Final Report

Master Internal - Marketing & Communications

21st June 2019

Team 8

Sebastian Thierfelder Guerrero | i6071904

Pierre Michel B. Hardy | i6117019

Christoph Mauer | i6109887

Robbert Lalisang| i6050220

Nardie Crijns | i6183567

Table of Contents

1. Introduction

Keeping students satisfied, retaining them throughout their Bachelor studies, and converting them to Master students are challenging tasks that the marketing departments at every university are facing. In general, it is highly desirable for any organization to retain their existing *customers* (students), as acquiring a new customer is approximately ten times more costly than keeping current ones (Berry & Linoff, 2004).

With this in mind, it is remarkable that the School of Business and Economics (SBE) currently does not use any form of marketing to convert internal Bachelor students to Master students, as Germano Giansante, Student Recruitment Advisor, pointed out to us. From an analytics perspective, this is even more surprising because, for internal Bachelor students, vast amounts of both demographic and academic data are available, which could be used for analyzing students' decision-making.

Therefore, this project aims to develop a smart service that leverages these vast amounts of data on internal Bachelor students, in order to support the marketing and communications team in improving the conversion to Master students.

To address this problem, our team presents a multi-layer solution consisting of a predictive model, a marketing dashboard, and a Student Portal extension. The dashboard is based on insights from our predictive analytics and helps the marketing team in doing personalized marketing for different types of students, depending on their characteristics and their computed probability of staying at Maastricht University. The Student Portal extension offers the unique opportunity to collect high-quality data on leads and to communicate more effectively with internal students.

To motivate and explain our service and its implementation, the remainder of this report is structured as follows: First, we are extensively elaborating the service concept, followed by an explanation of the analytics component, including data cleaning. Next, we describe how to implement the service from a data management perspective and give an evaluation proposal. Finally, we discuss possibilities for disruption and conclude with options for future evaluation moments and improvements.

2. Service Concept

During the service concept stage of the project, we explored the problem at hand by conducting interviews with both students and UM staff members, which allowed us to gain a holistic view of the current situation. With a clear understanding of the case, we designed a concept that solves the needs of internal Bachelor students and the UM marketing team, developing prototypes for each of the two groups. Finally, we validated them by exhibiting and discussing the prototypes with the corresponding target groups. The individual steps of this phase are elucidated in the following.

2.1. Exploration

To get a fundamental understanding of the underlying problem, we approached several UM staff members. Talking to them was crucial for the development of our service concept, as these interviews revealed invaluable insights.

We started by informally interviewing Monique Merckx, Head of Marketing and Communications at Maastricht University School of Business and Economics. From this interview, three main issues became evident: First, it is not clear how well UM Bachelor students are informed about available Master programmes. For the staff members, assessing the impact of the currently used communication channels (social media campaigns, open days, flyers, etc.) on internal Bachelor students is very challenging, if not impossible. Second, despite the geographical proximity and digital contact between the marketing department and UM students, there is little interaction between the two groups. So, although UM has the unique advantage of possessing students' information and being able to reach them via e.g., the student portal, there are no concrete ideas on how to interact with them in a way that brings about excitement for UM

Master programmes. Third, and most important, it is not clear how students make the decision to continue with a Master programme at Maastricht University.

Consequently, UM staff members do not know which Bachelor students to target with their advertisements and campaigns. This can result in a considerable amount of money wasted, simply because it might be spent on advertising to the wrong students (e.g., those who are already determined on studying somewhere else) or through ineffective channels (e.g. spending large amounts of money on AdWords campaigns for an audience of students that have already made the decision to stay at Maastricht University). Already from this initial interview, it was undeniable that data analytics can be used to significantly improve SBE's marketing efforts for internal Bachelor students.

To further deepen our understanding of the situation, we scheduled an interview with Germano Giansante, Student Recruitment Advisor at SBE, and Eva Schneider, former Recruitment Advisor at SBE (see Appendix for the interview notes). This interview confirmed all the insights mentioned above and shed more light on the problem at hand. Clarifying with the marketing and communications team who the main stakeholders in our project should be, helped us to get the holistic picture we aimed to have. From Germano's point of view, the most important stakeholders are students, academic staff, employers, and the marketing department. With regard to students as key stakeholders, it is essential to note that "internal students" are (1) SBE Bachelor and Master students, (2) UM Bachelor and Master students from other faculties, and (3) exchange students, as all these groups are in a unique position to directly interact with the staff members.

Interestingly, around 50% of current Master students did their Bachelor at UM, while 30% of all Bachelor students continue with a Master programme at SBE. These numbers mean that there is still much room for improvement, considering that 70% of all undergraduate SBE students do not return for a Master programme. However, as mentioned before, it is currently unknown to the marketing team how students make this decision and, therefore, not possible to effectively improve the recruitment funnel. Additionally, it was pointed out to us that parts of the currently available data, such as information flyer orders and events attendance (leads data), are almost useless as they are very inconsistent and include a vast number of missing values Finally, it is worth mentioning that the marketing team faces a particular challenge when targeting internal Bachelor students: On the one hand, the university wants to send the message that students can and should explore the world and all its educational and professional opportunities, while, on the other hand, incentivise them to stay in Maastricht to do their Master.

In sum, it became clear from the interviews that all marketing strategy for internal students are largely based on intuition and that it is not clear what students currently "know" about available Master programmes, nor how they decide to continue studying at or to leave Maastricht University.

To obtain answers to these open questions, we conducted a combination of desk and field research. After the first milestone presentation, we spoke to Jan Nijhuis (researcher and educational policy advisor at SBE). From him, we obtained the report, "Perception and Position of Bachelor Graduates Cohort 2016-2017", published by the Policy Development and Quality Assurance department of SBE, which presents a deeper understanding of graduating bachelor students. The report provides this information by asking questions on the students' situation after their Bachelor graduation through a survey, including students that go on studying a master at Maastricht University. Additionally, we conducted qualitative interviews to amplify the insights of internal master students and to go beyond the findings uncovered in the report.

The report states that in April 2018, 22% of graduating bachelor students continue to study a master programme at UM, while another 20% registers for a master elsewhere. They find that the choice to continue studying immediately after graduation, is much more common for Dutch students, wherein 55% choose to continue their study directly, 41% of which at UM. In comparison, for German students, the report indicates that only 25% continue their study, of which 11% at UM. German graduates prefer to take on internships (47%) as opposed to immediately resume with their education. These are essential insights which underline the importance of student data such as nationality and pre-education.

The report also provides retention percentages of students per Bachelor major or programme. It finds that students in Accounting and Econometrics have an overall higher retention rate in comparison to other programmes, 46% and 55% respectively. Students that graduated in Finance or Emerging Markets typically don't continue their study at UM. These findings suggest that the Bachelor programme of the student is of importance when regarding student retention. Naturally, this would also imply the importance of the chosen Master programme, since some programmes would typically be a better fit for a student who already has a particular background.

Another insight which the report provides is that students were asked what their main reasons were to continue their studies. This question was asked to students who continued their studies at Maastricht and students that started a Master programme elsewhere. The results are displayed in Figure 1. In both cases, soft psychological factors were often involved in the decision to choose a Master programme. For example, under 'Table 4.4 Reasons for study elsewhere' in Figure 1, it is shown that 2 out of the top 4 reasons mentioned include soft psychological arguments, i.e., experience something different from Maastricht and experience something different from SBE. Currently, these features are not yet tracked nor incorporated in any type of marketing effort by Maastricht University.

Table 4.3 Main reasons for studying at University Maastricht (n = 41)

Reason	Frequency	% most important
Quality	37	77%
Familiar with the system	32	67%
Reputation	23	48%
International aspect	14	33%
1 year program	13	27%
Expected career chances	14	29%

* Multiple answers per student possible.

Table 4.4 Reasons for study elsewhere (n = 37)

Importance for study elsewhere	mean	(s.d.)	1 / 2 NOT IMPORTANT	4/5 IMPORTANT	Main reason, frequency
Content of the study/program	4.44	(.7)	11%	89%	14
Experience something different from Maastricht	4.02	(1.0)	16%	71%	6
Quality of the university	4.10	(1.1)	7%	75%	6
Experience something different from SBE	3.61	(1.4)	2%	61%	5
Ranking of the university	3.63	(1.3)	16%	78%	4
Location	3.68	(1.4)	16%	73%	4
Personal reasons (family friends)	2.64	(1.4)	52%	34%	3
Teaching method	3.27	(1.2)	16%	42%	1

scale (1: not important — 5: important)

Figure 1. The Reasoning for studying a Master programme

While these are already intriguing findings, we conducted qualitative interviews with students to validate or potentially enrich the research. Interviews have been held with a mix of internal and external students to see if any significant differences arise between these groups. Additionally, the interviews held with Germano Giansante generated questions we wanted to include to provide even more insights. A full analysis of the interviews can be seen in Appendix A4.

To gain a better view of why students decide to continue doing a Master programme, we asked students various questions about their thought process. External students often specify that they want to gain more knowledge and expertise on a particular topic, while Internal students specifically choose Maastricht University since they know the culture and are familiar with the university. When asked about potential improvements of information channels, some students indicate they want to encounter registered students or even have a recreation of a 'typical day in the life of X student' as opposed to sales-like talks from professors or coordinators.

Students were also asked to specify at what point in time they decided to take a Master programme. Figure 2 shows the distribution of when students made the decision. Most students choose to register for a Master programme during the third year of their Bachelor. Additionally, students were asked how long the gap was between the end of the Bachelor programme and the beginning of the Master programme, also displayed in Figure 2. These findings give an indication on which moment students would potentially be most responsive to targeted marketing by the university.

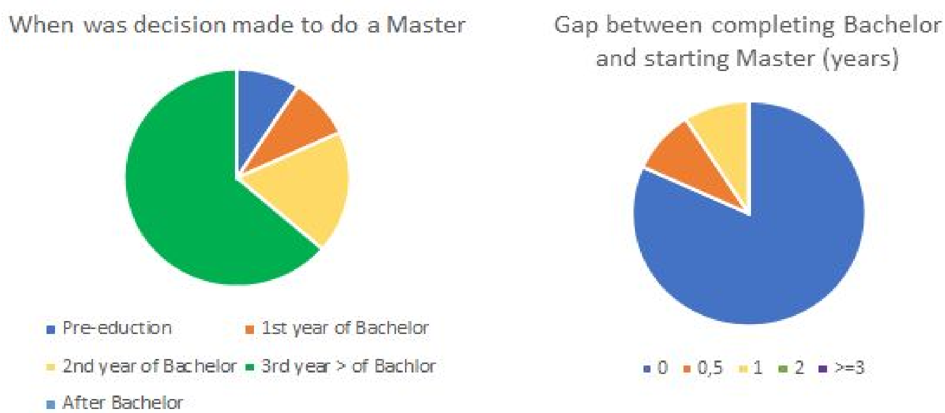


Figure 2. Partial interview results on Master programme decision and transition

In the interviews, we found that there are often softer psychological factors involved in a student applying for a Master programme. Assume a student is statistically very likely to continue a Master; however, this student wants to explore more parts of the world and goes a year abroad. It then becomes improbable to see this student in the next academic year. Psychological factors like these could have high importance in predicting whether a student chooses to stay at or leave the university after graduation. However, the university does not yet have any data on these softer psychological features.

To summarize the main research findings, it was of crucial importance to us to get a holistic view of the problem. Three interesting findings emerged from the interviews. First, internal students generally make use of the same information channels as external students. This finding indicates that internal and external students are similarly approached in terms of marketing; this includes attending open days and ordering brochures and flyers. As a result, this could potentially lead to the impression that Maastricht University does not care whether or not internal students stay.

Second, students are not aware of all the information available. When asked about different, currently available channels, like Facebook live streams or the opportunity to order flyers, many interviewees either did not use them or did not know about them at all. Third, there seems to be no difference in when internal and external students decide to apply for a Master programme. Often, the decision is made in the third year, making this the ideal time for targeted marketing on particular programmes.

Together, these insights led us to a service concept that solves both internal students' and UM staff members' needs. This concept is explained in the following section.

2.2. Concept Design

Focusing on internal students comes with the unique advantage of being able to harness data about students that is highly available for Maastricht University staff. These data include, non-exhaustively, demographics, course selection, and academic performance. As the marketing and communications department is currently completely unaware of a student's likelihood of continuing with a Master programme at SBE, the first part of our service uses data analytics to predict this probability. By taking into account students' grades, gender, nationality, age, and other available data, our models, which are explained in detail in Section 3, are designed to predict each student's likelihood of doing a Master programme at Maastricht University. With this information, the marketing department will be able to assess which students need to be targeted through various channels. For example, when our model predicts that a student born and raised in Maastricht with a GPA of above 8 in his second year has a 98% probability of staying, while students from France always have a probability of about 30% of staying, the marketing and communications team can make more informed decisions when defining target groups for advertisements, which significantly decreases waste of resources. In the form of a dashboard, which will be shown in the next section, the staff members will see (1) which attributes (like nationality, grades, etc.) affect this decision and (2) the exact probabilities that specific individuals and groups do a Master programme at UM.

However, that is not all: To address the previously discovered student needs and to substantially increase the quality of the data that is available to UM decision makers, we want to propose a second component to our service. This second component involves an addition to the Student Portal, which students already use daily to access course materials, view grades, and read about career-related offers within the tab called “My Career.” Within this currently existing section, our concept would allow students to browse through the available information on UM Master programmes, request brochures, and sign up for events. As all this information can then be traced back to the individual student, the data-quality issue regarding leads information will be solved. As a result, leads data can be reliable used for our predictive models mentioned above, and to then make highly-informed marketing decisions. Furthermore, UM staff members can use this Student Portal extension to better communicate and interact with current students by sending push notifications and displaying personalized information about events and Master programmes.

We are confident that this two-in-one solution solves both students’ and UM staff members’ needs. On the one hand, the marketing department will be able to make data-driven decisions about which students to target. With the new, high-quality data, they can make more informed decisions about which channels to use. Moreover, it will be effortless and efficient to communicate with internal students through the student portal. On the other hand, students can easily retrieve all the information available and register for events through a tool that they use daily so that they can find the perfect Master programme at UM. Next to that, they receive tailored information and, as a consequence, will feel more valued, since they are not treated like externals by the marketing department any longer. In the following section, we proudly present the development of our prototypes.

2.3. *Prototypes*

2.3.1. *Marketing and Communications Dashboard*

The information to be found in the internal students' data can only be as value-adding as our means to communicate it to our stakeholder is effective. Thus, for our primary service concept prototype, it is not only of critical importance that the data insights have a potential contribution *per se*, but also that a dashboard visualizes it appropriately. In Figure 3 below, our first prototyped Marketing and Communications Dashboard, which was created in *Tableau,* is presented. After conducting interviews and receiving feedback at presentations, we took on multiple iterations and switched from Tableau to Power BI, since the SBE recently started using this software..

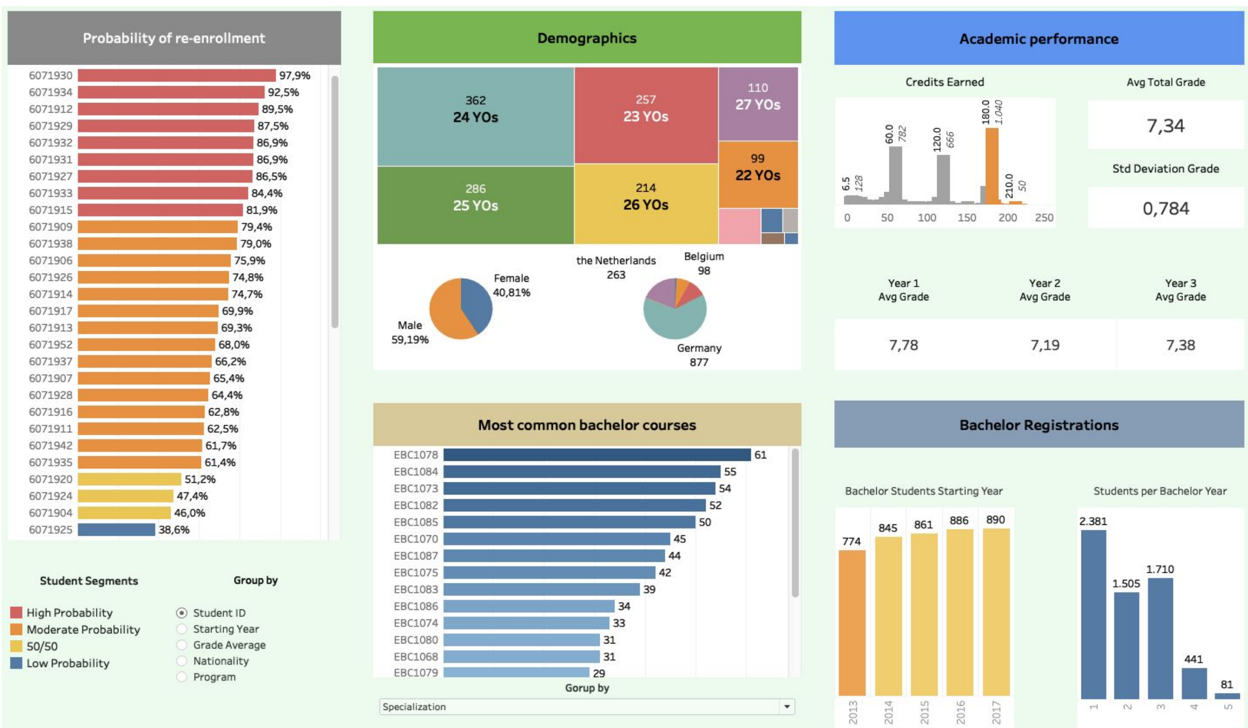


Figure 3. First Marketing and Communication Dashboard Prototype

The Tableau dashboard can be deconstructed into five different information boxes, four of which enclose visualizations and analytics relating to different student data (*Demographics*, *Academic*

Performance, *Bachelor Courses*, and *Registrations*). A fifth information box on the far left side of the dashboard exhibits the results from our predictive model (see section 3), namely the probability corresponding to each current Bachelor student of enrolling for a Master programme. Depending on these probabilities, presented in the far left horizontal bar chart, segments of students can be formed. In this case, we opted to divide them by their likelihood of re-enrollment: High / Moderate / 50-50 / Low probability segments. These segments can easily be re-defined within Tableau.

Furthermore, all five different information boxes are not to be viewed in isolation: Given that the *Student_ID* joined the underlying data, all the presented graphics and numbers are interrelated. Clicking on any of the graphics (a bar in a bar chart, a displayed number, or a piece of the pie chart corresponding to a single nationality, for example), will cause all other of the dashboard's graphics to readjust and visualize only the corresponding graphics for the selected data. This adjustment means that, for example, a predefined segment based on the predictive analytical output can be selected (e.g., High Probability Students), with the segment's descriptive analysis being produced automatically and instantly within the Dashboard's visualizations.

In our second prototype, after the first iteration, the transition from Tableau to Power BI was made, which Figure 4 depicts. To make the implementation as smooth as possible, Power BI was chosen over Tableau since the Marketing and Communications department already used this software. Additionally, the R scripts corresponding to the different Random Forest models were integrated, since Power BI allows R scripts to be a source to get data.

Figure 4. Marketing and Communications Dashboard after First Iteration

Overall, this dashboard worked similarly compared to the dashboard in Tableau. Most notably, this dashboard included a black background, to make the dashboard easier on the eyes of the end users. Additionally, we put a more significant focus on statistics like average grades and the standard deviation of grades. However, the main feedback we received at the presentation on the 29th of May was that the black background introduces a sloppiness to the dashboard. It also gave the impression that most areas seemed to be cramped up on each other.

After taking a fresh start, the third dashboard was largely built from scratch. We agreed on making different tabs within the dashboard so that each tab comes across more subtle than one big overview. Additionally, the different tabs could be focused on one specific subject. Ultimately, we agreed on having three different tabs, including Overview Classes, Characteristics, and Individual Student.

Figure 5. Dashboard: Overview Classes

Figure 5 displays the landing tab of our dashboard as part of our service. The four classes to which students can belong to each have their overview. These classes are a part of our segmentation and personalized marketing strategy, discussed in section 2.5 and include Loyalists, Hostages, Defectors, and Mercenaries. For each of these classes, the same section presents a distinct approach and strategy.

The left-hand side of Figure 5 presents the student numbers with the corresponding probability of staying indicated by our models. The bars can be clicked to take a closer look at one specific student since all presented graphics and numbers are again interrelated. The right-hand side shows general statistics per group, including average grade, mean age, average grade improvement, and the number of students in this class. The average grade improvement corresponds to the difference in grades obtained during the last and first years of their Bachelor, which will be further discussed in chapter 3.

Generally, students that are most likely to stay for their Master, the loyalists, have the lowest average grade. This average gradually increases as the probability of staying decreases. In contrast, these students experience the most considerable improvement in their grades. Students that belong to the mercenaries class are generally the youngest. This finding could indicate that younger students want to further explore other possibilities before starting a master. Of course, this could potentially be the result of much softer variables and considerations that are at stake. Hence, we opt for the generation and inclusion of softer psychological variables to investigate these differences further.

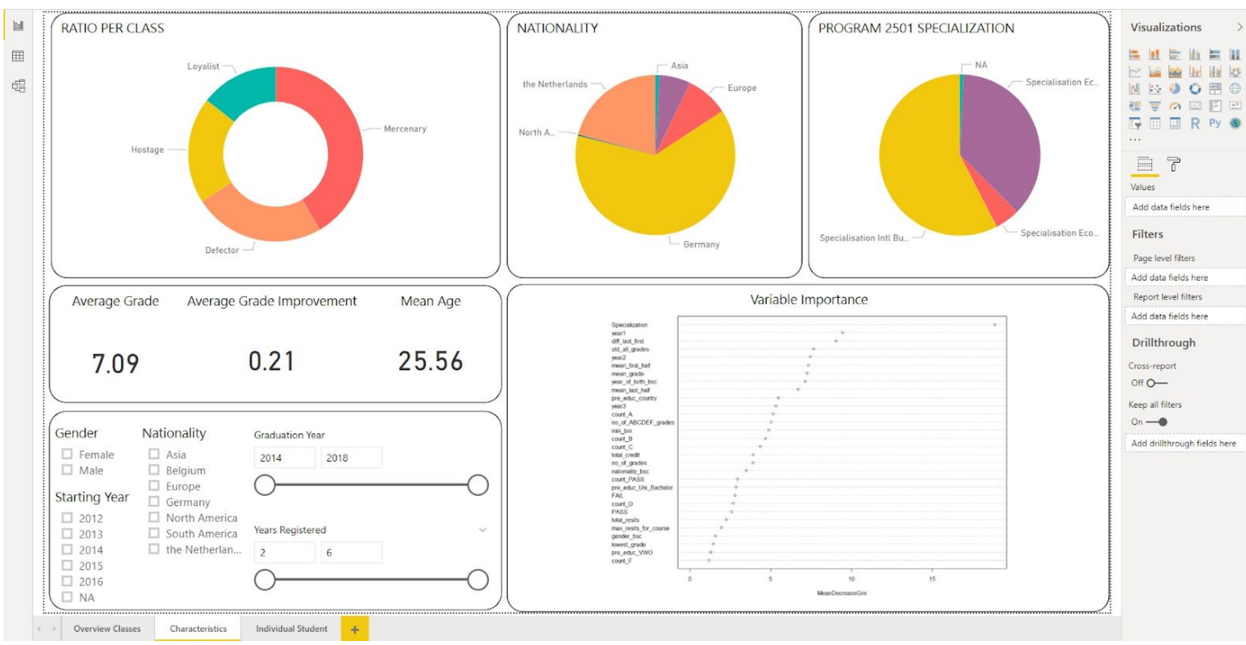


Figure 6. Dashboard: Characteristics

The characteristics of the various groups of students can be assessed in the tab ‘characteristics’, displayed in Figure 6. In the bottom left corner, the user can filter on various variables to gain more insight into a specific group. Based on the customization applied here, the rest of the dashboard will modify itself accordingly. Additionally, the charts at the top of the tab can also serve as a way to filter the data. We found that one of the most critical variables in deciding whether a student will apply for their Master includes the type of specialization the student chose, which the variable importance plot displays in the Analytics part of the report. Therefore, we incorporated an extra pie-chart that shows the distribution of specialization of course 2501.

The use of this tab lies in the user's ability to explore the characteristics of students filtered on various variables easily. It also serves as a method to better understand the data by visualizing and highlighting different groups of interest.

The final tab in our dashboard concludes the 'individual student' tab, displayed in Figure 7 and 8. This tab focuses on presenting an overview of one specific student by collecting statistics on this student. On the left-hand side, all the students are displayed with their corresponding probability of staying. This section can be filtered on by either clicking on a student or searching for a student number. After a student is specified, the rest of the dashboard's visuals will change according to that student. In the middle, various demographics are shown, including the average grade and the class of the student. More insights on the probability of staying for this student lies on the right-hand side of the dashboard.

Figure 7 exhibits a student with a probability of 27% of staying after completing his Bachelor. Around this value, the top 5 variables that make this student more likely to either leave or stay are shown. For example, this student has a specialization in International Business Economics, which is an essential indicator for students that remain at the university (increasing the probability of staying by 10,1%). However, this student performed worse during his Bachelor, indicated by a low average grade in the last half of his Bachelor (decreasing the probability of staying by 9,1%) and an average grade improvement of -0,41 (reduce the likelihood of staying further with 6.8%).

Figure 7. Dashboard: Individual Student with a low probability of staying

Figure 8 displays another example of a student with a probability of 77% of staying. This student has many grade related variables that push his likelihood of staying more towards 1. In contrast, this student has only one value for a specific variable that drives his probability of staying towards 0 (having a maximum of 5 resits for a course decreases the likelihood by 3%).

These plots should give the user of the dashboard more insights on why a student has a certain probability of staying. Our models are considered black box models referring to the fact that often there are no insights on what is going on in the modeling itself. With this tab in the dashboard, we try to overcome this fact by actively showing which values for what variables have a relatively high influence on the probability of staying.

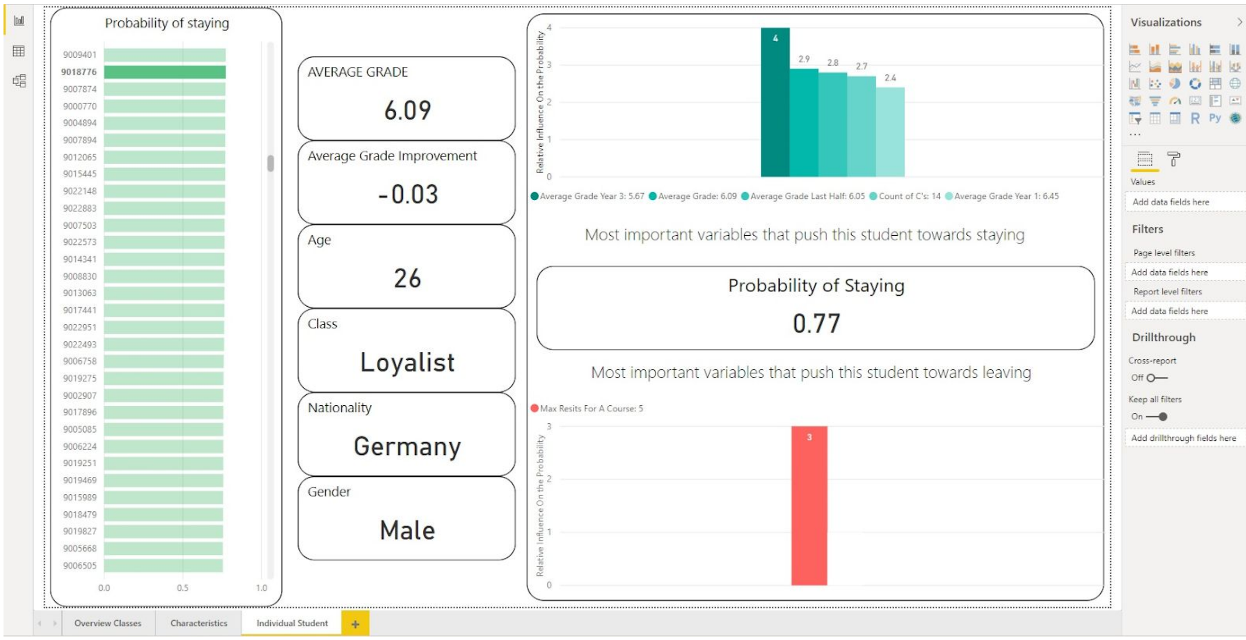


Figure 8. Dashboard: Individual Student with a high probability of staying

With this final dashboard, the Marketing and Communications team can, within seconds, review important student segments to target them with marketing campaigns for student retention and visualize descriptive and predictive information on these segments to make these campaigns more effective. As the visualized information relates (with potential causality) to internal bachelor students' choice to stay or leave for their Master studies, a more precise understanding of what makes students opt for a Master study at UM is brought about: This does not only answer the question of *which* bachelor students are most likely to stay/leave but also *why* they might do so.

2.3.2. Student Portal Extension

This section explains and visualizes the prototype for the student portal extension. The prototype is currently online and can be tested out by actively "clicking-through" the website. In the following, the concept is explained step-by-step with screenshots from the prototype.

Students log into the student portal with their student numbers and corresponding passwords. This functionality will give us the unique advantage of tracing back the effectiveness of different channels to the individual student.

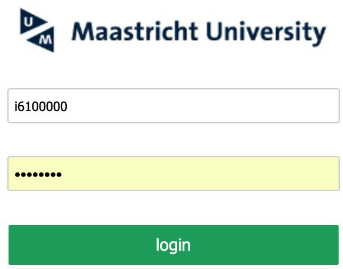


Figure 9. Login

After the login screen, students currently have different tabs to choose from, like “My Courses,” “My Timetable,” and “My Webmail.” Under the recently introduced section “My Employability” or as a separate tab, we want to add a new section called “My Career.”

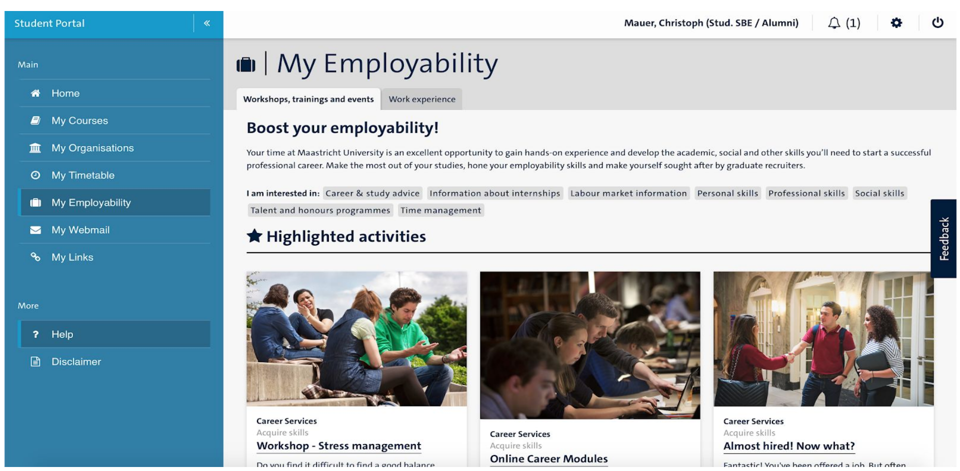


Figure 10. The current Student Portal.

When opening the proposed section “My Career”, students will see a short greeting and an introduction that explains the concept. This greeting will give them a feeling of comfort and being valued.

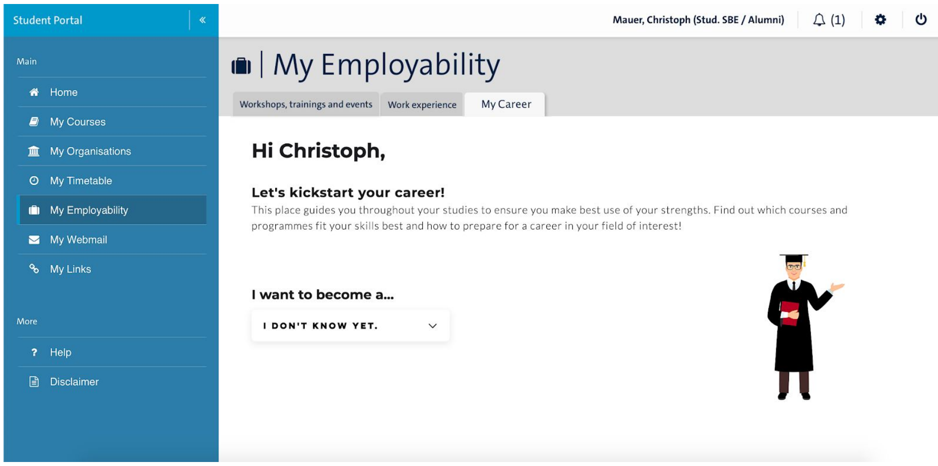


Figure 11. The start screen of the proposed section "My Career".

As a form of data collection, students will be able to indicate their current career preferences (if they already have some). This way, the marketing department can send out tailored information about available programmes that fit exactly these preferences.

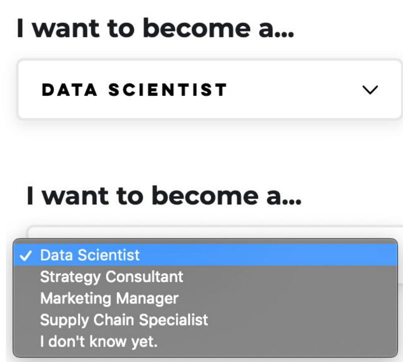


Figure 12. Choosing a career preference

When scrolling down, students could directly view information on Master programmes that fit their interests and that match their academic performance. The groundbreaking advantage of this solution is that the marketing department can again use this to collect high-quality data about individual students. They can trace back who ordered brochures, attended events, or browsed through the information available online and see how these activities relate to the decision to follow a Master programme at Maastricht University. For the students, this will be highly useful as well, as they can easily retrieve information or sign up for events that match their interests while having the option to browse through other available programmes always available. It needs to be mentioned that this recommendation algorithm, although extremely easy to implement, can not be developed based on the current highly anonymized data, as there is only one master programme in the data set. With access to raw data, this algorithm is a simple extension of the previously mentioned probability model, as explained in Section 3.

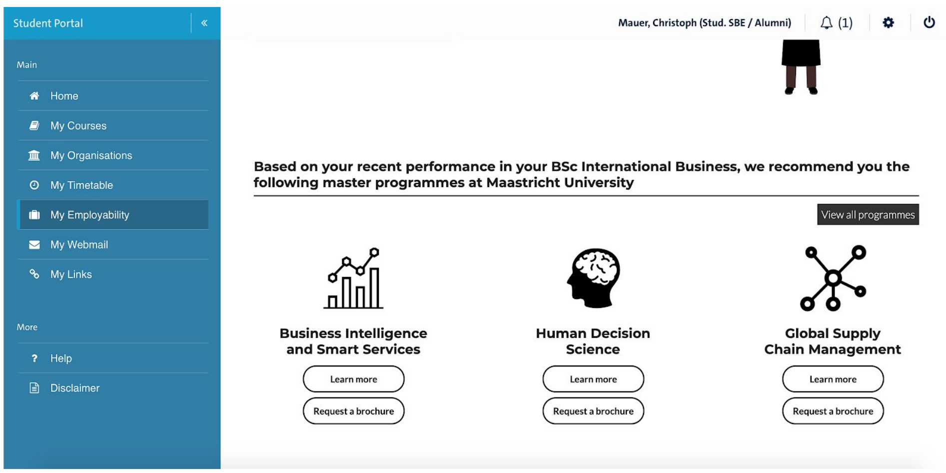


Figure 13. Information Retrieval

Since students pointed out that it is not always clear where information is available, this section can be used to display upcoming events and offer the option to sign up for them. For students, this is very convenient, as it is perfectly integrated into the student portal, which they open almost every day. For the marketing department, this is again a fantastic way of collecting leads data to gain more insight into the effectiveness of channels.

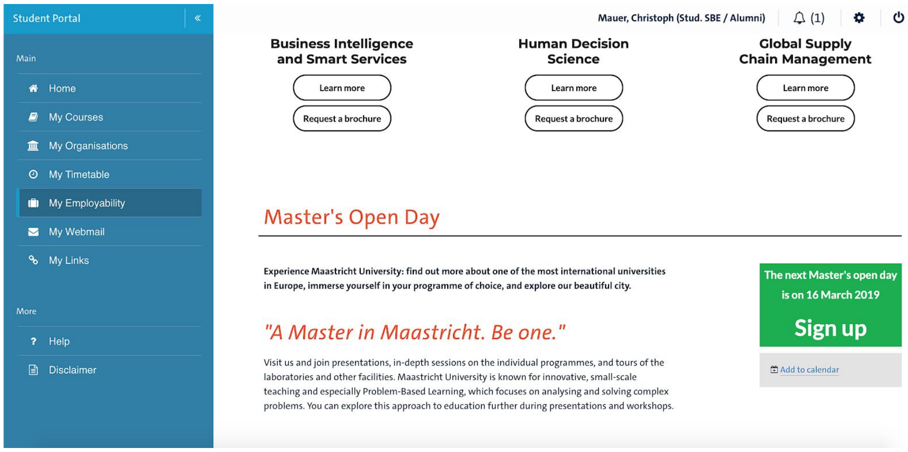


Figure 14. Events

In sum, we are highly confident that this two-in-one service concept solves both students' and UM staff members' needs with an easy-to-implement but a highly sophisticated solution. The next chapter describes how business analytics makes this service concept works by introducing and explaining the underlying models.

2.4. GDPR

With the implementation of the General Data Protection Regulation (GDPR) in 2018, privacy protection has become a necessary and essential consideration when developing data-driven services. Important to note at this point is that the data used in our service, like grades, gender, and nationality, are already available at Maastricht University. What is changing is the purpose for which data is processed, which is crucial because the GDPR states that data subjects (i.e., students) must consent to the analysis of their data: "GDPR consent must be freely given, specific, informed and evidenced by clear affirmative action. [...] Processing of data is fair only if it is transparent and this means there must be openness in data processing through effective communication with individuals including in the use of information notices" (Goddard, 2017).

To this end, we can leverage the Student Portal extension once more. When a student opens the new page “My Career” for the first time, she can opt-in to our service. Figure 15 shows the opt-in screen in our prototype.

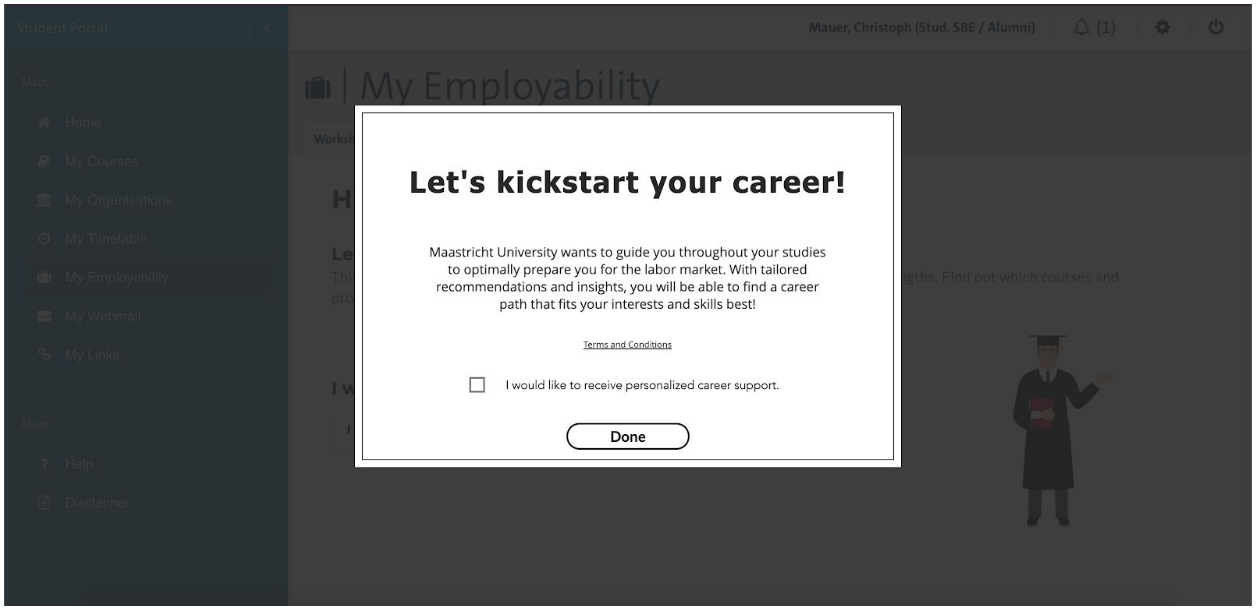


Figure 15. GDPR Regulation Screen

We are confident that students will openly opt-in for this service, as it is in their best interest to get support for their careers. Next, GDPR requires us to offer the option to opt-out of the service and to inspect the collected data (Tankard, 2016). This option could quickly be done in the settings of the Student Portal.

In practice, this can be very easily implemented. In our R script that models our data, one could e.g., filter for the student numbers that opted out of the service and do not fulfill the analysis on their data.

Even in those opt-out cases, one can use fundamental triggers from the data analysis of other students for tailored marketing. If the model shows that people with specific characteristics (i.e., a combination of gender, nationality, and GPA) have a certain likelihood of choosing a master programme at UM, one can react to students possessing these characteristics without having to calculate an individual probability of staying. The next section explains in detail our recommended way of using our service.

Since we are considering collecting new data via the Student Portal (especially concerning leads, e.g., brochure orders and attended events), there are a few other considerations to make. Beckett (2017) points out that an organization that processes data should document where data is stored, how sensitive personal information is handled, etc. Since these regulations also affect the data and analytics held and performed by the university in general, we are more than confident that our service will not cause any disruptions to the already existing legal spectrum.

2.5. Recommended Use - The Service in Practice

Although students' likelihood of choosing a Master programme at Maastricht University and factors that contribute to this decision are intuitively desirable to obtain from a marketing point of view, it might be less intuitive to "use", i.e. to act upon the insights gained through our service. From our second interview with Germano Giansante, Student Recruitment Advisor at SBE, it became clear that the marketing department currently does not use any tailored marketing strategies. This finding means that every student, regardless of his characteristics, academic performance, and history at UM, is treated the same way by offering a generic mix of marketing materials and communication channels. Germano confirmed that, therefore, the marketing department needs guidance on how our service can contribute to the tailoring marketing efforts to the previously mentioned insights derived from our service.

Before elaborating on our recommended use, we want to highlight why personalized marketing is essential and superior to the generic marketing that is currently in place. In general, customer heterogeneity is a crucial problem to consider and handle in marketing (Hahn et al., 2002). To address these differences, segmentation helps organizations to "divide large, heterogeneous markets into smaller segments that can be reached more efficiently and effectively [to] match their unique needs" (Armstrong & Kotler, 2015, p.199). According to Tran (2017), personalized marketing is beneficial for both the marketer and the customer. From the customer perspective, in the case of a university the student perspective, tailored marketing guarantees exposure to relevant information only, resulting in less time wasted and higher satisfaction (Bleier and Eisenbeiss, 2015). From the marketing department's point of view, personalized marketing is more cost-efficient, as no money is wasted on advertising to customers (students) who are not interested in the offering (Kim et al., 2001).

For our recommended course of action, we combined two streams of research: (1) Customer churn management and (2) Student retention interventions. On the one hand, customer churn management is concerned with, among others, predicting customers' likelihood of being loyal to a company and taking action upon these predictions (Provost & Fawcett, 2013). On the other hand, student retention interventions aim at keeping students satisfied and preventing them from leaving the university (Thammasiri et al., 2014). We believe that by combining these two streams of research, we can recommend a way of using our service that helps the marketing department attracting more internal Bachelor students to do a Master programme at UM.

Figure 16 displays the integration of data mining into decision-making, as proposed by Lejeune (2001).

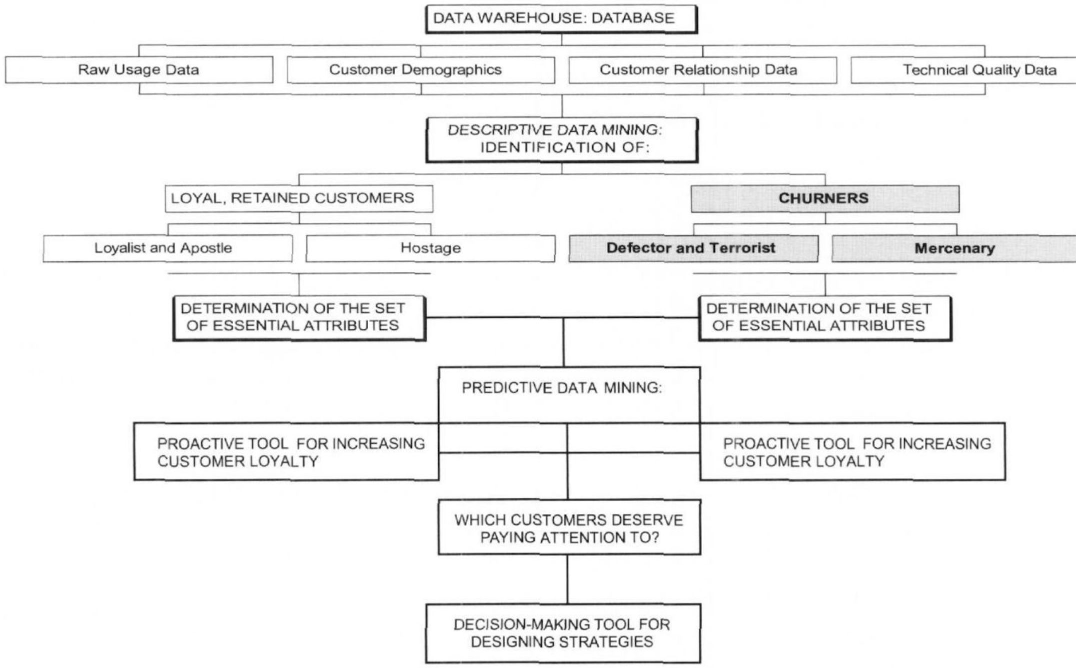


Figure 16. Integration of data mining into decision-making

What stands out is that based on data mining, customers are divided into four segments, namely: (1) Loyalists and Apostles - loyal and satisfied customers, (2) Hostages - not entirely satisfied; can either stay or leave, (3) Defectors - can be transformed into loyalists; need special attention , and (4) Mercenaries - are out of control for the company (Lejeune, 2001).

The idea here is that customers are segmented to decide then which segments to handle in what way. Based on this workflow, we want to propose segmenting students into four groups and then using insights from student retention research to supplement this with concrete actions. In practice, this could look as follows:

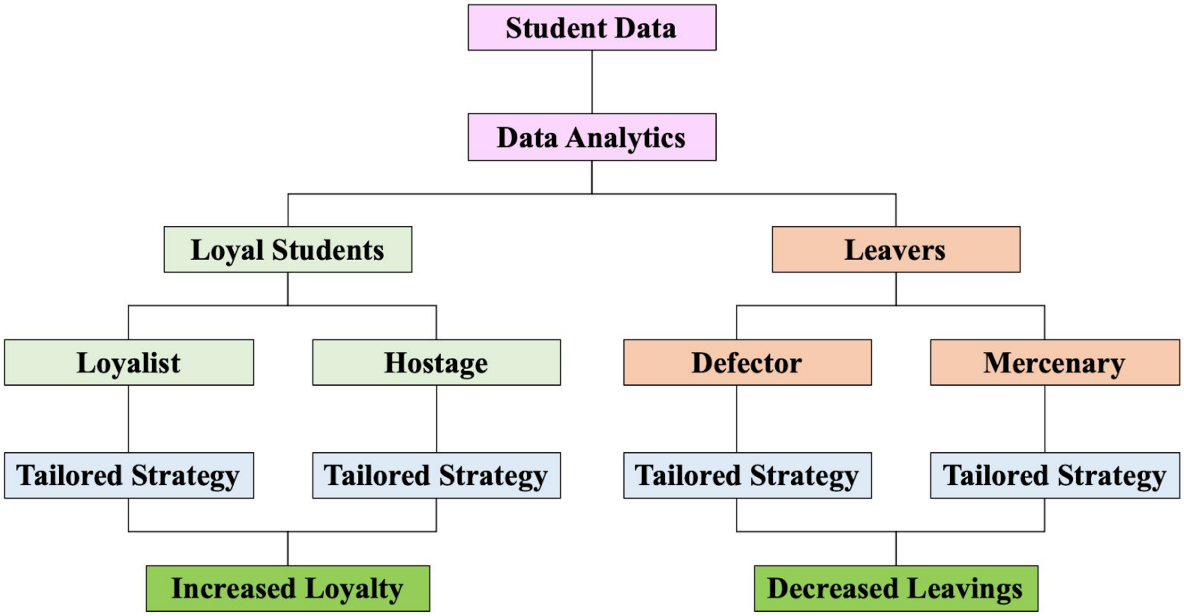


Figure 17. Adapted workflow of integrating data mining into personalized marketing

In the first part, student data is mined and analyzed using machine learning, which chapter three extensively explains. Based on the expected probability of staying (in brackets), students are divided into four groups: Loyalists (100-75%), Hostages (74-50%), Defectors (49%-25%), and Mercenaries (24%-0%). In the dashboard, this could look like this:

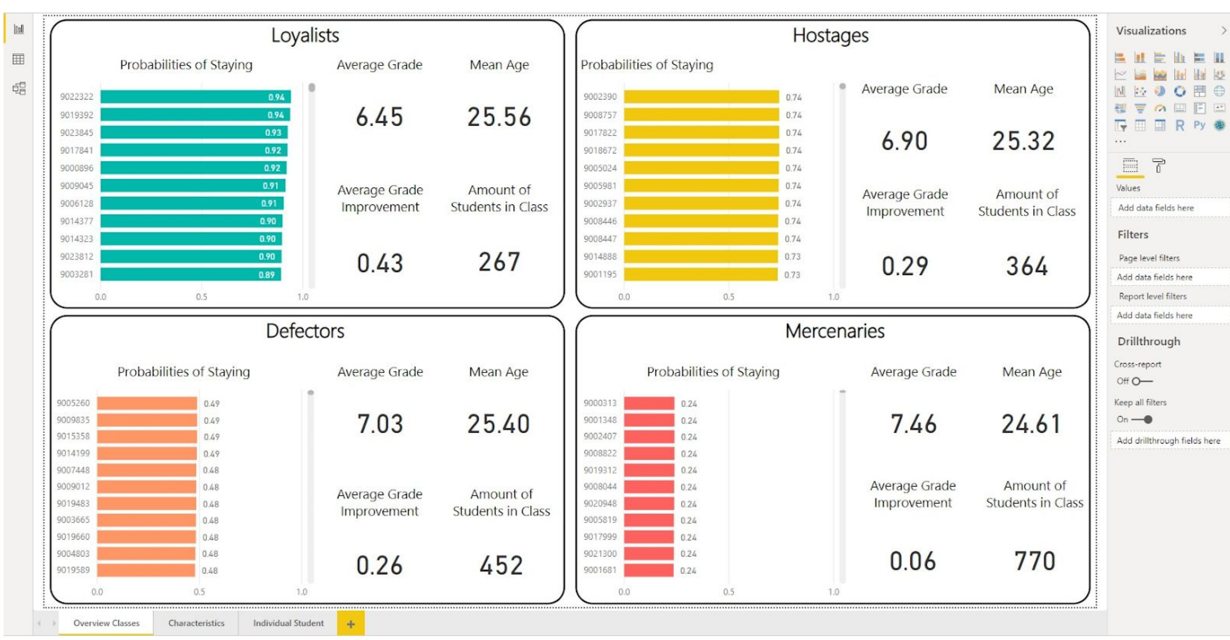


Figure 18. Dashboard: Overview Classes

Of course, over time, as the marketing team gets more experience with the effectiveness of personalized marketing, these segments could be extended and fine-tuned, potentially including different sub-segments.

Based on this segmentation, insights from student retention intervention research can be used to get as much value as possible from internal Bachelor students. Our suggested interventions are largely based on Hanover Research (2014). The expertise of the marketing department can supplement these suggestions and fine-tune over time.

The **loyalists** are likely to be very satisfied with Maastricht University and are therefore have a high probability of doing a Master programme. These students could be informed about upcoming open days or available brochures and other information via the Student Portal or email newsletter. Also, they could receive one or two brochures about relevant Master programmes, based on e.g., their academic performance and field of study, to their home address automatically during the second and third year. Thereby, these loyal students get the feeling that they are valued and that the university cares about them. For the marketing department, sending the brochures only to a limited segment can save costs, a significant advantage of segmented marketing.

The **hostages** are not fully satisfied and are less likely to follow a Master programme at Maastricht University. To this end, they need to be convinced more strongly than the loyalists through various channels. They could, for example, receive significantly more notifications through the Student Portal about the available programmes. One could even think about email campaigns (e.g. presenting one Master programme per week for a certain period) that are directly targeted at these students. At first glance, one could think that it makes sense to send this email campaign to all students. However, Germano Giansante pointed out that the marketing department does not want to “annoy” students who already made a decision. Therefore, this could be an effective way of targeting only students who are likely to be affected by this.

The **defectors** tend to not do a Master programme at UM. However, Lejeune (2001) states that they could still be convinced with high effort. To convince these students, one could think about intervention mechanisms like a mentoring programme. In this programme, trained coaches (like coaches in the Study Coaching Trajectory) could have office hours for internal Bachelor students to talk about their concerns or to give them extra information on Master programmes in 1-1 conversations. The defector-segment could be informed about these potential mentoring sessions regularly via the Student Portal. According to Hanover Research (2014), these "student support services can have a measurable, significant, positive impact on student retention" (p.3). Even if a student cannot be convinced, the marketing team could still gain essential insights on what drives her decision in a one-on-one conversation.

The **mercenaries** can most likely not be convinced to stay (Lejeune, 2001). To this end, the marketing department could use uncostly channels like generic email marketing or notifications only, to not waste any resources on trying to retain these students.

Important to notice is that the retention interventions should not be exclusive to specific segments, to avoid that students feel treated differently or even unfairly. In other words, this means that e.g., every student could have access to the buddy/mentoring programme and that every student could receive notifications about open days via the Student Portal. The key idea is, however, that each segment is frequently exposed to marketing channels that are most effective for this specific segment. While the hostage-segment could receive a student portal notification about a potential mentoring programme e.g., twice per semester, the loyalist-segment does not obtain any particular notifications about it, but could still find it on the website when needed.

We are convinced that this is a cost- and resource-effective way of doing personalized marketing to internal Bachelor students. As mentioned before, the marketing department can, of course, use the insights gained from our service for other retention mechanisms. Our goal here was to demonstrate how the service could work in practice, as the SBE does not currently use any personalized marketing.

In general, from our desk research, it seems that student retention is a complex and difficult-to-execute topic, which requires manpower and resources: "Many institutions lack a full-time coordinator for these programmes. Even institutions with retention coordinators seldom give these coordinators authority to launch new programs or fund new initiatives" (Hanover Research, 2014, p.3). Together with our service, we, therefore, recommend considering adding the full-time position "Retention Coordinator" to the SBE staff portfolio. This way, student retention inside the Bachelor and Master programmes, but also during the transition from Bachelor to Master could be improved. With this additional manpower and allocated resources, our service could become even more useful for the marketing department, as experts can continuously work out retention initiatives based on the insights from our service.

2.6. Critical Reflection

We believe that reflecting critically on the solutions and insights is provided to guarantee a smooth implementation of our service at SBE. Therefore, we dedicated a separate subsection of each chapter towards critical reflection.

In general, we are highly confident that our service addresses the needs of the marketing department very well. Nevertheless, there are some important considerations to ensure that it can be implemented fluently. From employees working at the SBE, we heard that the team that is developing and maintaining the Student Portal is relatively slow in adding new features. Therefore, it is advisable to involve them as early as possible to avoid delays in development. Next, it might be desirable to engage the legal department early in the process too. Although we did extensive research on any GDPR related issues, having the confirmation from UM lawyers is essential also. Furthermore, it is necessary that the marketing team recognizes the vast potential for personalized marketing and communicating with internal students by actually acting upon the insights provided. This potential could also require some iteration over the next years to find out which marketing efforts are perceived best by the different segments. Finally, the availability of some form of analytics support might be necessary in case the marketing team encounters any technical issues with the dashboard.

All in all, receiving extremely positive feedback from our main stakeholder Germano Giansante was very motivating and made us believe that our service will be a valuable addition to marketing for and communicating with internal students.

3. Analytics Stage

Based on the fully-developed service concept and the prototypes, we started with selecting methods to put the smart component of our service into practice. Before building and validating the actual models, we cleaned the data extensively to guarantee high-quality results. To account for the vast number of hours that were spent on data cleaning, a separate section explains this process. The following elucidates the individual steps of this phase.

3.1. Data Cleaning

Our group is developing a service for the Marketing & Communications department of the School of Business Economics at Maastricht University. More specifically, our main task was to focus on the Marketing & Communication channels for internal MSc student recruitment.

A necessary component in data cleaning is in understanding the context and requirements underlying the project. The first step involved meeting with our stakeholders of the Marketing & Communications department and discussing our questions and their pains and gains. We concluded from the meeting that internal students are currently targeted similarly to how they target external students. However, the stakeholders also mentioned a doublethink dilemma wherein they want to apply traditional marketing tactics to upsell bachelor students into Master students and stay in UM, but at the same time, part of their marketing strategy communicates the message of exploring the world. The Marketing & Communications department was operating with a cognitive dissonance that if detected, could be criticized for irony. A possible workaround to this dilemma is employing more subtle and targeted marketing tactics with the help of data insights that could enable the marketing department to retain their public message but still pursue converting bachelor students into master students.

In the current recruitment funnel, many exciting channels are present, but, there are only limited channels in place that take advantage of the fact that internal students are already within reach and more straightforward to target. One example of an existing channel within the current internal recruitment funnel is a non-personalized newsletter in the student's email. Besides this, available information and data on the internal students are currently not exploited to improve the channels in place continuously. Systems do not seem to communicate with each other in each situation. The recruitment data (the 'leads' dataset with brochure and events data) had, for example, major data quality issues. For a considerable number of students, it is unknown whether they requested a brochure for a BSc or MSc programme at Maastricht University. According to the recruitment data, only 282 internal students attended an event at Maastricht University (such as the open day). Proper identification of the student (by student number) was only possible for 111 of these 282 students. By making better use of internal channels that are already in place (precisely, the Student Portal), we could solve these problems to the benefit of both students and internal stakeholders. In conclusion, there is room to improve the current internal recruitment funnel for MSc students.

On the one hand, we were lucky with our target group (internal students) since many data was made available to us about these students. On the other hand, because our goal was to use every bit of the provided information, it meant that there was quite some cleaning work to be done. We ended up cleaning and using every provided dataset, and we transformed this data into variables that could be of interest for predicting whether an International Business BSc-student or an Economics and Business Economics BSc-student would end up applying for an International Business MSc programme at Maastricht University.

We see this project as a proof-of-concept study that would be easily extendable to other BSc programmes and other MSc programmes.

3.1.1. Original Data

This section will discuss the data cleaning process in more detail. To better understand the data, we were in close contact with Stan van Hoesel and Nico Rasters and asked them to clarify any uncertainties in the data.

Initially, data of 23,563 records, consisting of confirmed students and those who applied, in two different BSc programmes (IB and EBE) and one MSc programme (IB) was made available to us. For our service, we were mainly interested in BSc students to predict whether they are likely to continue their MSc at Maastricht University. In the dataset, 16,065 (of these 25,585 students) were bachelor students. Only 1,592 of these bachelor students also appeared in the Master's dataset.

3.1.2. The Need for Extra Data

First of all, while cleaning the academic work data of bachelor students, we stumbled on a problem: for many BSc students, only data for 1 or 2 years of data were available. We decided only to include students for which data on the total BSc programme was available to us (3 academic years or more and ≥ 180 credits). An alternative approach which was also considered is to check whether they applied and got accepted in the same academic year. The alternative approach was extracted by filtering out the students in the admissions sheet that applied for a main programme, did not apply for just a re-registration, got accepted, and appeared at least once in the registrations sheet. The alternative approach included students that did not finish their bachelor in UM but still applied for masters in UM. Both approaches are being used and applied concurrently to give greater choice in the future during modeling.

We only included students for which we have a full BSc programme because we cannot construct reliable variables for students that only have one year of data about grades (such as mean grade, or grades for specific courses). It will create a skewed view of the effects of other variables (like the number of applications, years registered, and credits earned) if the data on every variable is not available in all the years that a bachelor student was studying. So, we constructed a variable to check how many credits a student achieved within our time frame of data (2013-2017), and if this number was ≥ 180 credits, we include this student in our dataset. Theoretically, we could include students who started in 2013, 2014 or 2015 with their BSc (see Figure 19).

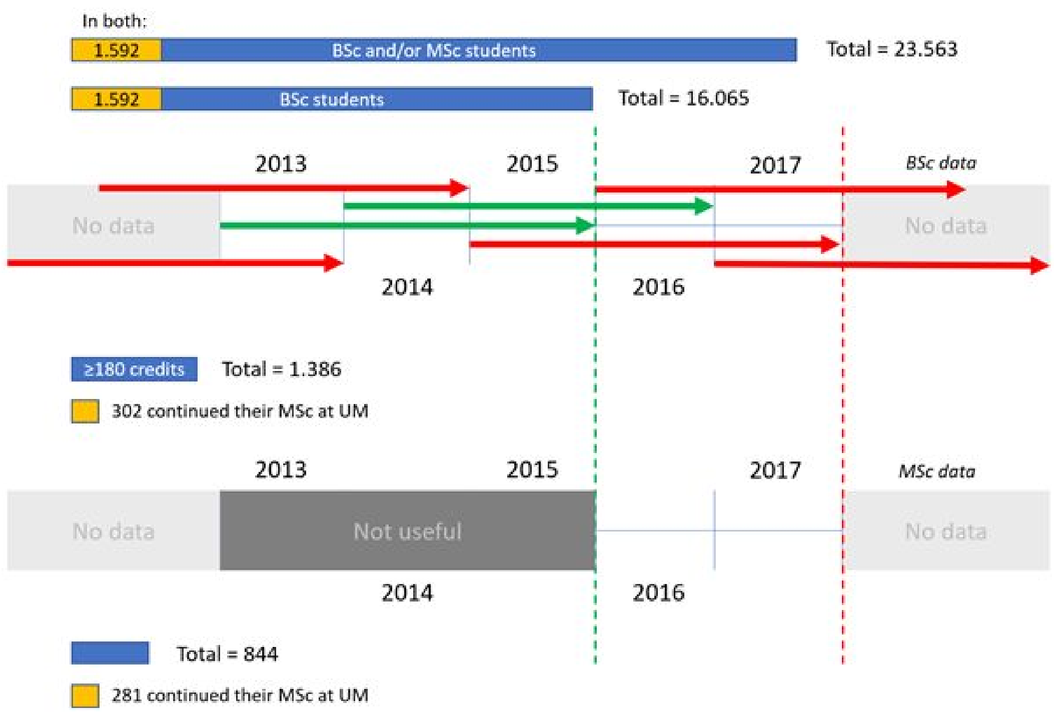


Figure 19. Data shortcomings

Using this approach, only 1,386 BSc students (of the 16,065) for which we had data on the whole BSc programme remained. Only 302 of the 1,386 bachelor students also appeared in the Master's dataset. However, we discovered another issue with this approach: students are included now who graduated in 2017, but we do not have MSc data of the year 2018, so these students were never able to start an MSc programme within our timeframe (2013-2017). Therefore, we also had to exclude the students who graduated in 2017 and beyond since they will not give us any informative value. The number of students we could use dropped to 844, of whom 281 continued their MSc at UM. The first year that included students that could possibly graduate was 2016 in our dataset, and this made the Master's data of the years 2013, 2014 and 2015 not useful for our purposes (see Figure 19).

In our opinion, the drop in the number of students we could include was too big and prompted us to think about ways to solve this problem (see Figure 20).

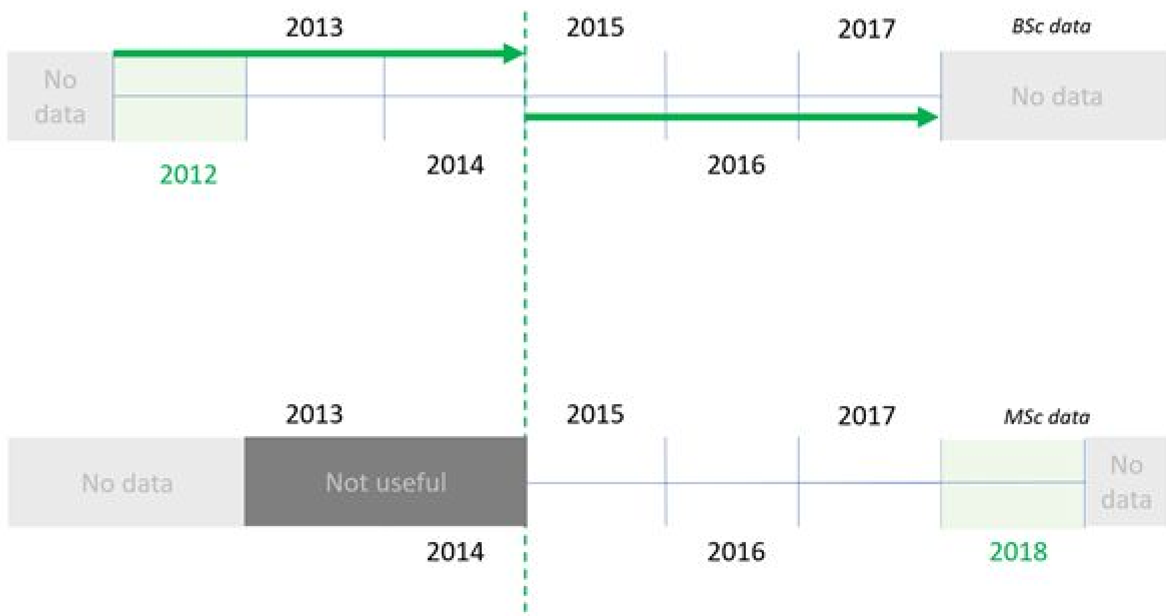


Figure 20. Need for extra data

If we could get BSc data of the year 2012, this would give us another full batch of BSc-students that can be used (BSc from 2012 till 2014) plus it would make the already provided MSc data of the year 2015 useful. Also, with MSc admission data of the year 2018, we could add another batch of BSc-students to our dataset (the one we had to exclude earlier because they were not able to apply and register for an MSc programme within our timeframe). An additional concern that we raised is that despite using ≥ 180 credits as a reliable proxy for whether a bachelor student graduated, a dataset stating directly if a bachelor student graduated would be more accurate. We, therefore, requested BSc data of the year 2012, MSc data of the year 2018, and data of BSc graduates. All our requests were honored.

This new data was of tremendous value for us because this step enabled us to include 2,355 BSc students in our models, an increase of 279% over the previous number. Of these bachelor students, 902 continued doing an IB MSc programme at Maastricht University. So, 38% of the IB & EBE BSc students continued their MSc at Maastricht University. This figure is a realistic number since it is close to the estimate mentioned during the interview with the Marketing & Communications stakeholders (30%) and the numbers we got from the earlier mentioned report ("Perception and Position of Bachelor Graduates Cohort 2016-2017"), which also confirms that our approach was valid and reliable (see Appendix for the interview notes). The figure also means that our potential training dataset is not too unbalanced. Figure 21 below shows the Sankey diagram of the journeys of all the recorded bachelor students and prospective bachelor students in the dataset. The figure gives a clear picture of how much of the bachelor students advanced into the next stage.

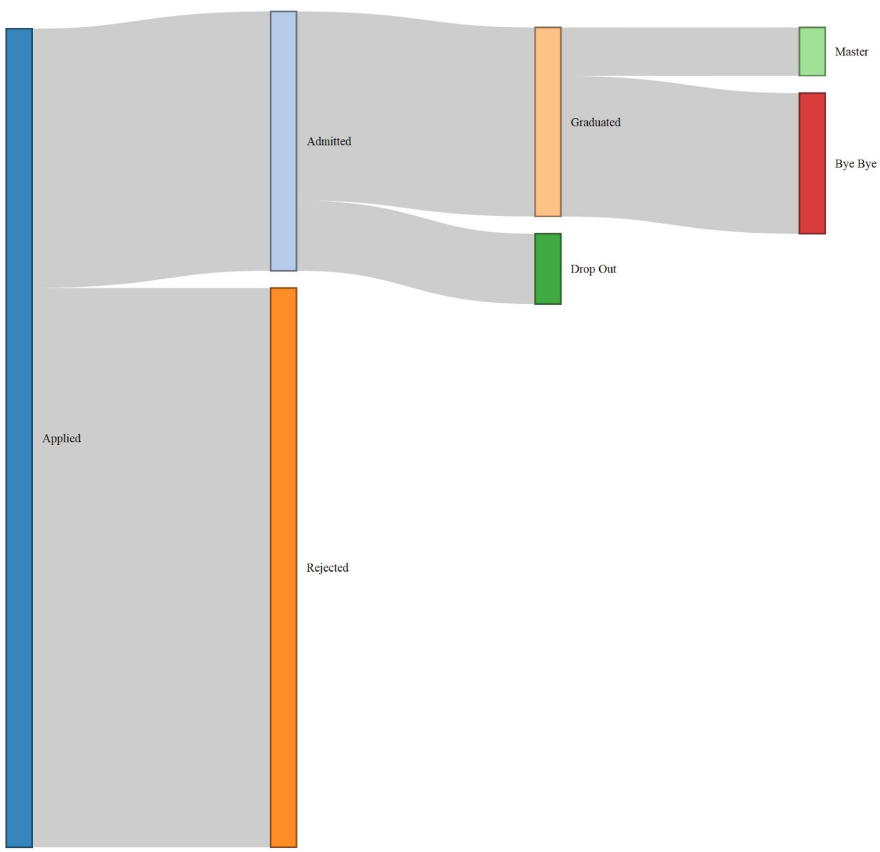


Figure 21. Sankey diagram of the student journey

We will now discuss the cleaning process in more detail, file by file. Our primary goal was to construct as many variables as possible to find any associations with our target variable (whether a bachelor student with a complete record also appeared in the master registration sheet).

3.1.3. BISS data - Bachelors - Student version.xlsx + Related Extra Data

We will start by discussing this file since this dataset contained the most important information for our purposes. The outline was to create one row per student number, which is the main identifier in the dataset.

For the cleaning of the Academic Worksheet, included were only BSc students for which we had data on the whole BSc programme (≥180 credits). Regarding the academic worksheet, we used grades for appraisal type 7055 and appraisal type 10. Appraisal type 7055 is a composite grade for the whole course (exam, participation, papers, and so on) and is on an A-B-C-D-E-F scale. Appraisal type 10 was the final grade for the course, but for most courses, this appraisal type only had PASS/FAIL. Therefore, we preferred appraisal type 7055 over appraisal type 10 and only used type 10 if no appraisal type 7055 was available for that specific course (PASS/FAIL courses). So, we removed all the PASS/FAIL grades if there was also an A-B-C-D-E-F grade for that specific course available since this grade gives us more information than only a pass or fail. For each student, a large number of variables were constructed using this dataset (see Table 1). We also included general variables regarding how the admissions process of each student went and how many years they registered as bachelor students in UM. The variables constructed using the 'BISS data - Bachelors - Student version.xlsx '-file, along with the extra data related to it, in Table 1 with an orange background.

Table 1: Variables

Dataset	Variable	Explanation
‘*BISS data - Bachelors - Student version.xlsx* ’ + *‘BISS data - Bachelors 2012 - Student version.xlsx’* + *‘BISS data - Bachelors Academic work 2012 - Student version.xlsx’* + *‘BISS data - Bachelors Admissions 2012 - Student version.xlsx’*	Student ID	This is the main identifier of the dataset. One row represents one student.
	Number Applications	This is the number of times a student applied for a bachelor’s degree in UM. This column doesn’t count the times a student applied for re-registrations and an additional programme. It counts only when a student applies for the first time to be a freshman in the programme.
	Approved	One means that the student’s bachelor application was approved. Zero means the student either rejected or has withdrawn. This doesn’t count how many applications a student was approved in. Some students were approved in two applications, but it would only be written as one.
	Years Registered	This counts how many years the student has registered as a bachelor student. This is calculated by counting the number of times a student number appeared in the Bachelor Registrations Sheet.
	Starting Year	This is a proxy on when the student’s first bachelor year in UM was. This is calculated by collecting all the years a student was registered as a bachelor student and getting the earliest year.
	Graduation Year	This is a proxy on when the bachelor student graduated from UM. This is calculated by collecting all the years a student was registered as a bachelor student and getting the latest year and adding one. This assumes a graduation that takes place the year after the start of an academic year.
	program.x	Number of BSc programmes (1 or 2)
	program.y	Which BSc programme(2501, 2504 or both)
	total_credit	Total number of ECTS achieved during time frame (≥180)

	gender_bsc	Gender of student
	nationality_bsc	Nationality of student
	year_of_birth_bsc	Year of birth of student
	count_A	Number of times an A was scored for a course
	count_B	Number of times a B was scored for a course
	count_C	Number of times a C was scored for a course
	count_D	Number of times a D was scored for a course
	count_E	Number of times an E was scored for a course
	count_F	Number of times an F was scored for a course
	count_EXCELLENT	Number of times an EXCELLENT was scored for a course
	count_PASS	Number of times a PASS was scored for a course
	count_FAIL	Number of times a FAIL was scored for a course
	count_NO_GRADE	Number of times a NO GRADE was scored for a course
	n_of_grades	Total number of grades in timeframe
	mean_grade	Mean grade for total BSc programme
	highest_grade	Highest grade for a course
	lowest_grade	Lowest grade for a course
	std_all_grades	Standard deviation of grades for all courses (variation in grades)
	mean_first_half	Mean grade calculated over the first 50% of grades
	mean_last_half	Mean grade calculated over the last 50% of grades
	diff_last_first	Difference between the last and first 50% of grades (positive number means better performance in last half)
	year1	Mean grade in year 1 of BSc programme
	year2	Mean grade in year 2 of BSc programme

	year3	Mean grade in year 3 of BSc programme
	year4	Mean grade in year 4 of BSc programme
	year5	Mean grade in year 5 of BSc programme
	year6	Mean grade in year 6 of BSc programme
	more_than_three_years	Dataset contains grades for more than 3 BSc years for this student
	more_than_four_years	Dataset contains grades for more than 4 BSc years for this student
	more_than_five_years	Dataset contains grades for more than 5 BSc years for this student
	PASS	Total number of courses passed (aggregate of A, B, C, PASS, EXCELLENT)
	FAIL	Total number of courses failed (aggregate of D, E, F, FAIL, NO GRADE)
	total_resits	Total number of times a course was re-taken during the BSc programme because of a fail
	max_resits_for_course	Maximum number of times that one specific course was re-taken during the BSc programme because of a fail
'BISS data - BSc Graduates - Student version.xlsx'	Graduated	This tells us if the bachelor student managed to graduate. One means yes, zero means no. This is taken directly from the Graduates dataset.
	Program Graduated	This is the programme that a bachelor student graduated from. This is directly taken from the Graduates dataset.
BISS data - Masters - Student version.xlsx + *'BISS data - Masters 2018 - Student version.xlsx'*	Number Applications Master	This tells us how many times a student applied for a master in UM. This was computed by counting how many times a Student Number appeared in the Master Admissions Sheet

	Master Approved	One means that the student's master application was approved. Zero means the student was either rejected or has withdrawn. This doesn't count how many applications a student was approved in. Some students were approved in two applications, but it would only be written as one.
	Years Registered Master	This counts how many years the student has registered as a master student. This is calculated by counting the number of times a student number appeared in the Master Registrations Sheet.
	Starting Year Master	This is a proxy on when the student's first master year in UM was. This is calculated by collecting all the years a student was registered as a master student and getting the earliest year.
	Bachelor and Master Present	This tells us that a student who did their bachelor in UM also did their master in UM. It also counts the bachelor students that applied for a master in UM but either got rejected or withdrew the application. Zero means no, one means yes. This is the target attribute. This was calculated by selecting the student numbers that appeared both in the Bachelor Admissions Sheet and Master Admissions and Registrations Sheet.
	in_master_um_file	Binary outcome variable (1 = continued MSc at Maastricht University, 0 = otherwise)
	Mean_Grade_Master	The average grade during a student's master
	new.target.var	The tertiary target variable taking into account the master performance of the student. (0 = did not continue their master in UM, 1 = continued but with mediocre grades, 2 = continued but with good grades).
BISS data - Ranking - Student version.xlsx (Ranking IB & Ranking EBE)	n_bins	Number of times a student has a bin (possible proxy for number of times a student applicated for a programme)
	max_bin	Best bin a student was placed in (the lower the better)
	min_bin	Worst bin a student was placed in (higher is worse)

BISS data - Leads - Student version.xlsx	Bachelor	Requested a brochure for one of the BSc programmes (1 = yes, 0 = no)
	Master	Requested a brochure for the of the MSc programme (1 = yes, 0 = no)
	BA.ECONOMICS.AND.BUS	Requested a brochure for this specific BSc programmes (1 = yes, 0 = no)
	BA.INTERNATIONAL.BUS	Requested a brochure for this specific BSc programmes (1 = yes, 0 = no)
	MA.Business.and.Economics	Requested a brochure for this specific MSc programmes (1 = yes, 0 = no)
	Bachelor.Open.Day	This tells us if the student attended a Bachelor Open Day. This is generally an uninformative variable and may be unselected.
	Master.Open.Day	This tells us if the student attended a Master Open Day. This is generally an uninformative variable and may be unselected.
BISS data - Pre-education - Student version.xlsx	applied_both_bsc_programs	We have pre-education information for both BSc programmes (applied both programmes) (1 = yes, 0 = no)
	pre_educ_foreign	Pre-education abroad (1 = yes, 0 = no)
	pre_educ_HBO	Did HBO as pre-education (1 = yes, 0 = no)
	pre_educ_VWO	Did VWO as pre-education (1 = yes, 0 = no)
	pre_educ_Uni_Bachelor	Uni Bachelor as pre-education (1 = yes, 0 = no)
	pre_educ_Uni_Master	Uni Master as pre-education (1 = yes, 0 = no)
	pre_educ_Other	Other pre-education (1 = yes, 0 = no)
	pre_educ_Country	Country of pre-education (1 = yes, 0 = no)
Specialisations - Student Version.xlsx	Specialization	The specialization of the bachelor that the student took.

We also constructed a dataset with each student as a row and all possible courses as columns. We constructed a lowest and highest grade for each student and each course in the dataset (NA if the student did not take this specific course). In Figure 22, one can see how this dataset looks for both BSc programmes (programme 2501 and 2504). The rows represent individual students (IDs not shown because of readability), and columns represent courses that have at least 10% non-missing values (at least 10% of the students took this course). We can see that (almost) every student in both the programmes only took the same two courses (ABC1096 and ABC1245).

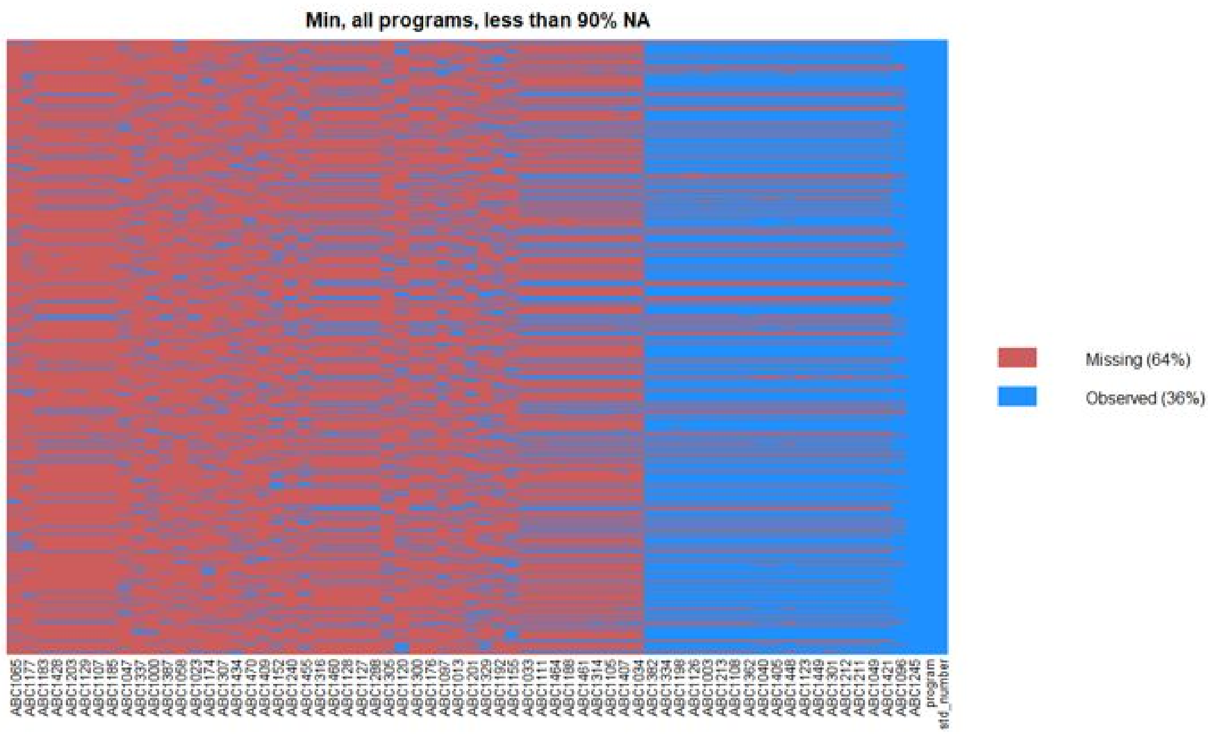


Figure 22. Missing course data of both bachelor programmes

We hypothesized that there would be more courses that every student took if we would split the dataset by programme (one dataset for programme 2501 and one dataset for programme 2504). In Figure 23, we filtered for BSc programme 2504, and we can see that there are 20 courses that (almost) every student took in this programme.

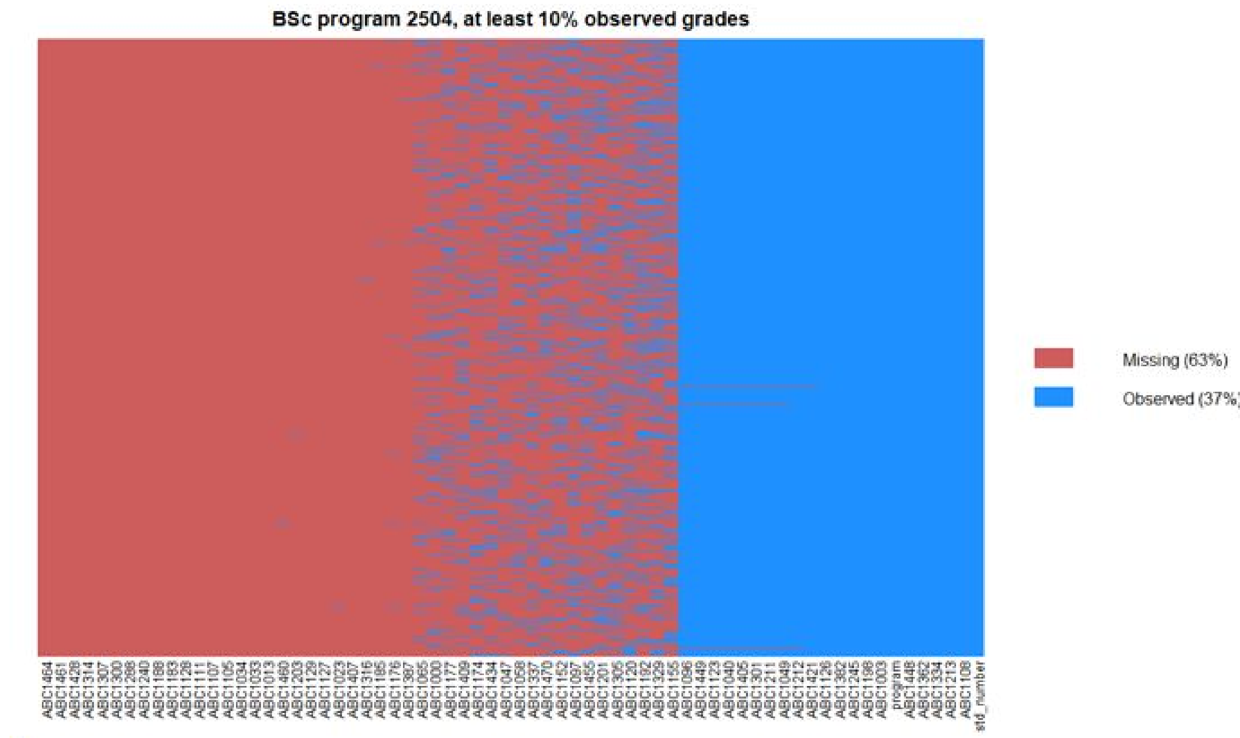


Figure 23. Missing course data of bachelor programme 2504

In Figure 24, we filtered for BSc programme 2501, and we can see that there are fewer courses that every student took in this programme compared to programme 2504. We can see specialization tracks in this BSc programme with similar patterns of courses emerging in the data. Speculation is that Figure 23 represents the International Business track, wherein students take the same courses for two years then specialize in the third (thus the blue and red part in the middle). Figure 24 is speculated to be the Economics and Business Economics track, wherein students take the same course for only the first year and start to specialize as soon as the second year. We can observe that by seeing that the blue part of Figure 24 is smaller than Figure 23. On top of that, there are four discernible columns in the red and blue region (highlighted with a yellow box) that could indicate the four specializations within the track (Economics and Management of Information, Economics, Emerging Markets, and International Business Economics).

In light of this observation, we were prompted to request additional data showing the specializations taken by each student in the university. As we will explain later in the modeling section of this report, requesting this data has greatly improved our model. The specializations, though only available for the 2501 programme, has become the most important predictor of whether a student will continue their master at UM.

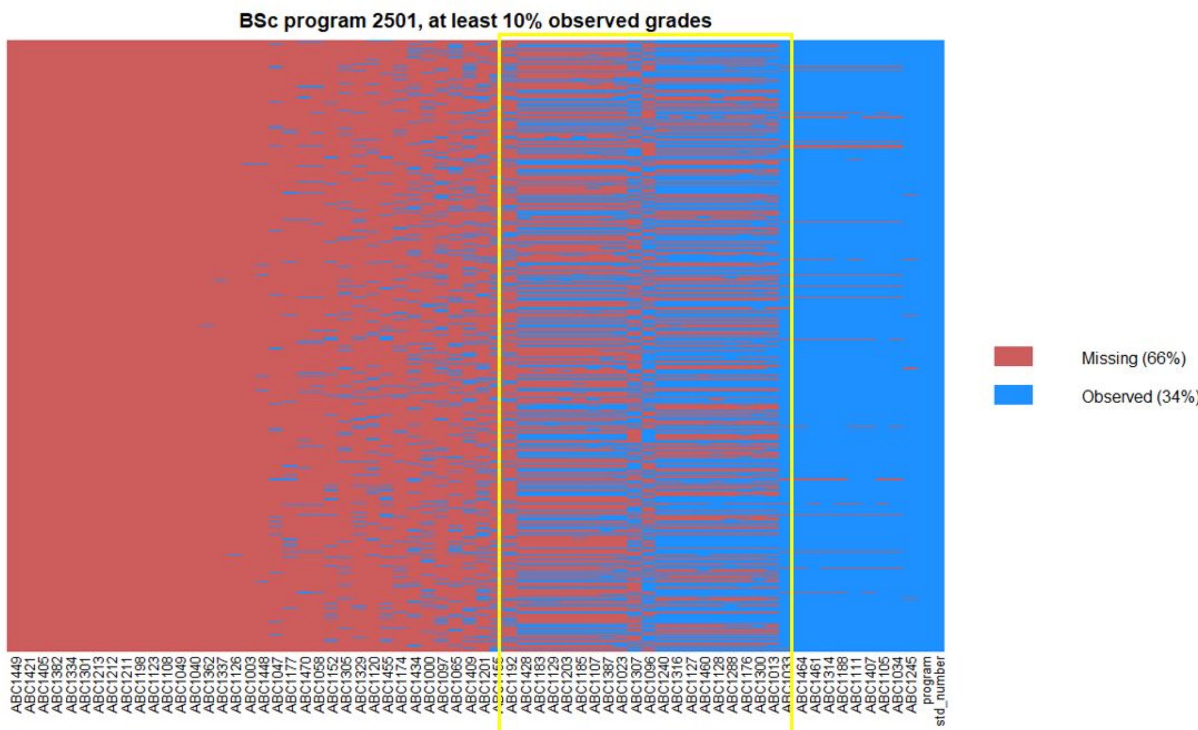


Figure 24. Missing course data of bachelor programme 2501

When observing this, we came to the conclusion that it would be a good idea to create three separate datasets for modeling (and possibly three models also): (1) a dataset with both BSc programmes and without data for specific courses (since there were only two courses that all students in the two programmes took). (2) A dataset for programme 2504 (3) A dataset for programme 2501. The latter two datasets would include variables with grades for specific courses and variables for which specialization (which set of variables) a student took within the BSc programme.

3.1.4. BISS data - BSc Graduates - Student version.xlsx

This dataset was used to have a more reliable way of knowing which bachelor students graduated. The dataset also clearly states the graduate's programme. Table 1 with a purple background further explain the variables constructed using this dataset.

3.1.5. BISS data - Masters - Student version.xlsx + Related Extra Data

This dataset was used to check which BSc-students continued their MSc at Maastricht University. We used the Admissions and Registrations sheets for this purpose. We checked which student number that was in the bachelor file also appeared in the master file to construct our binary outcome variable. We also included general variables regarding how the admissions process of each student went and how many years the master students registered in UM. Table 1 with a green background further explain the variables constructed using this dataset.

3.1.6. BISS data - Ranking - Student version.xlsx

We used this dataset to construct more explanatory variables. We used the IB ranking and EBE ranking information and constructed variables to check whether the bins, in which students were placed, were associated with our outcome variable. Table 1 with a grey background further explain the variables constructed using this dataset.

3.1.7. BISS data - Leads - Student version.xlsx

This dataset contained information on leads (whether a student attended events such as open days or requested brochures). After discussing this dataset with Nico Rasters and Stan van Hoesel, we came to the mutual conclusion that this dataset had severe data quality issues. We nevertheless cleaned this data and tried to construct reliable variables that could be associated with our outcome variable. Table 1 with a yellow background further explain the variables constructed using this dataset.

3.1.8. BISS data - Pre-education - Student version.xlsx

The final dataset we cleaned was the dataset with pre-education information. We only included pre-education data before starting a BSc programme and not before starting an MSc programme, because this information would not be available to us at the time of prediction. Table 1 with a light blue background further explain the variables constructed using this dataset.

3.1.9. Specialization - Student Version.xlsx

This dataset shows the bachelor specializations of the students. Table 1, with a pink background further explain the variable constructed using this dataset.

3.1.10. Total Dataset

Figure 25 shows the missing values in the total dataset. The rows represent individual students (labels not shown), and the columns represent variables. We can see that we have complete information for most of the variables. We can also see the data quality issues in the leads file since the variables we constructed from this dataset have many missing values. Mean grade for year four, five, and six is also missing for most students, and this makes sense since most students graduate within three years. Missing ranking values (n_bins, max_bin, min_bin) are also present for some students. In the modeling section, we will discuss how we will deal with missing values.

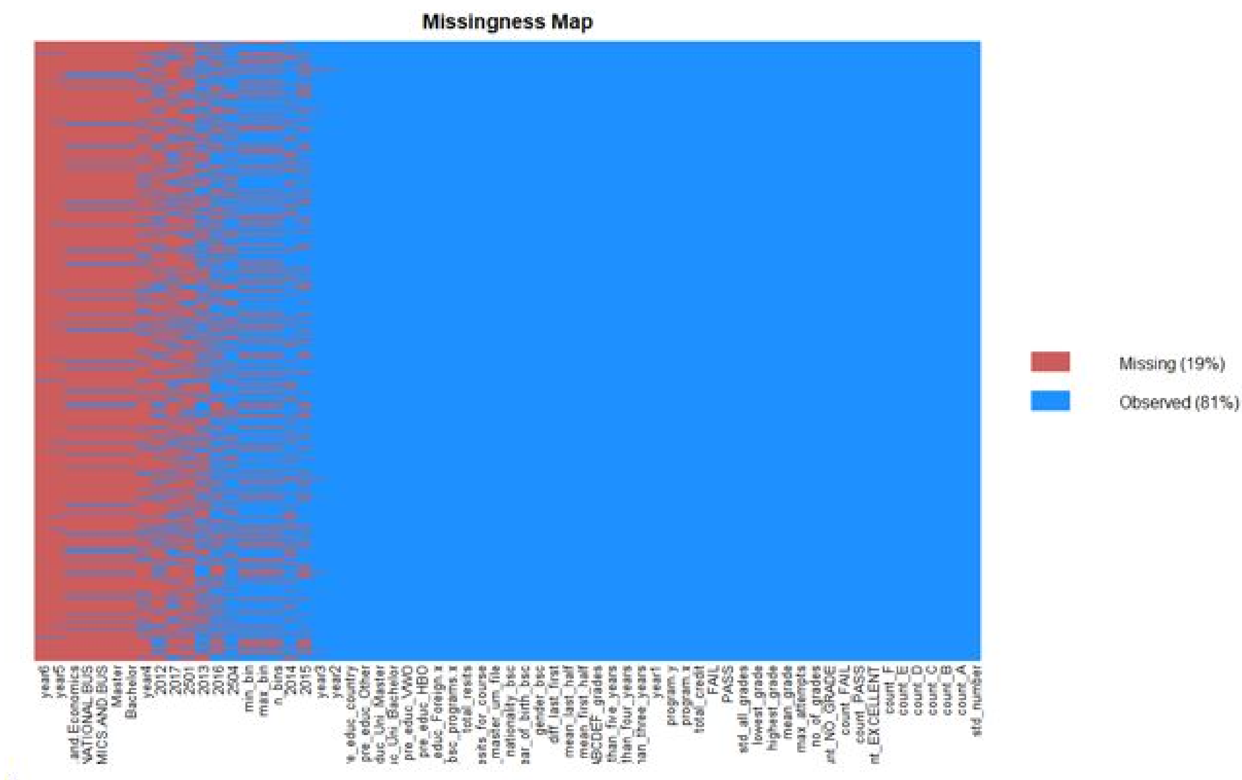


Figure 25. Missingness Map

Subsequently, we need to develop mathematical or statistical models that establish relationships between different constructs and propose algorithms to assess these relationships and obtain insights from the data. The chosen methods and techniques will be at the heart of the smart service.

3.2. Modeling

3.2.1. Method Selection

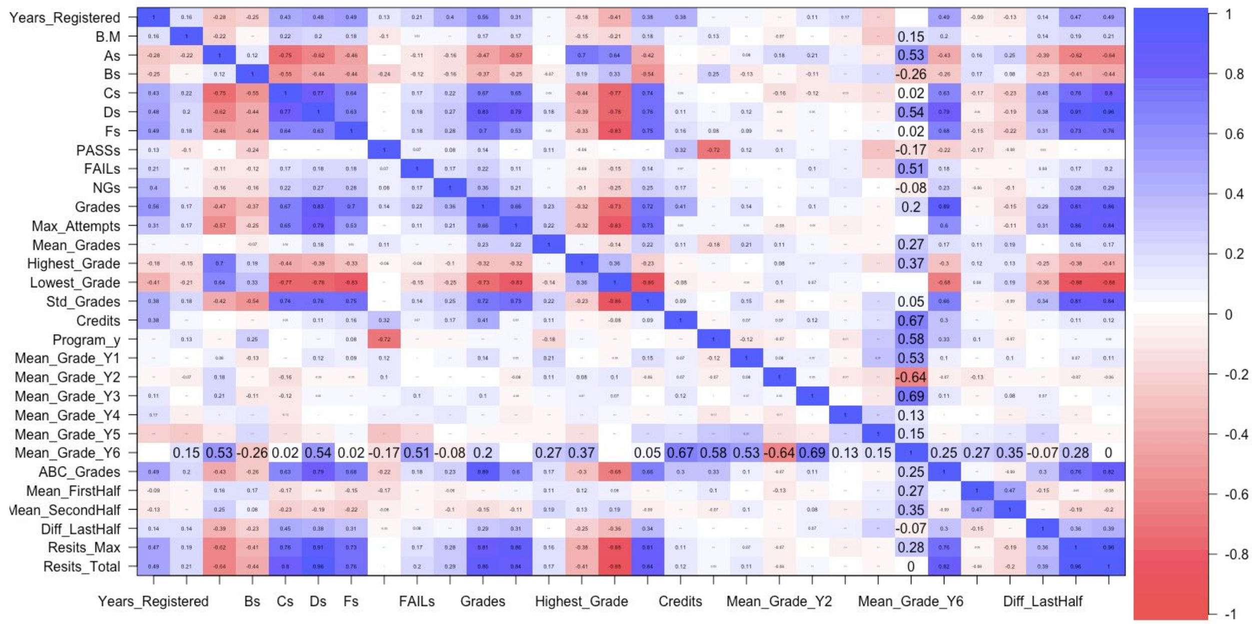
To derive these relationships analytically, we chose to code our analysis within the R Studio programming environment. All variables were coded and scrutinized for collinearity, missing values, and their relevance was reassessed in light of these findings. Once this pre-modeling analysis was completed, and variables were removed, different models were tested and compared. The model presenting the best balance of interpretability (i.e., usability), and accuracy was then selected.

3.2.2. Multicollinearity

Given the relatively large amount of data with the potential to lead to predictive insights on the re-enrollment of internal Bachelor students, data-related issues arose at the onset of the modeling. Firstly, we pondered over the complexity and interpretability of a model with 80 explanatory variables. Especially as our service's most critical aspect is to be its actionability, this aspect needed to be addressed in the first place. Secondly, contingent on the algorithm chosen, such a high variable count could proof technically infeasible, or lead to inaccurate predictions.

To tackle the issues mentioned, the first step taken before modeling was to plot the correlations between all variables, including the predicted variable (Bachelor and Master present). This plot infers 1) if (and which) variables potentially have the most predictive power over the outcome variables, and 2) if (and which) explanatory variables have the most predictive power over each other. Inferring 1) is desirable, while 2) is not, as the latter indicates the presence of potentially superfluous variables. The corresponding correlation matrix and its corresponding R code can be seen in Figure 26.

The matrix excludes three nominal variables (Gender, Nationality, and pre_educ_Country), which were deemed not to propose a high risk of severe collinearity. Leveraging the correlation matrix, some of the original variables were removed, arriving at a more simple, reduced data set. The removed variables related to identifiers, dates, underrepresented categorical variables (e.g., mean grade in a Bachelor's 6th year), superfluous variables, variables from low quality data sources (i.e. Leads data), and variables that were deemed to have no likely predictive power over the outcome variable (e.g. maximum amount of resits per single course).




```r
# Correlation Matrix
cor.ci(Data_Multicollinearity, method = 'spearman')
```

Figure 26. Variable correlation matrix

3.2.3. *Imputation of Missing Values*

A second data related issue, arguably still part of the data cleaning process, was the missing values (NAs), which were more or less frequent depending on the variable's original data source. The average number of missing values was not too problematic due to their information gaps, but presented technical barriers, as some of the modeling algorithms are unable to output a model based on data with NAs. To solve this problem, imputation of the missing values was required, which was done with the missForrest package in R. Before describing the imputation procedure, it is worth to point out why the imputation was done after the reduction of the variable count: Firstly, depending on the imputation procedure chosen, highly multicollinear data with NAs do not produce reliable imputation estimates, resulting in a high normalized root mean squared error (NRMSE, i.e., imputation error) — removing superfluous variables before imputation ameliorates this problem. Secondly, some of the removed variables were removed due to low data quality or irrelevance, meaning that these would have been removed after imputation nonetheless.

The imputation with the missForrest package computes missing variables by comparing data instances with NAs to data with no NAs and imputes values from instances that are most similar to the row with the missing values. Underlying this imputation procedure is the k-nearest neighbors algorithm, and thus, as aforementioned, it is subject to error. Given the relatively low count of NAs within the reduced dataset, this risk was assumed to be immaterial.

3.2.4. Model testing

At this point in the project, we developed a base data frame with 23 potentially significant explanatory variables with no NA values in either the explanatory or outcome variable(s). This data frame served as a perfect setup to produce models, switch variables in and out, and to compare the created models. The constructed models in this project include Logistic Regression, Support Vector Machine (SVM), Artificial Neural Networks (ANN), Random Forest, and Gradient Boosted Regression Model.

The model choices were influenced by studying what other academics have done on problems similar to ours. A study by Aulck et al. (2016) has attempted to predict student dropout using logistic regression, random forest, and k-nearest neighbors, with logistic regression performing the best at 67% accuracy (as you will read later on, our group has achieved a model with better accuracy). Furthermore, the study was able to demonstrate the feasibility of predicting student attrition after just one term and the importance of GPA as a predictor. A separate paper by Reason (2003) studies the different variables that would predict student retention. Some of the variables mentioned to have informative value is gender, ethnicity, socioeconomic status, pre-education grade, and bachelor grade. The current dataset contains most of these variables. Finally, a study by Thammasiri et al. (2014) has also studied student attrition with the focus on data imbalance problems. They employed logistic regression, decision trees, neural networks, and support vector machines. Their method for tackling imbalanced datasets has been of use to us too, as we will explain further in this report.

Before the actual modeling, the data frame was divided into a train set (70%) and test set (30%) to validate our models on unseen instances. By making these distinct data sets fixed, we ensure that all models are not only validated but also compared on the same unseen instances.

The first developed model used Logistic Regression, which computes new data instances' (i.e., students') value for the output variable, namely whether a Bachelor student will continue his studies at UM. The output of this regression is transformed to only include values between 0 (leaves UM) and 1 (stays), which can be interpreted as that student's probability of continuing his studies at UM. Given that logit models are more sensitive to overparameterization, we derived three different subsets of the aforementioned 23-variable data frame and modeled each one separately — the main differences within the subsets related to the choice of variables related to students' academic performance. Once settled on a Logit model with a satisfactory accuracy (67%), we moved on to test with another model type. (Other performance measures than accuracy were evaluated for different models and are shown in Figure 29; for the sake of clarity we only directly mention accuracies of the different models in the text)

Second was the Support Vector Machine model, which, as a machine learning model, is also able to classify new data instances in a linear and non-linear fashion. In this case, we didn't face the same degree of overparameterization risk, and, while including all 23 pre-selected variables did not yield the best balance of an accurate and parsimonious model, upon reducing the dimensionality it outputted good results, with an accuracy of 64%.

Third, in line was the ANN model. For this model, we prioritized avoiding redundant and highly collinear variables and tried three different datasets with varying amounts of variables. As it turns out, the best dataset is the one with the least amount dictated by the significant values highlighted by logistic regression. The reduced dataset is also normalized before training the ANN model. The resulting model outputted an accuracy of 59%, which is the worst accuracy of all the models.

Fourth, the Random Forest model creates many decision trees and concludes the predicted outcome variable by looking at the mode of all trees in the forest. To enhance the model, we looked into its main properties: Ntree and Mtry. Ntree refers to the number of trees the Random Forest creates to achieve the minimum possible amount of out of bag (OOB) error. The model can no longer decrease its error after approximately 350 trees, as displayed in Figure 27. Additionally, the Mtry property of the Random Forest describes the number of variables tried at each split. The plot showed us that the model performed best with two variables tried at each split (as depicted in Figure 27).

To conclude how vital a variable is in the Random Forest, we calculated the mean decrease in Gini, which refers to how important the value is for estimating the outcome variable. Variables with high importance in this model include any variable which consists of a variation of the mean grade (e.g., diff_last_first, std_all_grades or mean_grade) and max_bin. Upon enhancing the model with setting a predefined Ntree and Mtry it was able to reach an accuracy of 68%.

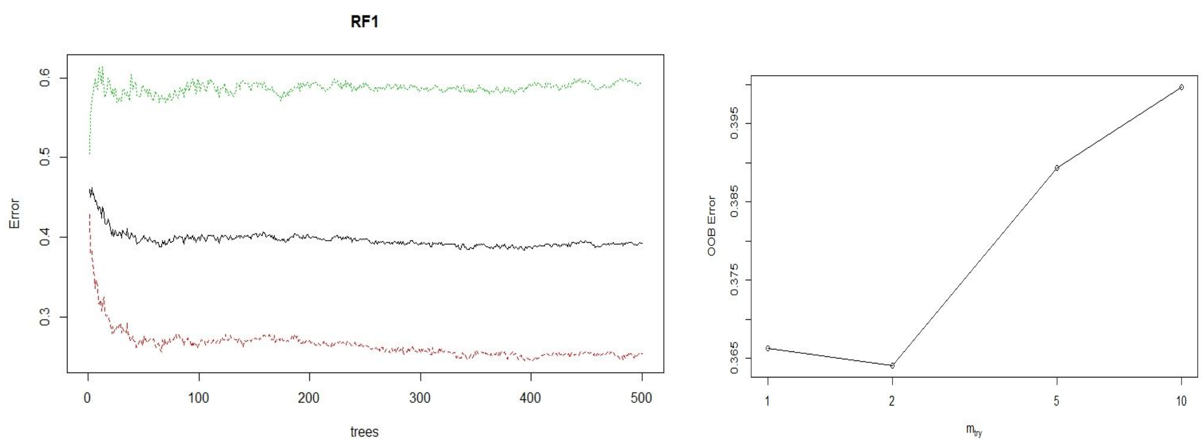


Figure 27. Random Forest complexity

Finally, we created a Gradient Boosted Regression model. Caution was taken to guard against the model's tendency for overfitting by finding the optimal number of iterations or trees for the model to produce. We used out of bag sample and cross-validation methods for this. The left graph on Figure 28 shows the out of bag method, and the right shows the cross-validation method. The blue dotted line indicates the sweet spot for the number of trees. The black line in the graphs refer to the training set while the green line in the cross-validation graph refers to the test set. We found that the optimal number given by the cross-validation method works better with large datasets. With the optimal parameters taken, the resulting Gradient Boosted model outputted an accuracy of 65%.

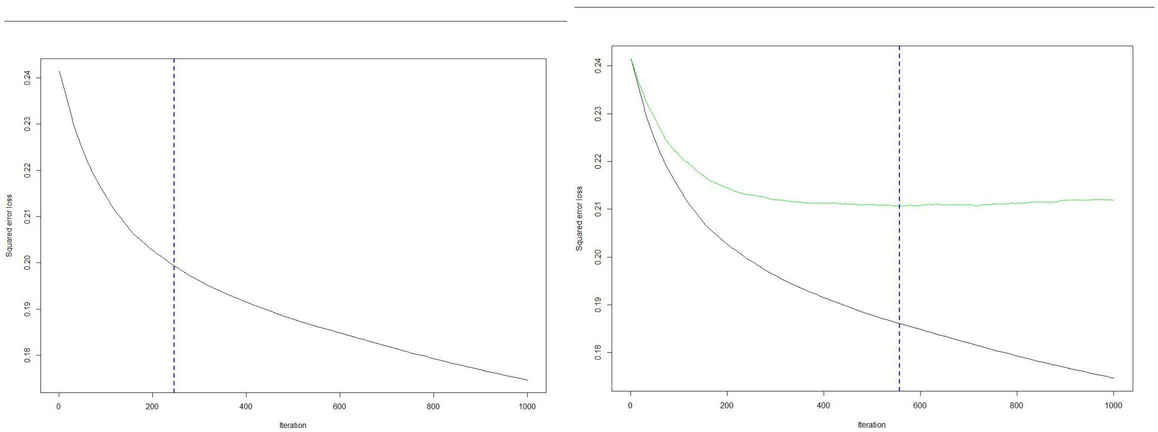


Figure 28. Gradient Boosted Regression model optimal trees

Figure 29 shows a summary of the evaluation metrics of each model. Considering that the Random Forest has the highest accuracy and recall, and a decent specificity and negative predictive value, we chose to pursue this model for additional fine-tuning and analysis.

	model	accuracy	true.positive.rate	true.negative.rate	positive.predictive.value	negative.preictive.value
1	Logistic Regression	0.6655052	0.5780142	0.7500000	0.6906780	0.6479290
2	SVM	0.6443662	0.5958904	0.6611374	0.3782609	0.8254438
3	ANN	0.5862069	0.4646925	0.7555556	0.7259786	0.5031712
4	Random Forest	0.6786942	0.6310680	0.7047872	0.5394191	0.7771261
5	Gradient Boosting	0.6531690	0.4130435	0.8165680	0.6050955	0.6715328

Figure 29. Evaluation Metrics Summary

3.2.5. Random Forest Fine Tuning: Separate Models per Programme

This section is dedicated to optimizing the Random Forest model we have selected from the preliminary analysis of different machine learning models. The direction we immediately decided to take for the Random Forest model is to create two separate models per programme (2501 vs. 2504). We split the dataset between the two bachelor programmes.

The resulting 2501 dataset had 528 instances, and the 2504 dataset had 1325 instances. The 2501 dataset is imbalanced with a 69% vs. 31% proportion of 0 and 1 target attribute. The 0 attribute refers to students that did not continue their master studies in UM, and the 1 attribute relates to students that did continue their master studies in UM. The 2504 dataset is more balanced than the 2501 dataset with a 55% vs. 45% proportion of 0 and 1 target attribute.

To fix 2501 dataset's imbalance problem, we employed oversampling. This method to solve imbalance problems were also used in a study by Thammasiri et al. (2014). We tried two different ways: Randomly Over Sampling Examples (ROSE) and Synthetic Minority Over-Sampling Technique (SMOTE). In the end, we chose to continue with the SMOTE method since ROSE created negative values for many numeric columns, which is unrealistic in real life. The resulting SMOTEd 2501 dataset has a more balanced 51% vs. 49% proportion of 0 and 1 target attribute.

We start by training a Random Forest model for both programmes. The train and test dataset are split as it was earlier (70% training, 30% testing) and will be the same throughout. We ran multiple iterations with different specifications for the random forest model that took a tremendous amount of time to identify the optimal number of trees (250-500 have low OOB-error) and optimal mtry (3), which is the number of randomly selected variables as candidates at each split of the tree. We decided to do this time-intensive task for better assurance of optimal parameters, which is done alongside the automatic checking of the optimal number of trees, as shown in Figure 30.

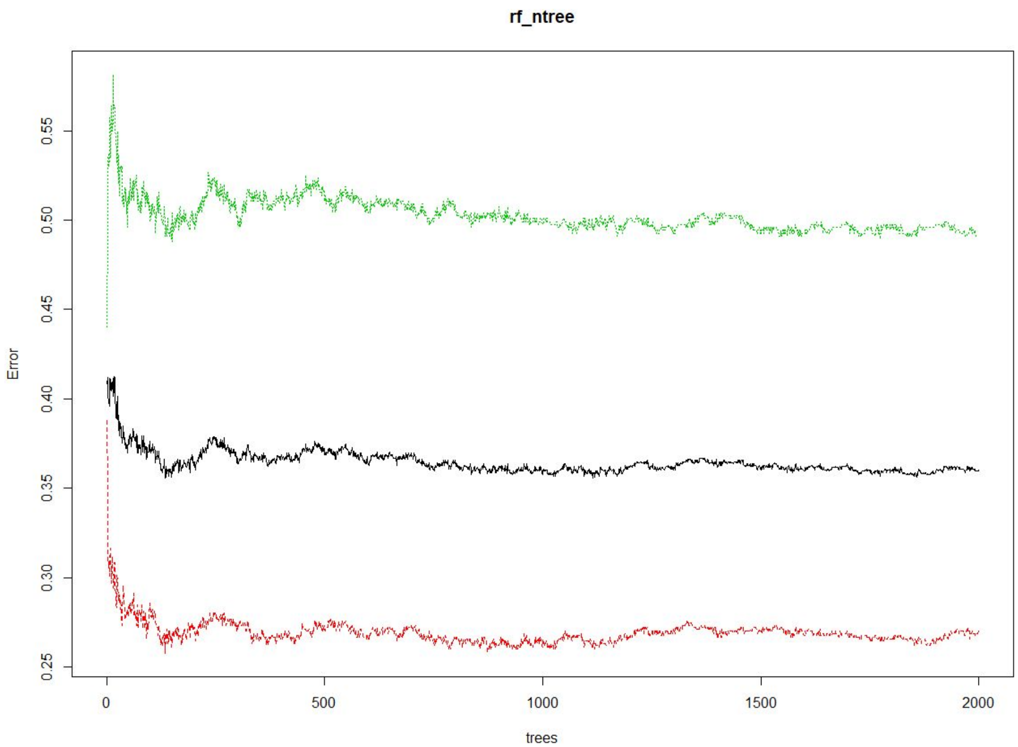


Figure 30. The Optimal number of trees for fine-tuned RF model of both programmes

The manual fine-tuning of the Random Forest model has borne fruit since it has managed to lift the accuracy from 68% to 72%. It is also worth noting in Figure 31 is the variable importance plot from this model. As we can observe, the most crucial variables are all related to the grades of the student during their bachelor years.

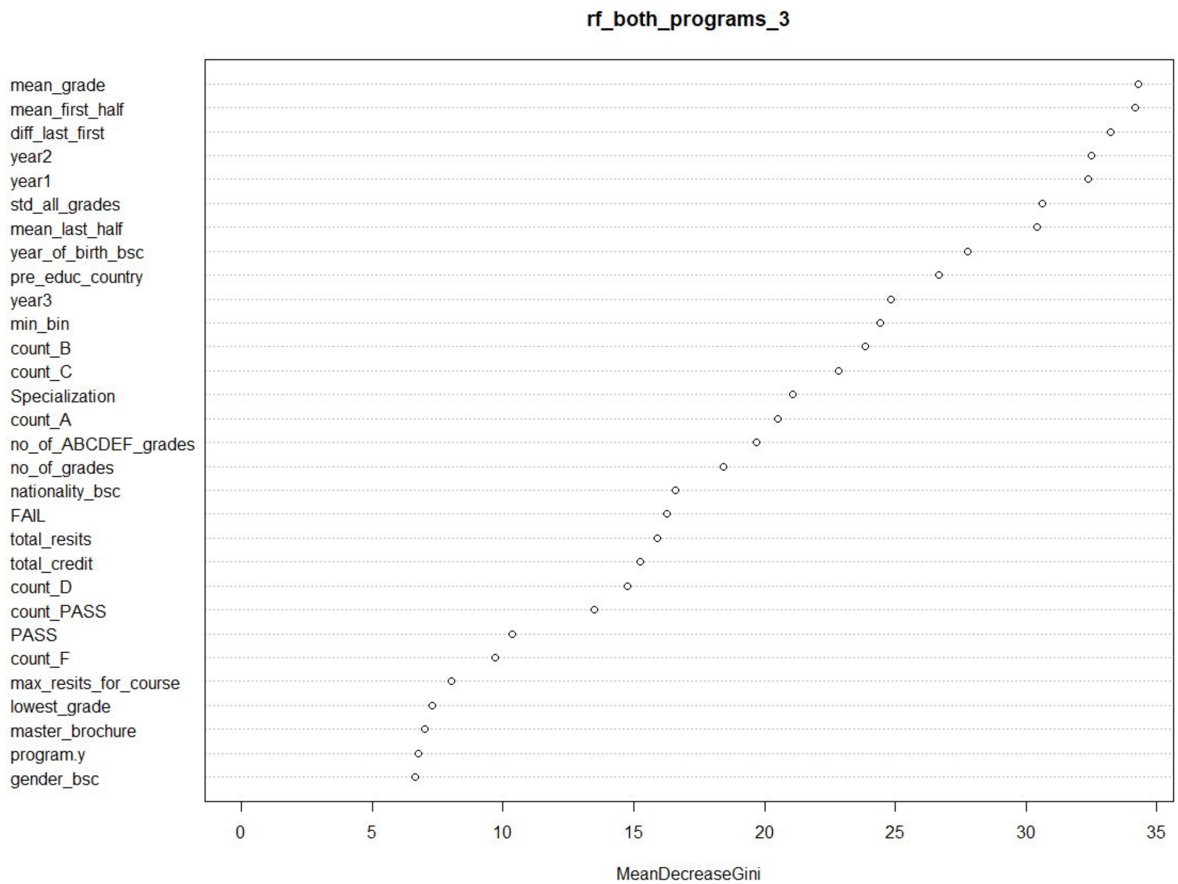


Figure 31. Fine-tuned RF model both programmes Variable Importance Plot

A problem with models like Random Forest is that it tends to be a black box, which means that the model performs well, but it is not intuitive and we have no idea how the model works and what factors affect its output. To demystify this black box, we use the prediction_breakdown() function in R. It utilizes SHAP values, which measures the impact of each variable to the final prediction of the model (Casas, 2019). The output of the prediction_breakdown() function is plottable and is exhibited in Figure 32. To display this plot, we give two examples: one for a student with a high likelihood and one with a low likelihood of staying.

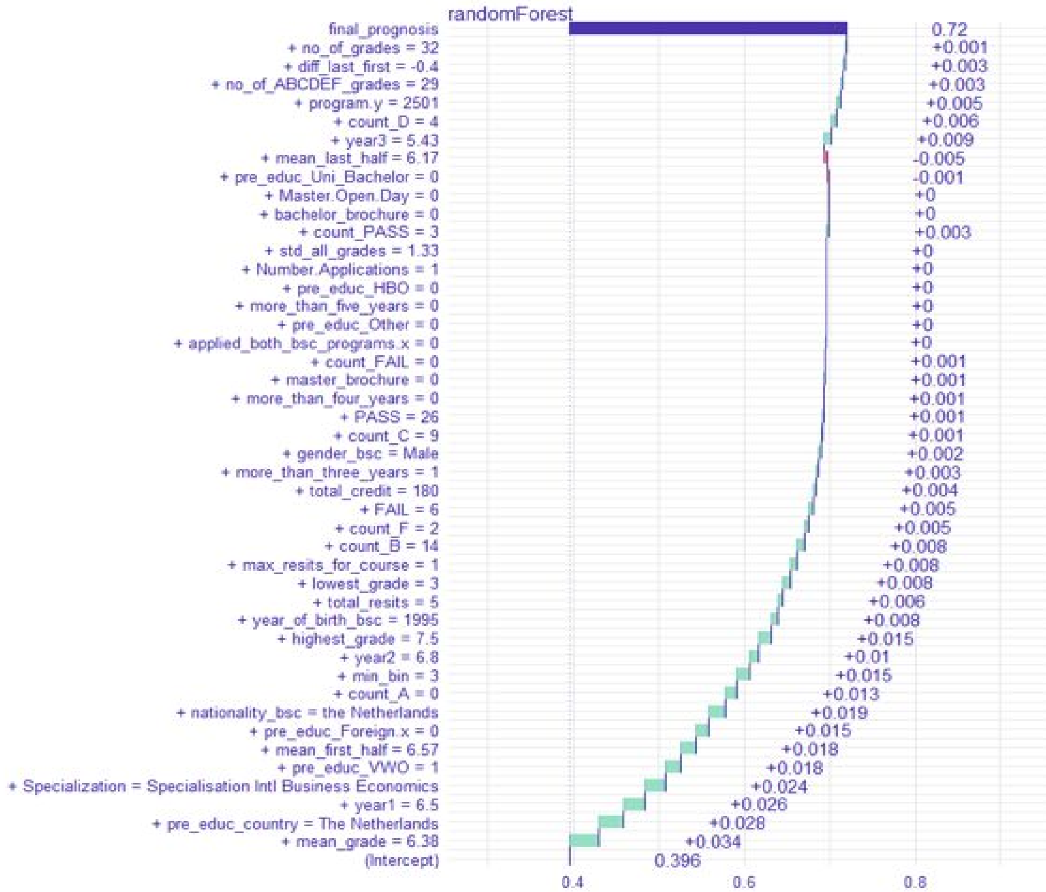


Figure 32a. Prediction Breakdown Plot for a student likely to stay.

Figure 32a shows the breakdown for the prediction of a student that is likely to stay. Figure 32a has opened the black box that was the Random Forest model made from both programmes. Form the plot; we can see how much particular variable affects the prediction and in which direction it does so. The plot starts at zero at the bottom and progresses upwards until it reaches the final prediction on top. From Figure 32a, we can see that the student being Dutch increases the chances of the student from staying. Additionally, we can see that since the student specializes in International Business Economics, it also pushes the prediction more to the right (more likely to stay and do their masters in UM).

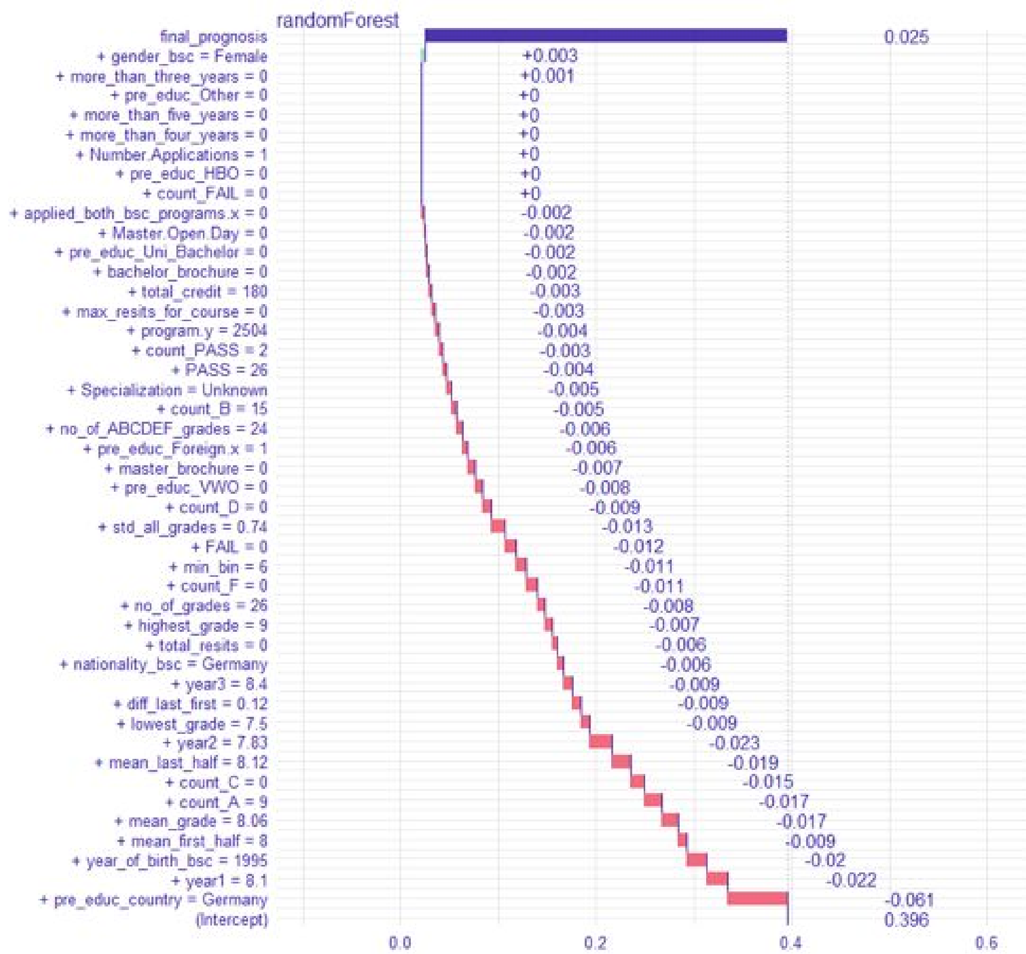


Figure 32b. Prediction Breakdown Plot for a student unlikely to stay.

Figure 32b shows the prediction breakdown plot for a student unlikely to stay in UM. The intuition is similar to Figure 32b. We can observe that mere fact that a student is German pushes the prediction towards the left by quite a lot (makes the student less likely to stay). We can also observe that having an average grade of 8.12 during the last half of their bachelor study makes them less likely to remain too. While the prediction breakdown plot does a great job at demystifying the model, an obvious limitation is that it is time intensive in computation and only for a single particular student. We will set aside this plot for now and continue with fine-tuning the Random Forest model. We will, however, attempt to incorporate these plots in the PowerBI dashboard.

The second Random Forest model we will train will be specifically for the programme 2504. The same time-intensive procedure was used as the model before in finding out the best parameters for the model. The optimal settings were found to be 200 trees and a mtry of 2.

The 2504-specific Random Forest model achieved an accuracy of 68%. The variable importance plot of this model shown in Figure 33 is similar to the one in Figure 31 for both programmes.

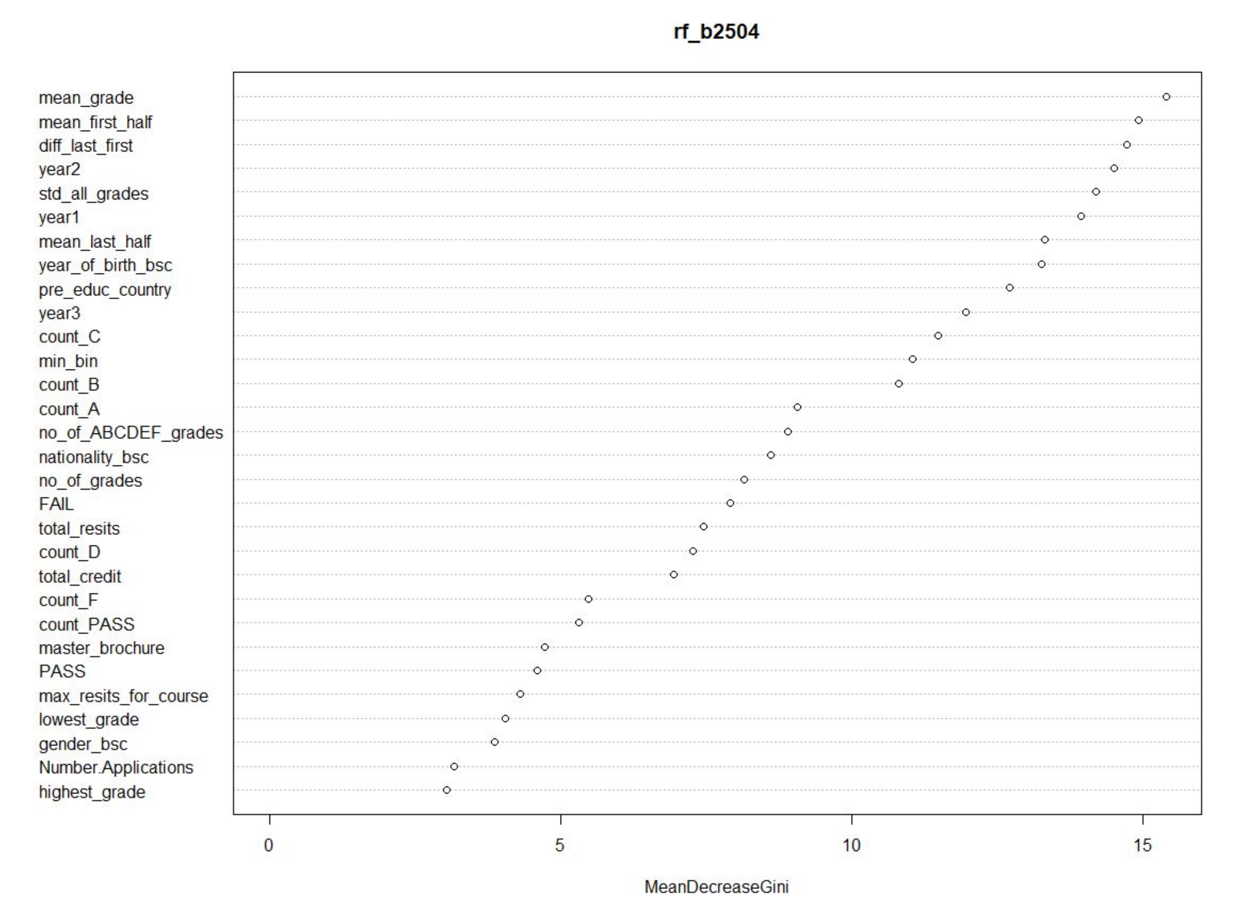


Figure 33. Fine-tuned RF model for 2504 programme Variable Importance Plot

A similar procedure was undertaken for training a model for 2501. We first trained a model under the original unbalanced 2501 dataset. The optimal number of trees is found to be at 400 and the mtry at 9.

This model has achieved a staggering 79% accuracy, which is well above the measure of any other model we have made. Taking a look at the variable importance plot in Figure 34 can explain the cause of this rise in accuracy. The importance of the specialization variable is clearly shown in Figure 34. It seems that this attribute has a tremendous amount of information value in predicting whether the student will continue their master in UM. The 2504 programme did not have such an attribute.

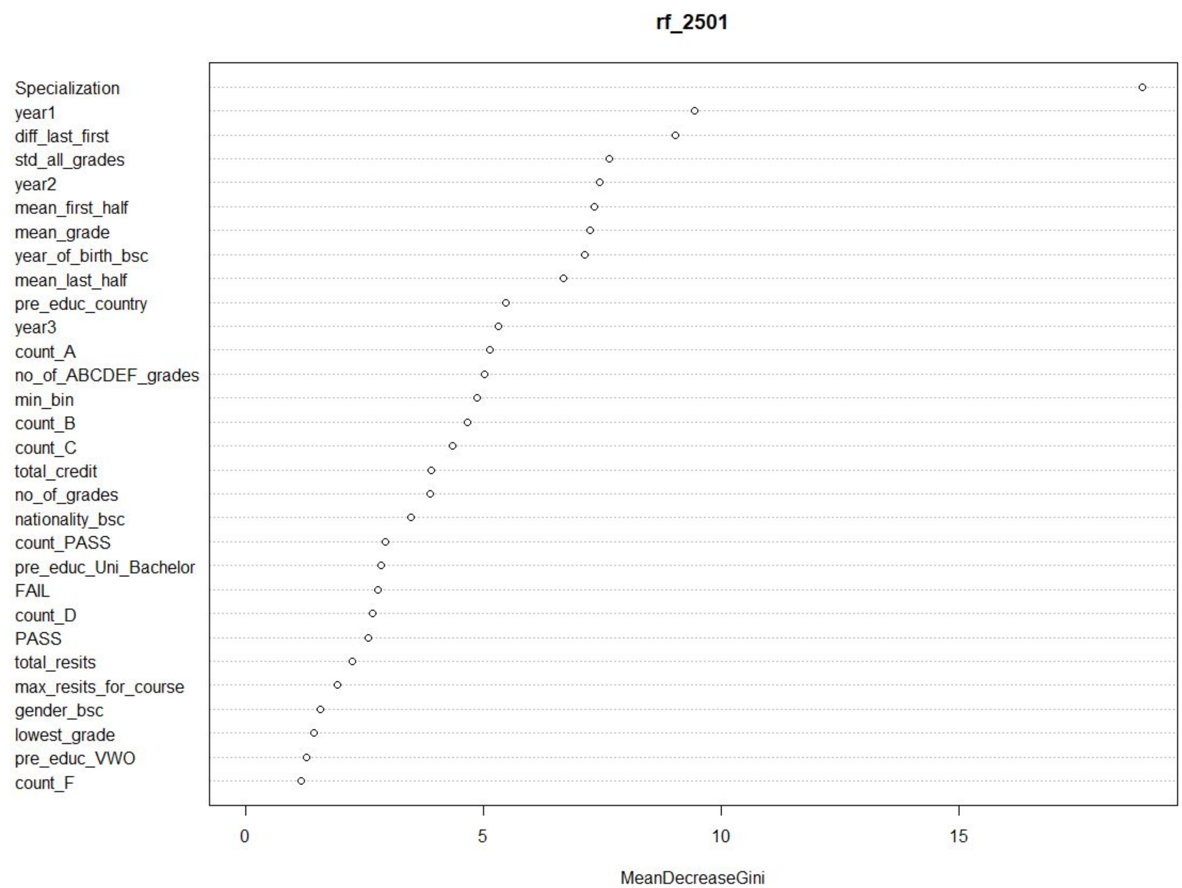


Figure 34. Fine-tuned RF model for 2501 programme Variable Importance Plot (unbalanced)

Investigating further, this made sense since the for the dataset being used, we only analyze the number of students taking the International Business master in UM, and the 2501 programme is the Economics and Business Economics bachelor in UM. One of the specialties under 2501 is an International Business specialty. This is an illuminating finding that suggests that a critical factor for student continuing their masters in UM is that it is related to the study they are doing. Therefore, it is worth pursuing an initiative to collect data that defines the content of their studies.

Moving on, we created another model very similar to the model just discussed, with the critical difference being that we use the oversampled and balanced 2501 dataset. Same procedures were taken and identified the optimal parameters of 100 trees and a mtry of 9.

The resulting model achieved an accuracy of 76%. While it may be lower than the model with the unbalanced dataset, it is still good. More than anything, this may indicate that the previous model is overfitting a bit. The variable importance plot in Figure 35 shows slight differences with Figure 34, but the most informative attribute is still specialization.

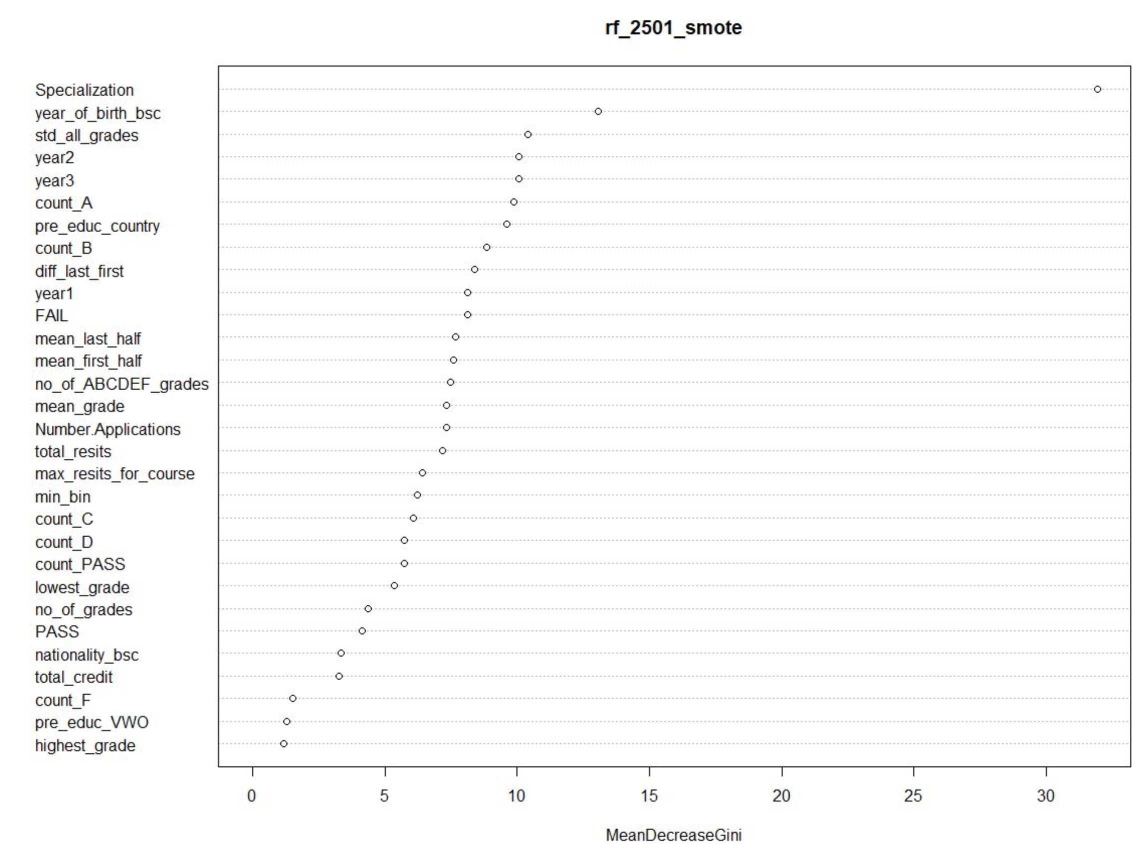


Figure 35. Fine-tuned RF model for 2501 programme Variable Importance Plot (balanced)

3.2.6. Additional Analysis: Considering Master Grades

The mid-semester feedback given to us inspires the final part of the analysis. The feedback suggested including the grades of the master student into the investigation. The highlighted possible need for this direction in the study is the aim to recruit more talented students into the master. While our main stakeholder in the Marketing Department does not highlight such a need, we find that it is still worth considering.

The method used was by transforming the old binary target attribute into a tertiary target attribute. We differentiated between students who continue their master studies into a student with high average master grades (above 7.5) and mediocre (below 7.5). We identified the student who did not continue as 1, those who stayed with a mediocre average master grade as 2, and those who continued with an excellent average master grade as 3.

The reclassification resulted in an unbalanced dataset of 70% vs. 15% vs. 15% between the target attributes 1 vs. 2 vs. 3, respectively. We employed oversampling using SMOTE once again to create a balanced dataset. The more balanced dataset has 36% vs. 27% vs. 36% between the target attributes 1 vs. 2 vs. 3, respectively

We used the same procedures as the previous Random Forest modeling, including the time-intensive sequence to identify the best parameters. We created two models for this last part: one for the unbalanced and one for the balanced. The optimal settings for the unbalanced dataset are 200 trees and mtry of 2. The optimal parameters for the balanced dataset are 400 trees and mtry of 3.

The unbalanced model's accuracy was 71%. While this figure is not too shabby, investigating confusion matrix within the code shows that it is not a good model as it tended to predict almost everything as 1, as seen in Figure 36. This result, however, is expected from an unbalanced dataset. This exercise highlights the importance of using a balanced dataset when modeling and the dangers of over-reliance to high-level evaluation metrics.

	Reference		
Prediction	1	2	3
1	326	60	69
2	8	9	1
3	0	1	0

Figure 36. Confusion Matrix for RF model, tertiary class target, unbalanced

The last model we made, which is the balanced model achieved an accuracy of 69%. While it is lower than the unbalanced dataset, the performance is much better. Figure 37 shows the confusion matrix of this model.

	Reference		
Prediction	1	2	3
1	294	40	49
2	28	27	12
3	12	3	9

Figure 37. Confusion Matrix for RF model, tertiary class target, balanced

While the potential of creating a model that discriminates against future good master students and mediocre master students is present, as of now, the data is still too few to create a functional and reliable model and the need not raised to justify recommending to focus on this future step.

3.3. Critical Reflection

This analytical exercise was all about extracting as much insight as possible with what data was given. A widespread criticism was that there was not enough data provided, and a favorite counter-argument is that part of the exercise is to be able to employ analytical tools to extract insights creatively either way. Our group sees merit to both arguments. However, we do not believe that the quantity of the data is a hindrance at all but the variety. It is worth to note that this master trained us to view data critically and not to rely on pure and hard analytics alone. We were also taught to take a softer approach, which made it imperative to include all specializations within a group. The more we explore the problem using the teachings from service design, the more we realize more possibilities of a more robust model given a different kind of data. While we report these possibilities in the following chapters, we would like to hedge and explicitly say that these are not complaints about the lack of data provided to us (of which we believe the opposite) but avenues for future approaches.

The most promising next step, which became evident after numerous interviews with our primary marketing department contact, Germano Giansante, is the potential of incorporating data detailing the personalities of each student. This idea is further highlighted in the report by Nijhuis (2018) that quality and familiarity with the system are the biggest reasons for staying in UM and that a vast number of students choose not to remain in UM simply because they wish for something different. Figure 38 shows screenshots from the report. The study complements what Germano told us that the counseling department conducted a survey that identified different types of students and their propensity to stay in Maastricht. For example, one of the personalities identified is the "pioneering spirit, " which is the same kind of people who choose to study elsewhere just because they wish for something different.

Another external study that would support this future direction is a study by Schreiner (2009) that shows that student satisfaction is an essential variable for student retention. This study highlights the potential of including softer data points from students like their satisfaction with their bachelor programme. Additionally, a conceptual model made by Girves & Wemmerus (1988) as shown in Figure 39 which shows variables affecting the degree of progress of a student. This framework could be extended to the progression from bachelor to master in the same university. While the dataset provided shows the information about grades the most, it is just one kind of data on the whole dynamic. Figure 38 shows that student characteristics affect degree progress, which is what we propose in this section for the university to attempt to capture data of next. The empirical model could even suggest even more possibilities, like obtaining student financial status.

In-line with this future direction is the congruence of the finding in our analysis and the report. The report mentioned that 89% of students considered the content of the master programme to be very important. This insight is similar to our finding that students studying the International Business specialization are more likely to continue with their master in UM, which in the dataset's case, an International Business master.

Table 4.3 Main reasons for studying at University Maastricht (n = 41)

Reason	Frequency	% most important
Quality	37	77%
Familiar with the system	32	67%
Reputation	23	48%
International aspect	14	33%
1 year program	13	27%
Expected career chances	14	29%

* Multiple answers per student possible.

Table 4.4 Reasons for study elsewhere (n = 37)

Importance for study elsewhere	mean	(s.d.)	1 / 2 NOT IMPORTANT	4/5 IMPORTANT	Main reason, frequency
Content of the study/program	4.44	(.7)	11%	89%	14
Experience something different from Maastricht	4.02	(1.0)	16%	71%	6
Quality of the university	4.10	(1.1)	7%	75%	6
Experience something different from SBE	3.61	(1.4)	2%	61%	5
Ranking of the university	3.63	(1.3)	16%	78%	4
Location	3.68	(1.4)	16%	73%	4
Personal reasons (family friends)	2.64	(1.4)	52%	34%	3
Teaching method	3.27	(1.2)	16%	42%	1

scale (1: not important — 5: important)

Figure 38. Tables 4.3 and 4.4 from the Maastricht University report on Perception and Position of Bachelor Graduates Cohort 2016-2017

That point brings us to the most significant limitation of this project: external validity. This project is a proof-of-concept, at best, and its actual findings are to be treated with a grain of salt. One must always keep in the back of their mind the dataset only includes one master, which is the International Business master, which is a minimal population relative to everyone else taking their masters in UM. For example, two of our members did their bachelor in UM and continued their master in UM too, but in the dataset, we would be marked to be students that did not keep their master in UM. This observation highlights the limits of our findings. We do not identify which students are not likely to take their master in UM in general but only identify which students are more likely to take the International Business master only. This limitation, in itself, does not affect the viability of the project. We prove that it is possible and straightforward tweaking will amend this (by including more masters), it's just that the raw results our model is to be taken with caution.

Another limitation of the model is that, despite its good evaluation metrics, is that the best model has been trained on an unbalanced dataset. In this case, the best solution is indeed to increase the quantity of the data.

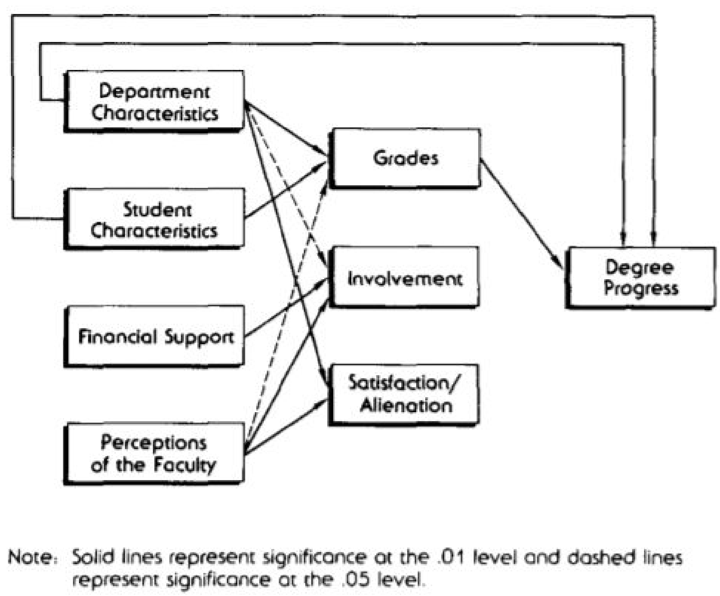


Figure 39. Figure 2 from Girves & Wemmerus (1988)

4. Data Management

Now that our service and models have been elaborated, we move on to data management and flow of the project. In this section, the importance of the Data Warehouse architecture is explained along with the data flow diagram (DFD). They act as the foundation of data storage and data transferring.

4.1. Data Warehouse Architecture

The Data Warehouse (DW) exists at the core of our service since it is responsible for the storage and collection of data. It should be considered that the Power BI dashboard relies on the DW as the source of the displayed information, which stresses the importance of the DW. Therefore, the uncovered insights and decision support realized by using the dashboard are also dependent on the data stored in the DW. Klepmann (2018) finds that the main advantage of having a DW is that users can issue queries without affecting the ongoing Online Transaction Processing (OLTP) operations. Thus, it speeds up time and can serve large scale needs of the end user. In the context of the university, the OLTP is the SAP platform that they to support daily operations, and the DW is the Be Informed Data Warehouse.

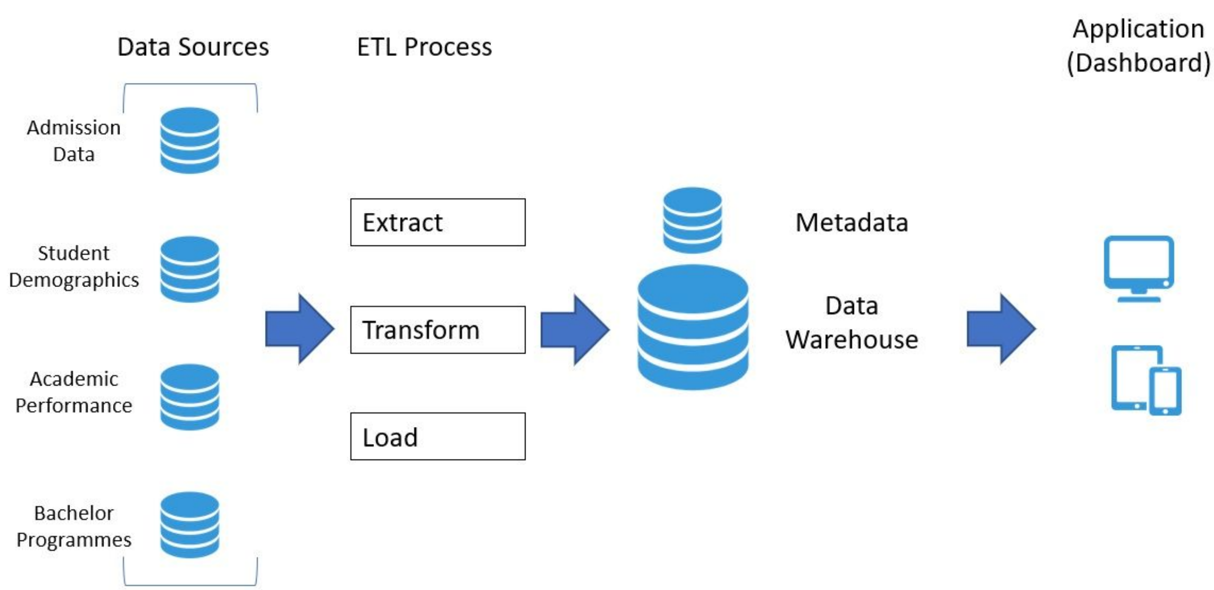


Figure 40. Data Warehouse framework

Figure 40 depicts a structure for filling the DW with the necessary data. The first part contains the collection of data from the various data sources. New data sources can be added anytime as the list displayed in Figure 40 is not exhaustive. For example, external data sources could also be added to the available data sources. Then, the data gets extracted from these data sources and enters a staging area. Before the data gets loaded into the DW, the Extract, Transform, Load (ETL) process cleanses and consistently structures the data. After the ETL process is complete, the DW is established.

Additionally, metadata gets linked to the data for clarification, which is extensively touched upon in chapter 3. The metadata will allow for better understanding and use of the data by the end users. Finally, a user interface will enable access to the DW so that the end users can interact with the data through various devices.

4.2. Data Flow Diagram

Figure 41 shows the data flow diagram of our service and the ecosystem around it. Figure 41 shows the journey of the data, from where it is housed and its mission going into our system and into the final output and the transformations it undergoes. Our service starts in the processes marked in green. The procedure shown in Figure 41 is to be done only once every period. The diagram has been made in cooperation with our contact in the data management team, Nico Rasters, to ensure that it reflects the current workflow in the university and the most practical implementation of our service. Interview notes can be found in Appendix A3. The data flow diagram is constructed following the guidelines detailed in Hoffer & George (2008)'s textbook, chapter 7.

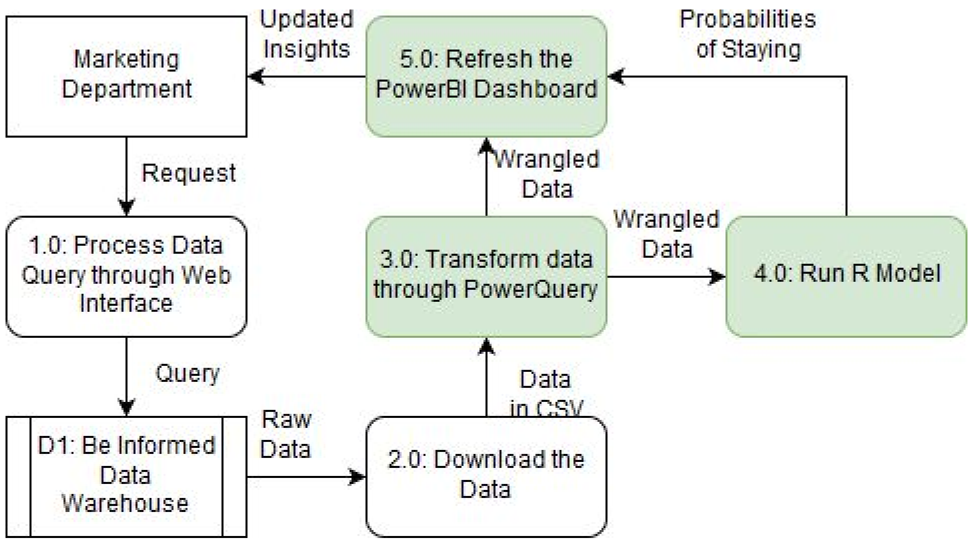


Figure 41. Data Flow Diagram

It starts with the marketing department going to an online interface to interact with the university's DW. This data warehouse is in D1: BeInformed Data Warehouse, which is the brand of the DW. All the data required by the service is housed in the DW. While it is possible to retrieve the data from the SAP platform, which serves as the OLTP of the school, the refresh rate of the DW is daily. The daily refresh rate ensures that the data in the DW is fresh. It would also be better to retrieve the data from the DW directly as the ETL process would ensure that the data is in a consistent format.

Through the online interface, the marketing department can slice and dice the data as required by our service in a user-friendly manner, which we will go further into in the next section. The online interface would translate this query into a language that the DW would understand and retrieve the queried data as a CSV file downloaded into the computer. This whole process involves utilizing the OLAP operations of the DW, which the next section will discuss in more detail. The next step is to wrangle the data, which can be done in two ways: (1) PowerQuery or (2) R code.

For our service to do the necessary analysis, it needs data preprocessing. The steps of which were described in section 3.1 and within part one of our R codes. The R code we made transforms the data in a step by step manner that can be done in PowerQuery. Using PowerQuery's main advantage is that this is a standard tool to the university already, requiring little change and learning. Another advantage is that it only needs to be set up once and it can be executed at each refresh by simply pressing a "run" button.

The second way is to use R directly, which involves incorporating the R code directly into PowerBI. PowerBI allows R script to be a data source. It does this by running the R code within the dashboard and identifying the data frames created by the R code. These data frames are then treated as it would any other datasets, and the necessary visualizations can be applied to it to populate the dashboard in a manner most productive to the end users.

After the initial setup, this entire process would just be downloading a CSV file from the DW and putting it in the correct directory in the computer. Everything else will be refreshed automatically. Since new academic data comes in after each period, the dashboard can be refreshed six times per academic year.

4.3. OLAP Operations

Online Analytical Processing (OLAP) operations are used to retrieve answers to the questions the Marketing and Commutations team currently has. Sharda et al. (2018) claim that OLAP is a method that uses multidimensional analytical queries on the DW to gain answers to managerial questions. Figure 42 displays the data cube used in these OLAP operations, which can be used to access subsets of the data. The user can obtain answers to managerial questions by transforming and navigating through the orientation of this data cube.

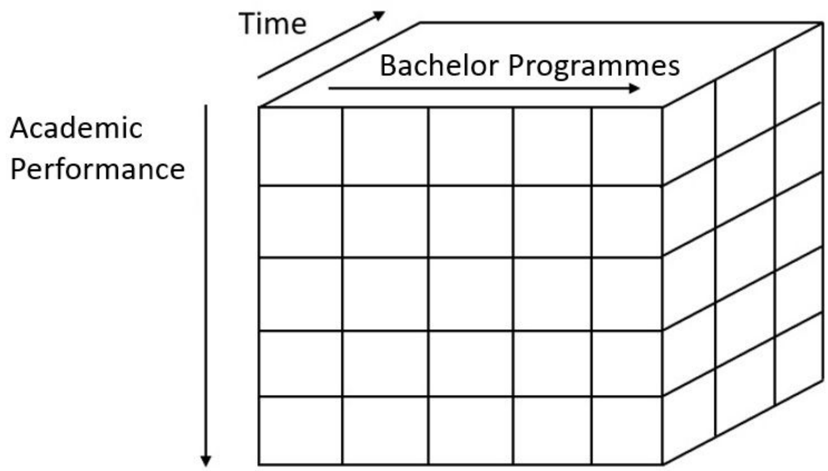


Figure 42. Datacube for OLAP operations

Many different OLAP operations can be used on this data cube that go beyond the scope of this project. However, to give more insight into the data cube, we provide a few examples.

Slicing is one the primary OLAP operations which refer to taking a subset of the multidimensional data set which corresponds to a single value on one of the dimensions. For example, a user could ask for student information for a specific value in Academic Performance on the variables Time and Bachelor Programme. This example would retrieve a two-dimensional array of the variables Time and Bachelor Programme that correspond to a specific value in Academic Performance. Similarly, a user might also ask for a two-dimensional array which corresponds to one Bachelor Programme on the variables Academic Performance and Time. These two tasks are both performed by slicing the data cube. However, they are executed differently.

Dicing is another typical OLAP operation. It’s similar to the slicing procedure, however, it is a three-dimensional subset of the original data cube. Thus, it takes a smaller cube from the original data cube. This operation can be useful when a user is interested in a range of values instead of one specific value like in slicing.

Finally, drilling up and down concludes the last OLAP operation discussed here. The interests of the user in this operation lie in the navigation among the variety of levels in data changes. Thus, a user could first look into Bachelor Programmes and then drill down further to look at results from a certain specialty within the same programme.

All these operations are likely to be required to extract the data from the DW in a manner that was similar to the way that the datasets were presented to us. However, we cannot give a more specific direction in this part since it requires being more intimate with the university's DW.

4.4. Critical Reflection

To conclude this chapter, we once again reflect on our methodology and the contents discussed in it. In this chapter, we took a broad overview by first mentioning the DW architecture and the data flow derived from it. From there, we went on and took a more specific focus to discuss some OLAP operations. During the interviews with Nico Rasters, we found that there already is a functioning workflow and DW in place.

We mainly discuss the new processes and tasks that flow from the implementation of our service and how it interacts with the established workflow. First, a new OLAP process needs to be implemented to extract the data to the current DW. Additionally, we visualized and explained the new data flow that emerged from our service. As a result, we made sure to highlight the parts where our service would interact with the data.

To review the use of this chapter is to have an overview of the new tasks and processes regarding data management. We took the same route as the original workflow to discover where there was a need for knowledge and tools. During this process, we made sure to keep the ultimate goal of making managerial decisions with data as our main driver.

5. Evaluation Proposal

5.1. Service Evaluation

Evaluating a new service is highly essential to assess its feasibility and to improve it continuously. This section is dedicated to providing recommendations for assessing the project when implemented. In general, it would be wise to approach this from the bottom up, which means that, since the marketing strategies rely on the predictions of our model, it would make sense to validate the predictions of the model before evaluating the performance of the recommended marketing strategies.

Testing the model performance and improving our predictions can be seamlessly integrated into the process. Since the primary purpose of our data analytics efforts is to predict students' likelihood to do a Master programme at UM, there are two essential and astonishingly simple feedback mechanisms. First, every year, new students start their Master. These new students can be added to our training set, which, as a consequence, can grow every year and make our predictions more accurate. Second, the registration of new Master students in itself can function as an evaluation of the model, as our predictions can always be compared to the actual new students. For example, if a student had a likelihood of 96% of staying and she starts her Master in UM, this provides validation for our models.

Over time, we can thereby also capture and improve on students that do not do their Master immediately, but who have a delay of e.g., doing an internship or traveling. The longer the service with its analytical component is in place, the better our predictions can get, and we can directly confirm them through comparing our predictions to actual registrations.

To test the previously proposed segmentation, it is recommended to check the students that the model classified as "loyalists" and "mercenaries." Testing students between these two groups is not practical since the students in that region are predicted to be in between the fences whether to stay or not. It is best to test the model on predictions that it is reasonably certain on, like how it is quite sure that loyalists would stay and mercenaries would leave. The prediction must take place in the first period of the school year. The evaluation mechanism would then see how many of the students labeled as loyalists stayed and how much of the mercenaries did not.

A step further would be to reach out to selected students to personally interview them on their choices. Their answers could be compared with the prediction breakdown plot for that particular student that our model could produce.

Once satisfied with the results, the marketing department can now implement our recommended marketing strategies and try out different ones over time. What is most important here is that the results of the marketing efforts can be evaluated by the standard and basic statistical analysis that takes into account the conversion rate of internal Bachelor to Master students. If the conversion rate increases in a statistically significant way after implementing our service, this gives a good indication of its effectiveness.

5.2. Critical Reflection

Although it is useful and essential to evaluate our service by comparing model predictions to actual registrations and by assessing the conversion rate over time, there are a few critical considerations to make during evaluation.

Despite that fact that the evaluation of our service can be automated to a large extent, it will be more beneficial when the marketing team invests time into its assessment. In practice, this could mean, for example, personally following up on a student that had an extremely high likelihood of staying at UM, but did not, to fully understand students' motivations. The more time and resources spent on actually understanding students, the better the outcomes and personalized marketing strategies can be.

Furthermore, it is vital that our service also depends on effective personalized marketing strategies. Although we already suggested several for each segment, it is essential that new ones are developed continuously. From an evaluation perspective, this also means that one has to be cautious while evaluating the service. If, for example, the conversion rate did not increase, it might be that the marketing strategies were not executed well, although the predictions of the service were valid.

All in all, we think that our service offers much room for continuous evaluation, which can be used to fine-tune marketing mechanisms at SBE.

6. Disruptive Business Model

This section aims to outline how our service can disrupt the way marketing for internal students and student retention is handled at SBE.

6.1. Opportunities for Disruption

As mentioned before, from working together with Germano Giansante, it became clear that there is much room for improving the way internal Bachelor students are treated at Maastricht University. To change the current direction of marketing, to recruit, and to retain students for the better, our previously presented Student Portal extension comes into play. This part of the service offers room for two kinds of fundamental changes. First, from a communications perspective, it is an incredibly efficient way to communicate with students because they are using the platform daily. Second, from a data collection perspective, this closer communication can not only be used to guide the students throughout their studies and into their Master at UM but also to collect valuable and difficult-to-obtain data playfully. This soft data is essential since Germano pointed out that many reasons for doing a Master programme at a specific university are personal. This insight was confirmed by the Perception and Position of Bachelor Graduates Cohort 2016-2017 report. This personal data has the potential to improve our predictive models significantly.

To tackle this problem, we, in cooperation with the marketing team, want to offer “mini-games” like “Which Master programme fits your personality” or “Which city is the best for your career” on our Student Portal extension. This could look as follows.

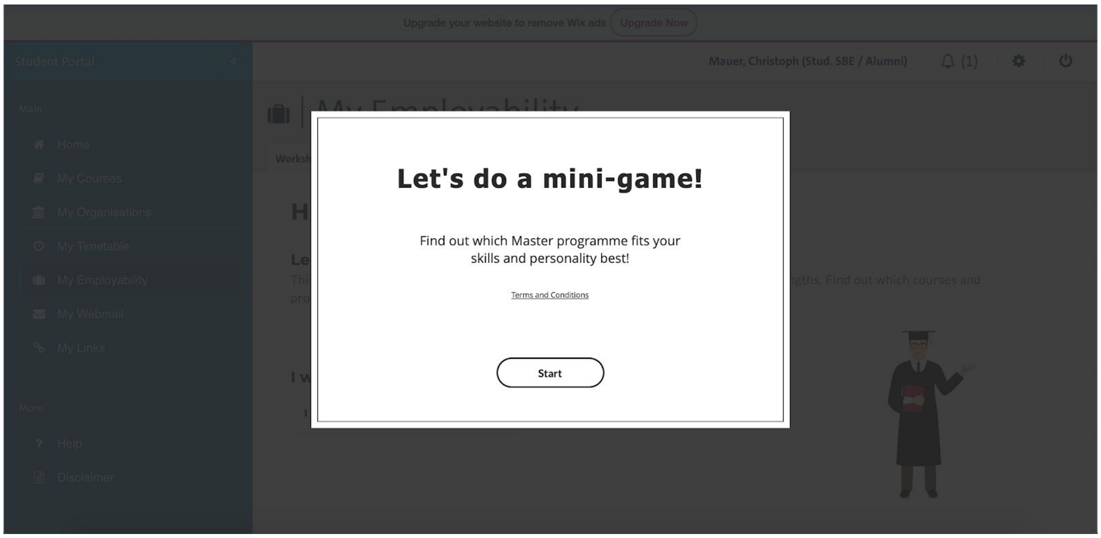


Figure 43. Mini-Game

Potentially difficult to imagine, but these types of games, called “Quizzes and Trivia” are highly popular for millennials and Generation Z. There are even whole businesses built around these data-collection gold mines, like *playbuzz* is shown below.

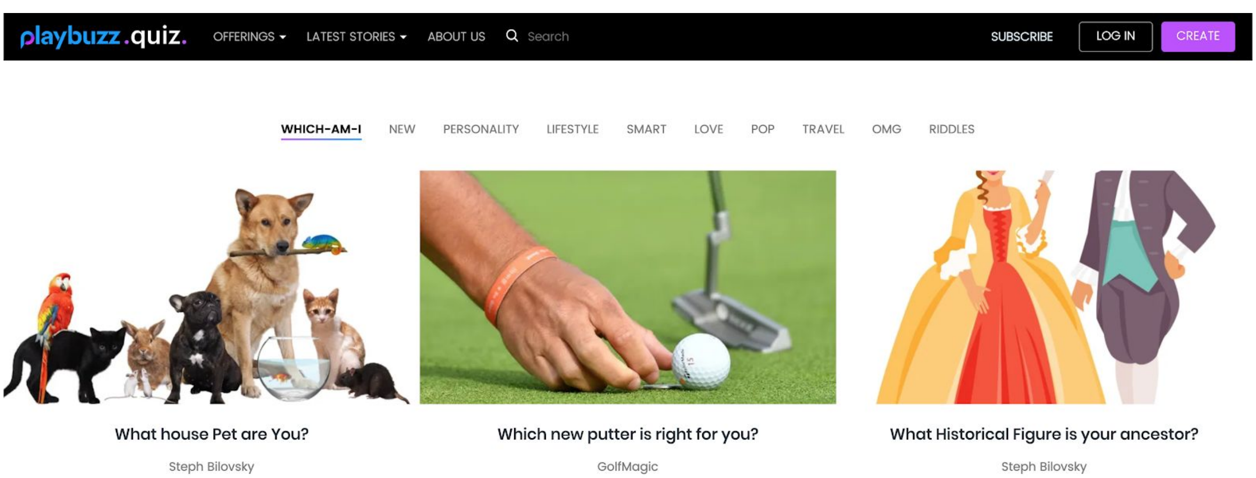


Figure 44. playbuzz.com

From a student perspective, this Student Portal add-on is both an exciting and fun experience that even has some value for your career. Students get to know themselves better and get the feeling that they are guided throughout their studies. We would personally all use these types of games if offered by the university. Moreover, from a marketing perspective, this can indeed be a data gold mine to learn more about students and their reasons for making decisions.

When thinking about the personal reasons for students (not) studying a Master programme at Maastricht University (see Figure 38), many of them could be captured by simple questions. For example, “Experiencing something different than Maastricht” was mentioned as the second most frequent reason for choosing a different university. In a quiz like “Which city fits my career interests best?”, one could easily find out about students’ preferences, while giving them a fun and helpful experience. Moreover, believe it or not, millennials love these types of games.

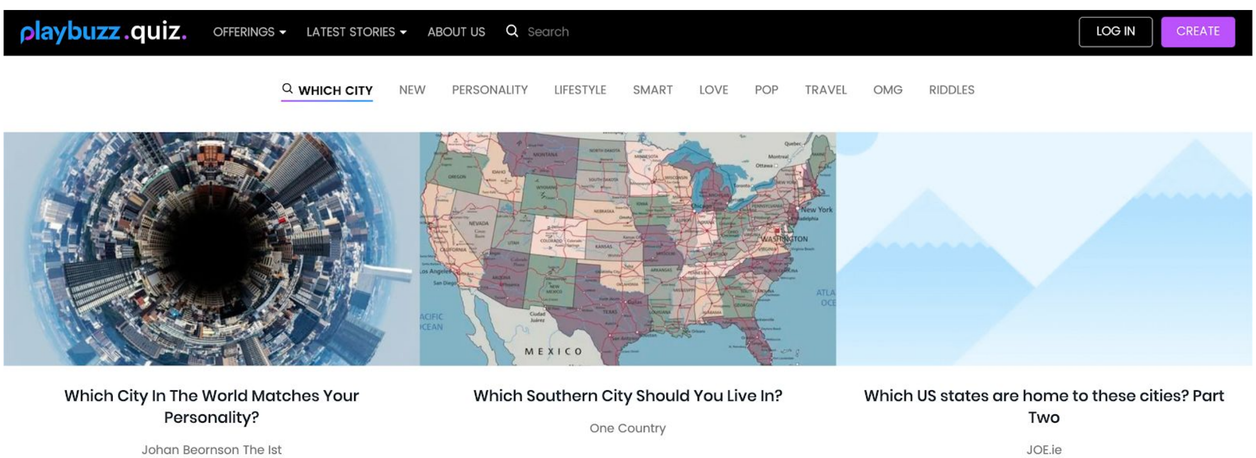


Figure 45. playbuzz.com City Quizzes

To make this disruptive idea more tangible, we created an example quiz that could help the marketing team to find out more about students’ likelihood to stay in the city of Maastricht, thereby covering a highly important reason from the *Perception and Position of Bachelor Graduates Cohort 2016-2017* report. The example quiz is shown below.

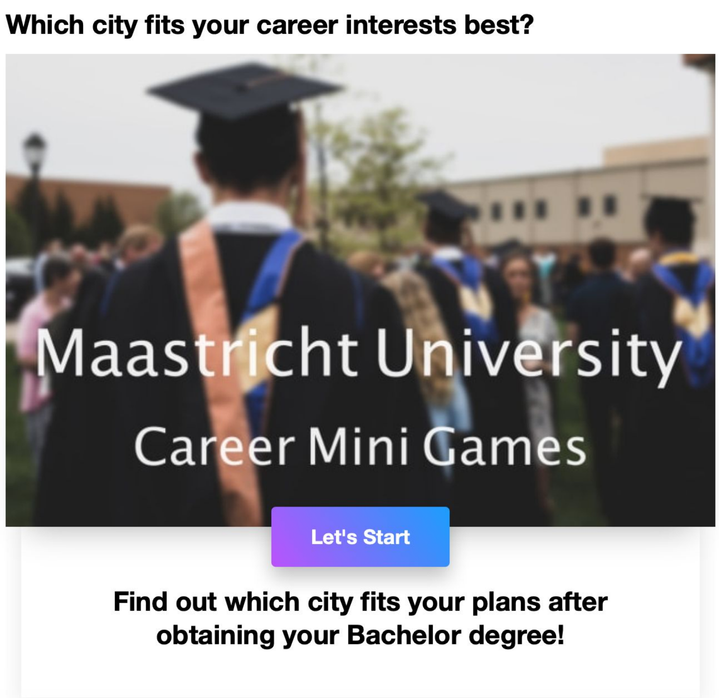


Figure 46. Example Quiz

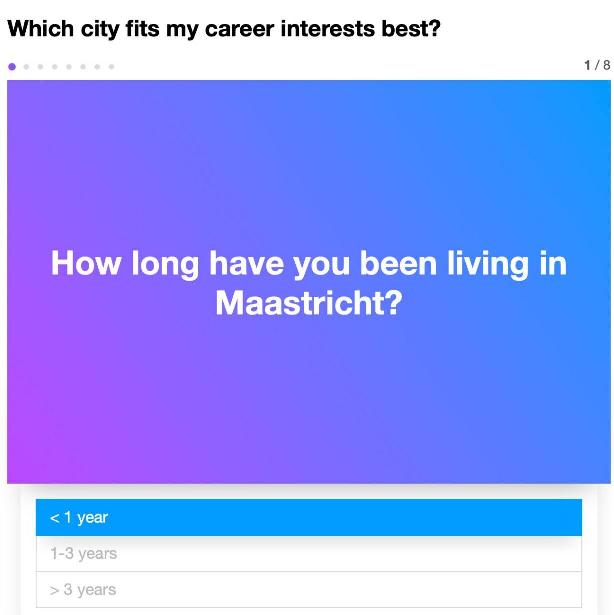


Figure 47. Example Question

But not only the marketing team can benefit from this, but also other staff members are interested in learning more about students' personality. For example, study advisors like Joey Mak and his team are very much interested in learning as much as possible for students to help them more efficiently and sustainably.

6.2. Suggested Execution

We suggest implementing this mini-game add-on after a sufficient number of students has adopted the Student Portal extension. Over time, quizzes could be added regularly. Beneficial for the adoption would be a two-way communication, meaning that the students can give feedback on the usefulness and entertainment level of these mini-games. If these games are well-perceived, one can offer the same quizzes every year to the new students. Necessary is the cooperation of marketing and data management and analytics. To be useful for retrieving information about students, it is essential that the results can be codified, i.e., used for modeling. Rating students could easily do this on several attributes, like "openness," "ambition," etc. For each of these attributes, students could receive a score based on their input to the mini-games which are then used to improve the models.

All in all, we are confident that this is an excellent way of collecting and using previously missing data about students' personality to improve student satisfaction, retention, and conversion to Master programmes.

6.3. Critical Reflection

Although the implementation of this disruption seems to be straightforward and beneficial for the marketing team and study advisors, this add-on needs to be carefully managed. If it is not communicated well how this service also benefits students, they might get a negative perception of the data collection process. Furthermore, one needs to be careful to find the right balance between entertainment and usefulness. In a university setting, these mini-games need to have the appropriate seriousness and do not focus too much on the entertainment factor.

Nevertheless, this disruption offers substantial opportunities to learn more about students' decision-making in a seamlessly integrated way that can support not only our service but also marketing and communications at SBE as a whole.

7. Conclusion

This project aimed to develop a smart service that assists the marketing and communications team in improving the conversion of internal Bachelor students to Master students. Since the University already possesses a vast amount of data on students, this service is uniquely suited to make use of predictive data analytics. Currently, there are no marketing efforts that aim at converting internal students and no forms of personalized marketing.

After doing extensive desk research, multiple interviews with stakeholders, and focus group interviews with students, we developed a smart service consisting of two layers. First, we developed an actionable PowerBI dashboard based on data analytics that can be used by the marketing team to create personalized marketing strategies for different types of students, based on their likelihood of doing a Master at UM and other characteristics. Through this tailored marketing, the SBE staff can efficiently and cost-effectively market to different segments. This can dramatically improve the conversion of internal students. Second, we proposed a Student Portal extension that gives the marketing team the unique opportunity to directly communicate with internal students, while also being able to improve leads data drastically.

Not only did we develop the service concept and the necessary analytics, but we also established an efficient strategy to implement the service that does not require any additional IT infrastructure. Next, we gave recommendations on how to evaluate the service through feedback loops. Also, we presented opportunities to disrupt further how marketing is done at SBE by collecting data about students' personalities, which play a huge role in their decision-making process.

We see this project as a proof-of-concept study that would be easily extendable to other BSc and MSc programs. Our results are already very promising on a subset of the SBE students. We are confident that by using the data from all programs and by improving the quality of the data even further (in ways as suggested by us), our concept will prove its value even more. We are looking forward to the implementation of this service, as we believe that both the marketing department and SBE students will benefit from this. After working closely with our primary stakeholders and IT experts from Maastricht University, we are confident that this service will improve the conversion of internal Bachelor students to Master students and disrupt how marketing is done at the School of Business and Economics.

References

Armstrong, G., & Kotler, P. (2015). *Marketing : An introduction* (Twelfth edition, Global ed.). Boston, Massachusetts: Pearson Education.

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. *arXiv preprint arXiv:1606.06364*.

Beckett, P. (2017). Gdpr compliance: Your tech department's next big opportunity. *Computer Fraud & Security*, 2017(5), 9-13.

Berry, M., & Linoff, G. (2004). *Data mining techniques : For marketing, sales, and customer relationship management* (2nd ed.). Indianapolis, Ind.: Wiley Pub.

Bleier, A., & Eisenbeiss, M. (2015). The importance of trust for personalized online advertising. *Journal of Retailing*, 91(3), 390-409.

Casas, P. (2019, March 18). *A gentle introduction to SHAP values in R.* Retrieved from https://www.r-bloggers.com/a-gentle-introduction-to-shap-values-in-r/

Girves, J. E., & Wemmerus, V. (1988). Developing models of graduate student degree progress. *The Journal of Higher Education*, *59*(2), 163-189.

Goddard, M. (2017). The eu general data protection regulation (gdpr): European regulation that has a global impact. *International Journal of Market Research*, 59(6), 703-705.

Hahn, C., Johnson, M., Herrmann, A., & Huber, F., (2002). Capturing customer heterogeneity using a finite mixture pls approach. Schmalenbach Business Review, 54(3), 243-269.

Hanover Research. (2014). Strategies for Improving Student Retention. *Academy Administration Practice*.

Hoffer, J. A., George, J. F., & Valacich, J. S. (2008). *Modern systems analysis and design*. Upper Saddle River, N.J: Pearson Prentice Hall.

Kim, J., Lee, B., Shaw, M., Chang, H., & Nelson, M. (2001). Application of decision-tree induction techniques to personalized advertisements on internet storefronts. *International Journal of Electronic Commerce*, 5(3), 45-62.

Kleppmann, M. (2018). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems.* Sebastopol, CA: OReilly Media.

Lejeune, M. (2001). Measuring the impact of data mining on churn management. *Internet Research: Electronic Networking Applications and Policy*, 11(5), 375-387.

Nijhuis, J. (2018). Perception And Position Of Bachelor Graduates Cohort 2016-2017 (Rep.). Maastricht: Maastricht University.

Provost, F., & Fawcett, T. (2013). Data science for business (1st ed.). Sebastopol, Calif.: O'Reilly.

Reason, R. D. (2003). Student variables that predict retention: Recent research and new developments. *Naspa Journal*, *40*(4), 172-191.

Schreiner, L. A. (2009). Linking student satisfaction and retention. *Noel-Levitz, Coralville, IA*.

Sharda, R., Delen, D., & Turban, E. (2018). *Business intelligence, analytics, and data science: A managerial perspective.* Boston: Pearson Education.

Tankard, C. (2016). What the gdpr means for businesses. *Network Security*, 2016(6), 5-8.

Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, *41*(2), 321-330.

Tran, T. (2017). Personalized ads on facebook: An effective marketing tool for online marketers. *Journal of Retailing and Consumer Services*, 39, 230-242.

Appendix

Appendix A1: Interview Notes With Germano Giansante 26-2-2019

Objectives

- Nurturing students during their bachelor
- Have an impressive master portfolio

Main Challenges

- The message they send is a doublethink: EXPLORE THE WORLD but stay in UM please.
- Conflicting strategies from central marketing department and faculty specific marketing team.
- Striking the balance between upselling the masters to bachelor students but not to oversell. In other words, it is a sales problem but sales solutions/mindset are not applicable.
- How to improve the recruitment funnel? Why did the student convert to the next stage or why did they not?

Some Figures

- ~50% of master students took their bachelor in SBE
- ~30% of SBE bachelors continues their master in UM

Stakeholders

- Alumni
- Students
- Academic staff
- Employers
- Marketing Department - the operational center
  - Faculty specific
  - Country teams

Who is internal?

- Current bachelor students (including other faculties)
- Current master students (including other faculties)
- Exchange students

What are they doing now?

- Channels active in
  - Facebook
  - Instagram
  - Fairs
- Internal communication (analysis performed AD HOC)
  - Email newsletters (known to have poor results)
    - Contact person - Michel Saive (m.saive@maastrichtuniversity.nl) - collects KPIs of email newsletters

    - Newspapers
- Content creation for channels
- Study advisors

Ideas from the Interview

- Interview each other - draft the group's journey from Bachelors to Masters
- Addition of a feature in the student portal that gauges student happiness\

Additional Notes

What is the idea of this project?

- Service to solve the challenges. What challenges do we want to work on? We are treated as consultants and may choose on what challenges we want to work.
- Internal masters do not have a lot of tools available to them.

Comments on first idea:

How are you liking your Bachelor so far? Happiness could relate instantly towards studying here for a master or not. Data would get more valuable over time. The department should be responsible for creating interesting content.

Appendix A2: Interview Notes Germano Giansante 29-4-2019

Main Findings

- Clueless on how to use the insights gathered from data. We need to include marketing strategies that leverage data (maybe look on how other companies do it.)
  - His hunch, borrow from political strategy: target those who are almost sure to vote for a candidate. (almost sure to take their masters in SBE.)
  - Maybe target the undecided?
  - Target those who have the "pioneer spirit," which are the students who wants change.
  - Tailored strategies per "student persona."
- Current strategy
  - Expanding product offerings (more masters)
  - Cultivate positive bachelor experience.
- Smarter students is not a priority for Germano.
- Providing information to students must be improved to make it easier to find.
- The inclusion of the counselor, Joey Mak (j.mak@maastrichtuniversity.nl), to get to know the depth of student reasons for staying or leaving like:
  - Community
  - City
  - Close interaction
  - International environment
  - Doesn't like the corporate life
- To keep in mind for implementation:
  - Cost (time & money)
  - Plan to evaluate
  - Change management

Other Findings

- GMAT filter during admissions has no significant effect in terms of the quality of students in the university.

Questions raised

- Why do people leave?

Potential interview questions to ask students:

- At what point in time were you sure on choosing your master program.
- Which moment would you have liked to have to better decide?
- At what moment did you have enough information to decide on your master? What channel was most important to you?
- What would you have done if your chosen program wasn't available at MU? Would you have gone to another university?

Appendix A3: Interview Notes Nico Rasters 9-5-2019

Nico "The Man" Rasters - Functional Information Manager

Details

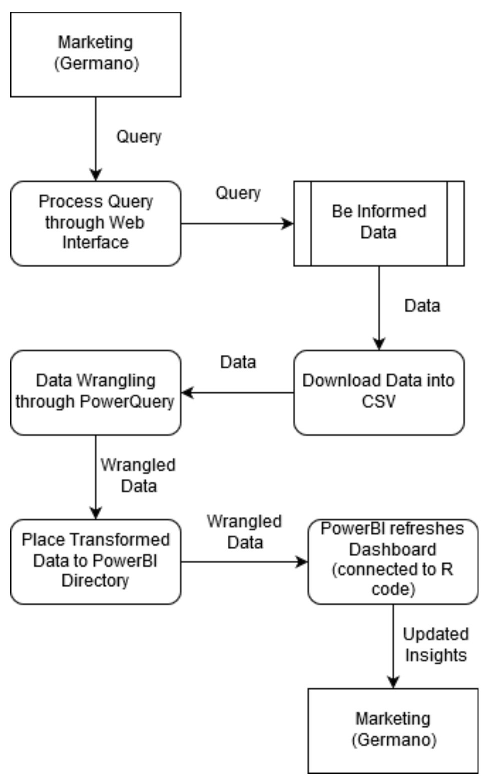
- PowerBI pilot is starting this month. It is headed by Drs. Rob J. Maessen EMFC (rob@maessenadvies.com; +31 6 38 899 640)
    - They plan to use just the free version. It is more than sufficient.
    - Otherwise, the pro plan costs 8 euros per month and will enable the publication of the dashboards in the cloud. Enable mobile version.
- Planned refresh rate: once per period (after re-sits).
    - Due to this, automatic refresh is not needed. Manually updating is the most sensible solution.
- GDPR must be considered.

Tools Used

- Dashboard - PowerBI (Currently in trial)
- OLTP - SAP
- OLAP/Data Warehouse - Be Informed (BI for short)
    - Refreshing daily.
- Data wrangler - PowerQuery in Excel
- Other details
    - Data are in flat format (CSV)

Data Flow Diagram

* Correction: Be Informed Data WAREHOUSE



Appendix A4: Qualitative Interview analysis with Students

Page 1: 12 Interviewees on the first 5 subjects

1	Information Channels		Before applying to Bachelor		
2	Awareness of Information channels provided by University	Must-have information channels	What would make the student apply to one university over another	Completed tasks when applying to Bachelor	User pains in applying (13-15)
3	Interview 1: External Master student, Male, 26, Dutch				
4	Website, open-days, brochures, webinars, social media pages, counselling.	Open-days, website and social media pages	A university that provides great communication and has great reputation.	HBO-nights (provided by high school), open-days, study counselling at school and private, summer school, website visits.	The large offer that was provided by multiple schools. I felt that my choice was going to affect the rest of my life. This had more to do with me being young than it had with any university.
5	Interview 2: External Master student, Male, 22, Dutch				
6	Zuyd Intranet	Their main visitors page. When they are describing their bachelor program they should also include possibilities for a master in Maastricht.	I would first pick on quality and then distance.	Websites on internet.	Not knowing how to apply directly
7	Interview 3: External Master student, Female, 23, Dutch				
8	University website with info, brochures, open days	Online marketing, these days students do so much things on the internet, info at the different bachelors. This was lacking at my bachelor institution	The course program, name/image of the university, language, the possibility to go the a university from HBO, the city and student life outside uni	Not much, I wanted to go to the hotelschool. This was a last minut cancel and facility management was the closest in terms of courses an knowledge compared to the hotelschool. So I did little research.	Everything went very smooth. However, I had a Belgium education so they had to perform an extra check if I could enter Zuyd. I knew this before and Belgium education is harder, so entering any university should not be a big issue. // Timing. It takes a long time to really be sure if you are able to enter a university. At that time you are sometimes to late too think about other options. Even though, you can start earlier with checking everything, but I did not wanted to go anywhere else so I did not think about the other options.
9	Interview 4: External Master student, Female, 23, Dutch				
10	SCOPE events mostly and information boards in the faculties	Information boards in the schools, posters, e-mail, social media messages	The evaluation rate, such as 'Top Master Program 2019 in Maastricht University' and the type of education. The PBL is at far most the most efficient way to learn I have experienced.	Open Days and information events at my High School	
11	Interview 5: External Master student, Male, 23, Dutch				
12	Website , Social media, folders, banners, open days, studiegids / keuze etc.	Website, banners, open days, folders	First: is the program of good quality, decided by reviews of previous students & online reviews. Second: Method of teaching	Receive info folder at high school about possibilities in the area, then visit open day and pick a program	no obstacles, easy for Dutch students if you meet the minimum requirement. // Vague information, but this is normal, you can't asses a study without seeing into the daily life
13	Interview 6: Internal Master student, Female, 24, Dutch				
14	Study counselling, website of the university, Lectures provided by the university at my high school, Open days/Brochures	Social Media pages. But these weren't very relevant yet when I registered. However, open days are definitely an important aspect as well. Interaction between student and university is the most important.	Distance from my home, Reputation	I went to open days and visited the website. Beforehand I knew that I was going to study Law. After high school I studied HBO law for one year where I got my propodeuse. This allowed me to admit to the university.	No I did not. However, It was kind of annoying that you had to upload various documents in different formats e.g. pictures that had to be a certain size and resolution. Those were only minor issues though.

1	Information Channels		Before applying to Bachelor		
2	Awareness of Information channels provided by University	Must-have information channels	What would make the student apply to one university over another	Completed tasks when applying to Bachelor	User pains in applying (13-15)
15	Interview 7: Master Student/Marketeer, Male, 25, Dutch				
16	Wesbite, Communication platofrms (Phone/mail etc), Option to visit different courses with students	Social media pages, Website	Preferred composition of courses in a bachelor or master programme. Employment statistics after graduation	Google courses in the program. Visited the university under the guidance of a student. Visited universities on open days	At several universities there were no student present, so you would only hear the opinion of the teachers.
17	Interivew 8: Current Bachelor Student, Male, 25, Dutch				
18	After graduating Zuyd University sent me a magazine with information regarding masters. Furthermore, I am aware there are tools helping you to make a choice, for example the 'Keuzegids'.	LinkedIn and other social media.	The location of the university is most important for me.	I went to an information evening in where you could go to short presentations of several bachelors. Furthermore, I did some field research and read about other people's experiences.	
19	Interview 9: External Master Student, Male, 27, Dutch				
20	UM website, Open days, Info lectures, social media	A website, open days, social media	Way of teaching, what courses the program includes	um website	The broad variety and many similarities that the different programs have.
21	Interview 10: Internal Master Student, Male, 22, German				
22	open days, web site, word of mouth from other master students	word of mouth, web site	good reputation, easy application process	Asked friends who studied there before // I checked the web site	finding the right study
23	Interview 11: Internal Master Student, Male, 23, Belgian Filipino				
24	Master open days.	Student Portal	Prestige; Language; Costs; Location	Googling mostly	Relative to the Philippines, UM was really smooth so I didn't notice anything
25	Interview 12: Internal Master Student, Male, 21, German				
26	Open Days, Brochures	Social Media, open days	Reputation, Location, Available Programmes	Attended Bachelor open day, searched website	crowded website
27					

Page 2: 12 Interviewees on the subjects 6 - 10

	During Bachelor		After Bachelor	
User gains in applying	**Amount of counselling or guidance used during Bachelor (per year)**	**Additional support mechanisms used during Bachelor**	**Things that would be done differently in retrospect**	**Potential improvements for current/future information channels**
	2 hours	None	No, I wouldn't. It got me to the point where I am today.	A view on a typical day of a student to gain a 'real world' indication of their tasks.
	NONE	NONE	d have chosen a bachelor than had more relationship with m	More information on the website.
I cannot really remember.	I received a new counsellor, which was not really a person w	I went to the dean once for advice	nore insight about my other options but at the end I am happ	I do not really know. Maybe in terms of time management. be faster or give some insight in how long everything takes. For example, do not say you can maybe get more info about your submission in about 3 to 4 months. That is a crazy long time
because I took a pre-master everything was arranged for us.	None			Nothing at the moment, I'm not that connected with the University, since I only take my Master there.
A short introduction provided me with better information	None		No I wouldn't have, i was extremely lucky to chose the one I did	Maybe they can offer a way to think about what people are most passionate about. It's not really up to the university to figure that out though, but perhaps it increases the chance of applying.
I cannot really remember any initiatives provided by the university.	Once for study counselling during my entire Bachelor. Around 20 minutes per year.	No, I didn't other than the study counselling.	I'm happy with my choice. In retrospect I might have applied to a university which had a higher rating in Law. MU has a great reputation, but not specifically for law.	Social Media pages since these are way more important in comparison to when I applied e.g. more posts on social media so that potential students can find it more easily

	During Bachelor		After Bachelor	
User gains in applying	**Amount of counselling or guidance used during Bachelor (per year)**	**Additional support mechanisms used during Bachelor**	**Things that would be done differently in retrospect**	**Potential improvements for current/future information channels**
Before classes started the university organized a weeklong event so new students could get used to the city and build a social circle in an unknown environment.	2 hours per year	While I was writing my bachelor thesis I made appointments with a writing guide for advice.	Nothing, during high school students of the radboud university visited my school and told about their courses. One of these was communication and information sciences and it sounded like a perfect program for me. Because of this I did not search for very long.	I would like it if universities would explain more what kind of employment opportunities are generally available to graduates.
	Once in my entire career. 15 minutes per year		There was an open day organized by Zuyd University which I did not go to at the time, that is something I would have done differently.	Social Media strategy
The availability to assist students by phone or by email	2-3 times per year. Roughly 2 hours per year	When I ran into problems, I used to contact the uni either by phone or by email.	I would to went to the open day to get more insights and speak to people about the program.	Social media, especially when I compare this with UM's social media. Zuyd does not put a lot of effort in this.
Sharing success, which is a web site to get to know other students	-	mystatlab, studentportal	Nothing, I'm satisfied with my bachelor program and the university I chose	MyUM
I like that they cancelled my requirements for an English test when I went to admission to personally prove I can speak it. And point out in the transcript of records that I took English courses in my old university.	-			It was confusing that the school and the department of Education had different websites
sharing success portal				better overviews, timelines, one-on-one guidance

Page 3: 12 Interviewees on the subjects 11 - 16

Master Stage

Reasons for applying to a Master program at Maastricht	Information channels which were used in applying to a Master 24-25	Information channels that would have helped with deciding on a Master	User pains	User gains	(Hypothetical) What if Master wasn't available, would UM still be chosen as university?
I wanted to further specify my knowledge and gain an extra edge in something I find interesting.	Lectures on programs at my university, watched general webinar, website, interviewed current students // I found the lectures on different programs very informative. Additionally, I liked browsing the website since it contains complete information on all the different programs.	Some indication of a typical day of a student. Something like: A day in the life of a ... student.	I did, but this was mainly due to the university I did my Bachelor at. Because of misaligning deadlines, I couldn't register for a master program right after my bachelor ended. This is why I had a gap of 6 months between the end of my Bachelor and start of Master.		I would have stayed in Maastricht and pick another program. I did a pre-master course so I was already connected to MU
Because that was the only pre-master offered by my college institute.	Looking through websites and talking to a student advisor // Internet / websites	More information on the regular website or on intranet on the pre-master			I would go to another university. However, I would have kept in mind the travelling distance. Or I would pick a similar Master program.
I wanted to gain more knowledge and grow as a person to gain more in my (working) life.	I wanted to go towards the behaviour direction. So I looked online and searched for the different masters. // Online information and mostly the open days	More online information and more open days as well		The easy switch/admission when you finished te premaster to the university.	I for sure would go to another university if possible. Since I liked the program. If there would be no similar program at other universities available I would look at other options. // My first choice was a master at the University of Utrecht, however, they did not want to have HBO Bachelor students. So I went for my second choice, with similar courses.
Because I wanted to gain more knowledge and according to my experience with HBO level managers during my intership, I preferred not to end up there but wanted to start higher on the managerial ladder.	The channels I received in my Bachelor program, such as the Board of Admissions and the open day in the SBE faculty.				No. The pre-master gave you the possibility to only apply to Maastricht Univeristy.
To increase my knowledge, did not feel ready for the job market.	Look in the area & which programs were best on internet. // Information sheet on website,	Talking to actual students, no sales talk.		Short introduction which provides quick information	If similar program was available at UM i'd evaluate my options. Really depends on the programs we're talking about. If nothing is available in my area I would not pick UM.
It was close to my home town and I studied one year at Zuyd. This school had a lot of connections with Maastricht which also pushed me in the direction of Maastricht university.	Open days and talks with professors. I already did my Bachelor at UM, so I wanted to do my master here as well.	Not any really. During my bachelor I already had my choice clear. In Dutch law there only, a small amount of choices for you master. During my Bachelor you find out what field you really like, and you base your choice on that.	Yes, the transition between my Bachelor and Master didn't went smooth. Officially they registered me one month too late when I started my Master. This caused some problems for my DUO and student account.		I would have gone to another university. I specifically wanted to do Dutch Civil Law.

Master Stage

Reasons for applying to a Master program at Maastricht	Information channels which were used in applying to a Master 24-25	Information channels that would have helped with deciding on a Master	User pains	User gains	(Hypothetical) What if Master wasn't available, would UM still be chosen as university?
The master program was a continuation of my bachelor program so there was a good fit. Furthermore, I already had housing and a social circle in Nijmegen. I thought it would be easier to stay in Nijmegen.	Meetings with my student councillor. // Employment statistics and an in-depth description of the courses.			An afternoon in which all the new students met and got a glimpse of the master program.	I would have chosen a different University. I enjoyed my master program and would have gladly attended a university in a different city in order to pursue it.
After my bachelor I wanted to further specialize/develop myself. Specifically Maastricht since I already lived here and did a premaster which let me apply to UM	Open day, info lecture at zuyd, UM website, social media (online lectures) // Um website, info lecture.	More information on the UM website, with student experiences.	Uploading the required document was a struggle since I had to reupload documents I already uploaded a few times.	The assistance that they provided through the SSC, by email and by phone and by actually going there for help.	I would have applied to a different program at MU
Easy transition // I was used to the university already // the program seemed very promising	Website of Maastricht University	The internet in general	There were none, Studielink made registering very easy	The Facebook live seminar was something I liked	I would have chosen a different program in the same direction, but still at Maastricht University.
Peer pressure	Google, website and studentportal	-			I would have probably go somewhere else.
convenience, familiarity with university, specific programme	searched website, requested brochure, website	clearer website, consolidated information in one place, overview of all channels			I would have applied to a different university

Page 4: Demographics on the interviewees