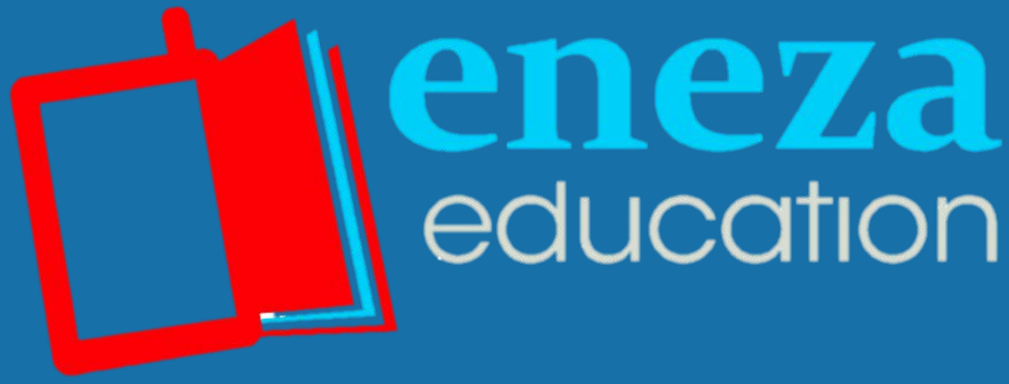


Data Science For Good:
Accelerating Mobile Learning in Kenya

About Us

Sara Hooker -
Data Scientist at Udemy,
Executive Director at Delta Analytics

sara@deltanalytics.org

Steven Troxler -
Data Scientist at Stitch Fix,
Project Lead at Delta Analytics

steven@deltanalytics.org

- Why data for good?
- Who is Delta?
- Technical Deep dive - Eneza Education

Why data for good?

Why data for good?

- Data revolution
- Skills gap is larger than ever

Data Revolution

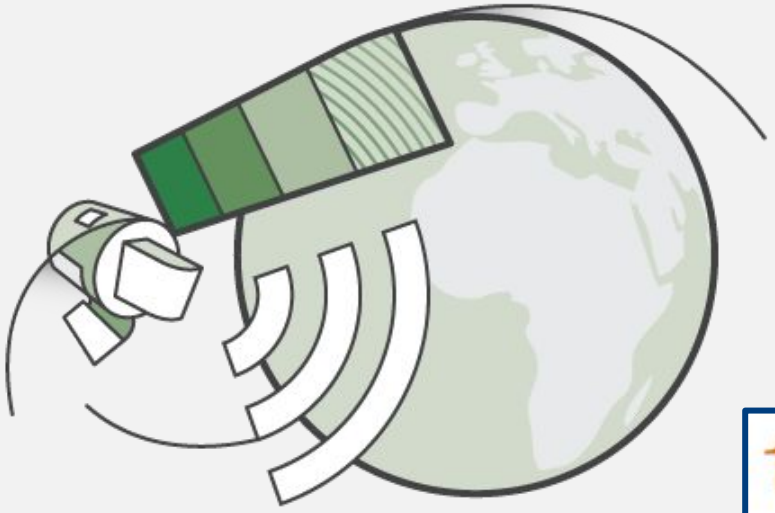
New tools for data collection

Focus on accountability

Data Revolution

Spread of mobile phones across industries

Complex Problems Now Within Reach



SpaceNet

1000 Genomes Project and AWS

Common Crawl on AWS

Public big data

Data Revolution

Non-profits are pushing the hardest for more data:

- focus on accountability
- desire to understand impact

Skills Gap is Larger Than Ever

Filling the data skills gap

Skills Gap is Larger Than Ever

Other organizations working in this space

Eric & Wendy Schmidt
Data Science For Social Good
Summer Fellowship
THE UNIVERSITY OF CHICAGO

BAYES IMPACT

DataKind SF Bay Area

Skills Gap is Larger Than Ever

And most importantly… you.

Who is Delta Analytics?

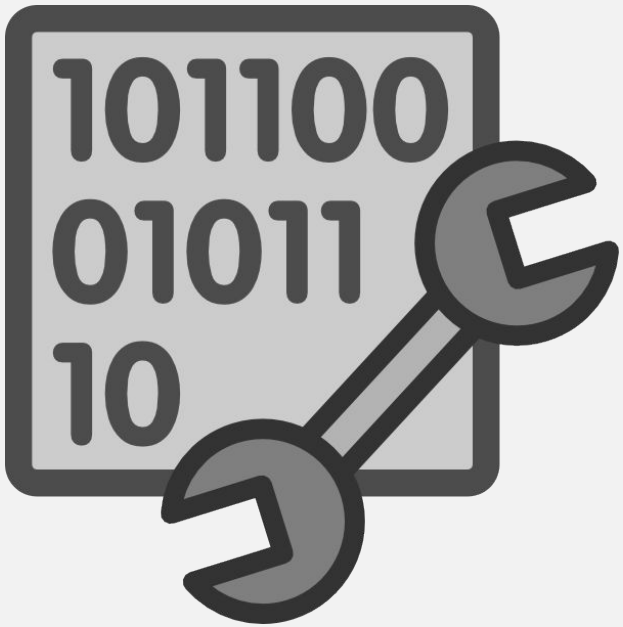
Who is Delta Analytics?

Delta Analytics collaborates with non-profits and other public service organizations to generate **positive social impact** through **key data insights and management services**. Driven by a passion for numbers and dedication to community engagement, we help public service organizations with all their data-driven needs.

Our mission, quite simply, is data for change.

Main constraints for Non-Profits

- **Resources** (staff or specialized skills) for data work
- **Infrastructure requirements**
- **Longitudinal data** collection
- **Analyzing the data itself**



Who is Delta Analytics?

19 projects with non-profits and social impact organizations

60 Fellows volunteering part-time over 3 years

Over **15,000** hours donated

$0.00 charged for services

11 US and 8 International projects (Tanzania, UK, Kenya, and more)

Which sectors do we serve?

Analytics

Community Engagement

United Way
UNITED WAY OF THE BAY AREA

TENANTS TOGETHER

foundation for sustainable development

ASHBY VILLAGE

Education

AAH
ASSOCIATION OF ART HISTORIANS

ROOTS OF SUCCESS

eneza education
spreading education everywhere

KUOW.ORG

Economic Development

Village Enterprise

TECHNOSERVE
BUSINESS SOLUTIONS TO POVERTY

Environmental

LIVELYHOODS

BioLite

KARIBU

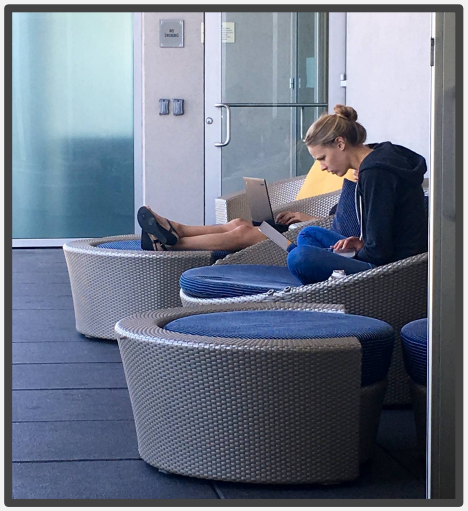
Where do Delta Fellows Work?

And many more!

What does the fellowship look like?

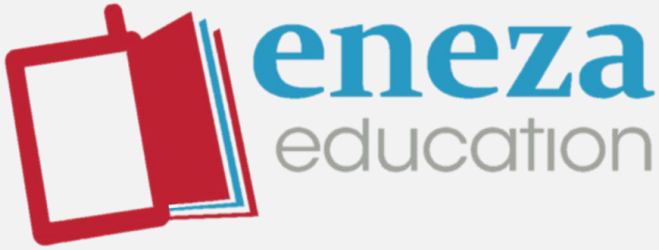
Monthly program-wide **hackathons** and ongoing social events

6 month engagement between non-profit and teams of 3 to 4 full-time data professionals

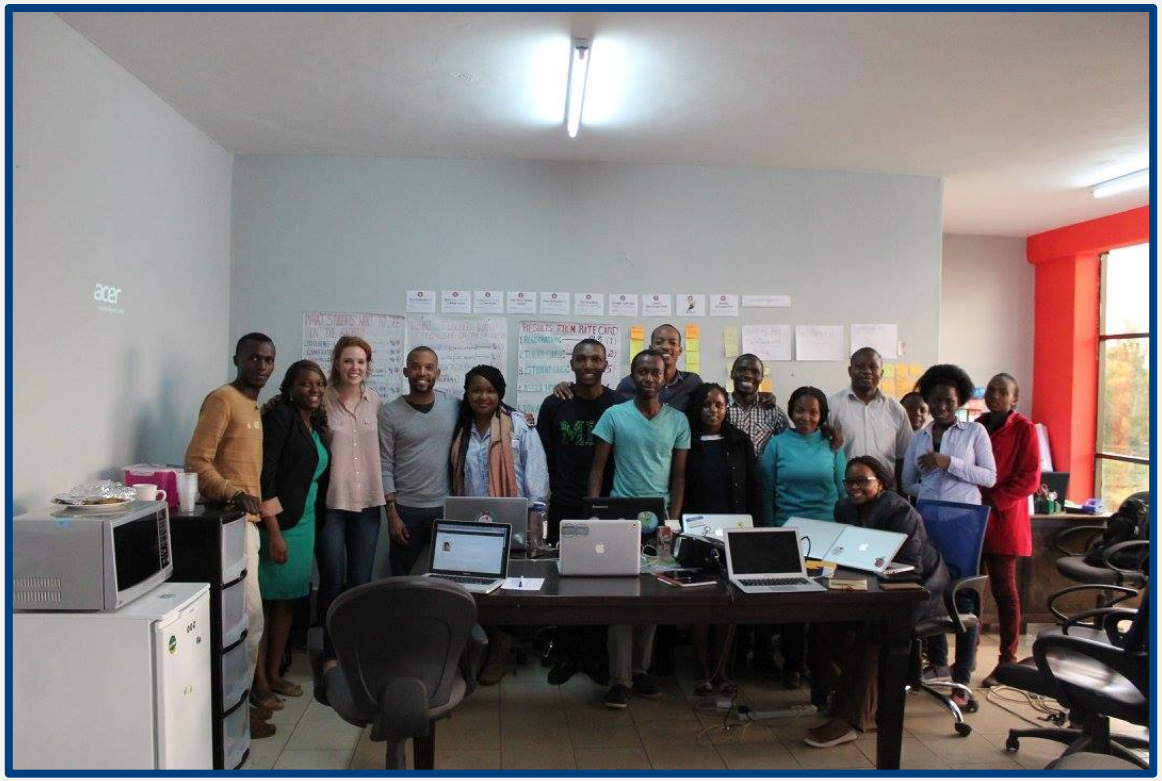


External **speakers and trainings** for ongoing technical growth and skill development

Technical Deep Dive -



Eneza Education



Eneza Education enables access to education on a low cost mobile phone.

Eneza Education

2012
Year Eneza Went Live

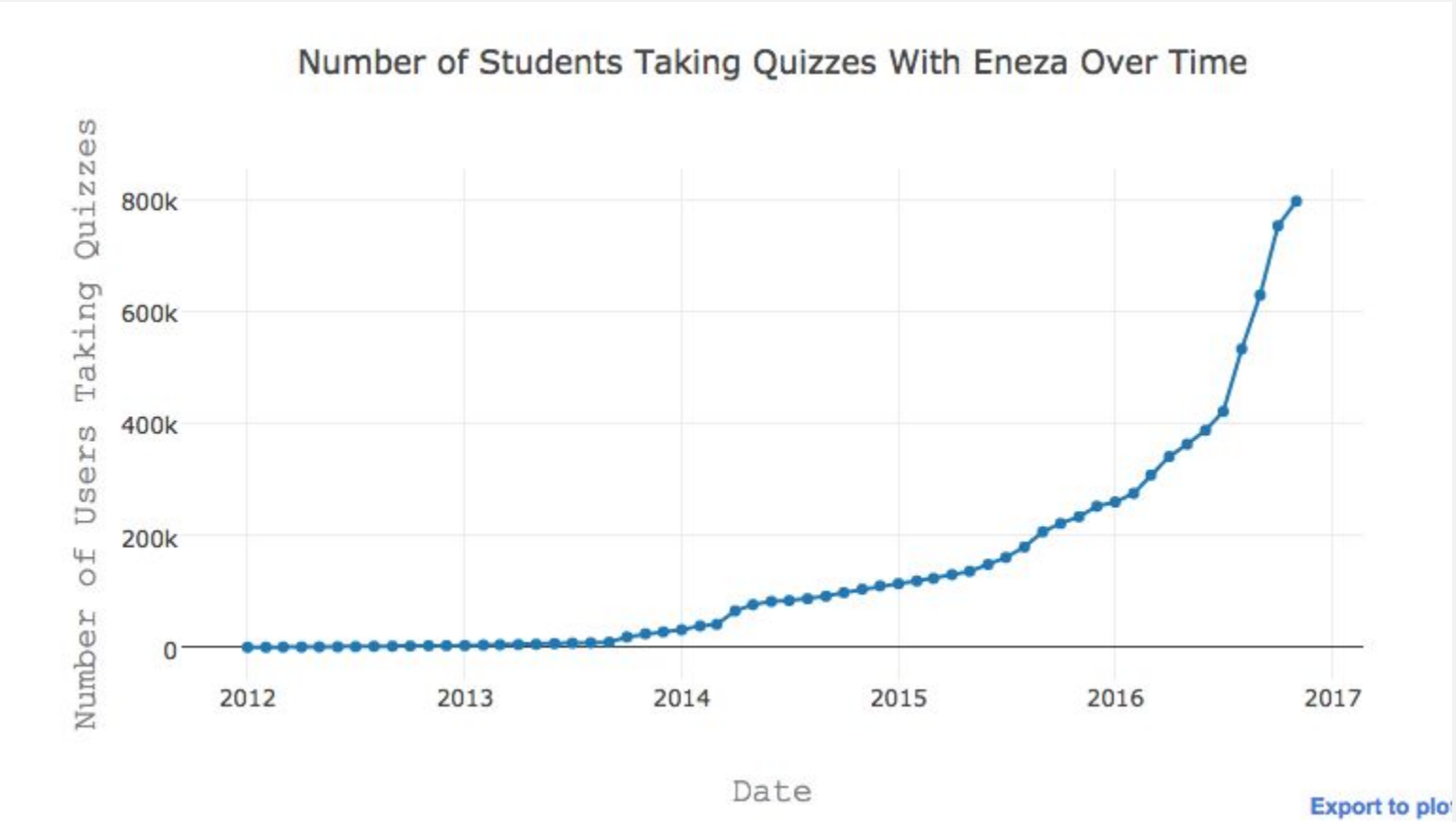
1,563,077
Number of Students on Eneza

10,144
number of quizzes available to students

3,993,463
number of questions answered in October

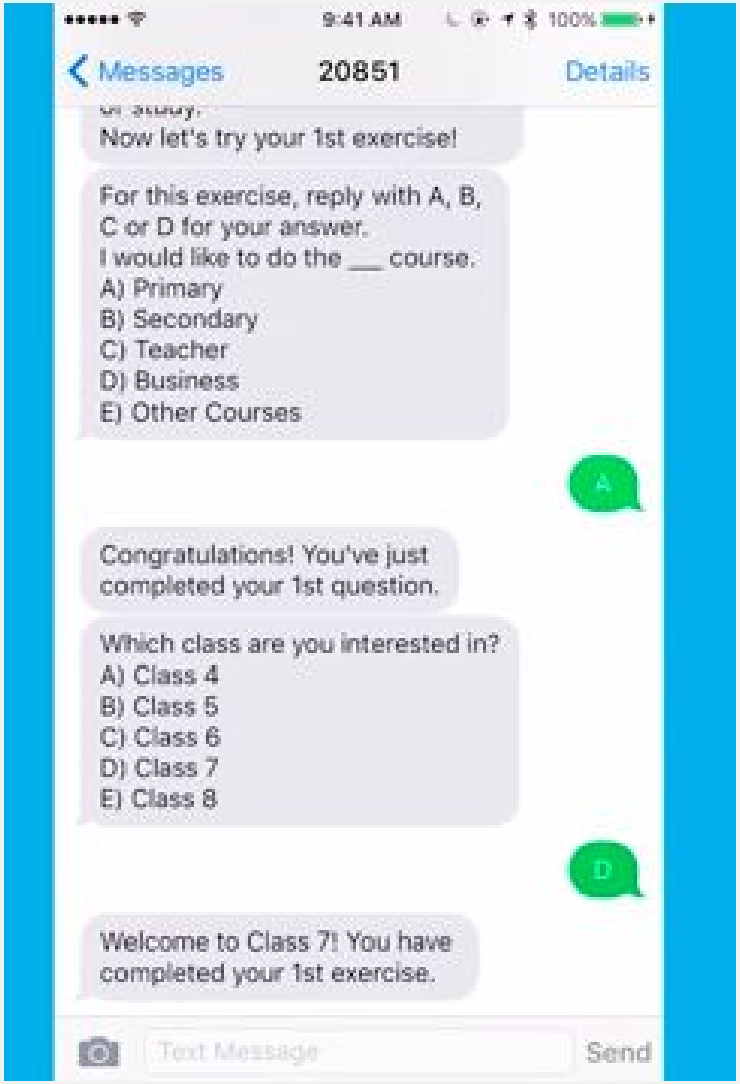
41.20
% of female students

Eneza Education



Incredible growth in number of students actively learning each month.

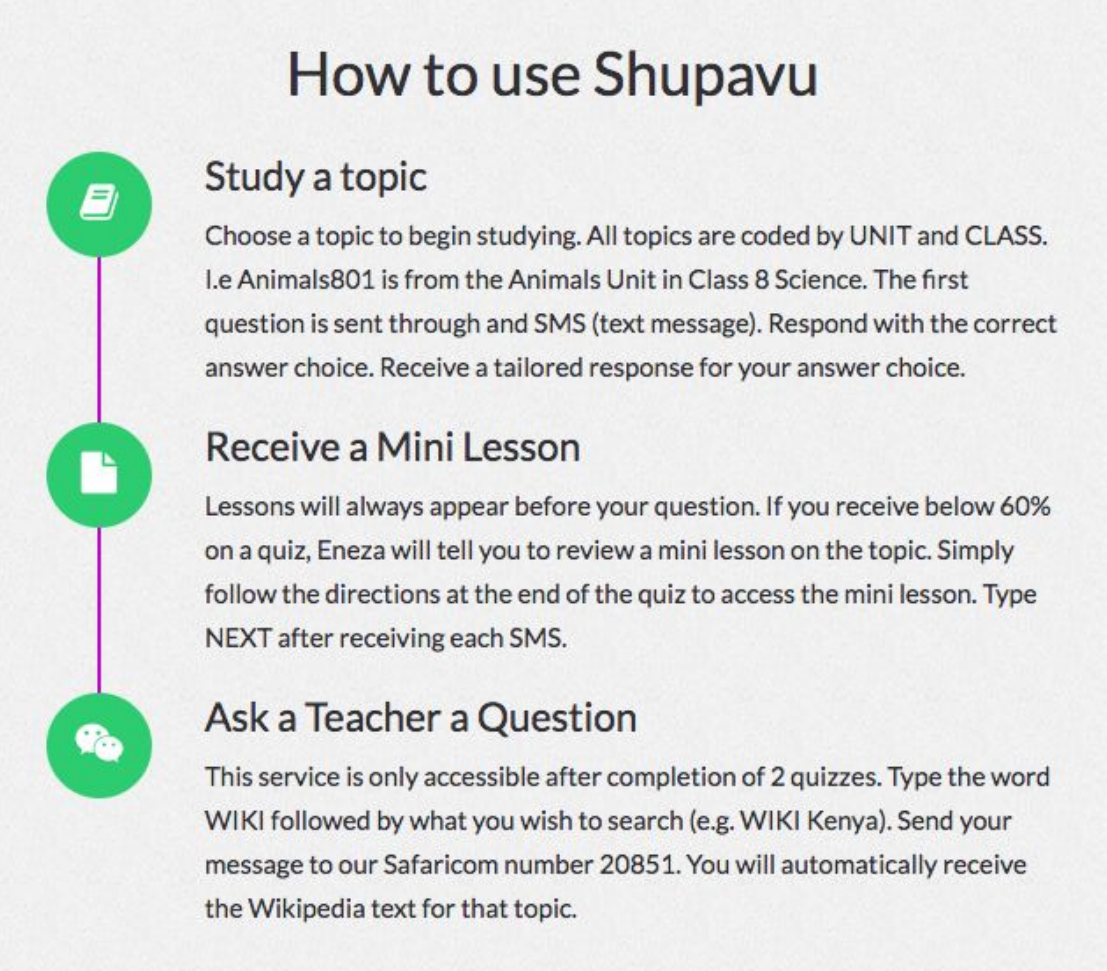
Eneza Education



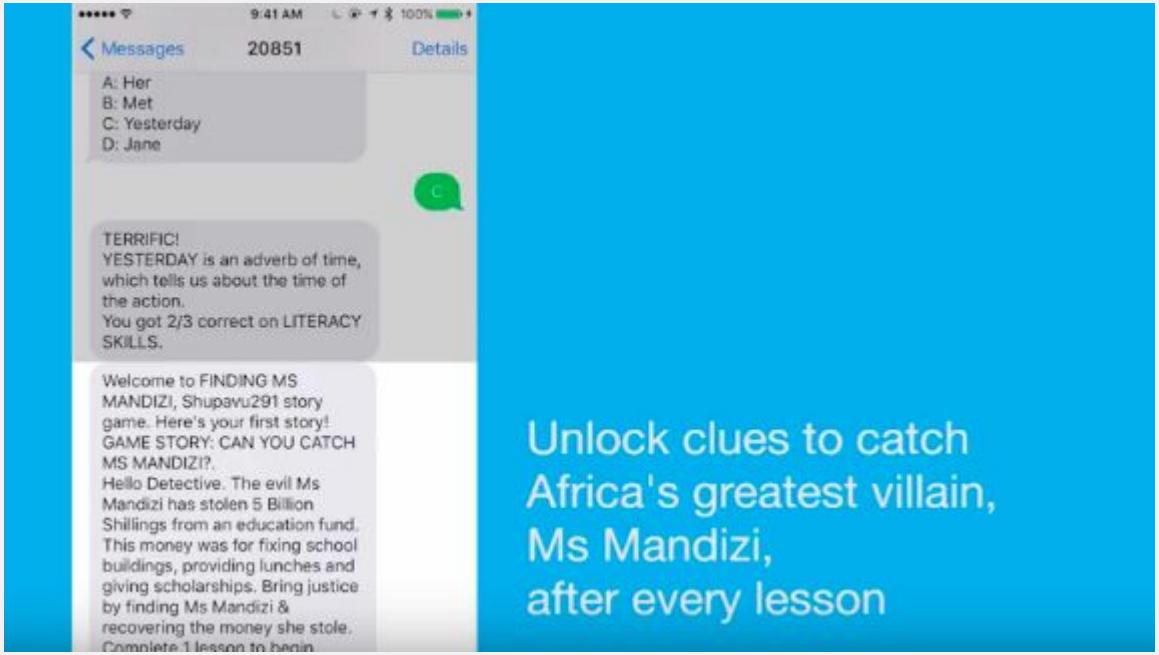
simple text based model with instant feedback to questions

pre smart phone portal =KES 10 for a weekly subscription ($0.098).

Eneza Education

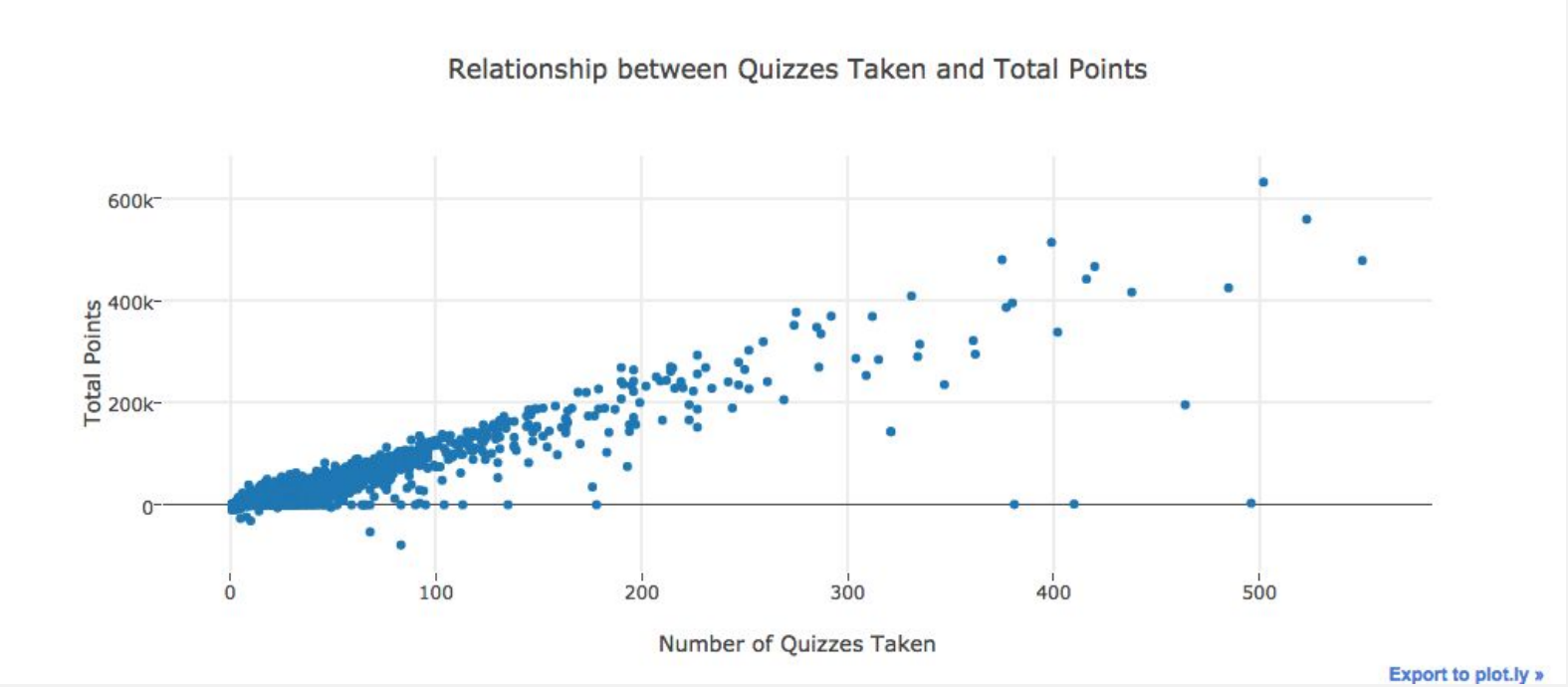


Eneza Education



Gamification techniques to improve retention with students.

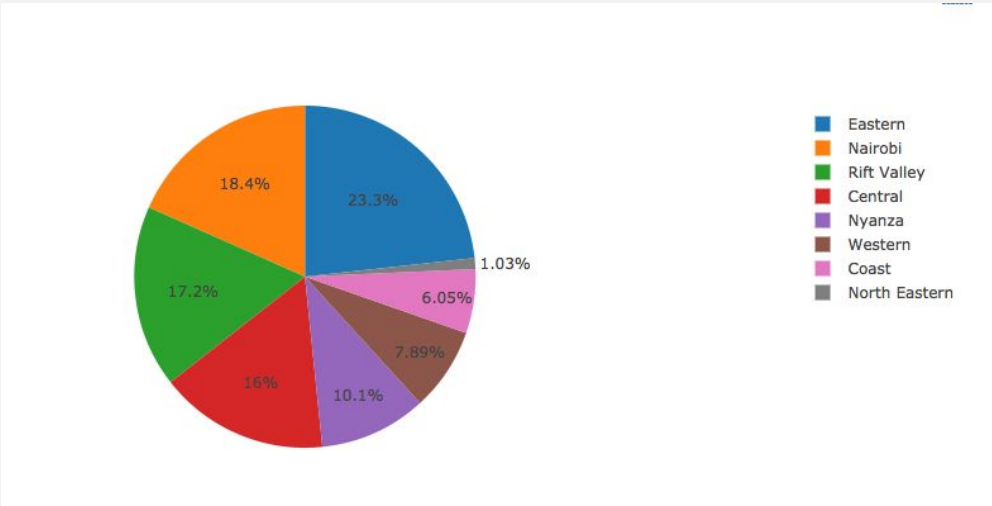
Eneza Education



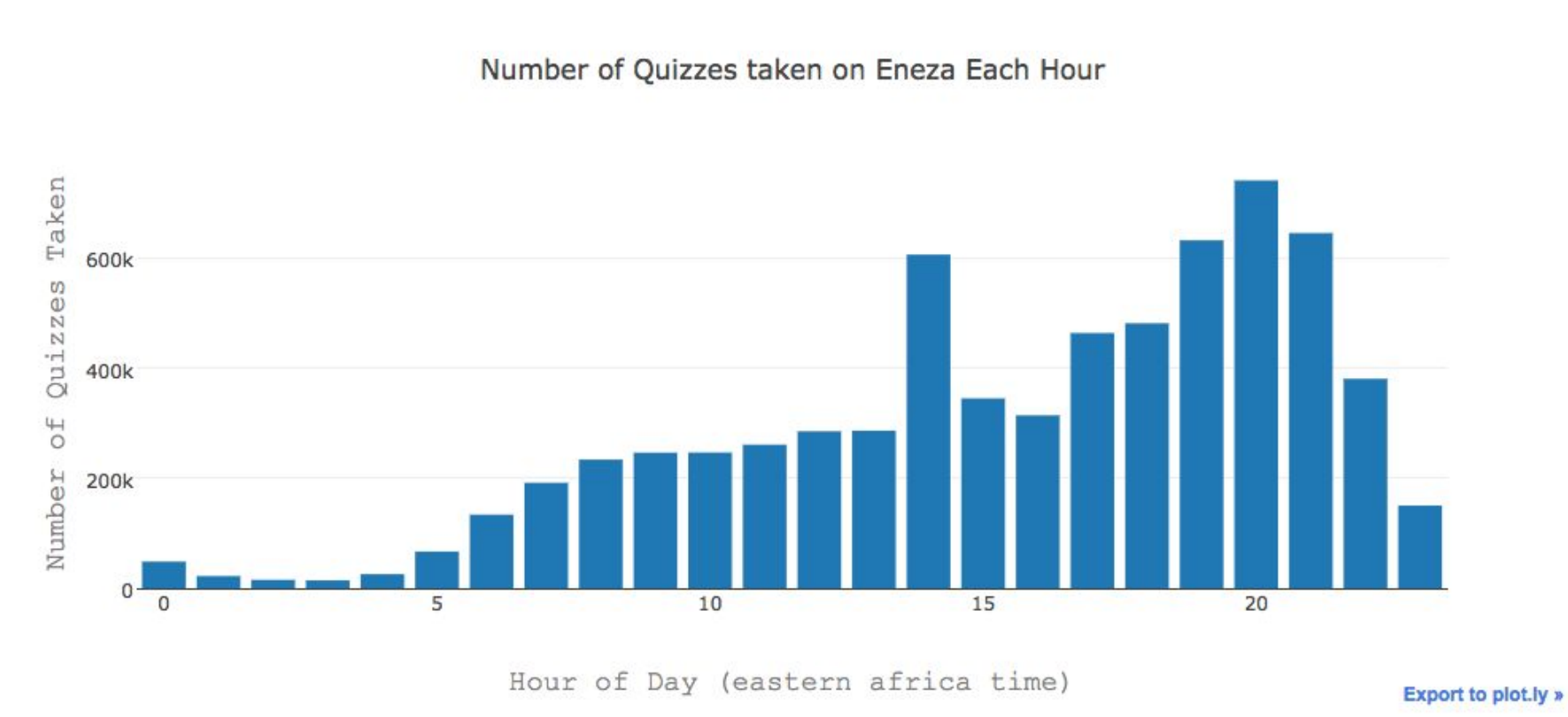
Gamification techniques to improve retention with students.

Eneza Education

Eneza started in Kenya, branching out to Tanzania and Ghana.

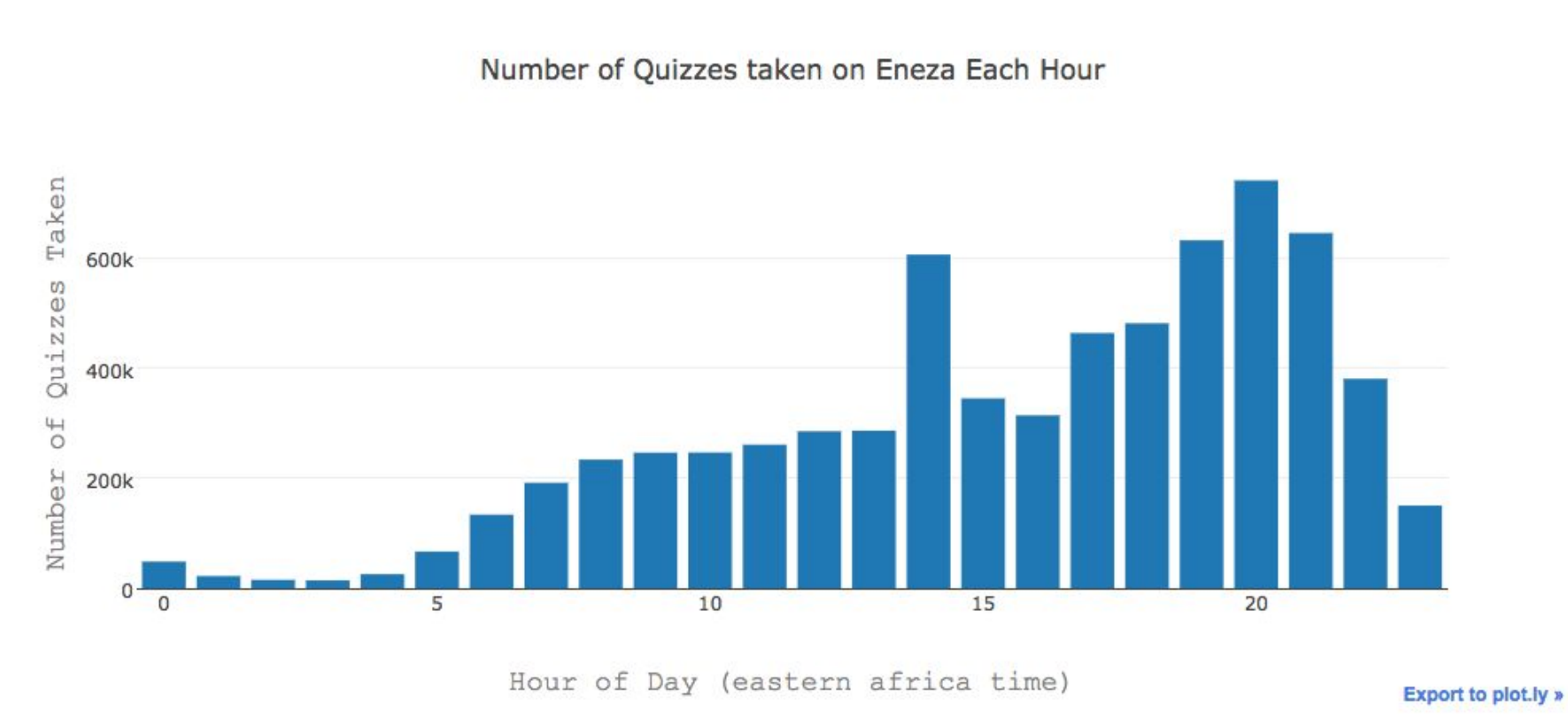


Eneza Education



Why is there a spike in activity towards the evening?

Eneza Education



Students mainly use their parents' cellphones to access Eneza.

Eneza Education

category	title
parent	How do I prepare for a future crisis or problem?
parent	Business
parent	Risks in Business
parent	What is Credit?
parent	How do I get my first customers?
parent	What is a contract?
parent	How do I keep my customers happy?
parent	What is VAT?
parent	How to budget for a Special Occasion?
parent	What is a cheque?

Because students have to use parents phone, Eneza creates quizzes for parents to also stay engaged with the platform.

Eneza Education

title_of_quiz
The Body is the Temple of God
TAS and God's creation
Concepts of God
Working for God
Overcoming fear, relying on God
God's Purpose
Christian teaching and God's creation
Response to God's creation
Work of Christians for God
Ways of reconciling with God

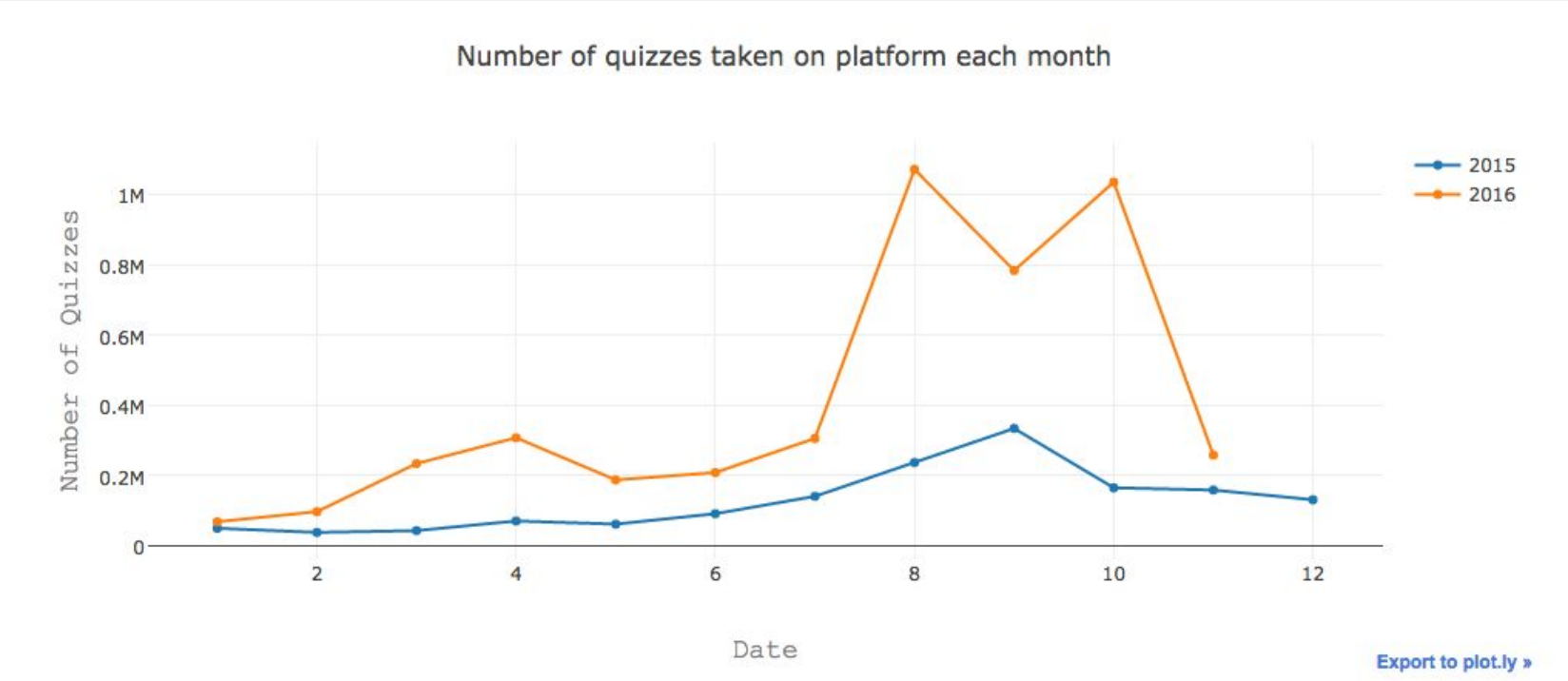
The types of quizzes help us understand how Eneza is connected with the region of the world it solves. 81% of Kenyans are Christians.

Eneza Education

title_of_quiz
Islam Along the Coast of Kenya
Roots of Islam
Roots of Islam
Obligations towards the dead: Following and ca...
Islamic teachings on circumcision
Islamic rituals during funerals
The role of Muslim leaders
The coming of Islam along the Kenyan Coast
The agents of spread of Islam along the Kenyan...
Hajj as a pillar of Islam

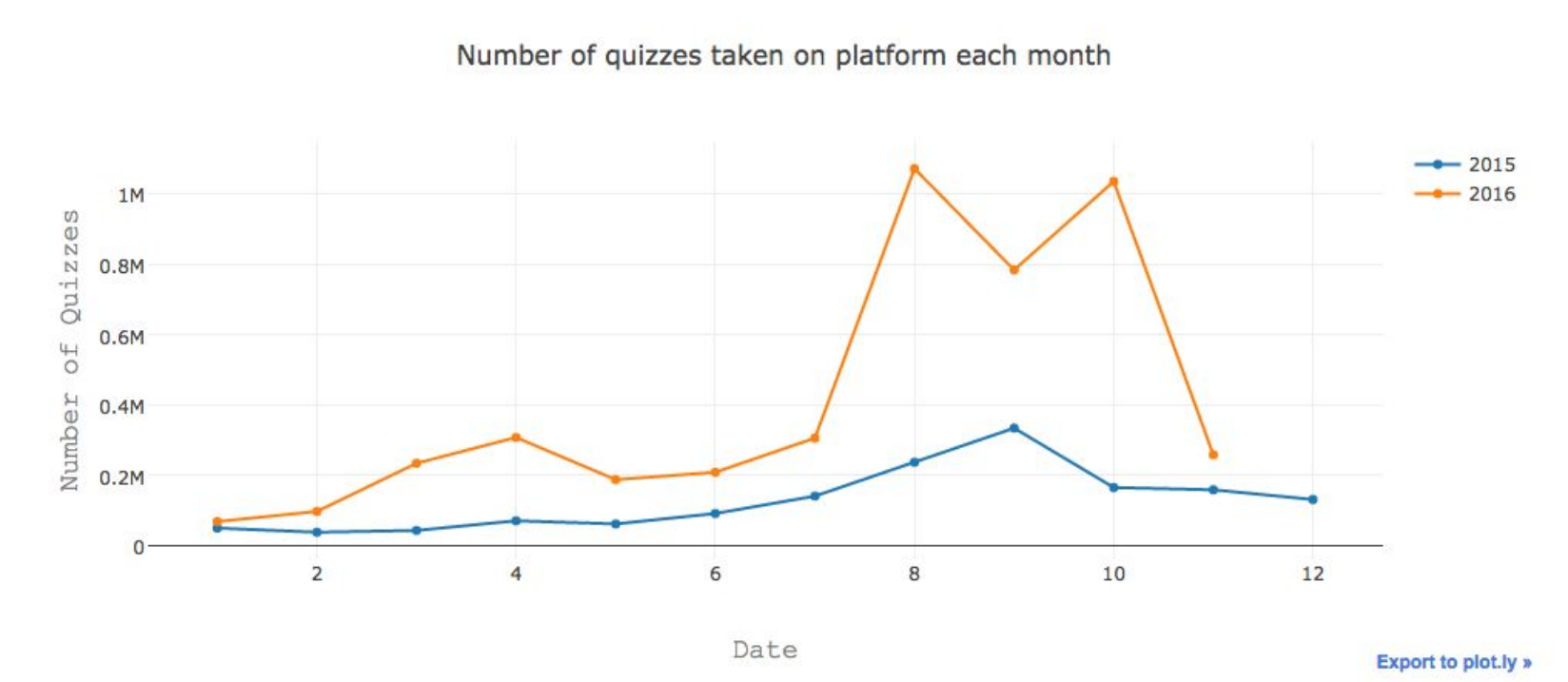
The types of quizzes help us understand how Eneza is connected with the region of the world it solves. 11% of Kenyans are Muslim.

Eneza Education



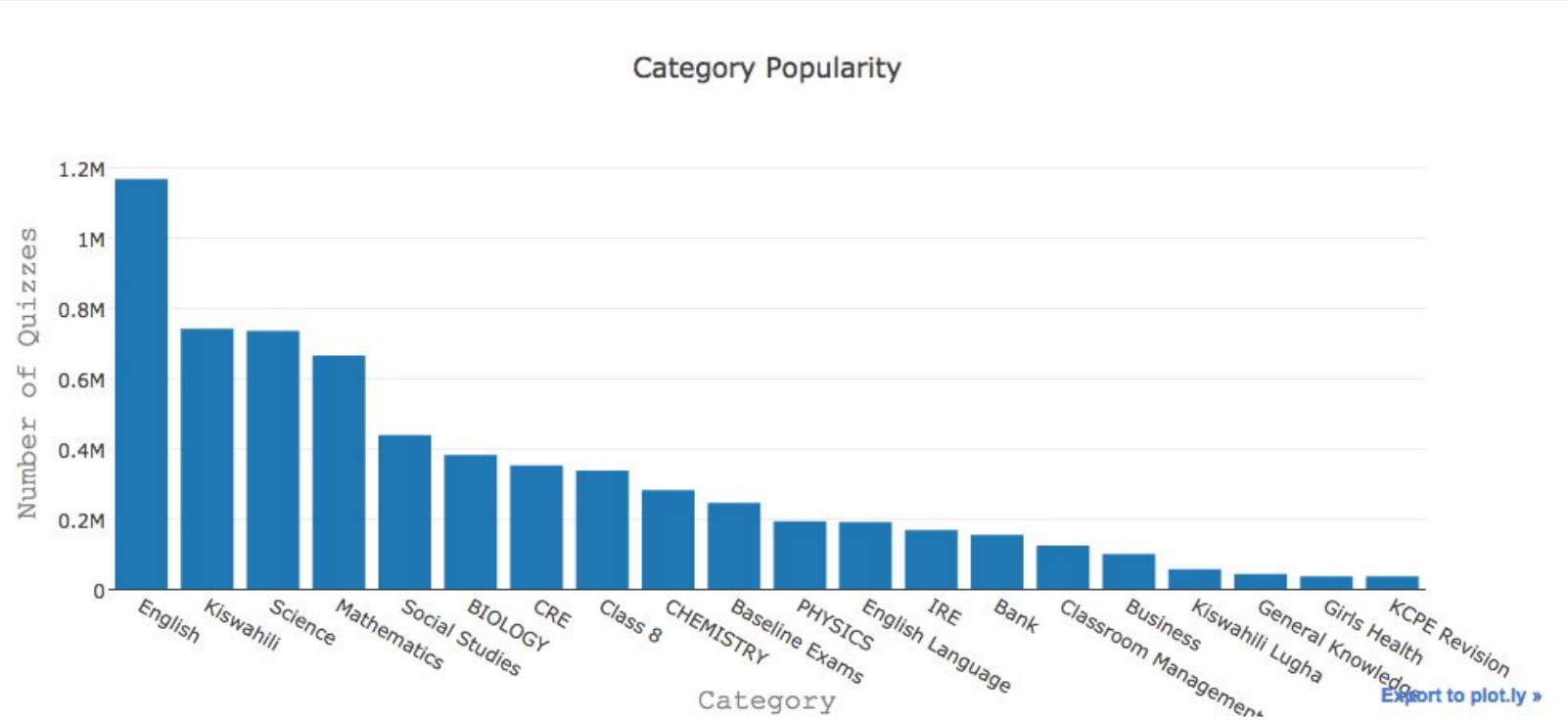
Why are there big spikes in engagement towards the end of the year?

Eneza Education



Activity peaks close to the Kenyan annual exams in October; it is at its lowest in January after Exams.

Eneza Education



Kiswahili and English are the national languages in Kenya. English, Kiswahili and Mathematics are all obligatory subjects for the secondary certification exam. This helps explain their popularity.

Lead you through one of the data products we built for Eneza.

Student engagement at Eneza: a churn analysis

Student engagement at Eneza: a churn analysis

Final product:

- We will build a model to help Eneza identify quizzes that are associated with high churn rates.
  - allows Eneza to quality-check the difficulty and content of quizzes that look like they may not engage students

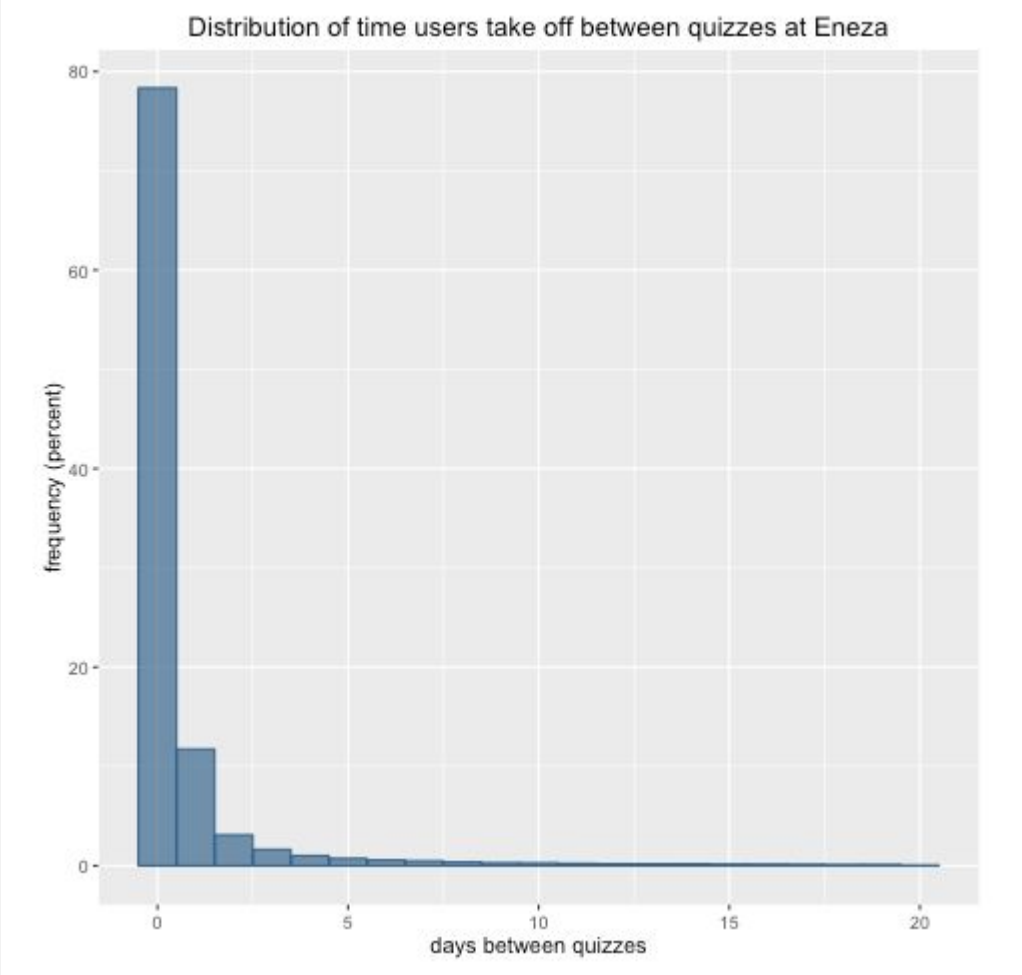
Student engagement at Eneza: a churn analysis

Questions we need to answer to build final product:

- How frequently do students "churn" (not come back to the site)?
- How does it relate to student's passing quizzes?
- What does the distribution of the number of quizzes students take before churning look like?
- Does "churn" behavior change as students take more quizzes?
- Does churn behavior vary as students take more quizzes?

Student engagement at Eneza: a churn analysis

Firstly let's define churn:

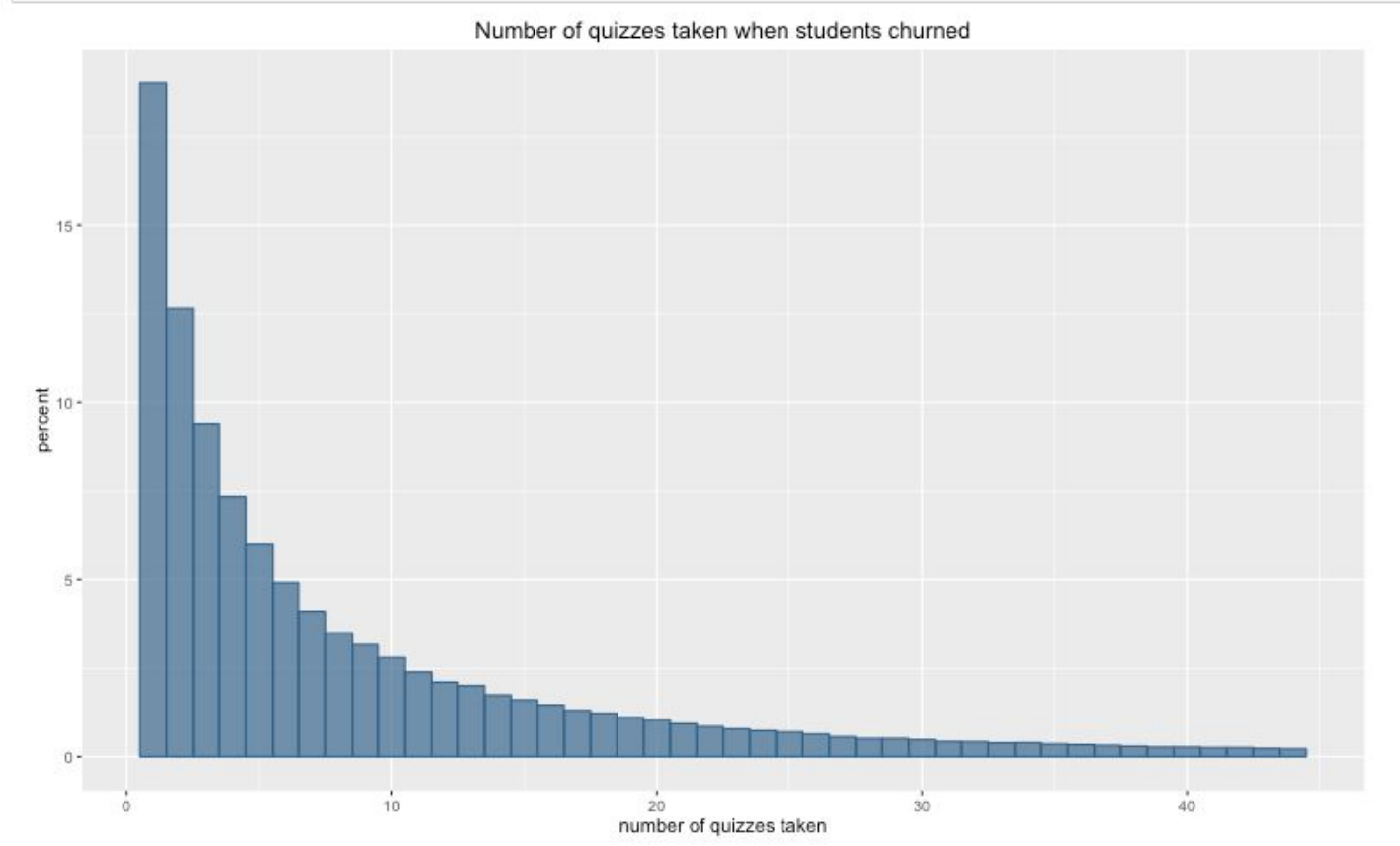


If a user returns for another quiz, they are most likely to do so on the same day. Based upon this plot we can conservatively define churned students (students who never come back) as:

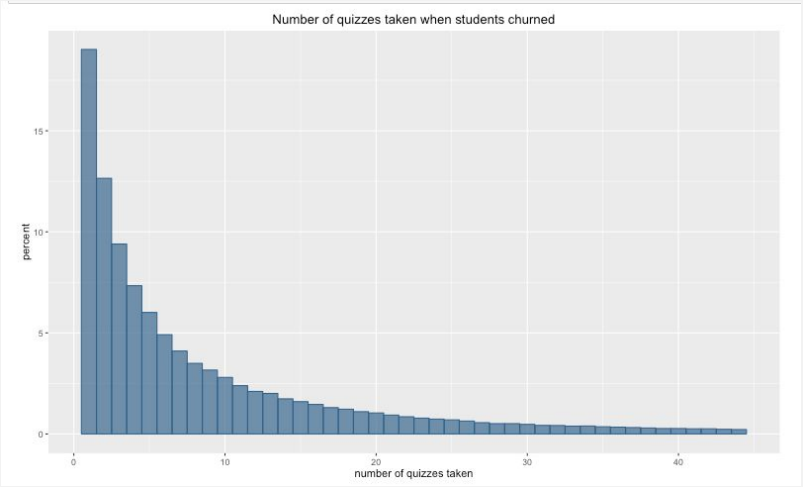
a student has churned if they not returned to the site within 30 days since the last quiz.

Student engagement at Eneza: a churn analysis

Let's understand the number of quizzes taken when students churned:



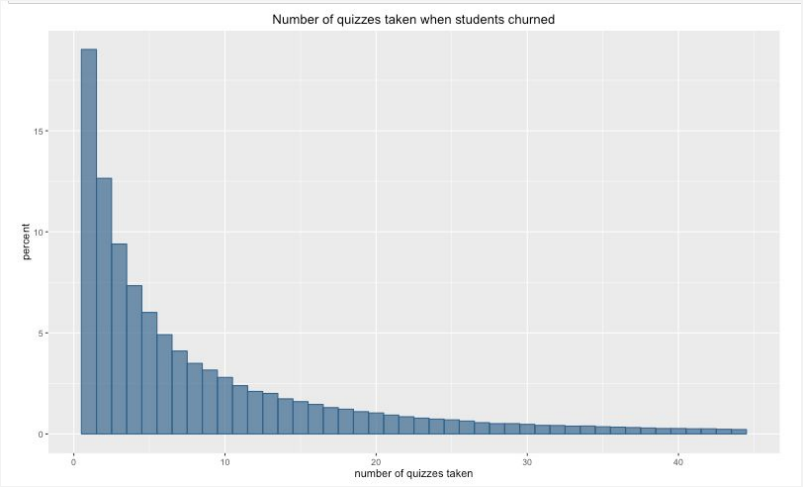
Let's understand how many quizzes students take before they churn:



Most students who churn only take a few quizzes before doing so.

This suggests the first few quizzes a student takes are very important for their experience, but also may be because Eneza is simply not a good fit for all students (natural dropoff).

Let's understand how many quizzes students take before they churn:

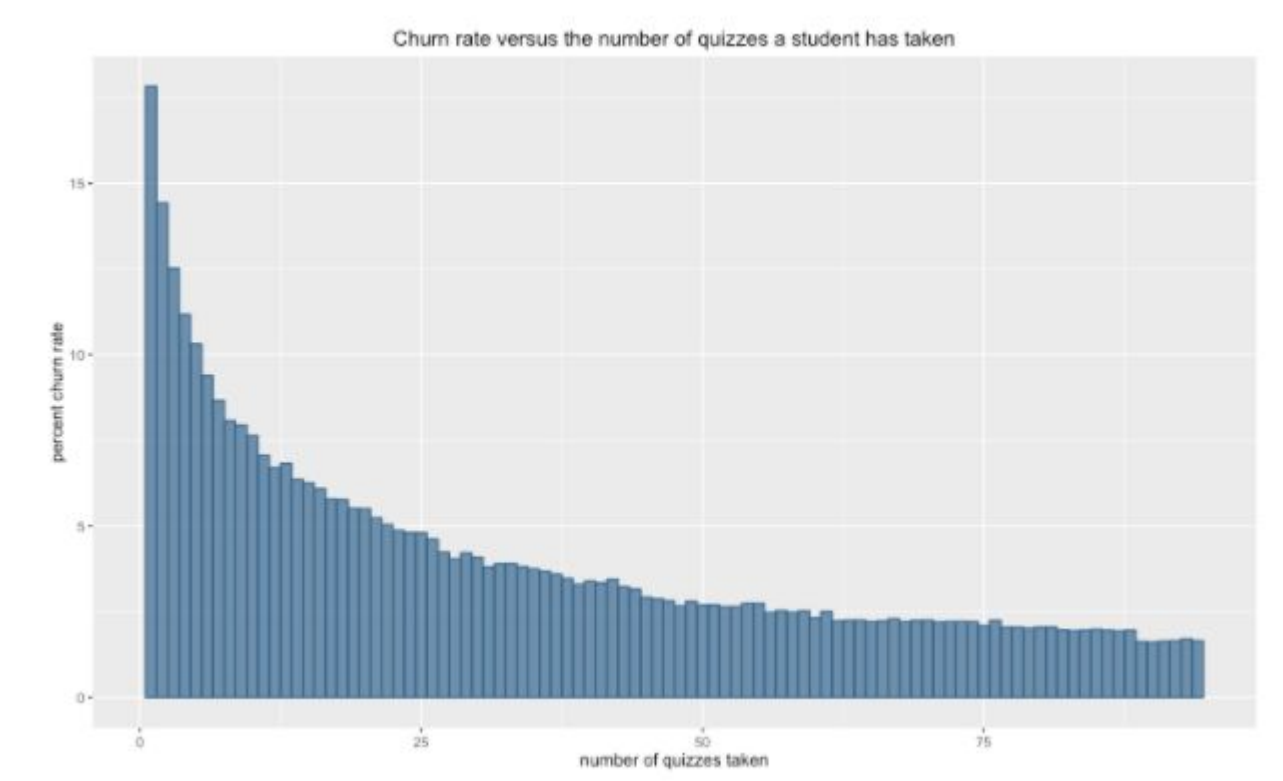


Most students who churn only take a few quizzes before doing so.

This suggests the first few quizzes a student takes are very important for their experience, but also may be because Eneza is simply not a good fit for all students (natural dropoff).

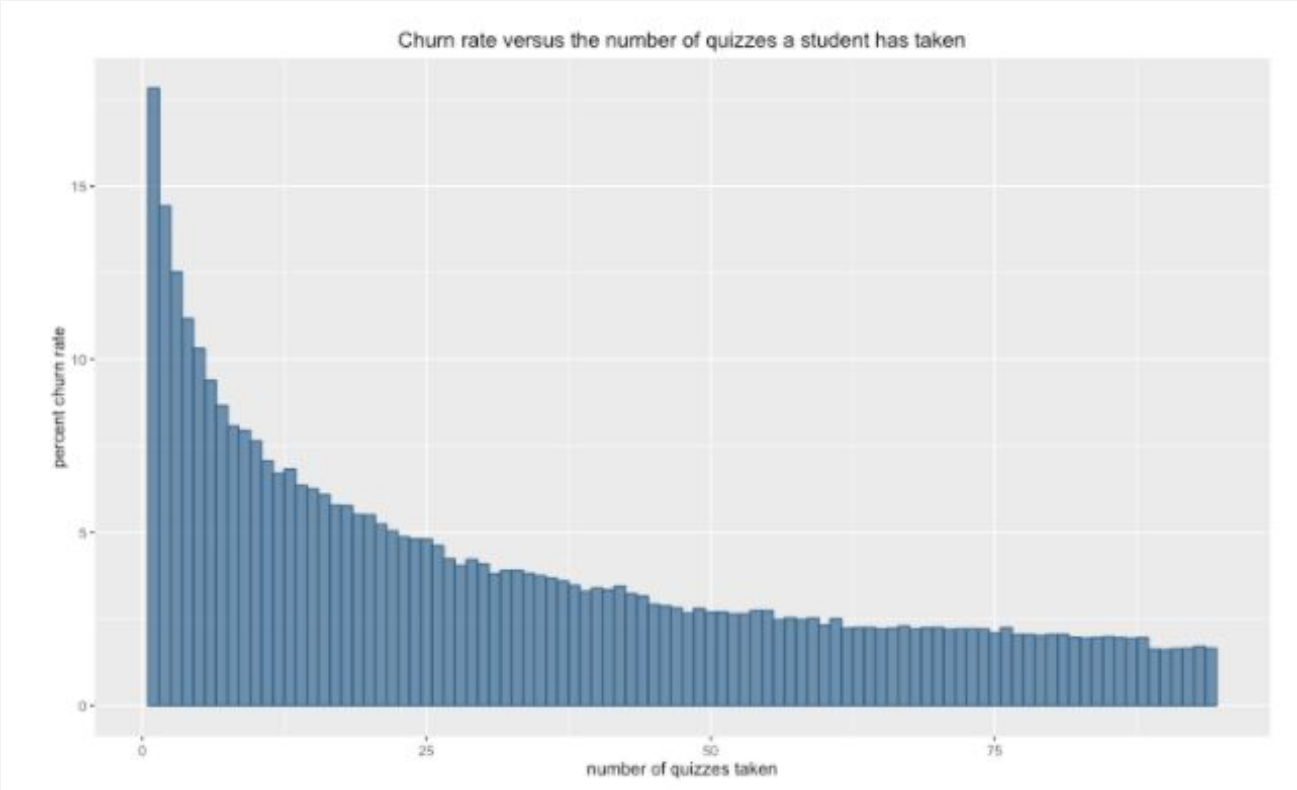
Student engagement at Eneza: a churn analysis

We can also think about this as a churn rate out of all students that take a number of quizzes.

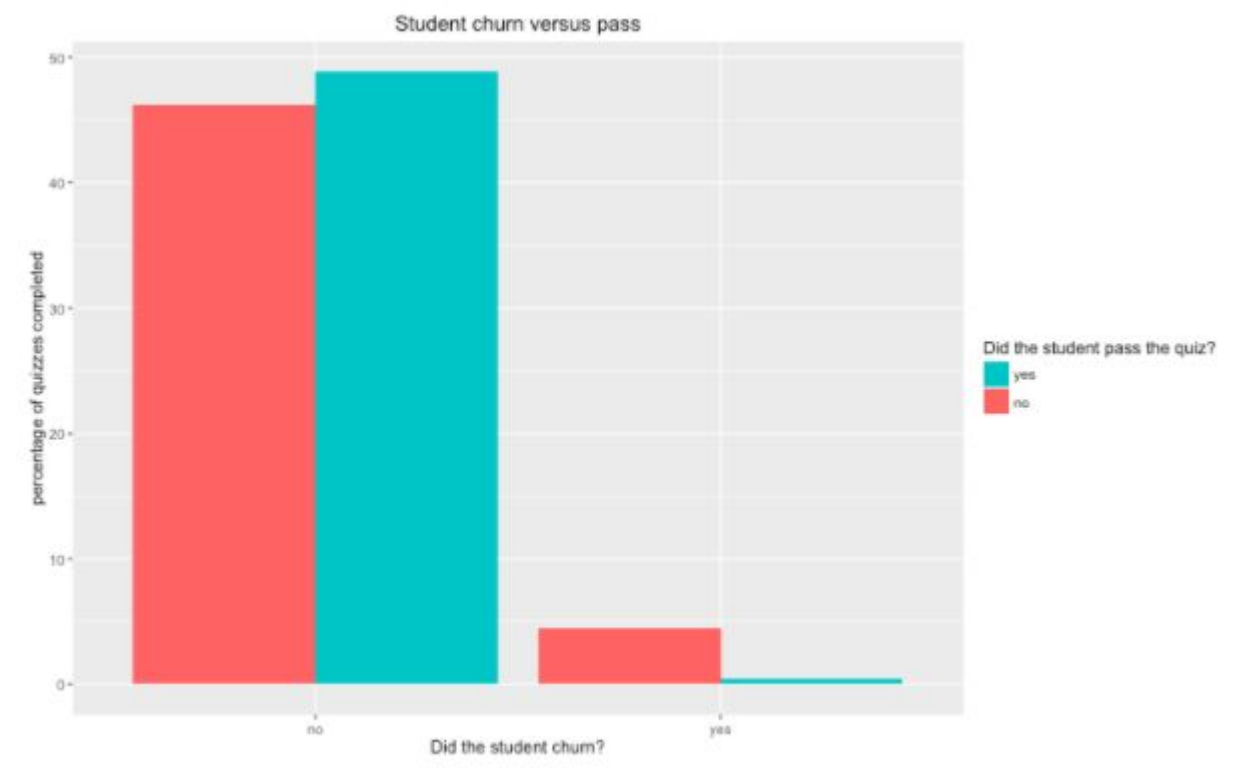


Student engagement at Eneza: a churn analysis

The churn rate is highest at the first quizzes, but as students become more "sticky" and take more quizzes it falls to ~3%.

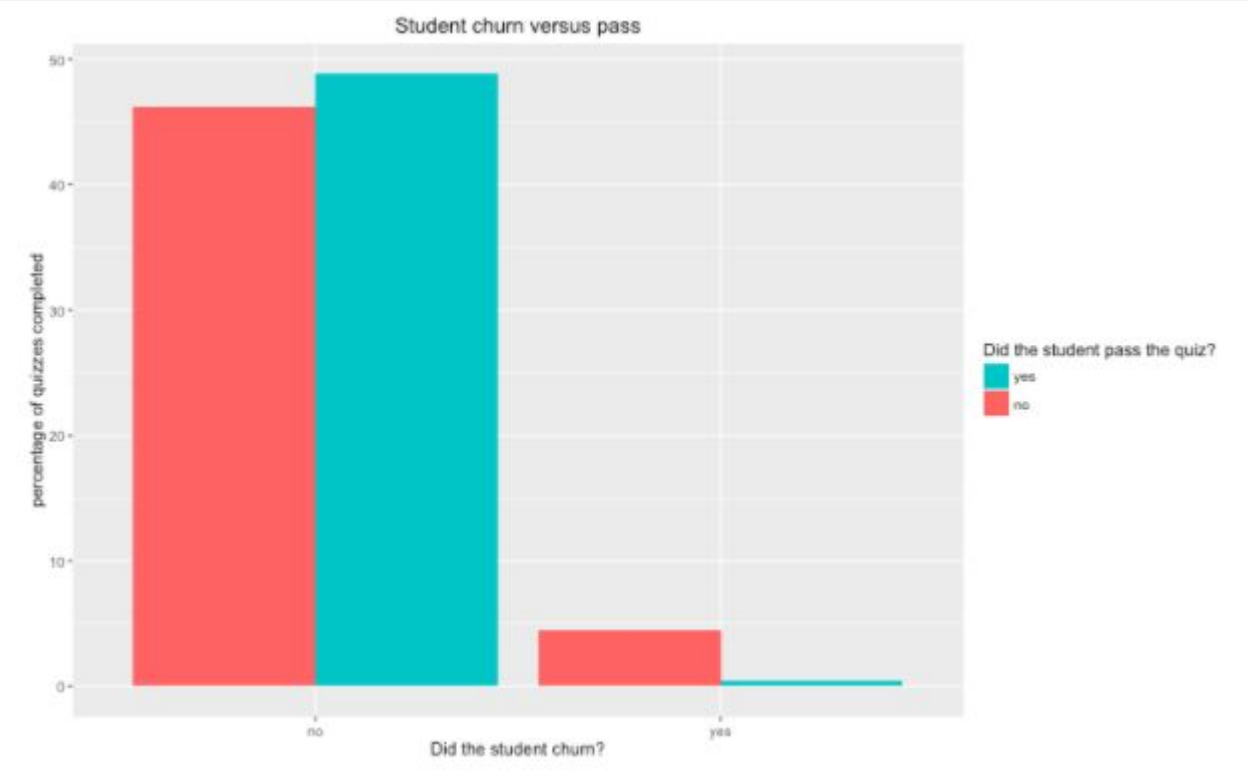


How is churn related to students passing quizzes?



Student engagement at Eneza: a churn analysis

Students who have churned are overwhelmingly more likely to have failed a quiz. This suggests difficulty of quiz is a big factor in understanding churn.



Student engagement at Eneza: a churn analysis

Conclusions based upon the finding that difficulty of quiz is linked to churn:

- Students may need positive feedback of a win (passing the quiz) to stay engaged
- One recommendation is that Eneza should think carefully about the first few quizzes a new student takes

Student engagement at Eneza: a churn analysis

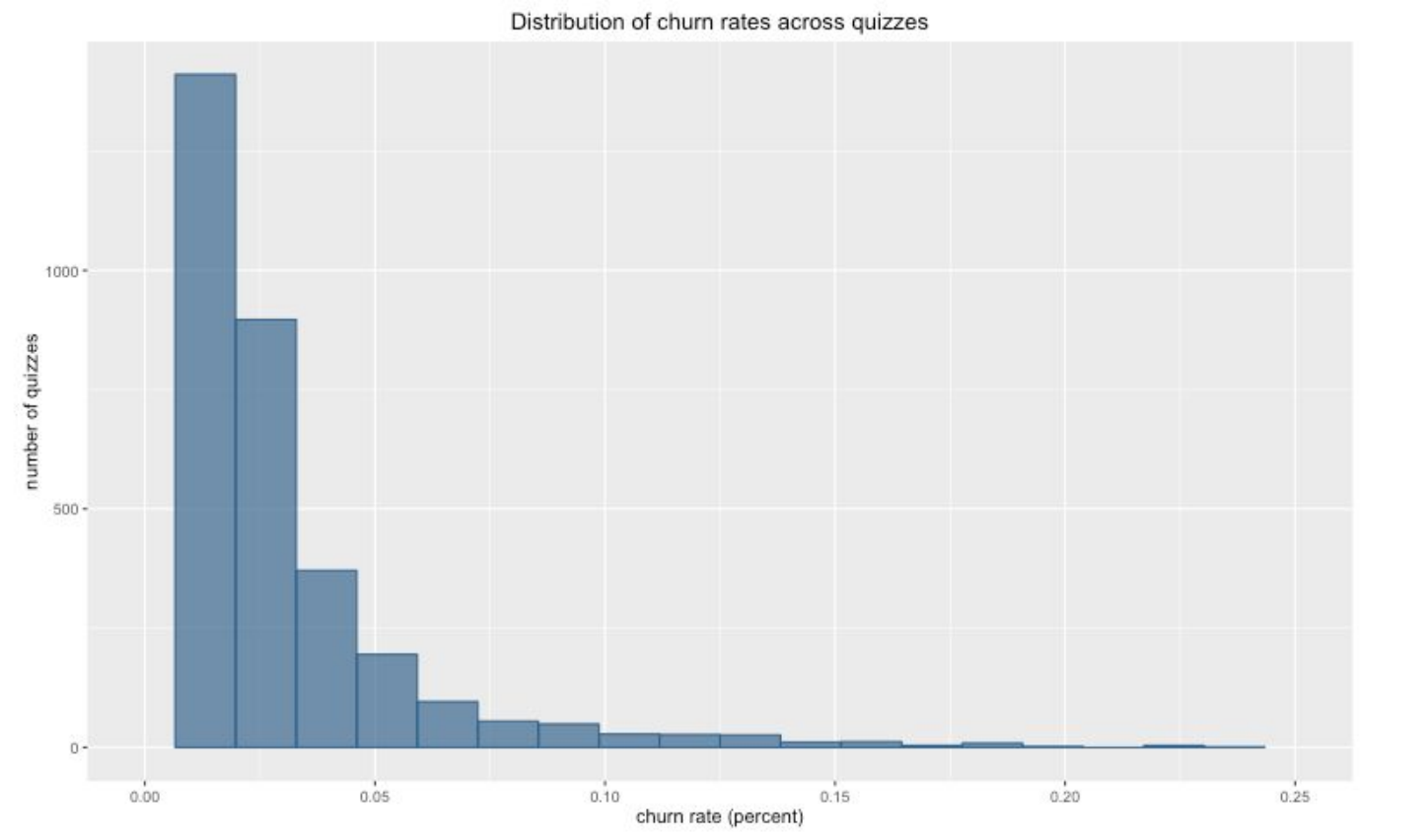
Now, let's isolate potentially problematic quizzes for Eneza. These are quizzes with very high churn rate.

We can start by just looking at the raw churn rate:

quiz churn rate=the number of students last quiz before churn/number of students who took quiz

Student engagement at Eneza: a churn analysis

Most courses have very low high churn rate but there is a long tail which suggests some courses are problematic.



Student engagement at Eneza: a churn analysis

Problems with using raw churn rate:

- Simply reporting the quizzes with the highest churn rate is problematic because many quizzes have small sample size.
- We could address this sample size issue by setting a minimum sample size threshold before calculating churn rate. But this isn't what we want either, because the earlier Eneza knows that a quiz might have problems the earlier they can take action.

Student engagement at Eneza: a churn analysis

Our solution is to instead take an Empirical Bayesian approach:

- imagine that the churn rate for each quiz is some unknown probability, drawn from a distribution of probabilities
- then, fit a model to the data that estimates this distribution and uses it to smooth out the empirical churn rates based on how much data is available.

Student engagement at Eneza: a churn analysis

Applying this shrinkage technique we expect:

- the model predictions to look similar to the averages where there is good data
- we expect the predictions to be reasonably close to the overall average if the sample size is small (this is quite important, since the goal of the model is to be actionable, especially on new quizzes, which it will not be if the regularization is poor)

Student engagement at Eneza: a churn analysis

Δnalytics

We can sanity check our results by looking at the churn prediction for small sample sizes:

	n_taken	n_churned	churn_rate	churn_rate_smooth
quiz_id				
7012	1	1	1	0.046387
8100	1	1	1	0.046387
8368	2	2	1	0.077815
8452	1	1	1	0.046387
8843	1	1	1	0.046387

If we just used the raw churn rate we would have predicted 100% churn. We have succeeded at reducing noise in the small sample size.

Student engagement at Eneza: a churn analysis

A look across a sample result shows model is performing well across different sample sizes.

	n_taken	n_churned	churn_rate	churn_rate_smooth
quiz_id				
2669	313	3	0.009585	0.013206
492	3042	85	0.027942	0.027909
259	956	26	0.027197	0.027140
7909	48	1	0.020833	0.023979
3540	62	4	0.064516	0.048127

So, given we are happy with our model. What are the most problematic quizzes?

	n_taken	n_churned	churn_rate	churn_rate_smooth	quiz_passmark	quiz_title	quiz_type	total
quiz_id								
6569	154	117	0.759740	0.712282	60	Abas3m Mmuab)	MULTIPLE	5
990	49808	28980	0.581834	0.581706	60	Types of animal feeds e,g pastures, fodder cro...	MULTIPLE	5
6935	224	112	0.500000	0.474650	60	Significance of death rites	MULTIPLE	5
7593	50104	18160	0.362446	0.362348	60	Literacy Skills	MULTIPLE	3
7002	2477	778	0.314090	0.312246	60	The Oxidizing Properties of Non-metals	MULTIPLE	5
5982	327	106	0.324159	0.310243	60	Factors That Improve Productivity	MULTIPLE	5
7247	400	119	0.297500	0.286581	60	Types of Friction	MULTIPLE	5
7580	16347	4381	0.268000	0.267742	0	How to Enjoy Shupavu291!	MULTIPLE	5
7899	619	156	0.252019	0.245514	60	The Strengths, Weaknesses and Contribution of ...	MULTIPLE	5
7758	852	213	0.250000	0.245276	60	Uses of Oxygen in Daily Life	MULTIPLE	5

Student engagement at Eneza: a churn analysis

Eneza will take a closer look at these problematic quizzes and try and understand why they are causing high churn with students.

	n_taken	n_churned	churn_rate	churn_rate_smooth	quiz_passmark	quiz_title	quiz_type	total
quiz_id								
6569	154	117	0.759740	0.712282	60	Abas3m Mmuab)	MULTIPLE	5
990	49808	28980	0.581834	0.581706	60	Types of animal feeds e,g pastures, fodder cro...	MULTIPLE	5
6935	224	112	0.500000	0.474650	60	Significance of death rites	MULTIPLE	5
7593	50104	18160	0.362446	0.362348	60	Literacy Skills	MULTIPLE	3
7002	2477	778	0.314090	0.312246	60	The Oxidizing Properties of Non-metals	MULTIPLE	5
5982	327	106	0.324159	0.310243	60	Factors That Improve Productivity	MULTIPLE	5
7247	400	119	0.297500	0.286581	60	Types of Friction	MULTIPLE	5
7580	16347	4381	0.268000	0.267742	0	How to Enjoy Shupavu291!	MULTIPLE	5
7899	619	156	0.252019	0.245514	60	The Strengths, Weaknesses and Contribution of ...	MULTIPLE	5
7758	852	213	0.250000	0.245276	60	Uses of Oxygen in Daily Life	MULTIPLE	5

Student engagement at Eneza: a churn analysis

Analytics

This is a simple model which can be operationalized easily and helps Eneza take action as soon as quiz is live.

	n_taken	n_churned	churn_rate	churn_rate_smooth	quiz_passmark	quiz_title	quiz_type	total
quiz_id								
6569	154	117	0.759740	0.712282	60	Abas3m Mmuab)	MULTIPLE	5
990	49808	28980	0.581834	0.581706	60	Types of animal feeds e,g pastures, fodder cro...	MULTIPLE	5
6935	224	112	0.500000	0.474650	60	Significance of death rites	MULTIPLE	5
7593	50104	18160	0.362446	0.362348	60	Literacy Skills	MULTIPLE	3
7002	2477	778	0.314090	0.312246	60	The Oxidizing Properties of Non-metals	MULTIPLE	5
5982	327	106	0.324159	0.310243	60	Factors That Improve Productivity	MULTIPLE	5
7247	400	119	0.297500	0.286581	60	Types of Friction	MULTIPLE	5
7580	16347	4381	0.268000	0.267742	0	How to Enjoy Shupavu291!	MULTIPLE	5
7899	619	156	0.252019	0.245514	60	The Strengths, Weaknesses and Contribution of ...	MULTIPLE	5
7758	852	213	0.250000	0.245276	60	Uses of Oxygen in Daily Life	MULTIPLE	5

Thank you! Questions?

Eneza Education

We are working with Eneza to share all of our code publically. Watch our public repo for when it becomes available.

A tutorial of how to use the different libraries we used is available publically here: https://github.com/DeltaAnalytics/python_tutorials.

Ways to be involved:

1. Fellow
2. Mentor
3. Guest Speaker

Visit our website at deltanalytics.org for more information.

Fellowship Program 2017

Preview of some of the organizations currently in our application process