17MDC82 – Decision Support Systems

Revitalizing EduQuest - An EdTech Company

Team members:

2033012 Genga K

2033017 Kavya Sree R J

Introduction:

EduQuest, once a promising player in the EdTech industry, encountered challenges leading to its decline. To reignite its success, a comprehensive resurgence strategy was devised. This project delineates the methodology employed to analyze data, identify shortcomings, and implement innovative solutions using dynamic programming, greedy algorithms, and Business Intelligence (BI) tools such as Power BI.

Aim:

The objective is to revive EduQuest by scrutinizing historical data, implementing dynamic learning trails, time management features, and integrating gamification principles to enhance student engagement and academic performance.

Objectives:

- Analyze data to discern past performance and unearth areas for enhancement.
- Implement dynamic learning trails tailored to individual learning preferences, pacing, and proficiency levels.
- Develop a time management feature to optimize study schedules and refine time allocation for students.
- Infuse gamification principles to heighten student engagement and motivation.
- Utilize quizzes to pinpoint weaker areas in students' knowledge and recommend targeted learning concepts.

System Description:

The system amalgamates data simulation and analysis, dynamic programming, greedy algorithms, and Power BI to bolster decision-making. Historical data is generated using simulation, while Power BI is employed for visualizing insights. Dynamic learning trails, time management, and gamification features are integrated into the EduQuest platform to enrich the learning experience.

Functional Requirements:

- Data simulation, cleansing, and preprocessing.
- Dynamic learning trail formulation based on historical data and student preferences.
- Time management feature for optimizing study schedules.
- Integration of gamification elements for augmented engagement.
- Quiz functionality to identify weaker areas and recommend focused learning concepts.

Modules:

1. **Data generation and Management:** The Data Management module handles a diverse dataset comprising manually curated sample data and simulated information. This dataset is meticulously organized in CSV format for easy accessibility and manipulation. Leveraging SQLite3, this module efficiently collects, cleans, and preprocesses the extensive dataset, ensuring its quality for subsequent insightful analysis. To facilitate insightful analysis, the program starts by loading the initial sample data from an Excel file and then generates a simulated dataset of approximately 3500 rows using statistical techniques for data-driven decision-making throughout the EdTech revitalization initiative.

2.**Dynamic Learning Trails:** EduQuest employs dynamic programming methodologies to curate adaptive learning trails customized to each student's unique learning preferences, pace, and proficiency levels. By analyzing historical data and performance metrics, the system dynamically adjusts the sequence and complexity of learning modules, ensuring an optimal learning trajectory.

3.**Time Management:** Through the application of dynamic programming algorithms, EduQuest facilitates effective time management for students. By considering variables such as academic objectives, available study time, and impending deadlines, the software generates optimized study schedules, minimizing time wastage and maximizing productivity.

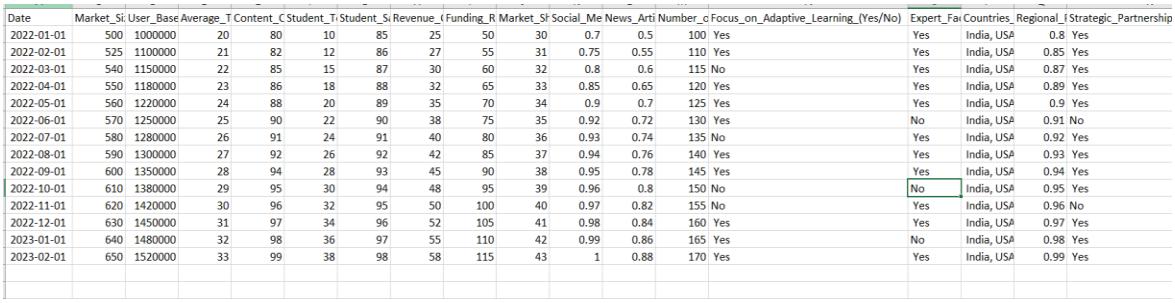
4.**Gamification Integration:** EduQuest integrates gamification principles into the learning process, utilizing greedy algorithms to optimize game mechanics for heightened engagement and motivation. Through gamified challenges, rewards, and progress tracking mechanisms, students

are incentivized to actively participate in their learning journey, while experiencing instantaneous feedback and rewards

5.**Quiz and Assessment:** Identify weaker areas in student's knowledge and recommend targeted learning concepts. The Quiz and Assessment module facilitates the creation of quizzes with various question formats. It incorporates intelligent assessment algorithms to pinpoint areas of weakness, providing valuable insights for personalized learning plans.

Dataset/Database Description:

The dataset utilized for this project comprises a combination of manually gathered sample data and simulated data generated through statistical techniques.



Date	Market_Si	User_Base	Average_T	Content_C	Student_Ti	Student_Si	Revenue_(	Funding_R	Market_Sh	Social_Me	News_Arti	Number_o	Focus_on_Adaptive_Learning_(Yes/No)	Expert_Fa	Countries_	Regional_I	Strategic_Partnership
2022-01-01	500	1000000	20	80	10	85	25	50	30	0.7	0.5	100	Yes	Yes	India, USA	0.8	Yes
2022-02-01	525	1100000	21	82	12	86	27	55	31	0.75	0.55	110	Yes	Yes	India, USA	0.85	Yes
2022-03-01	540	1150000	22	85	15	87	30	60	32	0.8	0.6	115	No	Yes	India, USA	0.87	Yes
2022-04-01	550	1180000	23	86	18	88	32	65	33	0.85	0.65	120	Yes	Yes	India, USA	0.89	Yes
2022-05-01	560	1220000	24	88	20	89	35	70	34	0.9	0.7	125	Yes	Yes	India, USA	0.9	Yes
2022-06-01	570	1250000	25	90	22	90	38	75	35	0.92	0.72	130	Yes	No	India, USA	0.91	No
2022-07-01	580	1280000	26	91	24	91	40	80	36	0.93	0.74	135	No	Yes	India, USA	0.92	Yes
2022-08-01	590	1300000	27	92	26	92	42	85	37	0.94	0.76	140	Yes	Yes	India, USA	0.93	Yes
2022-09-01	600	1350000	28	94	28	93	45	90	38	0.95	0.78	145	Yes	Yes	India, USA	0.94	Yes
2022-10-01	610	1380000	29	95	30	94	48	95	39	0.96	0.8	150	No	No	India, USA	0.95	Yes
2022-11-01	620	1420000	30	96	32	95	50	100	40	0.97	0.82	155	No	Yes	India, USA	0.96	No
2022-12-01	630	1450000	31	97	34	96	52	105	41	0.98	0.84	160	Yes	Yes	India, USA	0.97	Yes
2023-01-01	640	1480000	32	98	36	97	55	110	42	0.99	0.86	165	Yes	No	India, USA	0.98	Yes
2023-02-01	650	1520000	33	99	38	98	58	115	43	1	0.88	170	Yes	Yes	India, USA	0.99	Yes

Fig 1.1 Sample data

Sample Data

The initial dataset consists of 20 rows as shown in fig1.1, meticulously curated from web articles, news pages, and other reputable sources within the EdTech industry. These sources provided insights into various metrics relevant to EduQuest's performance and market dynamics. The data covers a diverse range of features including:

- Market Size (in millions USD): The estimated market size in terms of monetary value.
- User Base (Number of Users): The total number of users utilizing EduQuest's platform.
- Average Time Spent (in minutes): The average duration users spend per session.
- Content Completion Rate (%): The percentage of completed educational content by users.
- Student Test Score Improvement (%): Improvement percentage in students' test scores attributed to EduQuest.
- Student Satisfaction (%): Satisfaction level of students using EduQuest's services.
- Revenue Growth (Year-over-Year %): The annual growth rate of EduQuest's revenue.
- Funding Round Amount (in millions USD): Amount of funding raised by EduQuest in various funding rounds.
- Market Share (%): EduQuest's market share within the EdTech industry.

- Social Media Sentiment Score: Sentiment analysis score derived from social media discussions.
- News Article Sentiment Score: Sentiment analysis score derived from news articles.
- Number of Courses Offered: The total count of courses available on the EduQuest platform.
- Focus on Adaptive Learning (Yes/No): Indicates whether EduQuest emphasizes adaptive learning techniques.
- Expert Faculty (Yes/No): Presence of expert faculty members on the platform.
- Countries of Operation: The countries where EduQuest operates.
- Regional Performance Score: Performance rating across different regions of operation.
- Strategic Partnerships (Yes/No): Engagement in strategic partnerships.
- Government Initiatives (Yes/No): Participation in government initiatives.
- EduQuest's Future Plans: Future plans or strategies of EduQuest.

Simulated Data

To enrich the dataset and provide a more comprehensive analysis, simulated data was generated using statistical techniques. The simulated dataset includes approximately 3500 rows and covers the same set of features as the sample data.

The program starts by loading the initial sample data from an Excel file.It then proceeds to generate a sequence of dates spanning from January 1, 2015, to December 1, 2023. For numerical columns, it calculates the mean and standard deviation of each attribute from the sample data and generates random values based on these statistics. Categorical column values are randomly selected from the sample data. This process ensures that the simulated data closely resembles the statistical properties of the original dataset. Finally, the generated dataset is obtained.

```
# Numerical columns
numeric_columns = sample_data.select_dtypes(include=[np.number]).columns
for column in numeric_columns[1:]:
    mean = sample_data[column].mean()
    std = sample_data[column].std()
    simulated_data[column] = np.random.normal(loc=mean, scale=std, size=len(all_dates))
```

Fig 1.2 Numerical Data simulation

```
# Categorical columns
categorical_columns = sample_data.select_dtypes(include=['object']).columns
for column in categorical_columns:
    simulated_data[column] = np.random.choice(sample_data[column], size=len(all_dates))
```

Fig 1.3 Categorical Data simulation

Data Format

Both the sample and simulated data are organized in tabular format, with each row representing a specific date and corresponding metrics for that date. The data is stored in CSV (Comma-Separated Values) format, facilitating easy access, manipulation, and analysis using various data processing tools and techniques.

Tools and Techniques:

- Programming: Python for data analysis and algorithmic implementation.
- Database: SQLite3 for data management.
- BI Tool: Power BI for visualization and reporting.
- Algorithms: Dynamic programming and greedy algorithms for trail generation and gamification.

Data Simulation using Python:

```python
import pandas as pd
import numpy as np

# Load the sample data from Excel
sample_data = pd.read_excel("Data_for_bi_initial.xlsx")

# Generate dates between 2015-01-01 and 2023-12-01
all_dates = pd.date_range(start='2015-01-01', end='2023-12-01', freq='D')

# Create a DataFrame with all dates
simulated_data = pd.DataFrame({'Date': all_dates})

# Numerical columns
numeric_columns = sample_data.select_dtypes(include=[np.number]).columns
for column in numeric_columns[1:]:
    mean = sample_data[column].mean()
    std = sample_data[column].std()
    simulated_data[column] = np.random.normal(loc=mean, scale=std, size=len(all_dates))

# Categorical columns
categorical_columns = sample_data.select_dtypes(include=['object']).columns
for column in categorical_columns:
    simulated_data[column] = np.random.choice(sample_data[column], size=len(all_dates))

# If "Market_Size_(in_millions_USD)" is missing, generate random values for it
if "Market_Size_(in_millions_USD)" not in simulated_data.columns:
    mean_market_size = sample_data["Market_Size_(in_millions_USD)"].mean()
    std_market_size = sample_data["Market_Size_(in_millions_USD)"].std()
    simulated_data["Market_Size_(in_millions_USD)"] = np.random.normal(loc=mean_market_size, scale=std_market_size, size=len(all_dates))

# Display first few rows of simulated data
print(simulated_data.head())
all_dates = pd.date_range(start='2015-01-01', end='2023-12-01', freq='D')
simulated_data['Date']=all_dates
# Save simulated data to a CSV file
simulated_data.to_csv('simulated_data_final.csv', index=False)
```

Fig 1.4 Code for simulation

The code provided in Fig 1.4 represents the simulation process utilized to generate synthetic data to complement the initial sample dataset. It demonstrates a systematic approach to mimicking the statistical characteristics and distributions observed in the original data. Through

statistical techniques such as normal distribution and random selection, the code efficiently generates simulated values for both numerical and categorical attributes. Additionally, the code incorporates a mechanism to handle missing data, ensuring the completeness and integrity of the simulated dataset. Finally, the simulated dataset is exported to a CSV file, enabling further analysis and exploration of the synthesized data. This code serves as a crucial component in expanding the dataset for comprehensive analysis and strategic decision-making.

BI Dashboard:

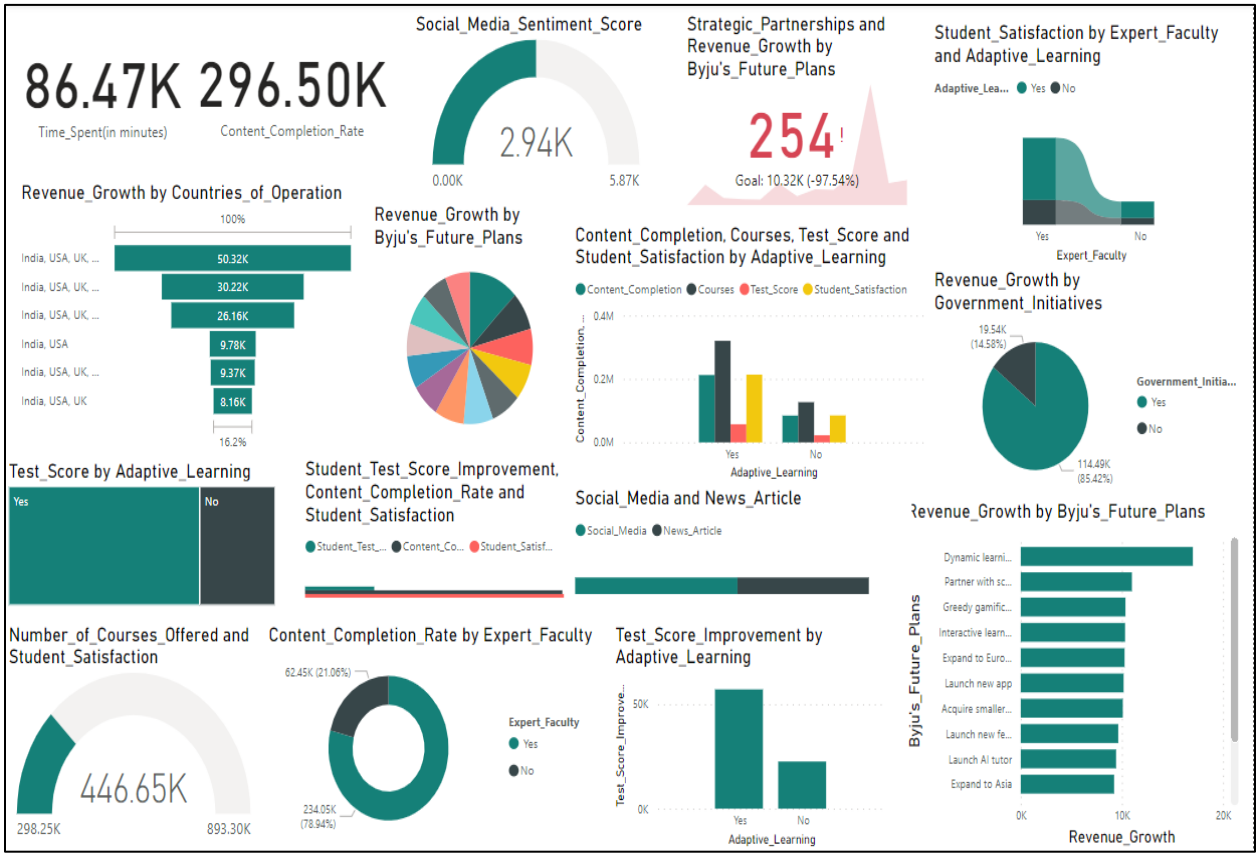
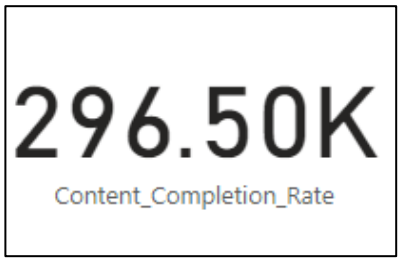


Fig 1.5 Dashboard

The above fig 1.5 showcases the complete dashboard designed to visualize the dataset, facilitating insightful analysis and interpretation of the data.



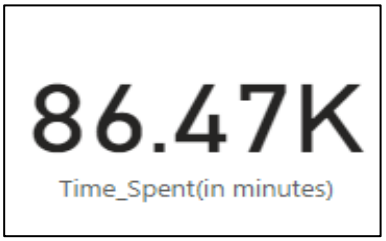


Fig 1.6 Course completion rate

Fig 1.7 Time spent

The above fig 1.6 and 1.7 shows the course completion rate and time spent by each student in completing a course.

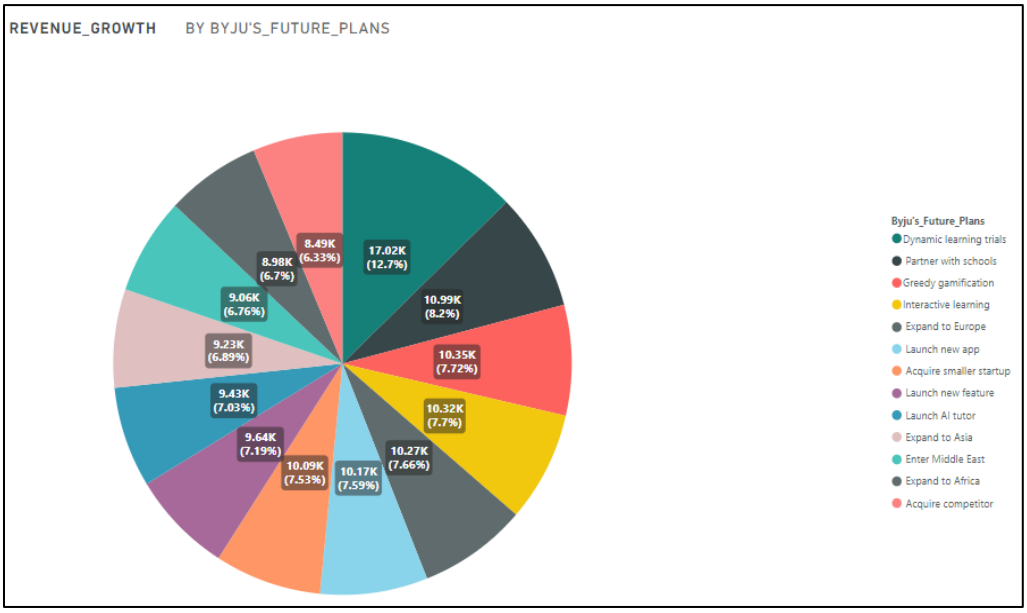


Fig 1.8 Revenue by Future plan

The figure 1.8 illustrates the future strategies designed to augment the company's revenue. Among the top four future plans, the highest contributor is Dynamic Learning, projected to contribute 13% to revenue growth. Following closely is Partnering with Schools, anticipated to yield an 8% increase in revenue. Additionally, the implementation of Gamification Techniques is expected to account for 7.72% of revenue growth, while Interactive Learning is projected to contribute 7.7%. Collectively, these top four strategies are estimated to drive approximately 40% growth in revenue. These future plans represent strategic initiatives aimed at leveraging innovative educational methodologies and partnerships to enhance the company's financial performance.

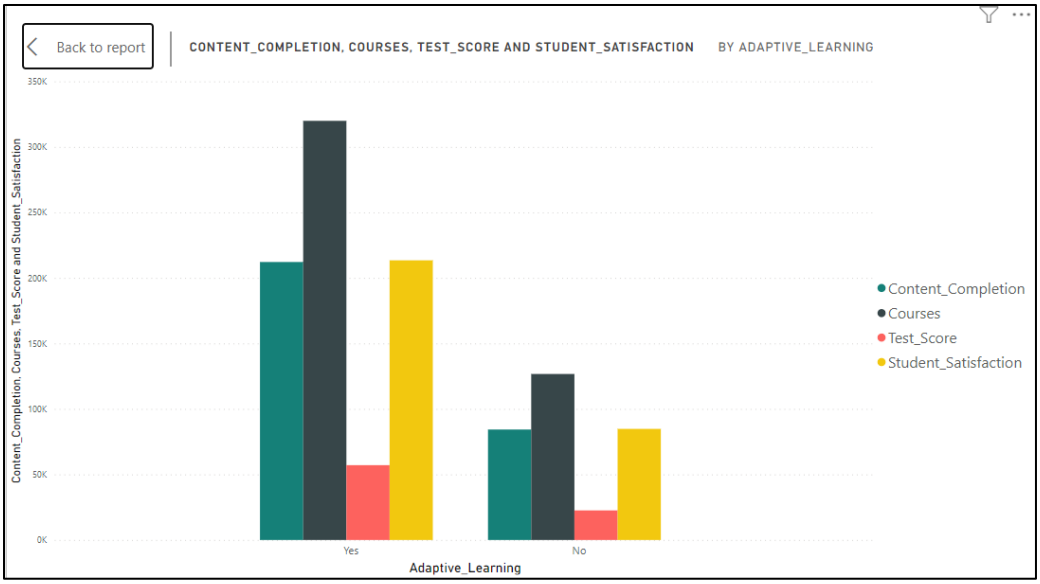


Fig 1.9 Range of Adaptive Learning of a student

The above fig 1.9 shows that adaptive learning plays an important role in improving course completion and student satisfaction. Hence we should improve our focus on implementing adaptive learning strategies.

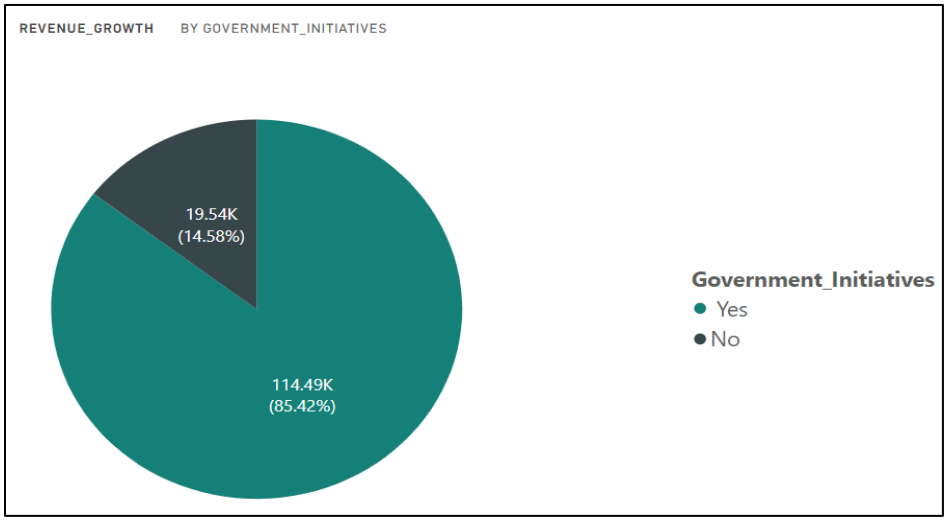


Fig 1.10 Revenue by Government initiatives

This chart shown in Fig 1.10 depicts the revenue generated due to the initiatives taken by the government is 85.42%. The revenue generated without the initiatives by the government is 14.58%. Thus government initiatives must be inculcated more to generate better revenue.

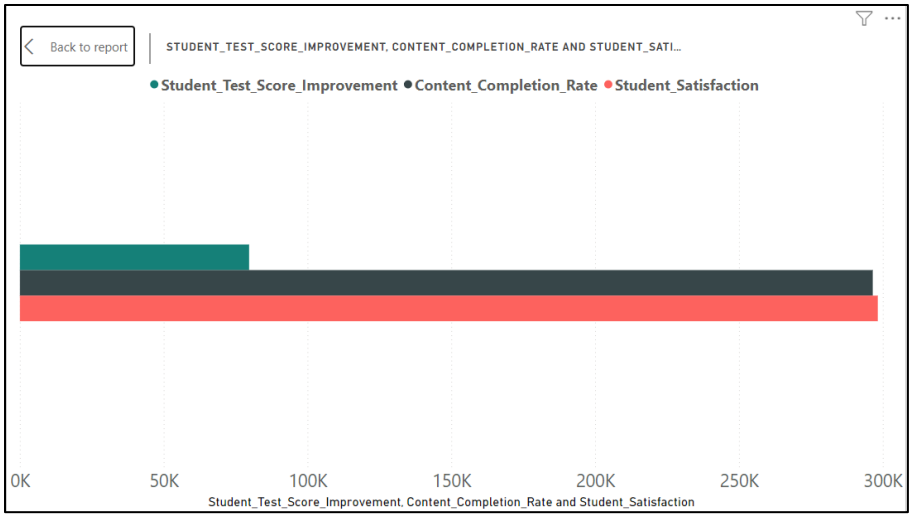


Fig 1.11 Rate of Student score, Content completion and Student satisfaction

From fig 1.11 it is identified that course completion and student satisfaction almost align with each other.

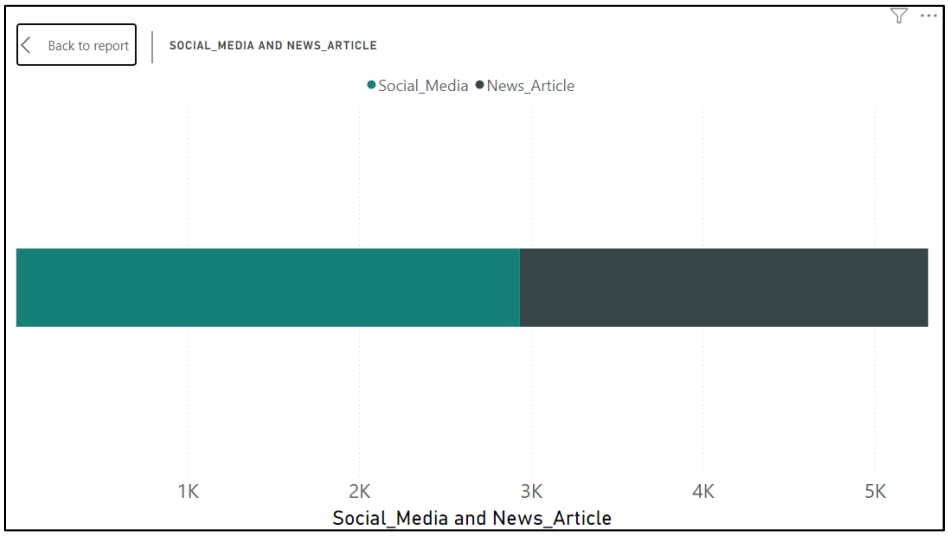


Fig 1.12 Marketing trend analysis

Based on the findings from Figure 1.12, which indicate positive reviews in both news articles and online articles, it is evident that the company is receiving favorable attention in the media. Consequently, it is inferred that the current marketing efforts are yielding positive results. Therefore, it is advisable to reallocate resources from marketing activities towards enhancing the quality of content and addressing user requirements. By prioritizing these aspects, the company can further improve user satisfaction and consolidate its positive reputation, leading to sustained growth and success in the long term.

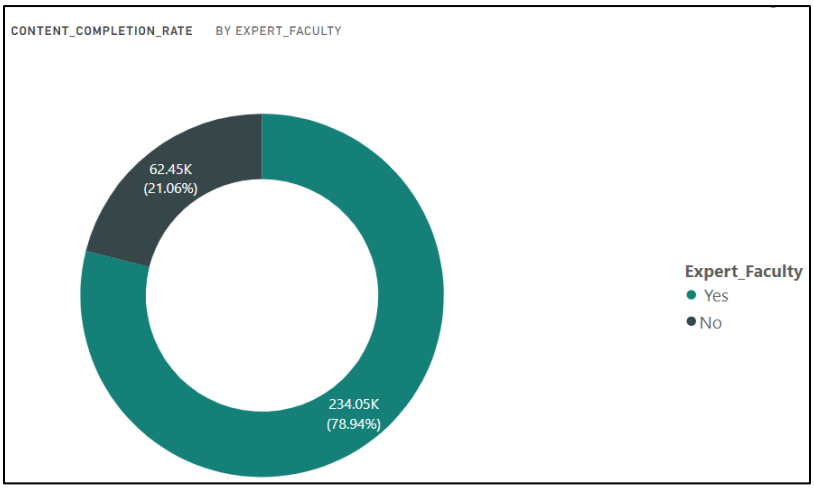


Fig 1.13 Content completion rate by Expert faculty.

The above fig 1.13 implies that the instruction of expert faculty helps in significant increase in content completion.

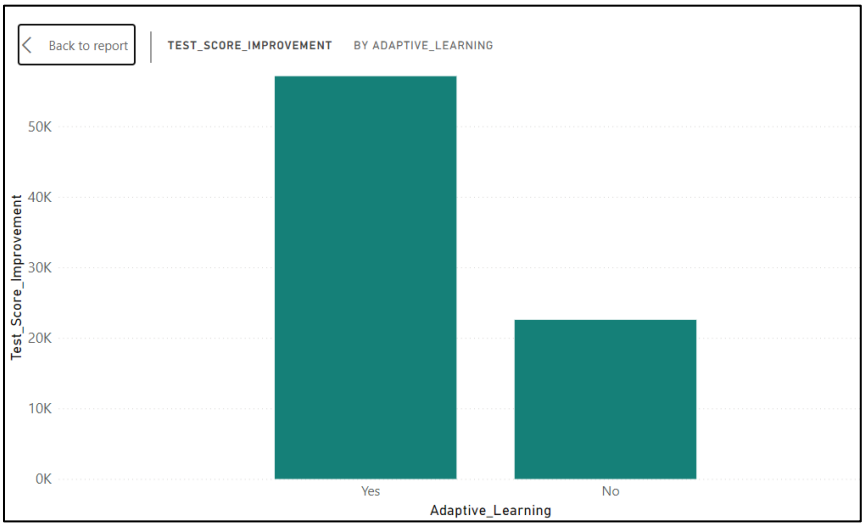


Fig 1.14 Test score by Adaptive Learning

The fig 1.14 indicates that adaptive learning strategy helps in improving test score secured by students.

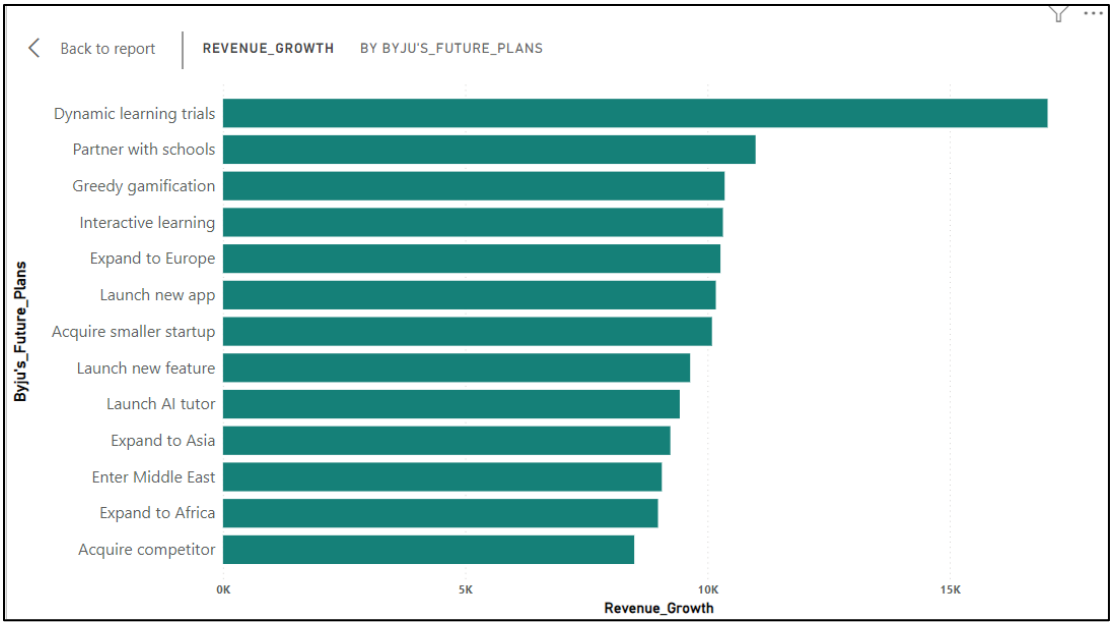


Fig 1.15 Revenue growth by Future plan

The above fig 1.15 indicates the top strategies that can be implemented in the future to improve the revenue of the company.

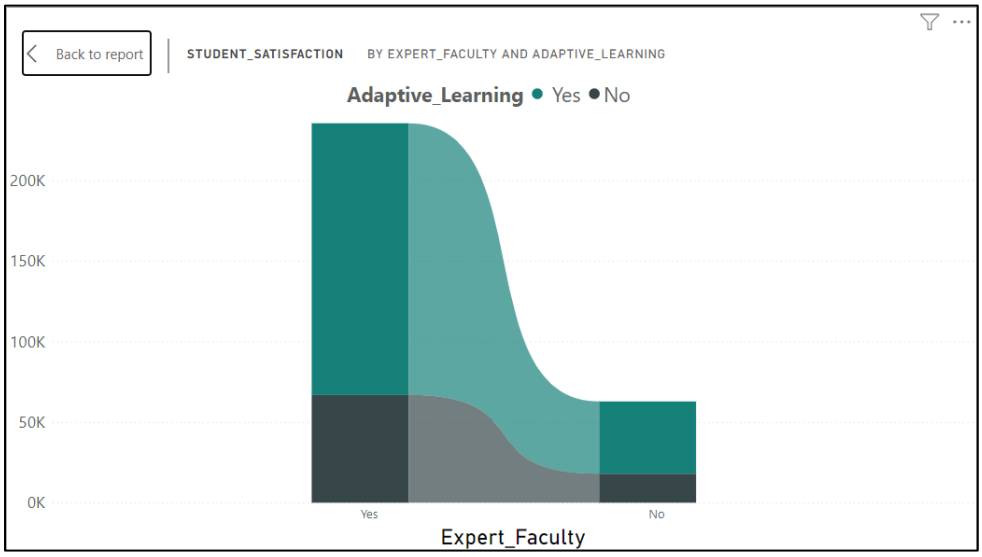


Fig 1.16 Student satisfaction by Expert faculty and Adaptive Learning

The above fig 1.16 tells how student satisfaction is affected by expert facility and adaptive learning. This plot tells us that adaptive learning with the help of expert faculty can improve student satisfaction.

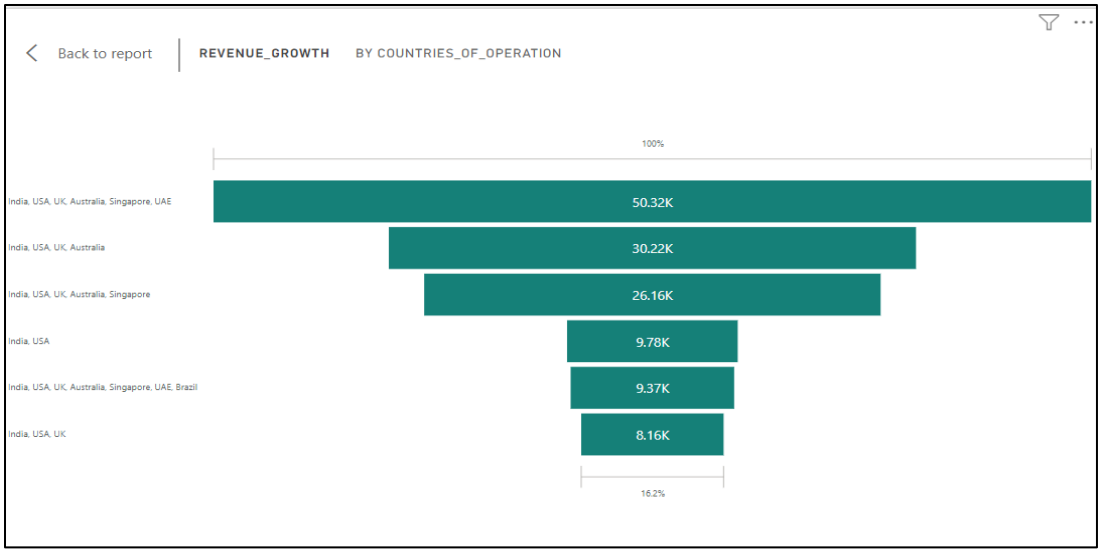


Fig 1.17 Revenue growth by Countries of operation

The above fig 1.17 shows the revenue generated across each country and tells us which countries to focus upon more.

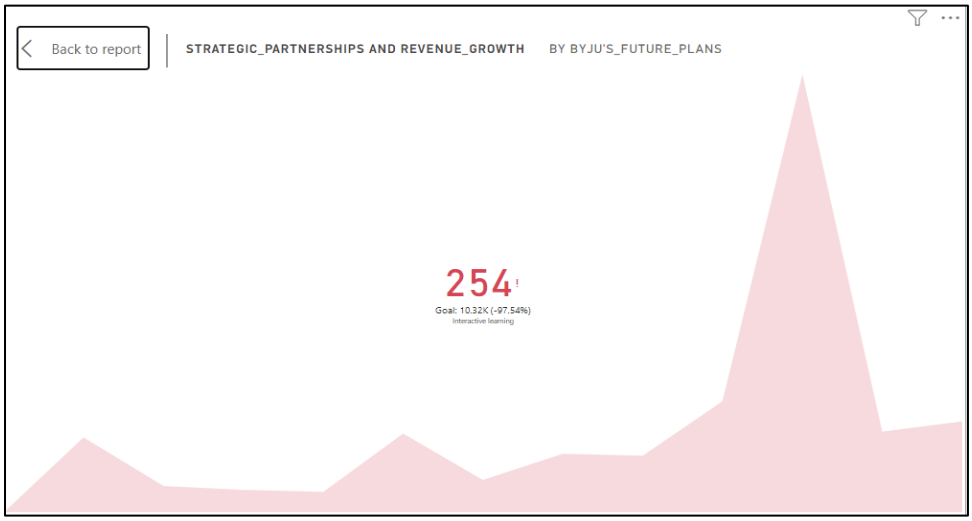


Fig 1.18 Revenue growth by Future plan

The provided graph illustrates the revenue growth achieved through the development of strategic partnerships and the implementation of future plans. While it indicates positive growth trends, it also highlights potential areas for further improvement. Despite the progress made, there remains untapped potential to enhance the organization's revenue further. Therefore, it is

imperative to continue refining strategic partnerships and executing future plans effectively to capitalize on additional revenue opportunities and drive sustained growth for the organization.

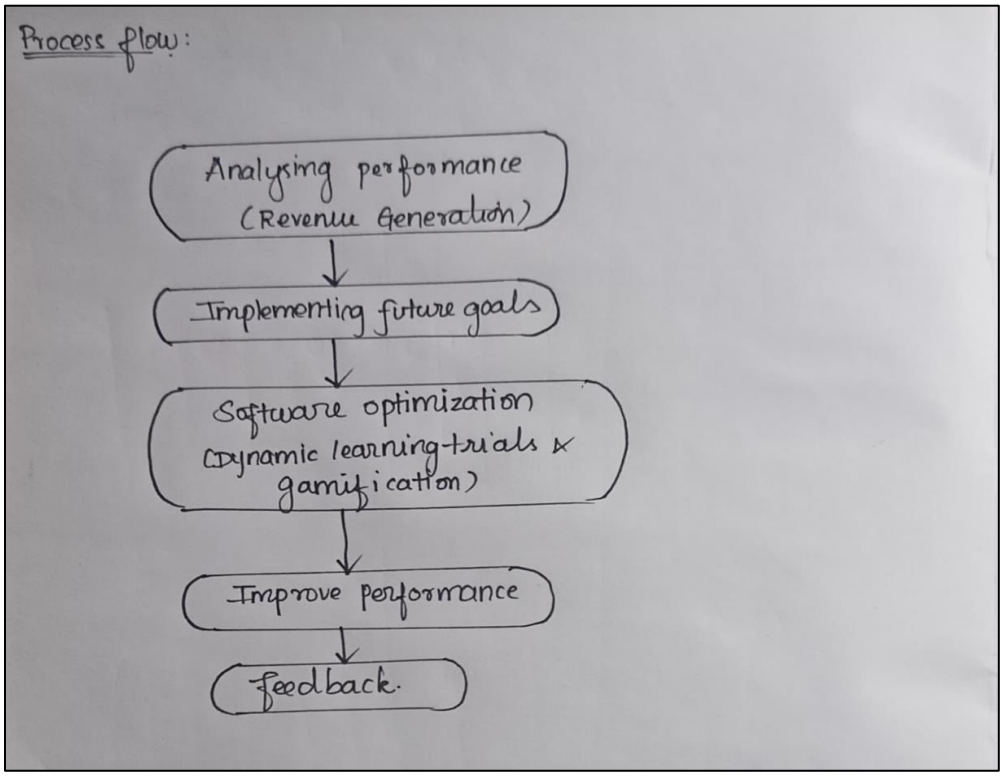
Overall inference:

From all these analyses, it is identified that the adaptive learning strategy aids in improving student satisfaction, consequently increasing content completion rates. Furthermore, expert faculty assistance significantly boosts student performance. Interactive learning, adaptive learning, and competitive learning are identified as the most promising implementations to rejuvenate the company, consequently enhancing revenue. Additionally, it is recommended to prioritize adaptive learning over marketing efforts for optimal results.

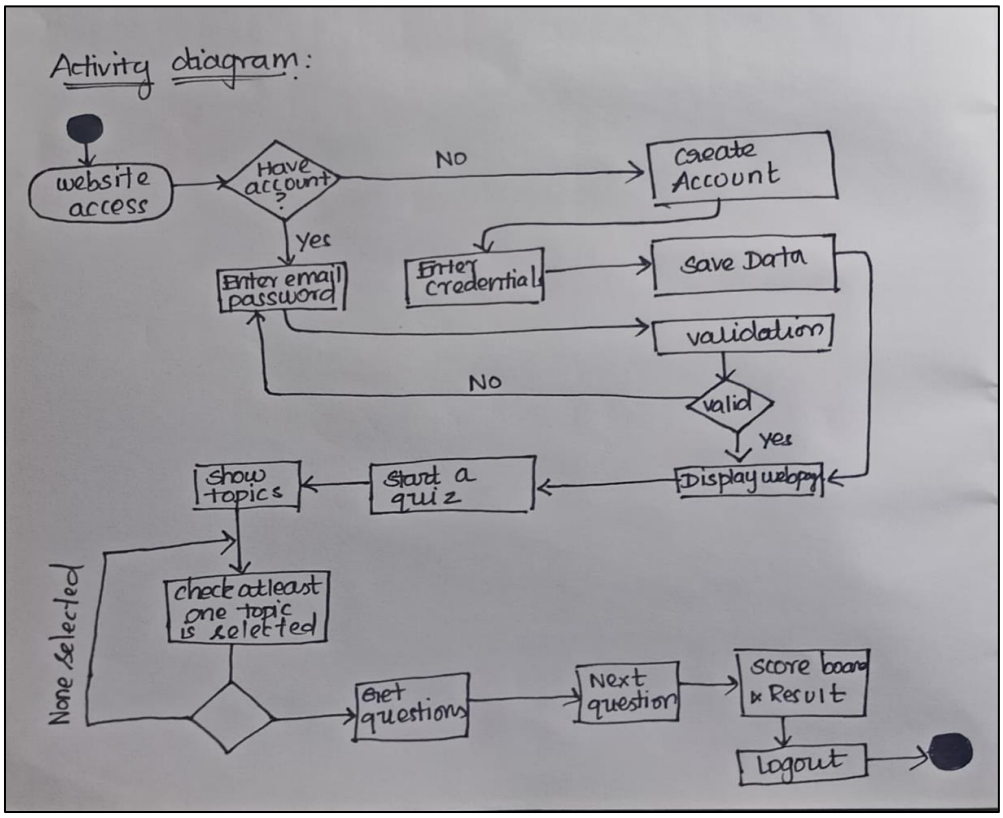
Technology Stack:

- Backend: Python, SQLite3
- Visualization: Power BI
- Frontend: Web-based platform for EduQuest

System Flow/Process Flow:



Activity Diagram:



Final Output:

The culmination is a revitalized EduQuest platform featuring dynamic learning trails, refined time management tools, gamification elements, and quiz functionality. Students embark on personalized learning journeys, optimized study schedules, heightened engagement, and targeted support for weaker areas.

Feedback obtained from users will be utilized to implement Agile methodologies aimed at enhancing user satisfaction. By adopting Agile practices, such as iterative development, continuous feedback loops, and collaborative problem-solving, the organization can quickly address user concerns and iterate on product improvements. This Agile approach allows for flexibility and responsiveness to evolving user needs, ultimately leading to higher levels of satisfaction and improved product quality. Through the implementation of Agile principles, the organization can achieve greater alignment with user expectations and drive ongoing enhancements to its offerings.

Challenges:

- Data Integrity: Ensuring the accuracy and completeness of historical data.
- Algorithmic Complexity: Designing and implementing dynamic programming and greedy algorithms.
- User Adoption: Encouraging students to embrace new features and gamified elements.
- Technical Integration: Integrating Power BI with the existing platform for seamless data visualization.
- Continuous Enhancement: Iteratively refining the system based on user feedback and performance metrics.

Conclusion:

The resurgence strategy for EduQuest epitomizes a holistic approach to rejuvenating the company's operations and elevating student outcomes. Through data-driven insights and innovative features, EduQuest aspires to reclaim its position as a leading provider of personalized and engaging learning experiences in the EdTech domain.