

Проект e-learning: вариант 2

Anastasiia Sibiriakova (a_sibiriakova)

Поток 15.08

Что считаем курсом

Курсом в данном исследовании будем считать связку "предмет в семестре". Предмет с одинаковым именем, но в разных семестрах будем считать разными курсами.

Почему так:

1. Средняя длительность предмета в семестре 256 дней - вполне достойный срок, чтобы считать такой предмет курсом.
2. Итоговым тестом по каждому предмету является финальный экзамен. То есть у нас нет предметов, где достаточно зачета для окончания. Это тоже говорит в пользу того, чтобы считать предмет курсом.
3. На предмете в среднем начинают обучение 1340 студентов (учитываем только тех, кто либо завершил предмет, либо отменил регистрацию уже после старта обучения), весомое количество студентов дает нам еще одно основание считать предмет курсом.
4. От семестра к семестру содержание предмета перерабатывается/изменяется.

1. Сколько студентов успешно сдали только один курс?

(Успешная сдача — это зачёт по курсу на экзамене) (7 баллов).

Я бы уточнила формулировку вопроса: здесь под "успешно сдали только один курс" будем понимать "закончили только один курс, успешно сдав все экзамены предполагаемые программой курса"



ОТВЕТ: 2709

ВЫВОДЫ:

Не смотря на то, что у нас большое количество курсов (22) и данные за 4 семестра, очень маленькое количество студентов успешно завершило только 1 курс, и нет студентов, которые успешно завершили 2 и более курсов.

2. Выяви самый сложный и самый простой экзамен:

найди курсы и экзамены в рамках курса, которые обладают самой низкой и самой высокой завершаемостью*. (5 баллов)

*завершаемость = кол-во успешных экзаменов / кол-во всех попыток сдать экзамен



ОТВЕТ:

Самый сложный: экзамен на курсе DDD в семестре 2013B 84% завершаемость

Самый простой: экзамен на курсе DDD в семестре 2014B 93% завершаемость

ВЫВОДЫ:

Из предобработки помним, что количество тестов в этих двух курсах разное:

- DDD в семестре 2013B: CMA-7 Exam-1 TMA-6
- DDD в семестре 2014B: CMA-0 Exam-1 TMA-6

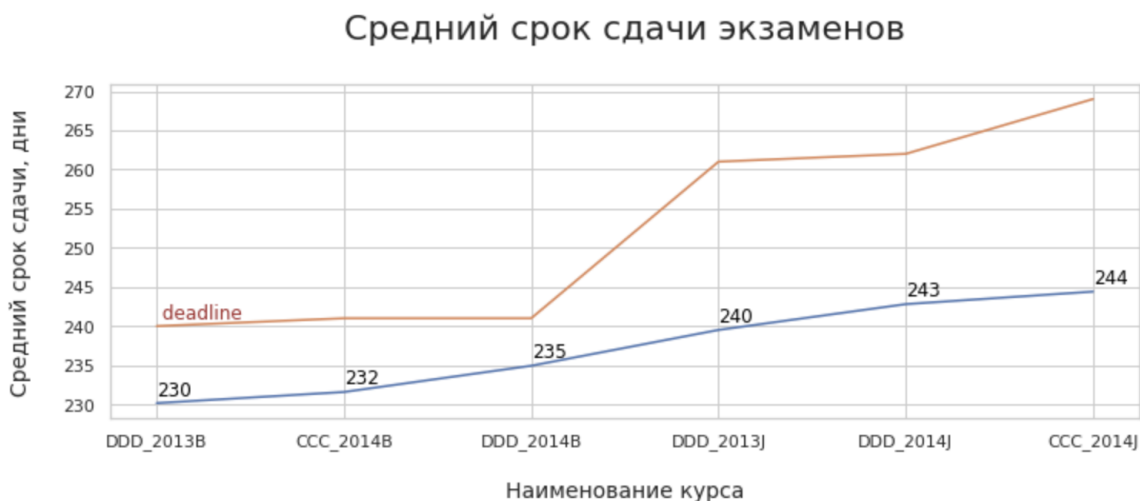
Получается, что на курсе, где не было промежуточных тестов с компьютерной оценкой, сдать экзамен оказалось проще.

Возможно в семестре 2013B была более сложная программа курса с более сложным экзаменом, поэтому предусмотрели промежуточные тесты с компьютерной оценкой, чтобы студенты лучше усвоили материал. А потом программу упростили и не стали делать промежуточных тестов с компьютерной оценкой, потому что экзамен достаточно простой.

3. По каждому предмету определи средний срок сдачи экзаменов

(под сдачей понимаем последнее успешное прохождение экзамена студентом). (5 баллов)

Как мы помним курс у нас состоит из одного предмета в семестра. Предмет с одинаковым именем, но в разных семестрах будем считать разными курсами. Поэтому считать средний срок сдачи экзаменов будем в разрезе курсов, тем более что у нас есть курсы, где предусмотрено несколько экзаменов.



ОТВЕТ:

Средний срок сдачи экзамена:

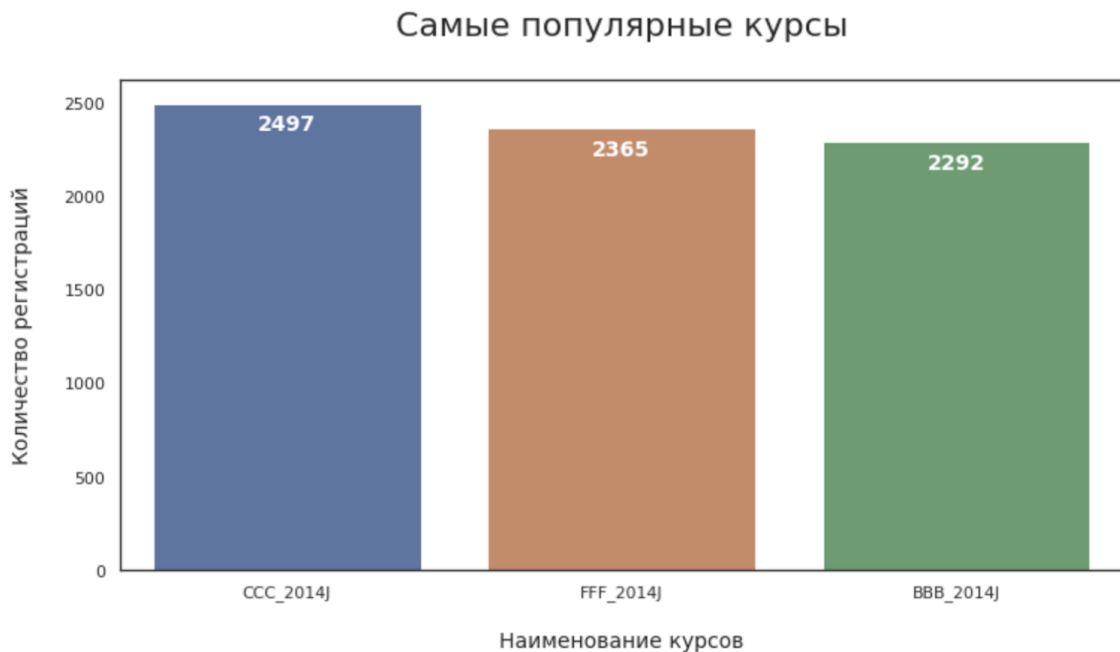
- курс DDD в семестре 2013B — 230 дней
- курс CCC в семестре 2014B — 232 дней
- курс DDD в семестре 2014B — 235 дней
- курс DDD в семестре 2013J — 240 дней
- курс DDD в семестре 2014J — 243 дней
- курс CCC в семестре 2014J — 244 дней

ВЫВОДЫ:

Студенты в среднем укладываются в дэдлайны по сдаче экзаменов

4. Выяви самые популярные предметы (ТОП-3) по количеству регистраций на них. А также предметы с самым большим оттоком (ТОП-3). (8 баллов)

Как мы помним курс у нас состоит из одного предмета в семестра. Предмет с одинаковым именем, но в разных семестрах будем считать разными курсами. Поэтому считать ТОП-3 будем в разрезе курсов.



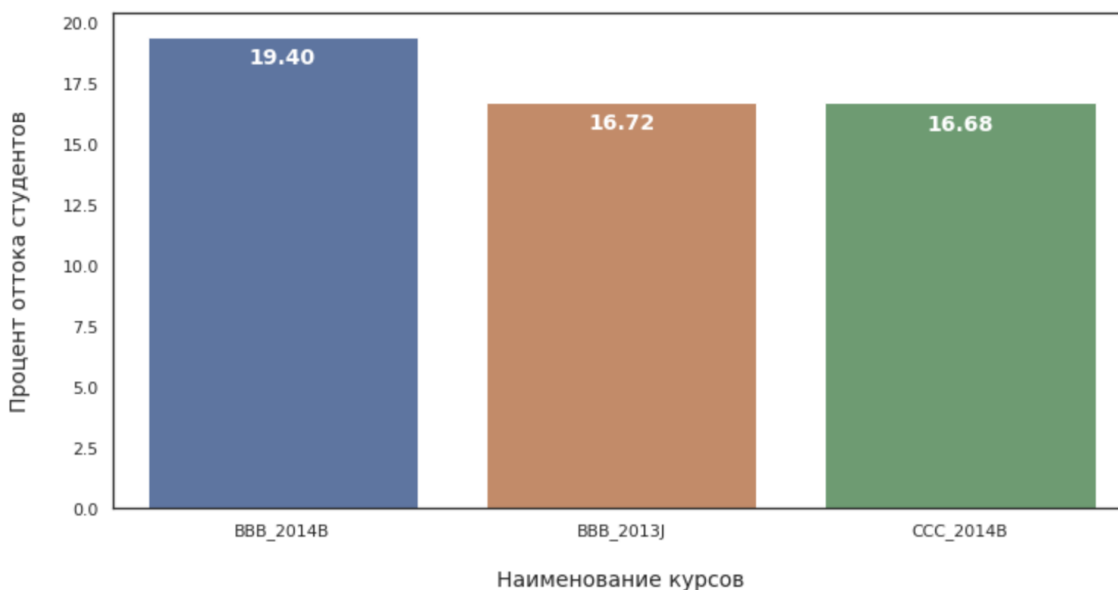
Отток

Что можно считать **оттоком**:

- когда студент отменил регистрацию на курс до начала обучения
- либо в день начала обучения
- либо до первой проверочной работы. Посмотрим минимальную дату теста в таблице `assessments_df` (12 дней с начала семестра)

Отток - это когда студент либо отменил регистрацию до начала обучения, либо в первые 11 дней после старта обучения

Курсы с самым большим оттоком



ОТВЕТ:

Самые популярные курсы:

- курс CCC в семестре 2014J — 2497 регистраций
- курс FFF в семестре 2014J — 2365 регистраций
- курс BBB в семестре 2014J — 2292 регистраций

Курсы с самым большим оттоком:

- курс BBB в семестре 2014B — 19.40% оттока
- курс BBB в семестре 2013J — 16.72% оттока
- курс CCC в семестре 2014B — 16.68% оттока

ВЫВОДЫ:

С самых популярных курсов не очень большой отток студентов, так как самые популярные курсы не входят в ТОП-3 курсов по оттоку.

Сам процент оттока мне не кажется драматически высоким, в около 20% студентов максимум отменяют регистрацию до начала либо в первые дни старта курса.

5. Используя pandas, в период с начала 2013 по конец 2014 выяви семестр с самой низкой завершаемостью курсов и самыми долгими средними сроками сдачи курсов. (10 баллов)

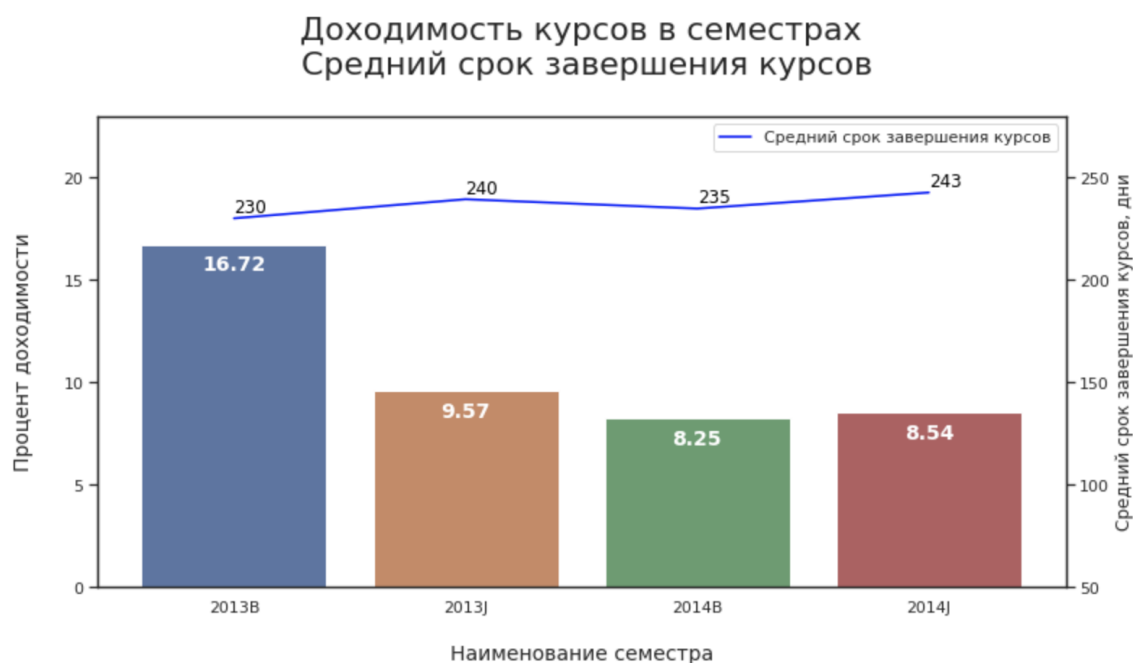
Под "завершаемостью курсов" здесь будем понимать их "доходимость" - сколько процентов обучающихся прошли до конца, сдали экзамен.

COR (Completion Rate) - доходимость

*количество студентов сдавших экзамен по курсу / количество студентов, начавших обучение на курсе

Под "сроками сдачи курсов" будем понимать дату сдачи экзаменов по курсу.

Нам в данном случае не так важна оценка на экзамене как сам факт, что студент доучился до момента сдачи экзамена и хоть как-то, но сдал.



ОТВЕТ:

Семестр с самой низкой доходимостью курсов: 2014B — 8.25% доходимость

Семестр с самым долгими средними сроками завершения курсов: 2014J — 243 дня

ВЫВОДЫ:

У нас очень низкая доходимость курсов, доходимость в 2014ом году снизилась в 2 раза от максимального значения в 2013ом, что-то изменилось в программах курсов или в организации учебного процесса, что так сильно повлияло на доходимость.

6. Часто для качественного анализа аудитории использую подходы, основанные на сегментации. Используя python, построй адаптированные RFM-кластеры студентов, чтобы качественно оценить свою аудиторию.

В адаптированной кластеризации можешь выбрать следующие метрики: R - среднее время сдачи одного экзамена, F - завершаемость курсов, M - среднее количество баллов, получаемое за экзамен. Подробно опиши, как ты создавал кластеры. Для каждого RFM-сегмента построй границы метрик recency, frequency и monetary для интерпретации этих кластеров. Описание подхода можно найти тут. (35 баллов)

Предлагаемый Василием вариант сегментирования студентов:

- R - среднее время сдачи одного экзамена,
- F - завершаемость курсов,
- M - среднее количество баллов, получаемое за экзамен

Не согласна с такими метриками. Для описания студентов чаще всего используют такие описательные метрики как: **трудолюбие и способности**.

К трудолюбию можно отнести:

- дисциплинированность - как вообще студент учится в течении курса
- целеустремленность - как хорошо студент достигает поставленных целей (с разрезе обучения: насколько часто завершает начатые курсы)

К способностям:

- средний балл за все оцениваемые работы в течении курса. По всем видам проверочных тестов, а не только экзаменов.

Таким образом, буду использовать следующий вариант сегментирования студентов:

- R - дисциплинированность - сколько тестовых работ студент сдал / сколько предусматривается программой курса (среднее по всем курсам). Без оценивания на сколько баллов сданы тесты, важно что студент сдавал тесты, а баллы будем смотреть в метрике M - способности
- F - доходимость - сколько курсов закончено / сколько курсов начато. Без оценивания на сколько баллов сдан экзамен, важно что студент дошел до конца обучения, а баллы будем смотреть в метрике M - способности
- M - способности - среднее количество баллов, получаемое за тесты во время обучения (все виды тестов)

дисциплинированность (recency):

- 1 — не сдал ни одного промежуточного теста
- 2 — сдал несколько, но не все
- 3 — сдал все промежуточные тесты

доходимость (frequency):

- 1 — не завершил ни один начатый курс
- 2 — завершил несколько, но не все
- 3 — завершил все начатые курсы

среднее количество баллов (monetary):

- 1 — невысокие баллы
- 2 — средние баллы
- 3 — высокие баллы

Описание сегментов

Мне понравилось описание типов студентов от преподавателя АлтГУ Людмила Комиссарова вот из этой статьи

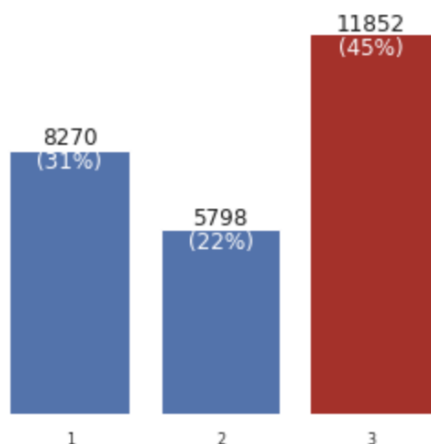
<https://rekportal.ru/persons/8-tipov-studentov-ot-zvezdy-do-edinorozhki/>

Решила воспользоваться им, в статье как раз используются такие описательных характеристики студентов как: способности, трудолюбие, мотивация и самооценка.

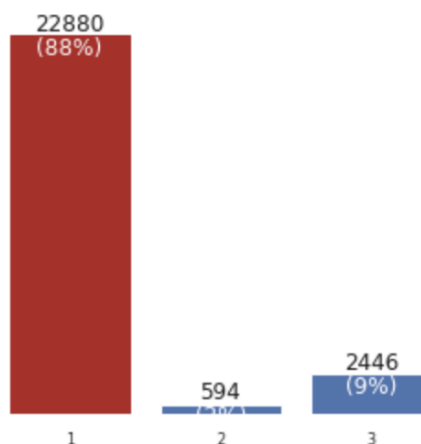
Сегмент	Характеристика	Краткое название	Описание
Низкая дисциплина			
111	троечник с низкой дисциплиной и мотивацией	Потребитель	Неизвестные способности. Высокие амбиции. Отсутствие трудолюбия. Неадекватно высокая самооценка.
112	хорошист с низкой дисциплиной и мотивацией	Талантливый лентяй	Блестящие способности в потенциале. Кривая мотивация. Импульсивная трудоспособность. Прыгающая самооценка. Заядлый прокрастинатор.
113	отличник с низкой дисциплиной и мотивацией	Талантливый лентяй	Блестящие способности в потенциале. Кривая мотивация. Импульсивная трудоспособность. Прыгающая самооценка. Заядлый прокрастинатор.
121	троечник с низкой дисциплиной и средней мотивацией	Талантливый лентяй	Блестящие способности в потенциале. Кривая мотивация. Импульсивная трудоспособность. Прыгающая самооценка. Заядлый прокрастинатор.
122	хорошист с низкой дисциплиной и средней мотивацией	Маркетолог	Средние или низкие способности. Высокие или средние амбиции. Нормальное или низкое трудолюбие. Адекватная самооценка.
123	отличник с низкой дисциплиной и средней мотивацией	Принц / принцесса	Высокие или средние способности. Высокие амбиции. Низкое трудолюбие. Высокая самооценка.
131	троечник с низкой дисциплиной и высокой мотивацией	Крутышка	Низкие способности. Огромные амбиции. Высокое трудолюбие. Самооценка зашкаливает.
132	хорошист с низкой дисциплиной и высокой мотивацией	Маркетолог	Средние или низкие способности. Высокие или средние амбиции. Нормальное или низкое трудолюбие. Адекватная самооценка.
133	отличник с низкой дисциплиной и высокой мотивацией	Принц / принцесса	Высокие или средние способности. Высокие амбиции. Низкое трудолюбие. Высокая самооценка.
Средняя дисциплина			
211	троечник со средней дисциплиной и низкой мотивацией	Талантливый лентяй	Блестящие способности в потенциале. Кривая мотивация. Импульсивная трудоспособность. Прыгающая самооценка. Заядлый прокрастинатор.
212	хорошист со средней дисциплиной и низкой мотивацией	Маркетолог	Средние или низкие способности. Высокие или средние амбиции. Нормальное или низкое трудолюбие. Адекватная самооценка.

Сегмент	Характеристика	Краткое название	Описание
213	отличник со средней дисциплиной и низкой мотивацией	Принц / принцесса	Высокие или средние способности. Высокие амбиции. Низкое трудолюбие. Высокая самооценка.
221	троечник со средней дисциплиной и средней мотивацией	Крутышка	Низкие способности. Огромные амбиции. Высокое трудолюбие. Самооценка зашкаливает.
222	хорошист со средней дисциплиной и средней мотивацией	Маркетолог	Средние или низкие способности. Высокие или средние амбиции. Нормальное или низкое трудолюбие. Адекватная самооценка.
223	отличник со средней дисциплиной и средней мотивацией	Принц / принцесса	Высокие или средние способности. Высокие амбиции. Низкое трудолюбие. Высокая самооценка.
231	троечник со средней дисциплиной и высокой мотивацией	Крутышка	Низкие способности. Огромные амбиции. Высокое трудолюбие. Самооценка зашкаливает.
232	хорошист со средней дисциплиной и высокой мотивацией	Ботан	Блестящие способности. Низкие притязания. Большое трудолюбие. Самооценка средняя или ниже среднего. Самый перспективный студент.
233	отличник со средней дисциплиной и высокой мотивацией	Ботан	Блестящие способности. Низкие притязания. Большое трудолюбие. Самооценка средняя или ниже среднего. Самый перспективный студент.
Высокая дисциплина			
311	троечник с высокой дисциплиной и низкой мотивацией	Крутышка	Низкие способности. Огромные амбиции. Высокое трудолюбие. Самооценка зашкаливает.
312	хорошист с высокой дисциплиной и низкой мотивацией	Маркетолог	Средние или низкие способности. Высокие или средние амбиции. Нормальное или низкое трудолюбие. Адекватная самооценка.
313	отличник с высокой дисциплиной и низкой мотивацией	Принц / принцесса	Высокие или средние способности. Высокие амбиции. Низкое трудолюбие. Высокая самооценка.
321	троечник с высокой дисциплиной и средней мотивацией	Крутышка	Низкие способности. Огромные амбиции. Высокое трудолюбие. Самооценка зашкаливает.
322	хорошист с высокой дисциплиной и средней мотивацией	Ботан	Блестящие способности. Низкие притязания. Большое трудолюбие. Самооценка средняя или ниже среднего. Самый перспективный студент.
323	отличник с высокой дисциплиной и средней мотивацией	Ботан	Блестящие способности. Низкие притязания. Большое трудолюбие. Самооценка средняя или ниже среднего. Самый перспективный студент.
331	троечник с высокой дисциплиной и высокой мотивацией	Крутышка	Низкие способности. Огромные амбиции. Высокое трудолюбие. Самооценка зашкаливает.
332	хорошист с высокой дисциплиной и высокой мотивацией	Ботан	Блестящие способности. Низкие притязания. Большое трудолюбие. Самооценка средняя или ниже среднего. Самый перспективный студент.
333	отличник с высокой дисциплиной и высокой мотивацией	Звезда	Блестящие способности. Высокие амбиции. Большое трудолюбие. Адекватная самооценка. Мечта любого преподавателя.

Распределение "Дисциплинированность"



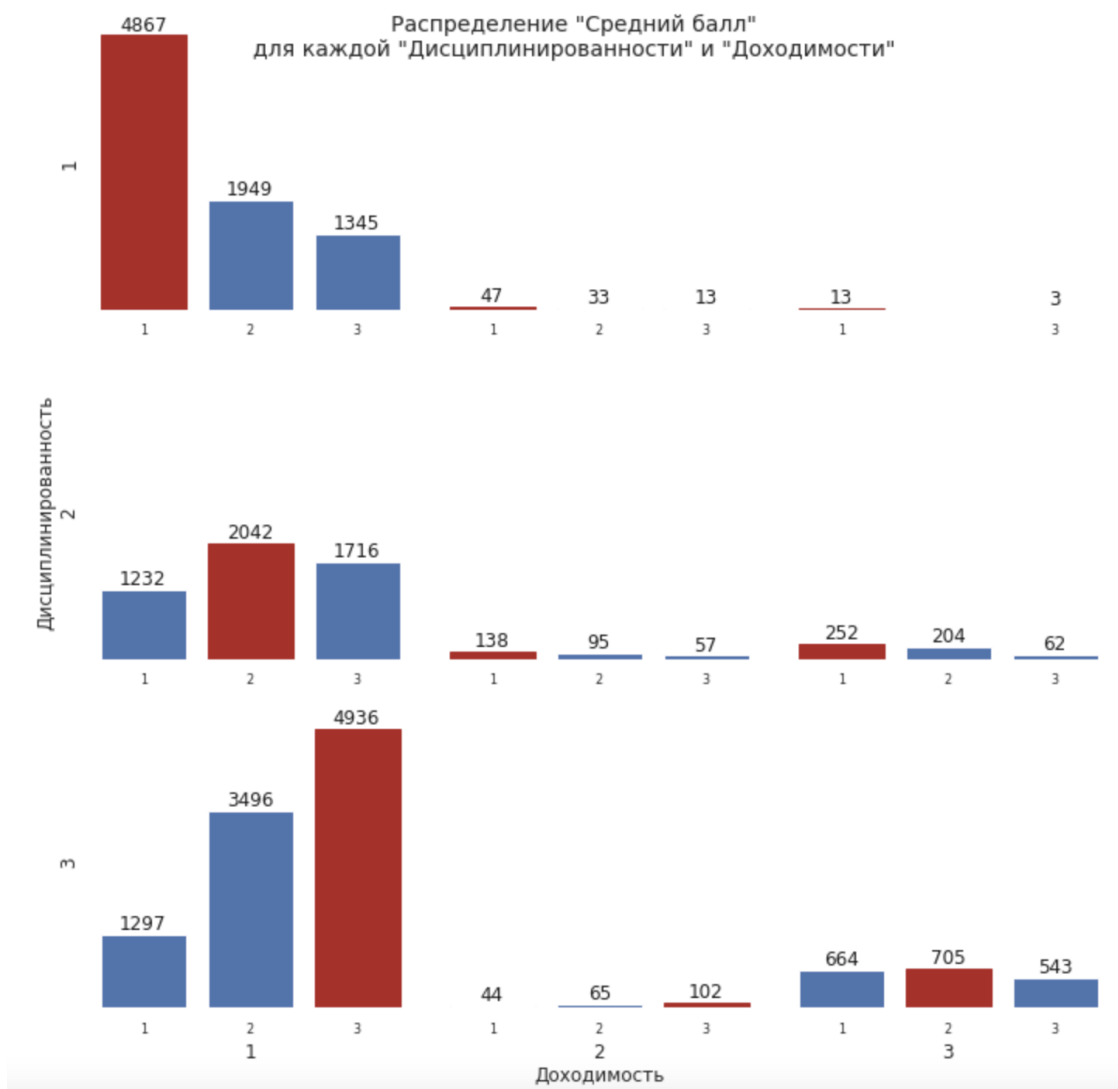
Распределение "Доходимость"



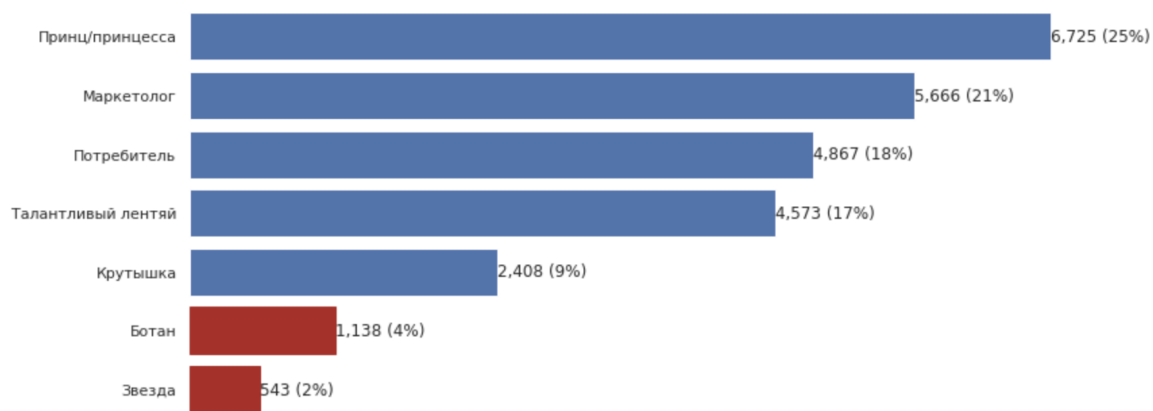
ВЫВОДЫ:

У нас 45% студентов, которые сдали все промежуточные тесты, и только 9%, которые сдали итоговый экзамен. То есть с дисциплиной у студентов порядок, но нет особой мотивации сдавать итоговый экзамен. Возможно нужно ввести дополнительный стимул для сдачи экзамена, например, выдачу сертификата о прохождении курса или добавления в отдельное сообщество выпускников с дополнительными плюшками.

Распределение "Средний балл" для каждой "Дисциплинированности" и "Доходимости"



Распределение количества студентов по сегментам



ВЫВОДЫ:

Сегмент	Что делать
Принц / принцесса	Самый многочисленный сегмент - 25% Высокие или средние способности. Высокие амбиции. Низкое трудолюбие. Высокая самооценка. Чтобы повысить их доходимость, нужно таким студентам уделять больше внимания, высылать напоминанлки и подбадривающие сообщения, для поддержания их самооценки.
Маркетолог	Второй по численности сегмент - 21% Средние или низкие способности. Высокие или средние амбиции. Нормальное или низкое трудолюбие. Адекватная самооценка. Таким студентам во время обучения можно рассылать напоминанлки о дэдлайнах и минимальных порогах сдачи.
Потребитель	Третий по численности сегмент - 18% Неадекватно высокая самооценка. Неизвестные способности. Отсутствие трудолюбия. За таких студентов можно не бороться, они всегда найдут к чему придраться и еще и негативные отзывы оставят.
Талантливый лентяй	Четвертый по численности сегмент - 17% Блестящие способности в потенциале. Кривая мотивация. Импульсивная трудоспособность. Прыгающая самооценка. Заядлый прокрастинатор. Таким студентам можно отправить истории успеха выпускников, чтобы повысить мотивацию.
Крутышка	Пятый по численности сегмент - 9% Низкие способности. Огромные амбиции. Высокое трудолюбие. Самооценка зашкаливает. Таким студентам можно рассылать письма с ссылками на доп материалы для изучения, чтобы повысить их успеваемость.
Ботан	Шестой по численности сегмент - 4% Блестящие способности. Низкие притязания. Большое трудолюбие. Самооценка средняя или ниже среднего. Самые перспективные студенты.Таким студентам можно высылать анонсы новых курсов.Особые предложения по оплате.
Звезда	Ожидаемо самый малочисленный сегмент - 2% Блестящие способности. Высокие амбиции. Большое трудолюбие. Адекватная самооценка. Мечта любого преподавателя, гордость и слава альмы-матери. Таким студентам можно высылать анонсы новых курсов.Особые предложения по оплате.