

Project Title: Retention Analysis: Predicting Student Retention Rate in Higher Education

Group Number: 04

Group Members:

Drusya Chandra

Ganesh Gundekarla

Domain:

Diving into our project, this idea revolves majorly across analyzing the dataset which provides us with a detailed view of the undergraduate students who are currently being enrolled in various branches at a higher education sector. This dataset will help us examine how many factors can play a key role in influencing the student retention rate and level of academic success.

When it comes to the type of domain we are dealing with, this project we are doing sits at the combination of Education Analytics and Data visualization. By carefully analyzing our dataset, we are aimed to find patters and insights that can help us come up with answers on why students tend to stay or leave from their academics and how much level of contribution they provide to their academic achievements.

Problem Statement:

Our primary goal by doing this project is to explore and understand the essential key factors that are influencing the undergraduate retention outcomes, this means understanding how some factors are making the students to “drop out ” or “graduate successfully” or “remain enrolled” .

This project makes the use of “data visualization “ to observe the trends and help the institutions to make actions which are based on the insights in order to improve the students outcomes.

Patterns we aim to understand by studying vast variables such as :

- a. Demographic info: which includes student’s age , gender and marital status .
- b. Socioeconomic factors: this includes scholarship status , application method.
- c. Performance in his academics: number of approved credits / enrolled units or grades etc.
- d. External key economic factors: Note that this factor can be a crucial step as the nation’s unemployment trend , inflation season or GDP rates can be a deciding step for the

student's plan to either plan or discontinue his/her studies , hence this factor is also being considered in our project.

Understanding this patterns between different variables can help us identify at risk students who tend to drop out early and this can help us come up with improving educational strategies to help retention of students and academic support for the effected students.

By using a interactive form of data visualization , we aim to improve the exploration of datasets by :

- a. Pattern discovery through filtering data which is done real time.
- b. Allowing educational institutions to engage with the data interactively .
- c. Allows "data driven decision making " in order to help students achieve their academic success.

Methodology :

- a. Data Collection

Source : Kaggle dataset : Higher Education Predictors of Student Retention

- b. Data prepping and cleaning the data :

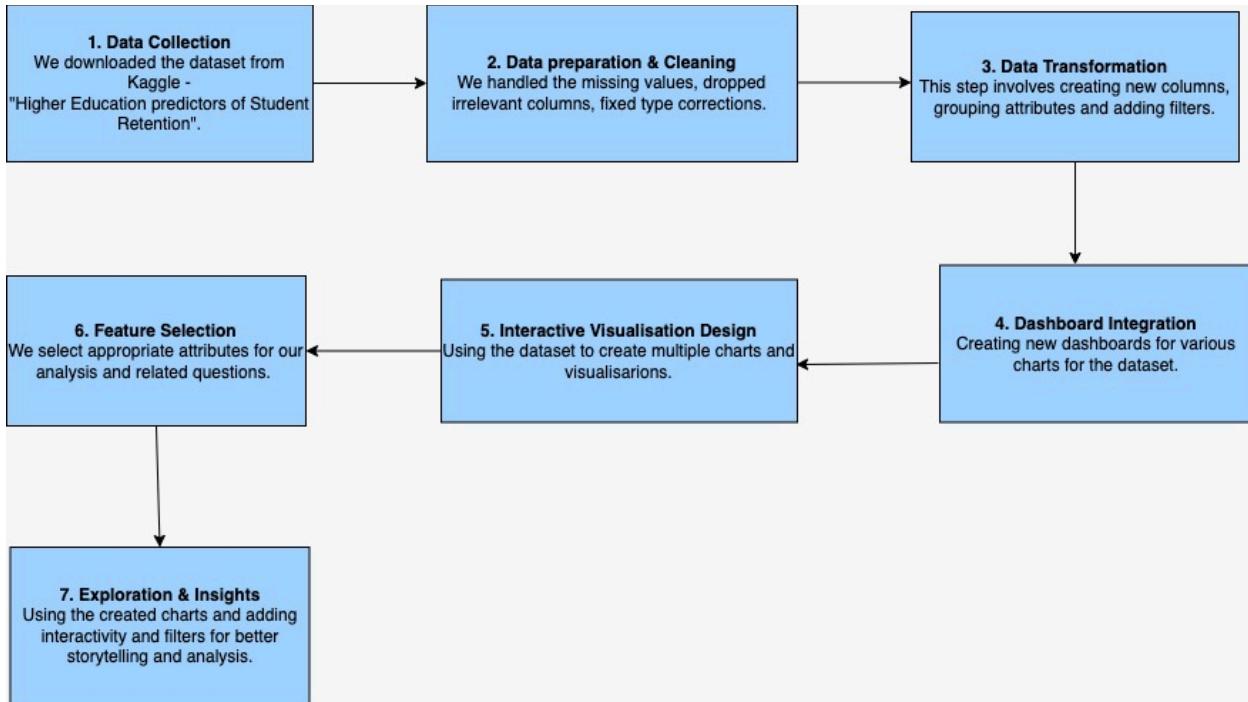
- 1. Handling missing values .
- 2. Data type corrections .
- 3. Removing irrelevant columns if necessary .

- c. Data transformation for creating success rate of the students.

- d. Selection of Features.

Chosen attributes include :

- 1. Demographics.
 - 2. Socioeconomics indicators.
 - 3. Academics
 - 4. Economics.
- e. Interactive Visualization design: Building multiple charts.
 - f. Dashboard integration
 - g. Exploring the data and providing insights.



Data Abstraction:

Type of dataset: the dataset we used is a structured one .

Attributes:

- a. Marital status: marital status of the student which is a categorical one .
- b. Application mode: mode of application used by the student.
- c. Application order in which the student has applied.
- d. Course: kind of course taken by the student.
- e. Attendance: attendance recorded by the student.
- f. Qualifications of the student: previous qualifications recorded by the student.
- g. Nationality: nationality of the student.
- h. Mothers and Father's qualifications and occupations.
- i. Tuition fees.
- j. Gender : gender of the student.
- k. Scholarship holder : whether the student is a scholarship holder or not.
- l. Age at enrollment : age at the time of enrollment.
- m. International enrollment : whether the student is an international student.
- n. Curricular units for 1'st sem :
 - a. Credits
 - b. Enrolled.

- c. Evaluations.
- d. Approved.

Data source : Kaggle

<https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention>

Data Transformations :

- 1. There was no need of any data cleaning of the dataset we have taken .
- 2. There are no missing values to remove those columns which would have become unnecessary.
- 3. There were no outliers too.

Task Abstraction

The main goal of our project is to not only identify the patterns or trends / relationships in the dataset we have chosen but also to help us explain the reasons and predict the student retention outcomes like

- a. Drop outs .
- b. Whether student is enrolled or not .
- c. Successful graduates.

This project aims to answer some of the key questions like :

- a. Are the students from a specific age group tend more likely to drop out ?
- b. Does academic / attendance (day or night time) / credits / approvals are more likely to correlate with graduation?
- c. Are the students who are awarded with scholarships tend to perform better?
- d. Also, any existence of trends which are by gender / application mode or parental education related factors?

These insights we aim to uncover will certainly help the institutions to build a better system for the success of the student.

Actions :

Methods we have performed accordingly:

- a. Filtering and slicing: we used power Bi capabilities to isolate groups of data.
- b. Aggregation: summarizing the values which includes like total dropout count for number of courses.
- c. Grouping: grouping of the students based on attributes like age range , gender.
- d. Correlation analysis: we performed visualizations which are based of relationsip between different variables.
- e. Clustering of the data using visual approach: we used maps and stacked bar charts to see the clusters and outcomes of a specific data points.
- f. Line trends analysis and charts.

Implementation using Tools:

For this project we primarily used Microsoft Power BI as our data exploration and as a core visualization tool .

Implementation steps:

We performed

- a. Data importing and transformation.
- b. Interactive visualization.
- c. Real time filtering.
- d. Dashboard creation.

With Power BI :

- a. Importing and cleaning the csv dataset.
- b. Transformations of the columns for better mapping.
- c. Created pie charts, bar charts, line graphs and stacked columns, tree maps etch
- d. Used filters and slicers for more better user driven exploration.
- e. Interactive dashboards for better understanding and perspectives and many more.

With Power BI visualizations, it was very precise on how we gather these so many insights and explore many hypotheses, by using a more interactive dashboards, we could see how trends were depending on several key factors.

Insights we gained:

- a. Dropout vs graduate distribution using pie chart.
- b. Impact of the scholarship on the retention rate.
- c. Academic performance correlation.
- d. Attendance and age of enrollment.

- e. Course selection patterns were discovered.
- f. Attendance patterns on daytime vs nighttime.
- g. Nationality and international students.

Results and Analysis:

Screenshots of visualizations and what insights our visualizations provide us with :

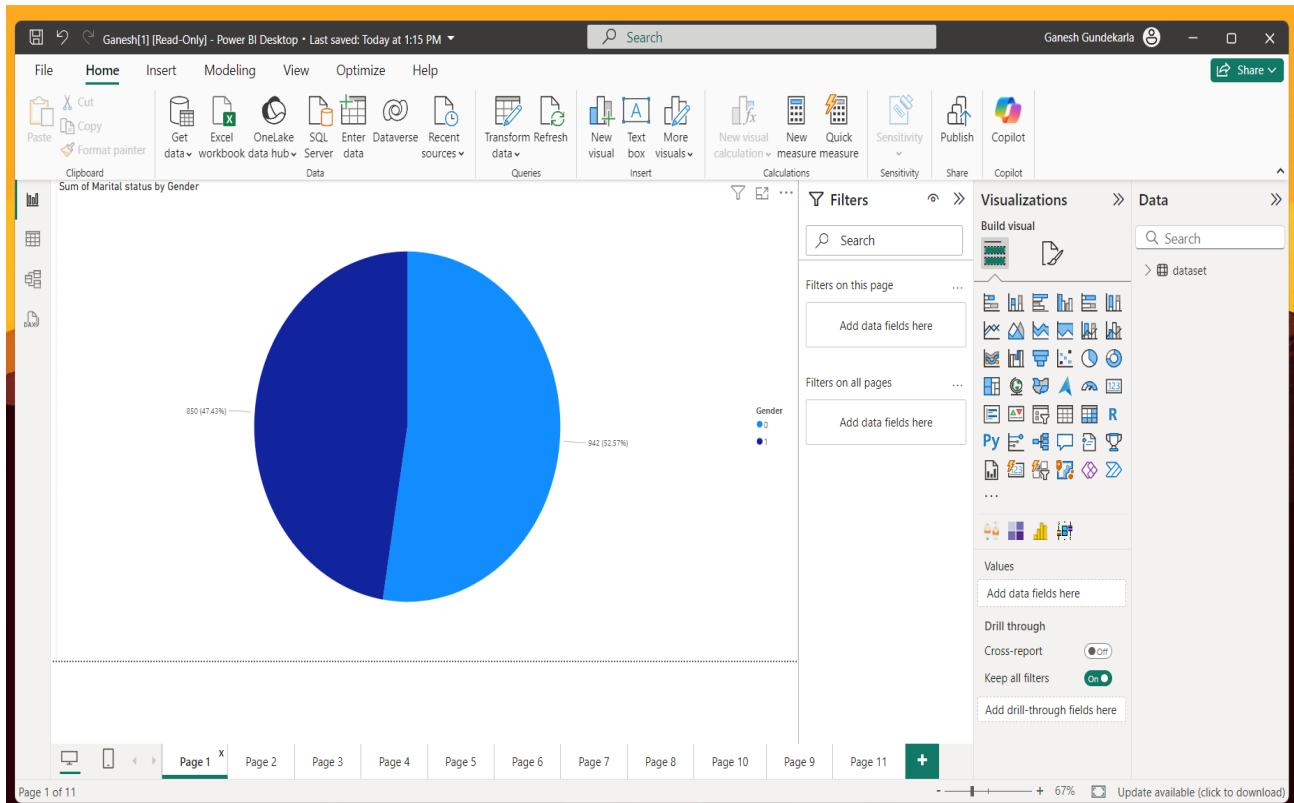
1st visualization:

Visualization insights:

This pie chart will provide us with composition of the marital status by gender , and this aims to highlight a good representation of distribution of the genders in the dataset . One gender accounts for a amount of 942 records with a overall percentage of 52.56 percent , whereas the other constitute to a percent of 47.44 percent of 850 records . This is well balanced and nothing but a slight variation of 5 percent. This kind of close proportion tells us that the dataset we have used is a demographically well balanced in terms of gender .

What could this mean ?

“Unbiased Analysis of marital status” trends could be observed , this visualization could effectively convey the split of the gender and this could mean that this ensures fair representation which is very crucial when we are considering to analyze relationships / patterns in the whole marital status distributed across different demographics in the dataset.



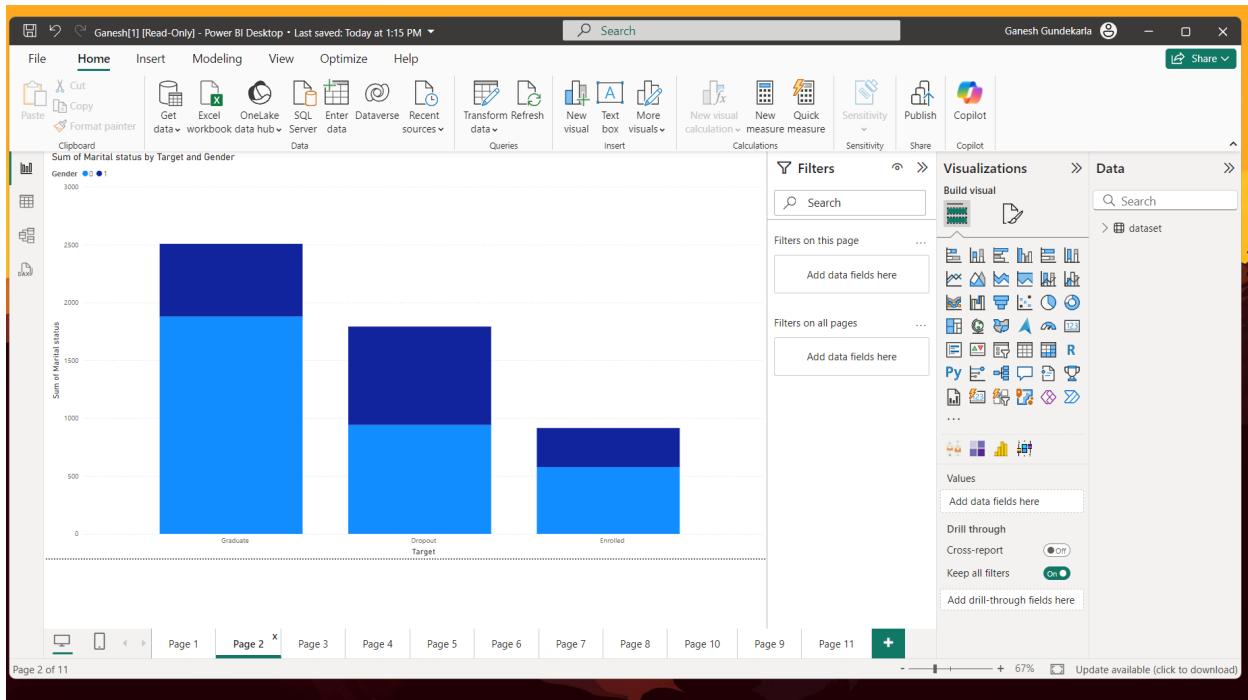
2nd visualization and it's insights :

Bar chart :

This bar chart represents the whole sum of marital status which is well segmented by the outcomes like (graduates, dropouts or the enrolled ones) and this one's is also divided by the gender. When we look at the graphs, graduates most like to have the highest marital status value across both genders. This indicates us that a **"strong correlation between the graduation and the marital status "**. Here, dropouts basically are showing a much moderate count, whereas the enrolled ones (current) have the lowest figures. When considered across all the categories where we can see, we observe that we provided with both genders, we could observe that one gender is consistently has a higher value (especially if we look it at graduates and dropouts).

Key point:

This graph shows us that the marital status might be a key factor in influencing / or tied to the academic outcomes which makes it as a important relevant demographic factor in the analysis for educational institutions.

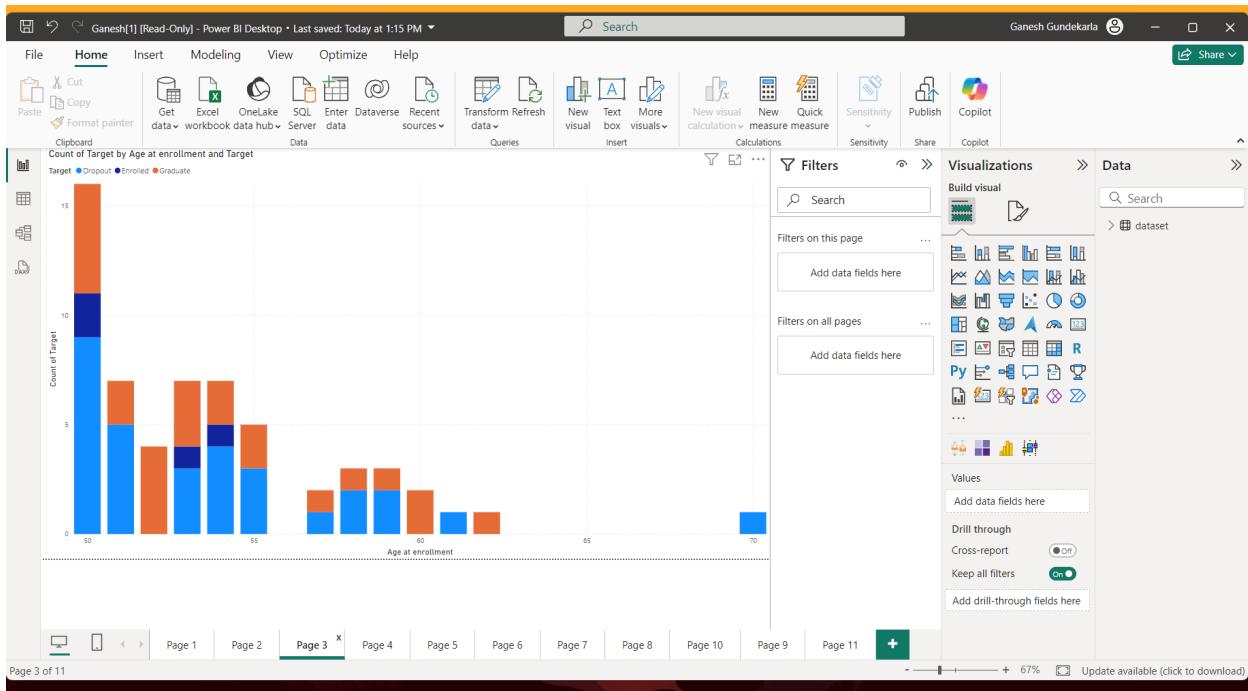


3rd Visualization : Bar chart:

This bar chart clearly shows us with the count of the target outcomes of the three categories (dropout students, enrolled students , graduate students) while the age of the enrollment . This will reveal us some good key trends in the educational attainment varied across different age groups . The highest concentration of students (particularly observed across dropouts and graduates) is mostly observed in the age groups of 48-52 range . **This indicates that this age group is much more active in the education .**

Key insights :

- Enrollment and graduation rates are declining as the age increases . While the dropout rates are much more sustaining / prevalent in the young age groups.
- Mid aged students seem to be much more engaged and to be more successful in terms of education , while the old groups are showing us with some reduced participation and reduced graduation rate.



4TH Visualization insights:

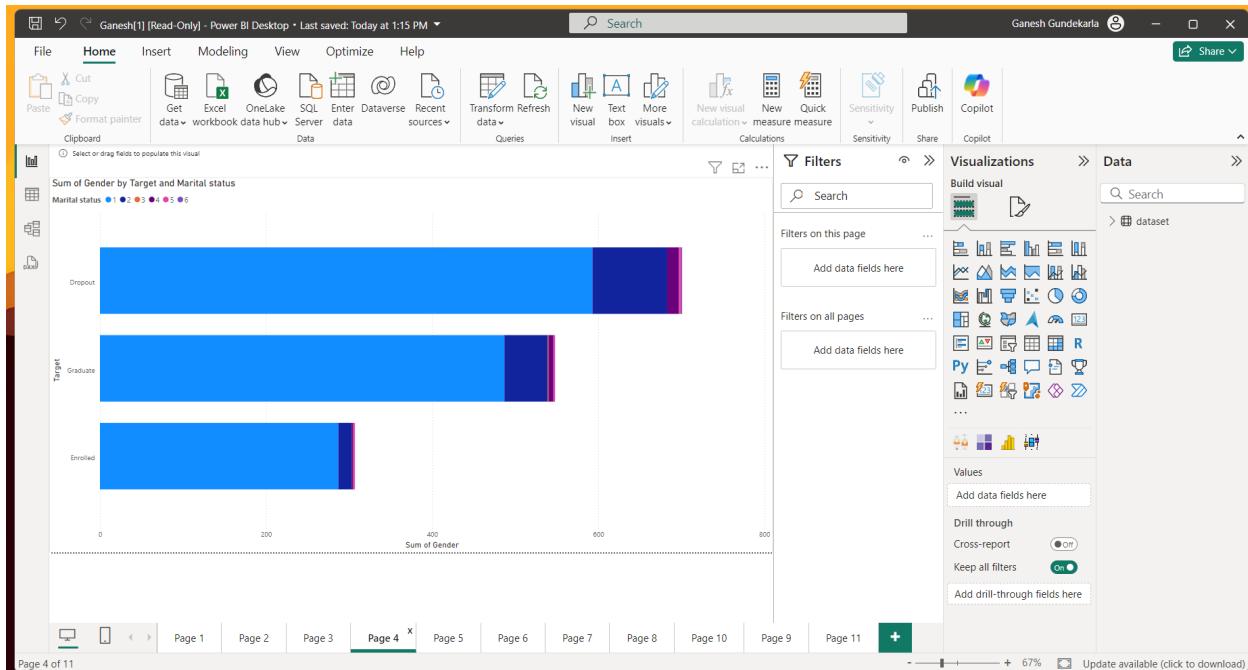
Sum of genders across a different kind of target outcomes – Dropouts , Graduates , Enrolled ones which is segmented by marital status.

The target status is broken into different marital groups where the bar chart combines the target status , gender and the marital status.

As we can see in the below graph, dropouts basically have the highest gender count which tells us that it is the overall higher representation in this category which is later followed by the graduates. Most individuals when regardless of the marital status are more likely to fall into the dropout ones.

Key points :

The blue color dominates suggesting that a particular gender and a marital status combination is significantly more prevalent. This highlights a much potential pattern when we could conclude saying that the marital status could be influencing the educational results , especially if we look into the dropout rates .



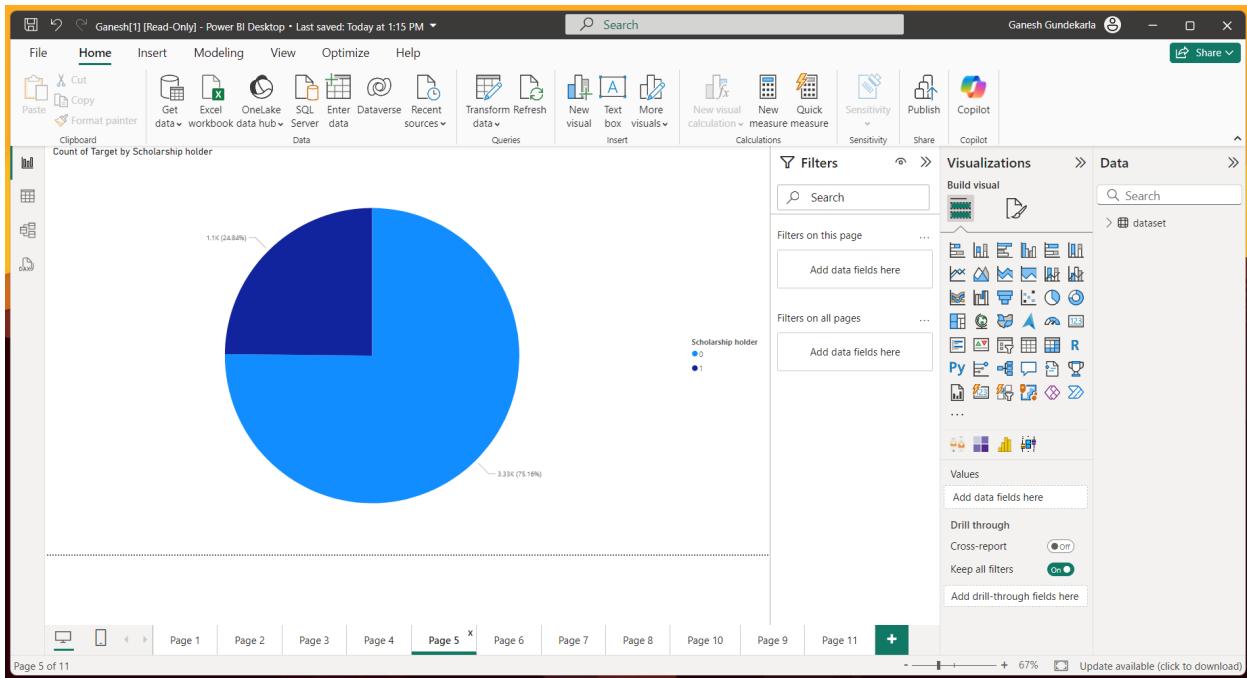
5th visualization insights:

Pie chart visualization of count of academic targets which is differentiated by scholarship holders

This pie chart reveals us that the most number of students / individuals (73.18%) are not holders of any scholarships, and only 26.2% are receiving the scholarships. This tells us that most the students in the educational institution have done their studies without any aid.

Key insights:

This tells us that not having some of the financial aid can be one of the important reasons for a student to be dropping out no matter the amount of target status. This tells us that the educational institutions should consider this for policy improvement, where improved scholarships could account for more academic success. Understanding how scholarships would impact graduation / dropout rates can help educational institutions design a much more better support systems for all students boosting overall educational success.



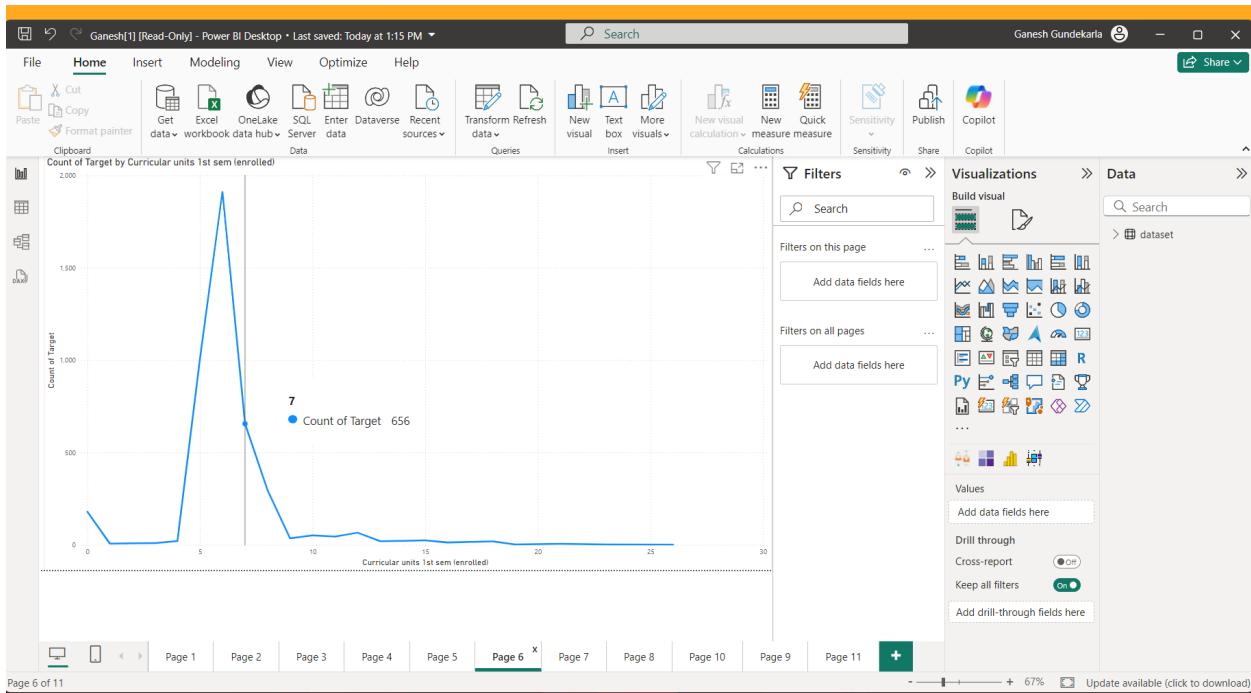
6th graph insights:

Line chart displaying count of academic targets based on number of curricular units taken in first semester:

Here, we could say, a clear peak is observed at 7-8 units with nearly 2000 students enrolled, it indicates us with that it has to be the most common course load. This line chart showed us with a steep rise and a gradual decline is observed afterwards, which tells us that students prefer to enroll with 8-9 units for their academic balance. This pattern presents us with a common standard enrollment behavior.

Insights to be gained:

We can see the overloading with course units can correlate with the dropouts , and underloading will be not satisfying the course graduation requirements , therefore institution should consider this when assigning the minimum credit hours.



7th graph visualization:

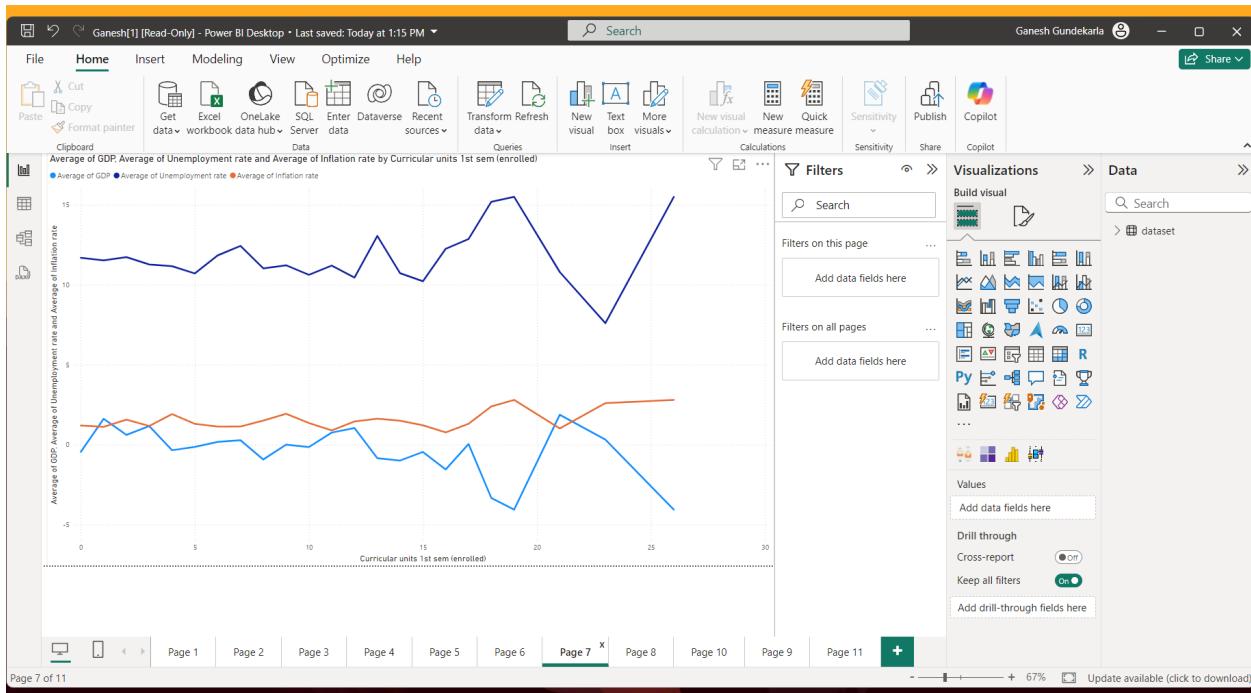
This graph displays the relationship between the number of curricular units enrolled in the first semester and in consideration with the economic indicators : GDP , unemployment rate and inflation .

As the curricular units increase , the average GDP tends to rise and which tends to say that there could be a potential link between the higher academic engagement and with a much stronger academic performance.

The unemployment line shows slight fluctuations in middle but it will increase with a much enrollment , showing us that better employment comes with more academic commitment .

The inflation rate line shows inconsistent trends (happens when there is a weaker correlation) .

Overall , higher academic workload is much more positively tied up with economic stability and growth.



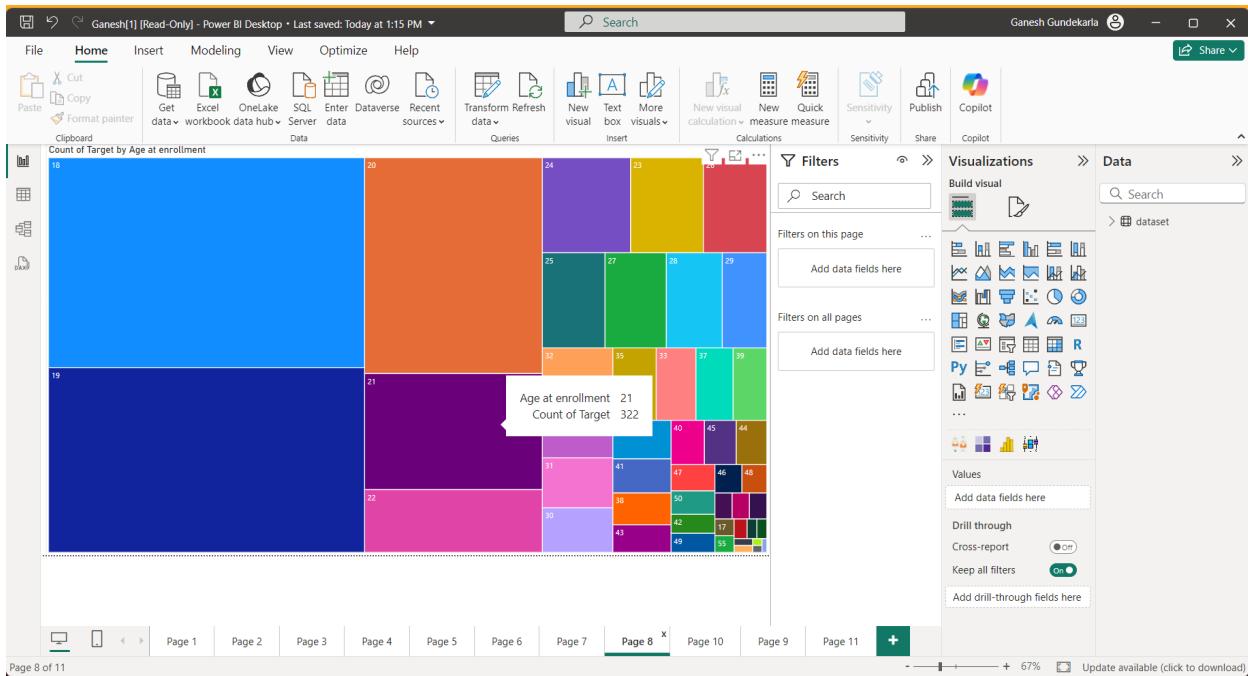
8th graph visualization:

Tree map visualizing the count of academic targets differentiated by age of enrollment:

This graph highlights most students who enroll between the ages 18-22, we can observe that the age 18 must be the highest representation. Core academic age group is in the ages 19,20,21. This tells us that most of the students immediately pursue higher education right after the secondary schooling. After the age 25, the enrollment suddenly drops with minimum over the age 35 .

Insights:

This visualization shows us that the younger age groups are most likely to pursue academic targets , this emphasizes the importance of early academic planning and need of adult education support in the older age groups.



9th visualization:

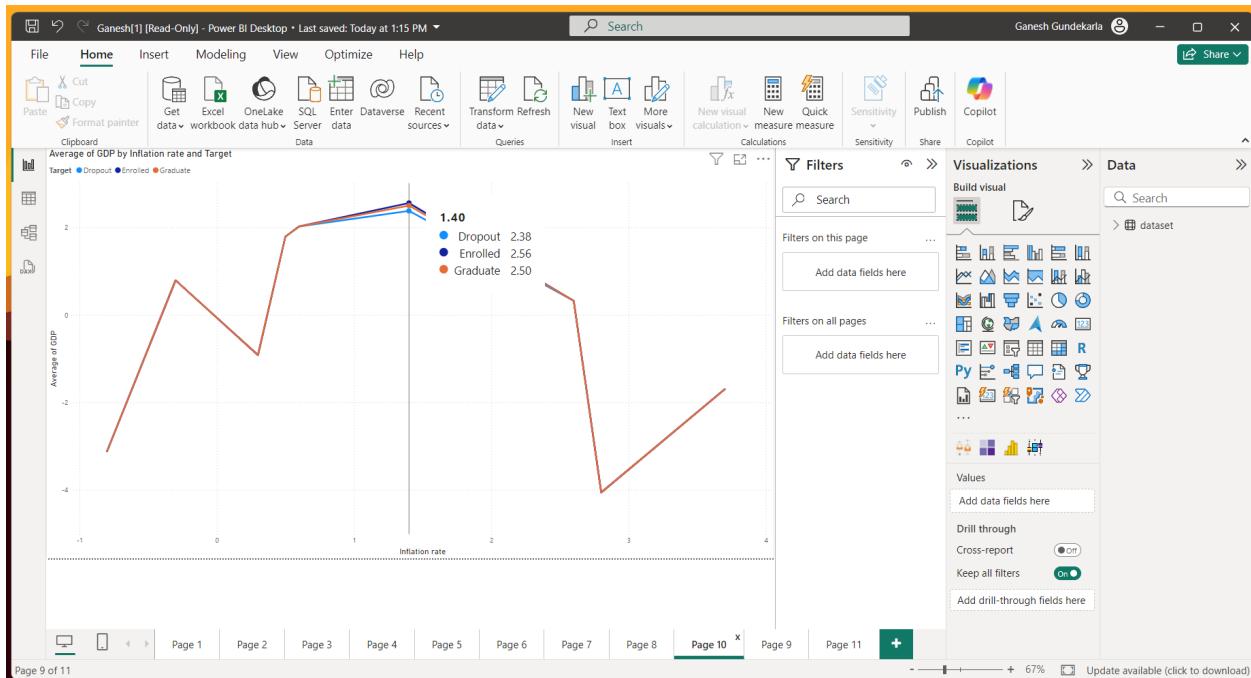
Line chart illustrating the relationship between inflation rate and average GDP across different academic groups – Dropout, Enrolled , and Graduate Students.

All three groups in the chart show a good positive GDP trend when the inflation rate is around 1.4 , with the current enrolled students who are reflecting the highest trend for the GDP average which is closely followed by the graduates.

Insights :

This trend suggests us that there is higher academic engagement particularly amongst the enrolled and graduated students which then correlates with a better economic performance under the inflation.

This visualization tells us that the education level might play a good amount of role in economic resilience especially when the inflation is mid , making the academic achievement a crucial contributor to the GDP growth .



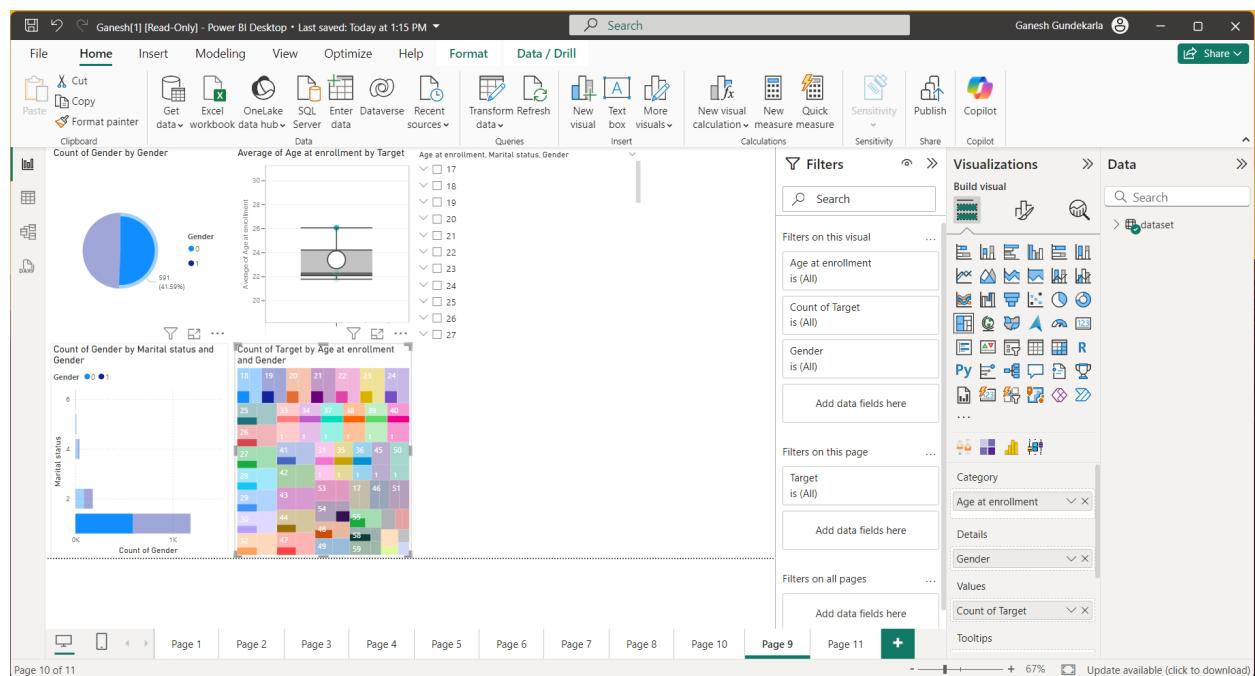
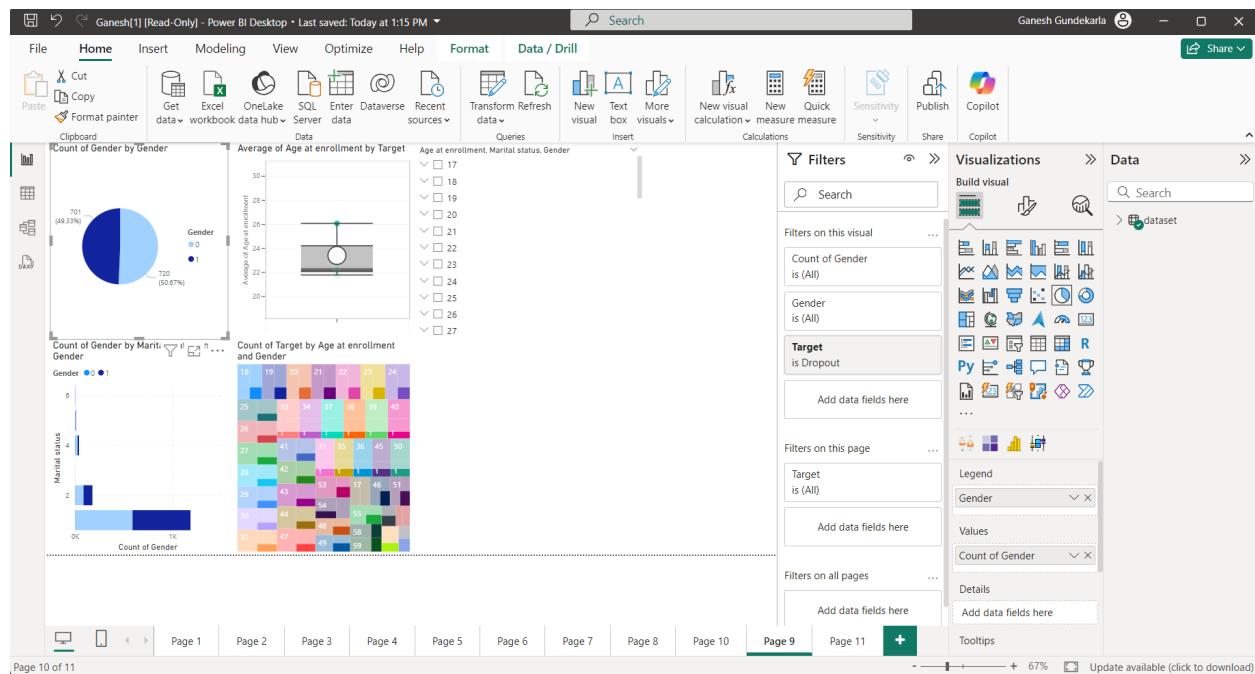
10th visualization :

Multiple visualization focusing on dropout category further segmented by gender , age of enrollment and the marital status :

- We can see that the gender distribution is even with males and females sharing nearly the same percentage which suggests a unbiased and a much balanced representation.
- The box plot here shows us the dropouts who are enrolled between the ages 20-24 .
- The heatmap highlights that the most dropouts occur amount students with clusters across the genders.
- The bar chart here reveals how the marital status 1c tends to be the most common amongst the dropouts , especially males .
- This overall suggests that most young and unmarried individuals are most likely to discontinue their studies.

Conclusion :

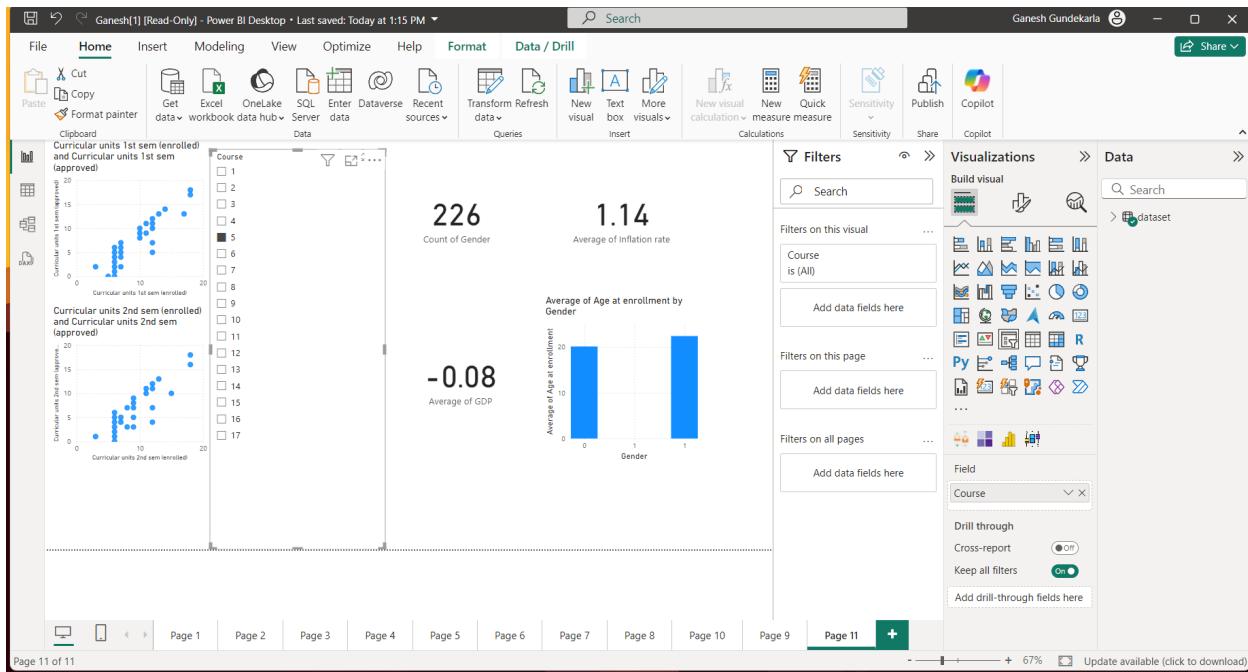
Here, most of the students enrolled are younger and unmarried and gender mix is even and balanced. Age and marital status might not be as much of a direct impact on output , but they affect it significantly .

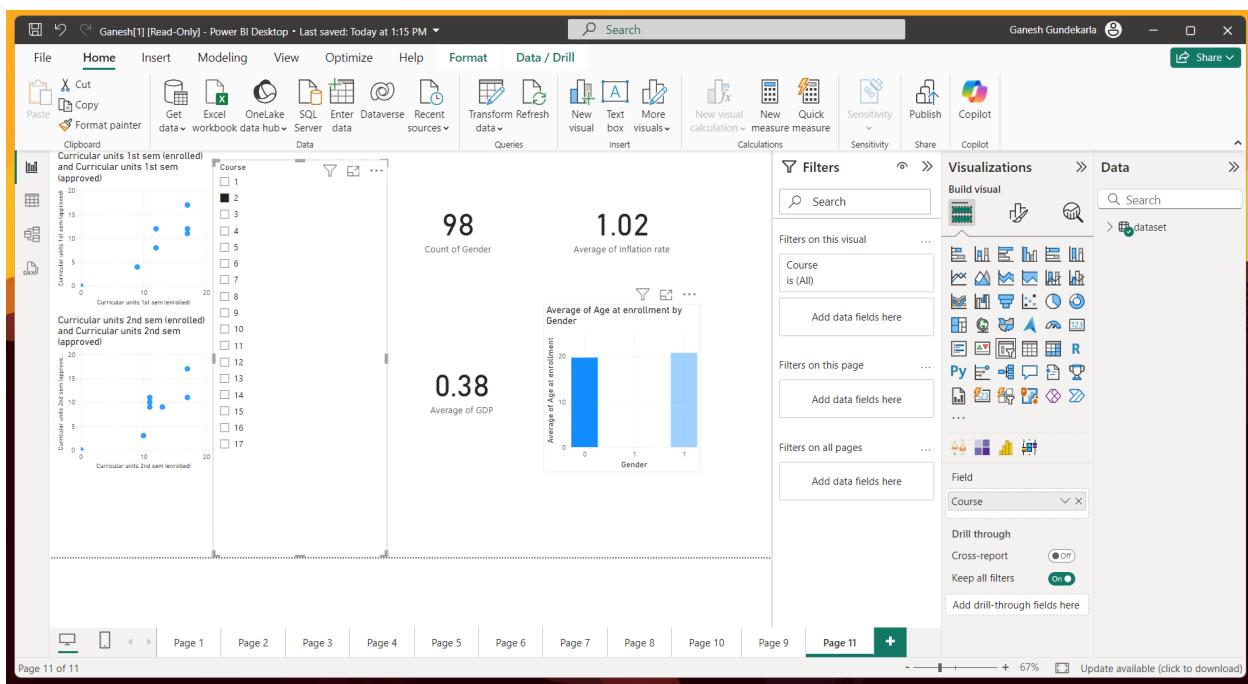
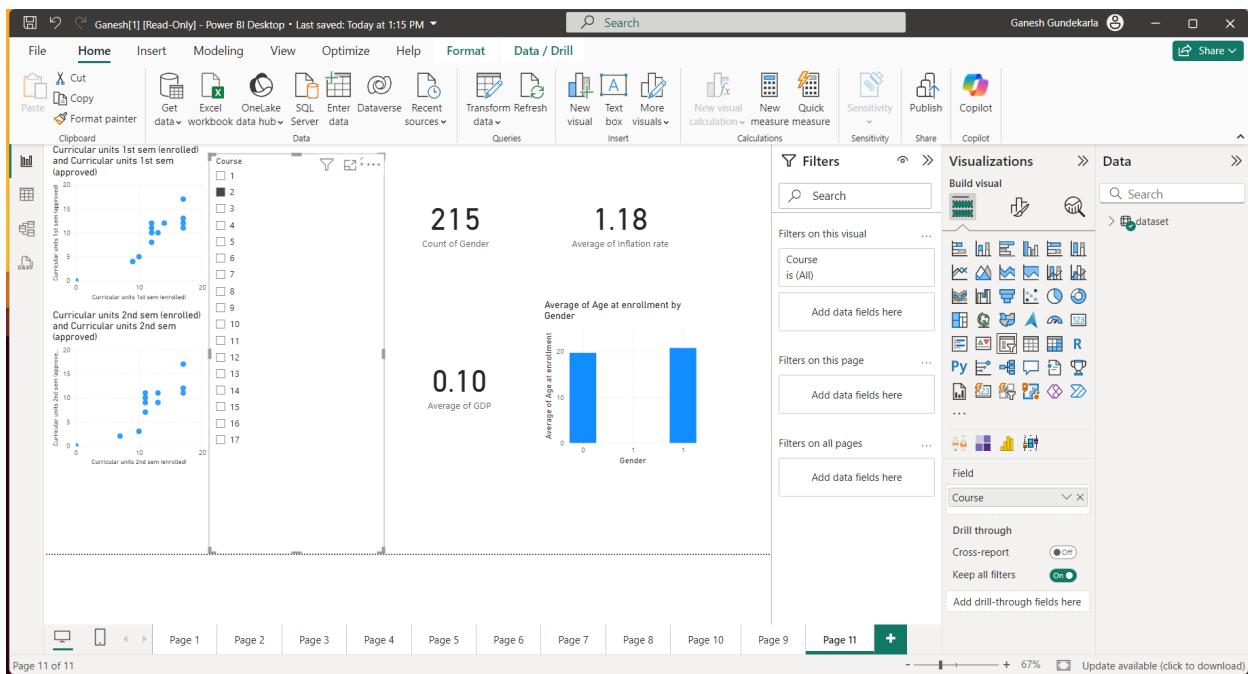


11th visualization.

Academic performance across semesters along with demographic and economic indicators taken into account:

The scatter plot here shows us with a good , strong positive correlation between units enrolled and approved in first and second semesters , indicating success. Also we can notice how the average age for enrollment is slightly high and inflation rate is 1.14 and also the GDP value suggesting us with economic instability. These graphs show us how these all the factors interplay for the outcomes. The goal of this dashboard is to provide a wide range of interactivity and see how different factors effect our outcomes .





Story telling approach :

Stage 1 :

- The data was structured and clearly labelled columns which covers everything .

b. It was a large one with more categorical columns.

Key variables that stood out :

- a. Target outcomes.
- b. Scholarship holder status.
- c. Curricular metrics.
- d. Age and attendance.
- e. Gender.

Stage 2 : Correlation Analysis and Transformation of the data :

- a. Relationship between variables :
 - a. A positive correlation is found between scholarship status and the graduation students.
 - b. A negative correlation occurred between curricular approvals and the retention.
 - c. Courses with high dropouts had low avg performance across the board.
- b. Patterns :
 - 1. Students struggling to perform in semester 1 had more dropout rates.
 - 2. International students showed a bit higher dropout rates.
 - 3. Younger age students showed most success.

Work Management:

Responsibilities and contributions :

- a. Drusya : 50%
 - a. Selection and exploration of dataset .
 - b. Data cleaning and transformation .
 - c. Designed visualizations for academic performance , patterns and scholarship analysis.
 - d. Created the whole methodology for our project.
 - e. Undertook whole documentation according to guidelines.
 - f. Took lead on what visuals to include in our documentation.
- b. Ganesh : 50%
 - a. Focused on dashboard design , organizing visuals like demographics ,academic outcomes and other factors.
 - b. Handled all the correlation analysis and the filtering analysis which led to identifying trends .
 - c. Creation of visuals based to gender , age and nationality insights.
 - d. Contributed to story telling approach and insights.

- e. Designed some final slides.

Next stages :

- a. Creation of interactive visualizations using D3.js or others.
- b. Using more datasets.
- c. Webpage integration.
- d. Try to refine our methodology a bit more.

Project Phase 2 Documentation:

The aim and objectives of our project remains the same. In this phase 2 of our project we pushed our story telling approach to a new scale by introducing different interactive visualizations using

Libraries and visuals :

- a. **Plotly library (wrapper class in python) : 9 Visualizations**
- b. **Altair library (python) : 8 Visualizations**
- c. **D3.js (extensive library in Js designed to provide users with interactive visualizations in webpages. 8 visualizations: 8 visualizations**

Goals on Visualizing the charts:

- a. We aim to plot multiple plots and ideas in different charts to get a deeper insight in our projects.
- b. We aim to explore from a basic bar chart to in deep with each factor levelling down in all of our visualizations.

Plotly visualizations :

Why plotly ?

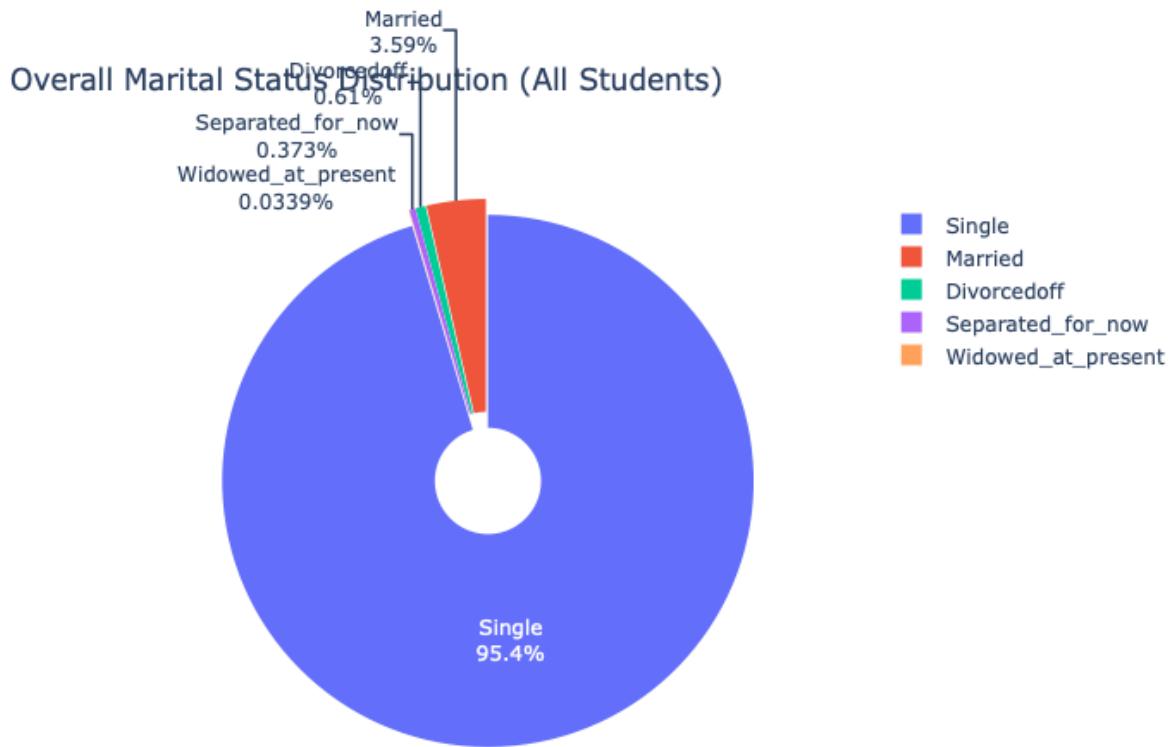
- a. Plotly is a graph library in python which enabled us to create interactive charts that can provide a comprehensive understanding of our data.
- b. This is an open source library rendered using Google Colab.
- c. With less code we can make an interactive charts than can help understand our data even better.

Visualization 1 : Donut Pie Chart.

Visualizing overall marital status distribution happening across students:

Purpose:

This chart here helps us to understand the overall distribution of marital status that's present in the dataset . We think that it is crucial to understand the demographic factors that can unlock different perspectives and patterns across the overall academic progress among students.



Storytelling and insights:

Upon looking at the chart we can find that most of the students belong to the 'single' category , only tiny margins account for married, divorced or separated etc. This helps academic institutions to focus majorly on the young students who are unmarried. This chart also highlights the cultural contexts of our dataset which tells us how diverse the students are . This helps us to identify and tailor the academic resources and support systems for all kind of groups .

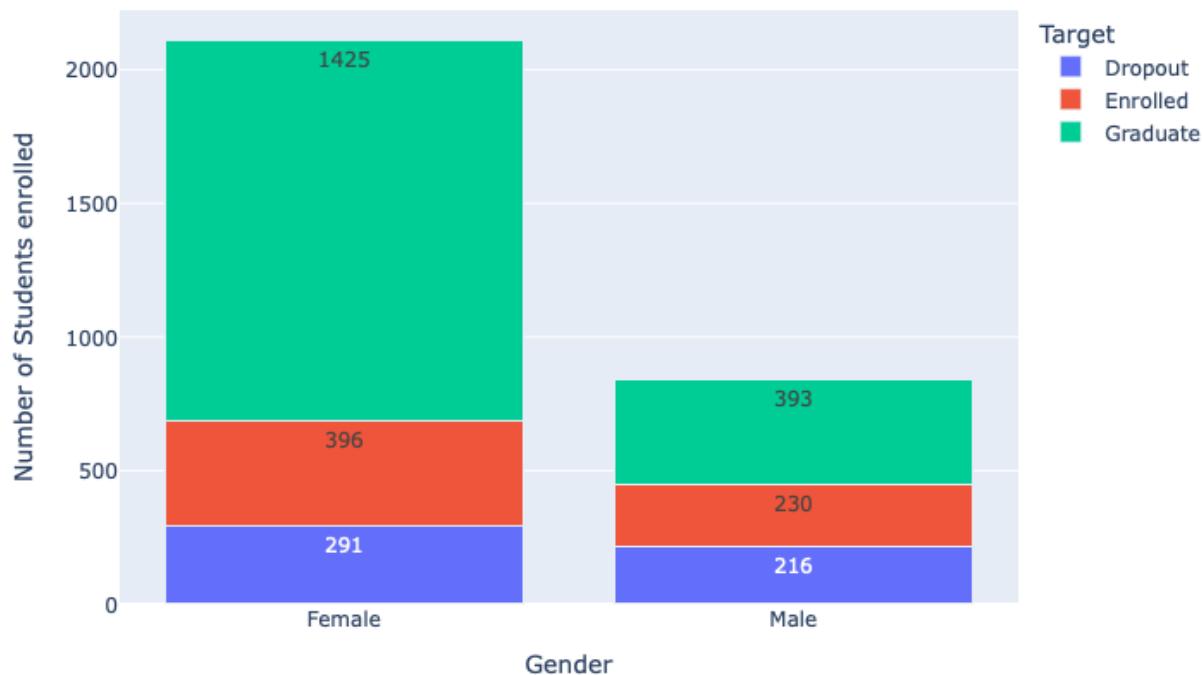
Visualization 2 : Stacked bar chart :

Title: Retention rate of the students targeting the gender

Purpose :

This chart is built to have a outcome based on retention (grad , enrolled or dropout students) which is compared across two genders(male or female) . This helps us to identify how gender can correlate with the academic retention rate and successful completion .

RETENTION RATES OF STUDENTS ACCORDING TO GENDER



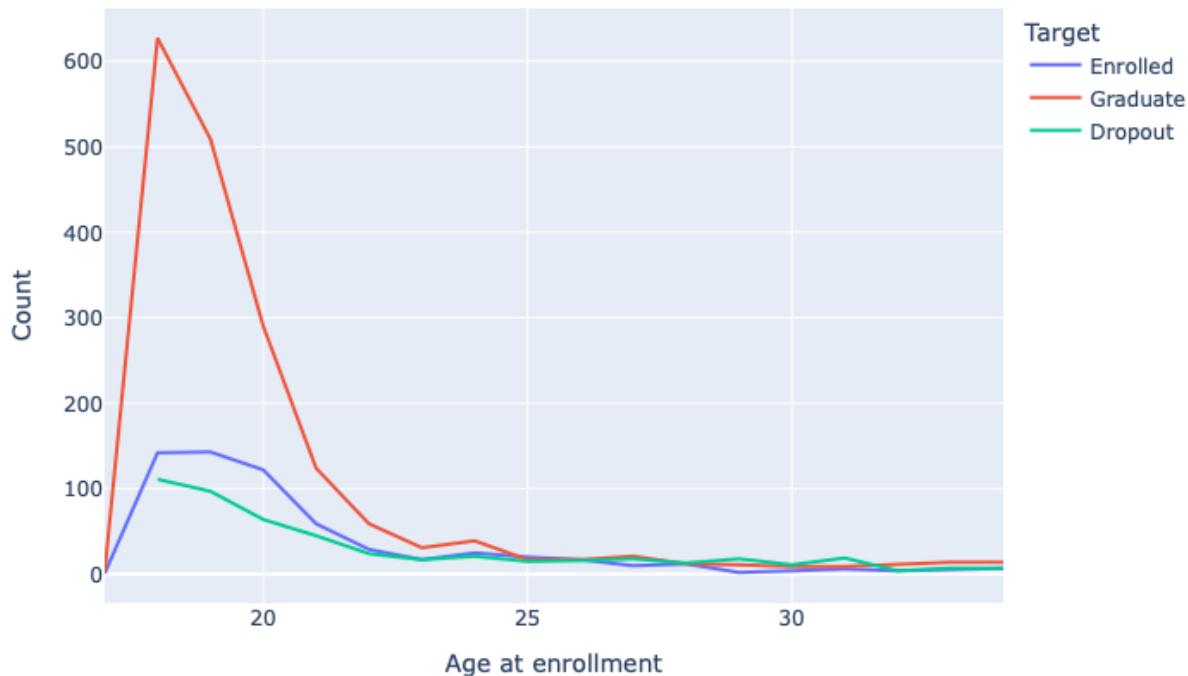
Insights:

- The stacked bar above gives us with a important perspective on how students with different gender gives out results.
- We can tell that the female students show us a very high graduation count when compared against with males.
- Male students are visibly lower in all the categories of retention.
- This chart reveals us with the disparities that exist among the outcomes of the students.
- This can help universities identify risk groups and help them re-evaluate their support systems
- This helps universities have more discussions on student engagement and prioritizing mental health.
- Stacked bar chart is chosen because it shows distribution happened across each category.

Visualization 3 : Line Chart

Title: Chart Visualizing Student Enrollment versus the Success in Retention Rate Outcome.

STUDENTS ENROLLMENT VS SUCCESS RETENTION RATE OUTCOME



Purpose:

This line chart is visualized to help us explore how age at enrollment can possibly correlate with the retention outcomes like whether or not the student graduates, drops out or keeps enrolled.

Questions this chart helps us to answer ?

- Which age groups of students are most likely to succeed?
- Which age groups most likely to graduate?
- Is there any groups of age which has most dropout rate ?

Insights:

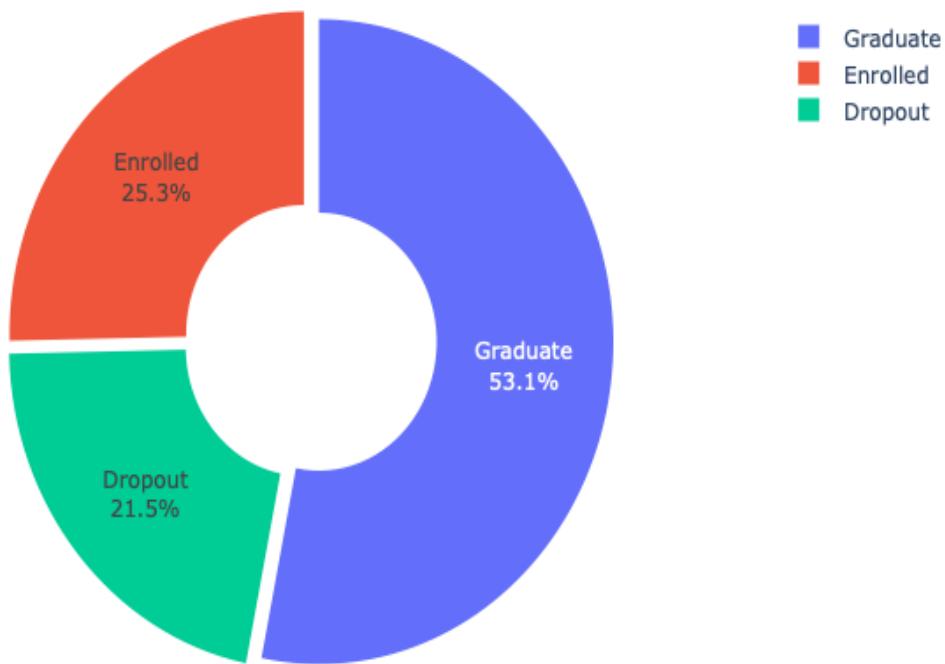
- This chart clearly shows us that most of the students that graduate are from younger age groups (17-19) .This age groups are most likely to show to long term commitment.
- The most amount of dropout rates are also at the younger groups but not that high showing us that the dropouts occur at younger age correlated with many factors like pressures, financial reasons etc.

- c. Older students tend to enroll in fewer quantities or more prepared when they enroll , and their outcomes and dropouts are balance out as well.
- d. Line chart can be a great way to compare these kinds of trends.

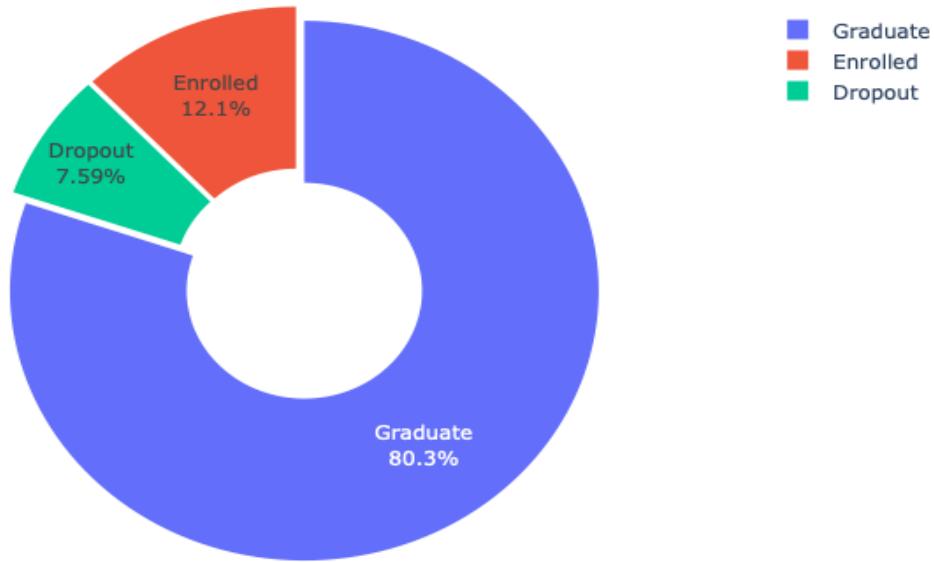
Visualization 4: Donut chart

Title: Retention outcomes of Scholarship issued vs Non-Scholarship issued students

Student Outcome by Scholarship status (receives): No



Student Outcome by Scholarship status (receives): Yes



Purpose:

This two side by side illustrated donut pie charts can easily reveal how a scholarship status among students can strongly correlate with the student outcomes like whether they tend to remain enrolled or gradually dropout or graduate successfully.

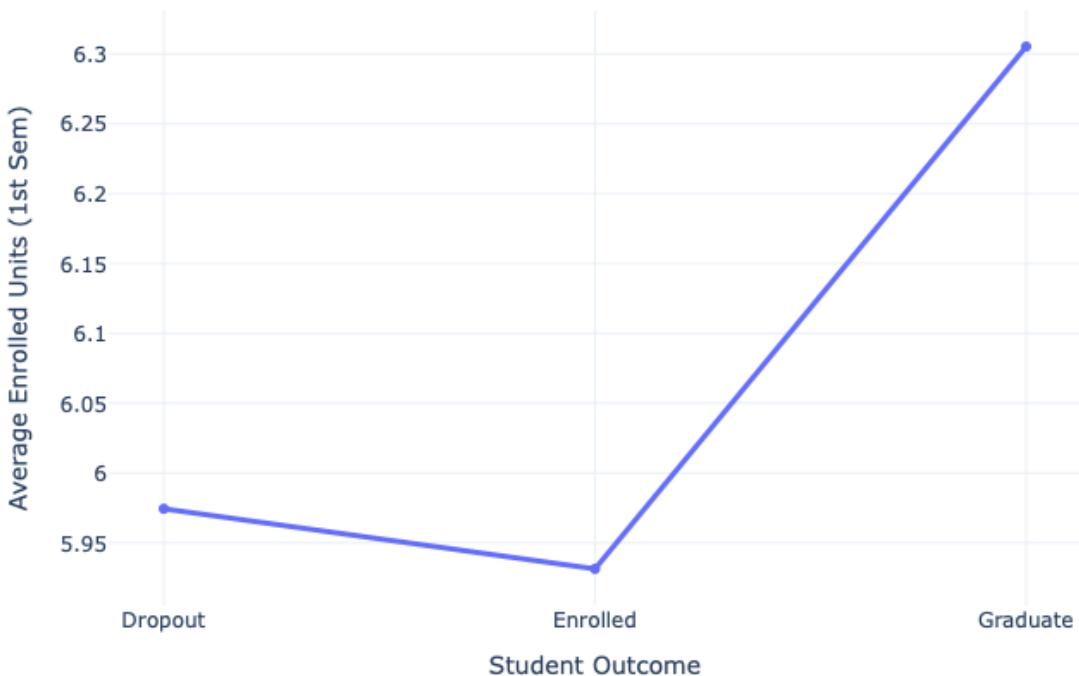
Insights:

- a. SCHOLARSHIP ISSUED students show an overall graduation rate of 80.3% which is significantly high and the dropout rate is too low like 7.6% .
- b. NON-SCHOLARSHIP ISSUED STUDENTS show graduation rate of only 50% which is way less than the issued ones.
- c. This tells us that financial support will allow these students to improve and focus on their studies leading to successful graduation.
- d. This tells us that academic institutions should focus on their financial support to the students to improve their overall graduation rate.
- e. This visualization is perfectly done using a donut pie chart which can accurately compare between two kinds of students.

Visualization 5: Line Chart

Title: Average number of Courses Enrollment Versus Student Outcomes(graduates, Enrolled, Dropouts) on the basis of Credit Hours taken.

AVG NO OF ENROLLED UNITS BY STUDENTS IN 1st SEMESTER/STUDENT OUTCOMES



This chart helps us to uncover one key point like how many credit hours are important for the student to have a successful outcome. This helps us understand key questions like?

- How many credit hours are ideal for a student?
- Does having enrolled in more ideal hours leads to more dropout rates?
- How can these insights be used to target the dropout rate in institutions?

Insights to be gained:

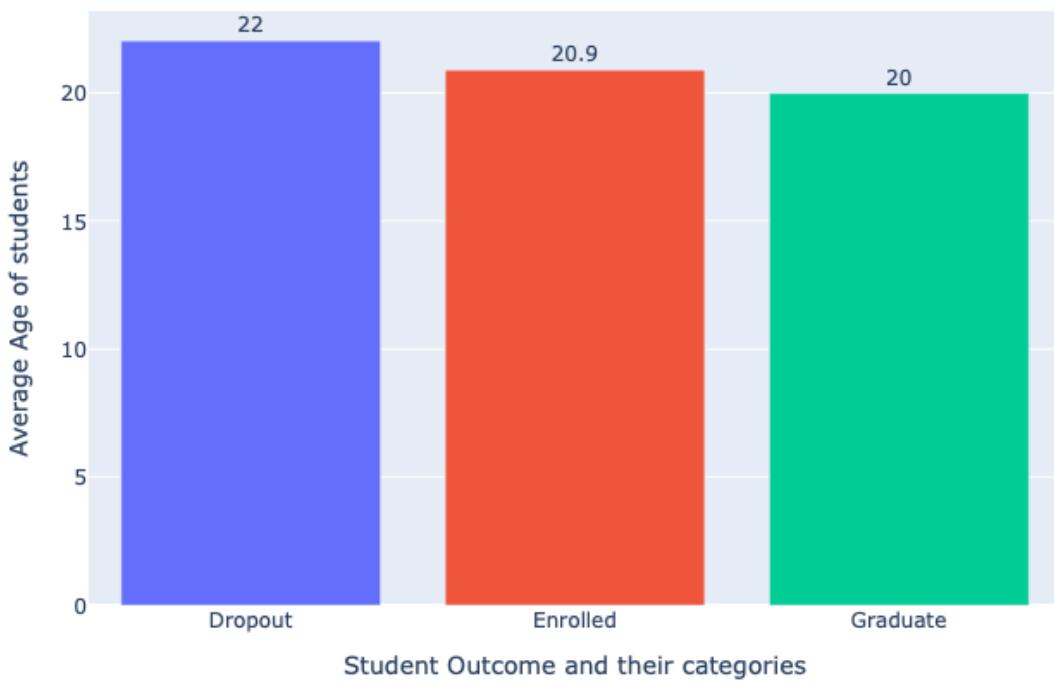
- This chart is calculated at the first semester with a near same credit hours.
- One key thing to understand here is how lower credit enrolled hours by students tend to easily dropout more.

- c. Those who graduate here show a different perspective by enrolling in more credit hours in the first semester.
- d. This shows us that graduates have more willingness to take academic responsibility very seriously.
- e. This trend can point out to institutions saying that people who takes less courses tend to dropout very soon.
- f. This can be a crucial visualization for institutions to take:
 - 1. Optimizing the career credit load.
 - 2. Counselling process.
- g. This helps in understanding the enrollment behavior using patterns and allowing academic colleges to target those students.

Visualization 6 : Bar Chart

Title: Average age of enrollment versus Student outcomes on Retention

AVG AGE OF ENROLLMENT AND THIER POSSIBLE OUTCOMES



Purpose: This chart is visualized how age of enrollment can serve as a warn sign for students who tend to drop out more likely.

Insights to gain:

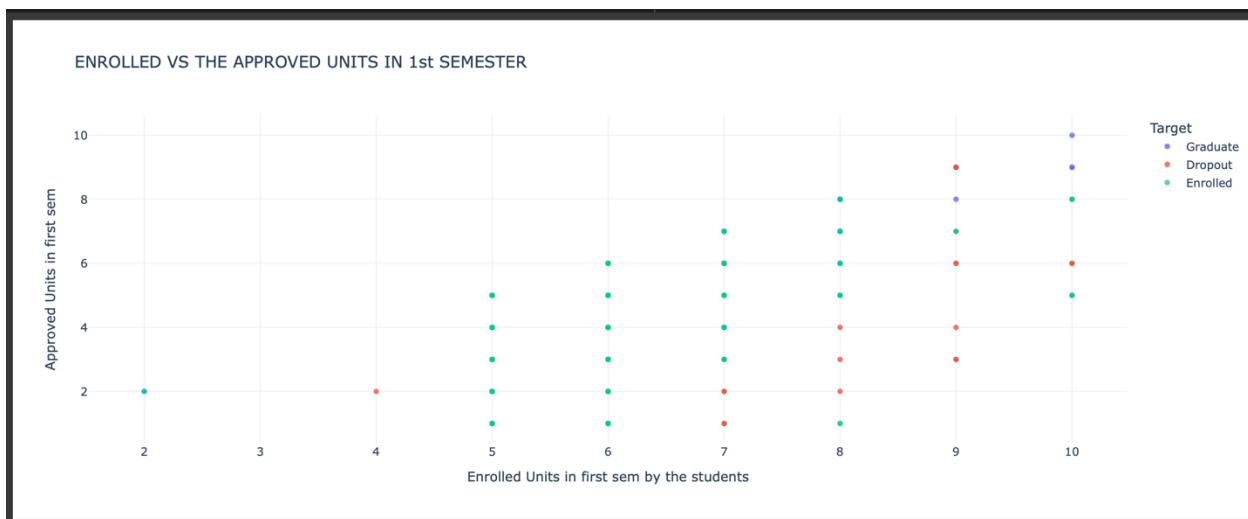
- Here, younger students tend to graduate more.
- Dropout age is more around 22 and graduate age is around 20.
- This says that 'success rate is correlated with starting college early'.
- Subtle age can matter in a close dataset like this.

Academic institutions can use these insights to:

- Implement support programs for older coming students.
- Implement different programs who are returning to the college after a break.
- Offering advising, help and learning options in advising offices to mitigate this issue early.
- This is a takeaway to make sure that the older students are not left behind and add encouragement to all kinds of students.

Visualization 7 : SCATTER PLOT:

Chart title: Exploring the workload versus the Success Rate on different categories.



Purpose:

This scatter plot helps us to identify a primary relationship that exists between the academic workload that the students tend to choose and how many of those actually complete those. This graph is mapped to all kinds of categories like,

- Graduates.
- Dropouts.
- Still Enrolled.

Insights:

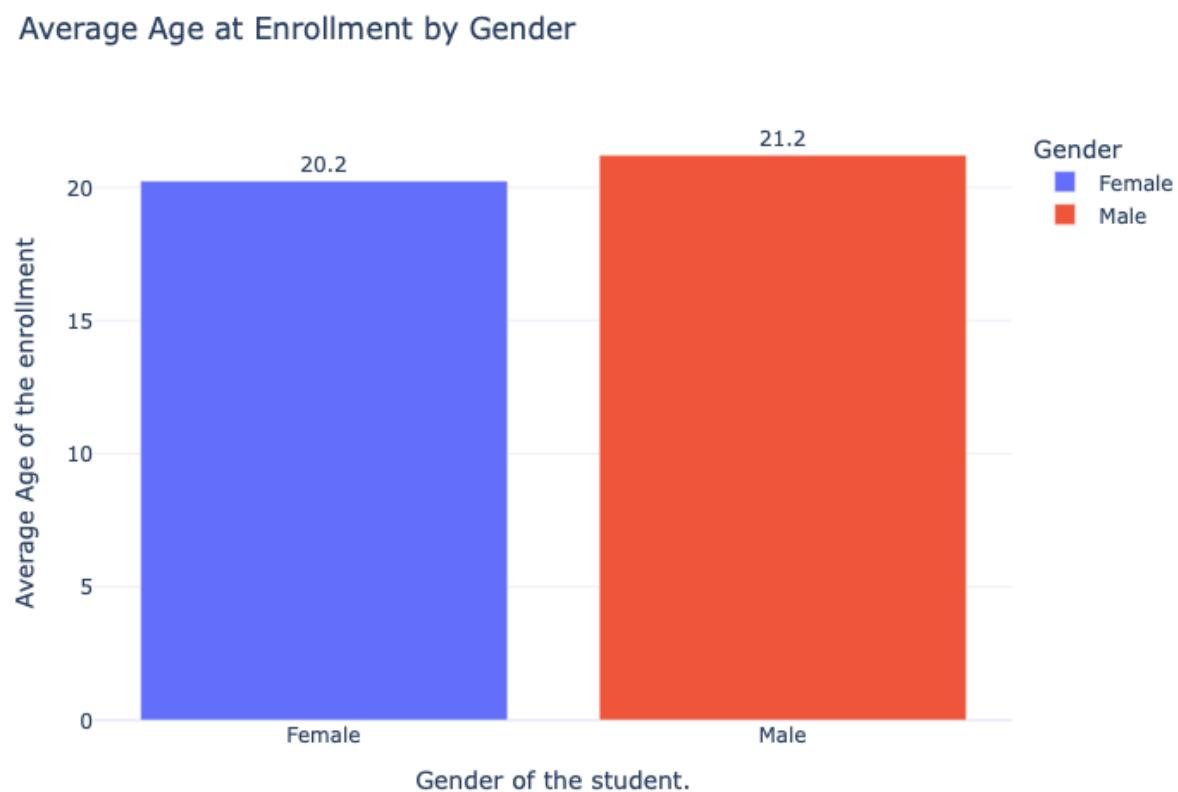
- a. We find that the graduates tend to cluster more around more credit hours and approval.
- b. Students who tend to enroll in 9-10 units (credit hours) often tend to have a successful completion.
- c. The dots indicate that there is a consistency and a good follow through.
- d. Dropouts : These students followed a wide spread with a improper completion.
- e. Dropouts are only enrolled in 2-6 credit hours .
- f. Enrolled students : these students show a consistency by keeping enrolled in 6-9 hours with moderate success and this outcomes may vary .

Key takes :

- a. These visualizations can help academic institutions to identify the potential dropouts early.
- b. This can help colleges to encourage students and come up with strategies to overcome these.

Visualization 8 : Interactive Bar Chart

Chart title: A Demographic Perspective on Gender and Enrollment age



Purpose:

This chart helps us by give a simple idea on how gender can have a impact on the average age at which the student tends to begin their academic journey.

Observations and key points:

- a. We could see that males tend to enroll slightly late.
- b. Males enroll a year later than girls on an average.

Does this matter:

- 1. We can see that from our previous charts which shows that enrollment at a older age may tend to have dropout risks.
- 2. Males on this chart can have a chance on this risk which can be a point out for academic institutions.

What does institutions do to prevent this?

- a. Can offer advisement which can tailored to specific age and gender.
- b. Can have awareness campaigns on importance of engagement and academic success.
- c. Provide flexibility support.

Visualization 9: TREE MAP

Chart Title: Outcomes on Students based on the Courses they enroll:

Purpose for this visualization:

This tree map can present us with a whole new understanding of how different courses taken up by students can pose an impact on the student's success rate and drop out rates.

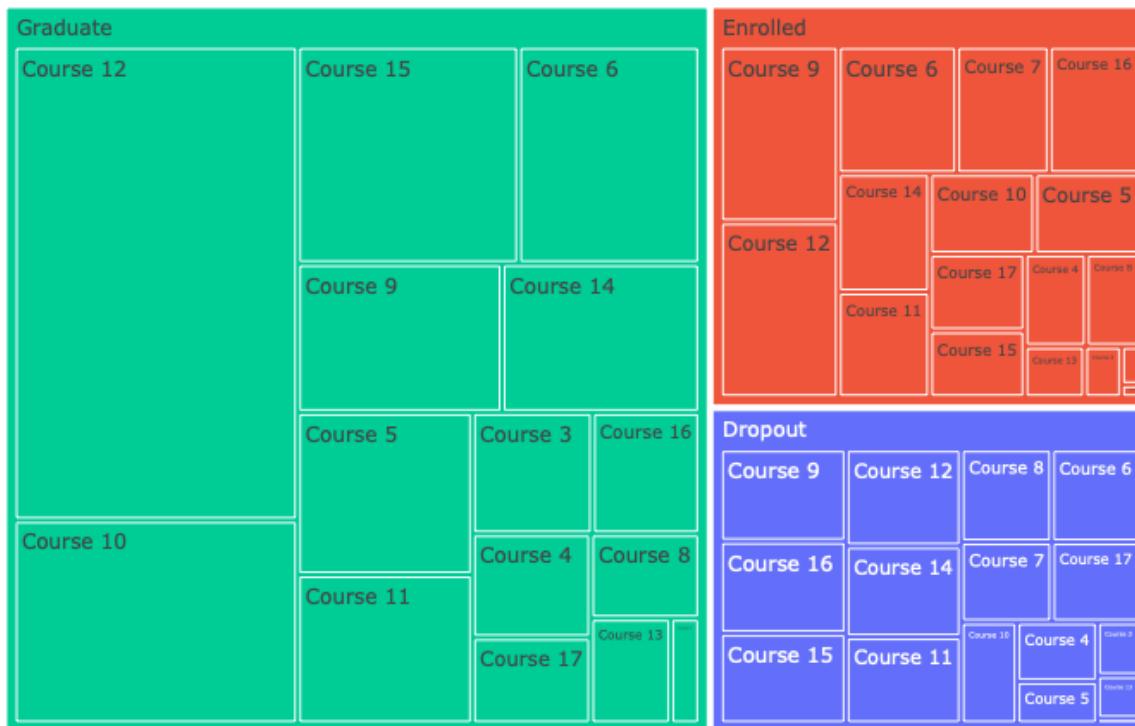
Key takeaways and points:

- a. Graduates heavy courses: Courses 10,12 tend to dominate more in the graduates section.
- b. These courses has a much more graduation outcomes suggesting that these courses are good overall.
- c. Courses 9,6 has been shown up in both enrolled and dropout categories suggesting that these courses do come up with challenges and students may succeed or fail in doing that.
- d. Potential Hard Courses : courses 9,12,8,16 seems to appear in a larger areas in the dropout categories.

Does this matter?

- a. Yes, the success or failure of the students can also be heavily relied on the courses they choose to take .
- b. Some courses tend to support students and some can create challenges which helps students grow.
- c. This can give the colleges a clear idea on how to help students who enrolled in those dropout courses and this tree map can be beneficial in understanding those.

Treemap of Student Outcomes on basis of Course



VISUALIZATIONS USING ALTAIR LIBRARY: (7 visualizations)

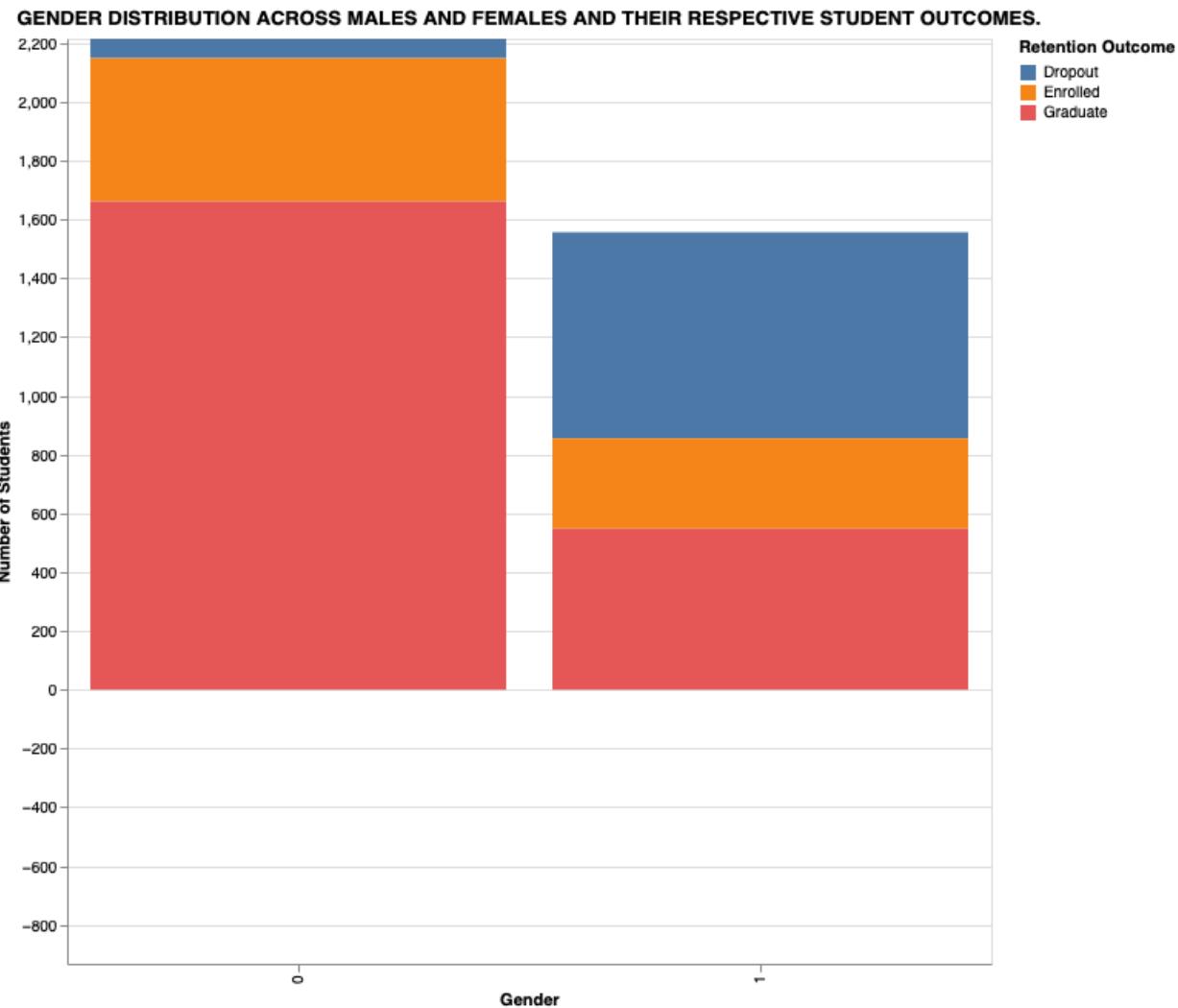
Visualization 1: Stacked Bar Chart

Chart Title: Plotting Gender Distribution against the Retention Outcomes

Purpose:

- a. This chart is aimed to compare the number of male and female students who are currently being enrolled or being dropped.

- b. To understand a demographic balance that helps us understand student retention.



Observations:

- The females denoted by 0 are being a large group overall.
- This group has significantly high graduate count when compared against dropouts and enrolled students.
- A few graduates in males when compared to females.
- This reveals us with a high proportion that exists in dropouts when compared with the female students.

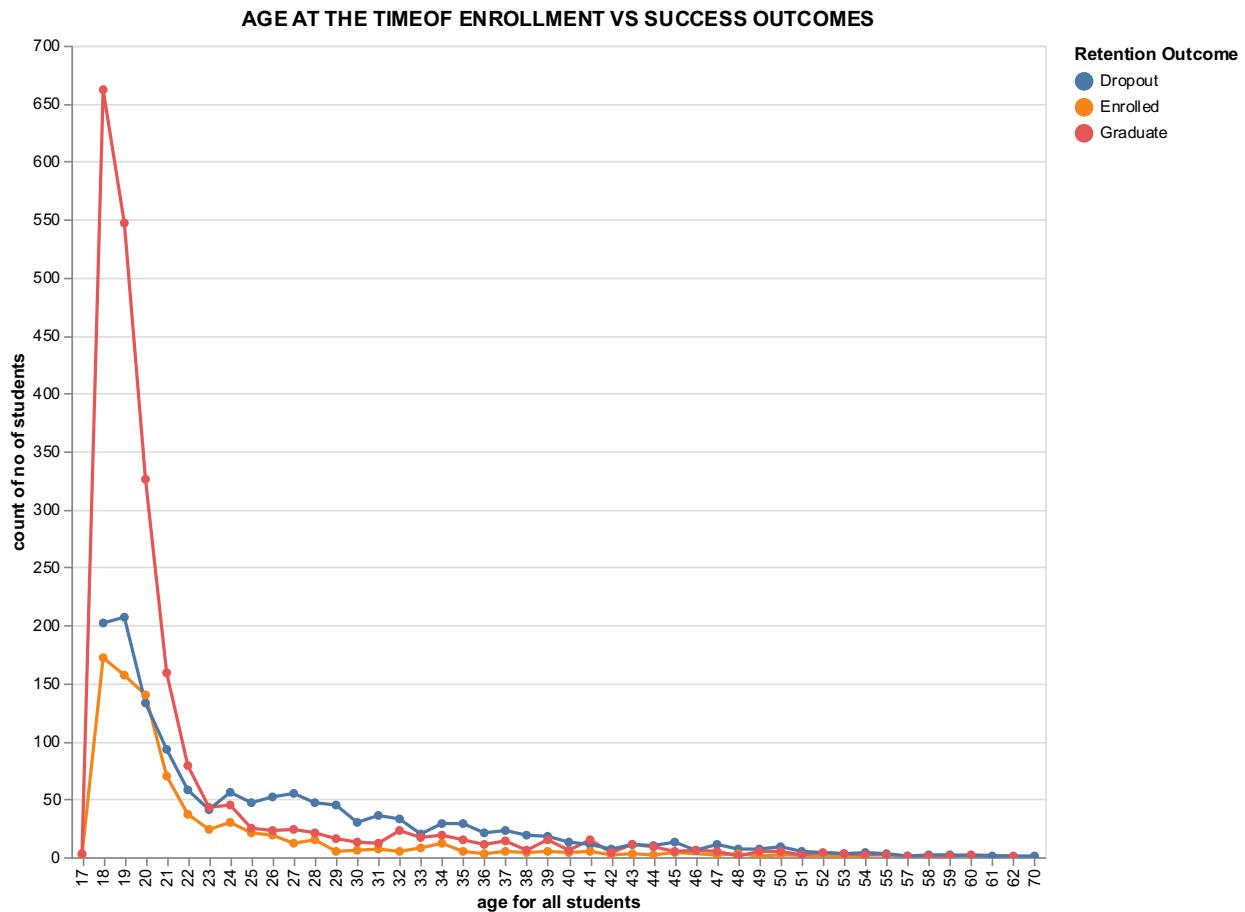
Insights:

- Female students not only enroll more but also tend to graduate more with a good rate when compared with males.
- Males pose a higher risk to dropping out based on our data.

- c. This tells us that 'demographic balance do matter'.
- d. Retention programs should target on male students early in their careers to hold them .
- e. Can provide us with how each gender performs in each category.

Visualization 2: Plotted using Line Chart

Chart title: age at the time of enrolling versus their success outcomes.



Purpose of this chart :

This interactive chart can help us analyze how a student's age at a given time of enrollment can 'correlate' with their academic success/ a risk of dropping out.

Observations and insights:

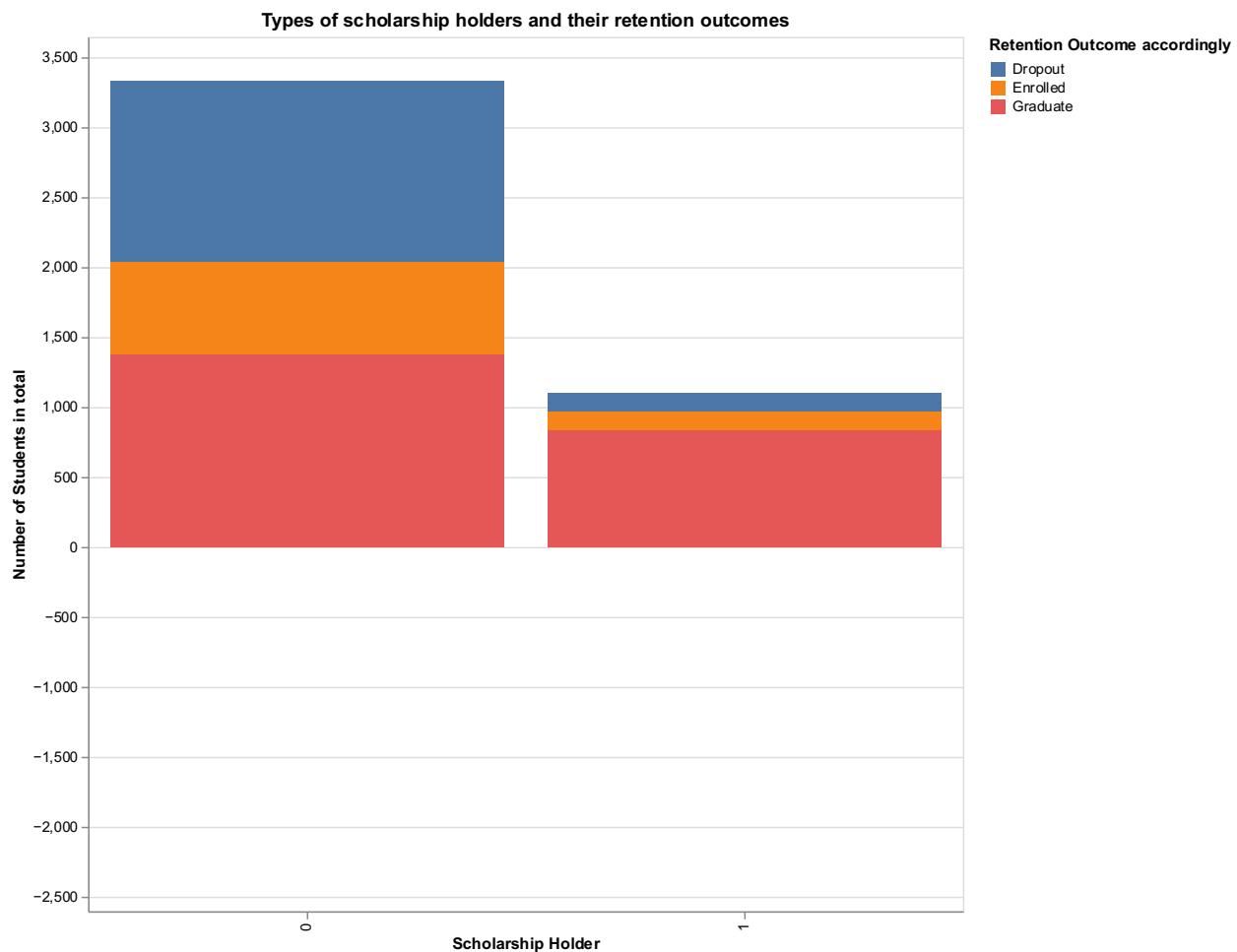
- a. This helps us identify specific age that requires support.
- b. Visualize trends.
- c. AGE can be significant predictor of outcomes .

Key observations:

- a. We can see that at the age of 17-21 shows younger students who has a high graduation rate.
- b. More academic momentum is observed at high school students .
- c. Older students with ages 21+ has an increased dropout rate: which shows that many factors like work, family pressures / financial problems could be a impact on this results.

Visualization 3: Stacked Bar Chart

Chart Title: Scholarship holder versus their success outcomes.



Purpose : this chart helps us to analyze whether students who are receiving a scholarship can have a impact on the likeliness to student's graduation / enrolled or potentially dropping out.

Key observations:

- a. Students who receive scholarship
 - 1. Tend to show a much high graduation rate.
 - 2. Dropout rates are very lower here
- b. Students who doesn't receive scholarship:
 - 1. Poses more dropout rates.
 - 2. Lower amount students tend to graduate here.

Insights:

Scholarships here tend to have a much positive influence on the retention rate and success of the students which helps them not to worry about the financial stress.

Visualization 4: Stacked Bar Chart

Chart Title: Marital status of Students and it's impact on the retention rate

Purpose:

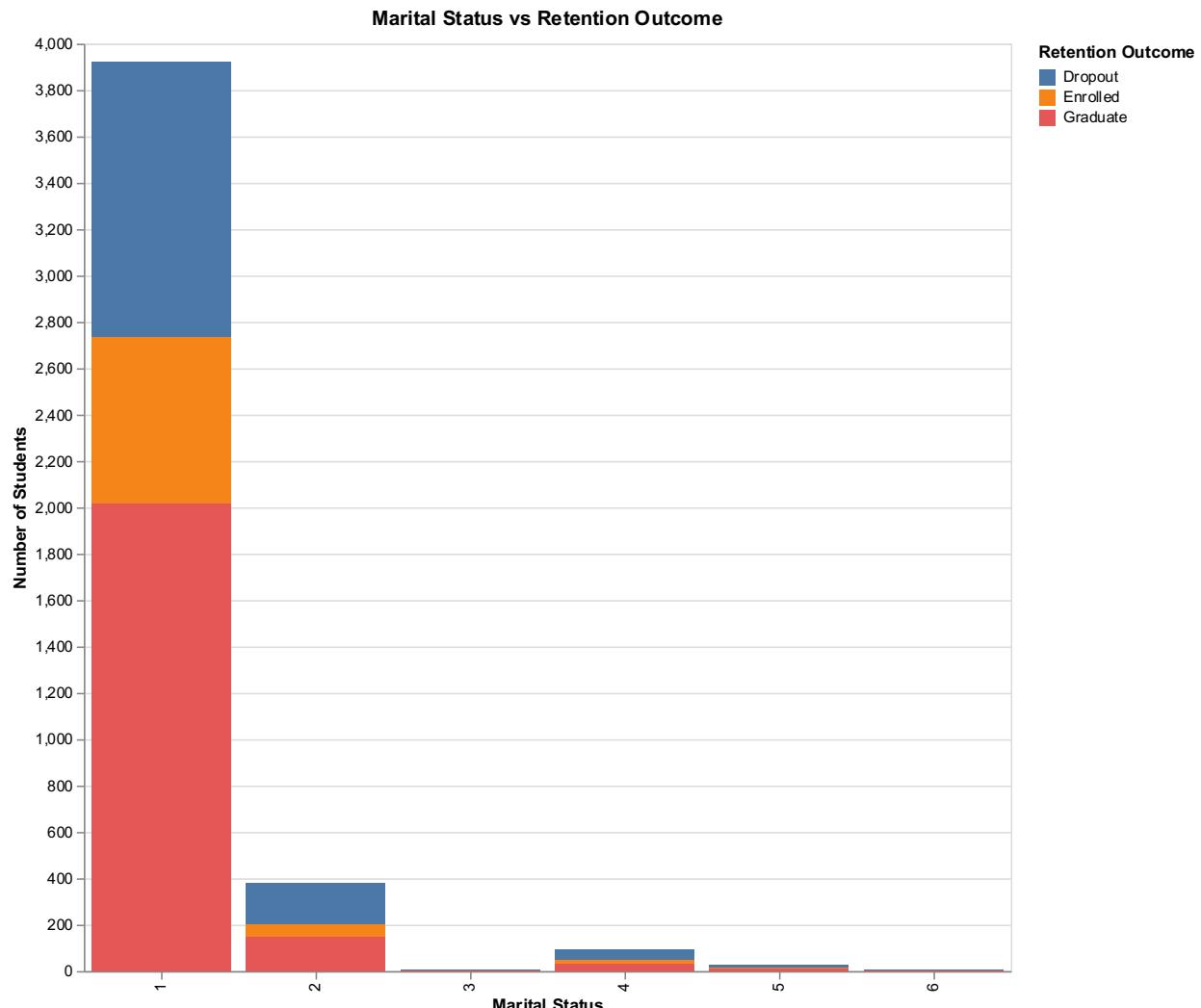
This is done to compare the retention rates and the dropout rates across all kinds of students. This helps us in understanding how marital status can help play a crucial role in the success.

Key Observations:

- a. Single students here represent a larger proportion of overall categories.
- b. Tends to show a much higher graduation rate.
- c. Drop out rates are few compared to the married students.

Key insights we learn :

- a. Complexity of the life can also bring out the academic outcomes and these are intertwined.
- b. Single students tend to perform way better in terms of a successful graduation , married students and other can gain extra benefit when offered with extra academic and support services.



Visualization 5: Scatter Plot:

Chart title: age vs 1st semester outcomes:

Purpose: this is prove “ AGE MATTERS AT ALL CAUSES , NOT AT AN ENROLLMENT BUT CONTINUES TO EFFECT THE OUTCOMES AFTER FIRST SEMESTER”.

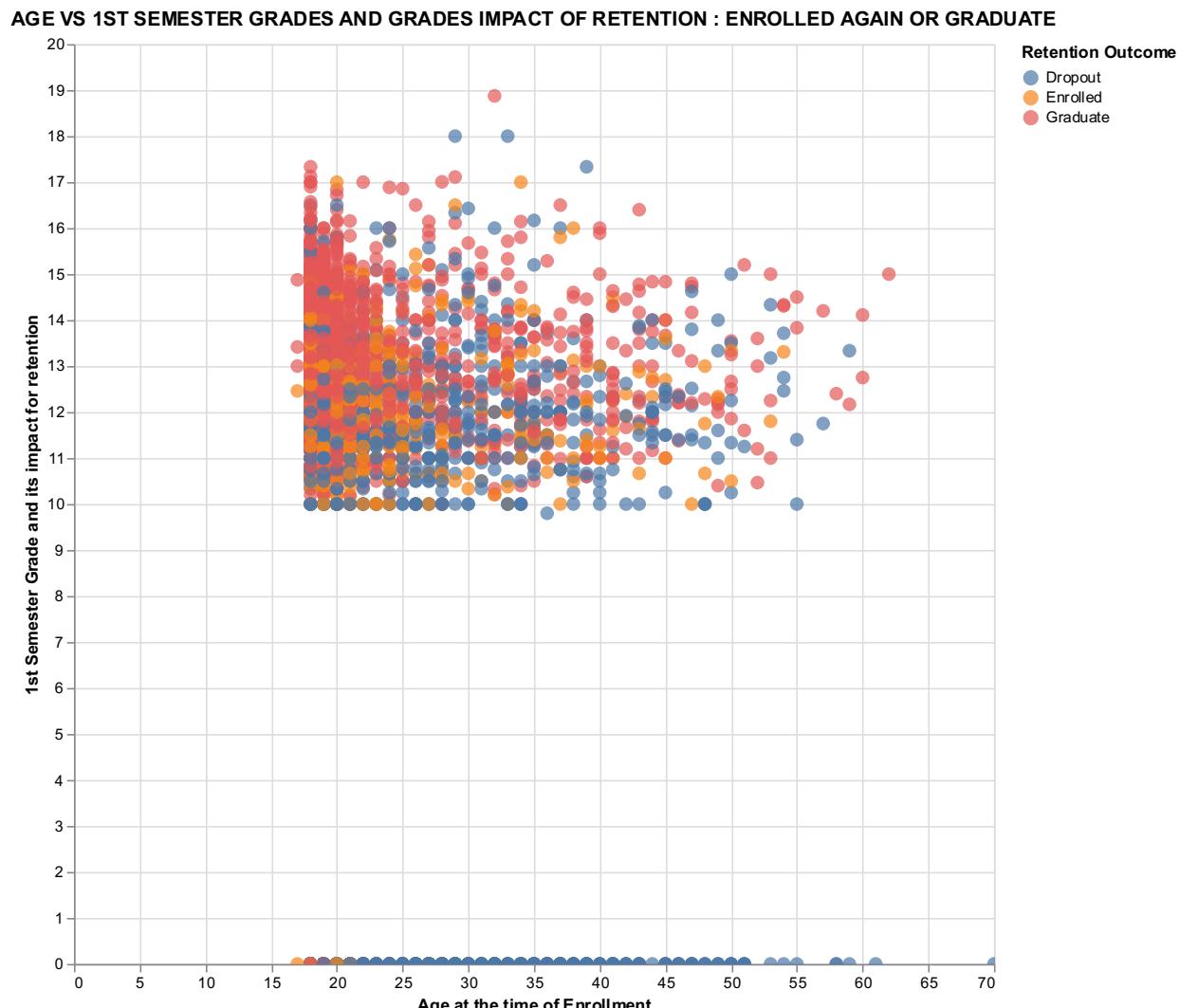
This chart is again visualized using scatter plot to really see how different categories of students scatter in the chart. This helps us a bit more in how age can impact the performance.

Observations:

- a. Younger students likely tend to cluster over successful outcomes (graduation or a continuous enrollment) .
- b. Older students shows us with more dispersion.

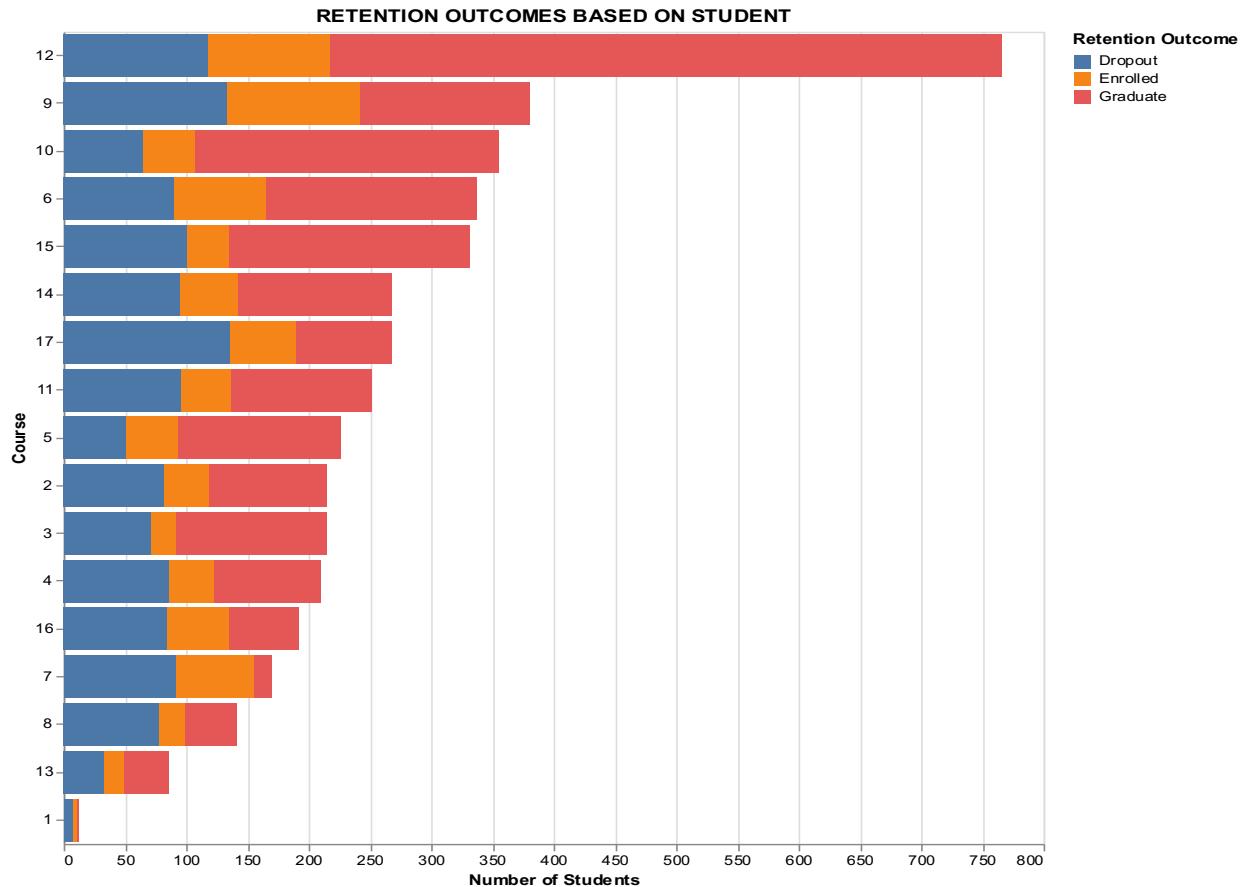
- c. A higher proportion of dropouts can be seen in students who are above 21 and near 22,23 student age.
- d. The beauty of scatter plot is that we can see that outcomes become very less predictable as the age increases , the reason is being many factors that come up in student's life which can have a good amount of impact.
- e. Color coding in this chart makes it clear how outcomes are linked.

Insight: first semester is proven to be crucial for older students , here in our chart age based clusters reveal that a intervention should not be a one size that can fit all. Support groups and mentorship, coaching could help us boast retention in older enrolled students.



Visualization 6: Stacked Horizontal Bar Chart

Chart Title: Retention outcomes based on the type of course:



Purpose:

- This aims to visualize more beautifully how courses have a high dropout / enrolment / graduation rates.
- Identify risk courses based on previous outcomes and suggest students about the course challenges.
- Help administrators redesign their courses to target retention .

Key Observations:

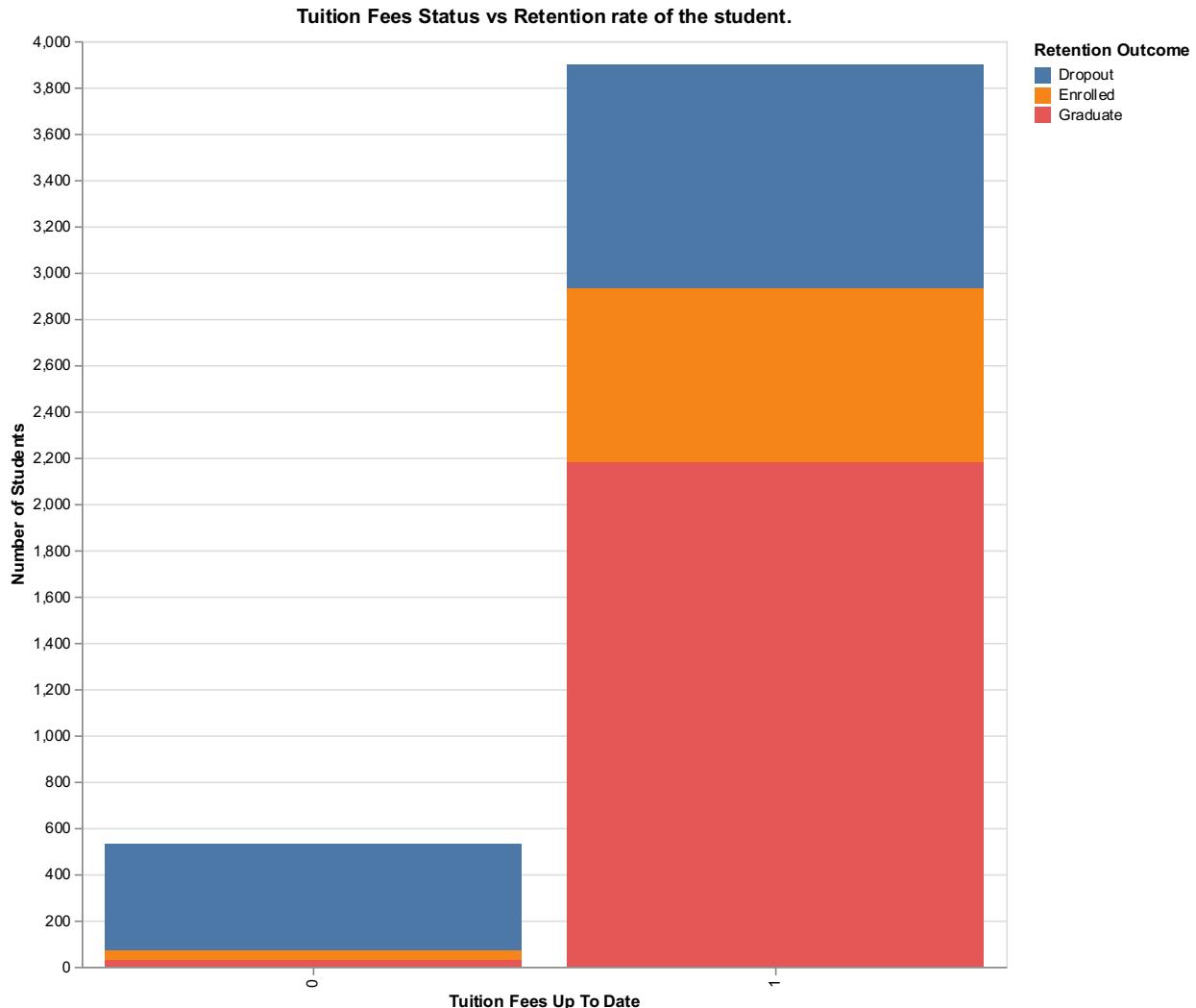
- Certain courses standout as high dropout based courses.
- Courses with high dropout rates has a better structure and support systems and aligns with student's interests.
- Relative size of the enrolled for each group shows us with popularity and demand of that course but 'that doesn't guarantees us with success'.

Insights: Just because a course is termed popular or high enrollment does not always mean that it is easy or comes with a guaranteed success. Institutions should allocate more resource on courses with high dropout rates to improve student's retention .

Visualization 7: Stacked bar chart showing us “Tuition Fee Status versus Retention Outcomes”.

Purpose:

- a. This is to help us examine a new relationship between the tuition fees status and the student retention outcomes.
- b. Can help us prove that ‘ financial stress has an influence over academic performance’ .
- c. Suggest advices to academic intuitions.



Key observations:

- a. Students who paid their tuition fees showed higher graduation rates and lower dropout rates.
- b. Students with unpaid / pending fees showed high dropout rates.

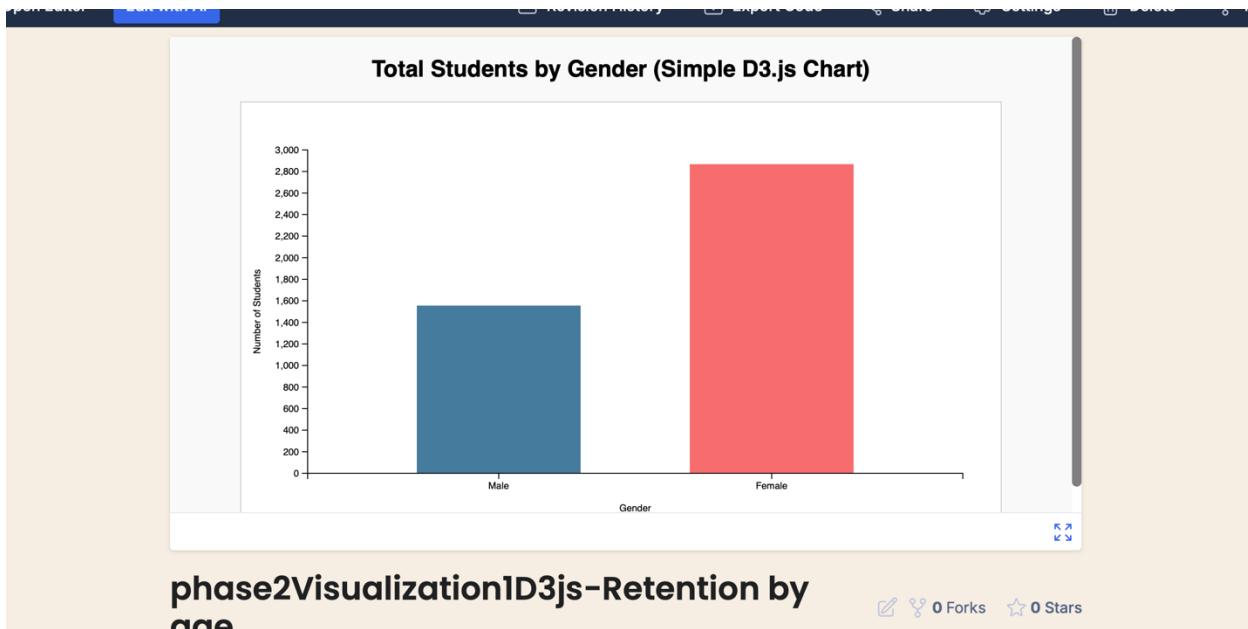
Insights: Having a financial stability could play a significant role in student's thriving nature and we can see that students relieved of the tuition fees pressures tend to more likely to thrive in academics.

Visualizations using D3.js (Viz hub links provided)

Visualization 1 : Bar Chart

Chart title: Total Number of Students by Gender

Link : <https://vizhub.com/gnevercodes/Visualization1>



Purpose :

- This is done to visualize the distribution of students based on a specific gender.
- A simple interactive chart to visualize it on a web page.
- To identify the kind of 'gender imbalance' in our existing data.
- To present us a decision using data driven approach.

Insights:

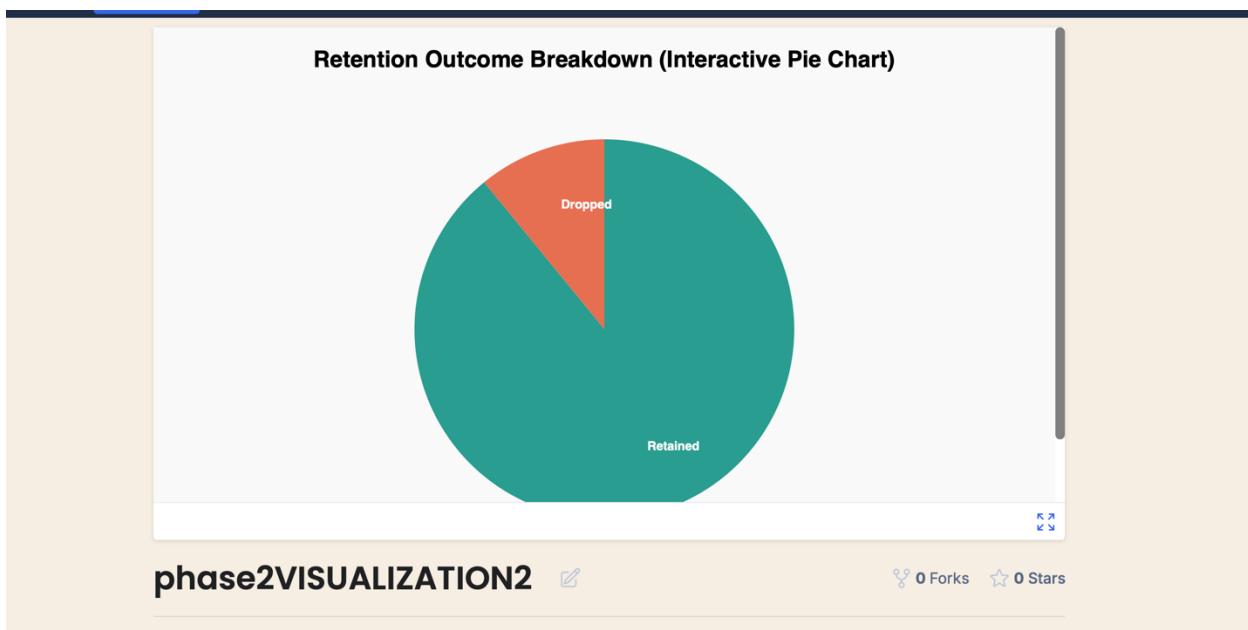
- A visible gender imbalance exists in our data and we can say that the females being the majority in ours.
- This suggests academic colleges to go for

1. Program tailoring : Colleges might want to consider reviewing their system to offer equity balance to exist between both genders.
2. Some of the existing support systems need to target the problem over here to improve engagement in male students.
3. Providing scholarships and outreach could be developed in colleges to retain the male students to balance out this concern.
4. We need to analyze more charts to evaluate to know what are the possible causes to this problem and mitigate this(cause the male percentage could be affected by multiple factors like stress , financial problems and many more) .
5. Display additional charts to reveal the reason behind this.

Visualization 2: Interactive Pie chart

Chart title : Retention breakdown in students.

Link : <https://vizhub.com/gnevercodes/visualization2>



Purpose:

- a. To visualize the overall percentage of students who were retained versus who dropped out of all the students.
- b. To quickly understand how this scales to the percentage of success/ challenge to the institutions.
- c. A good basic view for the administrators and decisions makers to get a whole good view.

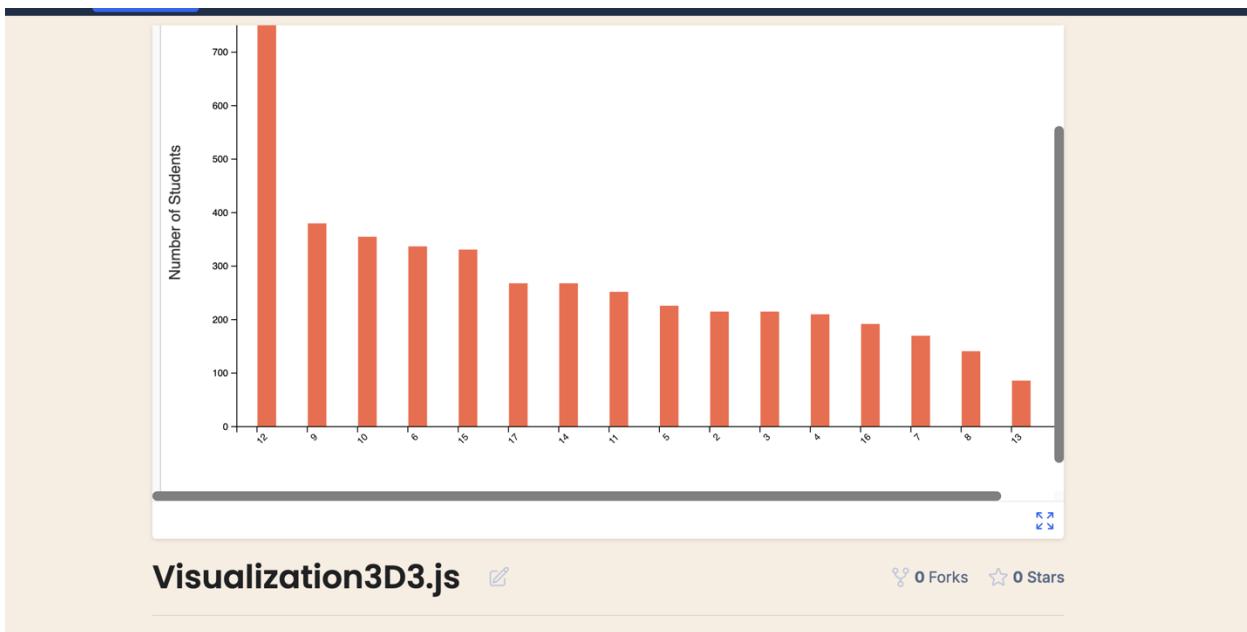
Insights:

- a. High retention rates indicates us with a positive one of overall institutional performance.
- b. Opportunities of improvement :
 - a. This chart helps us analyze the dropout part and can help us uncover which factors could lead on this decisions like the factors we mentioned before.
 - b. Design the system better early with this chart acting as a warning systems.

Visualization 3: Number of Students dropped out according to each course

Plotted using a bar chart with a tooltip

Link: <https://vizhub.com/gnevercodes/visualization3>



Purpose:

- a. This is aimed to analyze how dropout is distributed across different courses.
- b. To identify which courses are being the hardest to have high dropouts.
- c. To spot the academic risk zones and propose strategies.

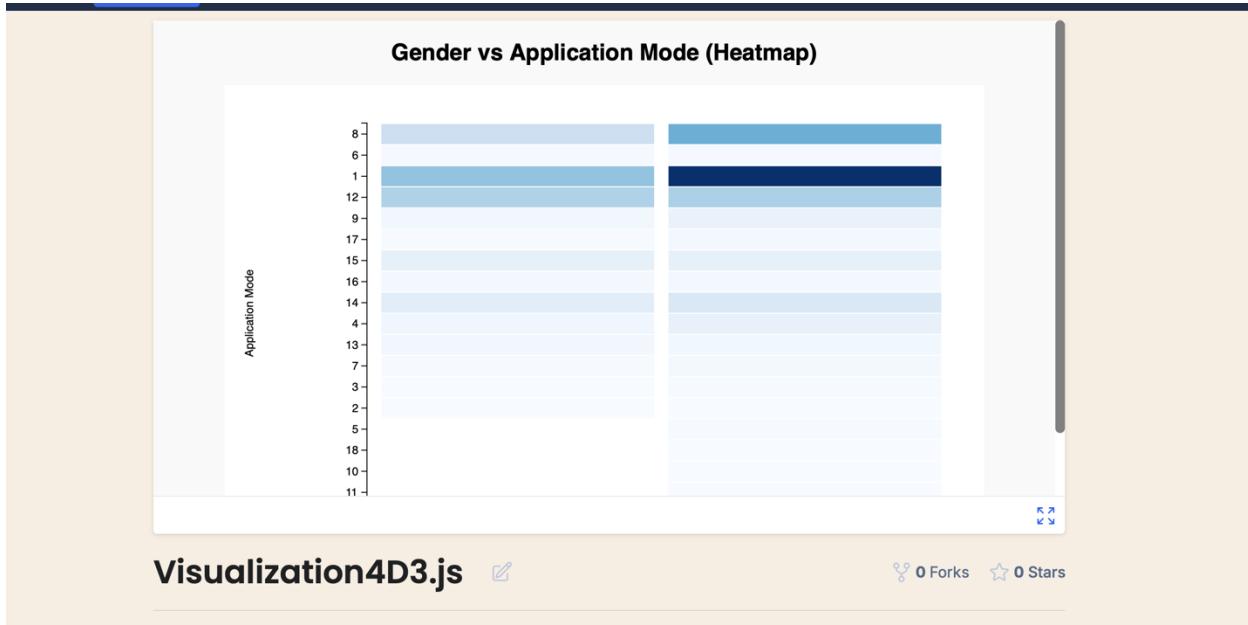
Insights:

Universities can significantly work to improve the overall retention rates by focusing (just) on some of the courses which has the high amount of dropouts. By this chart we can find out risk courses and having an early academic interventions for this handful of courses could lower the dropout rates.

Visualization 4: Heat Map

Gender vs Application Mode

Link: <https://vizhub.com/gnevercodes/visualization4>



Purpose:

This chart helps us to understand how different modes of application among the males and females are being used.

Insights:

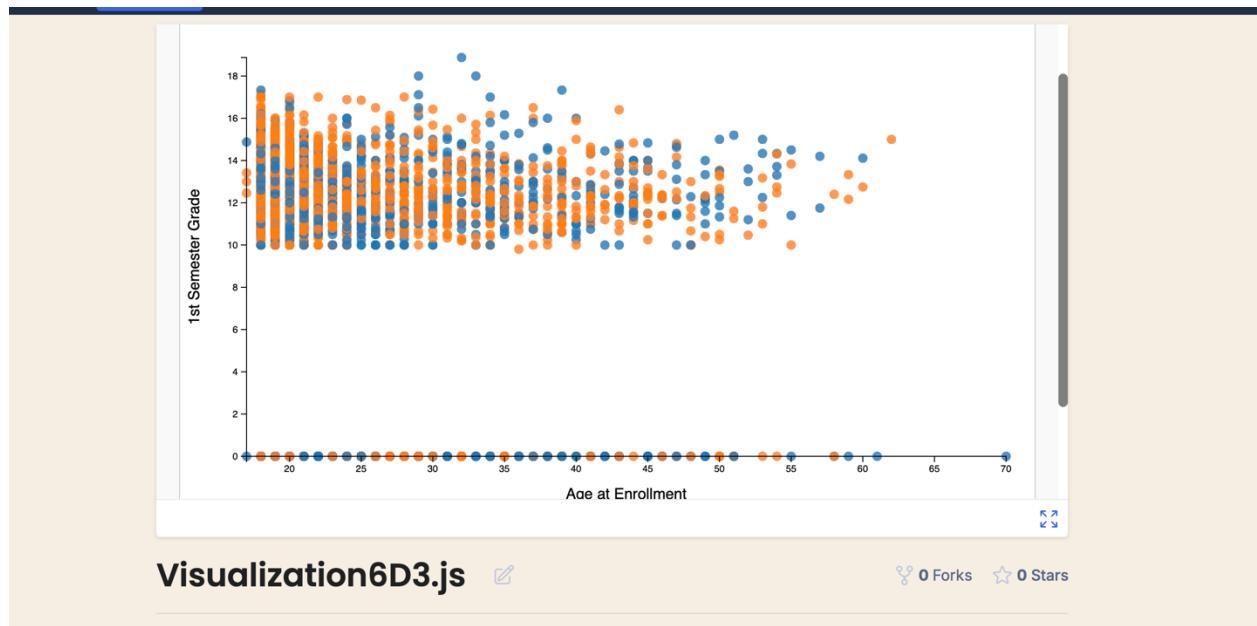
- This helps us to spot trends and preferences choice according to gender.
- To spot which application mode is used by a certain gender and to come with a whether this tells us if there is any 'gender bias'.
- This chart tells us that application mode 1 is the highest and tends to dominate.
- The reason for low color on the males section is due to the fact that the enrolled students on the male side is very low hence the color.

Crucial takeaways:

- This tells us that if the institutions continue to rely on certain modes , then this dependency on a specific one could be risky.
- Other modes should be encouraged too.
- Gender inclusivity check should be done to balance it out.
- Should focus on improving other application modes too.

Visualization 5: visualization using scatter plot

Link : <https://vizhub.com/gnevercodes/visualization5>



Insights:

This chart is aimed in clusters aka scatter plot to spot the trends of different age groups. We can say that the younger age groups tend to enroll more and most likely to succeed. This tells the academic institutions to focus more on the older groups as the dropout rate tends to increase more in that.

Institutions should offer:

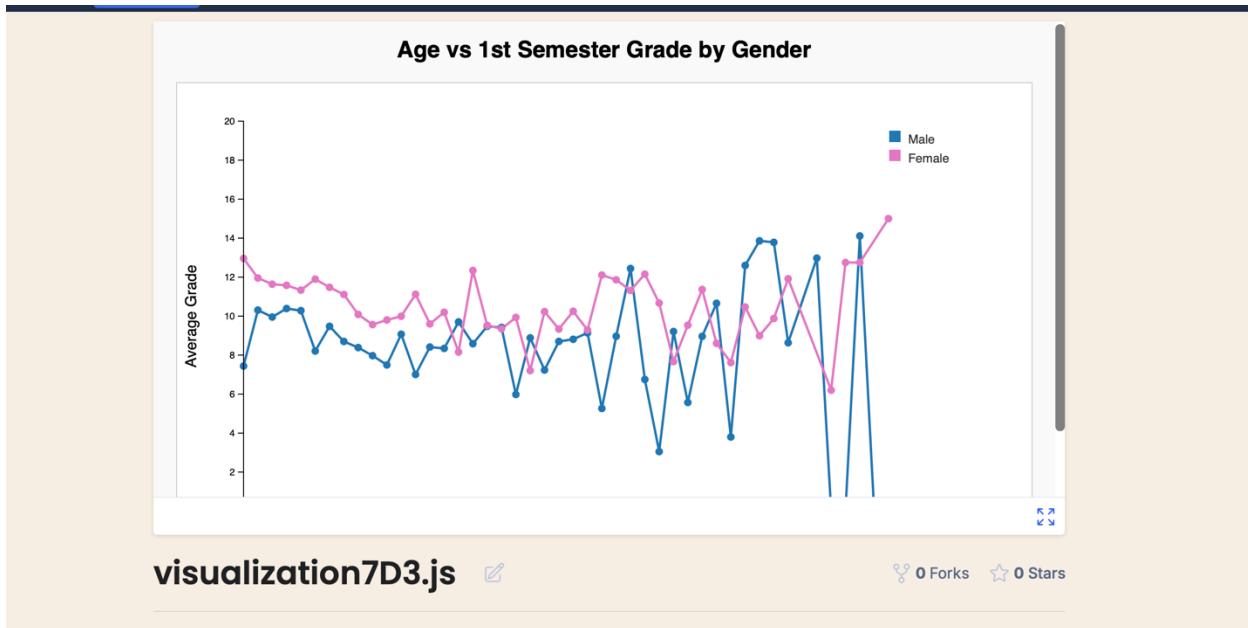
- Academic bridging performance.
- Flexible timings.
- Personalized tutoring.

Visualization 5: Line chart

Chart Title: Age vs 1st semester visualized on basis of Gender

Viz hub link: <https://vizhub.com/gnevercodes/visualization5>

Purpose: In order to show that there exists a trend on how grade changes on the basis of gender.



Key observation on the basis of graph :

- We can see that the female students has an higher average grades when compared against male across most ages.
- Also we can find a much more stable trend on female ones as the female grades stays around 11-13 average up the ages of 40.
- Even on older ages , female tend to dominate.
- Male students have an overall low avg grades to females .
- More fluctuations happen across ages 30-50 as expected because less data is concentrated around those ages and it does not reveal true pattern around those ages.
- Grades tend to scatter around old ages.

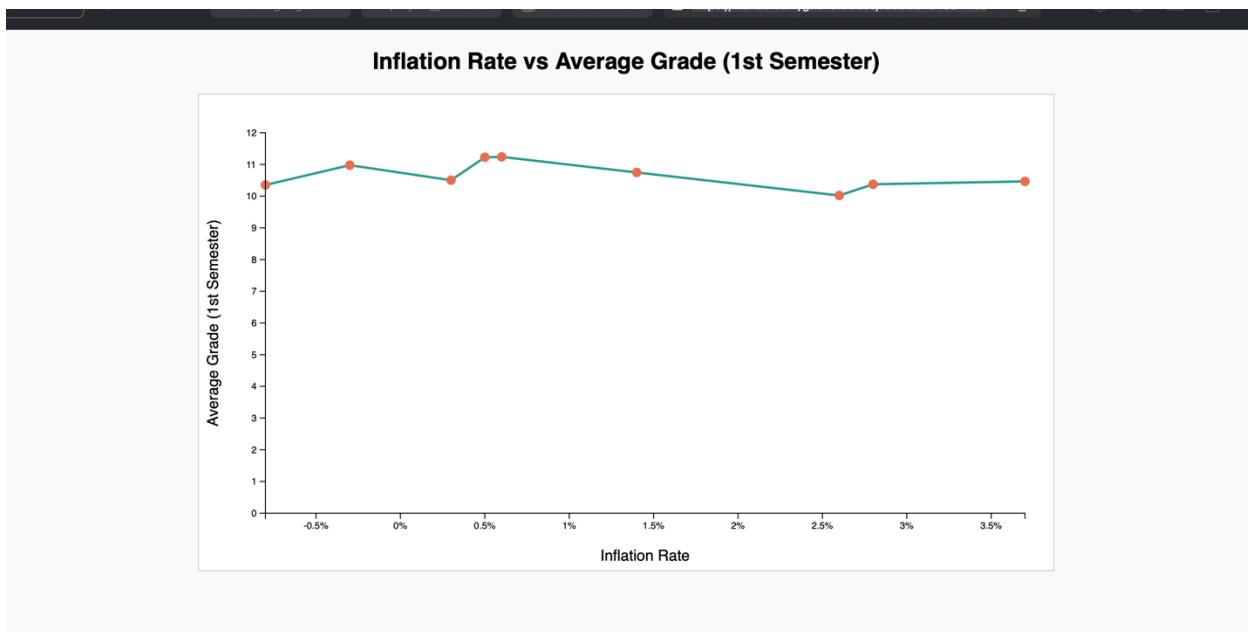
Insights:

- Females tend to adapt more to the academic environments regardless of any age.
- Males need support for old ages.
- Low and fluctuating grades happening across older ages could correlate with a higher dropout rate.
- Institutions should consider advisement and gender based sensitive support services.

Visualization 6: Line chart (trends)

Chart Title: Inflation rate versus the average grade comparison in 1st semester

Link: <https://vizhub.com/gnevercodes/vizualization6>



Purpose:

This is done to analyze the existing microeconomic factors (like inflation rate) and to prove that whether those could have an impact on the student's overall academic performance.

Key Observations:

- At a low or negative inflation rates , the average grades are slightly lower.
- As inflation tend to stabilize we could see that the average grades peak between 11-15.
- At a high inflation rate the decline of avg grades is observed.

Overall pattern: Grades are relatively stable but show a slight inverse relationship showing us that economic conditions can have a indirect but measurable impact on the overall performance.

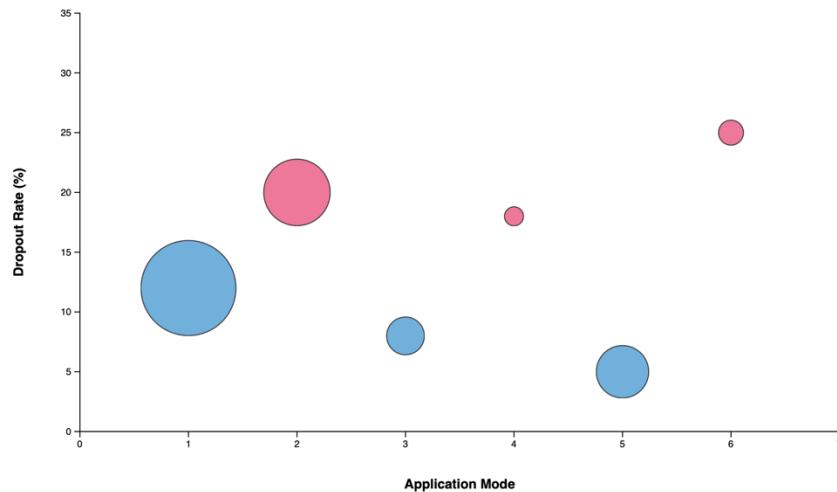
Visualization 7 : Bubble Chart

Chart title: Application mode used vs Dropout Risks with color based on gender domination

Blue = male , pink = female (dominated bubble for application modes)

Link : <https://vizhub.com/gnevercodes/visualization7>

Application Mode vs Dropout Risk (Bubble Chart)



Purpose:

- This aims to analyze how different types of admission application modes can have an influence on the dropout rates among the students.
- To visualize how the size of each application mode and gender dominance.

Observations:

- Application mode 1 has the largest number of applications with a moderate dropout rate with male domination.
- Application mode 2 and 4 have smaller number of applications but they have highest amount of dropout rates.
- Moderate is observed for mode 5

Interpretation :

Certain some of the less application modes pathways seems a bit riskier moves for student retention.

Insights:

- Aimed to prove 'admission method matters'.
- Gender factor is also observed.
- Strategic focus by universities on application modes.

Integrating Power Bi to a web page:

Html link to paste on a website:

```
<iframe title="ganeshdrusya_datavisproject" width="600" height="373.5"
src="https://app.powerbi.com/view?r=eyJrIjoiNWNiMzkzMDUtZTQ4Zi00MWZkLWFjZTMtNWVi
Mjk5MmZiNDIzliwidCI6IjcwZGUxOTkyLTA3YzYtNDgwZi1hMzE4LWExYWZjYmEwMzk4MyIsImMiO
jN9" frameborder="0" allowFullScreen="true"></iframe>
```

Link :

<https://app.powerbi.com/view?r=eyJrIjoiNWNiMzkzMDUtZTQ4Zi00MWZkLWFjZTMtNWViMjk5
MmZiNDIzliwidCI6IjcwZGUxOTkyLTA3YzYtNDgwZi1hMzE4LWExYWZjYmEwMzk4MyIsImMiOjN9>

Work management :

- a. Drusya: 50%
 1. Worked on Altair.
 2. Worked on documenting the charts.
 3. Worked on d3.js charts.
 4. Focused on what kind of charts best visualizes our data.
- b. Ganesh: 50%
 1. Focused on story telling and identifying the relationship between variables.
 2. Correlation analysis.
 3. Documenting the project
 4. Worked on plotly visualizations and some of the D3.js charts.

Insights gained from doing this project:

- a. How to properly create charts that has a unique story to tell.
- b. How to visualize using different libraries.
- c. How to filter and structure the data.
- d. How to perform correlation analysis.
- e. How to add interactivity to have a deeper insight and meaning to the visualizations.
- f. How to use appropriate colors.