Analysis of the behavior of students in MOOC's content of Mathematics in primary

Areli Moreno Fonseca, María Castro Valenzuela, Miguel Alfonso Feijóo

Master's Degree in Information Engineering, Universidad de los Andes
Bogotá, Colombia
{am.morenof, my.castro10, ma.feijoo1120}@uniandes.edu.co

***Abstract*— This document describes the development of an interactive visualization tool for the analysis of the course information "Mathematics Content in Primary" which provides statistics about the levels at which students complete the course activities, the number of students participating and the number of attempts they make to complete an evaluation. This information allows the team of the Faculty of Education "Una Empresa Docente" to identify the modules, subjects, and activities that are not completed by the students, the frequency of participation and the difficulty of the evaluations regarding the number of attempts to finalize it successfully. Some results found are that on average students do not complete all the activities of the course, and the average of attempts to complete an evaluation is 2, the same for all modules, which indicates that the difficulty of the tests is very similar, among other possible findings.**

This tool was carried out with the support of the team of professors Pedro Gómez and Carlos Velazco [1] of the "Una Empresa Docente" team. D3.js, principally its version 5.0, is used for the visual implementation and through Postgres, the pre-processing of the data provided by the client, followed by the framework of Tamara Munzner.

I. INTRODUCTION

"Una Empresa Docente" is a unit of the Faculty of Education of Universidad de los Andes, directed by Professor Pedro Gómez (put footnotes) and supported by Professor Carlos Velazco. They focus their activities on research, dissemination, and training in Mathematics Education. The team developed the virtual program 'Mathematics Education for Primary Teachers' that seeks to provide teachers with tools to improve their teaching practice. The program is offered free of charge through the Coursera platform, and its first component is the course 'Content of primary school mathematics'. The team of "Una Empresa Docente" is interested in making, analysis about the use and use of the contents of the course, and the difficulty of the evaluations designed in each module, to optimize its structure and content.

So far, the team of "Una Empresa Docente" only has the databases that can be downloaded through the Coursera platform and contain the demographic information of the participants, the interactions carried out during the course, the participation in the different activities and the type of student (new, old, retired or finished) in which the students are in each one of these. However, these databases are not easy to use, and specialized knowledge is required to carry out processing to obtain statistics of interest.

This document describes the development of an interactive information visualization tool that allows the team of "Una Empresa Docente" to perform analyzes on statistics such as the average percentage in which students complete the course activities, the number of students participating, and the number of attempts they make to complete an

1- "Una empresa docente" is a unit of the Faculty of Education of the Universidad de los Andes. His interests are focused on research and training in Mathematics Education. Its members are professors, masters and doctoral students of the Universidad de los Andes, together with the members of Gemad, the group of graduates of the Master's Degree in Mathematics Education from the Universidad de los Andes. Several professors from Spanish universities participate in its activities. "A teaching company" carries out training, dissemination and research activities.

evaluation successfully. This information allows the team to analyze the characteristics of the activities, lessons or modules with particular behaviors, with the aim of carrying out improvement actions, regarding their structure and content.

The first part of the document presents the context, the background. In the second one is the abstraction of the visualization concerning the Tamara Munzner Framework, in the third, the solution proposed, and the technologies used are described. The fourth presents the results of the evaluation of the tool and finally the conclusions.

II. BACKGROUND

As already mentioned, the course "Content of primary school mathematics" aims to provide teachers with tools to improve their teaching practice regarding the competencies, knowledge, and skills required. The course consists of seven modules that include the following topics: natural numbers and numbering systems, arithmetic of natural numbers, rational numbers, elementary geometry of the plane, and magnitudes and measurements. Each module is made up of up to six topics, in which different activities are found, including presentation videos and an explanation of the topics, learning and evaluation activities, conceptual maps and additional bibliography. The participants have several attempts to complete the evaluations successfully. Notwithstanding the above, they have not a must of filling all the activities following a particular order to pass it.

Currently, the team of "Una Empresa Docente" has the data provided by the Coursera platform. It consists of a series of relational databases that contain information about the contents of the course, the state of completeness of the activities, evaluations, qualifications of the participants, the type of participant in reaction to the state in each activity (New, Retired, Old or Completed) and sociodemographic information. Through SQL queries this kind of information is accessible. Coursera also provides SQL Scripts that allow you to load the exported data in a Postgres database.

The Coursera database is standardized, and although it is possible for users to make their queries by loading the .csv files into Excel, this is a wasteful activity. On the other hand, SQL queries require specialized knowledge, and the installation of the database engine is required locally. The only queries allowed at present are those that Coursera facilitates from the platform.

Coursera provides, some statistics related to the information described, as the total number of students in the active session of the course, the dates in which these students enrolled, all socio-demographic information of the active students, among other essential information. However, this information is not sufficient for the understanding of the use of the contents and the development of the participants.

III. DEVELOPMENT / CURRENT WORK

1. Objectives

- ***GENERAL***: Provide an interactive tool that allows "Una Empresa Docente" to make decisions about improvement actions on the content of the MOOC course 'Content of primary school mathematics'.

- ***SPECIFIC***:
  - Is willed to determine the magnitude of completeness of the activities per module or lessons in parallel, to know how to modify or improve the content of the course. For example, if lessons are identified within a module that are not completed by the students, it is an indicator that the structure or content of the lesson can be adjusted.
  - Identify the number of students per module, lesson or activity in order to have a measure of reception of the activities of this.
  - Identify the difficulty of the evaluations made for each module, regarding the number of attempts students make to complete them. An evaluation in which all end successfully in a single attempt may be an indication that it is too easy to evaluate.

2. Characterization - Abstraction

Next, the general abstraction is presented, according to the framework of Tamara Munzner, for the resulted visualization solution presented on page XX, of this document. The WHAT / WHY / HOW elements were extracted from the proposed objectives, needs and the data provided by the client 'Una Empresa Docente'.

2.1. WHAT

DATASET and ATTRIBUTES:

- ***Dataset type****:*
  RELATIONAL STATIC TABLE containing the following attributes:

- **Module**: Attribute that describes the module of the course in which a student participates. Type: Ordinal, there are 7 modules, which are suggested to be covered in a specific order.
- ***Lesson***: Theme or lesson in which a student participates, a module consists of several lessons. Type: Ordinal. The participant is suggested to cover the lessons in a particular order.
- ***Activity***: Activity or item in which a student participates, each lesson is composed of different activities. Type: Ordinal.
- ***Activity status****:* Type: categorical, indicates whether or not the student completed an activity.
- ***Type of student***: Represents the progress status of each activity for each student. Type: Categorical, includes the categories, New: that was registered in the current session of the course, Old: student that is registered in a previous session to the current one, retired: student that withdrew from the course, finished: Student that finished all the course activities.
- ***Completion date***: Indicates the date on which the student finished an activity. Temporary attribute. Claiming, although the date is used as it is explained in the following item to derive an attribute, it is not used as a variable to analyze changes in the time of any of the attributes, since it is not the client's objective, so this base of Data is not considered temporary.
- ***Completeness lesson***: Type: Quantitative - Sequential. Derived attribute calculated as follows: For each student, the number of activities that I complete from a lesson divided by the total number of activities that make up the lesson per hundred is calculated, then the average value of the students' value is calculated in each lesson.
- ***Module completeness***: Type: Quantitative - Sequential. Derivative attribute calculated as follows: For each student, the number of activities that complete a module divided by the total number of activities that make up the module per hundred is calculated, then the average of the value obtained by the students in each module is calculated.
- ***Number of students per module***: Type: Quantitative - Sequential. Derived attribute, counting the number of students participating in at least one module activity.
- ***Number of students per lesson***: Type: Quantitative - Sequential. Derived attribute, counting the number of students participating in at least one lesson activity.
- ***Number of students per activity***: Derived attribute, student count per activity.
- ***Number of attempts made by a student to answer the questionnaire of a Coursera module***: Type: quantitative, sequential.
- ***Average attempts per module***: Type: Quantitative sequential. Derived attribute that is calculated as the average number of attempts of the students who submitted the questionnaire for each module.

2.2. WHY

- ***MAIN TASK No. 1***

  Summarize the behavior of the students of the MOOC Mathematics course in terms of the completeness of the lessons and modules and according to the student's type [SUMMARIZE FEATURES]

- ***SECONDARY TASKS No. 1:***

  [1] Identify the modules that have a lower percentage of completeness. [BROWSE-IDENTIFY-OUTLIERS-FEATURES]

  [2] Identify the lessons that present a lower percentage of completeness. [BROWSE-IDENTIFY-OUTLIERS-FEATURES]

[3] Identify the topics with atypical behaviors concerning the completeness of their topics. [IDENTIFY-OTLIERS]

- ***MAIN TASK No. 2***

Identify the modules with the least number of attempts and the modules with the most attempts to answer, based on the average number of attempts of the qualifying activities (Coursera questionnaires) by the course participants, per type of student. [LOOKUP-IDENTIFY -FEATURES].

- ***SECONDARY TASKS No. 2***

[1] Compare the modules regarding the number of average response attempts of the course participants. In other words, compare the difficulty of the evaluations between the modules. [COMPARE - FEATURES].

[2] Summarize the number of average attempts of the course evaluations. [SUMARIZE - FEATURES].

[3] Identify in which modules the number of participants begins to decrease, in the case of retired students [IDENTIFY-FEATURES].

- ***MAIN TASK No. 3***

Compare the number of participants of the lessons or activities of a specific module, per type of student [COMPARE -FEATURES].

- ***SECONDARY TASKS No. 3***

[1] Identify the lessons or activities with the least participation in a specific module, [LOOKUP IDENTIFY - OUTLIERS / FEATURES].

[2] Identify the activity with fewer participants within the lesson with fewer participants, from the module with fewer participants. [LOOKUP IDENTIFY - OUTLIERS / FEATURES].

- ***MAIN TASK No. 4***

Build the following derived attributes. [DERIVE FEATURES]

* Completeness lesson
* Module completeness
* Number of students per module
* Number of students per lesson
* Number of students per activity
* Average of attempts per module

2.3. HOW (ENCODE and MANIPULATE)

This section presents the abstraction of the HOW abstraction, describing the Marks and Channels used for each of the visualizations presented in the solution, presented in the next section.

A. Visualization: HEATMAP

The idioms used to represent completeness are heatmaps. Two are constructed to represent completeness at module and lesson level respectively.

- ***Heatmap modules***: Vector of 7 columns by a row, aligned with colored area marks with color opacity (COLOR LUMINANCE), and ordered by the module attribute thematically [SEPARATE-ORDER-ALIGN]. The representation of the modules is done as columns (HORIZONTAL POSITION). The size of the area of the squares does not encodes any attribute. The quantitative completeness attribute is encoded using the opacity of the color.

- ***Heatmap lesson***: Vector of n columns by a row, where n is the number of lessons of a specific module, aligned with colored area marks with color opacity (COLOR LUMINANCE), and ordered by the lesson attribute thematically [SEPARATE-ORDER-ALIGN]. The representation of the lessons is done as columns (HORIZONTAL POSITION). The size of the area of the squares does not encodes any attribute. The quantitative completeness attribute is encoded using the opacity of the color. It is related to the heatmap of the modules so that by selecting the box corresponding to a specific

module, the heatmap of the lessons associated with that module is displayed. (reduces filter items, juxtapose - overview detail)

It is related to the heat map of the modules so that by selecting the box corresponding to a specific module, the heat map of the lessons associated with that module is displayed. (reduces filter items, juxtapose - overview detail)

B. Visualization: Bar Chart - # Average attempts per module

Based in line marks expressing the average value of the number of attempts required by the participants for the questionnaires per module [ENCODE-EXPRESS-VALUE], with alignment in the vertical spatial position, separated by the attribute (KEY) module with the horizontal position space. [ENCODE-SEPARATE-ALIGN]

C. ***Visualization: Bar Chart - Participation by module***

Based in line marks expressing the number of participants in the module [ENCODE-EXPRESS-VALUE] with alignment in the vertical spatial position, separated by the attribute (KEY) module with the horizontal spatial position. [ENCODE-SEPARATE-ALIGN]

D. Visualization: Bar Chart - Participation by lesson

Based in line marks expressing the number of registered participants, [ENCODE-EXPRESS], with alignment in the vertical spatial position, separated by the attribute (KEY) lesson with the horizontal spatial position. [ENCODE-SEPARATE-ALIGN]

It is related to the Bar Chart visualization - Participation by module, so that when a user clicks on a bar in the Chart bar by module, the bar chart associated with the lessons of that module is displayed. (reduces filter items, juxtapose - overview detail)

E. Visualization: Bar Chart - Participation by activity

Based in line marks expressing the number of registered participants, [ENCODE-EXPRESS], with alignment in the vertical spatial position, separated by the attribute (KEY) activity with the horizontal spatial position. [ENCODE-SEPARATE-ALIGN]

It is related to the Bar Chart visualization - Participation by lesson, so that when a user clicks on a bar in the Chart bar by lesson, the bar chart associated with the activities of that lesson is displayed. (reduce filter items, juxtapose – overview detail)

Additionally, all the visualizations can be observed for each type of student (old, new, withdrawn, finished) through a data filter. (REDUCE FILTER items)

3. Solution

According to the objectives and needs of "A teaching company" and the application of Tamara Munzner's framework, and once the additional attributes required for the understanding of the problem have been identified, a solution is proposed where a structure is first constructed of SQL queries for the processing of the data, and additional views that relate the module, lesson and activity fields in a simplified manner and the new attributes are generated. In the second part, based on the output resulting from the queries, the idioms are constructed according to the channels, brands and interactions that were defined in the abstraction stage and presented to the user in an interface where the visualizations are displayed in accordance to your demand.

3.1. Data Processing

Due to the fact that the area does not have the technical support for the generation of the information, we carry out the necessary technical tasks to obtain the information of the course. Coursera provides a copy of the database, for investigation of the course administrators, this copy contains database creation scripts and contains the data that will be populated in the tables, the database must be loaded in an engine of Postgres

database on the local server. In general, the structure of the database is the one presented below:

Illustration 1 Data Base Structure

For the generation of the information necessary for the visualization, the relationships of the tables provided were identified, which are shown in the following illustration:

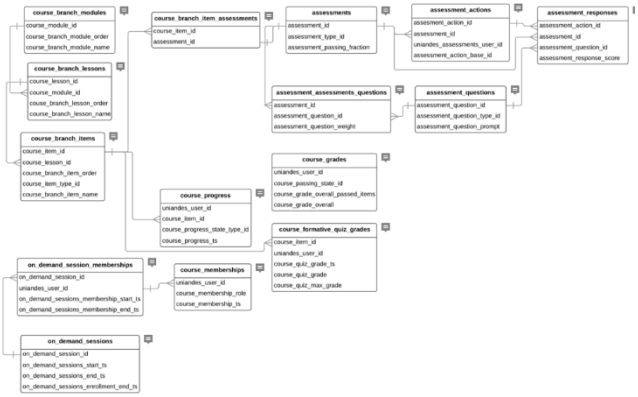

Illustration 2 Relations ideentified in tables of DB

Followed by the above, the *ua_v_item*_curso view was created in the database engine to identify the relationships between the modules, sessions and items. For the identification of active students, the *ua_v_estudiantes_actual* view was created, this to facilitate the creation of the respective queries as input of the presented visualization.

3.2. Interactive Visualization

The development of the visualization was done in D3 with its version 5, using the crossfilter library of JavaScript, the generation of the graphic interface was based on HTML5, using the front-end framework Bootstrap. The visualizations displayed follows the descriptions mentioned in the previous section.

The repository in which the development of visualization is located is https://bit.ly/2EfhEfL and the application is deployed at https://goo.gl/uHXgfH. In the links presented previously, you can have an interaction in the solution planned.

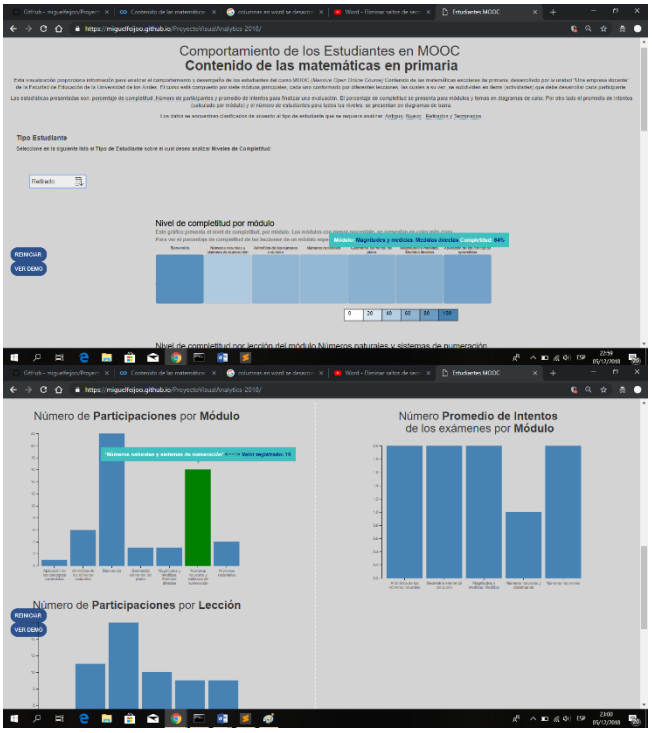

Illustration 3 MOOC Analysis Web Page Preliminary results

4. Evaluation

The evaluation was carried out initially, with the objective of analyzing if the selected idioms responded to the tasks and needs raised by "A teaching company". For this purpose, a

functionality test of a first version of the visualization was carried out in which the idioms were presented in a dashboard type scheme, presented in a very common visual tool which permits the navigation through a set of visualizations: TABLEAU.

The test was conducted at the Faculty of Education of the Universidad de los Andes on November 9, 2018. For the test, some tasks that were identified as tasks of greater difficulty were prioritized. The participants were made up of the two final users of "one teaching company" and four non-users of the subject and the application. In this test the assistants completed the formats of the ANNEX A, and a video was recorded, collecting the observations and suggestions of the two expert participants.
Some important results are listed below:

- It was suggested not to present them as a control board, because too much information is observed.
- Include clearer instructions for handling the application

In the second phase of the application, it was identified that although the content of the application in terms of the figures and variables presented is what is required and meets the requirements of the user, the presentation of the bar diagrams, in order of magnitude and not in thematic order, is what would facilitate the reading of information.

5. *Insights & Conclusions*

For a teaching company, the application is useful to answer your main questions, about the use of the contents, the participation of the students of the course and the difficulty of the evaluations. However, it is necessary to improve the graphic interface, initially providing the user with the option of ordering the data according to the magnitude of the variables.

Following are some of the main results obtained from the data and when exploring the application:

- For none of the types of students (old, new, withdrawn, or completed) is found that the modules are completely finished, since, on average, the maximum completeness is 91% and is observed in the welcome module for students Withdrawn
- The students who complete the course although they are students who pass do not complete the course activities, and it is observed that the lowest percentages of completion are in the last lessons of each module. This is an opportunity for "A teaching company" to review the content of these lessons and the order of activities suggested to the student.
- On average the number of attempts to pass the evaluations is 2, for all modules. This is an indicator that the evaluations have the same level of difficulty, contrary to what was expected, since the experts expected for some evaluations up to 6 attempts.
- The maximum percentage of modules other than Welcome and closing of old students is 58%. Given that for the date of the measurement there was only a week to finish the course, it can be concluded that the old students are not willing to finish the activities completely.
- The number of retired students that reached in the last module "Application of the learned concepts", is only 1, much smaller number of those that started in the welcome module (22).

6. *References*

[1] Contenido de las matemáticas de primaria. Universidad de los Andes Facultad de Educación Enlace del curso: *https://es.coursera.org/learn/contenido-matematica-escolar*

[2] M. Garey and D. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman, San Francisco, 1979

[3] L. Lamport, LaTex 2e: A Document Preparation System, Addison-Wesley, Reading, Mass., 1994

[4] K. Appel and W. Haken, “Every Planar Map is 4-colorable,” Bull. Amer. Math. Soc., vol. 82 (1976), pp. 711-712
[5] D. Knuth, Mathematical Writing, Mathematical Association of America, 1989
[6] J. Zobel, Writing for Computer Science, Springer-Verlag, New York, 1997
[7] Munzner, T., & Maguire, E. (2015). Visualization Analysis & Design. FL: Taylor & Francis Group. Obtenido de https://bit.ly/2FHJJxE