Investigating student flow bottlenecks during lunch hour at the University of Twente by means of timetabling data

KWAN SUPPAIBOONSUK, University of Twente, the Netherlands
BART THOMSON, University of Twente, the Netherlands

This paper aims to investigate the efficiency of manual timetabling at the University of Twente in managing room occupancy and population flows on campus. With methods of data preparation and visualization, given files containing scheduled activities over the course of 2013 to 2017 are used to create visualizations for useful insights. The methods and results are reported in this paper. With proper timetabling, there is potential for decreasing traffic jams in certain areas of campus during time periods such as lunch time. Further investigation in the possibility of using automatized timetabling system could help in improving this efficiency, although other performance indicators set by the University may be compromised.

Additional Key Words and Phrases: Timetabling, Student flows, Data preparation, Data visualization

1 INTRODUCTION

The planning and organization at universities has become an enlarged complexity due to the ever-growing student numbers. The development of course timetables has become rather complex. The assigning of students and lecturers to lectures which then must be assigned to rooms while not overlapping with other timeslots of other lectures. Furthermore, requirements of a various group of stakeholders should be satisfied. Key performance indicators (KPI's) often include lecturers' preferences, free hours, non-overlap of courses, student preferences, efficient use of rooms to provide efficient use of the buildings with minimal 'traffic jams'. Important KPIfis at the University of Twente (UTwente) include 4 to 6 student contact hours on any day, a compact time table, no classes on friday evenings, as well as a minimal occupancy of 70% of university rooms. The latter two KPI's will be checked in this paper.

Daskalaki, Birbas and Housos (2004) [1] came up with the University Course Timetabling Problems (UCTP), which includes satisfaction of KPIs proposed by the institution. Student flows in university buildings can be largely controlled by efficient timetabling. Based on the study of Vermuyten et al. (2016) [2], who looked further into a student flow problem at the business faculty of the university of Brussels, this study will monitor the efficiency trend from academic year 2013 to 2015 at UTwente, the Netherlands. Despite the complexity of UCTPs, UTwente still plans her timetables manually, which requires a lot of creativity and time of the planners. Besides trying to fulfill the needs of all stakeholders as well as possible, it is nearly impossible to incorporate the resulting student flows and to plan this efficiently at the same time.

Class transitions possibly cause a lot of unwanted stalling of student and/or staff flows in certain buildings. This paper will look into the spreading of certain class types over time as well as looking into finishing hours of classes in certain buildings. The ultimate goal of a research like this would be monitoring student flows on a heatmap, and thus getting an insight into which places are the bottlenecks. However, time did not allow the researchers to get into advanced data analysis which is necessary to achieve this. Therefore a simplified approach will be explored. More specifically, the researchers will look into the effect of finishing class hours, and their result on traffic jams at lunch. Another flow problem at UTwente is the crowdedness at the O&O-plein, there this will also be explored based on starting hours of classes in 'Hal B' and 'de Waaier'. This paper will gain an insight into the development of the efficiency of the planning of the UTwente, based on these results an advice will be given to the board.

2 METHODOLOGY

2.1 Overview of the provided data

The following CSV files contain an activity overview of the UT, or all the registered or booked classes throughout the years:

- $Activiteitenoverzicht\_2013-2014\_v2$
- $Activiteitenoverzicht\_2014-2015\_v2$
- $Activiteitenoverzicht\_2015-2016\_v2$
- $Activiteitenoverzicht\_2016-2017\_v2$

They were used as they contain multiple datasets that are consistent throughout four year, allowing for trend analysis. The fields that are given by these data are: Name of activity, Description of Activity, Hostkey, Hostkey_2, Activity Type, Date, Time from, Time til, Size, and Location. See table 2 in the appendix for a description of each of the fields. The descriptions for these fields can be found in the table. Out of these fields, the Hostkey fields were not taken into account for data processing, as distinction between Master- and Bachelor-level course can be identified in an easier manner through the 'Activity Name' field.

An important point to keep in mind is that data from 2016-2017 is not yet complete, as the school year has not yet been completed. Most of the incomplete data is mostly for the fourth quartile of the school year and some for the third.

2.2 Data processing & preparation

For improved analysis, some calculations are to be made from the given data. Also, new fields are created from the given fields, modifying and/or classifying the data to allow for simpler analysis. These new calculated fields are: RowID, Timestamp from, Timestamp til, Minutes occupied, Class type, Study year, Faculty, Building, and Academic Year. Descriptions of these fields can be found in table 1.

Processing in Pentaho Kettle

To process the data to get the desired fields as mentioned, the Pentaho Data Integration platform Kettle is used. The following steps were taken for each of the data file to acquire the desired data:

(1) Input the CSV data file and use 'Select data' with the value type of each field properly edited such that it does not display 'null', especially the date (yyy/MM/dd) and time (HH:mm:ss) fields. The 'Size' field should be set to Integer type. The 'Hostkey' fields are not selected as they will not be used.
(2) Use 'Filter rows' to filter out null dates, as it means that the activity was never properly booked; thus, it never occurred officially.
(3) Sort rows according to activity name, then use 'Unique rows' to remove any duplicate rows within the data.
(4) Use 'Add sequence' to generate 'RowID' field to count the row numbers in order to keep track of the data. This is a safety protocol to make sure that in the end, after multiple categorizing and calculating processes, the generated data will be reliable.
(5) With the rows 'tagged' with the 'RowID', use different streams to generate the following fields: 'Timestamp from', 'Timestamp til', 'Minutes occupied', 'Class type', 'Study year', 'Faculty', and 'Building'.
    (a) The first three fields can be calculated with a 'Calculator' step.
    (b) The 'Class type' field can be created by using the 'Value Mapper' step to categorize the 'Activity type' field accordingly. Value mapper is used, as opposed to 'Case/Switch' or 'Replace in String', because the 'Activity type' string values are exact and do not require extraction (search) from another string.
    (c) The 'Study year' field categorizes the courses into the study year level in which it is intended for, for example, Bachelor year 1 (B1), Bachelor year 2 (B2), etc. This information is found in the 'Activity Name' field and is specified with either letter 'B' for bachelor-level course or 'M' for master-level

Table 1. Field names of the analyzed data

RowID	This field is used to 'tag' the data, so that it can be eventually matched correctly after multiple processes of filtration, sorting, and calculation.
Timestamp from	Combined date and time from.
Timestamp til	Combined date and time til.
Minutes occupied	This calculated field gives the amount of time a room was occupied for or how long the class took.
Class type	This field sorted the 'Activity Type' and specifies the class into the following categories: College, Practical, Test, Self-study, Other. See table 5 in appendix for categorization conditions.
Study Year	From the 'Name of Activity' field, the target study year for the course is categorized into B1 (Bachelor year 1), B2 (Bachelor year 2), B3 (Bachelor year 3), and others.
Faculty	The study programs that are found in the 'Name of Activity' are categorized into their designated faculties at the UT. See table 3 in the appendix for categorization conditions.
Building	This field sorted the 'Specified Location' of the classrooms into buildings. Eg. NH 114 $\Rightarrow$ NH (Noordhorst building). Overview of all the buildings found in the data for each year dataset can be found in table 5 in the appendix.
Academic Year	This constant-value field tags the data in each file with the specified academic year, to allow for future merging of all the processed data file for each year.

course. Sometimes it is neither 'B' or 'M', but specified as 'MOD##' for module followed by the number. Modules 1 to 4 is considered to be B1, 5 to 8 is B2, and 9 to 12 is B3.
In order to extract and categorize the string values from the 'Activity Name' field, 'Strings Cut' and 'Case/Switch' steps are used. The 'Strings Cut' step is used to shorten the strings in the 'Activity Name' field, such that when the 'Case/Switch' step looks through the strings for comparisons, there is a lower chance of misreading. Case/Switch is used, and not 'Value Mapper' nor 'Replace in String', because it is necessary to read through the 'Activity Name' strings for the specific values. To reduce the chance for error 'Strings Cut' is used to cut the 'Activity Name' string and take only the first parts with the necessary information. Following the separation of rows by category, a new constant field of 'Study year' was added, then all the rows are merged again to contain all of the different categories in 'Study year'.

(d) The same methods are used to obtain the 'Faculty' field.

(e) As for the 'Building' field, after splitting the 'Location' field and filtering out unnecessary parts of the strings, such as the room numbers or the 'ZZ' indications, building information can be then considered processed. Although values for Therm and Sports Centrum could still be refined as the it sometimes contain extra '/T' following the desired values; thus, 'Replace in strings' to edit those values to remove it.

(6) Following those steps, it is necessary to combine all the newly calculated fields together. This is done by using 'stream lookup' to obtain the fields, with the lookup field being 'RowID', as it is a value that has not been rendered.

(7) Eventually, the processed data is output to a text file. This choice was made as opposed to outputting to a localhost database, as it is simplest for sharing between two laptops. Text files are also quickly and easily read by Tableau, the tool used for data visualization.

The Kettle flow in which all the steps above occur can be seen in figure 5 in the appendix. This is completed for all four years of data, generating four text files. For each year, the kettle flow took around three seconds to run. The text files can then be combined into one by using the command prompt or a text file merge tool.

Preparation in QGis
QGis 2.18.5 'Las Palmas' was used to create a custom map for visualizing class data seen in figure 4. Adding a new shapefile layer over a Google maps background allowed adding of objects, serving as latitude and longitude data for Tableau. By means of 'inner joining' the spatial file to the .csv produced by Kettle automated recognition of buildings was made possible.

2.3 Data Visualization

Data visualization was done in Tableau 10.2, dashboard functionality which has been created is visualized in figures 1 and 4. These timeseries visualizations have been realized by plotting the average size of the classes averaged per day (for the academic years 2014-2015 and 2015-2016). For the 'Waaier' lunch, solely the 'Waaier' and 'Hal B' have been taken into account, these are the buildings in the nearest vicinity of the actual lunch area in the 'Waaier'. A similar approach has been taken in the Horsttoren (figure 1). The created shapefiles with QGis have been linked to class type in pie diagrams, the sum of the classes is represented as the size of the pie charts. To allow academic year comparison, the pages diagram has been filled with the year, to provide for academic year comparison. In order to achieve these comparisons on a academic year basis, the data has been reorganized to financial years starting in september (basic Tableau functionality). Tableau data and shapefiles can be found in the added .zip file. Based on the visualizations in Tableau an analysis is performed, there will be looked into what the data can mean for the UTwente.

3 RESULTS

The upper row of figure 1 shows the average size of the buildings Hal B, de Waaier and Zilverling which finish at certain hours. The lower row depicts data in a similar fashion as the upper row, however only taking into account the sizes of the Horsttoren. Figure 2 pictures the starting hours of the classes in 'de Waaier', 'Hal B', 'Carre' and 'Zilverling' in the first row, and the finishing hours of these same buildings in the second row. Figure 3 are the results of satisfying the KPI's (a) no evening classes on friday and (top left) (b) minimal room occupancy of 70% (bottom center). The rooms displayed with corresponding occupancy rates are CU B200A, HT 1100 and NH 115. In this figure the comparison of the TOM model between two consecutive years is also displayed. This is a comparison of academic years 2013 and 2014 for all B2 study years of the respective academic year. Figure 4 is a display of the interactive time series dashboard visualizing class distribution over the academic years 2013-2015. On the left side, the class distributions a shown relatively to one another. Also faculty time series have been plotted in this figure.

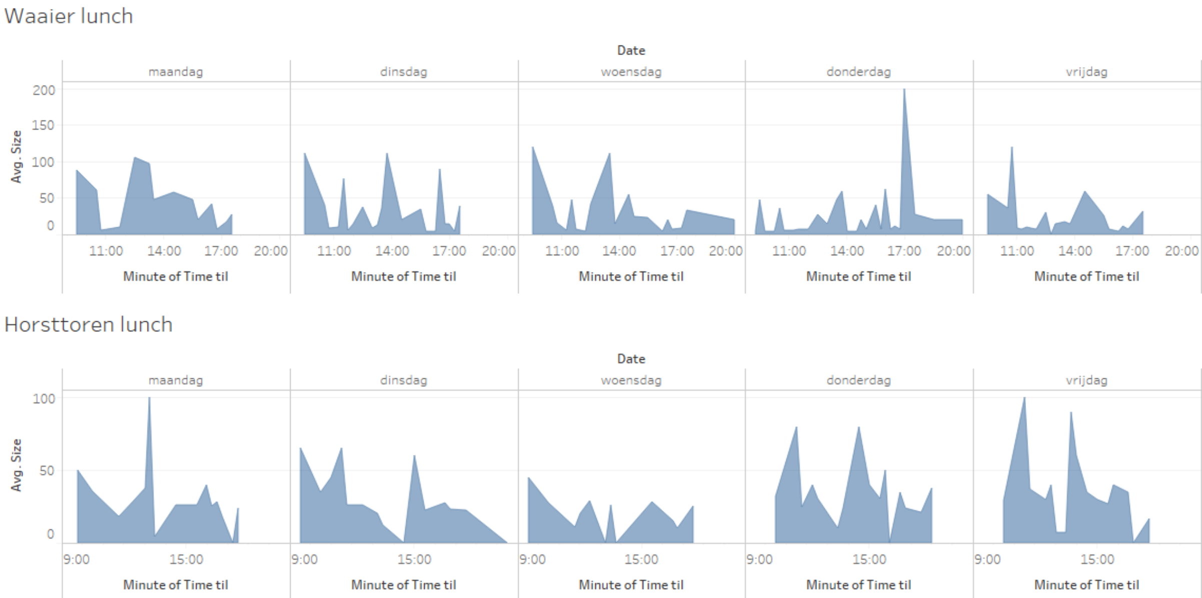

Fig. 1. Lunch dashboard produced in Tableau

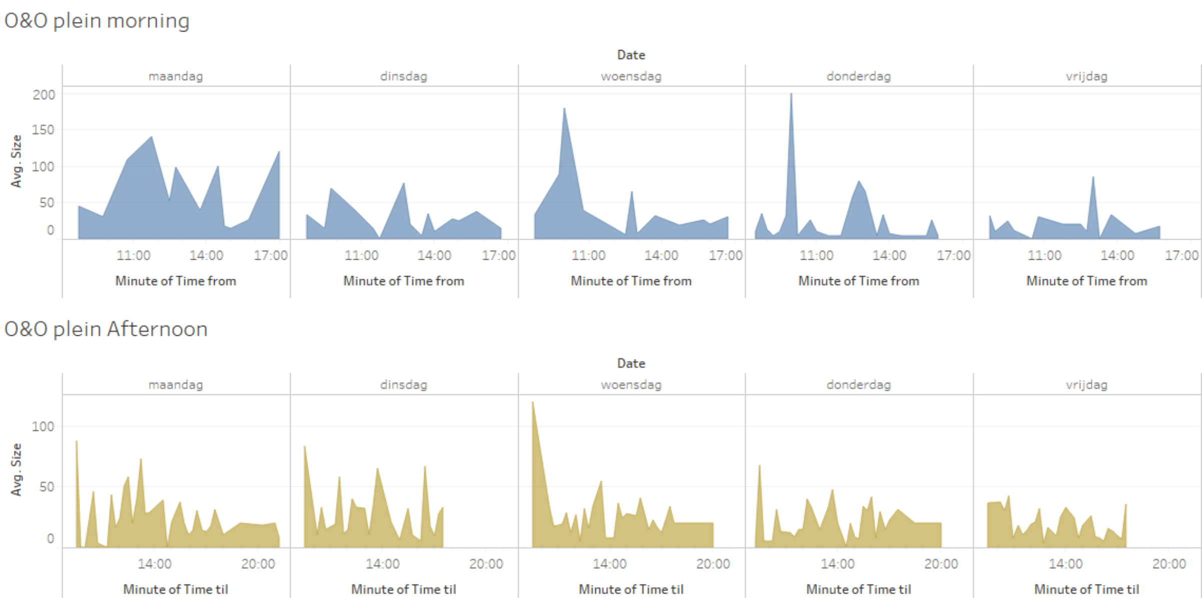

Fig. 2. Exploration of rush hour at the O&O plein

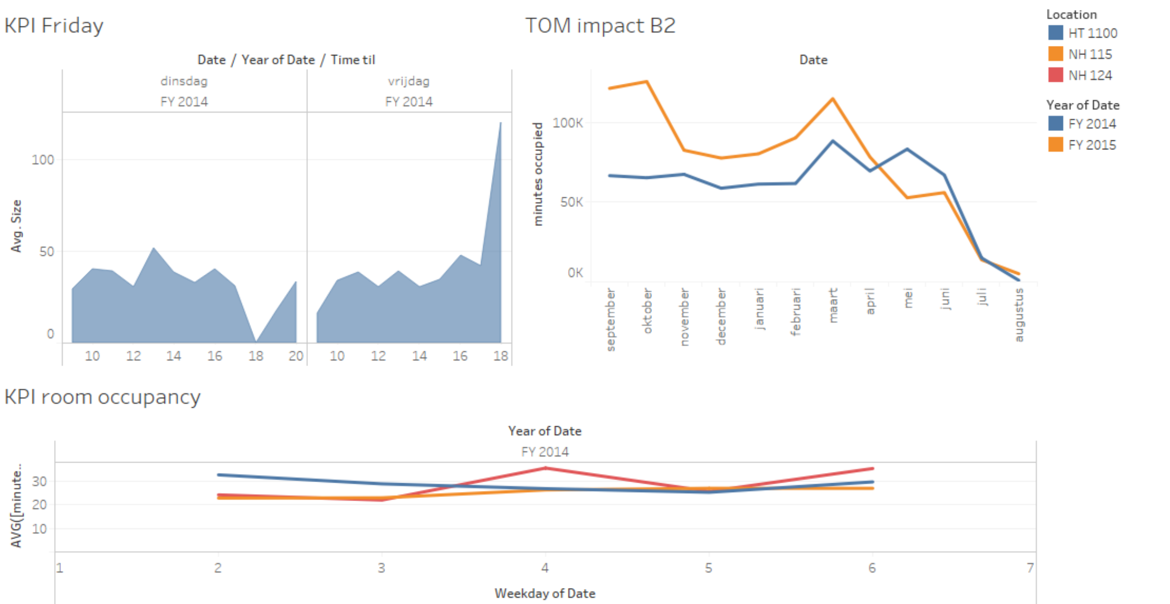

Fig. 3. Looking into certain KPI's at UTwente

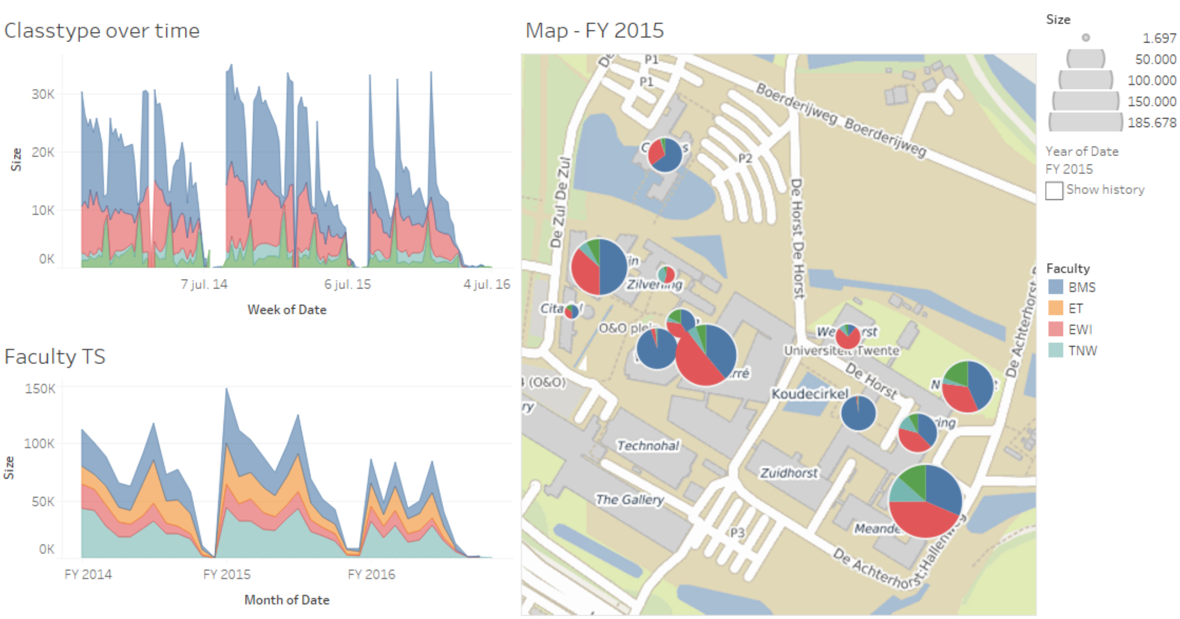

Fig. 4. Distribution map with timeseries of different class types and faculties

4 DISCUSSION

Timetabling is a very complex problem as described by Daskalaki, Birbas and Housos (2004) [1]. Investigating whether certain quality measures (KPI's) are fulfilled is already a very time consuming job, let alone planning of the timetables satisfying these KPI's. From figure 3 it becomes clear that there are no classes on friday after 18.00, thus satisfying this KPI. However, the room occupancy KPI is not satisfied, the average room (figure 3 only shows three rooms) only has a occupancy percentage of about 30%. The effects of the TOM model can also be distinguished, figure 3 displays an increase of total class minutes when comparing bachelor 2 of academic year 2013 (non-TOM) to bachelor 2 of academic year 2014 (TOM). Although the prepared data has been inspected extensively, some courses might not be represented in both academic years.

In data processing, the streams could have been simplified more. Following the results for data processing and preparation, it was realized that the the three calculator steps could have been combined into one and the categorization stream could have branched off after this, reducing the amount. Combining the calculator steps would affect very little in decreasing the run time, as they are simple steps, but decreasing the amount of branched of streams could potentially slightly improve the amount of time that it takes to run. When the latter step is used, there would be one less stream and one less 'stream lookup' step would be needed. Generally, these 'stream lookup' steps take up more time than steps such as the 'calculator'. Module numbers sometimes have 0 in front of them. Making sure to check data sorted into 'others' to make sure everything is sorted properly. Making sure to preview after every step to make sure that each step is working as intended.

Interesting matters for future research are the cause behind the rather low occupancy rates of the rooms at UTwente. It has to be noted that around 30% of the data had to be discarded in this study, due to improper formatting of the data. Reshaping this data to a readable format would have been too time consuming and would not have left space for other analytical steps. However, it is thought that this percentage is not able to explain the low occupancy numbers. Furthermore, spreading of class endings is an interesting matter to invest, smaller lines at 'de Waaier' lunchroom and the 'Horsttoren' elevators can possibly be consequences of this. Vermuyten et al. (2016) [2] have performed an extensive literature study about automated development of course timetables, an approach like this might be worth exploring.

5 CONCLUSION

This study has given a small look into the possible positive effects of timetabling alterations at UTwente, the Netherlands. It has been shown that this is a very complex problem in which data preparation is of utmost importance. Furthermore, it has been highlighted what a powerful tool timetabling is and what could be the effects on decreasing rush hours at certain points in UTwente. The researchers of this paper were amazed by the under occupancy of the rooms which was found when inspecting the KPI's. Further clarification of this finding is recommended, as well as exploration of possible automatic timetabling algorithms.

REFERENCES

[1] S. Daskalaki, T. Birbas, and E. Housos. 2004. An integer programming formulation for a case study in university timetabling. In *European Journal of Operational Research*. https://doi.org/10.1016/S0377-2217(03)00103-6

[2] Hendrik Vermuyten, Stef Lemmens, In??s Marques, and Jeroen Beli??n. 2016. Developing compact course timetables with optimized student flows. *European Journal of Operational Research* (2016). https://doi.org/10.1016/j.ejor.2015.11.028

6 APPENDIX

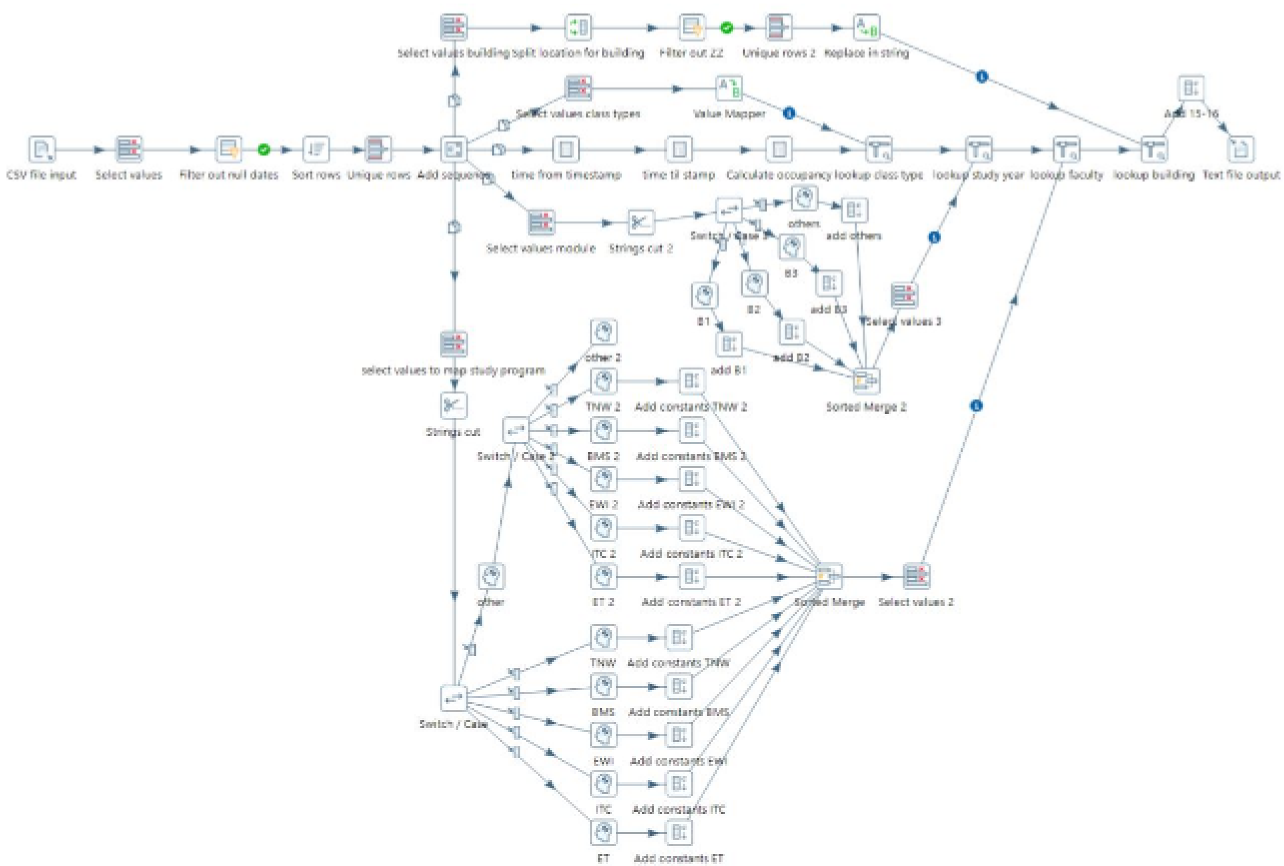

Fig. 5. Kettle flow

Table 2. Description of field in original data files

Field	Description
Name of Activity	This field is used to 'tag' the data, so that it can be eventually matched correctly after multiple processes of filtration, sorting, and calculation.
Description of Activity	Describes the activity. May vary from type of class to course code.
Hostkey	Course code for TOM course
Hostkey_2	Course code for master program course
Activity Type	Multiple tags for different types of activities
Date	Date in which activity takes place
Time from	Starting time of activity
Time til	Expected end time of activity
Size	Class size, as expected by teacher/staff who booked the activity
Location	Specific classroom in which activity is expected to occur

Table 3. Faculty and study programs classification

Faculty	Study programs
TNW	AT, BMT, CHE, GZW, ST, TG, TN, AP, BME, CE, HS, NT, TG/TM, TM, AT/ST, AT/ST/TN, AT/TN, TN/TW, ST/CHE, RT/EM/Stagevoorlichting
BMS	BSK, ESB, EPA, TBK, IBA, TB, CW, PSY, BA, ES, IEM, PA, CS, EST, PSTS, PST, SEC, EEM, MB, MBI, OWK, GW, WIJSB, ELAN, RT/SH/Skills, RT/SH/Bachelor, RT/EM/IBA
ET	CIT, EMS, E, IO, WB, CEM, CEM-CME, CME, IDE, ME, SET, IO/WB/TBK, IO/TBK, WB/IO, HP
EWI	BIT, CRE, EE, TI, TW, AM, CSC, EE, ES, HMI, SC, TEL, TI/BIT/CRE, TI/BIT, ATLAS, HMI, RT/PH/Create, RT/PH/TI
ITC	ITC-*

Table 4. Class types and activity types classification

Class type	Activity type
College	COL, HC, RESP
Practical	PRA, WC, WG, ZMB
Test	BOEK, DTOETS, PRS, TOETS, OVO
Self-study	PJB, PJO, ASM, ZGB
Other	ZZA ORATIE, ZZA OVERIG

Table 5. List of all buildings as found per year of dataset.

Dataset	Buildings
2013-2014 data	CR, CI, CU, EXTERN, HB, HR, HT, NH, OH, RA, SC/T (SC), SP, VR, WA, WH, ZI
2014-2015 data	CI, CR, CU, EXTERN, GY, HB, HR, HT, ITC, NH, OH, RA, SC/T (SC), SP, THERM/T (THERM), VR, WA, WH, ZI
2015-2016 data	CI, CR, CU, EXTERN, GY, HB, HR, HT, ITC, NH, OH, RA, SC/T (SC), SP, THERM/T (THERM), VR, WA, WH, ZI
2016-2017 data-study	CI, CR, CU, EXTERN, GY, HB, HR, HT, ITC, NH, OH, RA, SC/T (SC), SP, THERM/T (THERM), VR, WA, WH, ZI

Received April 2017; accepted April 2017