

Исследование рынка репетиторов по математике (г. Москва)

Цели исследования:

- Выделить и описать основные группы репетиторов по математике, работающих в г. Москве
- Определить основные факторы, влияющие на цену занятия по математике

Данные:

Данные для исследования представляют собой характеристики репетиторов по математике в городе Москва.

- ID_repetitor — идентификационный номер репетитора на портале
- price — цена 60 минут занятия
- experience — преподавательский опыт
- age — возраст репетитора
- reviews — количество отзывов о репетиторе
- education — образование репетитора
- video — наличие/отсутствие видеопрезентации о репетиторе
- photo — наличие/отсутствие фотографии репетитора
- diplomas — количество подтвержденных на портале дипломов
- students — количество подобранных учеников у данного репетитора
- years — количество лет после регистрации на портале
- last_activity — дата последней активности
- chapters — преподаваемые репетитором дисциплины
- additions — дополнительные преподаваемые дисциплины (ОГЭ (ГИА), ЕГЭ, подготовка к олимпиадам)
- home — район, в котором репетитор принимает у себя
- out — районы, в который репетитор выезжает
- male — если репетитор мужчина, то male = 1, если женщина, то male = 0

Данные собраны 24 марта 2023 года. Источник данных — портал repetit.ru

Создание дополнительных переменных:

Была проведена предобработка данных и добавлены следующие переменные:

- EGE = 1, если репетитор готовит к ЕГЭ по математике, иначе EGE = 0
- olympiads = 1, если репетитор готовит к олимпиадам по математике, иначе olympiads = 0

- $top_university = 1$, если репетитор закончил университет, входящий в топ-10 российских вузов с математической подготовкой (https://raex-rr.com/education/Russian_universities_by_directions/math/2020) + выпускники Московского педагогического государственного университета, иначе $top_university = 0$

- $students_per_year$ — количество учеников, подобранных репетитором за один год

Затем были удалены все неактивные репетиторы (не появлялись в сети больше полугода), а также репетиторы, которые за последние полгода не подобрали ни одного ученика.

1. Описательные статистики и графики распределений

В таблице 1 показаны описательные статистики выборки.

	число репетиторов	среднее	стандартное отклонение	min	25%	медиана	75%	max
ID_repetitor	3565.0	143825.976438	90824.172356	102.000000	62350.000000	139998.00	222159.0	290351.0
price	3565.0	1319.453015	628.776678	300.000000	1000.000000	1200.00	1500.0	9350.0
experience	3565.0	13.228612	11.659097	0.000000	4.000000	10.00	19.0	59.0
age	3565.0	37.116410	16.490761	18.000000	24.000000	31.00	50.0	86.0
reviews	3565.0	8.547546	14.374297	0.000000	1.000000	3.00	10.0	187.0
video	3565.0	0.044039	0.205211	0.000000	0.000000	0.00	0.0	1.0
photo	3565.0	0.930154	0.254923	0.000000	1.000000	1.00	1.0	1.0
diplomas	3565.0	1.183170	2.271511	0.000000	0.000000	0.00	2.0	35.0
students	3565.0	13.478261	22.174112	0.000000	1.000000	5.00	16.0	299.0
years	3565.0	5.400748	4.144072	0.083333	2.000000	5.00	9.0	15.0
male	3565.0	0.416269	0.493008	0.000000	0.000000	0.00	1.0	1.0
EGE	3565.0	0.647686	0.477758	0.000000	0.000000	1.00	1.0	1.0
olimpiads	3565.0	0.223843	0.416876	0.000000	0.000000	0.00	0.0	1.0
higher_math	3565.0	0.358766	0.479706	0.000000	0.000000	0.00	1.0	1.0
top_university	3565.0	0.548387	0.497723	0.000000	0.000000	1.00	1.0	1.0
students_per_year	3565.0	2.653330	3.823285	0.000000	0.363636	1.25	3.5	65.5

Таблица 1 – Описательные статистики выборки

На рынке репетиторов в г. Москва действуют 3565 активных репетиторов по математике, работающих на платформе repetit.ru.

Медианный репетитор по математике — это женщина возрастом 31 год со стажем работы 10 лет, не являющаяся выпускницей топового университета, которая готовит к ЕГЭ, но не преподаёт высшую математику и не готовит к математическим олимпиадам. На её странице repetit.ru опубликована фотография (но не опубликована видеопрезентация),

подтверждено 0 дипломов, имеется 3 отзыва. Через портал она подбирает 1.25 учеников в год. Её занятие стоит 1200 рублей в час.

Средние значения большинства переменных превышают медианные значения, что говорит о положительной асимметрии выборки, то есть существуют репетиторы, характеристики которых существенно превышают характеристики других (например, необычайно высокая цена занятия, или очень почтенный возраст репетитора)

Графики распределения количественных переменных представлены на рисунках 1 – 3.

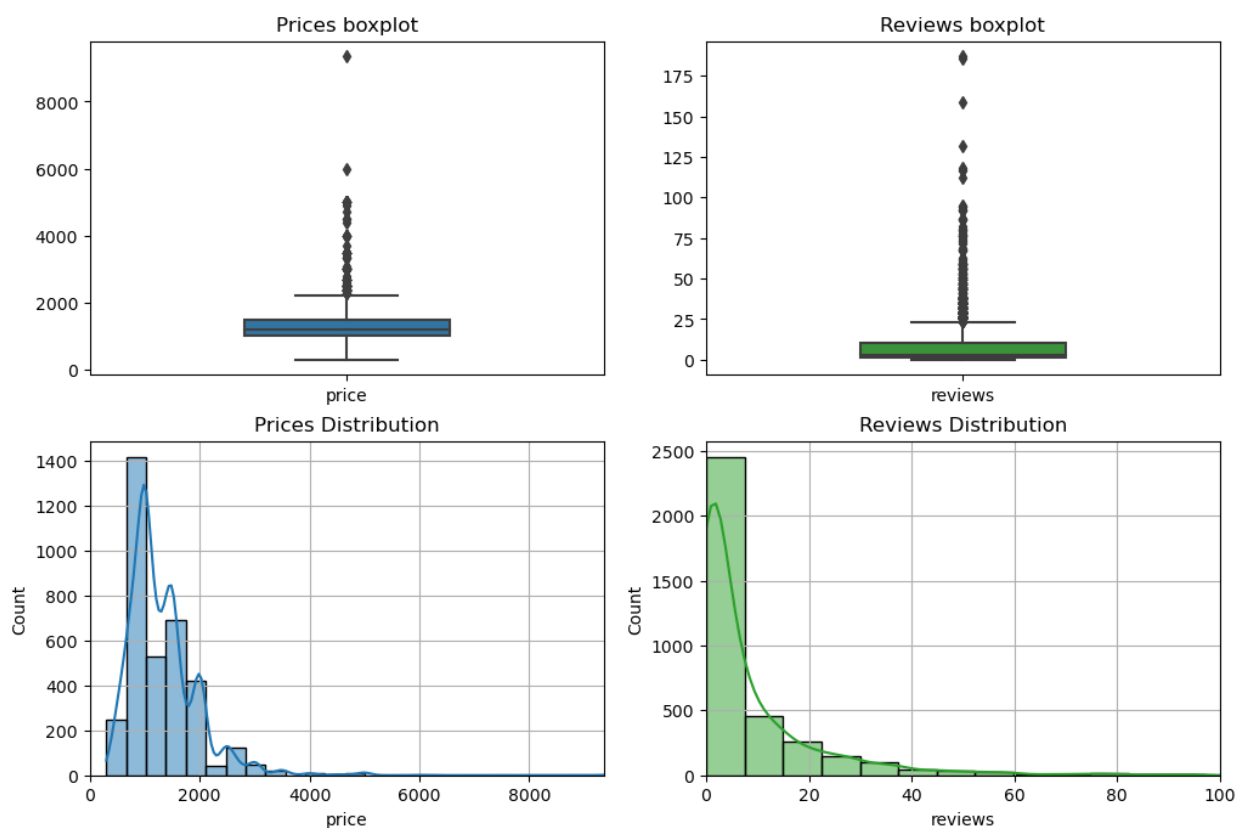


Рисунок 1 – Графики распределения переменных price и reviews

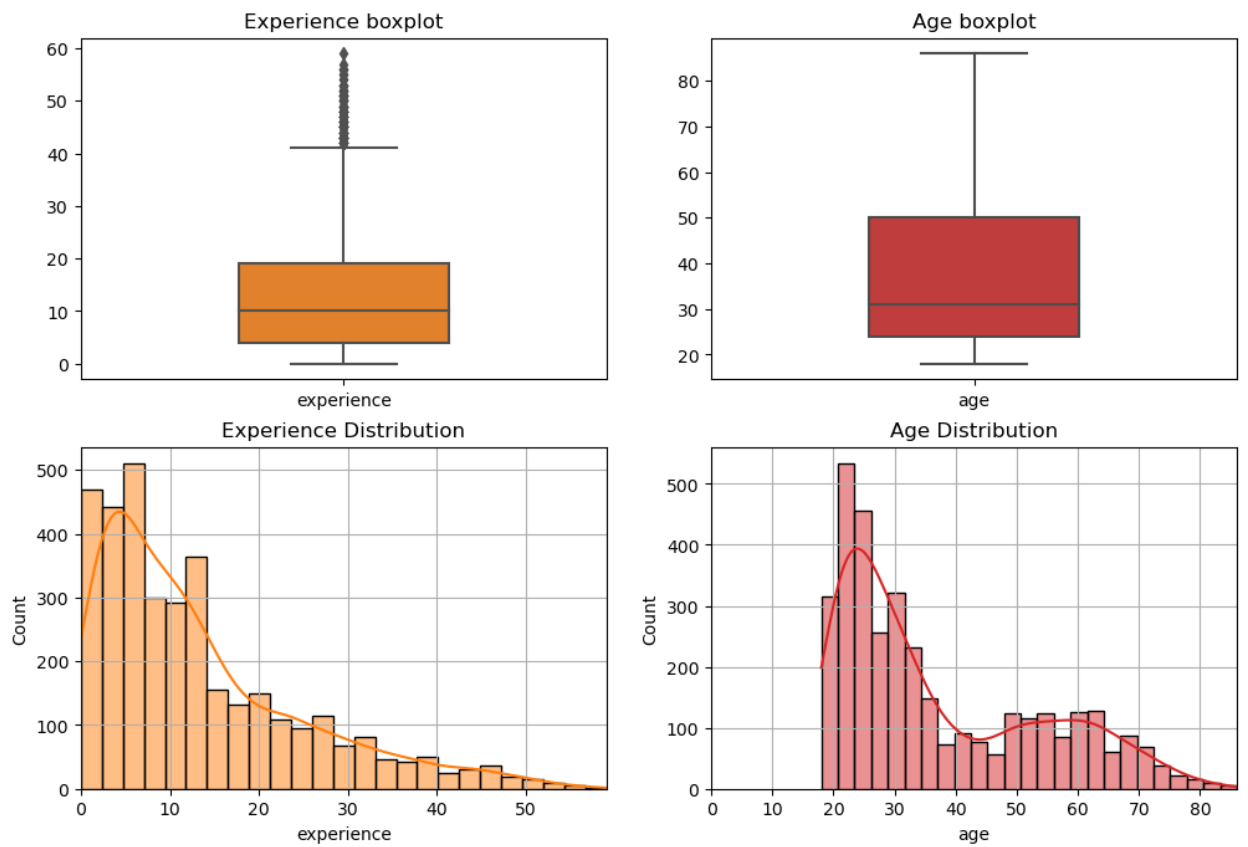


Рисунок 2 – Графики распределения переменных experience и age

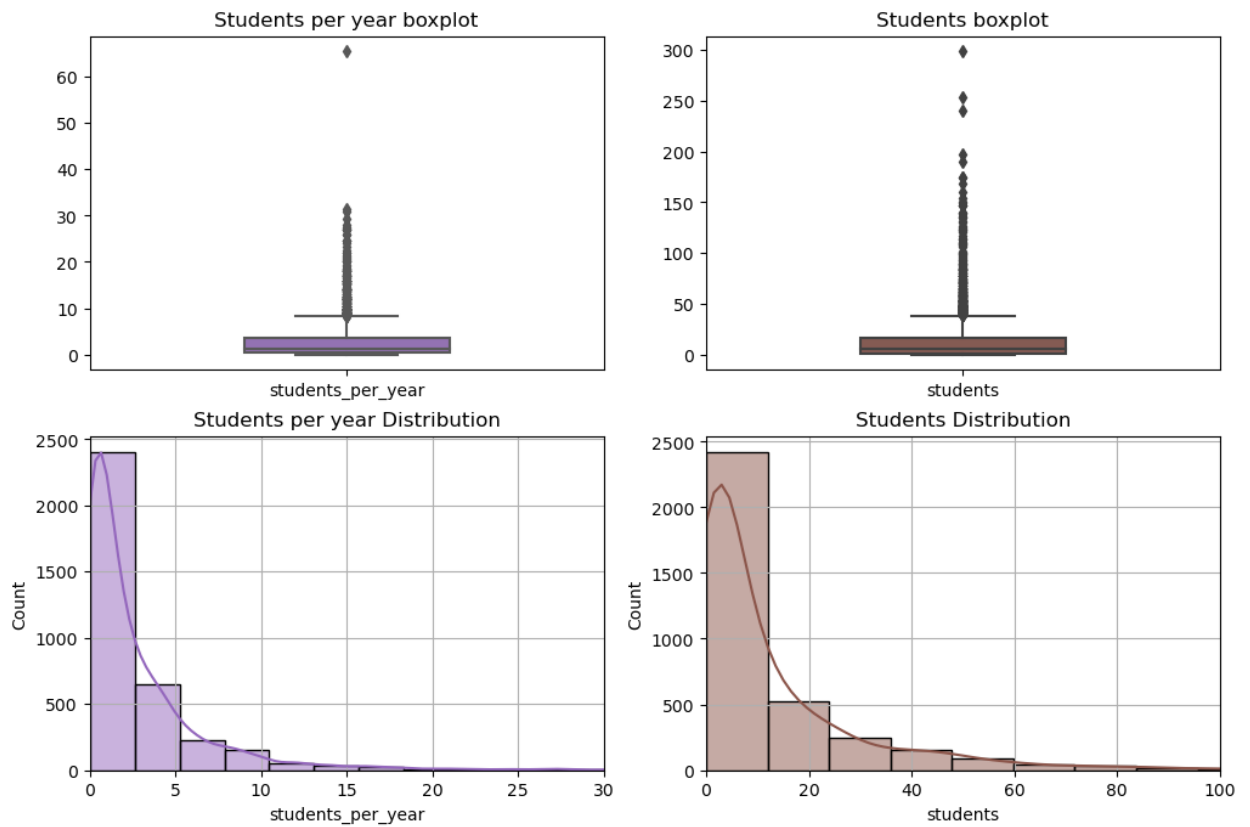


Рисунок 3 – Графики распределения переменных students_per_year и students

Практически все репетиторы преподают программу школьного курса (98% репетиторов), а также большинство репетиторов готовят к выпускным экзаменам ОГЭ

(82%) и ЕГЭ (64%). Достаточно много репетиторов занимаются такими разделами математики из старшей школы как алгебра (82%), геометрия (79%) и тригонометрия (20%). Из разделов высшей математики репетиторы наиболее часто преподают линейную алгебру (19%) и математический анализ (18%), наиболее редко топологию и теорию приближений (по 1%).

3. Распределение репетиторов по топовым университетам

На рисунке 4 изображено графическое распределение выпускников топовых вузов, преподающих математику в качестве репетитора.

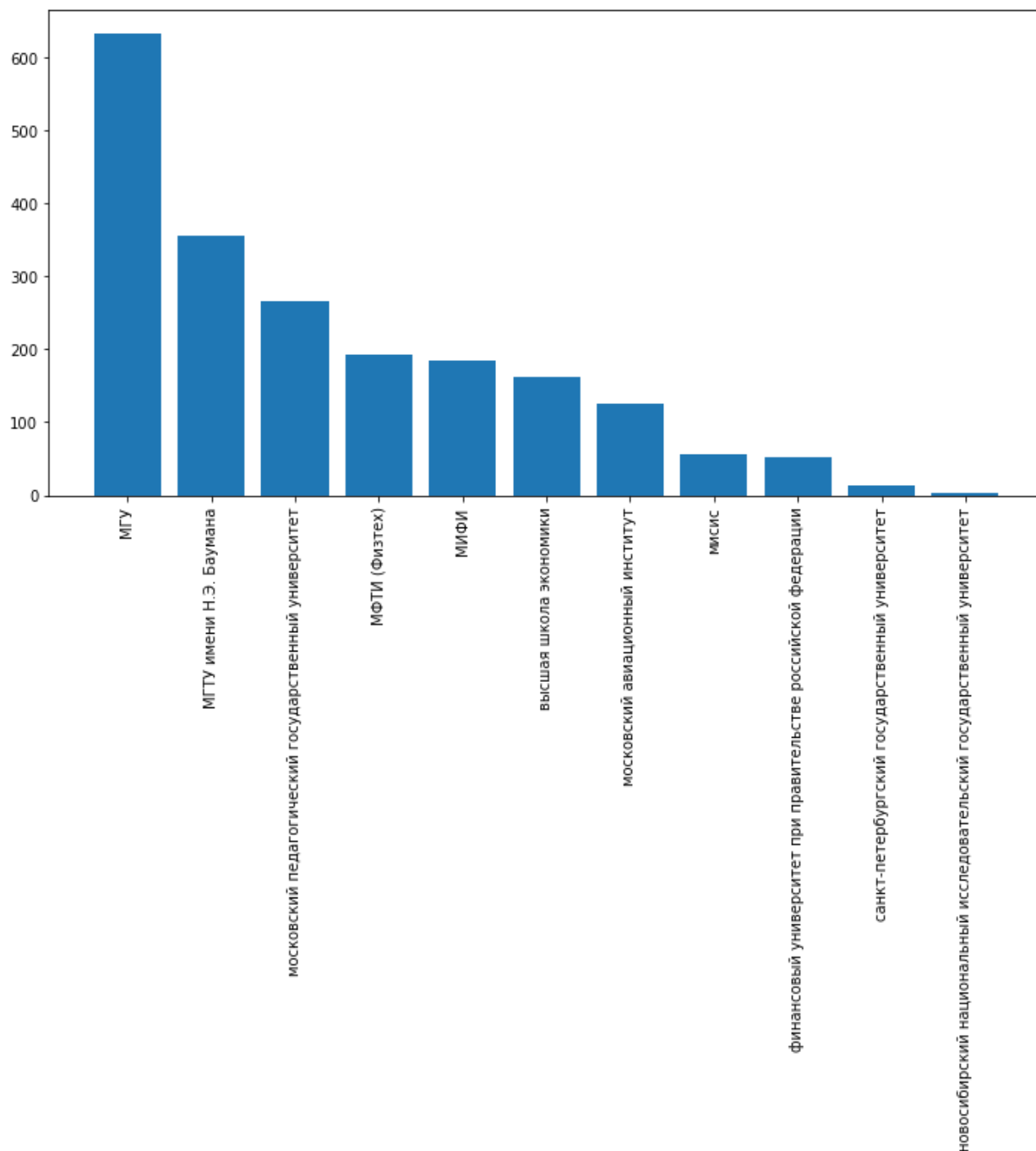


Рисунок 4 – Распределение репетиторов по топовым вузам

Наибольшее число репетиторов по математике получили образование в МГУ имени М.В. Ломоносова. Наименьшее число репетиторов из числа выпускников топовых вузов закончили СПбГУ и НГУ, что объясняется удаленностью этих университетов от Москвы.

4. Анализ ценообразования на рынке репетиторов по математике

На платформе repetit.ru репетиторы сами назначают цену на занятие исходя из собственных соображений. Однако, предполагается, что на выбираемую стоимость урока влияют некоторые объективные и сигнальные факторы:

Объективные переменные:

- Опыт работы
- Возможность преподавания ЕГЭ, олимпиадной и высшей математики
- Возможность выезда на занятие в любой район

Сигнальные переменные:

- Количество отзывов
- Наличие на странице фото или видео
- Образование топового университета
- Количество подтвержденных дипломов

Также проверяется гипотеза, что женщины и мужчины назначают различные цены занятия.

Таким образом, итоговая модель образования цены:

$$\begin{aligned} price = \beta_0 + \beta_1 * experience + \beta_2 * experience^2 + \beta_3 * reviews + \beta_4 * video + \\ \beta_5 * diplomas + \beta_6 * photo + \beta_7 * EGE + \beta_8 * olimpiads + \beta_9 * higher_math + \beta_{10} \\ * top_university + \beta_{11} * any_neighborhood + \beta_{12} * male + \varepsilon, \end{aligned}$$

где ε – нормально распределенная случайная ошибка.

Оценка производится с помощью метода наименьших квадратов.

Результаты оценивания модели представлены на рисунке 5.

OLS Regression Results						
=====						
Dep. Variable:	price	R-squared:	0.243			
Model:	OLS	Adj. R-squared:	0.241			
Method:	Least Squares	F-statistic:	81.26			
Date:	Tue, 25 Apr 2023	Prob (F-statistic):	5.54e-177			
Time:	00:11:46	Log-Likelihood:	-27534.			
No. Observations:	3565	AIC:	5.509e+04			
Df Residuals:	3552	BIC:	5.517e+04			
Df Model:	12					
Covariance Type:	HC3					
=====						
	coef	std err	z	P> z	[0.025	0.975]

Intercept	483.3087	37.780	12.793	0.000	409.260	557.357
experience	34.8235	2.734	12.739	0.000	29.466	40.181
experience_square	-0.6501	0.061	-10.643	0.000	-0.770	-0.530
reviews	3.3898	0.918	3.694	0.000	1.591	5.188
video	395.0398	57.435	6.878	0.000	282.470	507.610
diplomas	41.5321	5.870	7.075	0.000	30.027	53.038
photo	260.9179	26.612	9.805	0.000	208.760	313.076
EGE	180.8948	20.003	9.043	0.000	141.689	220.101
olimpiads	186.7472	28.381	6.580	0.000	131.121	242.374
higher_math	101.7517	22.010	4.623	0.000	58.614	144.890
top_university	112.6172	18.722	6.015	0.000	75.922	149.313
any_neighborhood	-32.3525	21.514	-1.504	0.133	-74.520	9.815
male	-19.7188	20.430	-0.965	0.334	-59.762	20.324
=====						

Рисунок 5 – Результаты оценивания модели ценообразования

По результатам оценивания модели можно сделать вывод, что значимыми для ценообразования факторами оказались все, кроме пола и возможности выезда на занятие в любой район Москвы (для них P-value > 0.10).

Вклад сигнальных факторов в ценообразование:

Из сигнальных факторов наиболее значимо повышают цену занятия наличие фото и видео: репетитор с фотографией на странице устанавливает цену на занятие в среднем на 261 рубль выше, чем репетитор без фотографии; репетитор с видеовизиткой устанавливает цену на занятие в среднем на 395 рублей выше, чем репетитор без нее. За каждый дополнительный отзыв репетитор повышает цену занятия в среднем на 3 рубля, а за каждый дополнительный верифицированный диплом - на 42 рубля.

Таким образом, сигнальные факторы значимо положительно влияют на цену занятия. Почему так происходит? Возможное объяснение состоит в том, что подавая сигнал (фото, видео, подтвержденные дипломы), репетитор сообщает потенциальным ученикам о своем профессионализме и честности. При этом сигналы, подаваемые другими учениками в виде отзывов, укрепляют репутацию репетитора. Все эти факторы создают заочное доверие между потенциальными учениками и репетитором, что снижает возможные риски взаимодействия между ними, а значит уменьшает и премию за риск для ученика. Так как

аналогом премии за риск в данном случае является дисконт со стороны репетитора, то уменьшение риска влечет за собой естественное увеличение стоимости занятия.

Вклад объективных факторов в ценообразование:

Объективные факторы также вносят значительный вклад в стоимость занятия: репетитор, преподающий олимпиадную математику устанавливает цену в среднем на 187 рублей выше, чем репетитор, не преподающий олимпиадную математику; репетитор, который готовит к ЕГЭ, в среднем назначает цену на 180 рублей выше, чем репетитор, который не готовит к ЕГЭ; репетитор, который преподаёт высшую математику, в среднем назначает цену на 102 рубля выше, чем репетитор, который преподаёт высшую математику.

Преподавательский опыт вносит вклад в назначаемую цену по следующей квадратичной зависимости:

$$\text{price} = \text{const} + 34,8235 * \text{experience} - 0,6501 * \text{experience}^2$$

График данной зависимости при $\text{const} = 0$ представлен на рисунке 6.

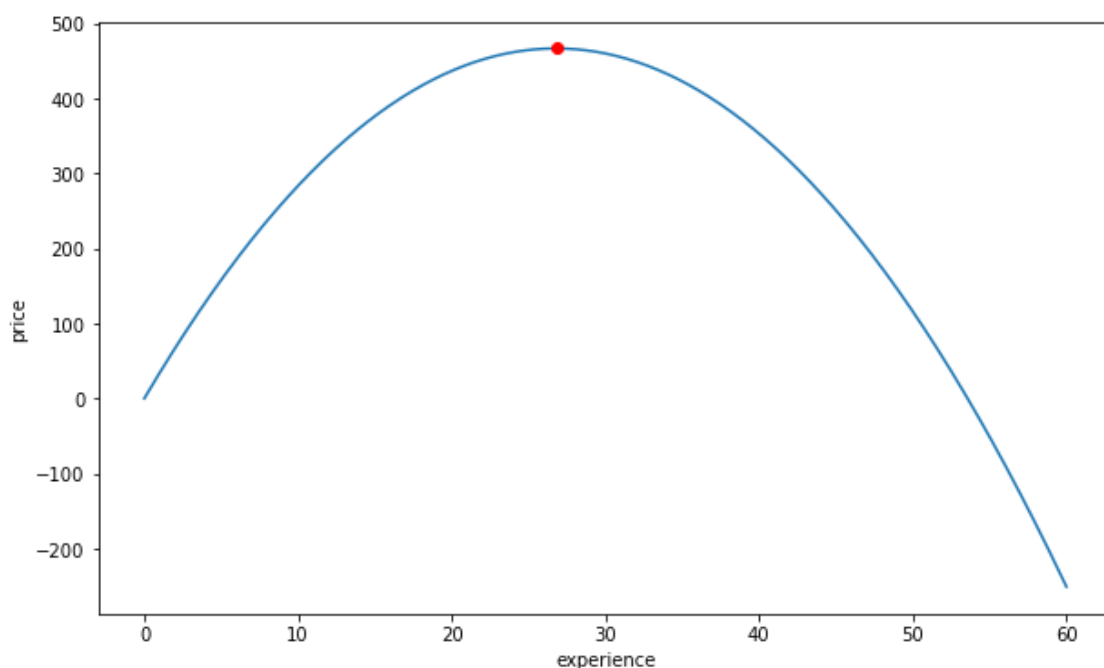


Рисунок 6 – График зависимости $\text{price} = 34,8235 * \text{experience} - 0,6501 * \text{experience}^2$.

Экстремум функции: $\text{experience} = 26.7831872019689$, $\text{price} = 466.342159763883$

Таким образом можно утверждать, что у репетитора по метематики существует "пик карьеры": при увеличении преподавательского опыта до 27 лет стоимость занятия увеличивается, однако после этого порога репетитор начинает снижать стоимость занятия. На "пике карьеры" репетитор устанавливает цену занятия в среднем на 466 рублей выше, чем репетитор без опыта работы.

Данный на первый взгляд парадоксальный результат можно объяснить, например, следующим образом: репетиторы с опытом работы больше 27 лет — это, вероятно, немолодые школьные или вузовские преподаватели, которые уже имеют основную работу и используют репетиторство как дополнительный источник дохода. Поэтому такие люди не устанавливают слишком высокую цену за занятие.

После удаления незначимых факторов пола и возможности выезда на занятие в любой район Москвы, получена формула, по которой можно узнать среднюю цену, устанавливаемую репетитором с заданными характеристиками:

$$\begin{aligned} price = & 465,4152 + 35,6535experience - 0,6639experience^2 + 3,3934reviews \\ & + 392,0863video + 41,6317diplomas + 260,2928photo + 177,4490EGE \\ & + 182,4560olimpiads + 96,5767higher_{math} + 110,7212top_university \end{aligned}$$

5. Анализ основных групп репетиторов по математике

Для идентификации основных групп репетиторов используется кластерный анализ методом k-средних.

Качество кластеризации определяется с помощью метода силуэта.

Силуэт — это мера того, насколько объект похож на свой кластер по сравнению с другими кластерами. Для каждого объекта этот коэффициент рассчитывается следующим образом:

- a. Вычисляется среднее расстояние между текущим объектом и всеми другими объектами в том же кластере (коэффициент "a").
- b. Вычисляется среднее расстояние между текущим объектом и всеми объектами в ближайшем кластере (коэффициент "b").
- c. Вычисляется коэффициент силуэта как $(b - a) / \max(a, b)$.

Данные шаги повторяются для различного числа кластеров и выбирается такое количество кластеров, при котором значение коэффициента силуэта максимально.

Оценки коэффициентов силуэта при различных количествах кластеров показаны на рисунке 7.

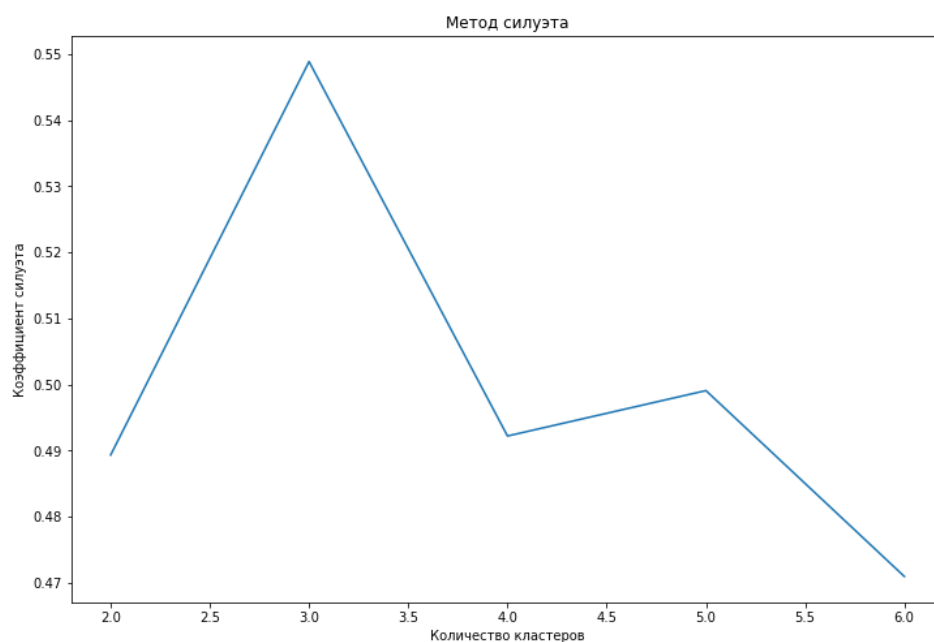


Рисунок 7 – Оценки коэффициентов силуэта при различных количествах кластеров

Таким образом, оптимальное количество кластеров – три.

В таблице 2 показаны результаты кластеризации методом k-средних ($k = 3$).

cluster	1	2	3
ID_repetitor	86165.662053	90129.025140	173405.036752
price	1407.554787	1833.240223	1208.205128
experience	28.892734	16.025140	6.997009
age	58.964245	41.594972	28.336325
reviews	9.414072	38.581006	3.631624
video	0.016148	0.245810	0.023504
photo	0.799308	0.986034	0.970085
diplomas	1.033449	2.416201	1.050000
students	15.173010	59.983240	5.735470
years	8.071799	7.758613	4.050356
male	0.335640	0.508380	0.432051
EGE	0.745098	0.868715	0.577778
olimpiads	0.229527	0.427374	0.190598
higher_math	0.319493	0.541899	0.345299
top_university	0.396770	0.620112	0.593590
students_per_year	2.101535	9.272663	1.845076
experience_square	937.005767	345.399441	73.630342
any_neighborhood	0.168397	0.262570	0.350855

Таблица 2 – Результаты кластеризации

Количество людей в первом кластере – 867, во втором – 358, в третьем – 2340.

Таким образом среди репетиторов по математике отчетливо выделяются три группы:

Группа 1: пожилые преподаватели

Средний возраст и стаж работы для репетитора из данной группы составляют 59 и 29 лет соответственно. Предположительно, большинство из них являются школьными или университетскими преподавателями. Средний репетитор из данной группы — это женщина, которая готовит школьников к ЕГЭ, но не готовит к олимпиадам и не обучает высшей математике. Она набирает в среднем 2 ученика в год и устанавливает цену за занятие 1407 рублей. Пожилые преподаватели — это группа с наименьшим среди других групп количеством выпускников топовых вузов и наименьшим количеством репетиторов, выезжающих на занятия в любой район Москвы.

Группа 2: профессиональные репетиторы

Наиболее малочисленной группой являются специалисты со средним возрастом и стажем работы 42 и 16 лет соответственно. Предположительно, большинство из них являются профессиональными репетиторами, то есть людьми, чей основной источник дохода — это репетиторство. В данной группе примерно равное количество мужчин и женщин. Средний репетитор из данной группы готовит к ЕГЭ и к высшей математике. Он набирает в среднем 9 учеников в год и устанавливает цену за занятие 1833 рубля. Профессиональные репетиторы обладают самым высоким среди других групп количеством отзывов, количеством подтвержденных дипломов; они наиболее часто являются выпускниками топовых вузов, чаще других готовят к олимпиадам. Также почти все видеопрезентации на портале выкладывают именно профессиональные репетиторы.

Группа 3: студенты и недавние выпускники

Наиболее многочисленная среди трех групп. Средний возраст и стаж работы для репетитора из данной группы составляют 28 и 7 лет соответственно. Предположительно, для данной группы репетиторов занятия являются подработкой во время учебы, или временной работой на начальном этапе карьеры. Средний репетитор из данной группы — это женщина, выпускница или студентка топового университета, которая готовит школьников к ЕГЭ, но не готовит к олимпиадам и не обучает высшей математике. Она набирает в среднем 2 ученика в год и устанавливает цену за занятие 1208 рублей. Репетиторы из данной группы наиболее часто выезжают на занятия в любой район Москвы, что, вероятно, является их конкурентным преимуществом перед более опытными специалистами из двух других групп.

Основные выводы:

Был проведен анализ рынка репетиторов по математике в городе Москва. Источник данных — портал repetit.ru

После удаления из выборки неактивных репетиторов было выяснено, что по состоянию на 24 марта 2023 года в г. Москва на платформе repetit.ru работают 3565 репетиторов по математике. Медианный репетитор — это женщина возрастом 31 год со стажем работы 10 лет.

Практически все репетиторы преподают программу школьного курса (98% репетиторов), а также большинство репетиторов готовят школьников к выпускным экзаменам ОГЭ (82%) и ЕГЭ (64%).

Было выяснено, что всех репетиторов по математике можно объединить в три различные группы: пожилые преподаватели, профессиональные репетиторы, а также студенты и недавние выпускники. Наибольшую среднюю стоимость занятия устанавливают профессиональные репетиторы (1833 рубля за 60 минут), а наименьшую — студенты и выпускники (1208 рублей за 60 минут).

На стоимость занятия, которую устанавливают репетиторы, также влияют следующие факторы: возможность преподавания ЕГЭ, олимпиадной и высшей математики, количество отзывов, наличие на странице repetit.ru фото или видео, образование топового университета, количество подтвержденных дипломов, опыт работы. Было также показано существование у репетиторов по математике "карьерного пика".

Итоговая формула, по которой можно узнать среднюю цену, устанавливаемую репетитором с заданными характеристиками, представима в следующем виде:

$$\begin{aligned} price = & 465,4152 + 35,6535experience - 0,6639experience^2 + 3,3934reviews \\ & + 392,0863video + 41,6317diplomas + 260,2928photo + 177,4490EGE \\ & + 182,4560olimpiads + 96,5767higher_math + 110,7212top_university \end{aligned}$$

В целом, методология, представленная в данном исследовании, является достаточно универсальной и поэтому может быть применена для анализа других репетиторов, в том числе работающих на других платформах.