Data Report 17
Early Adopter Student Profiles

Zachery McKinnon - 4/17/2019

SUMMARY

In this exploratory data analysis, I use a number of student features to group recent Saylor profiles. In particular, I group students based on their reason for joining, subject matter, country, and enrollment-completion dyads. This information is helpful for understanding the profile of students who have adopted Saylor Academy at this stage in our development and potential paths of least resistance going forward.

INTRODUCTION

Who is the current Saylor student? What does a successful Saylor student look like? Taking stock of our student base is helpful for establishing where we've been and where we could go in the future. Therefore, in this report, I look at the profile of Saylor students in 2018 based on their reason for joining, subject matter, country, and enrollment-completion dyads. It is important to note that these results represent who the Saylor student is, not necessarily who the Saylor student should be.

SAMPLE

The sample includes the last 40,000 confirmed users to register in 2018, which essentially includes those who enrolled (and verified their account) in the period from Mid-May 2018 to Dec 2018. The MySQL query used to collect their information is provided in the appendix.

METHOD

The tech team collectively decided on 6 potentially useful grouping strategies: by enrollment and completion, by geographical location, by reason for joining, by subject matter, by days to completion, and by types of activity. In this stage, I decided to look at the first four groups. Since I wanted to have a reasonably small number of groups for each grouping, I had to do some additional categorizing of continuous data, namely enrollments and completions, and abundant categorical data, namely all the countries in the world and subject matter. I basically used these aspect as pivots to look at the same information: enrollments/student, credit-possible enrollments/student, activities/student, quizzes/student, completions/student, credit-possible completions/student, and average days to completion among course completers, in addition to the grouping aspects themselves.

Reason for Joining

Reason for joining is a non-overlapping choice of the most applicable of 3 options: transfer college credit to my college/university, earn a certificate to advance my career, and learn for personal enrichment. Note that this question was not in effect for the entirety of the sample period, and only approximately 60% of the sample chose one of the three options.

Subject Matter

I also looked at the profile of students in 3 of our largest subject matters: BUS, CS, and PRDV. Unlike reason for joining, a student here can fall into multiple categories; for example, he can be enrolled in both business and computer science courses.

Enroller and Completer groups

For this grouping, I wanted to divide students based on how many courses they enrolled in and how many courses they went on to complete. Due to space and data considerations, I chose 4 categories for each aspect: 0, 1, 2-4, 5+, for a total of 16 possible combinations. However since enrollments is a necessary step of completions, there were 10 groups in actuality, marked with an x below:

Enrollment-Completion Dyads

COMP → EROLL ↓	0	1	2-4	5+
0	x			
1	x	x		
2-4	x	x	x	
5+	x	x	x	x

Location

I chose to single out all countries with at least 1000 users in the sample, which included the United States, India, Morocco, Algeria, Nigeria, and Egypt. All other countries were grouped together to represent a rest-of-the-world baseline. Furthermore, students had the option to leave this blank or were not given the option when they signed up, and this applied to approximately 20% of students. Their information is also reported.

RESULTS

Below are the results of my process of grouping and pivoting for the 4 features discussed above.

REASON FOR JOINING

Reason for Joining	Number of Students	Subject Enrollment	Enrolls./ Student (Credit Poss.)	Activities/ Student (Quizzes)	Comps./ Student (Credit Possible)	Avg. Days to Comp.
Earn a Certificate	12679	BUS: 23% CS: 23% PRDV: 9%	1.3 (0.6)	7.0 (1.0)	0.10 (0.02)	32
Personal Enrichment	8566	BUS: 18% CS: 17% PRDV: 5%	1.2 (0.6)	6.4 (0.9)	0.07 (0.02)	34
Transfer College Credit	3908	BUS: 27% CS: 16% PRDV: 4%	1.9 (1.3)	14.7 (1.8)	0.12 (0.07)	40

First note that most students chose to join to earn a certificate (50%), followed by personal enrichment (34%) and transfer college credit (16%). That said, students aiming to transfer college credit had higher levels of engagement in courses from beginning to end, enrolling in 46%-58% more courses and completing 20%-71% more courses on average, particularly credit courses. Those on the site for personal enrichment had the lowest investment in Saylor courses across the board.

Points worth considering:

- Students aiming to transfer college credit to some extent take care of themselves. They already engage with the site relatively well.
- In contrast, it appears we need to put in more work and provide more hand holding for students seeking personal enrichment and certification. Certification in particular represents the motivation of 50% of our students, so we should work on getting their level of engagement to the same level as credit seekers.

SUBJECT MATTER

Type of Student	Students	Reason for Joining	Enrolls./ Student (Credit Poss.)	Activities/ Student (Quizzes)	Comps./ Student (Credit Possible)	Avg. Days to Comp.
Business Student	8182	Cert: 53% Personal: 28% Credit: 19%	2.6 (1.6)	13.9 (2.0)	0.16 (0.08)	40
Computer Science Student	7392	Cert: 58% Personal: 30% Credit: 12%	2.2 (1.2)	11.3 (1.6)	0.17 (0.05)	33
PRDV Student	2542	Cert: 65% Personal: 27% Credit: 8%	3.9 (1.1)	19.9 (3.5)	0.52 (0.06)	31

Next, looking at students enrolled in specific types of courses (again, note that one student can fall into multiple buckets here), we see higher engagement numbers across the board for PRDV students. Of course, it is expected that PRDV students are more likely to finish a course given that PRDV courses tend to be shorter, but somewhat more surprising is that they enroll in and complete activities more often as well. Also note the average time to complete among PRDV students, 31 days, which suggests that these students are also going on to take longer courses or spend more time than expected on PRDV courses.

Points worth considering:

- Students appear to be encouraged by low-effort course completions (e.g., PRDV). Ultimately, they end up enrolling in the most courses and engaging with the most activities on our website.

ENROLLMENT-COMPLETION DYADS

Enroll, Comp	Students	Reason for Joining	Subject	Enrolls./ Student (Credit Poss.)	Activities/ Student (Quizzes)	Comps./ Student (Credit Possible)	Avg. Days to Comp.
Enroll 0 Comp 0	10636	Cert: 49% Personal: 38% Credit:12%	BUS: 0% CS: 0% PRDV: 0%	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	N/A
Enroll 1 Comp 0	20050	Cert: 51% Personal: 35% Credit: 14%	BUS: 24% CS: 22% PRDV: 4%	1.0 (0.6)	3.6 (0.6)	0.0 (0.0)	N/A
Enroll 1 Comp 1	1357	Cert: 54% Personal: 26% Credit: 20%	BUS: 15% CS: 36% PRDV: 23%	1.0 (0.3)	25.8 (2.5)	1.0 (0.3)	34
Enroll 2-4 Comp 0	5372	Cert: 51% Personal: 29% Credit: 20%	BUS: 36% CS: 30% PRDV: 10%	2.4 (1.3)	12.0 (1.6)	0.0 (0.0)	N/A
Enroll 2-4 Comp 1	650	Cert: 57% Personal: 23% Credit: 19%	BUS: 30% CS: 31% PRDV: 34%	2.6 (1.0)	48.0 (6.2)	1.0 (0.3)	37
Enroll 2-4 Comp 2-4	261	Cert: 57% Personal: 23% Credit: 20%	BUS: 38% CS: 30% PRDV: 36%	2.8 (1.0)	61.7 (8.5)	2.2 (0.8)	37
Enroll 5+ Comp 0	1271	Cert: 38% Personal: 29% Credit: 33%	BUS: 60% CS: 39% PRDV: 30%	8.9 (4.7)	25.7 (3.4)	0.0 (0.0)	N/A
Enroll 5+ Comp 1	199	Cert: 56% Personal: 21% Credit: 23%	BUS: 61% CS: 40% PRDV: 42%	8.7 (4.1)	115.6 (13.0)	1.0 (0.4)	35
Enroll 5+ Comp 2-4	155	Cert: 56% Personal: 19% Credit: 25%	BUS: 66% CS: 32% PRDV: 57%	10.0 (4.0)	99.6 (15.0)	2.5 (0.8)	39
Enroll 5+ Comp 5+	49	Cert: 43% Personal: 33% Credit: 23%	BUS: 65% CS: 37% PRDV: 65%	16 (5.8)	208.1 (34.6)	7.8 (2.8)	37

The first thing to note is that ¼ of our student sample did not even enroll in 1 course and ½ of our sample enrolled in just 1 course but did not finish it. That means that together, ¾ of our sample had minimal engagement with our site. That group tended to use the site for personal enrichment more than the other groups, which suggests that we may not yet have captured that group of students (or that that group is not capable of being captured in terms of completions). In addition, it is worth noting that students who enrolled in many courses overwhelmingly did so in business, whereas those who enrolled in many courses *and* completed many courses did so in PRDV as well.

Points worth considering:

- It may be difficult to get students using the site for personal enrichment to engage with our site and improve our KPIs.
- ¾ of our student body has no or very little engagement with our site. Therefore, getting students to the site may be less important than figuring out how to keep them. These students seek certifications and personal enrichment.
- High-enrollment students tend to take courses in business.
- High-enrollment and high-completion students tend to take courses in business and professional development.

LOCATION

Country	Students	Reason for Joining	Subject	Enrolls./ Student (Credit Poss.)	Activities/ Student (Quizzes)	Comps./ Student (Credit Possible)	Avg. Days to Comp.
US	9921	Cert: 29% Personal: 34% Credit: 37%	BUS: 27% CS: 13% PRDV: 6%	1.7 (1.0)	10.9 (1.5)	0.09 (0.04)	43
India	3061	Cert: 68% Personal: 27% Credit: 5%	BUS: 13% CS: 37% PRDV: 17%	1.3 (0.5)	9.1 (1.4)	0.23 (0.04)	31
Morocco	2982	Cert: 75% Personal: 18% Credit: 6%	BUS: 13% CS: 17% PRDV: 3%	1.0 (0.4)	3.2 (0.5)	0.02 (0.01)	30
Algeria	1796	Cert: 75% Personal: 19% Credit: 6%	BUS: 11% CS: 19% PRDV: 2%	0.9 (0.4)	3.6 (0.6)	0.04 (0.02)	45
Nigeria	1440	Cert: 73% Personal: 23% Credit: 4%	BUS: 25% CS: 22% PRDV: 8%	1.3 (0.7)	4.9 (0.9)	0.08 (0.02)	29
Egypt	1244	Cert: 65% Personal: 27% Credit: 8%	BUS: 18% CS: 18% PRDV: 6%	1.2 (0.5)	7.0 (0.8)	0.05 (0.01)	29
Other	10447	Cert: 55% Personal: 36% Credit: 9%	BUS: 23% CS: 21% PRDV: 6%	1.3 (0.6)	7.9 (1.0)	0.09 (0.03)	34
Blank	9109	Cert: 34% Personal: 48% Credit: 17%	BUS: 17% CS: 16% PRDV: 5%	1.1 (0.6)	5.4 (0.9)	0.08 (0.03)	34

Our top countries in terms of student location can be grouped into 3 categories that have unique characteristics: the United States, India, and Africa. The United States is the only major country where the pursuit of credit is a primary motivator of students (37%). This motivation represents less than 10% of students in all other countries. American students tend to enroll in the most courses and complete the most activities, but they are not particularly successful at completing those courses.

In the rest of the world, including both India and Africa, certification is far and away the main motivation of students. The difference between India and Africa lies mainly in the course selection (India is disproportionately more interested in computer science and professional development courses), as well as in the engagement with and eventual completion of courses (the average Indian students completes courses more than 2 times more often than students from anywhere else in the world, including the United States). Africa, in contrast, appears to be a very natural pool of Saylor students, but these students do not have a high level of engagement with our site.

Points worth considering:

- African students do not have a high level of engagement but do appear to come to our site naturally. We may want to consider them as low-activity organic traffic and limit advertising to them.
- Indian students, in contrast, have high engagement and very high completion rates, even compared to American students. These students want certificates, particularly in computer science and professional development.
- American students are the only group who want credit options more than certificates and are willing to put in time in courses (43 days on average to complete).

APPENDIX

The following is the MySQL query used to gather the data from our Moodle database.

```
Select s0.user_id, s0.domain, s0.city, s0.country, s0.language,
s0.first_accessed, s0.last_ip, s5.reason_for_joining,
s1.enrolled_courses, coalesce(s1.number_enrolled_courses,0)
number_enrolled_courses, s1.number_possible_credit_enrollments,
coalesce(s2.activities_completed,0) activities_completed,
coalesce(s3.quizzes_attempted,0) quizzes_attempted,
s4.completed_courses, coalesce(s4.number_completed_courses,0)
number_completed_courses, coalesce(s4.number_credit_completions,0)
number_credit_courses_completed, s4.total_sec_to_complete
From
    (Select u.id user_id, substring_index(u.email, '@', -1) domain,
u.city city, u.country country, u.lang language,
from_unixtime(u.firstaccess) first_accessed, u.lastip last_ip
    From mdl_user u
    where u.confirmed =1 and year(from_unixtime(u.firstaccess)) =
2018
```

```
    order by u.id desc
    limit 40000) s0
Left join
    (Select u.id user_id,
    group_concat(DISTINCT c.shortname ORDER BY c.shortname desc
SEPARATOR ", ") enrolled_courses, count(c.shortname)
number_enrolled_courses, count(cc.timecompleted)
number_completed_courses, sum(case when c.shortname = "BIO101" or
c.shortname = "BUS101" or c.shortname = "BUS103" or c.shortname =
"BUS105" or c.shortname = "BUS202" or c.shortname = "BUS203" or
c.shortname = "BUS204" or c.shortname = "BUS205" or c.shortname =
"BUS206" or c.shortname = "BUS208" or c.shortname = "BUS209" or
c.shortname = "BUS210" or c.shortname = "BUS303" or c.shortname =
"CHEM101" or c.shortname = "COMM001" or c.shortname = "CS101" or
c.shortname = "CS102" or c.shortname = "CS302" or c.shortname =
"CS402"or c.shortname = "ECON101" or c.shortname = "ECON102" or
c.shortname = "ENVS203" or c.shortname = "MA001" or c.shortname =
"MA005" or c.shortname = "MA121" or c.shortname = "PHIL103" or
c.shortname = "PHYS101" or c.shortname = "PHYS102" or c.shortname =
"POLSC101" or c.shortname = "POLSC201" or c.shortname = "POLSC221" or
c.shortname = "PSYCH101" or c.shortname = "SOC101"  then 1 else 0
end) number_possible_credit_enrollments
    From mdl_user u
    left join mdl_course_completions cc on cc.userid = u.id
    join mdl_course c on c.id = cc.course
    where cc.course != 370 and cc.course != 414 and
year(from_unixtime(u.firstaccess)) = 2018
    group by u.id
    order by u.id desc
    limit 40000) s1 on s1.user_id = s0.user_id
left join
    (Select u.id user_id, count(*) activities_completed
    from mdl_user u
    join mdl_course_modules_completion cmc on cmc.userid = u.id
    where year(from_unixtime(u.firstaccess)) = 2018
    group by u.id
    order by u.id desc
    limit 40000) s2 on s2.user_id = s1.user_id
```

```
left join
    (Select u.id user_id, count(distinct(qa.quiz)) quizzes_attempted
    from mdl_user u
    join mdl_quiz_attempts qa on qa.userid = u.id
    where year(from_unixtime(u.firstaccess)) = 2018
    group by u.id
    order by u.id desc
    limit 40000) s3 on s3.user_id = s1.user_id
Left join
    (Select u.id user_id, group_concat(DISTINCT c.shortname ORDER BY
c.shortname desc SEPARATOR ", ") completed_courses,
count(c.shortname) number_completed_courses, sum(case when
cc.timeenrolled > 1483228800 then cc.timecompleted-cc.timeenrolled
else 2678000 end) total_sec_to_complete, sum(case when c.shortname =
"BIO101" or c.shortname = "BUS101" or c.shortname = "BUS103" or
c.shortname = "BUS105" or c.shortname = "BUS202" or c.shortname =
"BUS203" or c.shortname = "BUS204" or c.shortname = "BUS205" or
c.shortname = "BUS206" or c.shortname = "BUS208" or c.shortname =
"BUS209" or c.shortname = "BUS210" or c.shortname = "BUS303" or
c.shortname = "CHEM101" or c.shortname = "COMM001" or c.shortname =
"CS101" or c.shortname = "CS102" or c.shortname = "CS302" or
c.shortname = "CS402"or c.shortname = "ECON101" or c.shortname =
"ECON102" or c.shortname = "ENVS203" or c.shortname = "MA001" or
c.shortname = "MA005" or c.shortname = "MA121" or c.shortname =
"PHIL103" or c.shortname = "PHYS101" or c.shortname = "PHYS102" or
c.shortname = "POLSC101" or c.shortname = "POLSC201" or c.shortname =
"POLSC221" or c.shortname = "PSYCH101" or c.shortname = "SOC101"
then 1 else 0 end) number_credit_completions
    From mdl_user u
    left join mdl_course_completions cc on cc.userid = u.id
    join mdl_course c on c.id = cc.course
    where cc.timecompleted is not null and cc.course != 370 and
cc.course != 414 and  year(from_unixtime(u.firstaccess)) = 2018
    group by u.id
    order by u.id desc
    limit 40000) s4 on s4.user_id = s1.user_id
Left join
    (Select u.id user_id, subq.data reason_for_joining
```

```
    From mdl_user u
    join (select * from mdl_user_info_data uid where uid.fieldid =
64) subq on subq.userid = u.id
    where year(from_unixtime(u.firstaccess)) = 2018
    order by u.id desc
    limit 40000) s5 on s0.user_id = s5.user_id
order by s0.user_id desc
```