduation	12LPA-15LPA	Yes	Yes	Yes	Govt job Exams, Professional Co	Yes	Aakash Insti	Coding Ni
7	After Graduation	9LPA-11LPA	Yes	No	Yes	Other	Yes	Other	Unacader
8	After Post Graduation	Above 15LPA	Yes	No	Yes	Competitive Exams	Yes	Other	Youtube
9	After 12th	5LPA-8LPA	No	No	Yes	Professional Course, Other	Yes	Other	Unacader
10	After Graduation	5LPA-8LPA	No	No	Yes	Other	Yes	Other	I haven't t
11	After 12th	9LPA-11LPA	No	No	Yes	Govt job Exams	No	Vajiram & R	Youtube,
12	After 12th	12LPA-15LPA	Yes	No	Yes	Other	Yes	Other	Youtube
13	After Post Graduation	9LPA-11LPA	Yes	No	Yes	Govt job Exams, Competitive Exa	Yes	Drishti IAS	Udemy
14	After 12th	12LPA-15LPA	Yes	No	Yes	Competitive Exams	No	Other	Youtube,
15	After Graduation	Below 5LPA	Yes	No	Yes	Govt job Exams, Competitive Exa	Yes	MADE EASY	I haven't t
16	After Graduation	12LPA-15LPA	Yes	No	Yes	Competitive Exams	Yes	Other	Youtube,
17	After Post Graduation	5LPA-8LPA	Yes	No	Yes	Govt job Exams	Yes	I would rathe	Youtube
18	After Graduation	5LPA-8LPA	Yes	No	No	Professional Course	Yes	Other	Udemy, C
19	After Post Graduation	5LPA-8LPA	No	No	Yes	Govt job Exams	Yes	Other	Other
20	After Graduation	5LPA-8LPA	Yes	No	Yes	Competitive Exams	Yes	Other	Youtube
21	After Post Graduation	12LPA-15LPA	Yes	Yes	Yes	Competitive Exams	Yes	Drishti IAS	Youtube
22	After Post Graduation	Above 15LPA	Yes	No	Yes	Competitive Exams, Other	Yes	Other	Other
23	After Graduation	12LPA-15LPA	No	No	Yes	Professional Course	Yes	TIME Institut	I haven't t
24	After Graduation	9LPA-11LPA	Yes	Yes	No	Competitive Exams	Yes	ALLEN Care	I haven't t
25	After Post Graduation	Above 15LPA	Yes	No	Yes	Professional Course	Yes	Other	Youtube,
26	After Post Graduation	9LPA-11LPA	No	No	Yes	Professional Course	Yes	Other	Youtube
27	After Graduation	9LPA-11LPA	No	No	No	Competitive Exams, Professional	Yes	MBA Guru	BYJU's, Y
28	After Graduation	5LPA-8LPA	Yes	No	Yes	Competitive Exams	Yes	Drishti IAS	Unacader
29	After Graduation	12LPA-15LPA	Yes	Yes	Yes	Govt job Exams	Yes	Rau's IAS st	Unacader
30	After Graduation	Below 5LPA	Yes	No	Yes	Govt job Exams	Yes	I would rathe	Youtube,
31	After Post Graduation	Above 15LPA	Yes	Yes	Yes	Competitive Exams	Yes	MBA Guru	Unacader
32	After Post Graduation	9LPA-11LPA	Yes	Yes	Yes	Govt job Exams	Yes	ALLEN Care	Unacader
33	After 12th	Below 5LPA	Yes	Yes	Yes	Professional Course	Yes	Other	Other

Form Responses 1

CLEANING THE DATA

Later the data was cleaned in the jupyter notebook. Basically, the columns 'Timestamp' and 'Name' were dropped as we didn't need them. Along with that all the column names were renamed accordingly.

DETAILED INFORMATION OF RENAMED COLUMNS OF DATASET

There are 33 columns after cleaning the dataset. Here is the detailed list:

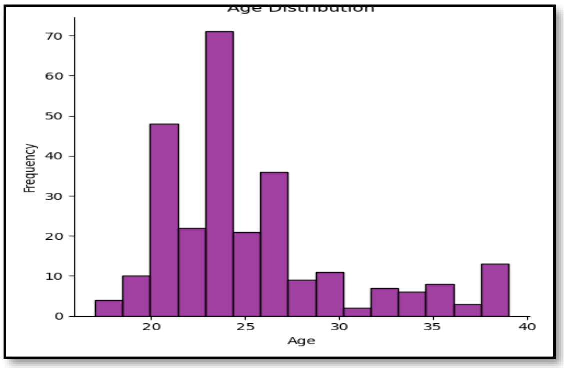
COLUMN NAME	DETAILS	TYPE
Age	Age of the users	Integer
Gender	Gender of the users	Object
Area	Current residence of the user	Object
Profession	Current profession of the user	Object
Education	Educational qualification of the user	Object
Stream	Educational background of the user	Object
Job_pref	Job preference of user	Object
Family_income	Family income of user	Object
Hx_coaching	History of coaching of user	Object
Current_coaching	Current coaching of the user	Object
Distance_matter	Coaching distance matter to user	Object
Exam_coaching	Which exam coaching does user prefer	Object
Price_matter	Coaching price matter to user	Object
Offline_institute	Which offline coaching the user prefer	Object
Online_platform	Which online platform the user prefer	Object
Coaching_days	Does user prefer coaching on week days or on weekends	Object
Study_hours	Hours user can study continuously	Object
Mode_of_study	User prefer online of offline mode of study	Object
Onlineplatform_met_expectations	Does the selected online platform met user's expectations	Object

Online_platform_flexible	Does the online platform flexible than offline	Integer
Most_convienent_platform	User’s most convenient platform	Object
Best_quality_content	User find the best quality content on which platform	Object
Recommend_onlineplatform_friend	Will user recommend the online platform to a friend or not	Integer
Online_value_for_money_than_offline	Does online platform is more valuable than offline platform	Object
Best_return_on_investment_platform	Does the preferred platform provides the best return on investment	Object
Best_aspect_online	Best aspect of online learning	Object
Best_aspect_offline	Best aspect of offline learning	Object
Online_platform_diversity	Does online platform more diverse than offline	Integer
Instructors_online_vs_offline	How are the instructors of online in comparison to offline	Integer
Most_interactive_platform	Which is the most interactive platform	Object
User_friendly_online	Does online platform more user friendly	Integer
Best_learning_material_platform	Platform with the best learning material	Object
Superior_support_service_platform	Platform with the most superior service	Object

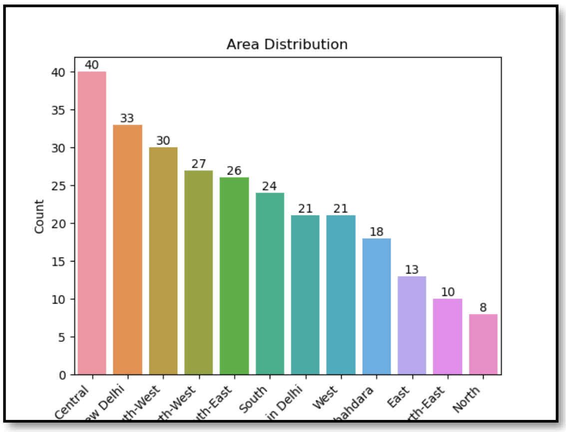
EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) is a critical step in the data analysis process, involving the thorough examination of a dataset to uncover underlying patterns, spot anomalies, test hypotheses, and check assumptions through visual and quantitative methods. In this project, EDA is employed to analyze the survey data collected on users' preferences and experiences with online and offline coaching. This includes data cleaning and preprocessing to handle missing values and outliers, followed by the use of descriptive statistics and visualizations such as histograms, bar charts, and box plots to summarize the demographic distribution and general perceptions of coaching among respondents. Through EDA, we aim to identify key trends and insights, such as the demographic factors that influence coaching preferences and the common reasons cited for favouring one mode over the other. This analysis not only helps in understanding the data better but also guides the feature selection and engineering process for the subsequent classification models. EDA provides a foundation for making informed decisions throughout the project, ensuring that the final models are built on a solid understanding of the underlying data patterns.

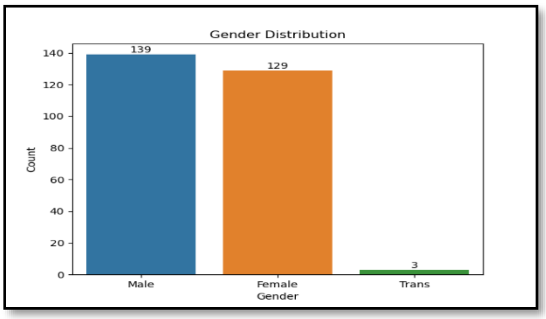
1) **AGE DISTRIBUTION-** This is a histogram representing the age of all users with their frequencies. According to this graph, the maximum quantity of users belongs between the age of ‘20-25 years’ which means it is mostly youth which has contributed to our study.



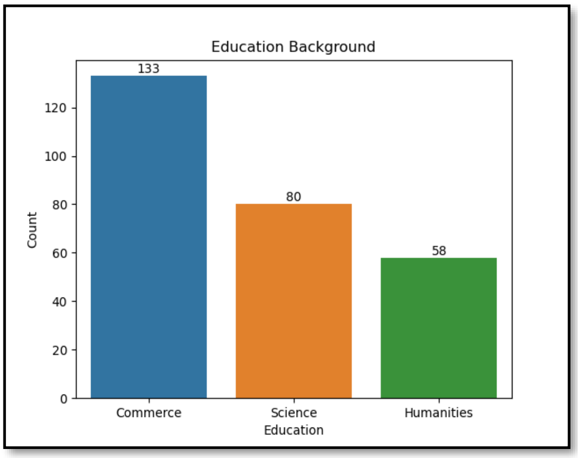
2) **AREA DISTRIBUTION-** This is a count plot representing the count of users according to the area which they belong. Here we can see that the maximum users belong to the ‘Central Delhi’ and minimum users belong to ‘North Delhi’. However, there are ‘21 users who don’t belong in Delhi’.



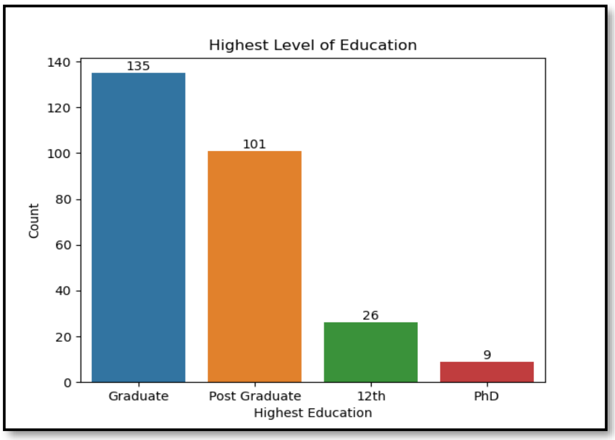
3) **GENDER DISTRIBUTION-** This is a count plot representing the count of users according to their genders. Here the maximum data is from 'Male' and the minimum from 'Trans'.



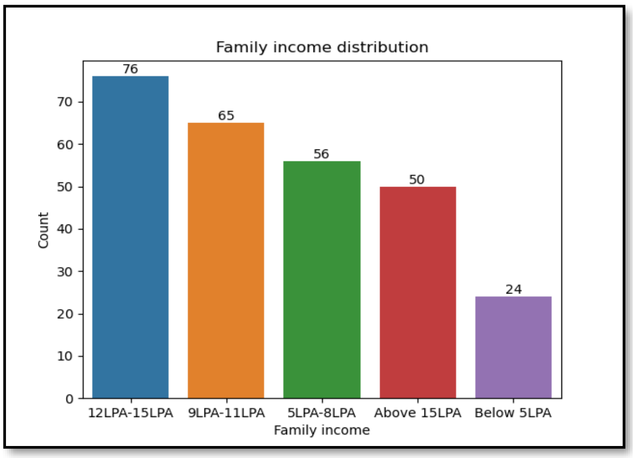
4) **EDUCATION BACKGROUND-** This is a count plot representing the count of users according to their respective background of education. Here the maximum data is from 'Commerce' background and minimum is from 'Humanities' background.



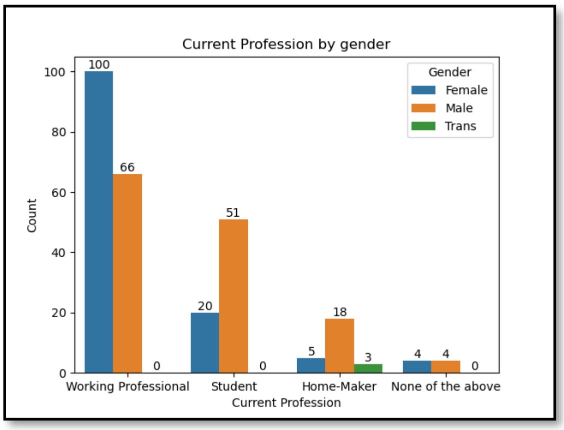
5) **HIGHEST LEVEL OF EDUCATION-** This is a count plot representing the count of users according to their highest level of education. Here the maximum data is of users who are 'Graduates' and the minimum data is of users who have 'PhD'.



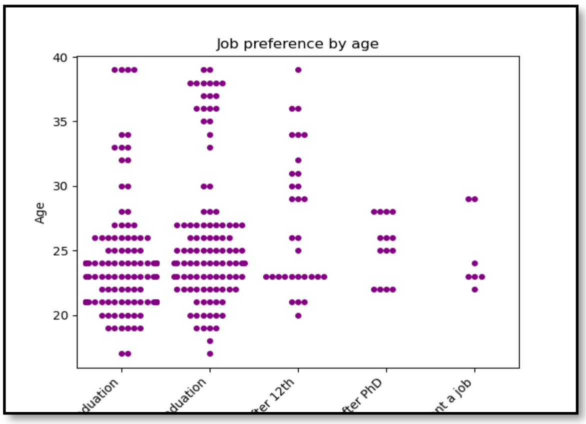
6) **FAMILY INCOME DISTRIBUTION-** This is a count plot representing the count of users according to their family income. Here the maximum data is of users who belong to '12LPA to 15LPA' earning family and least users belong to 'Below 5LPA' earning family.



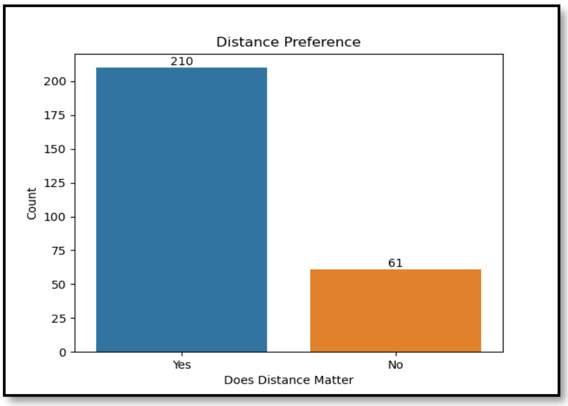
7) **CURRENT PROFESSION BY GENDER-** This is a count plot representing the current profession of the users by their gender. Here the maximum data is of 'Working professional', among them the highest gender of users is 'Female' followed by 'Male'.



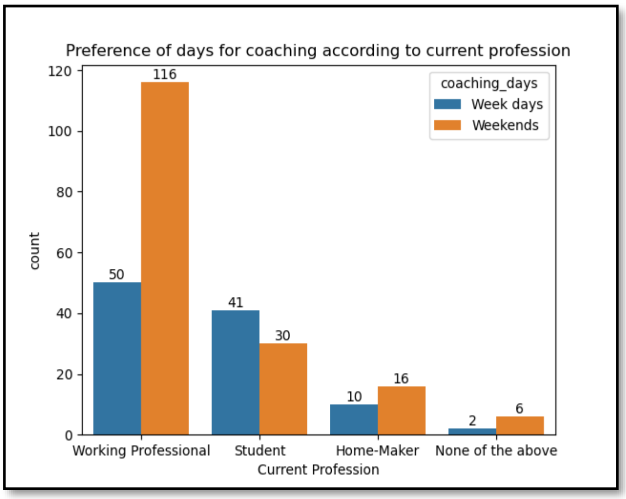
8) **JOB PREFERENCE ACCORDING TO AGE-** This is a swarm plot representing the job preferences (i.e. when should one do a job) of the users by their age. Here the maximum data belongs under the category of 'After Graduation' and 'After Post Graduation'.



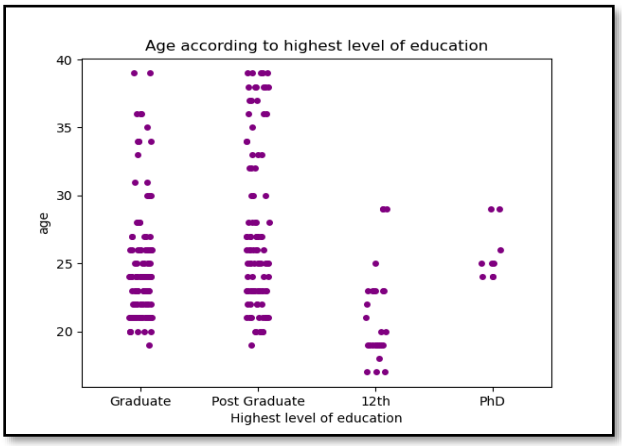
9) **DISTANCE PREFERENCE-** This is a count plot representing the distance preferences of user (i.e. whether the distance of coaching institute matter or not). Here the maximum users have voted that the distance does matter for them.



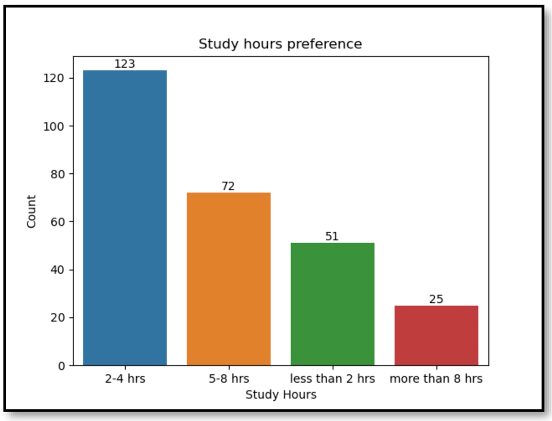
10) **PREFERENCE OF COACHING DAYS ACCORDING TO CURRENT PROFESSION-** This is a count plot representing the preferences of users for coaching days (i.e. week days or weekends) according to their current professions. The maximum data is of ‘Working Professional’, among them the highest preference is given for ‘weekends’ coaching rather week days.



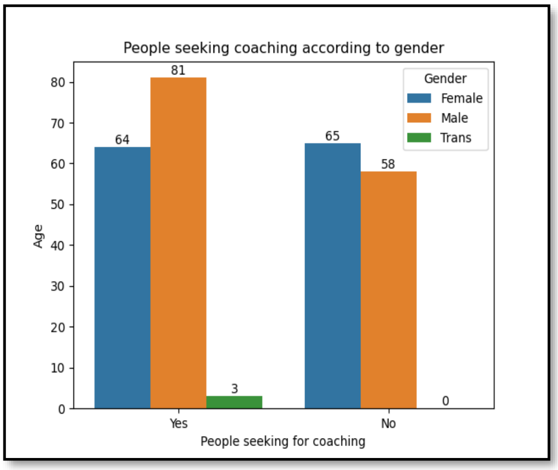
11) **AGE ACCORDING TO THE HIGHEST LEVEL OF EDUCATION-** This is a strip plot representing the highest level of education according to the age of users. Here we can see the spread of users by their respective ages.



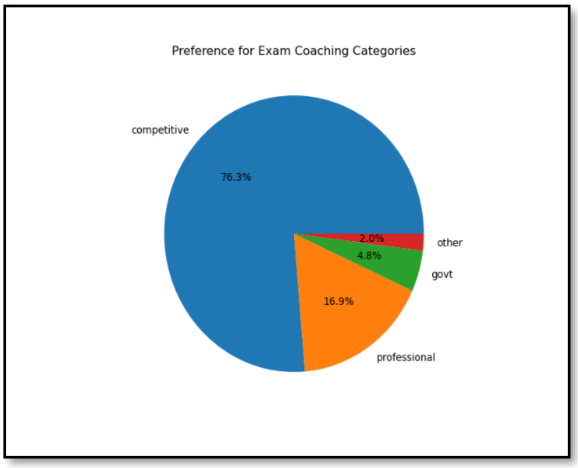
12) **STUDY HOURS PREFERENCE-** This is a count plot representing the hours of study preference of user. Here the maximum preference is given to '2-4 hours' of study and minimum preference is for 'More than 8hours'.



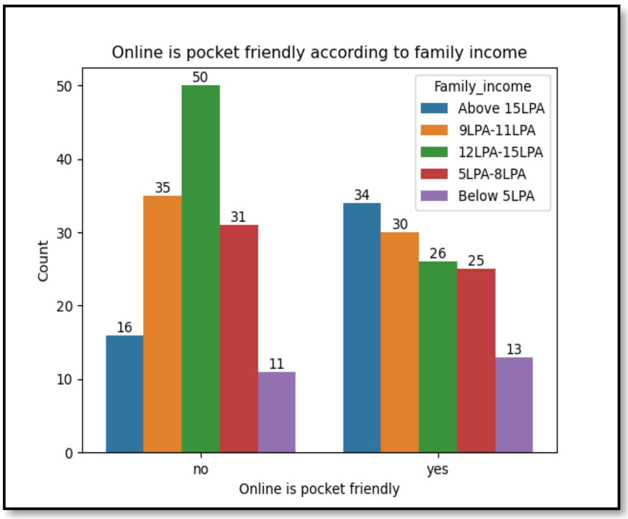
13) **PEOPLE SEEKING COACHING ACCORDING TO GENDER-** This is a count plot representing the users who are seeking for coaching at present according to their gender. Here we can see that the maximum people are seeking for coaching at present, among them the highest 'male' and the least is 'Trans'.



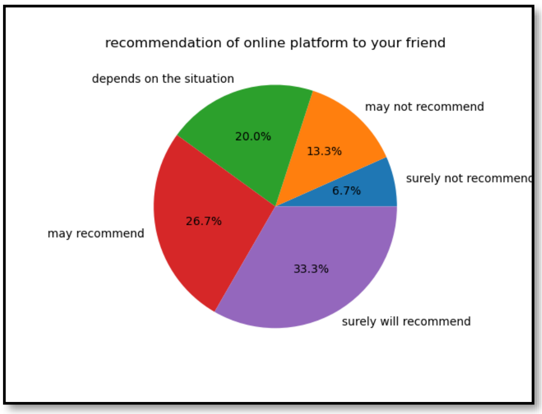
14) **PREFERENCE OF EXAM COACHING CATEGORIES-** This is a pie chart representing the exam coaching categories of users. Among which we can see that the maximum preference of the user is for 'Competitive' exams followed by 'Professional Courses'.



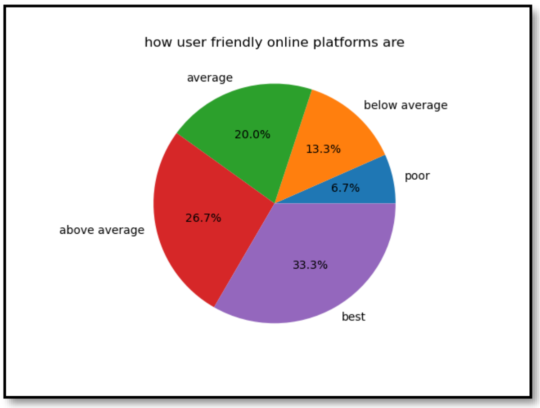
15) **IS ONLINE PLATFORM POCKET FRIENDLY ACCORDING TO THE FAMILY INCOME-** This is a count plot of users representing the preference whether the online platform is pocket friendly or not according to their respective family income. In our data maximum vote is for 'No' and among them the maximum users belong to the '12LPA to 15LPA' earning families.



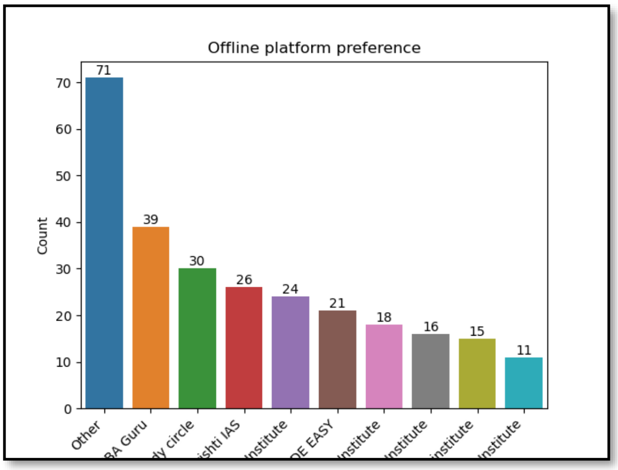
16) **RECOMMENDATION OF ONLINE PLATFORM TO YOUR FRIEND-** This is a pie chart representing whether the users will recommend the online platform to their friend or not. Here the maximum vote goes to 'surely will recommend' followed by 'may recommend'.



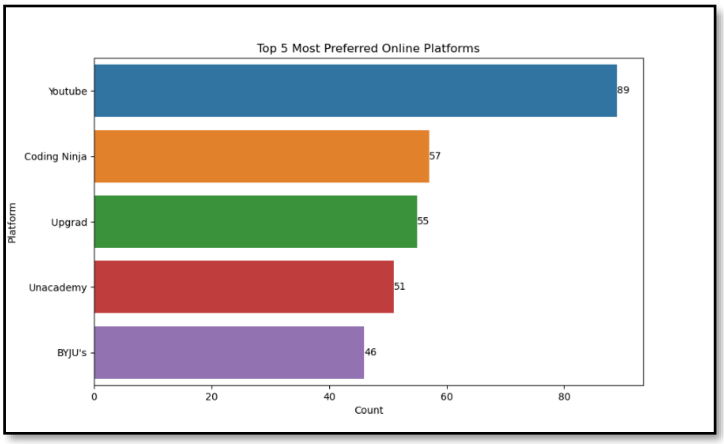
17) **HOW USER-FRIENDLY ONLINE PLATFORMS ARE-** This is a pie chart representing whether the online platforms are user-friendly. Here the maximum vote goes to 'best' followed by 'above average'.



18) OFFLINE PLATFORM PREFERENCE- This is a count plot representing the offline platforms user preferred. Here the maximum vote goes to 'Other' offline platforms followed by 'MBA Guru' and the minimum vote goes to 'Aakash Institute'.



19) TOP 5 MOST PREFERRED ONLINE PLATFORMS- This is a count plot representing the top 5 most preferred online platform by the user. Here the maximum vote is for 'YouTube' followed by 'Coding Ninja'.



CLASSIFICATION ALGORITHMS

In our project comparing online versus offline coaching, we employ various classification algorithms to analyse and predict outcomes based on collected data. Classification algorithms are a subset of supervised learning techniques in machine learning, designed to categorize data into predefined classes. By training these algorithms on labelled datasets, we can predict the class labels of new, unseen data. The goal is to understand the effectiveness and student outcomes from online and offline coaching methods, ultimately determining which mode yields better results under different circumstances. The use of multiple classification algorithms allows us to compare their performance and ensure the robustness of our findings.

LOGISTIC REGRESSION

Logistic regression is a statistical method used for binary classification problems, where the outcome is a dichotomous variable (i.e., it has two possible outcomes). It models the probability that a given input belongs to a particular class using the logistic function, which outputs values between 0 and 1. In our project, logistic regression can help us determine the likelihood of a student achieving a certain level of success based on features like study hours, mode of coaching, and prior academic performance. Despite its simplicity, logistic regression is powerful for interpreting the relationships between features and the target variable.

SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) is a robust classification algorithm that works well for both linear and non-linear data. It aims to find the optimal hyperplane that best separates the data points of different classes in a high-dimensional space. The points closest to the hyperplane, known as support vectors, are critical in defining the decision boundary. In the context of our project, SVM can be used to classify student success rates in online versus offline coaching by identifying the most influential factors and separating successful students from less successful ones. SVM is particularly effective when dealing with complex datasets where the classes are not easily separable.

RANDOM FOREST

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) of the individual trees. This approach helps to improve the model's accuracy and control overfitting. Each decision tree is built from a random subset of the training data and features, which enhances the model's generalization capabilities. For our project, Random Forest can analyse various attributes of online and offline coaching environments and predict student performance. Its ability to handle a large number of input variables and maintain high accuracy makes it ideal for our comparative study of different coaching methods.

MODEL DEVELOPMENT

- Before starting with the algorithms, we have to convert the categorical data into dummy variable i.e. (True/False).
- We have considered 'The Most Preferred Mode of Study' by the user to be the dependent variable.
- After that we will convert this data into integer format i.e. (0 to 1) where 0 is offline and 1 is online.
- We then split the data using Sklearn library into Training data (80% of dataset) and Test data (20% of dataset).

CLASSIFICATION REPORTS

LOGISTIC REGRESSION REPORT

Logistic Regression Report:

	precision	recall	f1-score	support
0	0.89	0.91	0.90	35
1	0.84	0.80	0.82	20
accuracy			0.87	55
macro avg	0.87	0.86	0.86	55
weighted avg	0.87	0.87	0.87	55

SVM REPORT

```
SVM Report:
              precision    recall  f1-score   support

           0       0.64      1.00      0.78        35
           1       0.00      0.00      0.00        20

    accuracy                           0.64        55
   macro avg       0.32      0.50      0.39        55
weighted avg       0.40      0.64      0.49        55
```

RANDOM FOREST REPORT

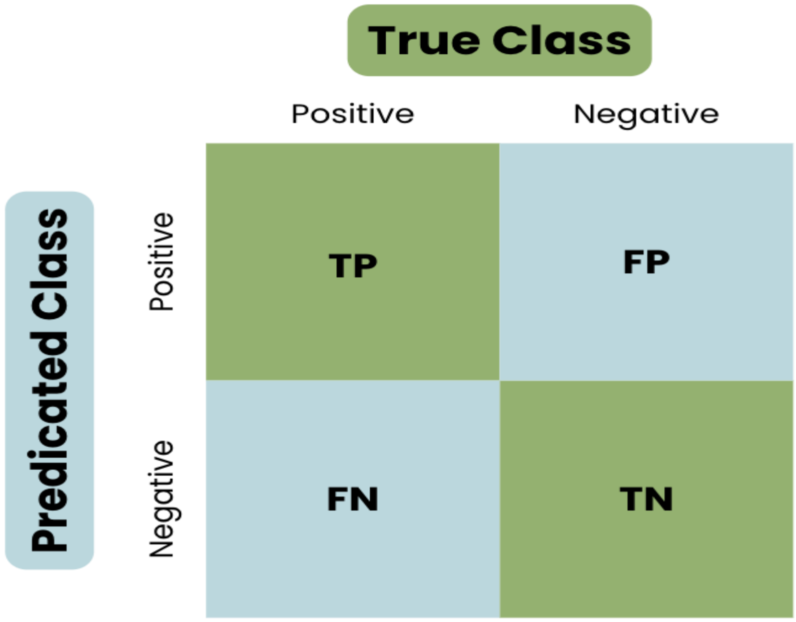
```
Random Forest Report:
              precision    recall  f1-score   support

           0       0.79      0.94      0.86        35
           1       0.85      0.55      0.67        20

    accuracy                           0.80        55
   macro avg       0.82      0.75      0.76        55
weighted avg       0.81      0.80      0.79        55
```

CONFUSION MATRIX

A confusion matrix is a performance evaluation tool used in classification tasks to visualize the performance of a machine learning model. It presents a summary of the predictions made by the model compared to the actual ground truth across different classes.



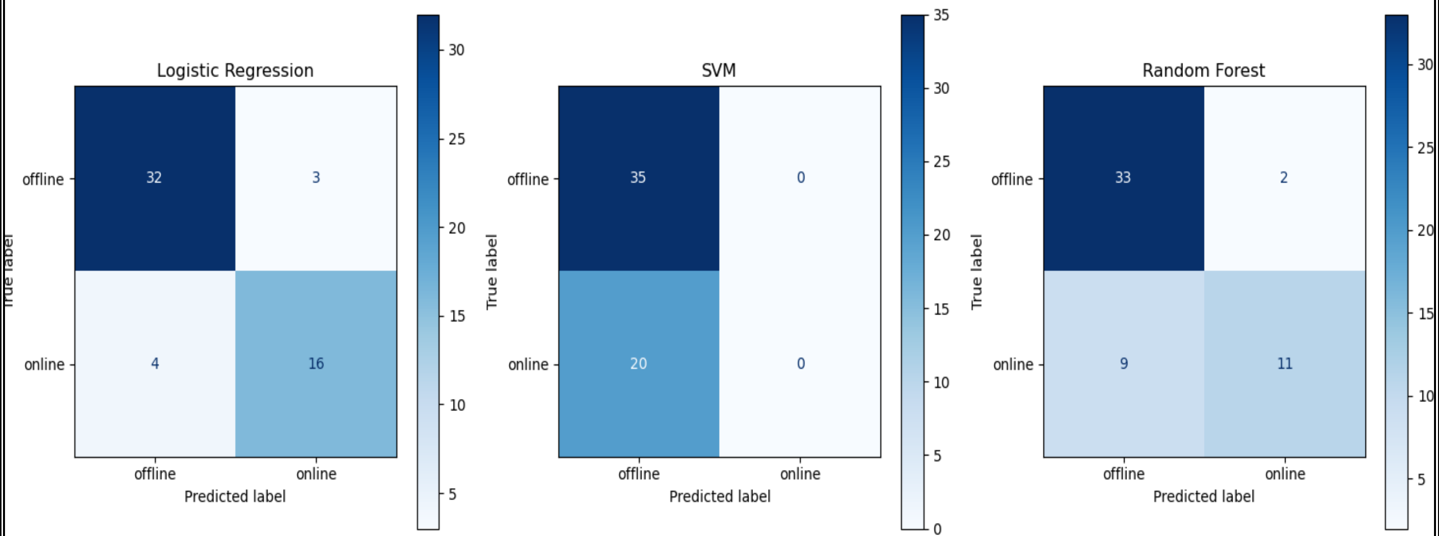
COMPONENTS OF CONFUSION MATRIX

1) True Positive (TP): The instances that were correctly predicted as positive (or belonging to the target class) by the model.

2) True Negative (TN): The instances that were correctly predicted as negative (or not belonging to the target class) by the model.

3) False Positive (FP): The instances that were incorrectly predicted as positive by the model when they actually belong to the negative class. Also known as Type I error.

4) False Negative (FN): The instances that were incorrectly predicted as negative by the model when they actually belong to the positive class. Also known as Type II error.



INTERPRETATIONS

After analyzing the performance of three classification algorithms—Logistic Regression, Support Vector Machine (SVM), and Random Forest, we can draw the following conclusions:

1. Logistic Regression:

- Precision, Recall, and F1-Score: The logistic regression model achieved a high precision of 0.89 for class 0 (offline) and 0.84 for class 1 (online). The recall values were 0.91 for class 0 (offline) and 0.80 (online) for class 1, resulting in f1-scores of 0.90 and 0.82, respectively.
- Overall Accuracy: The model attained an overall accuracy of 87%.
- Observations: This model demonstrates a balanced performance across both classes, making it a reliable choice for this classification task.

2. Support Vector Machine (SVM):

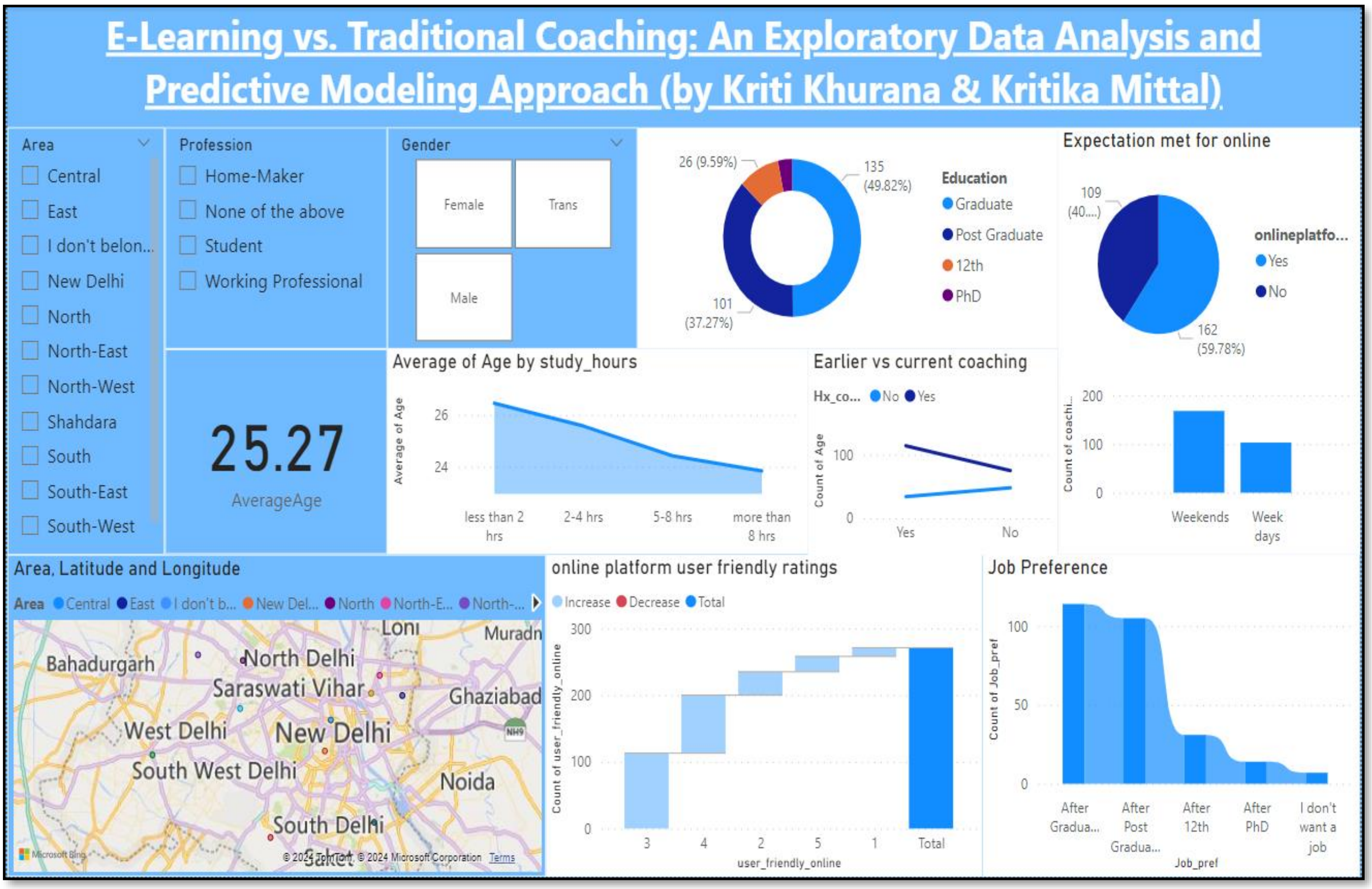
- Precision, Recall, and F1-Score: The SVM model showed a high precision of 0.64 for class 0 (offline), but it failed to predict class 1 (online) correctly, yielding a precision of 0.00. The recall was perfect (1.00) for class 0 (offline) but 0.00 for class 1 (online), leading to f1-scores of 0.78 for class 0 (offline) and 0.00 for class 1 (online).
- Overall Accuracy: The accuracy was 64%.
- Observations: The SVM model struggled with class 1 (online) predictions, indicating it may not be suitable for this dataset, possibly due to class imbalance or the need for parameter tuning.

3. Random Forest:

- Precision, Recall, and F1-Score: The random forest model achieved a precision of 0.79 for class 0 (offline) and 0.85 for class 1 (online). The recall values were 0.94 for class 0 (offline) and 0.55 for class 1 (online), resulting in f1-scores of 0.86 and 0.67, respectively.
- Overall Accuracy: The model attained an overall accuracy of 80%.
- Observations: This model performed well with a good balance between precision and recall, particularly excelling in class 0 (offline) predictions but less so in class 1 (online). Fine-tuning the model might improve its performance further.

POWER BI REPORT

Power BI is a powerful business analytics tool developed by Microsoft that allows users to visualize data and share insights across their organization. A Power BI report is an interactive and dynamic compilation of visualizations, such as charts, graphs, and tables, that provide a comprehensive view of business metrics and key performance indicators (KPIs). These reports are designed to be user-friendly and highly customizable, enabling users to drill down into data, filter information, and gain deeper insights through intuitive dashboards. Power BI reports can be connected to various data sources, including Excel spreadsheets, cloud-based services, and on-premises databases, ensuring seamless data integration and real-time updates. This flexibility makes Power BI an essential tool for decision-makers seeking to analyse trends, monitor business performance, and make data-driven decisions. Additionally, the collaborative features of Power BI allow team members to share reports and collaborate on data analysis, fostering a data-centric culture within the organization.



CONCLUSION

In this study comparing online and offline coaching preferences, we utilized logistic regression, SVM, and random forest algorithms to analyse and classify user data. The results show distinct preferences and trends among users based on various demographics and criteria.

The logistic regression model demonstrated a high overall accuracy of 87%, with precision and recall rates indicating that offline coaching (class 0) is slightly more preferred and accurately predicted compared to online coaching (class 1). The support values indicate more respondents favour offline coaching. In contrast, the SVM model showed a significant bias, favouring offline coaching with an accuracy of 64%, but failing to accurately predict preferences for online coaching. The random forest model provided a more balanced view with an 80% accuracy, indicating a reasonable prediction capability for both classes but with a higher precision for offline coaching.

Analysing the demographic distribution, the majority of users are aged between 20-25 years, indicating that young adults are the primary demographic engaging in coaching.

Distance is a crucial factor for coaching preferences, with most users indicating that proximity matters. The majority prefer coaching during weekends, especially working professionals. Study hours preference trends towards 2-4 hours daily. Gender-wise, more males seek coaching, with competitive exams being the primary focus.

Despite the higher inclination towards offline platforms like MBA Guru, there is a strong recommendation for online platforms, deemed user-friendly by most. However, a significant number feel online platforms are not pocket-friendly, especially among higher income families. YouTube emerges as the top preferred online platform, followed by Coding Ninja.

In conclusion, while offline coaching retains a robust preference, online platforms are gaining traction, especially among tech-savvy young adults. These insights highlight the need for both coaching modalities to adapt and cater to evolving user preferences and demographics.

BIBLIOGRAPHY

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10183331/#:~:text=The%20study's%20findings%20show%20that,digital%20learning%20is%20also%20increased

2. Pokhrel S., Chhetri R. (2021). A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Education for the Future,* 8(1), 133–141. 10.1177/2347631120983481 [CrossRef] [Google Scholar]

3. Zhu X., Liu J. (2020). Education in and after COVID-19: Immediate responses and long-term visions. *Postdigital Science and Education,* 2, 695–699. 10.1007/s42438-020-00126-3 [CrossRef] [Google Scholar]

4. Iivari N., Sharma S., Ventä-Olkkonen L. (2020). Digital transformation of everyday life—How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care? *International Journal of Information Management,* 55(102183), 102183. 10.1016/j.ijinfomgt.2020.102183 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

5. https://chatgpt.com/?oai-dm=1
6. https://oswalpublishers.com/wp-content/uploads/2022/10/online-and-offline-classes.webp
7. https://www.google.com/imgres?q=online%20classes%20merits%20and%20demerits&imgurl=https%3A%2F%2Fapi.edoxi.com%2Fassets%2Fstudyhub%2FAdvantages_and_Disadvantages_of_Online_Learning.webp&imgrefurl=https%3A%2F%2Fwww.edoxi.com%2Fstudyhub-detail%2Fadvantages-and-disadvantages-of-online-learning&docid=eipep0I79rqRDM&tbnid=8HR4kmqNkXNyGM&vet=12ahUKEwitgsjDxcSGAxXCcmwGHYKsBrwQM3oECGoQAA..i&w=1532&h=699&hcb=2&ved=2ahUKEwitgsjDxcSGAxXCcmwGHYKsBrwQM3oECGoQAA
8. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTbNs4sMjr8R_PRt4j4jBPWxB6-eYCDfNtN6g&s
9. https://www.google.com/imgres?q=online%20classes%20merits%20and%20demerits&imgurl=https%3A%2F%2Fiotcdn.oss-ap-southeast-

1.aliyuncs.com%2Fstyles%2Fa_share_fb%2Foss%2FCourse.jpg&imgrefurl=https%3A%2F%2Fieltsonlinetests.com%2Fwriting-correction%2Fadvantages-and-disadvantages-studying-online-course-corrected-essay&docid=NVfqM0UYgQoadM&tbnid=HAkdxoSgAIyE4M&vet=12ahUKEwitgsjDxcSGAxXCcmwGHYKsBrwQM3oECGAQAA..i&w=1200&h=630&hcb=2&ved=2ahUKEwitgsjDxcSGAxXCcmwGHYKsBrwQM3oECGAQAA

10. 1. Allen, I. E., & Seaman, J. (2011). "Going the Distance: Online Education in the United States, 2011." Sloan Consortium. https://onlinelearningconsortium.org/survey_report/2011-going-distance-online-education-united-states/
11. 2. Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). "Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies." U.S. Department of Education. https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf
12. 3. Bates, A. W. (2015). "Teaching in a Digital Age: Guidelines for Designing Teaching and Learning." Tony Bates Associates Ltd. https://opentextbc.ca/teachinginadigitalage/
13. 4. Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). "A Meta-Analysis of Blended Learning and Technology Use in Higher Education: From the General to the Applied." Journal of Computing in Higher Education, 26(1), 87-122. https://link.springer.com/article/10.1007/s12528-013-9077-3
14. 5. Kearsley, G., & Shneiderman, B. (1999). "Engagement Theory: A Framework for Technology-Based Teaching and Learning." Educational Technology, 38(5), 20-23. http://home.sprynet.com/~gkearsley/engage.htm
15. 6. Garrison, D. R., & Kanuka, H. (2004). "Blended Learning: Uncovering Its Transformative Potential in Higher Education." The Internet and Higher Education, 7(2), 95-105. https://www.sciencedirect.com/science/article/pii/S1096751604000156
16. 7. Tinto, V. (1997). "Classrooms as Communities: Exploring the Educational Character of Student Persistence." The Journal of Higher Education, 68(6), 599-623. https://www.jstor.org/stable/2959965
17. 8. Wentzel, K. R. (1999). "Social-Motivational Processes and Interpersonal Relationships: Implications for Understanding Motivation at School." Journal of Educational Psychology, 91(1), 76-97. https://psycnet.apa.org/doi/10.1037/0022-0663.91.1.76

18. 9. Xu, D., & Jaggars, S. S. (2013). "The Impact of Online Learning on Students' Course Outcomes: Evidence from a Large Community and Technical College System." Economics of Education Review, 37, 46-57. https://www.sciencedirect.com/science/article/pii/S0272775713000711
19. 10. Al-Qahtani, A. A., & Higgins, S. E. (2013). "Effects of Traditional, Blended, and E-Learning on Students' Achievement in Higher Education." Journal of Computer Assisted Learning, 29(3), 220-234. https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2729.2012.00490.x
20. 11. Bergmann, J., & Sams, A. (2012). "Flip Your Classroom: Reach Every Student in Every Class Every Day." International Society for Technology in Education. https://www.iste.org/resources/product?ID=2285
21. 12. Chen Hsieh, J. S., Wu, W. C. V., & Marek, M. W. (2017). "Using the Flipped Classroom to Enhance EFL Learning." Computer Assisted Language Learning, 30(1-2), 1-21. https://www.tandfonline.com/doi/abs/10.1080/09588221.2015.1111910
22. 13. Holmes, W., Bialik, M., & Fadel, C. (2019). "Artificial Intelligence in Education: Promises and Implications for Teaching and Learning." Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf
23. https://images.datacamp.com/image/upload/v1701364260/image_5baaeac4c0.png