

# What Factors affecting Students Dropout in Higher Education in United Kingdom

Siddharth Chikalkar  
220264730

28 September 2023

## Abstract

This dissertation delves into the intricate web of influences shaping student attrition in UK higher education. The study examines a range of factors contributing to dropout rates, seeking to enhance retention strategies. Through a framework of research questions, it investigates demographic variables, curriculum effects, student contentment, and sociodemographic factors. Utilizing quantitative analysis, surveys, and literature reviews, the research uncovers insights to guide institutions, policymakers, and educators in evidence-based strategies for improving student retention. By unravelling the nuanced relationship between dropout rates and these factors, the study aims to foster tailored interventions, equitable educational environments, and enriched academic experiences for all UK higher education students.

**Keywords:** student attrition, higher education, dropout rates, demographic factors, curriculum impact, student contentment, sociodemographic influences, retention strategies.

## Acknowledgements

I extend my heartfelt gratitude to my parents and friends, whose unwavering support has been a cornerstone of my academic journey. I am immensely appreciative of the guidance and mentorship provided by my supervisor, Professor. Helen Higson, whose expertise, and insights have been invaluable throughout the course of this dissertation

## Table of Contents

Introduction.....	6
Literature Review.....	6
Introduction.....	6
2.1 Literature Review on Student Dropout in Higher Education.....	6
2.2 Literature Review on Factors Influencing Student Dropout and Predictive Analysis: Enhancing Student Retention Strategies in Higher Education.....	7
2.3 Literature Review on Academic Advising, Online Course Effectiveness, Teacher Turnover, and Economic Effects of Student Dropouts.....	8
Problem Description .....	9
To what extent can demographic factors serve as predictors of student dropout in the realm of higher education? (Lovenoor Aulck, Nishant Velagapudi, 2016) .....	9
Is there a tangible correlation between student contentment and the occurrence of dropout within the realm of higher education? (Rovai & Barnum, 2012).....	9
What are the sociodemographic factors that exert influence over attrition rates within the realm of higher education? (Larsen, Sommersel, et al., 2013, p. 14). .....	10
Description of Higher Education Statistic Agency .....	10
Demographic Student Drop Analysis.....	11
To what extent can demographic factors serve as predictors of student dropout in the realm of higher education ? .....	13
Annually Each University analysis .....	16
Year 2014 -15.....	16
For year 2015-16.....	17
For year 2016-17.....	18
For year 2017-18.....	19
For year 2018-19.....	19
For year 2019-20.....	20
Understanding the Russell Group and Oxbridge universities dropout rates. ....	21
Russell Group.....	22
Interpretating of demographic and university student attrition.....	24
Is there a relationship between student satisfaction and dropout rates?.....	24
The Data set Definition. ....	25
Data Preparation.....	25
Scale 03: Assessment and feedback.....	26
Scale 04: Academic support.....	27
Scale 06: Learning Resources .....	28
Scale 8: Student Voices survey responses.....	29

Statistical Analysis overview of survey questions and attrition.....	30
Correlation Between Student Dropout and Dissatisfaction Survey Responses.....	33
What are the sociodemographic factors which affect dropout rates?.....	36
Negative Correlation with Dropout Rates:.....	36
Anglia Ruskin University: .....	37
The Manchester Metropolitan University: .....	38
The University of Central Lancashire: .....	38
Analysing these correlations leads to the following insights: .....	39
Unemployment Rate and Dropout Rates: .....	39
Additional Sociodemographic Factors:.....	39
Temporal Factors: .....	39
Conclusion: .....	40
Demographic Factors Influence Churn Rates: .....	40
Oxbridge and Russell Group Universities: .....	40
Unemployment Rate Correlation: .....	40
Temporal Factors Play a Role: .....	40
Limitations .....	41
Further Work .....	41
References.....	41
Appendix.....	43

## List of Figures

Figure 1: Academic year dropout trend.....	12
Figure 2: Top 20 demographic location with high mean percentage of dropouts .....	13
Figure 3: Top 20 locations with students who transfer to another HE .....	14
Figure 4: Top 20 locations with low mean dropouts rates .....	15
Figure 5: Student dropouts percentage by top 25 HE .....	16
Figure 6: Year 2015-16 .....	17
Figure 7: Year 2016-17 .....	18
Figure 8: Year 2017-18 .....	19
Figure 9: year 2018-19.....	19
Figure 10: Year 2019-20 .....	20
Figure 11: Oxbridge dropout average percentage 2014-20.....	21
Figure 12: Russell Group dropout percentage by years .....	23
Figure 13: Data preparation .....	26
Figure 14: Assessment and feedback .....	26
Figure 15: Academic support .....	27
Figure 16: Learning Resources .....	28
Figure 17: Student Voices .....	29
Figure 18: Overall Dissatisfaction Responses .....	32
Figure 19: Survey Dissatisfaction correlation.....	34

Figure 20: Anglia university dropout and correlation .....	37
Figure 21: Manchester Metropolitan University dropout and correlation .....	38
Figure 22: University of central Lancashire dropout and correlation .....	38

## List of Tables

Table 1: Overall student dropout statistics .....	11
Table 2: Repeated Universities in each year .....	21
Table 3: Oxbridge Universities dropouts percentage .....	22
Table 4: Russell group universities .....	22
Table 5: Overall dropout statistics in each year Russell group .....	23
Table 6: scale 3 SSAT and DISSAT .....	27
Table 7: scale 4 SSAT and DISSAT .....	27
Table 8: scale 6 SSAT and DISSAT .....	28
Table 9: scale 8 SSAT and DISSAT .....	29
Table 10: All scales and year statistics .....	30
Table 11: Descriptive statistics.....	31

## List of Abbreviations

1. HESA: Higher Education statistical agency
2. NSS: National Student Survey
3. SSAT: Student satisfaction score
4. DISSAT: Student Dissatisfaction score

## Introduction

In the realm of higher education, the pursuit of academic excellence often encounters challenges that lead to student attrition, a critical concern for educational institutions and stakeholders. This dissertation embarks on a thorough exploration of student dropout within the context of higher education in the United Kingdom. By closely investigating the factors contributing to student attrition, this study aims to uncover insights that can pave the way for more effective strategies to retain students and create a nurturing academic environment.

Focusing on the higher education landscape in the United Kingdom, this research aims to offer a comprehensive view of challenges and potential solutions. By uncovering the factors behind student attrition, this study intends to provide insights for institutions, policymakers, educators, and stakeholders to develop targeted strategies for enhancing student success and retention. Additionally, this research envisions the creation of a supportive academic environment that addresses students' needs and encourages resilience in the face of challenges.

A key aspect of this study involves survey analysis, where the voices of students and their experiences will be captured. This approach not only adds depth to the research but also provides a holistic understanding of the factors contributing to student attrition. The synthesis of these advanced techniques and qualitative insights aims to provide a comprehensive view of the complex dynamics at play. By utilizing machine learning predictive analysis, this study seeks to identify patterns that can foresee student dropout, offering insights into potential intervention strategies.

Furthermore, this research delves into the correlation between socioeconomic factors, such as unemployment rate and cost of living rate, and student attrition. By examining how these specific socioeconomic variables interact with dropout rates, this study aims to shed light on the socio-demographic influences that may play a role in students' decisions to continue or discontinue their education. The inclusion of unemployment rate and cost of living rate as distinct socioeconomic factors adds depth to the analysis, as they are recognized as significant determinants of students' financial well-being and overall ability to sustain their education.

## Literature Review

### Introduction

This literature review explores how students can succeed in higher education and stay in universities by looking at different reasons why some students drop out and using predictive methods to understand these factors. The review focuses on studies about the importance of advisors in helping students, how online courses work, why teachers sometimes leave, and the impact of students quitting on the economy. These studies, done by different experts, help us understand how to support students better and make education more effective.

### 2.1 Literature Review on Student Dropout in Higher Education.

In the landscape of higher education, the pressing need to uncover the intricate drivers of student dropout while employing predictive models for fortified retention strategies takes centre stage. A constellation of studies has illuminated key insights in these crucial domains. Notably, (Aulck, Velagapudi, Blumenstock, & West, 2016) researched extensive data and diverse machine learning methods like logistic regression, k-nearest neighbours, and random forests. Through these tools, the study pinpointed pivotal predictors, offering pathways for enhancing student persistence. In the EdTech sphere, (Paul & Rashmi, 2022) showcased

machine learning prowess to predict student contentment and attrition. With adept data preprocessing and algorithms like K-Nearest Neighbour and Support Vector Machine, personalized interventions emerged, spotlighting the potential of technology-driven tailored strategies. Transitioning to academia, (Grillo & Leist, 2013) explored academic support's role, employing a comprehensive database to delve into tutoring, learning assistance, and supplemental instruction's impact on outcomes. This emphasized tailored academic support's potency in boosting student persistence. Affirming support's value, (Coates & Ransom, 2011) revealed the symbiotic link between effective student services and retention, leveraging insights from the Australasian Survey of Student Engagement. This study underscored the correlation between support efficacy and sustained student commitment. (Li & Carroll, 2017) embraced a comprehensive approach, employing a theoretical model rooted in job satisfaction literature. This guided exploration of multifaceted factors governing student satisfaction and dropout. Moreover, the study extended its purview to equity groups, gauging the differential impact across diverse student populations. Larsen, Sommersel, et al. (2013) embarked on a comprehensive review of research into student dropout. Fuelled by The Swiss Council for Educational Research, their mission in 2012/2013 was to understand what causes students to leave and what universities can do about it. Using a software tool, they mapped out the research landscape, drawing insights from databases, research hubs, and key journals (Larsen, Sommersel, et al., 2013, p. 14).

Summarizing these findings, the studies collectively underscore the significance of predictive modelling, personalized interventions, and tailored academic support in enhancing student retention in higher education. To build upon this foundation, a potential research question could be: "How can a holistic approach integrating predictive analytics, personalized interventions, and academic support effectively reduce student dropout rates across diverse higher education environments?"

## 2.2 Literature Review on Factors Influencing Student Dropout and Predictive Analysis: Enhancing Student Retention Strategies in Higher Education.

(Lovenoor Aulck, Nishant Velagapudi 2016) delved into dropout prediction using a range of machine learning models. Their study prompts us to consider how these models can be further fine-tuned and optimized to enhance their predictive accuracy. Exploring the potential of machine learning in addressing evolving learning contexts, (Rinika Paul and MR. Rashmi 2022) focused on predicting student satisfaction and churn. Their work encourages us to adapt and apply similar algorithms to cater to changing educational landscapes. Michael C. Grillo and Cathy W. Leist 2013) cantered their research on the role of academic support in student retention and graduation. This opens avenues for further investigation into different types of support systems and how they contribute to student success. (Hamish Coates and Laurie Ransom 2011) emphasized the pivotal role of effective student support services in promoting retention. This highlights the importance of incorporating student perspectives to enhance support offerings. (Ian W. Li and David R. Carroll 2017) took a comprehensive approach by drawing on a theoretical model from job satisfaction literature to explore factors influencing student satisfaction and dropout. Their study encourages us to delve deeper into understanding diverse student populations and their unique experiences.

Collectively highlight the importance of refining machine learning models for dropout prediction, adapting algorithms to changing educational contexts, investigating various support systems, incorporating student perspectives into support services, and understanding diverse student experiences. Moving forward, these insights prompt the formulation of research questions that explore the

effectiveness of optimized machine learning models in predicting dropout, the impact of evolving educational landscapes on student satisfaction, the types of support systems that resonate with different student populations, and the interplay between student perspectives and retention strategies.

### 2.3 Literature Review on Academic Advising, Online Course Effectiveness, Teacher Turnover, and Economic Effects of Student Dropouts

Exploring the critical dimensions of academic success, a series of six studies provide valuable insights into the role of academic advising, online course effectiveness, teacher turnover, and the economic effects of student dropouts. (Young-Jones, Burt, Dixon, & Hawthorne, 2013) embarked on a study evaluating academic advising's impact on student success, employing a survey-based methodology to uncover factors such as advisor accountability, empowerment, and student self-efficacy, underscoring the substantial relationship between advising and student success. The effectiveness of online graduate courses was dissected by (Rovai & Barnum, 2012), who utilized ex post facto and correlational designs, employing perceived learning measurements and interactions records to discern the nuanced factors contributing to online course effectiveness. Investigating student connections and retention efforts, (Hoyt, 2021) calculated retention rates based on student interactions with campus programs, applying binary logistic regression to explore the relationship between program engagement and retention, revealing the multifaceted role of academic support services. Delving into the intricate challenge of teacher turnover, (Menzies, 2023) conducted a comprehensive study utilizing diverse research methods, emphasizing the factors contributing to teacher attrition and its consequences on educational outcomes. A comparative study by (Latif, Choudhary, & Hammayun, 2015) assessed the economic implications of student dropouts, employing a literature review and comparative analysis to uncover global patterns and causes of dropout rates. Additionally, (Aasma, Ali, & Asad, 2015) contributed to this understanding by conducting a comparative analysis of student dropout causes and rates across different countries, with a focus on the Subcontinent region, suggesting interventions to combat these challenges. Collectively, these studies illuminate the multifaceted landscape of academic advising, online course effectiveness, teacher turnover, and student dropout effects, providing a framework for institutions and policymakers to navigate these critical facets of education.

In their study titled "Cause Analysis of Students' Dropout Rate in Higher Education Study Program," (Liga Paura and Irina Arkhipova 2014) embarked on an in-depth exploration of the factors contributing to student dropout rates in higher education institutions. The authors focused their research on engineering study programs at the Latvia University of Agriculture. They utilized a robust data set comprising 677 full-time students from five faculties of engineering science who were enrolled during the 2011-2012 academic year. The data encompassed a range of variables, including students' study duration, gender, secondary school scores, the priority of the program to study, and the source of finance for their education. To analyse these factors, Paura and Arkhipova employed the Proportional hazard model of Survival analysis.

A more comprehensive perspective emerges academic advising, online course effectiveness, teacher turnover, and student dropouts are interconnected dimensions influencing academic success. While the current compilation highlights these facets, future research might delve into the interplay between these factors and their collective impact on educational outcomes. By scrutinizing potential research questions, educators, institutions, and policymakers can harness these findings to inform strategies that foster holistic student achievement.



## Problem Description

The primary objective of this research is to investigate and delineate significant factors influencing student dropout rates in higher education. By meticulously examining a variety of contributing variables through advanced statistical analyses, this study aims to convert these factors into discernible patterns and predictive insights. The focal point of this investigation revolves around a series of pivotal research questions, each designed to shed light on the intricate web of elements impacting student attrition in educational institutions. Through this comprehensive exploration, we seek to enhance our understanding of the complex interplay between demographics, curriculum dynamics, student satisfaction, and sociodemographic factors, ultimately paving the way for more effective and tailored student retention strategies.

**To what extent can demographic factors serve as predictors of student dropout in the realm of higher education? (Lovenoor Aulck, Nishant Velagapudi, 2016)**

This investigation seeks to ascertain the extent to which demographic variables can serve as predictors of student dropout within higher education. The study aims to unravel the relationships between demographic factors such as age, gender, ethnicity, socioeconomic status, and educational background, and their predictive capabilities in anticipating occurrences of student attrition. By examining these relationships, valuable insights will be generated, equipping academic administrators with a deeper understanding of the factors influencing student dropout. Consequently, this understanding will enable the refinement of strategies aimed at enhancing student retention. The investigation's outcomes are poised to facilitate the identification of at-risk student groups and the development of targeted interventions that address specific demographic-related challenges, thereby fostering a more supportive and inclusive educational environment.

**Is there a tangible correlation between student contentment and the occurrence of dropout within the realm of higher education? (Rovai & Barnum, 2012)**

This research inquiry delves into the presence of a tangible correlation between student contentment and the incidence of dropout in the higher education context. The central objective of this investigation is to uncover whether students' overall satisfaction with their educational experience is linked to their decision to drop out. By assessing factors such as course satisfaction, campus engagement, faculty interactions, and the availability of support services, the study aims to identify whether higher levels of contentment are associated with lower dropout rates. Such insights have the potential to inform institutions on the critical role of student well-being and satisfaction in shaping retention outcomes. Should a strong correlation be established, academic administrators can prioritize efforts to enhance student satisfaction, possibly leading to a reduction in dropout rates. Additionally, the findings could underscore the need for comprehensive student support services that address not only academic needs but also the broader aspects of student life, fostering an environment conducive to both academic success and personal fulfilment.

What are the sociodemographic factors that exert influence over attrition rates within the realm of higher education? (Larsen, Sommersel, et al., 2013, p. 14).

This inquiry aims to identify the sociodemographic factors that exert influence over attrition rates in the realm of higher education. The central focus of this investigation is to comprehensively examine the array of sociodemographic variables that may contribute to students' decisions to leave their educational pursuits. Through an analysis of factors such as age, gender, ethnicity, socioeconomic status, and first-generation college status, the study seeks to uncover patterns and relationships that could shed light on specific vulnerable groups. By pinpointing which sociodemographic characteristics are more strongly associated with higher attrition rates, this research can assist higher education institutions in designing targeted strategies and interventions that address the unique challenges faced by different student populations. The results of this investigation hold the potential to foster a more inclusive, supportive, and equitable higher education environment that prioritizes the success and retention of all students, regardless of their sociodemographic backgrounds.

Collectively, these research questions delve into various facets of student dropout in higher education, shedding light on demographic, curricular, contentment-related, and sociodemographic factors that contribute to attrition. The insights from these inquiries hold the potential to inform effective strategies and interventions for improving student retention and fostering an inclusive and supportive learning environment.

The demographic dropout analysis provides a comprehensive understanding of student attrition patterns in higher education, shedding light on factors contributing to students leaving their academic pursuits. A crucial tool in this analysis is HESA data, which offers invaluable insights into student outcomes and institutional performance within the higher education sector. By examining metrics such as continuation rates, transfers, and sector benchmarks, HESA data enables institutions and policymakers to make informed decisions aimed at improving student retention and overall educational quality.

The definition of HESA data explains the meaning of important measures which are in the provided dataset.

## Description of Higher Education Statistic Agency

Below is the definition of [HESA data](#) which is used to combined to analysis dropouts for multiple years.

1. Higher Education Statistics Agency data, provides information about student outcomes and institutional performance in higher education. Key metrics include:
2. Percentage who continues or qualify at same HEP (%): Indicates the percentage of students staying at the same institution or successfully advancing to the next level of their program.
3. Adjusted sectors continue or qualify (%): Compares an institution's performance in this regard to the sector average, revealing if it's better or worse than the norm.
4. Number who transfer to other UK HEP: Counts students who switch to different UK higher education providers.
5. Percentage who transfers to other UK HEP (%): Shows the percentage of students transferring to other UK institutions, reflecting mobility patterns.

6. Adjusted sector transfer to other UK HEP (%): Compares an institution's transfer rate to the sector-wide average.
7. These metrics offer insights into student mobility, academic progression, and institutional performance, allowing comparisons against sector norms.

The [data\(Merged data\)](#) spanning six years, from 2014 to 2020, encompasses information from 240 universities across the country, providing a holistic view of student retention challenges. What stands out prominently in this analysis is the impact of geographic location on student attrition. Now let's understand the demographics student dropout rate.

## Demographic Student Drop Analysis

The analysis conducted spans six years and encompasses data from 240 universities across England. Over this period, an average of approximately 356,621 full-time entrants were recorded annually, reflecting a dynamic higher education landscape. Notably, the continuation/qualification rate saw an average of about 227.86%, highlighting the variability in students continuing or qualifying at the same higher education provider. Additionally, an average of around 10,411 students transferred to other UK higher education providers, signalling a subset of mobility within the student population. The dropout rate, calculated as a percentage of students no longer in higher education, averaged around 0.12%, with institutions showing variation in their retention strategies. These insights underscore the diverse nature of student retention patterns, motivating further exploration into factors influencing retention rates and initiatives aimed at fostering student success.

*Table 1: Overall student dropout statistics*

Year	Total full-time entrants	Number who continue or qualify at same HEP	Percentage who continues or qualify at same HEP (%)	Number who transfer to other UK HEP	Percentage who transfers to other UK HEP (%)	Number no longer in HE	Benchmark (%)	Standard deviation (%)
2014-15	337750	302740	188.35	13736	8930	375	1348.2	1289.2
2015-16	353405	314735	249.88	15051.7	10170	448.6	1501.4	1499.1
2016-17	358420	318380	231.79	14721.2	10850	465.7	1539	1512
2017-18	359360	318095	244.58	15182.2	10880	483.1	1659.6	1634.9
2018-19	359505	318340	254.83	15261.3	11245	504	1625.3	1634.1
2019-20	371285	334720	197.7	14881.9	10395	446.9	1243.2	1271.1

Over a span of six years, encompassing data from 240 universities across England, a comprehensive analysis of student retention and attrition patterns was conducted. The average annual intake of full-time students stood at approximately 356,621, with a slight variance as indicated by a standard deviation of around 10,959. Amidst this, a dynamic continuum emerged, with an average of 317,835 students successfully continuing or qualifying within the same higher education institution. Concurrently, an average of 14,806 students opted for mobility, transferring to other UK-based institutions, showcasing

a subset of academic mobility. Notably, an average of about 454 students decided to discontinue their higher education journey each year, reflecting a nuanced picture. The calculated dropout rate, elucidating the proportion of students discontinuing in relation to total entrants, was characterized by an average of 0.12%, fluctuating within a narrow band from 0.11% to 0.14%. Concomitantly, institutional contributions to student retention, benchmarked at around 1,486.12%, interwove with the percentage of students who continued or qualified. Unveiling variance, a standard deviation of roughly 1,473.40% underscored the diverse retention landscapes across institutions and years. Amidst this array of observations, an average of approximately 523.57 higher education providers contributed to a contextual backdrop, where student retention and engagement strategies were implemented with a consciousness of this dynamic variation. These insights collectively highlight the intricate interplay of student persistence, mobility, and attrition, underscoring the need for tailored retention initiatives and a deeper exploration into the contributing factors across the higher education ecosystem.

Across the span of six years from 2014 to 2020, the transfer rate to other UK higher education providers displayed a relatively steady trend, with student numbers fluctuating between approximately 14,736 and 15,261. Meanwhile, the count of students no longer in higher education demonstrated a stable pattern, ranging between approximately 26,060 and 29,135. These trends shed light on potential patterns of student mobility and attrition, prompting considerations for institutional retention efforts.

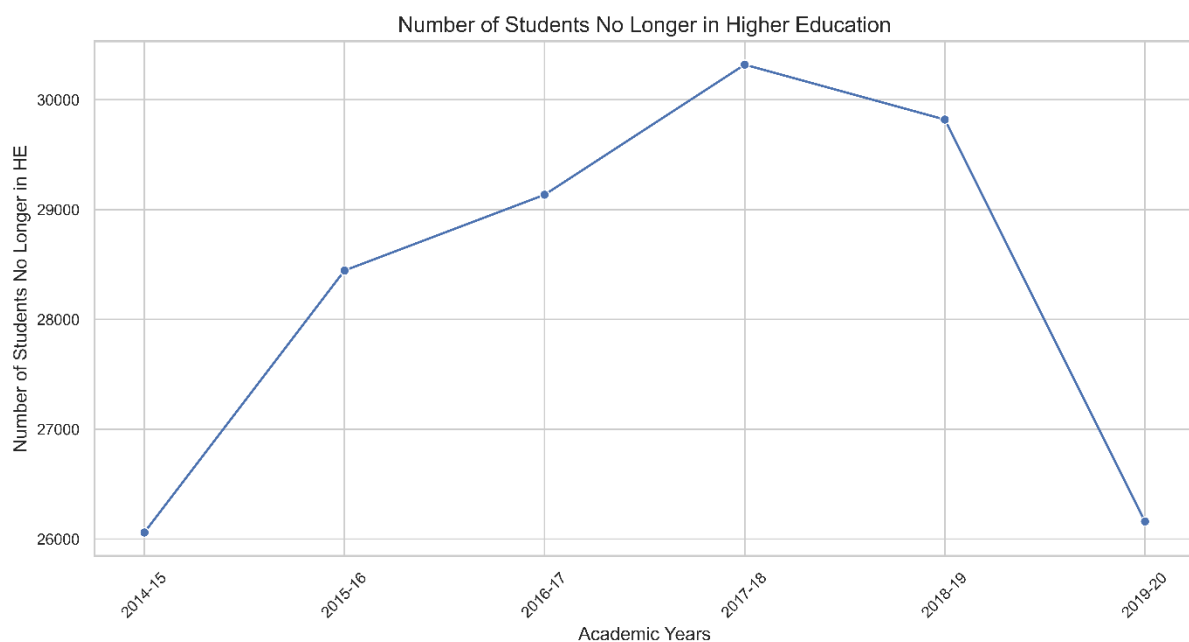


Figure 1: Academic year dropout trend

Across the academic years of 2014-2020, an insightful pattern emerges when analysing universities that exceeded the average number of students no longer in higher education for a given year. Notably, in specific years, certain universities surpassed the mean value of around 29,000 students who discontinued their studies annually. This observation suggests a potential need for deeper investigation into the contributing factors leading to higher attrition rates within these institutions. By identifying and addressing these factors, universities can strategically tailor interventions to enhance student persistence and success. This insight underscores the significance of context-aware retention strategies and the value of personalized support mechanisms to mitigate attrition challenges.

## To what extent can demographic factors serve as predictors of student dropout in the realm of higher education ?

The data we have merged along with the year and location enables the understanding of No longer students which are dropped out students from each university's location wise. Figure 3.1 explains the location wise student dropout rates average of over six years since 2014 to 2020 of top 20 UK locations.

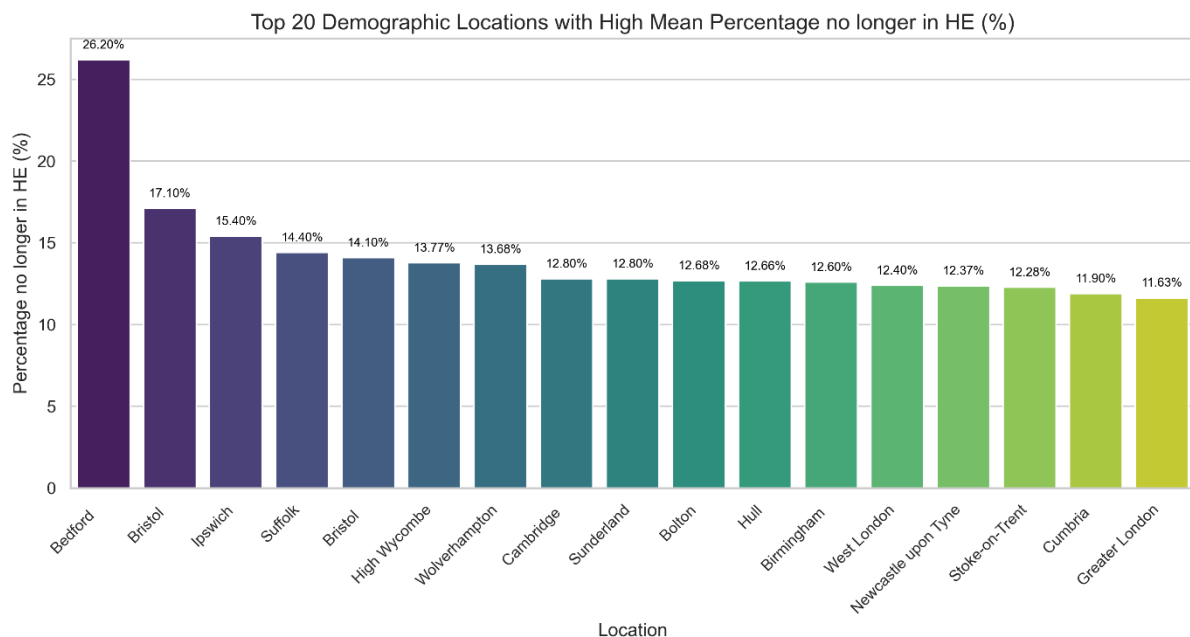


Figure 2: Top 20 demographic location with high mean percentage of dropouts

The meaningful insights we get from the location wise student Dropout. here are some important insights which understand by the figure 2

**Variation in Dropout Percentages:** The “Percentage no longer in HE (%)” indicates the percentage of students who have dropped out of higher education. These percentages vary across different locations, ranging from around 26% to 11.9%. This indicates that demographic factors might play a role in student retention as well as the no longer student percentage.

**Transfer Rates:** The “Percentage who transfer to other UK HEP (%)” shows the percentage of students who transfer to other UK higher education providers. These percentages range from around 7.8% to 0%, suggesting differences in student mobility among locations.

**High Dropout Locations:** Locations like Bedford and Luton have relatively high dropout percentages (26.2% and 22.2% respectively), potentially indicating challenges in retaining students.

**Transfer Patterns:** Locations such as East Anglia and Cambridge, England, have lower transfer percentages (3.2% and 0% respectively), suggesting that students in these areas are less likely to transfer to other institutions.

Urban Centres and Dropouts: Urban centres like Bristol and Birmingham show varying dropout percentages, indicating that the size and nature of urban environments may impact student persistence.

Potential Predictive Value: The differences in dropout and transfer percentages across locations suggest that demographic factors associated with these areas may contribute to student outcomes. However, other factors such as academic support, program quality, and personal circumstances also play a role.

Some students who are churned from the university had enrolled in other universities and understand their percentage also gets better understanding of actual students dropout and transferred student churning. Figure 2.2 is the representation of the six-year (2014 – 2020) average of no longer percentage and the transfer to other UK students.

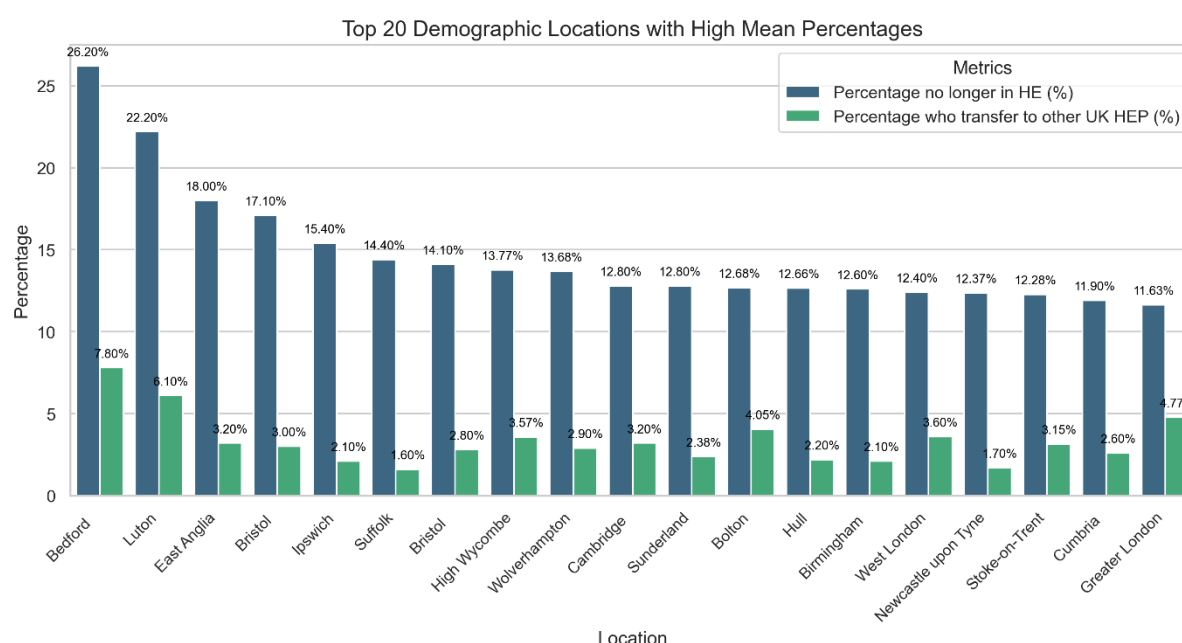


Figure 3: Top 20 locations with students who transfer to another HE

The meaningful insights from the top 20 demographics locations with Mean percentage.

Location Variation: Student dropout rates vary by location. Locations such as Bedford and Luton have relatively higher percentages of students leaving higher education (up to 26.2% and 22.2% respectively). In contrast, Birmingham, England, has a comparatively lower dropout rate of 11.9%. certain locations stand out as hotspots of student departures. Bedford and Luton draw attention with dropout rates of 26.2% and 22.2% respectively, reflecting the challenges these areas face in retaining students. In contrast, the determination of students in Birmingham, England, is evident as the city boasts a relatively lower dropout rate of 11.9%. While some students choose to leave, others embark on a journey of transfer to other UK higher education institutions. Cambridge showcases a unique trend with a dropout rate of 13.2% and no recorded transfers. On the other hand, Ipswich records a 15.4% dropout rate alongside a 2.1% transfer rate, underscoring the complexity of student mobility. The concept of continuation within the same institution also comes into focus. Locations like Cambridge and Bristol demonstrate a commitment to their institutions, with percentages exceeding 4.5% and 0.7% respectively. This finding suggests a strong bond between students and their educational establishments. Within

this geographical tapestry, some locations emerge as places of lower dropout rates, providing intriguing insights into potential factors that contribute to student persistence. Suffolk and Wolverhampton, with dropout rates of 14.4% and 13.675% respectively, hint at the influence of local dynamics on student choices

**Consideration of Other Demographics:** While location plays a role in dropout rates, other demographic factors such as age, gender, and socioeconomic background are not provided. These factors could influence dropout rates and should be considered for a comprehensive analysis.

The diverse demographic landscape across different locations plays a pivotal role in predicting student dropout rates. The data highlights considerable variation in dropout percentages, suggesting that geographical factors intersect with educational outcomes. Locations like Bedford and Luton exhibit higher dropout rates, which might stem from unique socio-economic challenges or insufficient support systems. In contrast, areas such as Birmingham boast lower dropout rates, potentially indicating a more conducive learning environment or stronger community engagement. These findings underscore the importance of considering location-specific dynamics when designing interventions to address student dropout, as local factors can significantly impact a student's educational journey and decision to remain in higher education.

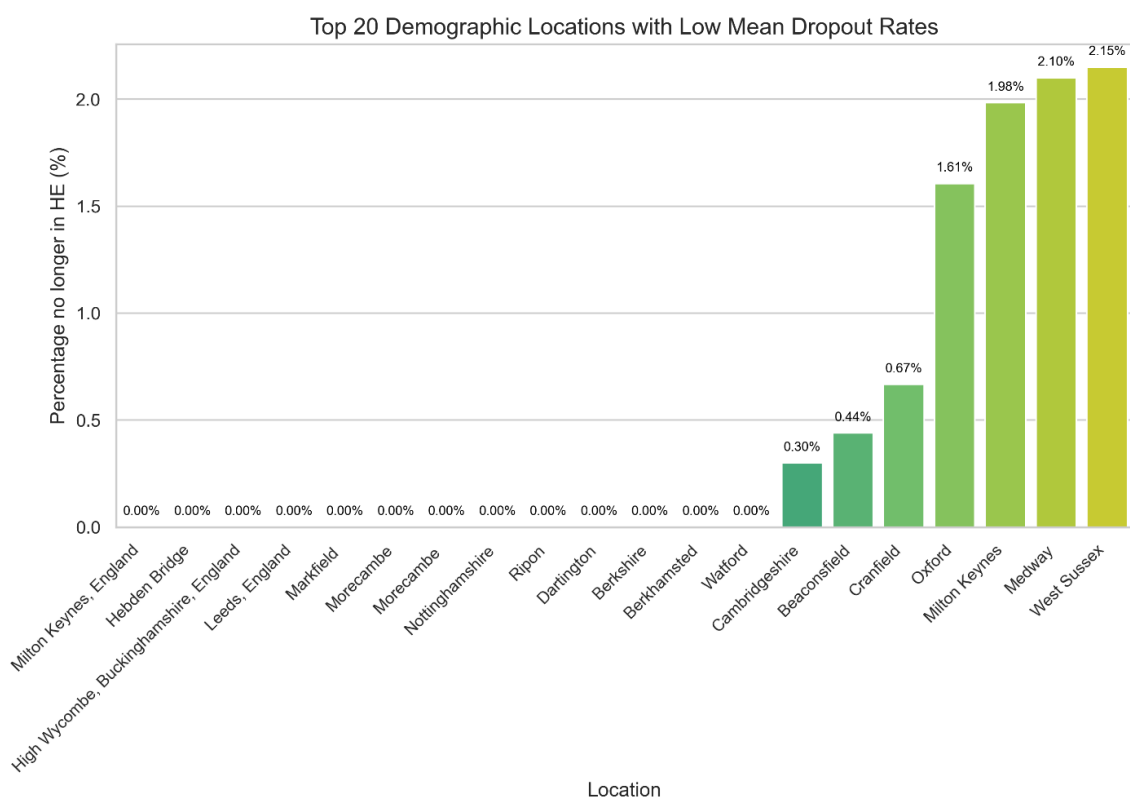


Figure 4: Top 20 locations with low mean dropouts rates

We can observe some interesting insights. The locations with the lowest dropout rates, such as Milton Keynes, Hebden Bridge, and High Wycombe, England, have a dropout rate of 0.00%, indicating that very few students are leaving higher education institutions in these areas. This could suggest that these locations might provide a supportive environment or have programs in place that contribute to student retention.



In contrast, when observing the locations with higher dropout rates, like Bedford, Luton, and East Anglia, we can see dropout rates ranging from 18.0% to 26.2%. These higher rates might be indicative of various factors influencing student retention, such as the availability of academic resources, support systems, or even economic conditions in the respective regions.

## Annually Each University analysis

Understanding each university distribution dropout and the transfer of students is the important insight and for that Year by year, our analysis focuses on understanding the distribution of students who are no longer enrolled in higher education and those who have chosen to transfer to other institutions. This examination provides valuable insights into the dynamics of student mobility and institutional retention strategies.

### Year 2014 -15

The figure 5 below illustrates the distribution of students who were no longer enrolled in higher education during the academic year 2014-15 across various universities.

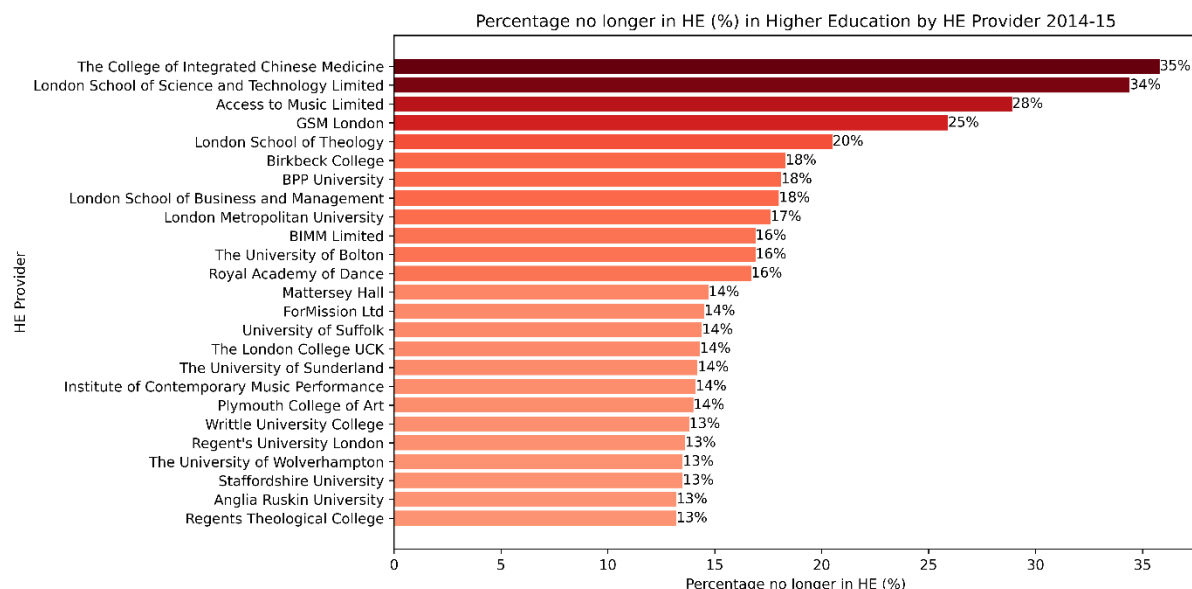


Figure 5: Student dropouts percentage by top 25 HE

In the data for the academic year 2014-15, we can observe significant variations in the percentage of students no longer in higher education across different universities. Notably, Kingston University and London Metropolitan University stand out with relatively high percentages of students no longer in higher education, at 38.89% and 30.82%, respectively. Conversely, De Montfort University and The University of Brighton have notably lower percentages, at 70.09% and 70.18%, indicating better retention rates.

These percentages specifically capture the portion of students who did not continue their higher education journeys at these institutions during that academic year. Such data provides valuable insights into the effectiveness of institutional retention strategies and the factors that influence students' decisions to leave higher education.



## For year 2015-16

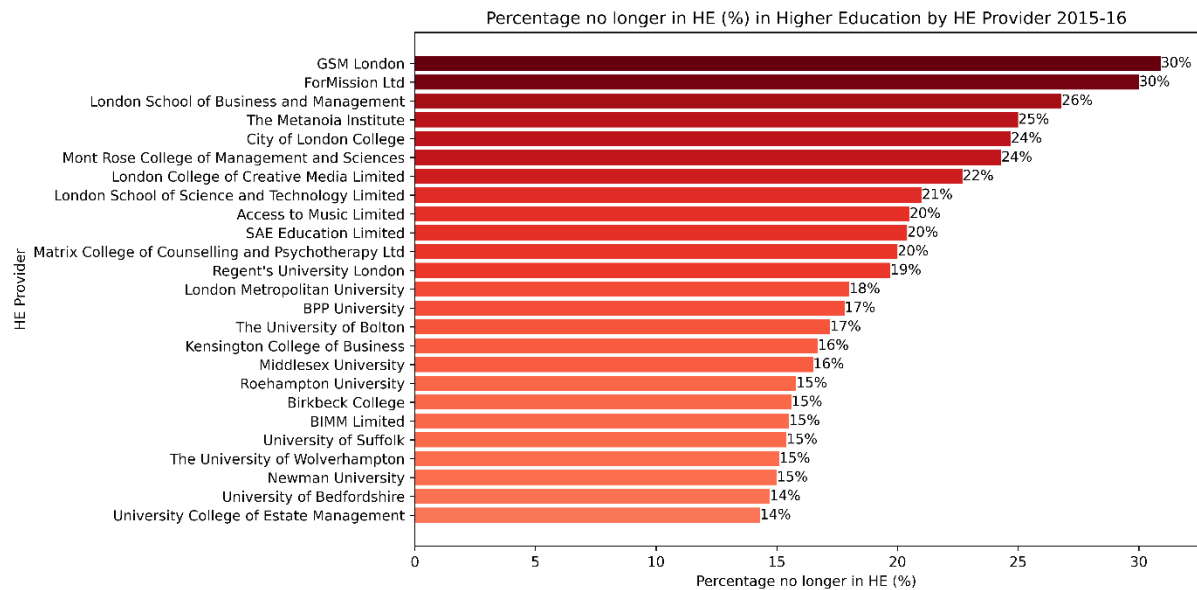


Figure 6: Year 2015-16

In the dataset for the 2015-16 academic year, there are noteworthy discrepancies in the percentages of students who discontinued their higher education across various universities. For example, University College of Estate Management and University of Bedfordshire demonstrated relatively lower percentages, standing at 14.18% and 14.09%, respectively. These figures suggest that these institutions have been more successful in retaining their students.

Conversely, London Metropolitan University and GSM London reported higher percentages of students no longer pursuing higher education, at 18.0% and 30.9%, respectively. These percentages specifically reflect the students who did not continue their academic journeys at these institutions during that particular academic year. This data provides valuable insights into the efficacy of institutional retention strategies and the underlying factors contributing to student attrition.

For year 2016-17

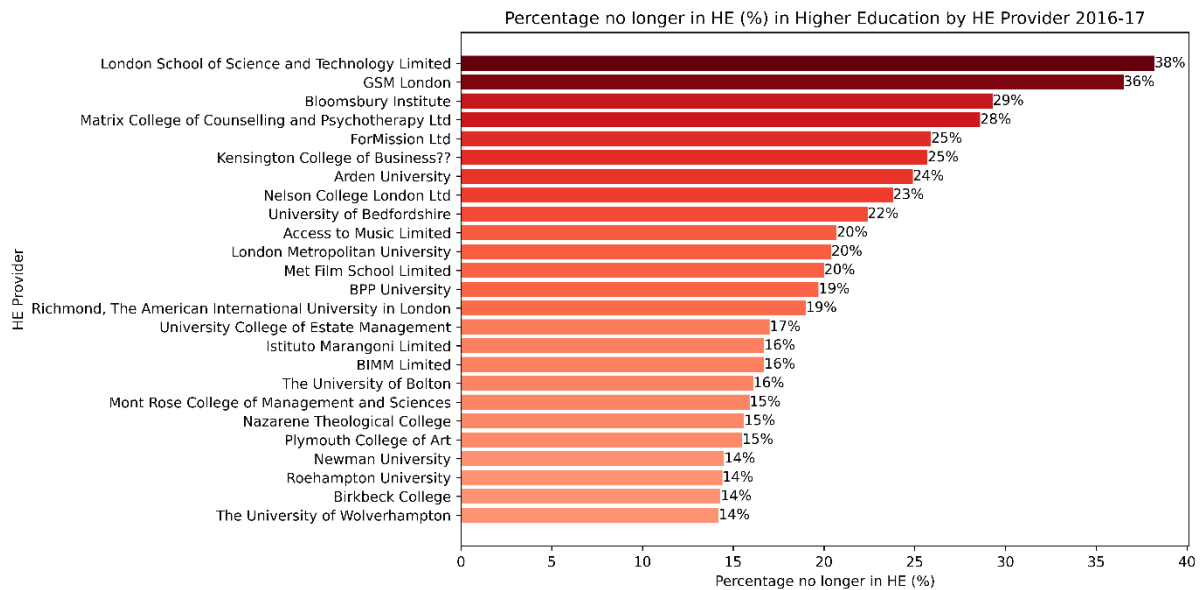


Figure 7: Year 2016-17

In the data for the academic year 2016-17, notable disparities in the percentage of students no longer in higher education are evident among various universities. For instance, University College of Estate Management and University of Bedfordshire reported relatively lower percentages, standing at 4.49% and 0.60%, respectively, which suggests more successful student retention efforts.

On the other hand, London Metropolitan University and GSM London exhibited higher percentages of students no longer in higher education, at 20.4% and 36.5%, respectively. These percentages represent students who did not continue their higher education journeys at these institutions during that academic year. This data provides valuable insights into the effectiveness of institutional retention strategies and factors influencing student attrition.

## For year 2017-18

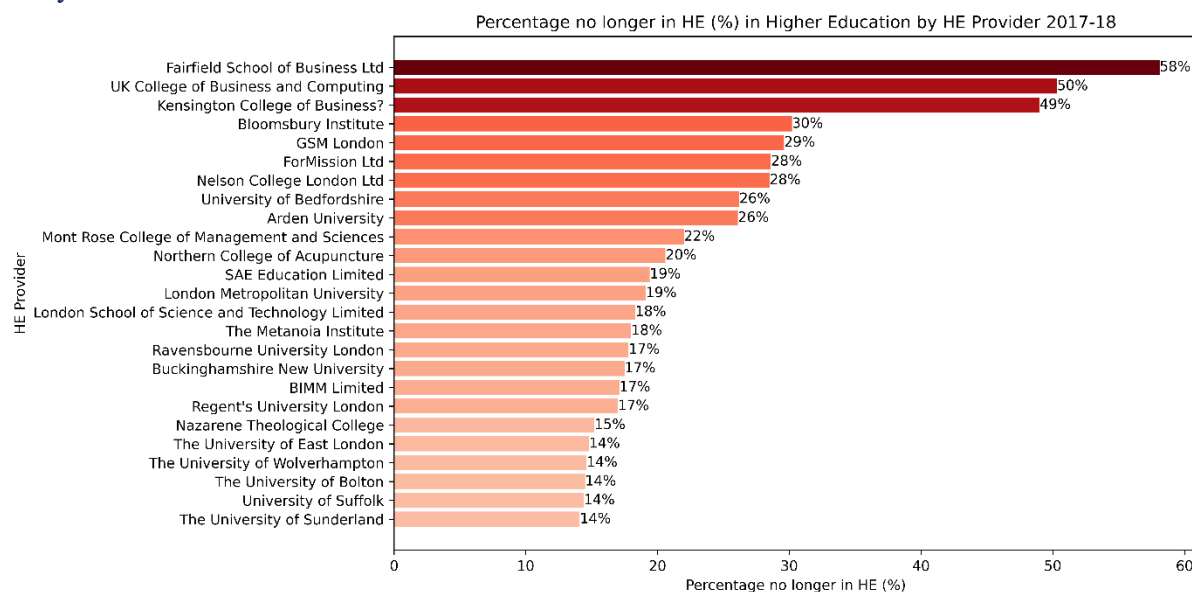


Figure 8: Year 2017-18

In the academic year 2017-18 data, there are noticeable disparities in the percentage of students no longer enrolled in higher education across various universities. The University of Sunderland and University of Suffolk recorded relatively lower percentages, at 14.1% and 14.4%, respectively, indicating effective student retention efforts.

Conversely, fairfield school of business and UK college Business and computing reported higher percentages of students no longer in higher education, at 58% and 50%, respectively. These figures represent students who did not continue their higher education journeys at these institutions during that academic year. This data continues to offer valuable insights into the effectiveness of institutional retention strategies and factors influencing student attrition.

## For year 2018-19

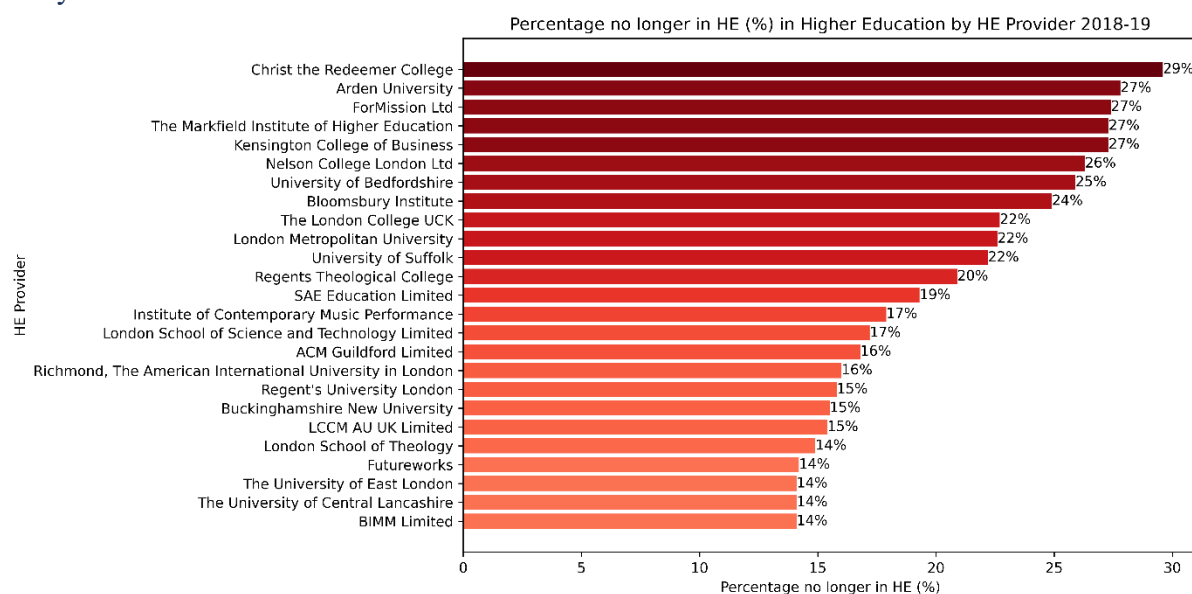


Figure 9: year 2018-19

In the academic year 2018-19 dataset, we uncover intriguing insights into student retention across various universities. These numbers shed light on the effectiveness of retention strategies and reveal some compelling stories within higher education.

BIMM Limited impressively retains a substantial 85.4% of its students, reflecting a well-crafted environment for student success. Conversely, London Metropolitan University faces challenges with a 0.71% lower retention rate, possibly indicating areas for improvement in student support and engagement.

The data also spotlights the diversity in retention rates across regions. Institutions in Greater London, for instance, contend with a higher churn rate of 22.6%, compared to East Anglia's 22.2%. This regional variation could be influenced by various factors, including local demographics, employment opportunities, or the availability of support services.

## For year 2019-20

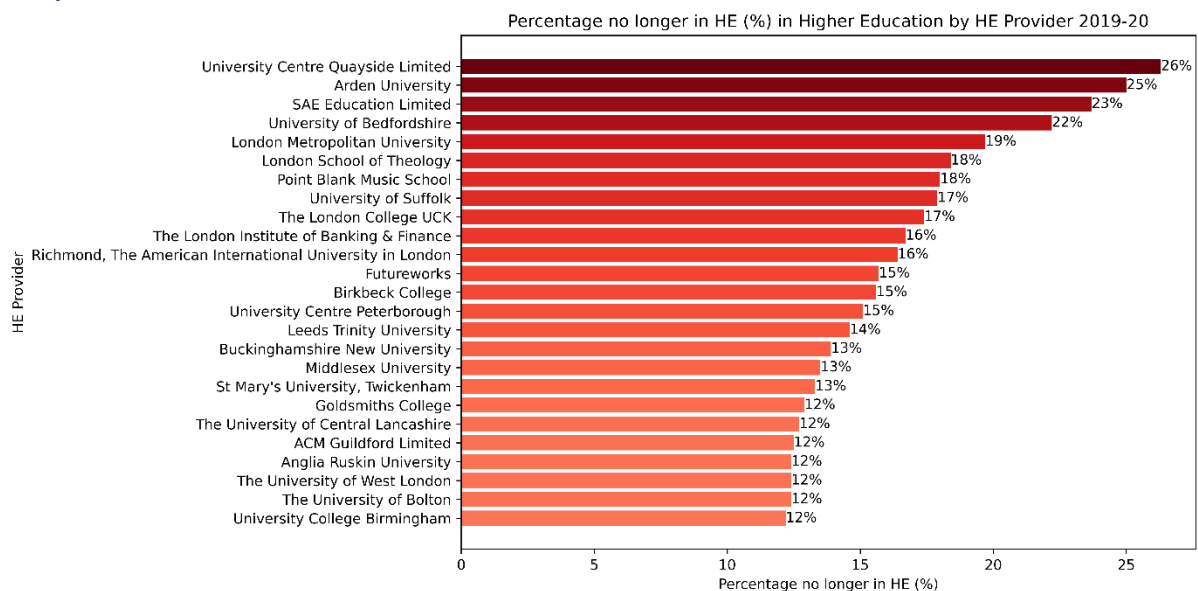


Figure 10: Year 2019-20

In the 2019-20 academic year, student retention rates among various universities in England continued to reveal intriguing patterns and challenges, with the added complexity of the COVID-19 pandemic impacting higher education institutions.

University College Birmingham stands out with an impressive retention rate of 1.01%, indicating a strong commitment to student success. In contrast, the University of Bedfordshire faces a retention challenge with a rate of 0.53%, suggesting the need for enhanced support strategies.

The impact of the COVID-19 pandemic is evident in the data. Some institutions, like Arden University, managed to maintain an 80.0% retention rate despite the challenges posed by the pandemic. Others, such as the London Institute of Banking & Finance, saw their rates dip, highlighting the varying responses and adaptability of universities during these unprecedented times.

we can observe a pattern where certain universities repetitively appear over a span of six years, from 2014 to 2020. So these universities are following:

Table 2: Repeated Universities in each year

Rank	HE PROVIDER	Repetition	Frequency
1	Anglia Ruskin University	6	3870
2	The Manchester Metropolitan University	6	2790
3	The University of Central Lancashire	6	1440
4	The University of Wolverhampton	5	1305
5	Leeds Beckett University	5	1700
6	London Metropolitan University	5	3035
7	Birmingham City University	5	3425
8	University of Bedfordshire	4	1650
9	GSM London	4	2845
10	De Montfort University	3	1850

In above table the repetition column explains the count of universities in year 2014 to 2020, the rank 1 university has count 6 which meaning that, that university has come in each year from 2014 to 2020 and the sum of no longer 3870 is the number of students that university has lost and churned within 6 years of time span.

### Understanding the Russell Group and Oxbridge universities dropout rates.

In order to understand the Oxford and Cambridge University dropout rates we have filtered the data and using group by method we can see the combined average percentage of both universities in each year. And we can observe both universities combined average dropout percentage is too low.

Figure 11 explains the aggregated average percentage of both universities, the band Oxbridge dropout percentage exhibits the prestigious university relatively low student attrition rate for each year.

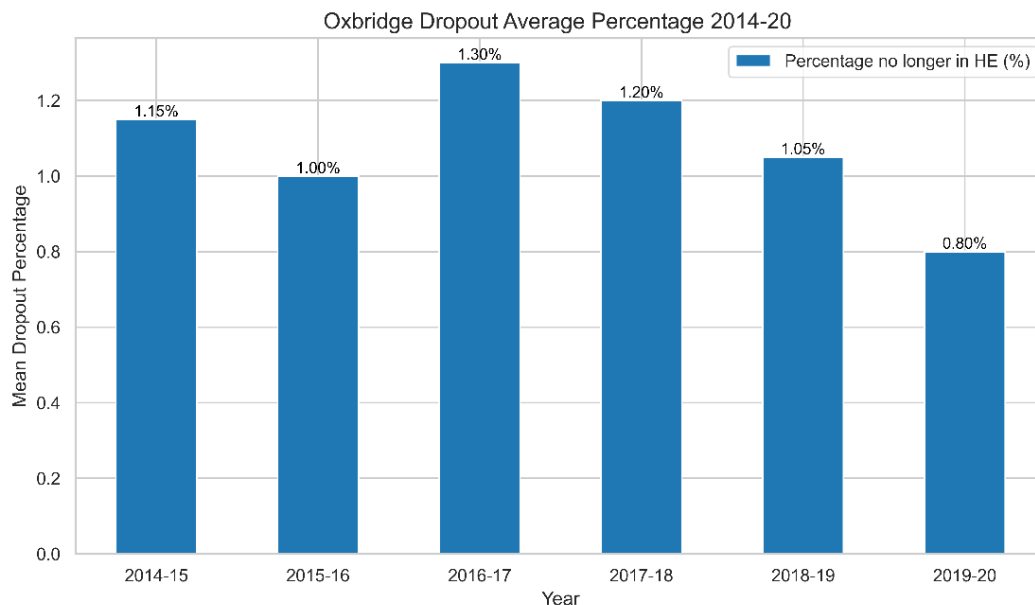


Figure 11: Oxbridge dropout average percentage 2014-20

Let understand further deep analysis of Oxbridge band. Below statistical table represent the six year's descriptive statistics of Oxbridge Universities. The count 6 means the six years 2014-20 and aggregated mean, standard deviation, etc.

Table 3: Oxbridge Universities dropouts percentage

Oxbridge Universities	Percentage no longer in HE (%)							
	count	mean	std	min	25%	50%	75%	max
The University of Cambridge	6	0.9667	0.175119	0.7	0.9	0.95	1.075	1.2
The University of Oxford	6	1.2	0.1788854	0.9	1.125	1.25	1.3	1.4

This is very interesting insights we can see here, in table, we can observe, The University of Cambridge experiences a lower dropout rate at 0.97%, with a narrow standard deviation of approximately 0.18%. Meanwhile, The University of Oxford demonstrates an average dropout rate of 1.2%, also with a similar standard deviation. Both universities exhibit relatively low variability in their dropout percentages.

Both universities exhibit relatively low variability in their dropout percentages. This indicates a degree of stability in student retention rates over the specified time frame.

## Russell Group

The Russell Group is an association of 24 prominent public research universities based in the United Kingdom. These institutions are widely acknowledged for their outstanding contributions to research, education, and the overall learning environment for students. They stand as some of the most esteemed and historically significant universities in the UK.

Figure 3 is the representation of the combined Russel group universities average percentage of each year. As we can observe the significantly low average percentage of all universities in each year, this is because of all universities are prestigious so we see the less students dropout rates.

Table 4: Russell group universities

Russel Group
1. University College London
2. University of Oxford
3. University of Cambridge
4. The University of Bristol
5. The University of Birmingham
6. The University of Manchester
7. The University of Manchester
8. Newcastle University
9. Imperial College of Science, Technology and Medicine
10. King's College London
11. The University of Leeds
12. The University of Exeter
13. University of Nottingham
14. The University of Sheffield
15. Queen Mary University of London
16. London School of Economics and Political Science
17. The University of Southampton
18. The Queen's Foundation for Ecumenical Theological Education
19. University of Durham

20. The University of York

21. The University of Liverpool

Table 4 shows all universities which are in the Russel group and above universities are only in England.

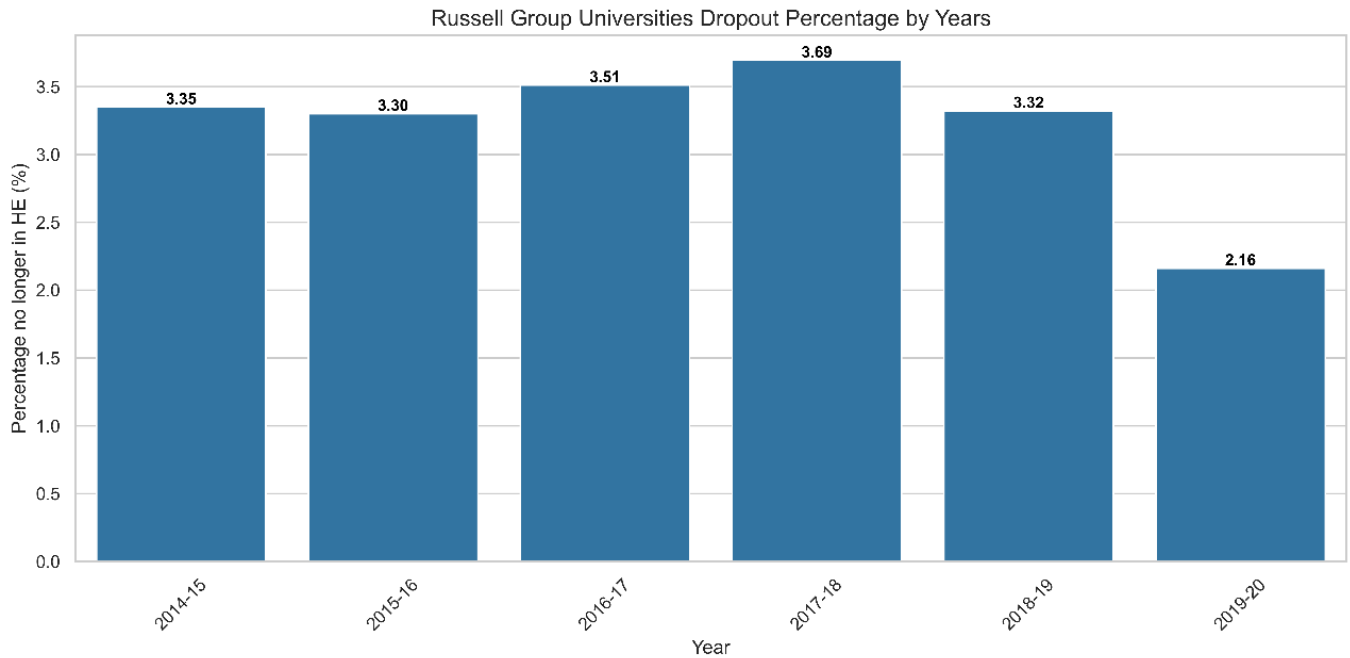


Figure 12: Russell Group dropout percentage by years

Table 5 explains the actual descriptive statistics of all Russel Group universities in each year wise and the mean column are used to plot figure 4.1.

Table 5: Overall dropout statistics in each year Russell group

Year	mean	std	min	25%	50%	75%	max
2014-15	3.35	1.18334	0	2.925	3.4	3.75	6.3
2015-16	3.3	1.301583	0	2.525	3.15	3.775	5.6
2016-17	3.511111	1.47085	0	2.625	3.6	4.175	7.4
2017-18	3.694444	1.267917	0	3.125	3.85	4.075	6.3
2018-19	3.322222	1.259111	0	2.7	3.55	3.925	5.9
2019-20	2.161111	0.747654	0	1.725	2.25	2.75	3.1

We can see the IQR, and standard deviation explains the lower dropouts of all these universities, now it will be more interesting to find out these universities locations and their dropout rates. The consistently low IQR values suggest that most universities maintain a relatively narrow range of dropout rates, indicating a certain level of consistency and stability in their educational programs. Even in 2019-20 in the COVID pandemic these universities dropout rate was 2.16.

## Interpretating of demographic and university student attrition

The research conducted over six years from 2014 to 2020, encompassing data from 240 universities across England, sheds light on the role of demographic factors as predictors of student dropout in higher education. Several key quantitative findings emerge from this analysis:

**Variation in Dropout Percentages:** The data reveals a considerable variation in student dropout percentages across different locations, ranging from 11.9% to 26.2%. This substantial disparity suggests that demographic factors associated with specific areas exert a significant influence on student retention.

**Transfer Rates:** The analysis also highlights variations in the percentages of students who transfer to other UK higher education providers, ranging from 0% to 7.8%. These differences in student mobility patterns based on location underscore the complex interplay between demographics and student choices.

**Lowest Dropout Locations:** Some locations, such as Milton Keynes, Hebden Bridge, and High Wycombe, exhibit exceptionally low dropout rates, all at 0.00%. These findings imply that these areas may provide a highly supportive educational environment or have effective retention programs in place.

**High Dropout Locations:** Conversely, areas like Bedford, Luton, and East Anglia experience higher dropout rates, ranging from 18.0% to 26.2%. These elevated rates could be indicative of various factors impacting student retention, including the availability of academic resources, support systems, and economic conditions specific to these regions.

**Oxbridge and Russell Group Insights:**

Understanding the Oxford and Cambridge university dropout rates involved filtering the data and using the group by method to calculate the combined average percentage of both universities in each year. The resulting statistics show that both universities exhibit relatively low variability in their dropout percentages over the six-year period.

**Russell Group:** The Russell Group is an association of 24 prominent public research universities based in the United Kingdom. These institutions are widely acknowledged for their outstanding contributions to research, education, and the overall learning environment for students. They stand as some of the most esteemed and historically significant universities in the UK.

In conclusion, demographic factors, in conjunction with geographic location, indeed serve as important predictors of student dropout in higher education. The substantial variations in dropout rates among different locations underscore the significance of considering local context and demographics when designing targeted interventions to address student attrition. While this analysis provides valuable insights, it is essential to recognize that other demographic variables, such as age, gender, and socioeconomic background, may also play crucial roles in predicting student dropout rates and warrant further investigation in future research endeavours.

## Is there a relationship between student satisfaction and dropout rates?

(Rinika and Rashmi, Date) and (Young-Jones et al, Date)

The NSS data provides the information about the students responses towards the university and the related entities. So, after we have the overall high student dropout rate university in each year 2015 to 2020, here we will be finding the student's satisfaction relationship on the high dropout rate university in this analysis we can understand the satisfaction and dissatisfaction over Higher education provider and students.



## The Data set Definition.

National student survey website provides the [data](#) annually the dataset structured in then given questions and scales(group of specific field questions) and their respected responses. Responses are columns, Answers1 to Answers5 which indicates the highly satisfied and dissatisfied responses as well as neutral responses.

Rating options	Rating options
Very good	To a large extent
Good	To some extent
Not very good	To a small extent
Not at all good	Not at all
This does not apply to me	This does not apply to me

Answers 5 and Answer 4 are the positive responses which represents the top two options, answers 3 and 2 are the dissatisfied responses and answer 1 has not been taken into consideration because this does not give any meaningful responses which are related to our research question.

The Satisfaction score formula and its significance to find out the satisfied and dissatisfied responses.

$$\text{SSAT Score (\%)} = \left( \frac{\text{Satisfied Responses}}{\text{Total Responses}} \right) \times 100$$

SSAT Score (Student satisfaction score) this gives the average score of satisfied responses, which are answer 4 and 5 additions. We have questions and their responses of 1 to 5. So, the 4 and 5 are the satisfied responses and 2 and 3 are the dissatisfied responses.

Where :

- "Satisfied Responses S " is the calculated addition of highest rating value on the scale (e.g., 4 or 5).

- "Total Responses N " is calculated by that question total responses.

The top university which was exceeded the average of student dropout in each year since 2014 to 2020 which was Anglia Ruskin University. This university has lost 3870 students in the spam of 6 years.

## Data Preparation

To gain a comprehensive understanding of the student survey responses, it is essential to perform data manipulation. This involves filtering the data within the timeframe spanning from 2014 to 2020. This process enables us to observe the distribution of responses on an annual basis. Furthermore, it allows for an examination of the aggregated distribution encompassing all six years. To ascertain an overarching perspective, it is prudent to compute the average over the six-year aggregated distribution, achieved by dividing the total by six, given the uniform duration of the data collection period.

In the demographics student dropout analysis, we have observed the university which have exceeded the student dropout rates more than that particular year's average student dropout rate. In order to find the factors from for that possible student dropout of those universities, in this research question we will analyse the student survey responses of the highest student dropout rate university. Anglia Ruskin University has lost 3870 students in the spam of six years. So, in this research question analysis we are analysing the survey responses of Anglia Ruskin University to interpret the factors.

In order to analyse the correlation between student dropout and survey analysis(Dissatisfaction score) we have to merge them, and we have similar data structure for survey dissatisfaction score table and dropout students data table for Anglia Ruskin University.

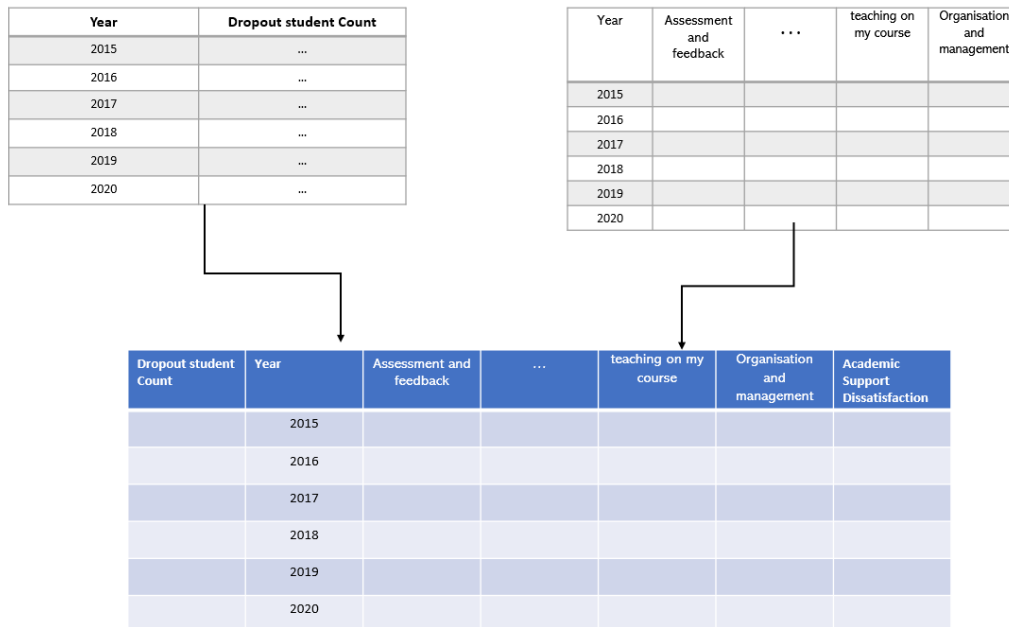


Figure 13: Data preparation

In above figure, these are two different datasets with same and have similar column of year so we can join two datasets to analyse the two different datasets in our student attrition factor analysis. We have taken only specific scales which are more relevant to student perspectives although the provided dataset has seven scales which consists of approximately 4 to 5 questions in each scale.

### Scale 03: Assessment and feedback

Survey scale which consists of Q08 to Q11 titled Assessment and feedback related question were asked to the students. And figure 2.1 explains the average six-year percentage of student satisfaction score (Answer 4 and Answer 5 addition) and Student Dissatisfaction score(Answer 2 and Answer 4 addition).

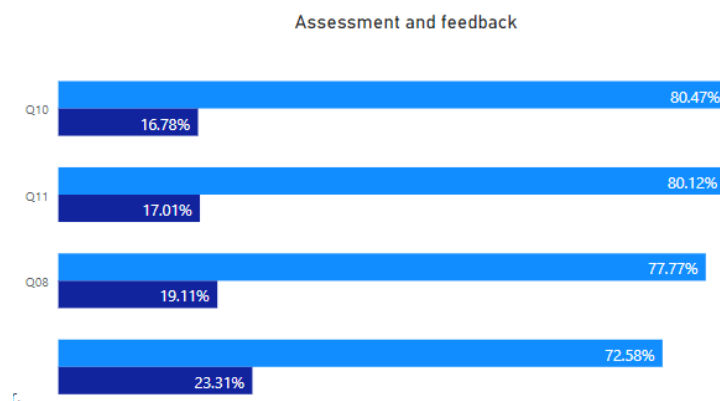


Figure 14: Assessment and feedback

Table 6: scale 3 SSAT and DISSAT

Question Number	Average SSAT	Average DISSAT
The criteria used in marking have been clear in advance.	80%	17%
Marking and assessment has been fair	80%	17%
Feedback on my work has been timely.	78%	19%
I have received helpful comments on my work	7%	23%

the average CSAT and DISSAT percentages across six years for different criteria. The criteria related to marking clarity and fairness exhibit consistent satisfaction levels of around 80%, indicating a steady positive sentiment over time. However, there is room for improvement in the timely provision of feedback, which averages at 78% satisfaction. Notably, receiving helpful comments on work records the lowest satisfaction at 7%, emphasizing a critical area for enhancement to enhance overall satisfaction across the years. high level of dissatisfaction (23%) with the feedback they receive on their work. This suggests a substantial need for improvement in providing helpful comments.

#### Scale 04: Academic support

Another survey data questions scale is for academic support and below is the six-year added percentage

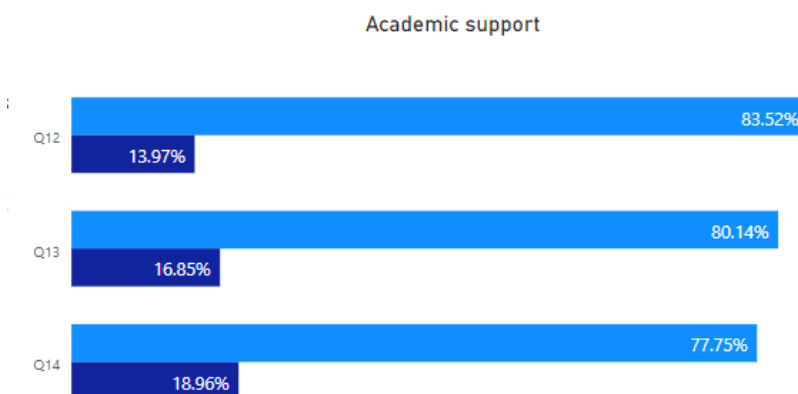


Figure 15: Academic support

The table represent the exact question description which associated with the Q12 to Q14.

Table 7: scale 4 SSAT and DISSAT

Question Number	Average SSAT	Average SDISSAT
I have been able to contact staff when I needed to.	84%	14%
I have received sufficient advice and guidance in relation to my course.	80%	17%
Good advice was available when I needed to make study choices on my course.	78%	19%

The data regarding academic support reveals valuable insights. Across the six-year span, students express a consistently high level of satisfaction (around 80%) with the clarity and fairness of academic support related to factors such as being able to contact staff when needed and receiving sufficient advice

and guidance in relation to their courses. However, there is room for improvement in some areas. The provision of good advice for study choices on courses records an average satisfaction of 78%, suggesting a modest level of dissatisfaction. Moreover, the category with the lowest satisfaction is "I have received sufficient advice and guidance in relation to my course," scoring 80% on average. Additionally, the dataset highlights that students express a higher degree of dissatisfaction (around 14%) with their ability to contact staff when needed, and a more pronounced level of dissatisfaction (approximately 19%) when it comes to good advice for study choices. These findings underscore the importance of enhancing communication channels and providing more comprehensive guidance to meet students' academic needs effectively and thereby improve overall satisfaction.

## Scale 06: Learning Resources

We can observe that the learning resources has got the good student satisfaction scores and very less dissatisfaction score but it still above the 10% of each certain facility related to learning available sources which indicate room for improvement.

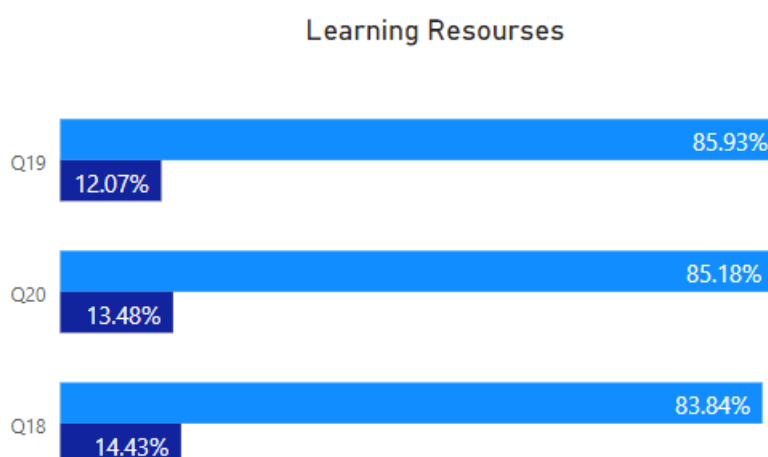


Figure 16: Learning Resources

In the figure 16 Q18 related to IT resources where the 19 is related to the university library resources and 20 is about the course-specific resources.

Table 8: scale 6 SSAT and DISSAT

Question	Actual SSAT	Actual DISSAT
The IT resources and facilities provided have supported my learning well.	85.94	12.07
The library resources (e.g., books, online services and learning spaces) have supported my learning well.	85.18	13.48
I have been able to access course-specific resources (e.g., equipment, facilities, software, collections) when I needed to.	83.84	14.43

IT resources and facilities have consistently supported student learning, with an average SSAT of 85.94 and an average DISSAT of 12.07. Library resources have also been generally supportive, with an average SSAT of 85.18 and an average DISSAT of 13.48. Access to course-specific resources has had slightly lower satisfaction levels, with an average SSAT of 83.84 and an average DISSAT of 14.43.

Overall, students seem to be relatively satisfied with IT and library resources, while satisfaction with course-specific resources is slightly lower but still positive.

## Scale 8: Student Voices survey responses.

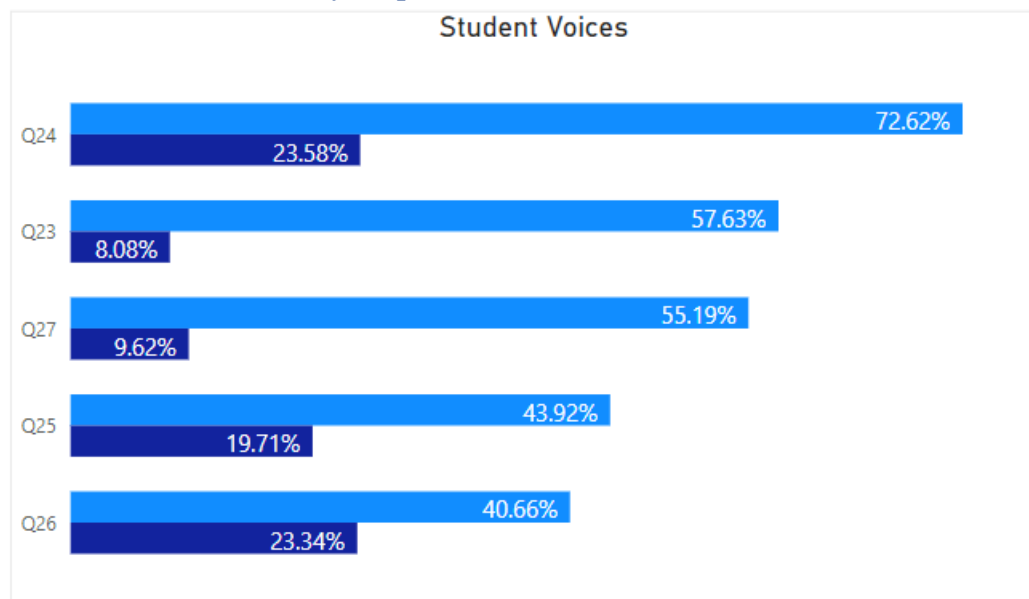


Figure 17: Student Voices

Table 9: scale 8 SSAT and DISSAT

Questions	Actual SSAT	Actual DISSAT
<b>Have had the right opportunities to provide feedback on my course.</b>	72.62	<b>23.58</b>
Staff value students' views and opinions about the course.	57.63	<b>8.08</b>
It is clear how students' feedback on the course has been acted on.	55.19	<b>9.63</b>
The students' union (association or guild) effectively represents students' academic interests	43.93	<b>19.71</b>
Overall, I am satisfied with the quality of the course.	40.66	<b>23.34</b>

## Interpretation of the survey results

**Limited Opportunity for Feedback:** A significant portion of students (around 23.58%) feels that they haven't had enough opportunities to provide feedback on their course. This lack of input channels might lead to frustration and disengagement.

**Lack of Staff Engagement:** A considerable number of students (approximately 8.08%) believe that staff do not sufficiently value their views and opinions about the course. This lack of recognition can impact motivation and commitment to the program.

**Ineffective Response to Feedback:** About 9.63% of students feel that feedback provided by students isn't adequately acted upon. This perception of unresponsiveness to concerns or suggestions can lead to student dissatisfaction.

**Weak Student Representation:** A substantial portion (around 19.71%) believes that the students' union or association doesn't effectively represent their academic interests. This disconnect might contribute to a sense of disconnection from the institution.

**Low Overall Course Satisfaction:** A significant number of students (approximately 40.66%) express overall dissatisfaction with the course's quality. This overarching discontent is a strong indicator of potential dropout or attrition.

To mitigate dropout, it's crucial to address these issues by enhancing communication channels for feedback, improving staff-student interactions, demonstrating responsiveness to feedback, strengthening student representation, and continuously improving course quality to meet student expectations.

## Statistical Analysis overview of survey questions and attrition.

In order to see the dropout rate year wise and the NSS survey questions average dissatisfaction score together I have merged the values using Pandas library by using the year attribute as joins. And below table is the statistical descriptions of all merged number no longer student and survey data.

The mathematical expression of each survey scale is like following. Here is the example of Academic and feedback scale overall six-year average dissatisfaction score.

$$\text{Academic Dissatisfaction} = \frac{\text{Count of Dissatisfied Responses for Q12, Q13, Q14}}{\text{Total Responses for CDISSAT Scale}}$$

In this equation:

“Count of Dissatisfied Responses for Q12, Q13, Q14” represents the sum of dissatisfied responses for the specified questions.

“Total Responses for CDISSAT Scale” represents the total number of responses for the CDISSAT scale.

Same formulas have been used to calculate other scales and below table exhibits the calculated values in each survey scale questionnaire.

*Table 10: All scales and year statistics*

	<b>Number no longer in HE</b>	<b>Year</b>	<b>Academic and Feedback CDISSAT</b>	<b>Academic Support Dissatisfaction</b>	<b>teaching on my course</b>	<b>Organisation and management</b>	<b>Learning Resources</b>	<b>Student Voices</b>
<b>0</b>	565	2015	0.75	0.5	0.42	0.38	0.42	0.32
<b>1</b>	670	2016	0.69	0.49	0.38	0.38	0.4	0.32
<b>2</b>	610	2017	0.76	0.48	0.53	0.61	0.36	0.92
<b>3</b>	680	2018	0.76	0.48	0.51	0.61	0.4	0.98
<b>4</b>	590	2019	0.73	0.48	0.53	0.57	0.39	0.96
<b>5</b>	755	2020	0.88	0.56	0.58	0.63	0.43	0.99

below tables explains the statistical properties of the merged table. Which explains the overall statistical interpretation of the merged table columns.

Table 11: Descriptive statistics

All scales	mean	std	min	25%	50%	75%	max
Number no longer in HE	645	70.14271	565	595	640	677.5	755
Academic and Feedback CDISSAT	0.761667	0.063692	0.69	0.735	0.755	0.76	0.88
Academic Support Dissatisfaction	0.498333	0.031252	0.48	0.48	0.485	0.4975	0.56
teaching on my course	0.491667	0.075741	0.38	0.4425	0.52	0.53	0.58
Organisation and management	0.53	0.117813	0.38	0.4275	0.59	0.61	0.63
Learning Resources	0.4	0.024495	0.36	0.3925	0.4	0.415	0.43
Student Voices	0.748333	0.332651	0.32	0.47	0.94	0.975	0.99

Attrition Rates (Number no longer in HE):

The average number of students no longer in higher education over six years is 645, with a notable standard deviation of 70.14. This variation suggests that student attrition rates fluctuate significantly within the university.

The range of attrition rates spans from a minimum of 565 to a maximum of 755, indicating significant variability in student retention.

Academic and Feedback Satisfaction (CDISSAT):

On average, students express a high level of satisfaction with academic and feedback experiences, with a mean score of approximately 0.762 and a low standard deviation of 0.064. This consistent positive sentiment reflects the university's effectiveness in these areas.

Academic Support Dissatisfaction:

Students report a moderate level of academic support dissatisfaction, with an average score of 0.498 and a low standard deviation of 0.031. This suggests relatively uniform dissatisfaction levels among students.

Teaching Satisfaction (Teaching on My Course):

Teaching satisfaction levels are moderate, with an average score of approximately 0.492. However, the higher standard deviation of 0.076 indicates considerable variability in teaching satisfaction among students.

Organization and Management Satisfaction:

Satisfaction with organization and management is moderately positive, with an average score of about 0.53. The relatively high standard deviation of 0.118 suggests diverse administrative experiences among students, indicating potential areas for improvement.

Learning Resources Satisfaction:

Students generally report positive satisfaction with learning resources, with an average score of 0.4 and a low standard deviation of 0.024, indicating consistency in satisfaction levels.

Student Voices Satisfaction:

Student voices are, on average, rated quite high at 0.748. However, the substantial standard deviation of 0.333 implies a wide range of opinions regarding students feeling heard, pointing to potential areas for engagement and communication enhancement.

In summary, these statistics provide a comprehensive overview of student attrition and satisfaction levels at the university. They offer valuable insights for the institution to identify areas of strength and opportunities for improvement in the pursuit of enhancing the overall student experience and potentially reducing student attrition.

If we find out the average dissatisfaction score only over each year, which gives meaningful insight of the dissatisfaction score trend. Below figure explains the dissatisfaction trend over 6 years span.

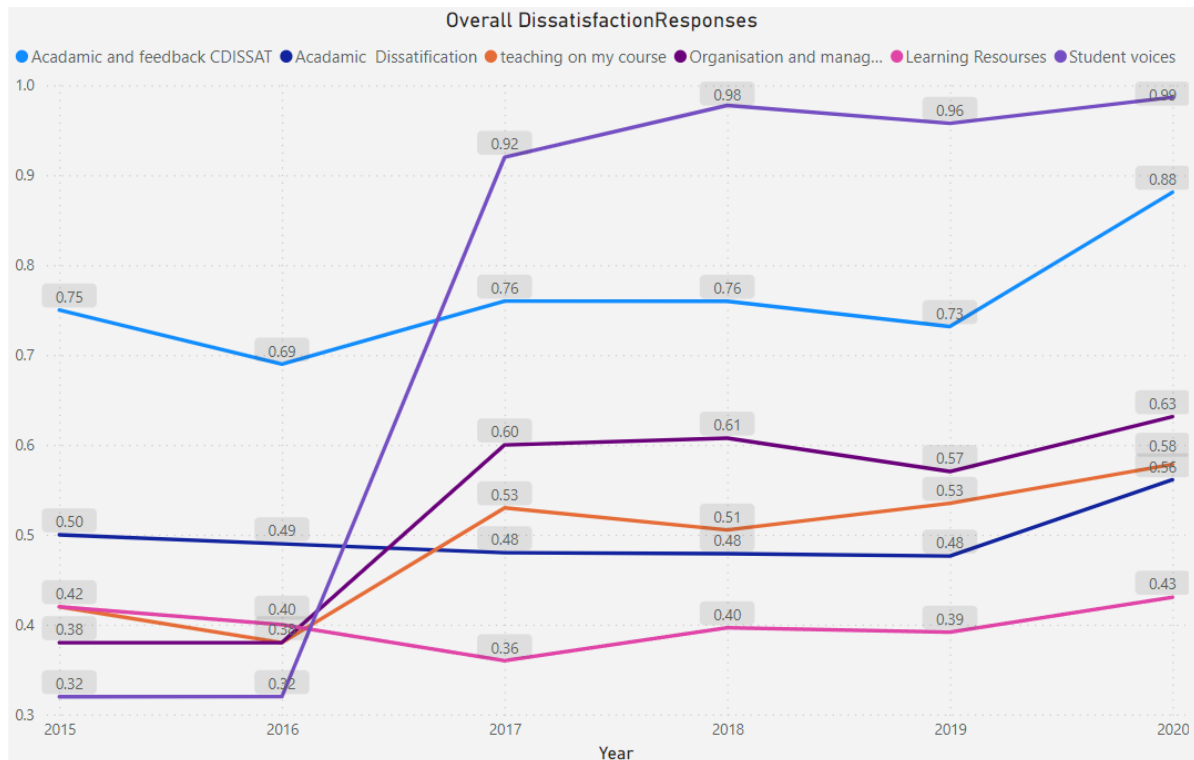


Figure 18: Overall Dissatisfaction Responses

**Overall Trend:** The data shows a general trend of increasing dissatisfaction from 2015 to 2017, followed by a fluctuating pattern until 2020.

**Academic and Feedback Dissatisfaction:** This aspect has remained relatively stable over the years, with scores hovering around 0.75 in 2015 and 2016, and then declining slightly. This suggests that academic and feedback-related concerns have not seen significant changes during this period.

**Teaching on My Course:** There is an increase in dissatisfaction related to teaching on the course from 2015 to 2017 (0.42 to 0.53). However, this score starts to stabilize from 2017 to 2019 and then increases again in 2020 (0.58). This may indicate fluctuations in teaching quality or student perceptions.

**Organization and Management:** Dissatisfaction with organization and management shows a noticeable increase from 2015 to 2017 (0.38 to 0.61). This score remains relatively high in subsequent years (2018 to 2020). This trend suggests that students may have concerns regarding how the institution is organized and managed.



Learning Resources: Dissatisfaction related to learning resources remains somewhat stable, with fluctuations but no significant upward or downward trend. This indicates that students' concerns about the availability and quality of learning resources have not drastically changed over these years.

Student Voices: Dissatisfaction with student voices, as indicated by a low score in 2015 (0.32), increases significantly from 2017 onwards, reaching a peak in 2018 (0.98) before declining slightly in 2019 and 2020. This suggests that students may have had concerns about their ability to voice their opinions or be heard by the institution, with some improvement in recent years.

Year-to-Year Variations: It's important to note that there are some year-to-year variations in these scores. These fluctuations might be due to changes in policies, leadership, or other factors within the institution.

Action Needed: The data indicates areas of concern, particularly in organization and management and student voices. Addressing these issues may help improve overall student satisfaction and the quality of education provided.

### Corelation Between Student Dropout and Dissatisfaction Survey Responses.

In order to analyse student retention trends at Anglia Ruskin University, the initial data was filtered to focus solely on this institution. The number of students no longer in higher education was summed up for each year. To enhance this analysis, survey data related to academic satisfaction, support dissatisfaction, teaching quality, and organizational management was merged based on the shared year index. To ensure fair comparison, the data was then normalized, scaling all variables to a common range. By calculating the correlation matrix after normalization, insights were gained into potential relationships between different factors. This comprehensive approach allowed for a more balanced evaluation of how academic support, teaching quality, and other aspects might be associated with the number of students leaving the institution over time.

The Pearson correlation coefficient, denoted by "r," quantifies the strength and direction of a linear relationship between two variables. It ranges from -1 to +1, where -1 indicates a perfect negative linear relationship, +1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship. The formula for calculating the Pearson correlation coefficient between variables "X" and "Y" is given by:

$$r = \frac{\sum(X_i - \bar{X})^2}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

In the equation  $X_i$  and  $Y_i$  are the individual data points of variables "X" and "Y" respectively and the  $\bar{X}$  and  $\bar{Y}$  are the means of variables "X" and "Y" respectively. The summations are taken over all data points.

In the figure 19 explains the Pearson correlation coefficient of student dropout and survey dissatisfaction scores.

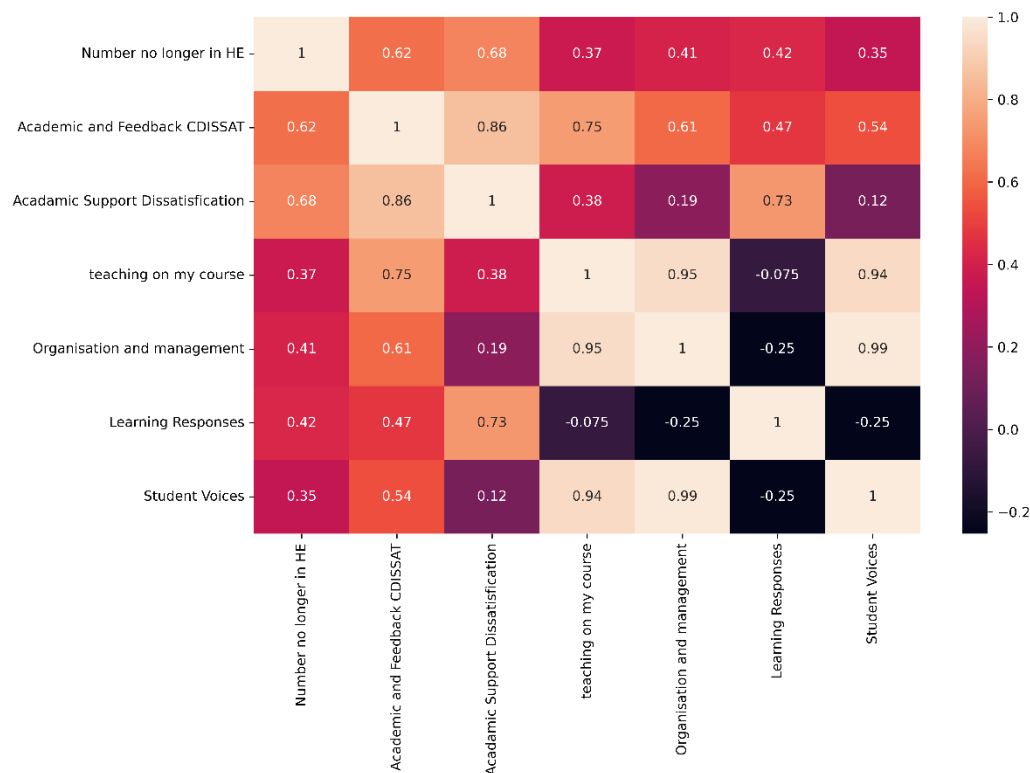


Figure 19: Survey Dissatisfaction correlation

1. Number no longer in HE" and "Academic and Feedback SDISSAT:

There is a moderate positive correlation of approximately 0.622 between the number of students no longer in higher education and dissatisfaction with academic feedback. This suggests that as the number of students leaving higher education increases, there is a corresponding increase in dissatisfaction with academic feedback.

2. Number no longer in HE" and "Academic Support Dissatisfaction:

There is a moderate positive correlation of about 0.680 between the number of students no longer in higher education and academic support dissatisfaction. This implies that students who leave higher education tend to experience academic support dissatisfaction, which could contribute to their decision to discontinue their studies.

3. Number no longer in HE" and "Teaching on My Course:

There is a weak positive correlation of around 0.371 between the number of students no longer in higher education and dissatisfaction with teaching quality. While the correlation is not very strong, it suggests that dissatisfaction with teaching may play a minor role in students' decisions to leave higher education.

4. Number no longer in HE and "Organisation and Management:

There is a weak positive correlation of approximately 0.413 between the number of students no longer in higher education and dissatisfaction with course organization and management. This indicates that students who have left higher education may have perceived issues with how their courses were structured and managed.

5. Number no longer in HE and "Learning Responses:

There is a positive correlation of approximately 0.425 between the number of students no longer in higher education and learning responses. This suggests that as the number of students leaving higher education increases, there is a corresponding increase in learning responses.

6. Number no longer in HE and "Student Voices:

There is a positive correlation of around 0.348 between the number of students no longer in higher education and student voices. This implies that as the number of students leaving higher education increases, there is a corresponding increase in student voices.

Overall, these correlations indicate relationships between the variable "Number no longer in HE" and various aspects of the student experience. As the number of students leaving higher education increases, there tends to be corresponding increases in dissatisfaction with academic feedback, academic support, teaching quality, course organization/management, learning responses, and student voices. These findings underscore the potential impact of dissatisfaction on student attrition and the importance of addressing these areas to improve student retention and overall satisfaction with the educational experience.

The number of students dropout in higher education appears to be intricately related to the survey responses. The correlations unveiled in the data reveal that as this number increases, there is a concurrent rise in dissatisfaction across multiple dimensions of the student experience, including academic feedback, support, teaching quality, and course management. This suggests that students who decide to leave higher education may have encountered various challenges and concerns during their academic journey, ultimately leading to their departure. These findings emphasize the significance of addressing areas of dissatisfaction highlighted in the survey to potentially stem the tide of student attrition. By proactively addressing the issues that students face, institutions can work towards creating a more conducive and supportive learning environment, ultimately fostering greater student retention and satisfaction.

## What are the sociodemographic factors which affect dropout rates?

(Larsen et al, Date)

Here, we have taken the sociodemographic factors [Data](#) from office for nation statistic such as unemployment rate and calculate the correlation(of five year) between no longer students in HE providers. Data set link has provided in references.

	University	Unemployment rate with no longer in HE	Correlation
0	University of Bedfordshire		-0.900903
1	Anglia Ruskin University		-0.147150
2	Leeds Beckett University		0.540947
3	The University of Wolverhampton		0.105895
4	The Manchester Metropolitan University		0.753797
5	The University of Central Lancashire		0.467255
6	London Metropolitan University		0.216919
7	Birmingham City University		-0.287719
8	Middlesex University		0.287672
9	Liverpool John Moores University		-0.010939

In table, the top 10 universities and their correlation of their number of no longer(dropout students) and the unemployment rate within six year of time spam (2014 – 2020).

The correlation values between university attributes and the unemployment rate of students no longer in higher education provide valuable insights into potential sociodemographic factors affecting dropout rates. Here's a concise interpretation of the correlation values:

### Negative Correlation with Dropout Rates:

The University of Bedfordshire shows a strong negative correlation (-0.900) with the unemployment rate of students who are no longer in higher education. This suggests that students from this university might have a lower likelihood of dropping out due to better employment prospects.

Birmingham City University also demonstrates a negative correlation (-0.288), indicating that students from this university might be less likely to drop out when they have more favourable employment opportunities.

### Positive Correlation with Dropout Rates:

The Manchester Metropolitan University exhibits a high positive correlation (0.754) between its attributes and the unemployment rate of students no longer in higher education. This suggests that there might be factors related to this university that contribute to a higher dropout rate among students.

Leeds Beckett University also displays a moderate positive correlation (0.541), indicating that certain university attributes might be associated with a higher likelihood of students leaving higher education.

### Limited Correlation:

Universities like Anglia Ruskin University, The University of Wolverhampton, The University of Central Lancashire, and London Metropolitan University show relatively lower correlation values (between -0.147 and 0.217), implying that the relationship between their attributes and dropout rates might not be as strong.

### Neutral Correlation:

Middlesex University has a moderate correlation (0.288) with dropout rates, suggesting a moderate association with the unemployment rate of students who have left higher education.

Liverpool John Moors University has a near-neutral correlation (-0.011), indicating a weaker link between its attributes and dropout rates.

### Anglia Ruskin University:

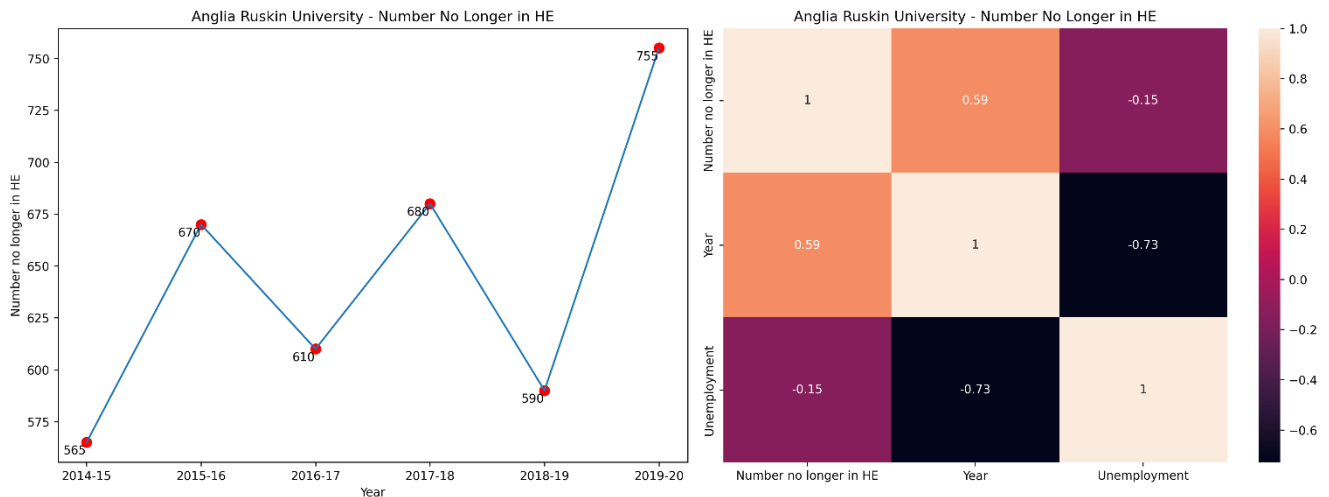


Figure 20: Anglia university dropout and correlation

The correlation matrix for Anglia Ruskin University indicates that the correlation between the number of non-continuation students and unemployment rate is -0.147, indicating a weak negative relationship. Additionally, the correlation between the number of non-continuation students and year is 0.594, suggesting a moderate positive association.

Considering these correlations, we can infer the following:

The weak negative correlation between non-continuation students and unemployment rate (-0.147) might suggest that other sociodemographic factors play a more significant role in influencing student dropout rates at Anglia Ruskin University.

The moderate positive correlation between non-continuation students and year (0.594) could indicate that other temporal factors, such as academic cycles or policy changes, might contribute to the fluctuations in student dropout rates.

## The Manchester Metropolitan University:

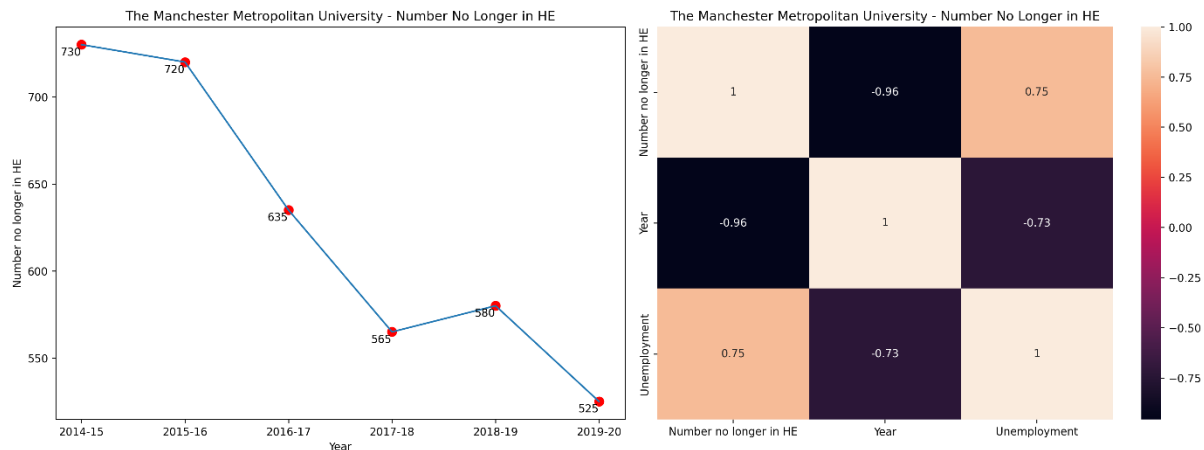


Figure 21: Manchester Metropolitan University dropout and correlation

The correlation matrix for The Manchester Metropolitan University shows a strong positive correlation between the number of non-continuation students and the unemployment rate (0.754).

The correlation between the number of non-continuation students and year is -0.957, indicating a strong negative relationship.

Interpreting these correlations yields the following insights:

The strong positive correlation between non-continuation students and the unemployment rate (0.754) suggests that the unemployment rate could have a notable impact on dropout rates at this university. Lower unemployment rates might be associated with higher retention rates.

The strong negative correlation between non-continuation students and year (-0.957) could indicate that specific temporal influences, such as changes in policies or support systems, have contributed to the university's consistent pattern of exceeding average dropout counts.

## The University of Central Lancashire:

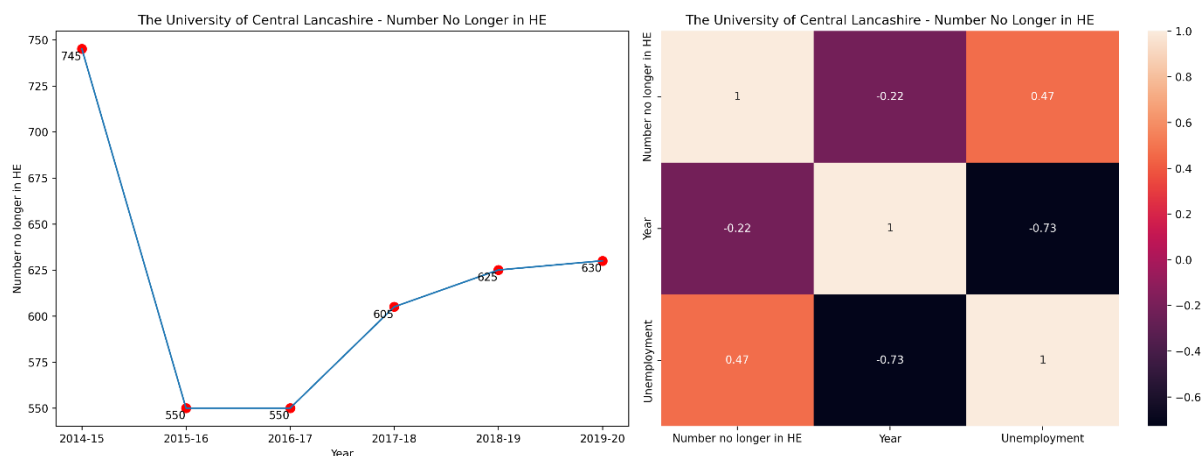


Figure 22: University of central Lancashire dropout and correlation

The correlation matrix for The University of Central Lancashire demonstrates a moderate positive correlation between the number of non-continuation students and the unemployment rate (0.467).

The correlation between the number of non-continuation students and year is -0.220, indicating a relatively weak negative relationship.

### Analysing these correlations leads to the following insights:

The moderate positive correlation between non-continuation students and the unemployment rate (0.467) suggests that, to some extent, the unemployment rate might influence student dropout rates at the university.

The relatively weak negative correlation between non-continuation students and year (-0.220) could imply that temporal factors might contribute less to the observed fluctuations in dropout rates compared to other influential variables.

In examining the correlation between various sociodemographic factors and student dropout rates, it becomes evident that the relationship between these factors and dropout rates is nuanced and multifaceted. The correlation analysis highlights that while certain factors exhibit correlations with dropout rates, their impact varies across different universities and contexts.

### Unemployment Rate and Dropout Rates:

The correlation analysis shows that the unemployment rate is not a singular determinant of student dropout rates across all universities. Its influence on dropout rates varies from weak to strong correlations, suggesting that while it can be a relevant sociodemographic factor, it doesn't solely dictate dropout trends.

Universities with strong positive correlations between unemployment rates and dropout rates could benefit from efforts to improve employment prospects to potentially enhance student retention.

### Additional Sociodemographic Factors:

The consistent patterns of universities exceeding average dropout counts, despite varying correlations, indicate the presence of underlying sociodemographic factors beyond the ones studied (e.g., food costs) that contribute to student attrition.

The complexity of these factors underscores the need for a more comprehensive exploration that includes variables like socioeconomic diversity, campus support systems, student engagement, and local influences.

### Temporal Factors:

The correlation between year and dropout rates, particularly strong negative correlations, suggests that temporal factors such as changes in policies, support systems, or academic cycles, might significantly impact student retention trends.

In essence, this analysis emphasizes that student dropout rates are influenced by a combination of socioeconomic, temporal, and university-specific factors. To gain a holistic understanding of the sociodemographic factors affecting dropout rates, it's crucial to consider a wide range of variables beyond the ones examined here. The intricate interplay of these factors necessitates further qualitative research, including student surveys, interviews, and a deeper exploration of campus dynamics, to fully comprehend the mechanisms driving student attrition. Such comprehensive insights will aid institutions in developing targeted strategies to mitigate dropout rates and enhance overall student success.

## Conclusion:

The exclusive conclusion has been explained at end of each research questions and below are the overall key findings that puts highlights on main areas of factors most possible reasons.

### Demographic Factors Influence Churn Rates:

1. **Geographic Variation:** The research highlights substantial geographic variation in student attrition rates across different UK locations. This suggests that local demographic and socioeconomic factors significantly influence the likelihood of student dropout. Certain areas may face unique challenges or offer specific support systems that impact retention rates.
2. **High Dropout Areas:** The study identifies specific locations with notably high dropout rates. These areas may require targeted interventions to address the underlying factors contributing to student attrition. Understanding the local context and implementing tailored retention strategies can be crucial in these regions.

### Oxbridge and Russell Group Universities:

1. **Oxbridge Excellence:** The University of Cambridge and the University of Oxford, collectively known as Oxbridge, consistently demonstrate low student attrition rates. This reflects their effective retention strategies, which may include rigorous academic support, robust mentorship programs, and a conducive learning environment.
2. **Russell Group Performance:** Russell Group universities, renowned for their research excellence, also exhibit low dropout rates. This suggests that institutions with a strong academic reputation and comprehensive support systems tend to retain students at higher rates.

### Unemployment Rate Correlation:

1. **Influential but Not Sole Determinant:** The correlation analysis emphasizes that the unemployment rate is a significant sociodemographic factor affecting student churn. However, it's important to note that while it plays a substantial role, it is not the exclusive determinant. Other factors, such as campus support services, academic engagement, and socioeconomic diversity, also contribute to student retention.
2. **Varying Degrees of Correlation:** Universities display varying degrees of correlation between the unemployment rate and dropout rates. This indicates that the impact of unemployment on attrition is nuanced and context dependent. Some universities may have additional support structures in place that mitigate the influence of unemployment on student retention.

### Temporal Factors Play a Role:

**Policy and Support Changes:** The study identifies temporal factors, including changes in policies, support systems, and academic cycles, as significant contributors to student retention trends. These temporal influences can lead to year-to-year variations in dropout rates. Understanding and adapting to these changes is crucial for universities seeking to enhance student retention.

The study underscores the multifaceted nature of student churn in UK universities, influenced by a combination of sociodemographic, temporal, and university-specific factors. While geographic location, unemployment rates, and prestigious university status (Oxbridge, Russell Group) play significant roles, they are not the sole determinants of student attrition. Other variables such as socioeconomic diversity, campus support systems, and student engagement also contribute to the complex churn dynamics.



## Limitations

- The data obtained from the HESA website lacks information pertaining to students' demographic backgrounds and the precise geographical locations of universities.
- Manually adding data for each university is limited by the absence of specific information regarding university locations
- Before we can conduct any meaningful analysis, we need to reorganize the datasets. They are currently organized by academic year, which makes it tricky to draw conclusions. Once we've carefully restructured the data through a process called feature engineering, we'll be able to proceed with our analysis.

## Further Work.

The profound study can be said when we have immanence amount of data so we can explore every perceptive that might have possibility to affect the student dropouts and lots of other factors are still can be studied on student dropouts which needs more data, more computation power, more complex statistical techniques. This concept calls Big Data Analysis, which is performed on the latest cloud Databases, Hadoop, Apache spark, Hive, Deep learning algorithms, Reinforcement training, Natural Language Preprocessing and more. In this AI thriving generation, we have multiple ways to captures more information from the student more advanced techniques like Facial recognition data where we can analysis the student behaviour while they are learning in lectures, Voice and Audio Analysis, Mobile App Usage Data, social media and Online Activity, Graph Analytics. These are the new high dimensional data which will give more propound meaningful insights. Moreover, we can use different architectures like this study suggest combining more different source which are relent and finding more meaningful insights.

## References

1. Aina, C., Baici, E., Casalone, G. and Pastore, F., 2022. The determinants of university dropout: A review of the socio-economic literature. *Socio-Economic Planning Sciences*, 79, p.101102.
2. Berge, Z.L. and Huang, Y.P., 2004. A Model for Sustainable Student Retention: A Holistic Perspective on the Student Dropout Problem with Special Attention to e. *Learning*, 13(5), pp.97-108.
3. B  ulke, L., Grunschel, C. and Dresel, M., 2022. Student dropout at university: A phase-orientated view on quitting studies and changing majors. *European Journal of Psychology of Education*, 37(3), pp.853-876.
4. Paura, L. and Arhipova, I., 2014. Cause analysis of students' dropout rate in higher education study program. *Procedia-social and behavioral sciences*, 109, pp.1282-1286.
5. Kehm, B.M., Larsen, M.R. and Sommersel, H.B., 2019. Student dropout from universities in Europe: A review of empirical literature. *Hungarian Educational Research Journal*, 9(2), pp.147-164.
6. Xenos, M., Pierrakeas, C. and Pintelas, P., 2002. A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. *Computers & Education*, 39(4), pp.361-377.
7. Aulck, L., Velagapudi, N., Blumenstock, J. and West, J., 2016. Predicting student dropout in higher education. arXiv preprint arXiv:1606.06364.
8. Paul, R. and Rashmi, M.R., 2022, October. Student Satisfaction and Churn Predicting using Machine Learning Algorithms for EdTech course. In *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 1-6). IEEE.

9. Grillo, M.C. and Leist, C.W., 2013. Academic support as a predictor of retention to graduation: New insights on the role of tutoring, learning assistance, and supplemental instruction. *Journal of College Student Retention: Research, Theory & Practice*, 15(3), pp.387-408.
10. [https://www.ncsehe.edu.au/wp-content/uploads/2017/03/03LiUWA\\_Formatted\\_FINAL.pdf](https://www.ncsehe.edu.au/wp-content/uploads/2017/03/03LiUWA_Formatted_FINAL.pdf)
11. <https://research.acer.edu.au/ausse/1/>
12. Young-Jones, A.D., Burt, T.D., Dixon, S. and Hawthorne, M.J., 2013. Academic advising: does it really impact student success?. *Quality Assurance in Education*, 21(1), pp.7-19.
13. Rovai, A.P. and Barnum, K.T., 2003. On-line course effectiveness: An analysis of student interactions and perceptions of learning. *International Journal of E-Learning & Distance Education/Revue internationale du e-learning et la formation à distance*, 18(1), pp.57-73.
14. Hoyt, J.E., 2021. Student connections: The critical role of student affairs and academic support services in retention efforts. *Journal of College Student Retention: Research, Theory & Practice*, p.1521025121991502.
15. Menzies, L., 2023. Continuity and churn: understanding and responding to the impact of teacher turnover. *London Review of Education*, 21(1).
16. Latif, A., Choudhary, A.I. and Hammayun, A.A., 2015. Economic effects of student dropouts: A comparative study. *Journal of global economics*, 3(2), pp.1-4.
17. Latif, A., Choudhary, A.I. and Hammayun, A.A., 2015. Economic effects of student dropouts: A comparative study. *Journal of global economics*, 3(2), pp.1-4.
18. <https://www.studyin-uk.com/uk-study-info/uk-university-map/>
19. <https://www.hesa.ac.uk/>
20. <https://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/>
21. <https://www.ons.gov.uk/>
22. <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-quickstart-learn-dax-basics>
23. HESA DATA SET:- <https://www.hesa.ac.uk/data-and-analysis/performance-indicators/non-continuation>
24. NSS DATA: <https://www.officeforstudents.org.uk/data-and-analysis/national-student-survey-data/nss-data-archive/>
25. <https://www.officeforstudents.org.uk/data-and-analysis/national-student-survey-data/nss-data-archive/> Economic Data

# Appendix

## Student Dropout and demographic analysis Research Question 1

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

In [2]: df = pd.read_csv('A_England.csv',encoding='latin1')

In [6]: df.drop('Column1',axis=1,inplace=True)

In [5]: #df.drop('+/-',axis=1,inplace=True)

In [6]: def change_int(column):
    clened = []
    for i in column:
        try:
            if math.isnan(i):
                clened.append(0) # or any other appropriate value for NaN
            else:
                val = float(i)
                clened.append(val)
        except:
            clened.append(int(i.replace(',','')) if isinstance(i, str) else None)
    return clened

In [7]: df['Number no longer in HE'] = change_int(df['Number no longer in HE'])

In [8]: df['Percentage no longer in HE (%)'] = df['Percentage no longer in HE (%)'].fillna(0)

In [9]: df['Total full-time entrants'] = change_int( df['Total full-time entrants'])

In [10]: df['Number who continue or qualify at same HEP'] = change_int(df['Number who continue or qualify at same HEP'])

In [7]: df['Number no longer in HE'] = change_int(df['Number no longer in HE'])

In [8]: df['Percentage no longer in HE (%)'] = df['Percentage no longer in HE (%)'].fillna(0)

In [9]: df['Total full-time entrants'] = change_int( df['Total full-time entrants'])

In [10]: df['Number who continue or qualify at same HEP'] = change_int(df['Number who continue or qualify at same HEP'])

In [11]: def clean_float(column):
    clened = []
    for i in column:
        if math.isnan(i):
            val = 0.0
            clened.append(val)
        else:
            clened.append(float(i))

    df['Percentage who continue or qualify at same HEP (%)'] = clened
    return clened

In [12]: df['Percentage who continue or qualify at same HEP (%)'] = clean_float(df['Percentage who continue or qualify at same HEP (%)'])

In [13]: df['Adjusted sector continue or qualify (%)'] = clean_float(df['Adjusted sector continue or qualify (%)'])

In [14]: df['Percentage who continue or qualify at same HEP (%)'] = clean_float(df['Percentage who continue or qualify at same HEP (%)'])

In [15]: df['Adjusted sector transfer to other UK HEP (%)'] = clean_float(df['Adjusted sector transfer to other UK HEP (%)'])

In [16]: df['Percentage who transfer to other UK HEP (%)'] = clean_float(df['Percentage who transfer to other UK HEP (%)'])

In [17]: df['Benchmark (%)'] = clean_float(df['Benchmark (%)'])
```

```

In [83]: top_20_locations.to_csv('top_20_locations.csv')

In [55]: import matplotlib.pyplot as plt
import seaborn as sns

# Assuming you have already calculated and stored top_20_lowest_dropout_locations

# Setting up the plot style
sns.set(style="whitegrid")
plt.figure(figsize=(10, 7), dpi=300) # Adjusted figure size and DPI

# Creating the bar plot
ax = sns.barplot(x=top_20_lowest_dropout_locations.index, y='Percentage no longer in HE (%)', data=top_20_lowest_dropout_locations)
plt.xticks(rotation=45, ha="right", fontsize=10) # Adjusted font size and rotation
plt.xlabel('Location', fontsize=12) # Adjusted font size
plt.ylabel('Percentage no longer in HE (%)', fontsize=12) # Adjusted font size
plt.title('Top 20 Demographic Locations with Low Mean Dropout Rates', fontsize=14)

# Adding data Labels above the bars
for i, v in enumerate(top_20_lowest_dropout_locations['Percentage no longer in HE (%)']):
    ax.text(i, v + max(top_20_lowest_dropout_locations['Percentage no longer in HE (%)']) * 0.02, f'{v:.2f}%', ha='center', va='bottom')

# Adjusting plot layout
plt.tight_layout()

plt.savefig('low_dropout_locations.png', dpi=300, bbox_inches='tight')

# Show the plot
plt.show()

```

The analysis has been done with lots of cells and manipulating data, for each diagram I have reported the code because of uploading it into the dissertation and saving that file. So, we have here lots of cells. Here is the link which has all code this is the Jupiter file on GitHub repository.

<https://github.com/Sid0925/Dissertation-calculations-and-Data-sep-2023/blob/main/Data%20Cleaning%20England%20and%20analysing%20bad%20and%20good%20HE%20Providers.ipynb>

## Research Question 2 code and analysis calculations on power bi

Student survey data analysis which has been done on only on that university which has reportedly come in six years. Below DAX formulas are used to create the diagrams which are mentioned in research question 2. Here is the dashboard link: <https://github.com/Sid0925/Dissertation-calculations-and-Data-sep-2023/tree/91d77bb814f476f7b40dc7c24cb6b0973832efeb> this will required power bi software to see the dashboard.

```
1 teaching on my course =
2 CALCULATE(
3     [CDISSAT],
4     FILTER(
5         'Main data',
6         'Main data'[Question Number] = "Q01" ||
7         'Main data'[Question Number] = "Q02" ||
8         'Main data'[Question Number] = "Q03" ||
9         'Main data'[Question Number] = "Q04"
10    )
11 )
12 )
13
```

```
1 Academic and feedback CDISSAT =
2 CALCULATE(
3     [CDISSAT],
4     FILTER(
5         'Main data',
6         'Main data'[Question Number] = "Q08" ||
7         'Main data'[Question Number] = "Q09" ||
8         'Main data'[Question Number] = "Q10" ||
9         'Main data'[Question Number] = "Q11"
10    )
11 )
12
```

```
1 Organisation and management =
2 CALCULATE(
3     [CDISSAT],
4     FILTER(
5         'Main data',
6         'Main data'[Question Number] = "Q15" ||
7         'Main data'[Question Number] = "Q16" ||
8         'Main data'[Question Number] = "Q17"
9    )
10 )
11 )
12 )
13
```

```
1 Academic Dissatification =
2 CALCULATE(
3     [CDISSAT],
4     FILTER(
5         'Main data',
6         'Main data'[Question Number] = "Q12" ||
7         'Main data'[Question Number] = "Q13" ||
8         'Main data'[Question Number] = "Q14"
9    )
10 )
11 )
12
```

```
1 Learning Resources =
2 CALCULATE(
3     [CDISSAT],
4     FILTER(
5         'Main data',
6         'Main data'[Question Number] = "Q18" ||
7         'Main data'[Question Number] = "Q19" ||
8         'Main data'[Question Number] = "Q20"
9    )
10 )
11 )
12 )
```

```
1 Student voices =
2 CALCULATE(
3     [CDISSAT],
4     FILTER(
5         'Main data',
6         'Main data'[Question Number] = "Q23" ||
7         'Main data'[Question Number] = "Q24" ||
8         'Main data'[Question Number] = "Q25" ||
9         'Main data'[Question Number] = "Q26" ||
10        'Main data'[Question Number] = "Q26" ||
11        'Main data'[Question Number] = "Q28"
12    )
13 )
14 )
15
```

## Research Question 3 Economical factors Analysis.

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

```
In [2]: df = pd.read_csv('Unemployment rate.csv')
```

### Data cleaning For Unemployment Rate Factor

```
In [3]: import pandas as pd

# Load the unemployment data
unemployment_data = pd.read_csv('Unemployment rate.csv', skiprows=5, names=['Date', 'Unemployment'])

# Define a function to convert the date strings to datetime objects representing the end of the period
def convert_date_string(date_string):
    # If the string contains 'Q', it represents a quarter
    if 'Q' in date_string:
        year, quarter = date_string.split(' Q')
        quarter = int(quarter)
        # Determine the last month of the quarter
        if quarter == 1:
            month = '03'
        elif quarter == 2:
            month = '06'
        elif quarter == 3:
            month = '09'
        elif quarter == 4:
            month = '12'
        return pd.to_datetime(f'{year}-{month}', format='%Y-%m')
    # If the string contains a space, it represents a specific month
    elif ' ' in date_string:
        month_dict = {'JAN': '01', 'FEB': '02', 'MAR': '03', 'APR': '04', 'MAY': '05', 'JUN': '06',
                     'JUL': '07', 'AUG': '08', 'SEP': '09', 'OCT': '10', 'NOV': '11', 'DEC': '12'}
        year, month = date_string.split(' ')
        month = month_dict[month]
        return pd.to_datetime(f'{year}-{month}', format='%Y-%m')
    # Otherwise, the string represents a year
    else:
        return pd.to_datetime(f'{date_string}-12', format='%Y-%m')

# Apply the function to the 'Date' column
unemployment_data['Date'] = unemployment_data['Date'].apply(convert_date_string)
```

### Loading England Churn File

```
In [5]: C = pd.read_csv('A_England.csv')
C.drop('Unnamed: 0', axis=1, inplace=True)
```

```
In [7]: C.tail()
```

```
Out[7]:
```

	HE PROVIDER	Total full-time entrants	Number who continue or qualify at same HEP	Percentage who continue or qualify at same HEP (%)	Adjusted sector or qualify (%)	Number who transfer to other UK HEP	Percentage who transfer to other UK HEP (%)	Adjusted sector transfer to other UK HEP (%)	Number no longer in HE	Percentage no longer in HE (%)	Benchmark (%)	Standard deviation (%)	HE provider's average contribution to benchmark (%)	A <sub>nurr</sub> pro comp
1405	The University of Wolverhampton	4045	3415	0.44	86.8	150.0	3.7	3.2	480	11.9	10.0	0.44	3.2	
1406	University of Worcester	2255	2015	0.55	90.0	40.0	1.9	2.5	200	8.8	7.5	0.55	2.4	
1407	Writtle University College	195	175	1.68	89.4	5.0	2.6	2.7	15	6.7	7.9	1.68	5.2	
1408	York St John University	1825	1605	0.61	89.1	60.0	3.3	3.0	165	9.0	7.9	0.61	1.5	
1409	The University of York	3735	3615	0.33	94.4	45.0	1.2	2.1	75	2.0	3.4	0.33	3.0	

### Finding the top 10 Universities

```
In [7]: df = pd.read_csv("A_England.csv")

# Group the data by university and calculate the sum of 'Number no longer in HE' for each university
university_churn_totals = df.groupby('HE PROVIDER')['Number no longer in HE'].sum()

# Sort the totals in descending order and select the top 10
top_10_universities_by_churn = university_churn_totals.sort_values(ascending=False).head(10)

top_10_universities_by_churn
```

```
Out[7]: HE PROVIDER
University of Bedfordshire      4260
Anglia Ruskin University        3870
Leeds Beckett University        3855
The University of Wolverhampton  3825
The Manchester Metropolitan University  3755
```

## Data merging and manipulating

Merging with the unemployment rate csv (Indexing matching)

```
A = C[C['HE PROVIDER'] == 'University of Bedfordshire'][['Number no longer in HE', 'Year']] # For line plot
```

```
uni = C[C['HE PROVIDER'] == 'University of Bedfordshire'][['Number no longer in HE', 'Year']]
year_uni = unemployment_data_yearly[-9:-3]
year_uni.set_index(uni.index, inplace=True)
uni_clean = pd.concat([uni, year_uni], axis=1)
```

```
A = C[C['HE PROVIDER'] == 'University of Bedfordshire'][['Number no longer in HE', 'Year']] # For line plot
uni = C[C['HE PROVIDER'] == 'University of Bedfordshire'][['Number no longer in HE', 'Year']]
year_uni = unemployment_data_yearly[-9:-3]
year_uni.set_index(uni.index, inplace=True)
uni_clean = pd.concat([uni, year_uni], axis=1)
# First subplot: Line plot and scatter plot
plt.figure(figsize=(16, 6), dpi=250)

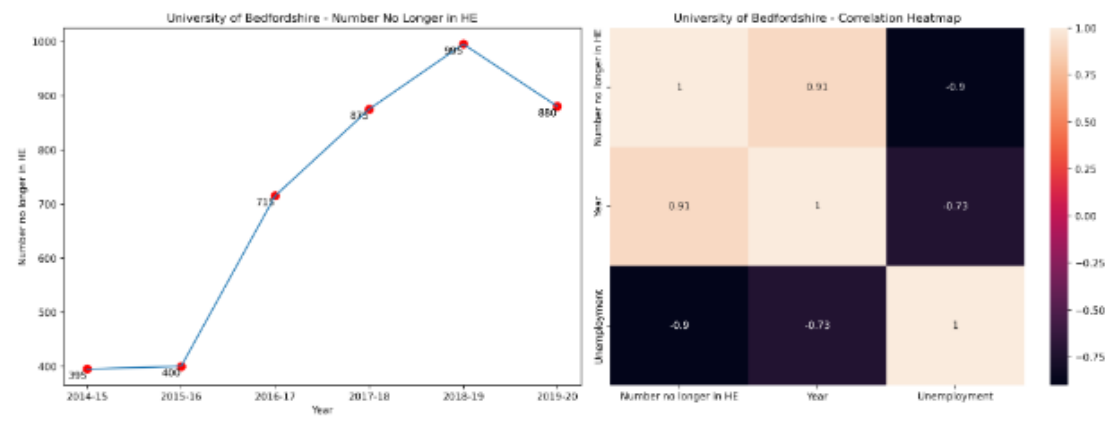
plt.subplot(1, 2, 1)
sns.lineplot(x=A['Year'], y=A['Number no longer in HE'])
sns.scatterplot(x=A['Year'], y=A['Number no longer in HE'], color='red', s=100)
for x, y in zip(A['Year'], A['Number no longer in HE']):
    plt.annotate(f'{y}', (x, y), textcoords="offset points", xytext=(-10, -10), ha='center')

plt.title('University of Bedfordshire - Number No Longer in HE')

# Second subplot: Heatmap
plt.subplot(1, 2, 2)
sns.heatmap(uni_clean.corr(), annot=True)

plt.title('University of Bedfordshire - Correlation Heatmap')

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



Similarly, above code we have multiple times of code for each university. Follow the link to see full code and analysis.

<https://github.com/Sid0925/Dissertation-calculations-and-Data-sep-2023/tree/main>