



Summer Fellowship Report

On

YAKSH - Data Analysis

Submitted by

Shivam K. Thakkar

Submitted to

Prof.Kannan M. Moudgalya
Chemical Engineering Department
IIT Bombay

July 6, 2018

Acknowledgment

I, Shivam K. Thakkar, a FOSSEE intern of the YAKSH Project is overwhelmed in all humbleness and gratefulness to acknowledge my deep gratitude to all those who have helped me accomplish the tasks assigned to me.

I am very thankful to my mentors Mr. Arun K P and Mr. Aditya Palaparthi for their valuable suggestions. They were always there to show the right track and with help of their brilliant guidance and encouragement, I was able to complete the assigned tasks.

Last but not the least, I wholeheartedly thank all the other colleagues working in different projects for helping me evolve better with their support and advices.

With Regards,

Shivam K. Thakkar
(CSPIT, CHARUSAT, Changa, Gujarat)

Contents

1	Introduction	3
2	Reports	4
2.1	Report 1 - Course Wise	4
2.2	Report 2 - State Wise / Institute Wise	4
2.3	Report 3 - Error Analysis	5
3	Implementation	6
3.1	Manipulating, Formatting and Displaying data	6
3.2	Cleaning data	7
3.3	Model level changes	7
4	Results	8
4.1	Moderator Dashboard	8
4.2	Student Error Report	11

Chapter 1

Introduction

YAKSH is a open-source MOOC platform for hosting various workshops and courses. It is maintained and managed by FOSSEE Team, IIT Bombay. Currently it has reached over 10,000 registrations (including various workgroups - students, faculty, industry professionals, e.t.c.) and has a lot of data.

A lot of data can provide a lot of meaningful insights which can help both the Students and Moderators. The aim was to perform analysis and cleaning of the present data of YAKSH to provide some useful visualizations which I presented in form of reports (discussed in the next section).

Chapter 2

Reports

Report is nothing but a visualization, provided in forms of different charts, graphs and tables depending on the behaviour and distribution of data.

There are 3 Reports in total -

2.1 Report 1 - Course Wise

This is a moderator side report and presents mainly two views

1. Overall stats of the courses taken by various students which provides numbers like enrolled counts, passed counts, A+ grade counts to get better idea about the performance of the students.
2. Visualization using grouped bar charts with various filters (categories - age, gender, position, grades) to provide a detailed overview and better understand the user distribution. It also provides various trends that will help us to get complete understanding of the user base we have.

2.2 Report 2 - State Wise / Institute Wise

This is a moderator side report and presents state wise distribution of enrolled students and institute wise stats, briefly described below

1. State wise visualization consists a bar chart to show the distribution of enrolled students over the states.

2. For the Institute wise stats we have numbers like enrolled counts and passed counts that present a picture of the overall performance of the students of that particular institute. Also followed by details of the courses taken by those institutes.

2.3 Report 3 - Error Analysis

This is a student side report which gives them a view of errors (Syntax and Logical) occurred in a quiz and also particular question. I have used two visualizations to present it.

1. Pie chart that shows distribution of the errors for over all quiz.
2. Percentage bars for distribution of the errors for a particular question.

Chapter 3

Implementation

The implementation had three major sub tasks

1. Manipulating, Formatting and Displaying data
2. Cleaning data
3. Model level changes

3.1 Manipulating, Formatting and Displaying data

The following packages and technologies were used in the whole process.

- pandas: A Python library to manipulate data using data frames. It was used at backend to manipulate and format data in JSON which is then sent to client.
- plotly.js: A JavaScript package used to render different charts and graphs. It was used to display the received data from client in different reports.
- canvasjs: A similar JavaScript package to as plotly.js used for the same purpose.
- JQuery: JQuery is a JavaScript library built to make most of the ease and simply client-side code.
- AJAX: It is basically a technique used to request a server using XMLHttpRequest object and asynchronously display the received data. It was used to make better UX while switching over different

reports in Moderator Dashboard.

- Django: Python framework for web application development. It follows MVC(Model View Controller) structure for managing the models and controlling the views

3.2 Cleaning data

There are some fields like - Position and Institute name in the Profile Model that need to be cleaned as the data present is misleading and noisy which makes the final reports uninformative and clumsy visualization.

- Cleaning Position Field: The purpose was met by writing a simple Python script that reads the present field value and corrects it using hard-coded rules written after analysing the present entries.

- Cleaning Institute Field: Currently this task is under process and will be implemented after assurance and testing. The idea is to use Levenshtein Distance and find the best match from the AICTE college list and update the present value to the match found.

3.3 Model level changes

We wanted to analyse some data that our system was not recording so certain model level changes were made to add those fields so we can record data. Some changes were made in following models.

- Course: Added a level field that takes value of - Basic, Intermediate and Advanced and group field that takes value of Self and Instructor. Giving a level tag to the course according to the level of difficulty and group tag to know the type of course.

- Profile: Fields age and gender were added to get better overview of the user base and find new trends that can be used to improve our system.

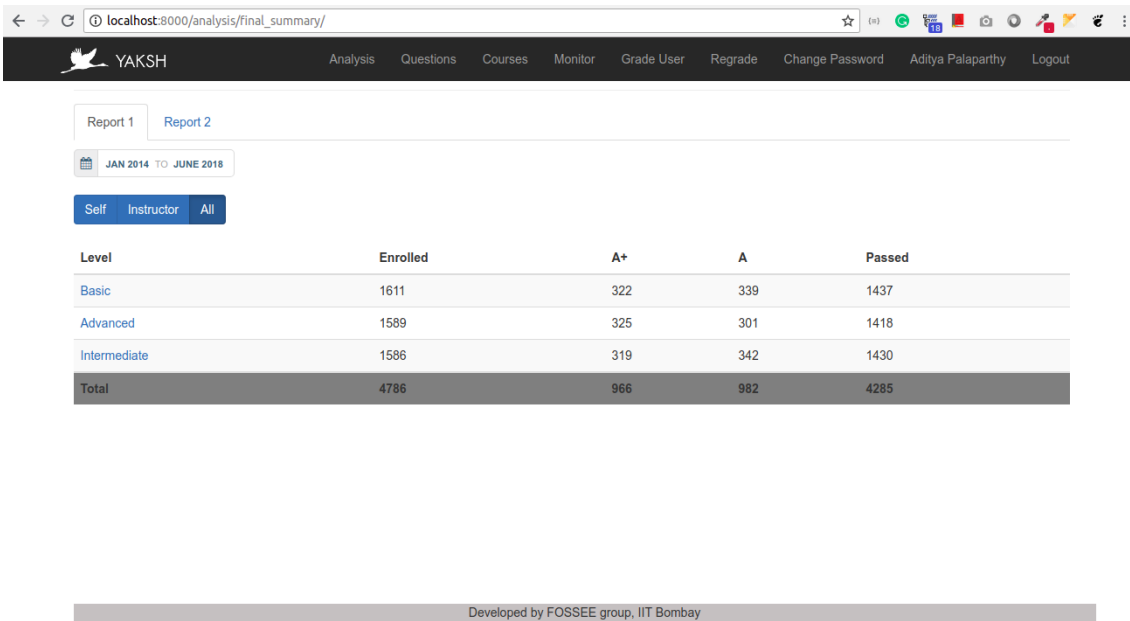
Chapter 4

Results

The two reports are merged as one and shown in Moderator Dashboard whereas the third one can be seen by students for each quiz they have attempted in the course.

4.1 Moderator Dashboard

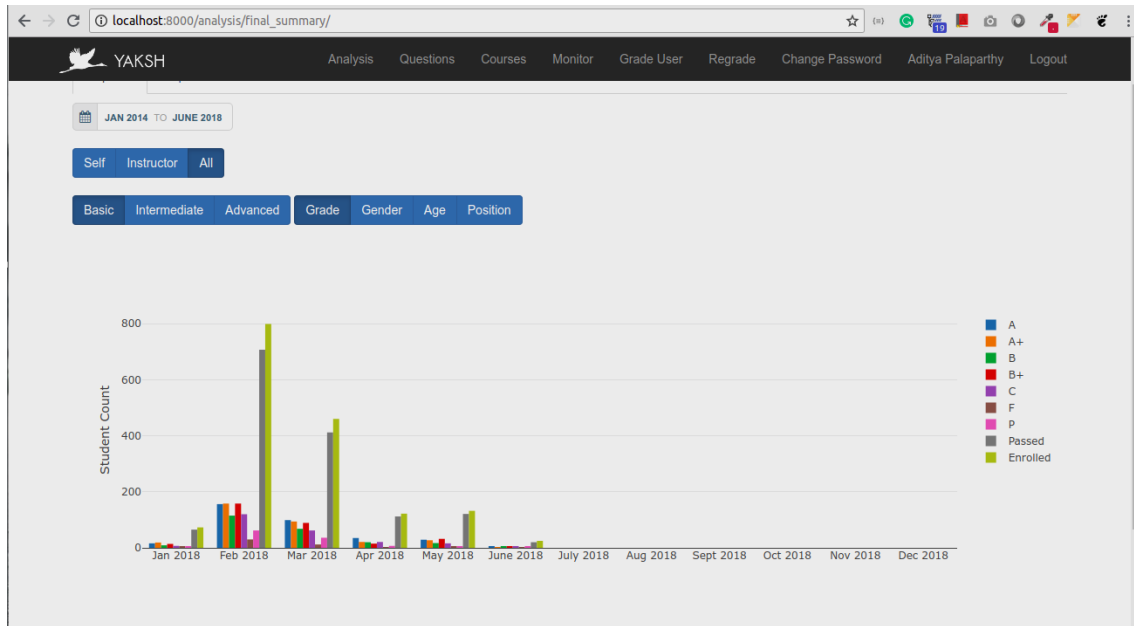
- Report 1



The screenshot displays the YAKSH Moderator Dashboard. At the top, there is a navigation bar with the YAKSH logo and several menu items: Analysis, Questions, Courses, Monitor, Grade User, Regrade, Change Password, Aditya Palaparthy, and Logout. Below the navigation bar, there are two tabs for 'Report 1' and 'Report 2'. A date range selector shows 'JAN 2014 TO JUNE 2018'. Below the date range, there are three buttons: 'Self', 'Instructor', and 'All'. The main content area features a table with the following data:

Level	Enrolled	A+	A	Passed
Basic	1611	322	339	1437
Advanced	1589	325	301	1418
Intermediate	1586	319	342	1430
Total	4786	966	982	4285

At the bottom of the dashboard, there is a footer that reads 'Developed by FOSSEE group, IIT Bombay'.



Here we have certain filters which can be applied to get particular visualization

Group: To jump between different course groups - Self and Instructor or All (both).

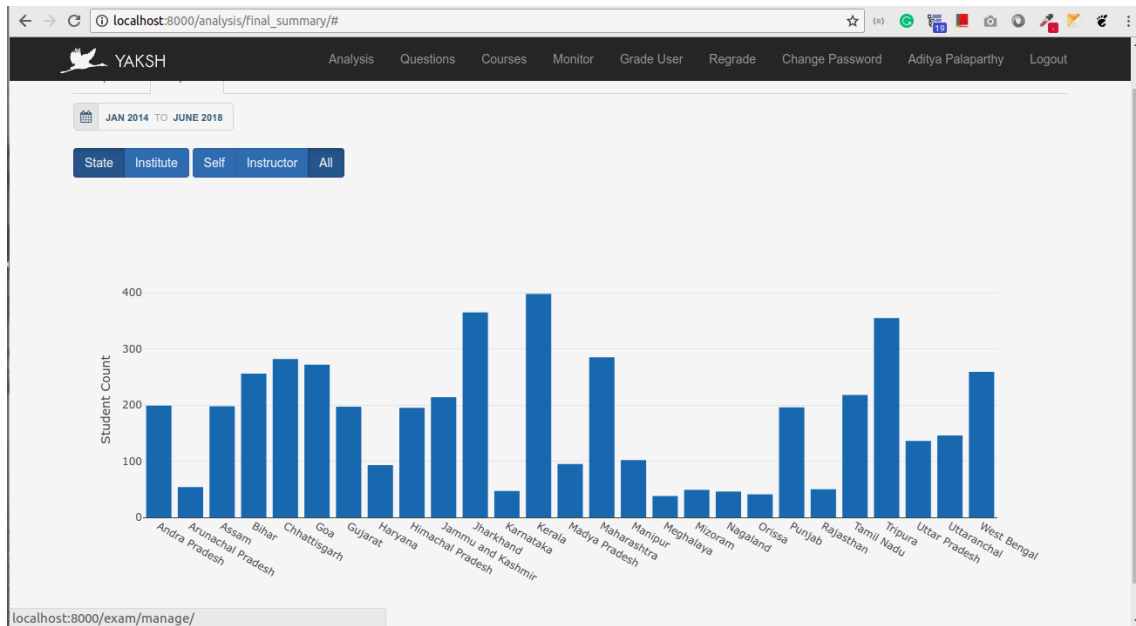
Level: To jump between different course levels - Basic, Intermediate and Advanced.

Category: To jump between different categories (group-bys) - Age, Gender, Position and Grade.

Also there is a month range picker which allows moderator to pick range so that only the data within the range is displayed.

- Report 2

State Wise



Institute Wise

The screenshot shows the YAKSH application interface for an 'Institute Wise' report. The table displays the following data:

Institute	Enrolled	Passed
AA	33	33
AB	44	43
AC	49	46
AD	48	48
AE	58	57
AF	53	52
AG	42	42
AH	59	57
AI	43	42
AJ	45	45
AK	45	41

localhost:8000/analysis/final_summary/#

YAKSH Analysis Questions Courses Monitor Grade User Regrade Change Password Aditya Palaparthy Logout

Report 1 Report 2

JAN 2014 TO JUNE 2018

State Institute Self Instructor All

Course	Enrolled	Passed
BPPy, G Pulla Reddy Engineering College(22-03-2018)	1	1
Basic Programming Using Python Jawaharlal Institute of Technology, Borawan	1	1
Basic Programming Using Python(01 - 30 June)	1	1
Basic Programming Using Python(01 - 30April)	1	1
Basic Programming Using Python(01 - 31 May)	3	3
Basic Programming Using Python(01 - 31March)	11	11
Basics Of Python, Poornima College of Engineering	2	2
Basics of Python, GEHU(29-01-2018)	1	1
Basics of Python, Ramdeo Baba College of Engineering and Management	1	1
Cambridge Institute of Technology(24/04/2018)	1	1
ISCP(fellowship), MBM Engineering College	1	1

We have filters to jump between State Wise and Institute Wise views.

4.2 Student Error Report

