MANAGEMENT PROJECT

Fontys IT Student Success Factors & Correlations

Authors
Joey Wolfs, Tim Backx, Peter Kreuwels, Roman Siabro

Obelisk Solutions
Data Driven Business Lab
09/01/2018

Foreword / Preface

The purpose of this project is to establish which datasets can (co-)determine a student's' study performance by looking at both internal and external data and process insight available (FODI) The goal is to find out and prototype a tool that gives the student better control of their own performance with real time insights and performance indicators based on not just static data but based on relevant sources and possibilities. Think about mobile collected data correlated to grades for example! Obviously, it is not just about what is technically possible but also about what is ethically reasonable and desirable. Hence a broad scoped research to determine what is relevant, important, and how such a tool should look like.

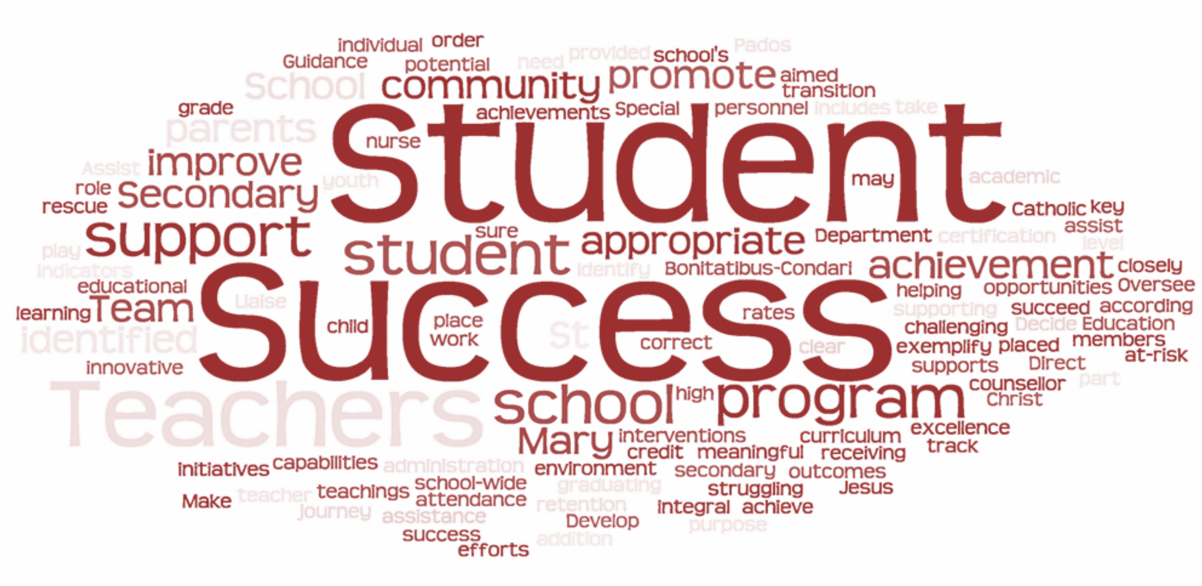


Table of Contents

Executive summary

This document serves as the central document including the research towards the possible solutions and documentation regarding the project mandated by Fontys IT group in investigating student

This paper considers the compilation of knowledge concerning the process of capturing data to visualizing factors indicating student success. The triumphs and tussles, discoveries and obstacles.
After a short glossary containing acronyms you might not know, we'll follow with an outline of the project. Something about the company and a description of the assignment.
Also to be found are the risks, phasing and approach, methodologies, tools & techniques.
Related article. Insights and recommendations to serve as solution to the project mandate, including cost breakdown and expected ROI.

Glossary

PBI	Is a business analytics service that provides interactive visualizations with business- intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on information technology staff or database administrators.
KPI's	Key Performance Indicators.
FODI	FODI = Fontys open data initiative: Is a centralized institutional that could be used to: • Derive insights from data that help drive institution policies • Deploy data-driven apps that positively impacts student success
AI	Artificial Intelligence
ML	Machine Learning
MVP	Minimal viable product
BPM	Business process modelling
PCA	Principal component analysis

Chapter 1: Outline

We live in a world where a lot of data is collected. Administration is digitalized, grades are collected database management systems rather than within offices with dozens of drawers stacked to the ceiling. The mobile devices evolved to high tech computers that people carry around 24/7. They send and receive copious amounts of data. We live in the era of Big Data!

Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating and information privacy. - Wikipedia

Within Fontys copious amounts of data is collected, yet the student is not yet benefitting from the opportunities that are present. We believe that it is possible to use all this information to see correlations and patterns, how variable X Influences Y by combining data sources and analysing these new datasets to support the question how to determine the success of a student at Fontys. How can this be linked back to the student? Can new discoveries be made?

Chapter 2: About the company

Fontys University of Applied Sciences is a Dutch university of applied sciences with over 45.000 students in several campuses located in the southern Netherlands. The three largest Fontys campuses are located in the cities of Eindhoven, Tilburg and Venlo. Fontys wants to highlight that it is a source of knowledge for students. Fontys offers 200 bachelor's and master's study programmes in the fields of economics, technology, health care, social work, sports and teacher training. A selection of these programmes is offered in German and English. Fontys International Campus Eindhoven is located in the South-East of the Netherlands, the so-called “Brain port region” which is an international centre of science and technology. This is the place where high-tech companies design and improve state of the art industrial and consumer products for customers all over the world.

Chapter 3: Assignment overview

3.1 Current Situation

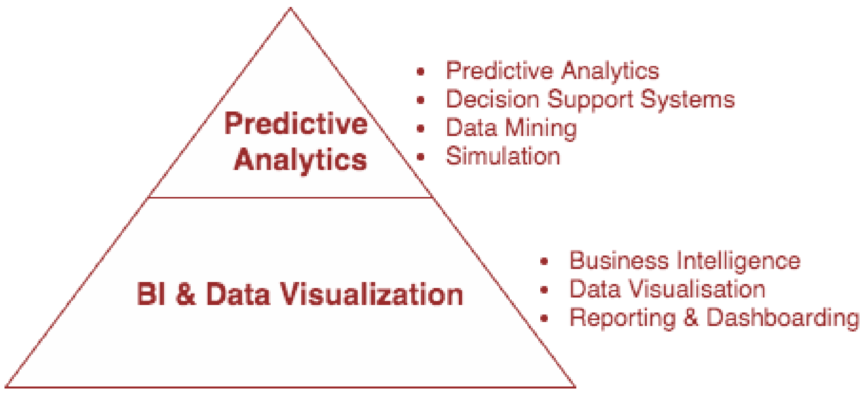
Currently Fontys students merely have access to their grades, schedules and courses. There are several mobile applications out there for students to see their schedule, grades or both depending on the application used. Up till the present no statistical analysis is used with the student in mind as end-users. We're going to research the potentials to change that. And to help students directing their own success within the Fontys university. What are factors which can be influenced to improve student graduation rates?

- Cost of education has been steeply increasing and as a consequence student debt has also been rising steeply
- Increase quality / value of education
- How do we ensure students graduate on time, don't drop-out, get better jobs?

3.2 Description of the assignment

The baseline of the assignment is that a student should be the director of his own success. The idea is that students have to be able to take control and specific action to steer their study achievement's. Hence, we are going to research which internal & external factors can (co-)determine a student's success and how this could be best linked back to them. This can be achieved by analysing the current data available e.g.(FODI) to find, determine and measure relevant success factors.

- Research and show data related to a student's activity and profile, predict a student's success
  - Identify key attributes that influence the time to graduate which will assist institution act on some of the identified attributes.
  - Create Dashboard for student's activity and profile.
- To introduce policies and tools that positively impact student success
- Study factors that determine student success from (static & dynamic) (un)structured data sources



"What is success"?

What would be your response? Most people would say: Succeeding, achieving your goals, being resilient, something one is working towards. And so on.

When they think of success they usually think of material benefits, happiness, and accomplishment. Certainly, success can include all of these things and more. Now here's a personal question, what is success for you? Therefore, it is crucial that to avoid any misunderstandings here is a simple delineation of what success defines within this project: The students grade compared to averages, the estimated time to graduate, graduation chance %, and the like. Measures that relate to how the student performs with regard to what is deemed 'student success' within the vicinity of Fontys Hogescholen.

3.3 Limitations

The main constraint of the project is timing; the project period starts on 28th of August 2017 and lasts until 14th of January 2018, so within 20 weeks we have to deliver the final product.

Constraint 1: Time

The project must be completed within 20 weeks.

Constraint 2: Integrity

The project is done within boundaries and respecting the rules.

Assumption 1:

The Project benefactor will participate in the timely execution of the Project Plan (i.e. timely approval cycles and meeting when required).

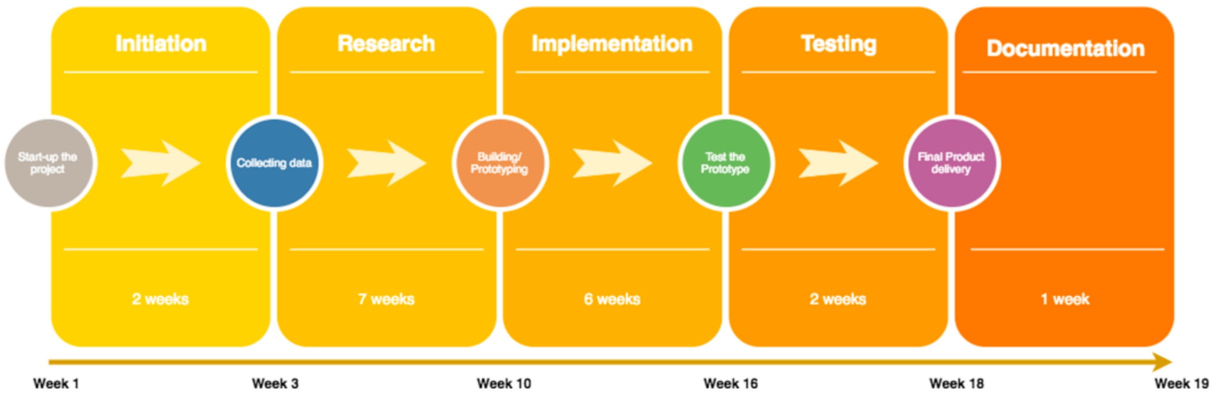
Assumption 2:

All required software/ hardware is provided to project team.

Assumption 3:

The Project Plan may change as new information and issues are revealed. And will deviate from the foreseen different scenario's.

3.4 Phasing



Phase 1: Initiation

Activity: Start-up the project

Duration: 2 weeks

Tasks:

- Kick- off meeting
- Declaring scope of the project
- Project planning
- Know- how

Phase 2: Research

Activity: Data collection

Duration: 7 weeks

Tasks:

- Reviewing available data
- Documenting findings
- Decide on project continuation

Phase 3: Implementation

Activity: Prototyping

Duration: 6 weeks

Tasks:

- Designing
- Creating a dashboard

Phase 4: Testing

Activity: Testing the dashboard

Duration: 2 weeks

Tasks:

- Create a test plan
- Discuss test plan with a client
- Document all succeeded also as failed tests and a list of improvements

Phase 5: Documentation

Activity: Final product delivery

Duration: 1 week

Tasks:

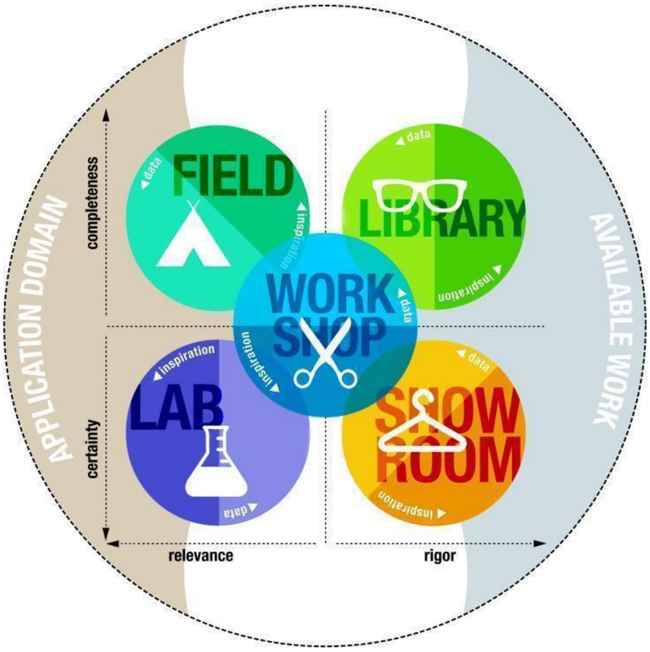
- Finalise Research Report
- Executive summary

3.5 Research methodology (DOT)

Throughout the duration of this project the Five Strategies Research Framework will be used, which is also known as the “Development Oriented Triangulation (DOT) framework” (Van Turnhout, et al., 2014). The framework consists of five strategies: Library, Field, Workshop, Lab, and Showroom. The Library strategy is used for investigating available resources, collecting information that is related to the “research questions” of the project. Field research is done to explore application context. The Workshop strategy also known as “hands on” strategy deals with the implementation and development - here researcher implements his or her ideas. The Lab strategy compares what has been developed as the result of the Workshop and Field strategies. Last is Showroom strategy - is aimed at demonstrating and justifying the results. Different research procedures are applied in various phases of the assignment. The research phase essentially is managed with use of Field and Library strategies. While applying the Field procedure the team is expected to gather the necessities of the tasks and have a clear overview of the assignment. The project team can utilize short meetings and observe users. With the assistance of the Library strategy we can look into the internet, literature and other documentations available in order to learn the systems Fontys IT uses and discover the tools to be used for the project.

3.6 Tools and techniques

Trello: it is a tool used to organize and prioritize projects, tasks - scrum. (Trello, 2017) A user can simply check the status and progress of each task. Obelisk Solutions team used this tool to set daily goals and to monitor performance of a group overall.



Draw.io: is a free online tool for creating different types of diagrams. The tool will be used by the team to create organizational charts and other diagrams.

Canva.com: is a free online graphic-design tool. It has an easy to use drag-and-drop interface and provides access to over a million photographs, graphics, and fonts. It is used by non-designers as well as professionals. The tools can be used for both web and print media design and graphics. Our team will use this tool to create infographic.

Power BI: is a business analytics service that provides interactive visualizations with business intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on information technology staff or database administrators.

Chapter 4: Procedures and Outcomes

4.1 Launch

The first couple of weeks were rough as project team was overwhelmed with possible scope of the assignment. Even though the team took its time to understand the assignment, within time we were able to break down the project in smaller, more manageable pieces and define the scope. During initiation phase, Obelisk team had to understand the assignment, current business processes, existing applications, client/server systems and business requirements of Fontys IT. During initiation of the project Obelisk Solutions team got introduced to the parties involved in the project, we became more aware of the assignment specifics and were able to complete the infographic (see Appendix A) were all relevant information so far (company structure, members, projects) is accumulated in one neat overview.

4.2 Investigation

Library.educause.edu says:
"Just 56 percent of bachelor's degree students graduate in six years and only 29 percent of associate's degree students graduate in three-years. Among low-income students, completion rates are even lower: 45 percent for full-time and 17 percent for part-time bachelor's degree seekers and 12 percent for full-time and 4 percent for part-time associate's degree students.
So what does it take for students to pass their courses, return to continue their studies each term, and graduate college with a degree, prepared for a career or further study, citizenship, and a lifetime of learning? This page is full of resources for institutional leaders, faculty, advisors, CIOs, and others who are interested in learning how their colleagues and peers are addressing these challenges and tackling student success. Check back often for new and valuable resources, and join the conversation on social media using the #StudentSuccess, #AdvisingReform and #DigitalLearning"
This strongly connects to our situation within FHICT where we see a similar situation & similar numbers.

We agreed that when analysing datasets in search of correlations there are three different types of analysis:

- Prognostic analytics: Statistical modelling and forecasting to understand the future and answer "what could happen"
- Descriptive analytics: Involves data aggregation and data mining techniques to provide insight into the past and answer "what has happened"
- Prescriptive analytics: The optimization and simulation of algorithms to advice on outcomes and answer: "what should we do"

We brainstormed about **internal** factors in Fontys that could contribute in determining relevant factors regarding the determination of a student's success.

- Static administration data
- Student satisfaction (by survey)
- History of student (identity, socio-family past, academic past, age, ethical background, race and gender)

- Student’s involvement in studies (participation in extracurricular/optional activities, meeting with lecturers) = ‘behaviour’?
- student’s progress (wwwProgess: grades, credits, courses taken, minor’s & internships)
- Canvas (and other platforms: grades, attendance, courses taken, course info)
- FODI (accumulation of sources)

Considering that the most promising results can be obtained by combining these sources with **external** (dynamic) ones to see how success gets influenced:

-Travel information (trains, traffic jams, road works)
-medical information (could be very hard to obtain; law & privacy)
-personal integrated agenda (reflect one individual's activities to the average)
-Foreign vs local applicants (could foreign students be more motivated?)
-When internship is done (influence of field experience)
-Changes in info (major, program, campus)
-% of homework submission (does homework and repetitive exercise work?)
-% Resits (course/semester)
-Submission speed (effect of dedication on results)
-Participation ratio (effect of following lectures on results)
-Popularity (effect of social status on results)
-Network activity (access point logistics)
-Facility use (printers, canteen)

Some correlations we’d like to check if possible or we expect Influence the relevant variables. Also this gives an explanatory insight into what kind of things we want to measure:

1. How higher the time present in classes the higher the grades. (participation vs results)
2. The more Feedback has been asked the higher the grades. (retro/introspect vs results).
3. The (fe-)male brain works different, each gender outperforming the other on different fields based on averages every time (might be taboo but true).
4. If a student has to spend a higher amount for travel time, the total time present is also higher. (forensics vs total time spent).
5. The more classes in a subject combined with attendance the higher the average grade. (diversity vs success).
6. The more afterschool studying the higher the grade. (participation vs self-study).
7. The higher the teacher scores are the higher the student grades are. (teacher quality vs student success).
8. The earlier the class the lower the grade. (research proves mornings are not productive)
9. If classes are only on Monday, the grades will be lower. (after weekend absorption capacity is lower).
10. The weather affects the grades. (surroundings vs morale).

11. Student who do a second education are more likely to pass (ambition vs performance).
12. Students below 22 are more likely to fail. (adolescence vs success).
13. Students with clear goals (extracurricular) perform better.
14. The higher the study results the better the student's feedback towards the institute.
15. Students with high grades have good job expectations.
16. There is software available that filters out the strongest correlations from various datasets combined, determining factor x that influences the variables between we see a correlation thus mining for us what influences success.

Looking at the above we think things need to be measured that currently are not collected (yet), or stored in such a way that It cannot be translated to the information desired, some examples of this are:
- Social economic situation (living, parents, friends, income)
- Average grade (compared to, class, lecturer, etc.)
- Damaged goods ("a backpack")
- Ethnicity/Nationality
- Gender (m/f/o)
- Previous education (domain: e.g. technical-background)
- Motivation (reason for studying)
- Personality (the big 5)
- Financial situation & debts

Some of these factors are in the system, others can be derived or made an estimate of using the available factors. Others are simply not obtainable due to privacy issues. **Strong sound BI solutions require a** new privacy regulation between Fontys and the student besides limited variables could prevent their transformation from data to insight. We see that data is stored for their designated purpose but maybe it should be stored with analysis in mind?

Online Articles

Following are articles underwriting to the insights sought from studentsucessjournal.org which gave us greater comprehension in the dominion of success and its measuring with technology. These articles used some of the previous mentioned measures.

Interactions among students' prior learning, aspiration, confidence and university entrance score as determinants of academic success

Theo Papakonstantinou - Monash University

Abstract

We studied the effect grade aspiration, confidence in achieving that grade, prior learning and university entrance ranking had on first year biology students' final grade. We hypothesised that (1) students with higher aspiration will achieve higher grades than those with lower aspiration; (2) students with prior biology learning will have a higher grade aspiration and a higher confidence of achieving that aspiration than those without such learning; (3) university entrance rank will impact students' final grade; and (4) students with prior biology learning will achieve a higher final grade than those without such study. We found that Hypotheses 3 and 4 were supported, Hypothesis 2 was partially supported, and that Hypothesis 1 was unsupported. If these results reflect broader patterns - that undergraduate student grade aspiration is not a predictor of their subsequent final grade - then targeted information and curricula scaffolding must be provided to better align student aspirations with their actual academic achievement.

Information

The purpose of this work is to develop metrics for evaluation of medical physics graduate student performance, assess relationships between success and other quantifiable factors, and determine whether graduate student performance can be accurately predicted by admissions statistics. A cohort of 108 medical physics graduate students from a single institution were rated for performance after matriculation based on final scores in specific courses, first year graduate Grade Point Average (GPA), performance on the program exit exam, performance in oral review sessions, and faculty rating. Admissions statistics including matriculating program (MS vs. PhD); undergraduate degree type, GPA, and country; graduate degree; general and subject GRE scores; traditional vs. non-traditional status; and ranking by admissions committee were evaluated for potential correlation with the performance metrics. GRE verbal and quantitative scores were correlated with higher scores in the most difficult courses in the program and with the program exit exam; however, the GRE section most correlated with overall faculty rating was the analytical writing section.
Students with undergraduate degrees in engineering had a higher faculty rating than those from other disciplines and faculty rating was strongly correlated with undergraduate country. Undergraduate GPA was not statistically correlated with any success metrics investigated in this study. However, the high degree of selection on GPA and quantitative GRE scores during the admissions process results in relatively narrow ranges for these quantities. As such, these results do not necessarily imply that one should not strongly consider traditional metrics, such as undergraduate GPA and quantitative GRE score, during the admissions process. They suggest that once applicants have been initially filtered by these metrics, additional selection should be performed via the other metrics shown here to be correlated with success.

The parameters used to make admissions decisions for our program are accurate in predicting student success, as illustrated by the very strong statistical correlation between admissions rank and course average, first year graduate GPA, and faculty rating (p < 0.002). Overall, this study indicates that an undergraduate degree in physics should not be considered a fundamental requirement for entry into our program and that within the relatively narrow range of undergraduate GPA and quantitative GRE scores of those admitted into our program, additional variations in these metrics are not important predictors of success. While the high degree of selection on particular statistics involved in the admissions process, along with the relatively small sample size, makes it difficult to draw concrete conclusions about the meaning of correlations here, these results suggest that success in medical physics is based on more than quantitative capabilities. Specifically, they indicate that analytical and communication skills play a major role in student success in our program, as well as predicted future success by program faculty members.

Finally, this study confirms that our current admissions process is effective in identifying candidates who will be successful in our program and are expected to be successful after graduation, and provides additional insight useful in improving our admissions selection process.

Exploring the Correlation Between Non-traditional Variables and Student Success: A Longitudinal Study.

Strickland HP, Cheshire MH.

BACKGROUND:

The purpose of this project was to determine whether a correlation exists between the traditional admission criteria of grade point averages with the potential admission criteria of emotional intelligence (EI) scores or critical thinking (CT) scores to predict upper division student outcomes.

METHOD:

A quantitative, longitudinal design was selected to examine the identified variables to predict undergraduate student success. The recruiting sample included a convenience sample drawn from 112 junior-level undergraduate nursing students beginning their first of a five-semester nursing program.

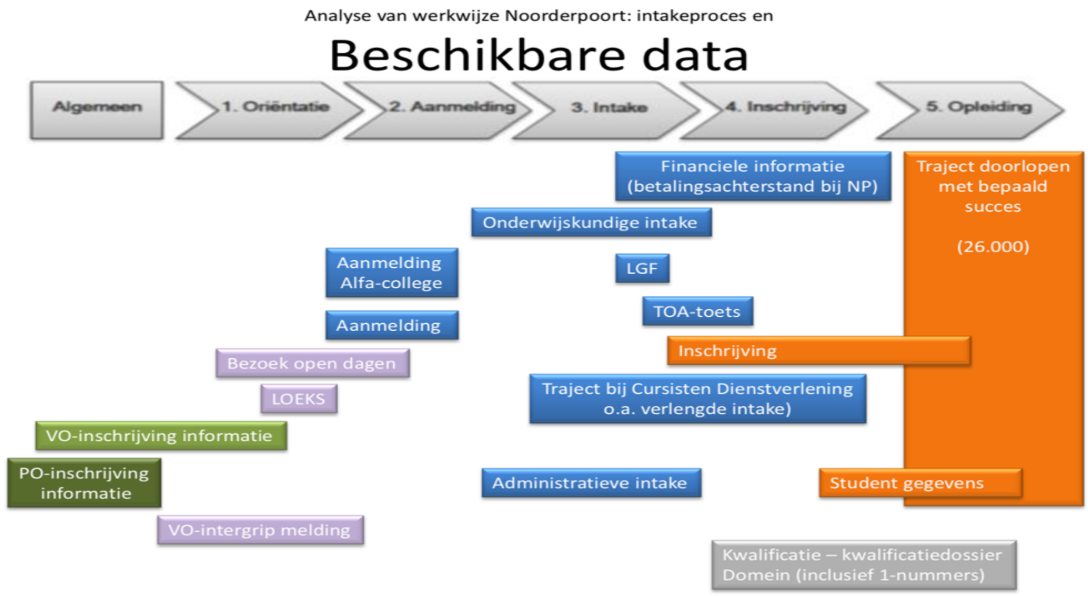
RESULTS:

EI and HESI® CT scores did not significantly correlate with main analysis variables.

CONCLUSION:

Although EI and CT scores were not significant in this study, it remains vital to incorporate EI and CT activities throughout the curriculum to develop students' ability to think like a nurse and, therefore, be successful in nursing practice.

Points of data creation in time by Martijn Broekhuizen (NOORDERPOORT) en Marius van Zandwijk (Kennisnet) research at ROC some years ago (found via google.com):



-This neatly shows when a 'student object' creates certain data throughout its lifecycle

An enlightening article by William Vorhies:

Summary: Higher education has been a little slow on the uptake to use advanced analytics to improve student success but now with the technology that allows us to marry and analyse structured and unstructured data, including streaming data, a number of successful projects are underway. What changes in the most current studies is the extensive use of unstructured data integrated with structured data. It wasn't until about 2007 that our ability to store and analyse unstructured data took off and now we have data from a variety of new sources.

Learning Management Systems is one of the most important new sources. These are the on line example when they submitted assignments relative to the deadline, how they interact with instructors and classmates in the chat rooms, and a variety of click stream data from library sites and the like.

Sensor and Wi-Fi data showing frequency and duration on campus or at specific locations like the library.

Student Information Systems.

These aren't necessarily new but greatly improved in level of detail regarding classes enrolled and completed with regular grade markers.

Social Media.

What is standard now in commerce is becoming a tool for assessment of progress or markers for concern. Positive and negative social media comments are evaluated for sentiment and processed as streaming data that can be associated with specific periods in a student's term or passage through to graduation.

The goals of each study are slightly different. Some are seeking better first year integration programs which are so important in student long term success. Some are focused on the transition from

Community College to four-year institution. But universally they tend to look at some similar markers that would allow counsellors and instructors to intervene. Some of those common markers are:

- **Predicting first term GPA.**
- **Predicting specific course grades.**
- **Predicting reenrolment.**
- **Predicting graduation likelihood, some focused on getting students through in four years, others getting them through at all.**

As in any data science project, each institution seems to have identified its own unique set of features drawn from both the traditional structured and new unstructured data sources. Paul Gore who headed one of these studies at the University of Utah had a nice summary of the categories that's worth considering. He says the broad categories of predictive variables fall into these six groups:

1. Measures of academic performance:
Academic engagement or academic conscientiousness: in other words, how seriously does the student take the business of being a student? Does the student turn in assignments on time? Attend class diligently? Ask for help when needed?
2. Academic efficacy: the student's belief and confidence in their ability to achieve key academic milestones (such as the confidence to complete a research paper with a high degree of quality, or to complete the core classes with a B average or better, or their confidence in their ability to choose a major that will be right for them).

3. Educational commitment: This refers to a student's level of understanding of why they are in college. Students with a high level of educational commitment are not just attending college because it is "what I do next" after high school (i.e., in order to attain a job or increase their quality of life); these students have a more complex understanding of the benefits of their higher education and are more likely to resist threats to their academic persistence.
Campus engagement: This is the intent or desire to become involved in extracurricular activities. Does the student show interest in taking a leadership role in a student organization, or participating in service learning opportunities, intramural sports, or other programs outside of the classroom?

4. Measures of emotional intelligence: EI/CT
5. Resiliency: How well does the student respond to stress? Do small setbacks throw the student "off track" emotionally, or are they able to draw on their support network and their own coping skills to manage that stress and proceed toward their goals?
Social comfort: Gore notes that "social comfort is related to student outcomes in a quadratic way -- a little bit of social comfort is a good thing, while a lot may be less likely to serve a student well, as this may distract their attention from academic and co-curricular pursuits." (aka too much partying).
Where the studies were willing to share, the fitness measures of the predictive models look pretty good, achieving classification success rates in the 70% to 80% range.
From our data scientist friends at Pivotal who are featured in the webinar we also learn that

administrators and counsellors are generally positive about the new risk indicators. There was always the possibility that implementation might be hampered by disbelief but there are some notable examples where there is good acceptance.
Some of the technical details are also interesting. For example, there are instances where the models are being run monthly to update the risk scores. This allows the college to act within the current term and not wait for the term to be over, which might be too late.
And there are examples in which the data is being consumed not only by administrators and counsellors but also being pushed directly to the students through mobile apps.
I originally thought to include a listing of the colleges that were undertaking similar projects but a Google search shows that there are a sufficiently large number that this is no longer a completely rare phenomenon. In its early stages to be sure but not rare.

Finally, I was struck by one phenomenon that is not meant as a criticism, just an observation. Where the research and operationalization of the models was funded by say a three-year grant, it took three years to complete the project. But where our friends at Pivotal were embraced by their client, four data scientists, two from Pivotal and two from the university had it up and running in three months. Just saying.

An interview on success, how it it more than passing your college and the future:

As student success becomes a priority for increasing numbers of institutions—and as the "success" part of that equation begins to encapsulate more than just completion—institutions need to find ways to leverage data to transform their student experience and support the broad success of their learners.
Though it's a high priority for every administrator at every college and university across the country, student success is a somewhat nebulous concept. It's hard to define and that has left postsecondary leaders struggling to define what it takes to achieve it. But it's not impossible. Drawing on recent research as well as his work with the Student Success Collaborative, Ed Venit shares his thoughts on what student success means in the modern era and discusses the impact Big Data and analytics can have on delivering a student experience that drives success.

The EvoLLLution (Evo): What are some of the key lessons you and your colleagues learned from your research on the evolution of student success, which included the development of this infographic?

Ed Venit (EV): There were really three things we learned from doing this. The first was that, unlike biological evolution, nothing seems to go "extinct" in this space. We can go back to the 1970s to see the first pioneering work on "student integration" theory. The principles student success leaders talked about then are still embedded in a lot of the student engagement programs that you might see through a student affairs office or in a first-year experience initiative. Over time additional layers of practice have emerged. In the 1980s we saw schools targeting support to specific student populations, including traditionally underserved students. In the 1990s we saw more work on the transition to college and the first-year experience. Lots of technology came into play during the 2000s, like early warning systems. What's interesting here is that these efforts seem to layer on top of one another, instead of new things replacing the old. The expansion of what folks consider to be important in this arena has meant that it's now a big table with lots of people around it all representing different areas of the university.

Another big conclusion is that the rate of change seems to be accelerating. We noticed that because we saw a large number of practices enter the success space roughly around the same time. It was around 2010 and we linked it to the Great Recession, combined with some demographic trends. Colleges and universities are continuously welcoming more students from lower-income backgrounds, partially because there are more college-age, lower-income students than in the past—but also because of gains made in access and opportunity, which has been a positive development. Additionally, following the recession, more students need more financial support. In many cases, these students also require more support not just to go to school, but in order to succeed, and that has resulted in a proliferation of student success efforts post-recession as schools seek to serve previously underrepresented student groups.

The third conclusion was that if you take all these things together, student success is now a story of what some have called the “return on education”—not quite return on investment, but it’s similar. It’s a focus on making sure the student gets out of their postsecondary experience what they want. And for many students that’s making sure they graduate in the shortest amount of time at the least cost possible with a positive post-graduate outcome. Orienting your student success strategy around ROE is far different—and broader—than orienting around a single metric like first-year retention.

Evo: What were some of the motivations that have now pushed colleges and universities to focus more on success when 10 years ago that focus was opening the door on access?

EV: Schools are motivated to focus on success largely because of the demands of constituencies like parents and students, as well as funding bodies. But when you actually speak to institutions about it, they’ll tell you student success is not an economic imperative but a moral one. This is the right thing to do—we have to deliver on the promise to these students. And if you think about it, that’s a huge credit to higher education that that is the number one pressure here.

Evo: What are the lessons that higher education can learn from e-commerce leaders like Amazon and what are the limits to those lessons?

EV: There’s actually a tremendous amount higher education can learn from e-commerce leaders. Universities are making investments in self-service with resources like student portals where students can see all the information they need like forms, guidance, campus news etc. The issue is that a lot of online student portals are not mobile-friendly. And they’re often organized around the administration lines of the university, not really around student problems. There’s a user experience issue here. Amazon and many other successful online businesses are designed around us, the users. So much of our experience online now is just-in-time stuff. Think about how often your phone pings you to do stuff—attend a meeting, follow up with someone, something is trending on Twitter, whatever it may be, but it’s in-the-moment stuff. We can apply that principle to student success and say, hey, it’s the beginning of the term, you should go see your advisor, here’s all the information you need to set up that appointment. You’re heading towards the middle of term, time to start thinking about declaring a major and selecting some courses, here’s what you need to do about that. Maybe one day we’ll be able to say to a student, hi we know you’re in English 101 and we know you have a paper coming due soon, a week in advance we’re going to send you a nudge to go to the writing center to help yourself out and get some advice on that. We’re nowhere near that level of personalization yet or at the level of timeliness but it doesn’t feel that far off and it would mimic an experience that we have in every other aspect of our lives.

Evo: What role do you expect Big Data to play in student support initiatives over the next decade?

EV: There are a lot of different ways, but the most obvious one is that if you're going to personalize and customize something you have to know something about the student. There's a lot of data that we already have on students about their academic performance, their background, things along these lines. We also have data collected from the student specifically for purposes of pushing guidance, such as an engagement survey designed to learn a little bit about them or their preferences and needs. When we get further into the future you could see customization at the level of the course using data from the LMS. We might even see customizations based on student behaviours around campus as captured via swipe care or GPS data. These innovations might seem creepy, or even a bit like Big Brother, but really they just mimic the collection and use of data that is already happening in our lives, mostly to our benefit. Higher education is just now thinking about how to apply these marketing principles to the student success challenge.

Evo: Is there anything you'd like to add about the integration of Big Data and analytics into the delivery of student support and service?

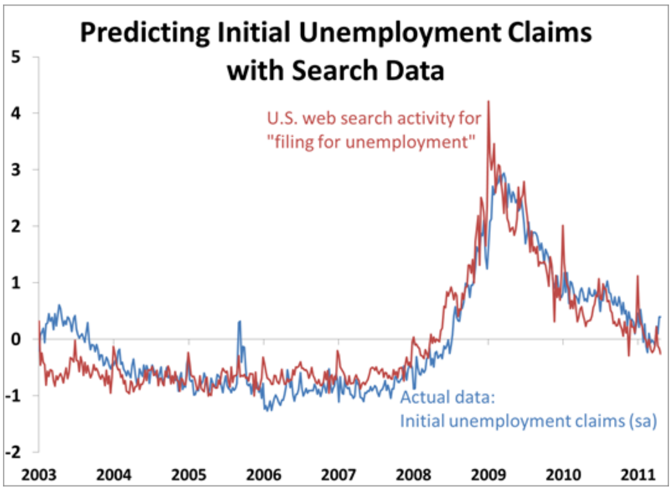
EV: Big Data is not a panacea. It's just another tool in the toolbox. And like any tool, you need to know how to use it correctly. At EAB, we believe strongly in change management as a necessary part of technology implementation. To explain why, let's take the example of one of our predictive models meant to predict a student's likelihood of graduation. If the model shows an advisor that a student has a 20-percent chance of succeeding, there are a lot of different ways to use that information, some good and some bad. We aren't trying to "drown the bunnies," and we definitely don't want to create self-fulfilling prophecies. To avoid this, we need to train advisors, and the institution at large, on what a risk prediction means, how to interpret it, and how to intervene and communicate with at-risk students in a way that promotes their success rather than undermines it. People need to know that a prediction is just a forecast, it is not fate. Higher education is dealing with a lot of questions like these for the first time, which is why the philosophical conversations are just as important as the technological innovations.

This interview has been edited for length and clarity.
source: evolllution.com

History on how to determine success with data.

Six years ago: May 25th 2011 by Justin Wolfers (http://freakonomics.com/2011/05/25/mining-for-correlations-it-works/)
Google's New Correlation Mining Tool: It Works!
You may have heard of Google Trends. It's a cool tool which will show you the ups-and-downs of the public's interest in a particular topic—at least as revealed in how often we search for it. And you may have even heard of the first really important use of this tool: Google Flu Trends, which uses search data to try to predict flu activity. Now Google has released an amazing way to reverse engineer the process: Google Correlate. Just feed in your favourite weekly time series (or cross-state comparisons), and it will tell you which search terms are most closely correlated with your data.
So I tried it out. And it works! Amazingly well.
I fed in the weekly numbers on initial unemployment claims—one of the most important weekly

economic time series we have. The search term that is most closely correlated? Crikey, it's "filing for unemployment." Indeed, the correlation is an astounding 0.91.



Given the latest Google Trends on "filing for unemployment," I'll forecast that initial unemployment claims will tick down in the next couple of weeks.
With an eye to earning a quick trading fortune, I also uploaded data on weekly returns on the S&P 500. But Google failed to find anything significantly correlated. Score one for the random walk hypothesis.

The deduction so far: Is that when technology was quite more 'embryonic' than it is now big players like google then government agencies played part into moving the frontier of what technically is deemed to be possible and now BI technologies are reachable for the back-end user.

Nowadays the threshold is allot lower. One thing seems to be forgotten in most of the articles we find as well in general where BI solutions are implemented into smaller initiatives. A correlation is not a causality yet! Most things don't translate back to the real world and all too often people seem to forget that. Therefore, first a strong correlation should be determined, then its causality, which in turn can be translated back to real-world-activities. Ideal would be a tool that recognized a correlation on its own and also provides visuals.

Here a list of tools that perform correlation mining/visualization that we looked in to:

1. Sisense (https://reviews.financesonline.com/p/sisense/)
2. Oracle data mining
3. RapidMiner (https://reviews.financesonline.com/p/rapidminer/)
4. Microsoft sharepoint (https://reviews.financesonline.com/p/sharepoint/)
5. IBM Cognos (https://reviews.financesonline.com/p/ibm-cognos/)
6. KNIME
7. Dundas BI (https://reviews.financesonline.com/p/dundas-bi/)
8. Board (https://reviews.financesonline.com/p/board/)
9. Orange

10. SAP Business Projects (https://reviews.financesonline.com/p/sap-businessobjects-lumira/)
11. Salesforce analytics cloud (https://reviews.financesonline.com/p/salesforce-analytics-cloud/)
12. DOMO (https://reviews.financesonline.com/p/domo/)
13. SPSS Modeler
14. Qlik sense (https://reviews.financesonline.com/p/qlik-sense/)
15. Birst (https://reviews.financesonline.com/p/birst/)
16. WEKA
17. R-Programming tool

compilation of tools found at: **https://financesonline.com/top-15-data-mining-software-systems/** and https://www.invensis.net/blog/data-processing/12-data-mining-tools-techniques/

During our investigation we discovered a company that guided another campus with realising a tool visualizing their success named:

Pivotal™

Thus we found a company called Pivotal. Among their clients we find big players like BMW. It appears as if they could build the application delivering the functionality meeting the project goals. Hence our project could completely be outsourced to a celebrated and professional company. They already set up similar environments for other Universities like Purdue University, a major research university located in Lafayette, Indiana known for discoveries in science, technology, engineering and more.

Purdue University has become a leader in using data and data science to help students increase student success rates, flag issues, and improve teacher effectiveness. With the help of Pivotal Big Data Suite, data mining techniques, and predictive analytics, the University can give students and teachers an early warning system in situations where students might have challenges.

"One project at a time, let's reimagine your business—pairing your organization's core expertise and values with our modern infrastructure software and development mind-set. That way, we can build great software together and reshape the world." – Pivotal
https://pivotal.io

Educational institutions have a wealth of information, which can be brought together in an institutional data lake to predict and influence student behaviour. In this webinar, one of Pivotal's principal data scientists discusses a recent collaborative project with a top university, in which many data sources were used to build a 360-degree profile of student activity on campus and help predict student success. Learn about the data science pipelines that Pivotal developed and how they are now being used to predict student metrics (such as GPA, course grade and time to graduate), and even as intervention tools to help prevent students from dropping out.
Webinar recording: https://youtu.be/SxXZBmAs1aE
Presentation: https://www.slideshare.net/Pivotal/how-data-science-is-preventing-college-dropouts-and-advancing-student-success

4.2.2 Outsourcing

As mentioned in the article by William Vorheis, Pivotal is a company that has helped the University to develop a similar solution for them. Namely increasing the rates by determining the student success factors, exactly what we also want to do!

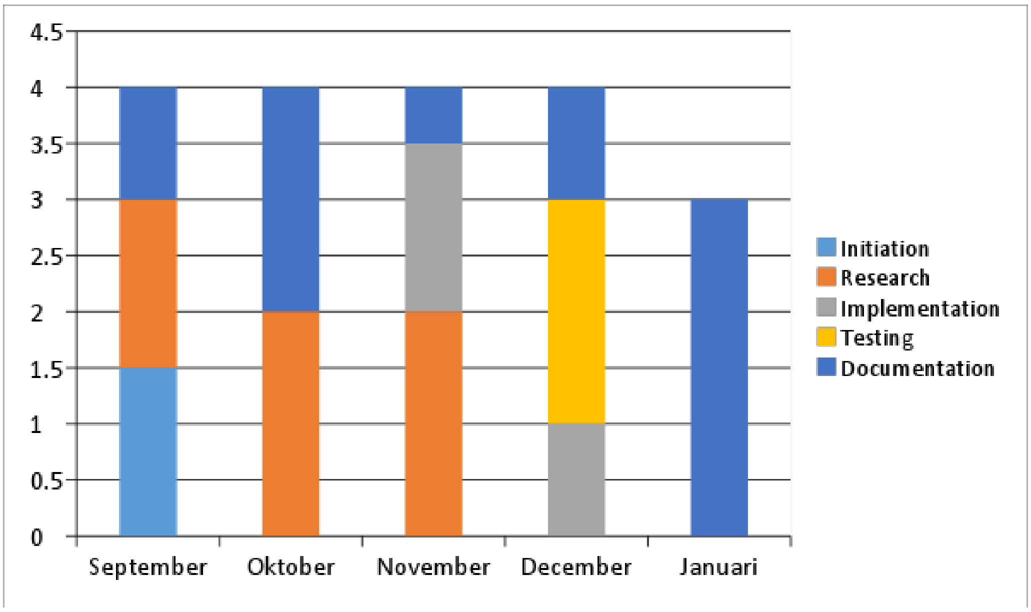
Our company has made contact with Pivotal, where we discussed the possibilities of implementing such a solution at Fontys University of Applied Science.
After a Zoom(like Skype) call with Les Klein on the 12th of October, executive at Pivotal it was clear to us that it's more a company that coaches employees and guides through insight as a service rather than a development company to which we could outsource the whole project. The company has

resources (tools and knowledge), that can help us to prepare different systems for the application, which will determine student success. It's very important for us, that all the data rights are still ours, if we work together with Pivotal. In order to implement such solution, Pivotal has to work closely together with Fontys IT, expecting that within 3-5 weeks there is already a working prototype (MVP to further enhancement).

Before Pivotal can realise this they need to do some research. Based on the results for the other university in the past. It comes down to that underneath the solution they need to set up a data warehouse and above that the analytic models.

4.2.3 Cost Breakdown

Below you'll find a breakdown of the different costs during this project.



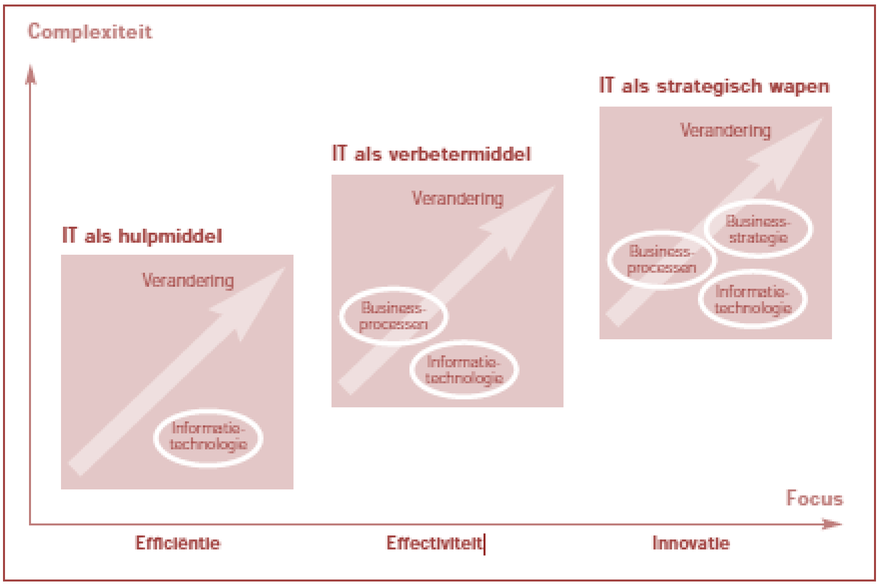
In the chart above you'll see the project costs in week per phase. This is the set amount of time in which the project should be completed.

	per year		
Project costs	2017	2018	2019
Fixed			
Pivotal	€ 40.000	€ 10.000	€ -
Servers	€ 5.400	€ 5.400	€ 5.400
Fontys IT (project member)	€ 28.800	€ 28.800	€ 28.800
Rent buildings	€ -	€ -	€ -
Total costs	**€ 74.200**	**€ 44.200**	**€ 34.200**
Amount of students succesfull passes			
	amount of students per year		
"Profit per student"	5	15	17
€ 8.000	€ 40.000	€ 120.000	€ 136.000

These costs and projected 'revenues' give the following ROI:

ROI	-46%	171%	298%
	Negative	Positive	Positive

4.2.4 Complexity



(Noordam & Zalm van der, 2004)

There are different complexity of IT investments and these are classified in three categories:

1. IT as a tool (efficiency)
2. IT as an improver (effectiveness)
3. IT as a strategic weapon (innovation)

Our project is "Provide students with insight in their own performance/ success to increase their efficiency." This project focuses on a student perspective rather than the business. With that in mind we can place it in category 1 (IT as a tool).

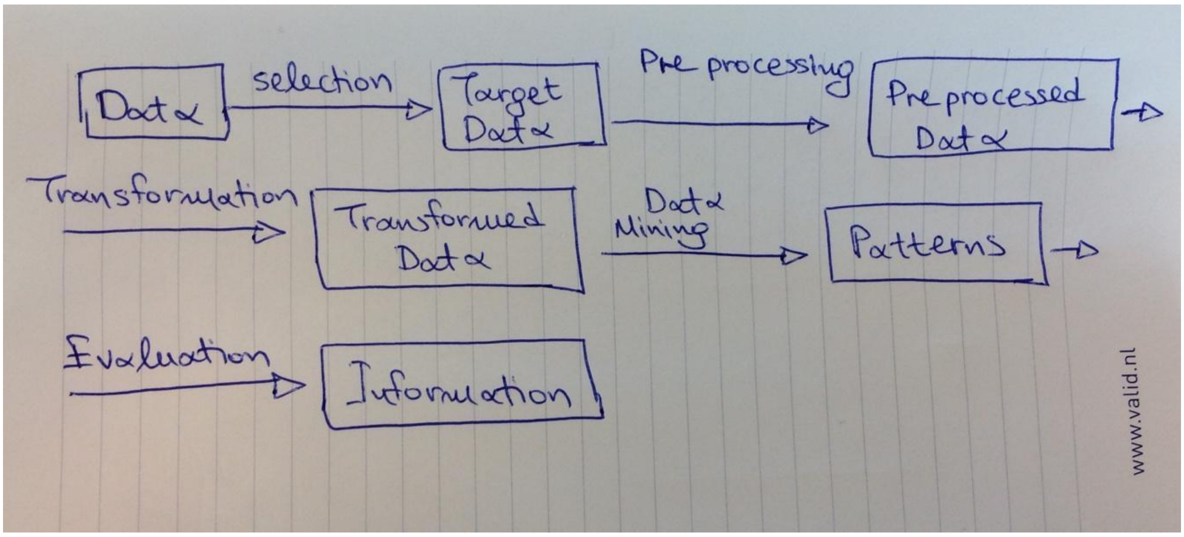
If the tool will not provide an insight to the students, then the business will not see this (positive) in the results. The amount of successful students won't rise while the investment costs are made. All in all, we want more students to succeed each year and to know, If the investment is feasible. As mentioned in the ROI above, we can see that project should generate a positive outcome for University.

Advice: Working together with Pivotal would provide a much faster complete solution than 'reinventing the wheel' yourself. As mentioned in the first paragraph, once the project is finished the entire solution and its data/ information rights are owned by Fontys. Pivotal can make a MVP within weeks and then Fontys can further develop throughout iterations on their own.

Chapter 5: Machine learning

On the 10th of November we participated in an Azure Machine Learning workshop on the vicinity of TU/e. Here we got through the hands on basics as well as technical knowledge about this corner of AI (Artificial Intelligence). The basics, principles of AI and concepts behind Machine Learning.

The following delineation is provided as an overview of and topical guide to machine learning: Machine learning – subfield of computer science (more particularly soft computing) that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed". Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model from an example training set of input observations in order to make data-driven predictions or decisions expressed as outputs, rather than following strictly static program instructions.
-https://en.wikipedia.org/wiki/Outline_of_machine_learning



Before a real data analysis can be done, there are some pre-steps;

- Data preparation;
- Treating missing and repeated values;
- Treating outliers and errors;
- Scaling.

Data preparation, means that we need to prepare the data in order to analyse. Data cleaning first. Because 99.9% of the cases the data are not structured or at the form we need them to perform our analysis. Data munging, as we call it, is the most time consuming part of a data science project. There are specific steps to prepare our data.   In Azure there are modules to treat the missing values. The module is called Clean Missing data. By using this module, you can either:

- Remove rows with missing data;
- Substitute with a specific value;
- Interpolate values (e.g. Linear);
- Forward/Backward fill.
  Then sometimes we need to perform some mathematical operations or just create calculated values in our dataset that don't exist in the source data. This is called Feature Engineering and there is a module in Azure called Apply Math Operation to do so. You can also write a script in SQL, Python or R to perform all these calculations.   Large datasets often include values that are errors or outliers, which will skew the relationships in the model. Finding outliers involves comprehensive exploration and visualization of the data, and you must be careful to ensure that the outliers you identify are genuine outliers that should be treated, and not indicators that there are some important subsets within the data that should be taken into account in the model. The first step to identifying outliers is often to visualize the relationships between important features and labels as a scatterplot, and looking for plot points that fall outside of the apparent pattern in the data,

To clean outliers and errors can be done by: o Censor (delete row); Trim (trim values over a range); o Interpolate (linear); o Substitute.

There is a module in Azure for this step and it is called 'Clip values'. Also can be done by using SQL, R or Python scripts to handle outliers. Then most of the datasets contain multiple numeric variables, which they need to be in similar scale. We always scaling after treating outliers. This technique is called normalize the data. It's really important after scaling not to lose the relative relationship between the numeric features. In Azure the respective module is called Normalize data. After cleaning and preparing the data, we are ready for our analysis and more specifically to create our predictive models by using Machine Learning (ML). There are different types of ML algorithms. The most common are:

o Classification; o Regression; o Clustering.

We use Classification algorithm to predict answers to Yes/No questions. We use Regression to predict real values and Clustering to find patterns of similar objects. The best models are simple models that fit the data well. We don't want to under/over fit the model. And what we need to achieve that? We need a balance between accuracy and simplicity. Both classification and regression are examples of supervised learning, in which a machine learning model is trained using a set of existing, known data values. Clustering on the other hand, is an unsupervised learning technique in which machine learning is used to group (cluster) data entities based on similar variables.

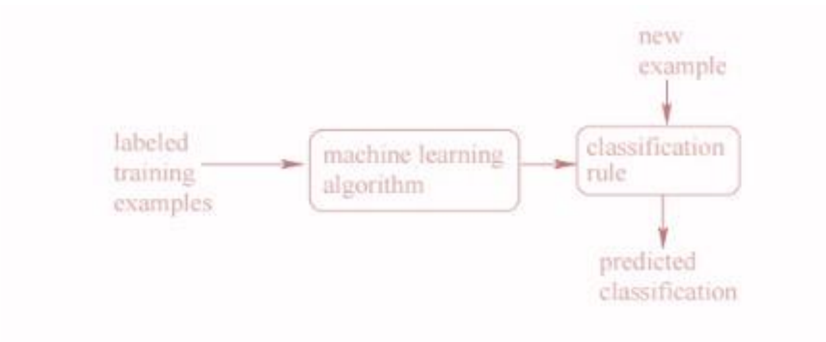
CLASSIFICATION ALGORITHM

Classify examples into given set of categories

Examples of classification problems: o Text categorization;

o Fraud detection; o Market segmentation (e.g. If client will respond to promotion).

How to evaluate the algorithm?



The evaluation metrics available for binary classification models that are used the most are: Accuracy, Precision, Recall, F1 score and AUC. In addition, the module outputs a confusion matrix showing the number of true positives, false negatives, false positives, and true negatives.

Accuracy is simply the proportion of correctly classified instances. It is usually the first metric you look at when evaluating a classifier. However, when the test data is unbalanced or you are more interested in the performance, accuracy doesn't really capture the effectiveness of a classifier. For that reason, it is helpful to compute additional metrics that capture more specific aspects of the evaluation.

There are 4 possible outcomes. The answer to this kind of problems is either Positive or Negative. So, either we predicted that the answer would be Positive and we were correct or wrong (2 possible scenarios) or we predicted that the answer would be Negative and we predicted correctly or wrongly (2 possible scenarios). So, in total we have 4 different outcomes. True Positive(TP), False Positive(FP), True Negative(TN) and False Negative(FN).

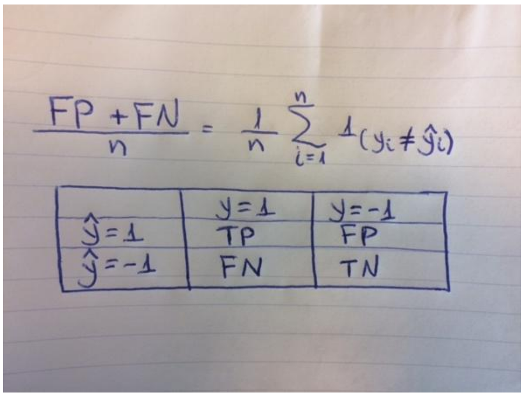
Misclassification error:
Accuracy = 1 – misclassification error

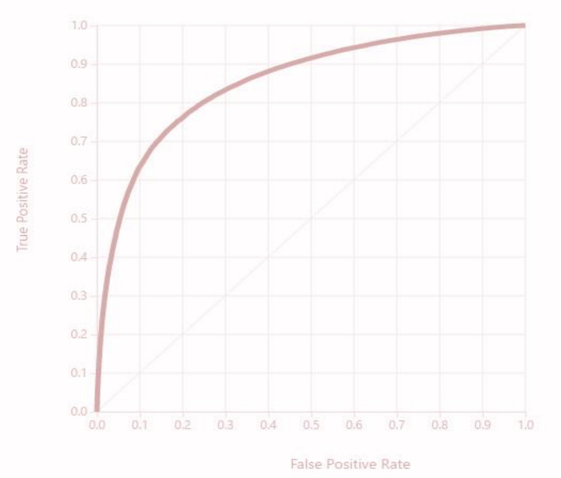
REGRESSION ALGORITHM

It is used to estimate real values based on continuous variables. Examples of regression problems:

o Predict sales; o Predict how long an employee will stay in a company.

Evaluation of algorithm: How we evaluate the closeness of predicted value to the real value? |f(xi) - yi| : error





Mean absolute error (MAE): ∑ |f(xi) - yi| where y: real value and f(x): predicted value
Sum Squared Error (SSE): ∑ (yi - f(xi))2 We have to measure how much of the variation of the yi's is explained by our model f(xi).

EVALUATE CLUSTERING ALGORITHM

It is difficult to evaluate clustering models, because there is no training set with known values that you can use to train the model and compare its results. That's why clustering is often used in the early phases of ML tasks, to explore the data and identify relationships that you might not logically derive by just browsing the data. In this way you can discover correlations that you didn't expect. A common algorithm for clustering is K-mean, in which data values are iteratively divided into K clusters based on distance from a centroid point. The K-means algorithm begins with an initial set of centres and then it iteratively refines the location of those centres.

K-means :

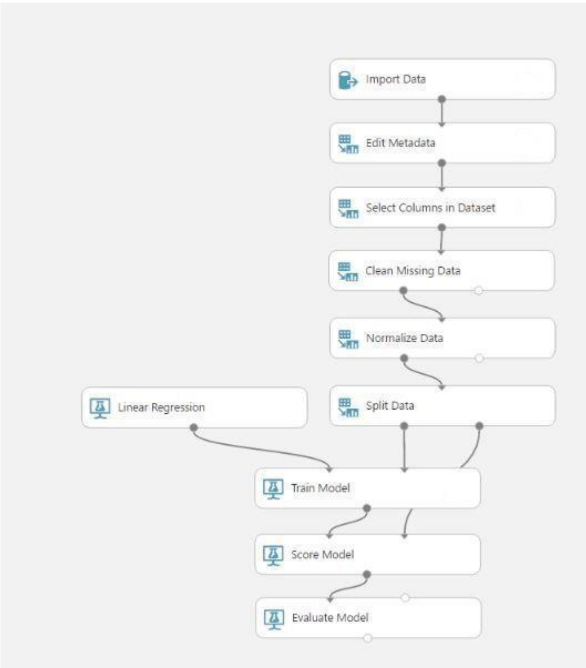
o Input number of clusters and randomly initialize centres; o Assign all points to the closest cluster centre; o Change cluster centres to be in the middle of its points; o Repeat until convergence.

BUILDING PREDICTIVE ML MODELS IN AZURE

(https://studio.azureml.net)

Azure ML modules

- Import data Load data from sources such as the Web, Azure SQL db, Azure Blob storage, csv, etc
- Edit metadata Edits metadata associated with columns in dataset
- Select columns Selects columns to include or exclude from a dataset in an operation
- Clean missing data Specifies how to handle the values missing from a dataset
- Normalize data Rescales numeric data to constrain dataset values to a standard range



• Split data Split the rows of a dataset into two distinct sets

After preparing the data, we are ready to create a ML model. There are 5 steps to do it:

o Split the data into two sets. One will be used to train our model and the other one to score it;

o Find which algorithm we are going to use (Regression, Classification,); o Train our model; o Score our model; o Evaluate our model.

Train model Provide data (60%) to the configured model to learn from patterns and create statistics that can be used for predictions
Score model Create predictions using the trained model
Evaluate model Measure the accuracy of a trained model or compare multiple models

IMPORTANCE OF VARIABLES

How to identify which variables are important for our predictions. Not all the variables in a dataset have the same predictive power. We can use the Permutation feature to

identify those variables. By using this module in Azure, we can see what kind of weight a variable has. This means how much influence a variable has in the prediction. It's easy to use since we just need to feed the module with our trained dataset and a test dataset. After using this module, we can really simplify our model by selecting only the variables which add predictive value.

DEPLOY PREDICTIVE EXPERIMENT AS A WEBSERVICE & USE IT IN EXCEL

https://rquintino.wordpress.com/2014/11/26/azureml-web-service-scoring-with-excel-and- power-query/

1. Click Setup Web Service, and then click Predictive Web Service (Recommended).

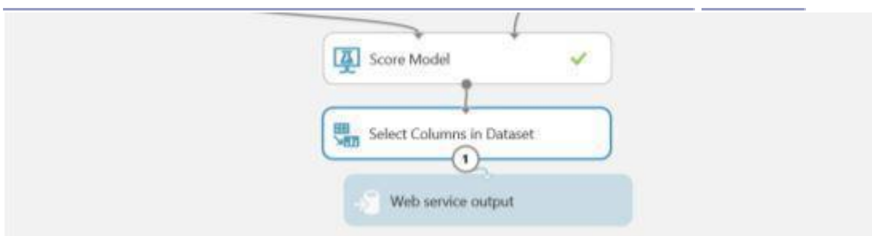
2. Then add a Select Columns in Dataset module, use the column selector to select only the Scored Labels and Scored Probability columns.

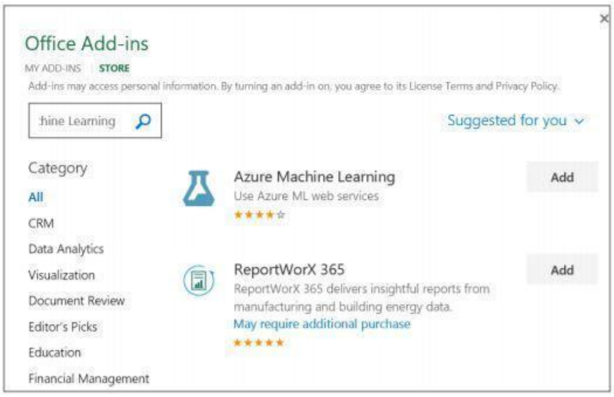
3. Verify that your experiment looks like this:

In order to use the Web Service

In a new browser tab, navigate to https://office.live.com/start/Excel.aspx. Then on the Insert tab, click Office Add-ins. Then select Store, search for Azure Machine Learning and add it as shown here:

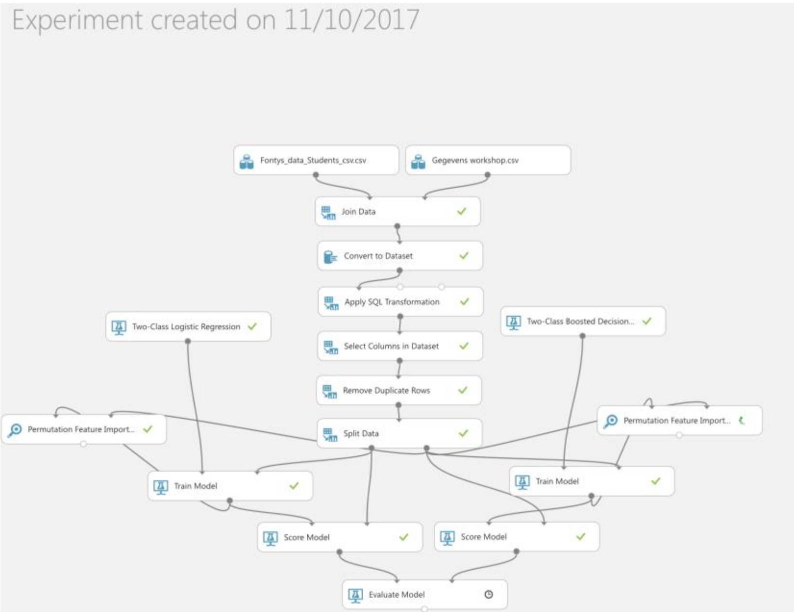
After it is installed, in the Azure Machine Learning pane on the right of the Excel workbook, click Add Web Service. Then you will see the boxes for the URL and API key of the web service.





On the browser tab containing the dashboard page for your Azure ML web service, right-click the Request/Response link and copy the web service URL to the clipboard. Then on the browser tab containing the Excel Online workbook, paste the URL into the URL box. On the browser tab containing the dashboard page for your Azure ML web service, click the Copy button for the API key in order to copy the key to the clipboard. Then return to the browser tab containing the Excel Online workbook and paste it into the API key box. After the web service has been added, its ready for use!

Output experiment ML **during workshop:**

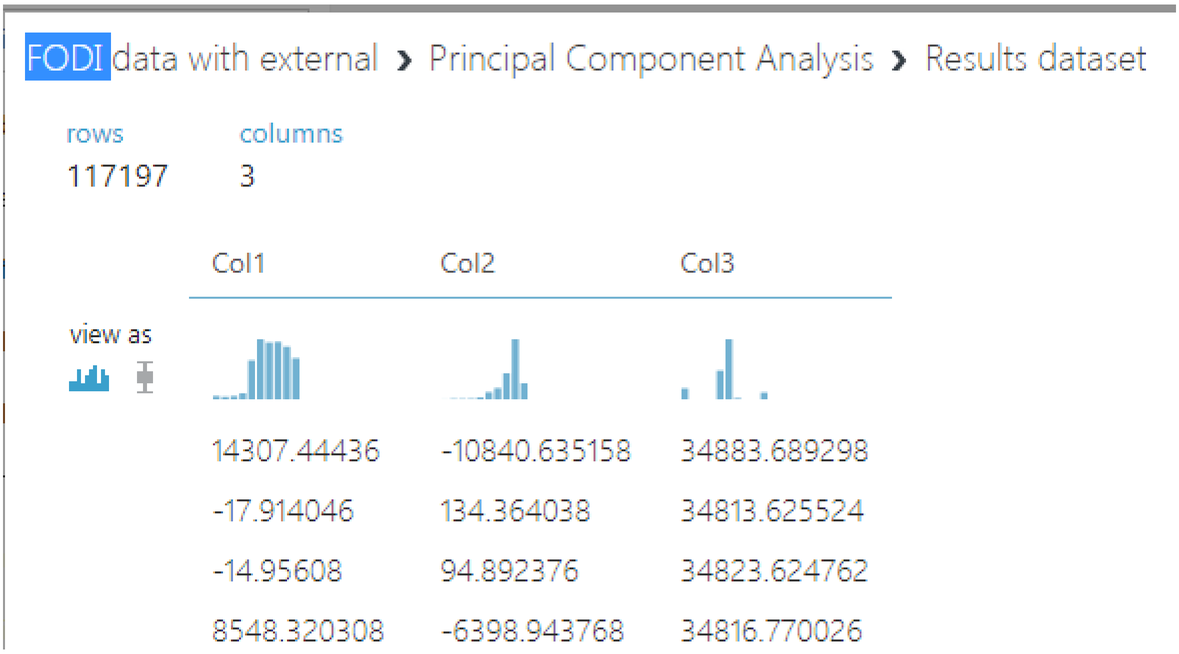


After the workshop

After a workshop with Azure ML we decided to continue our venture with this tool.
In the weeks that followed we used the Azure ML tool to mine for correlations in trained datasets. Quickly we discovered that there are allot of inaccuracies in the datasets. And that any discoveries we would make to visualize on our dashboard would stand far from the real world. This is not a problem when the goal is to gain technical know-how and experience but it's a disappointment for us

since we want to discover new insights into what determines a students' success.

Thanks to the PCA analysis that would identify 3 of our most important columns that we needed to use the Two-Way classification algorithm so that we can predict if the students will get his/ her diploma. The outcome of the analysis was that we received 3 columns with random based values that had nothing to do with our data.



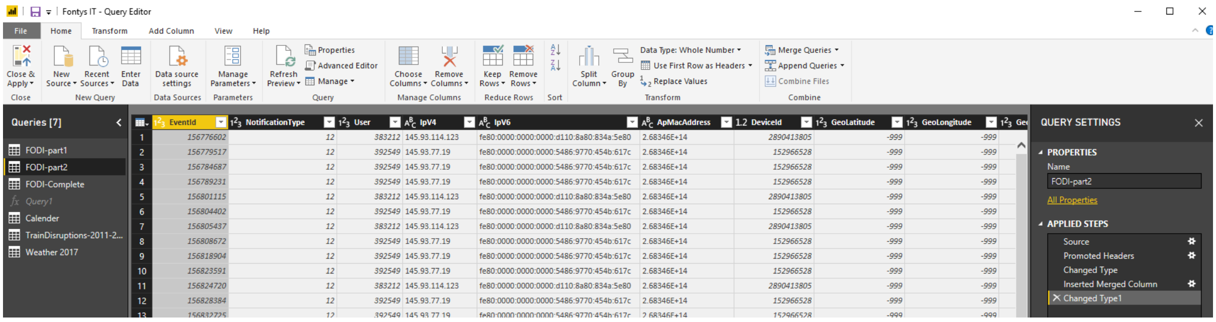
We received this result because the dataset is inconsistent for the methods we want to use. This and any discoveries we would make to visualize on our dashboard would stand far from the actual physical world. This is not a problem when the goal is to gain technical know-how and experience but it's a disappointment for us subsequently because want to discover new insights into what determines a students' success, change how this is measured and challenge what success really is. This delivers us little insight in that manner, but it is still an experience nonetheless.

5.2 Implementation

The implementation of the dashboard faced difficulties; The data quality is simply too poor to get articulate results that are reliable and/or correctly reflect on the real world. As mentioned above. Nonetheless we made a prototype dashboard in PBI showing technical realisation.

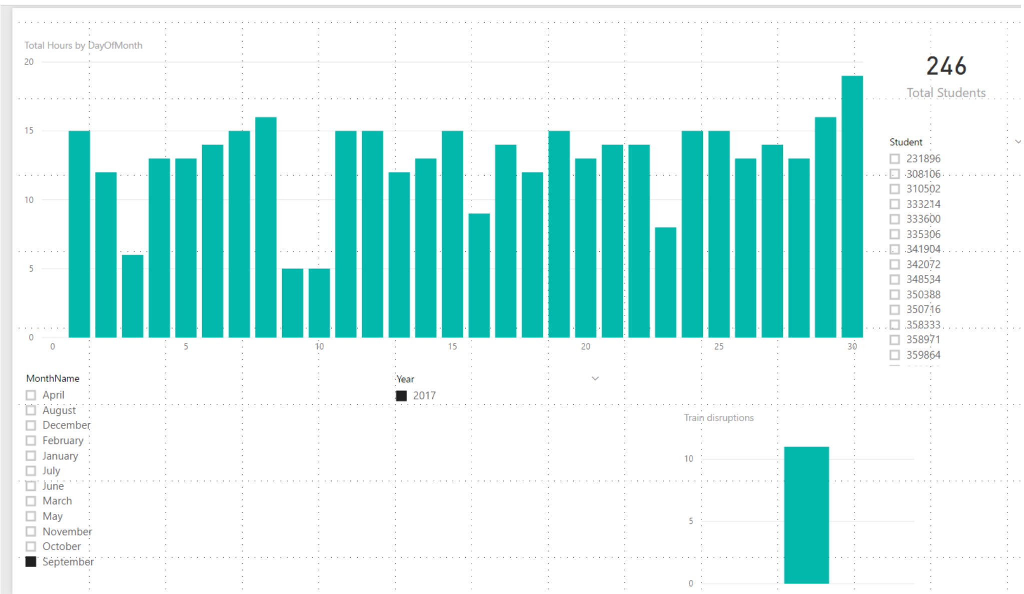
Below you'll find some visuals, with explanation, that show the developing process of the dashboard that has been conveyed as a framework for visualizing the factors we have in mind.

Part 1



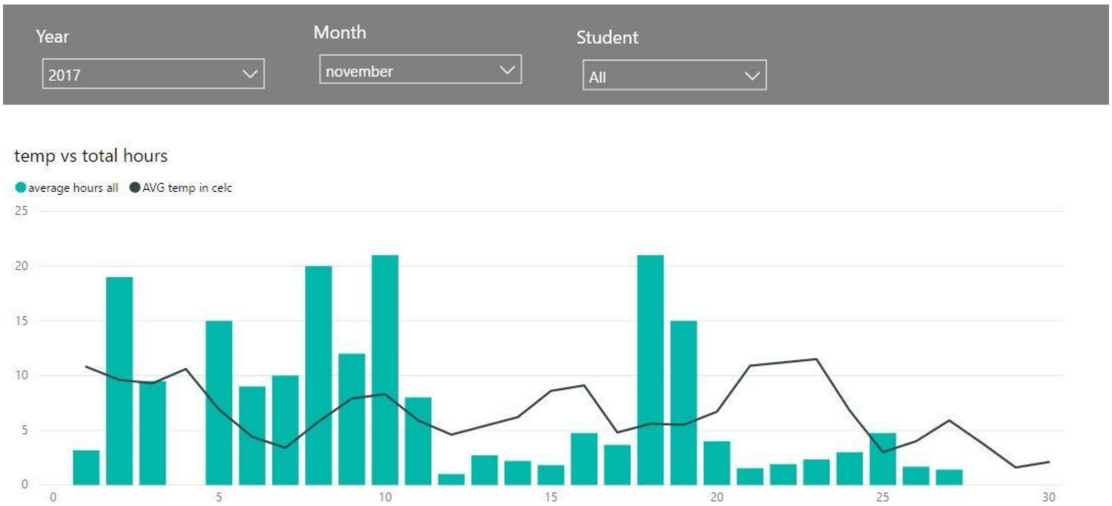
The dataset needed to be cleaned. Most of that has been done in Excel and changed to a .csv format. Still some changes needed to be made. A lot of date & timestamps needed to change from Dutch to US/UK settings, separate columns to create measures, etc.

Part 2



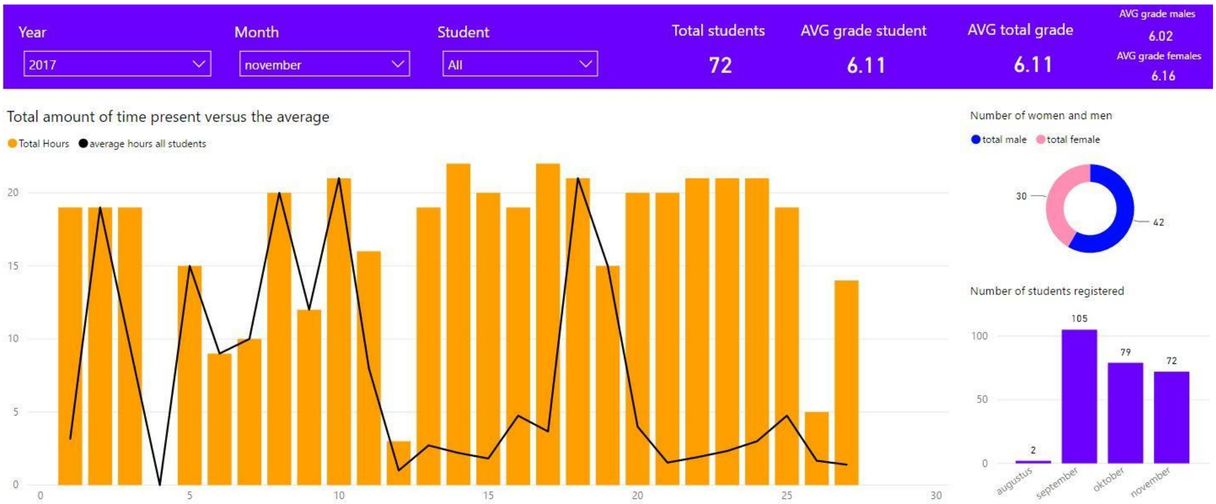
After the first dataset was imported and the measures where created we visualised the attendance of students. Based on this framework we implemented more datasets. This gave us a lot of errors. The solution for that took us 1,5 weeks and that was changing the 'relationship' between tables.

Part 3



When we combined all data with the right 'relations' we added more measures. As you can see on the image above there is no correlation between weather and presence. This contradicts the outcome of previous research done by students of course DS71_A in which they could predict presence based on the weather.

With the student presence and our datasets, it's not possible, in the current framework, to find new correlations to increase the number of students that successfully pass their semester(s). What we could show is the amount of hours compared to other students as well as grades.



We would like to stress once more that unpretentious insights require good quality data which is not collected by the institute within the current infrastructure. Please see recommendation 6.2

Chapter 6: Conclusion and Recommendations

6.1 Conclusions

This report is the result of combined effort of the Obelisk Solutions team and Fontys IT. The following project has been conducted for Fontys University of Applied Science. The time our team spent working on this project was challenging, nevertheless a project it's-self have been a rewarding experience where we learned a lot.

We would like to express our gratitude to parties involved as this assignment is a result of the collaboration of different people involved with diverse backgrounds. Without their participation, it would not have been possible to achieve our final product. First of all, we would like to thank Fontys organization for given opportunity to work on this project; we would like to thank Fontys IT team who supported us through the project and supplied with necessary data and information to complete the assignment and also those who shared a cup of coffee with us while brainstorming.

We found out that the architecture should be optimized to show dominant metrics, but also were able to connect the obtainable sources like FODI, access point data, and other sources like train disruptions to show how factors influence the students' performance. This should be done from a UX perspective where there is much care is taken for how the student gets influenced by being granted these insights. A student might as well get demotivated by seeing his/her performance and then such a solution would backfire. To get the desired outcome we requisite a top down look relying on use-cases to collect what is necessary to show what's required, there where data is now assessed bottom up.

6.2 Recommendation

We have two recommendations. A quick but costly one and a long yet almost priceless one.

I. Contact Pivotal and work with them on redesigning the infrastructure as well as a working prototype for students to get insight in the (for now unknown) success factors. This solution will cost some money but the amount of time the project needs with them if far lesser than letting students reinvent the wheel. It gives you a working prototype to build on and all the data and applications are yours only (the rights lay at Fontys). This is our proposed solution.

II. During the next semester launching a new project with students about redesigning how to collect the data. This should lead to the right data being captured. By then, the semester after that our project could be rerun to simply transact and visualize data that is specifically taken for this purpose. Instead of trying to find if what is anticipated is available in an effective way. This solution will take approximately 20 weeks. A less favourable but almost costless alternative.

Project Evaluation

After a semester of (learning how to) do(ing) research, contacting software vendors, mining correlations, machine learning and dashboard design we learned a lot. Not just about the tools and techniques, but about process management, people skills, resilience and optimism. Like most ventures, we close this project without the expected outcome we had in mind when we started, but instead the real problem unveiled.

It's unmanageable to get reliable and specific metrics from erratic data, architecture should be developed accordingly.

With that we mean that the data is stored for their distinct reasons. To weave them together in a prevailing way they might need to be stored from the perspective of visualizing certain KPI'S, instead of collecting data for their purposefulness, then try to correlate them. Nevertheless, we are proud to deliver a Dashboard showing the presence of the students based on their device data (access point).

OBELISK SOLUTIONS

References / Literature list

Fontys University of Applied Sciences. (2017, December 1). Retrieved from https://en.wikipedia.org/wiki/Fontys_University_of_Applied_Sciences#Eindhoven

Burmeister J , et al. (2014). Correlation of admissions statistics to graduate student success in medical physics. - PubMed - NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24423842

Strickland HP and Cheshire MH. (2017). Exploring the Correlation Between Nontraditional Variables and Student Success: A Longitudinal Study. - PubMed - NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28585983

Campbell AR and Dickson CJ. (1996). Predicting student success: a 10-year review using integrative review and meta-analysis. - PubMed - NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8583033

Student Success. (2017). Retrieved from https://library.educause.edu/topics/information-technology-management-and-leadership/student-success

Leveraging Big Data to Drive Student Success. (2017, March 31). Retrieved from https://evolllution.com/attracting-students/customer_service/leveraging-big-data-to-drive-student-success/
source:

DOI:

Interactions among students' prior learning, aspiration, confidence and university entrance score as determinants of academic success | Student Success. (2017, Sep 21). Retrieved from https://studentsuccessjournal.org/article/view/371

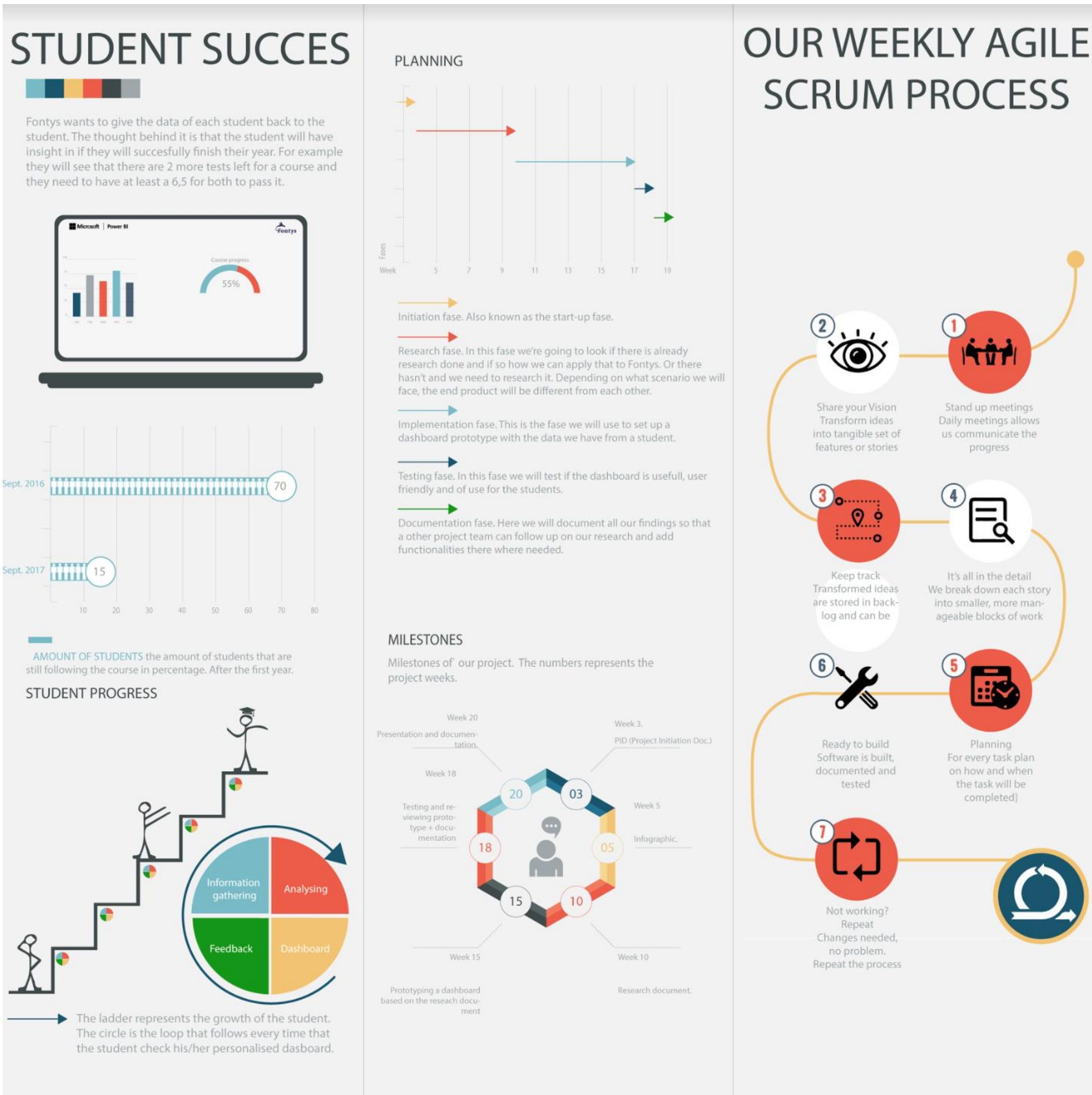
Full research:

View of Interactions among students' prior learning, aspiration, confidence and university entrance score as determinants of academic success. (n.d.). Retrieved from https://studentsuccessjournal.org/article/view/371/411

Correlation of admissions statistics to graduate student success in medical physics.
Burmeister J1, McSpadden E, Rakowski J, Nalichowski A, Yudelev M, Snyder M.
PMID: 24423842
J Nurs Educ. 2017 Jun 1;56(6):351-355.
doi: 10.3928/01484834-20170518-06.

Appendices

Appendix A. Infographic



Appendix B. Project Plan

Management summary

In one sentence:
This project aims to research (and develop a prototype) regarding potentials for students to get more control of their own success within Fontys by facilitating them with insights gained from the data that surrounds them.

This Documents purpose:
This document's objective is to define the project, to be the project management base and to summarize expectations and pronouncements.

The two main reasons for this document are:

- to guarantee a sound project base before inviting the Steering Committee for commitment;
- the Steering Committee and the Project Manager use this document to monitor progress and approve changes and project validity.

Time Window:
We will run this project for the duration of our minor (end December 2017) after which all insights and content will be available for further continuance of our research / development.

Outline

We live in a world where allot of data is collected. Administration is digitalized, school grades are collected in database management systems rather than within offices with dozens of drawers stacked to the ceiling. The mobile devices evolved to high tech computers that people carry around 24/7. They send and receive copious amounts of data. We live in the era of Big Data!

Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating and information privacy. - Wikipedia

Within Fontys allot of data is collected, yet the student is not yet benefitting from the opportunities that are present. We believe that it is possible to use all this information to see correlations and patterns, how variable X Influences Y by combining data sources and analysing these new datasets to support the question how to determine the success of a student at Fontys. How can this be linked back to the student? Can new discoveries be made?

Project definition

Company background

Fontys is a Dutch university of applied sciences with over 45.000 students in several campuses located in the southern Netherlands. The three largest Fontys campuses are located in the cities of Eindhoven, Tilburg and Venlo. Fontys wants to highlight that it is a source of knowledge for students. Fontys offers 200 bachelor's and master's study programmes in the fields of economics, technology, health care, social work, sports and teacher training. A selection of these programmes is offered in German and English.

Fontys International Campus Eindhoven is located in the South-East of the Netherlands, the so-called "Brain port region" which is an international centre of science and technology. This is the place where high-tech companies design and improve state of the art industrial and consumer products for customers all over the world.

Motive

Currently Fontys students merely have access to their grades, schedules and courses. There are several mobile applications out there for students to see their schedule, grades or both depending on the application used. Up till the present no statistical analysis is used with the student in mind as end-users. We're going to research the potentials to change that. And to help students directing their own success within the Fontys university and therefore be less likely to drop out.

Objectives

- Business knows what student needs to successfully complete the studies (success factors)
- Student is a director of his own success (insight into factors that (co-)determine success)
- A prototype in with the technical apprehension of aforementioned is showcased

Research questions

Main question:

- Which internal and external factors can co-determine the successful completion of the study and how can this information be linked back to the student to gain more control of their own success?

Other research questions:

- What is FODI and how can we make use of it?
- What student needs to successfully complete his/ her studies?
  - Internal/ External
- What are the correlations between data and student success?
- What external data is relevant for success?
- What are the relevant privacy regulations related to the student data?
- What are the minimal requirements regarding security when handling this data?

Research methods

- DOT Framework
- LSM Validation
- SBP (Scenario analysis)

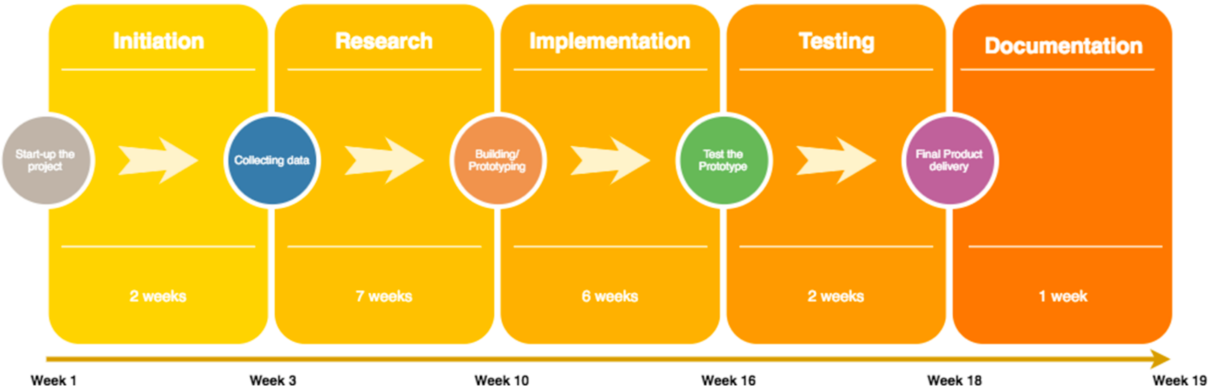
Scope

<u>Depends on SBP followed!</u>

- PID outlining projects prospects
- Extended Research document (including conclusions, recommendations, appendices)
- Dashboard prototype showcasing initial setup of initiative

Out-of-scope

- Redesigning architecture and data flow so that the required data is of high quality
- Prototyping the final application (not the dashboard)



- Development & enrolment of this new application

Project planning

Phasing roadmap

This is the general project phasing in the ideal scenario. Ignoring other scenario's and gives an overview of the ideal circumstances:

Phase 1: Initiation

Activity: Start-up the project

Duration: 2 weeks

Tasks:

- Kick- off meeting
- Declaring scope of the project
- Roadmapping

- Know- how

Phase 2: Research

Activity: Data collection

Duration: 7 weeks

Tasks:

- Reviewing available data

Phase 3: Implementation

Activity: Prototyping

Duration: 6 weeks

Tasks:

- Designing dashboard
- Creating a dashboard

Phase 4: Testing (optional)

Activity: Testing the dashboard

Duration: 2 weeks

Tasks:

- Create a test plan
- Discuss test plan with a client
- Execute plan
- Document all succeeded also as failed tests and a list of improvements

Phase 5: Documentation

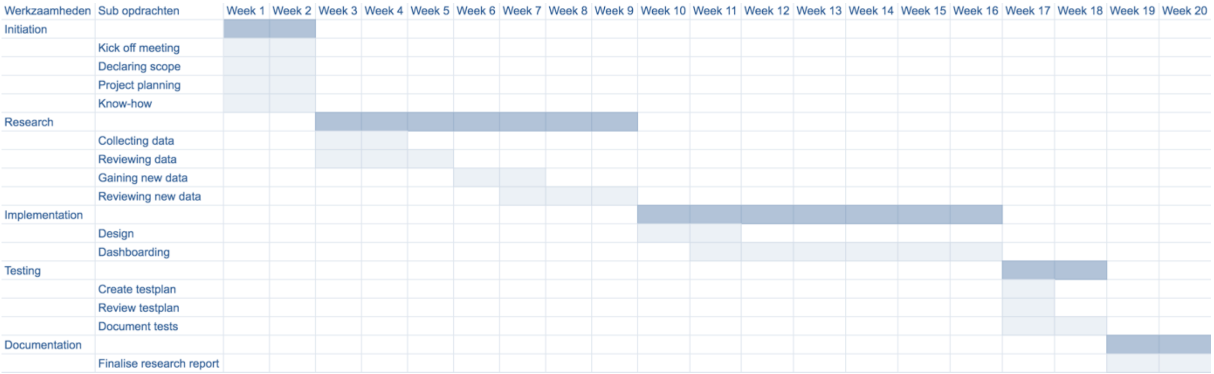
Activity: Final product delivery

Duration: 1 week

Tasks:

- Finalise Research Report (conclusion, recommendation, executive summary)

Global planning



Scenarios

As there is no way to forecast the outcome of this project we brainstormed about different futures, variations in outcome and different results. Above in the phasing we rely on the most desired setting in which all goes as planned. We highlight the three most feasible alternative outcomes on which we will anticipate in our project.

Scenario #1:
Research has already been conducted by a third party. (only development)

- ➔ Looking at core functionalities. Can this be applied for Fontys.
  - What are the corresponding privacy rules?
  - Define the (core) information needs.
  - Build prototype dashboard with the given information from the research questions.

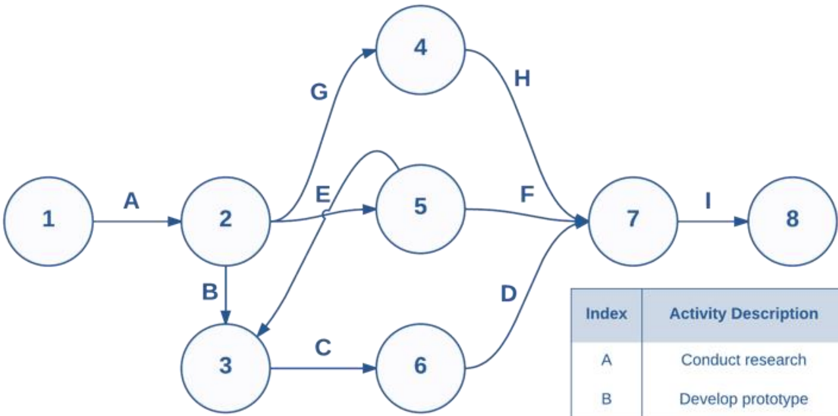
Scenario #2:
Research has not been conducted yet. (research + development)

- ➔ Find the needs of a select group of Fontys students.
  - Need to set up statistical research questions (Closed questions).
  - Distribute the research questions.
  - Review the answers.
  - Define the (core) information needs.

Scenario #3:
The prototyping isn't feasible for the project (only research)

- ➔ The time limit for the project is too short
  - Need to adjust scope of the project
  - Review with the customer
  - Set new scope for the project

➔ The needed data isn't present
  - Review other possibilities with the customer
  - Search other data sources



Index	Activity Description	Required Predecessor	Duration aproximation (weeks)
A	Conduct research	(none)	+/- 5
B	Develop prototype	A	2
C	Refining prototype	B	5
D	Testing beta product	C	3
E	Continue research	A	5
F	Deliver fundamental research	E	1
G	Product exists already	A	-
H	Adapt for Fontys	G	5
I	Deliver results	D/H/F	1
-	Margin time	(none)	+/- 5