**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**

**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**

**(РУТ (МИИТ)**

Институт/факультет «Академия “Высшая Инженерная Школа”, АВИШ»

Специальность/Направление подготовки Информатика и вычислительная техника

**КУРСОВАЯ РАБОТА**

**по дисциплине:** Технологии хранения больших данных

|  |  |
| --- | --- |
| **на тему:** | "Аналитика успеваемости студентов с помощью методов Big Data " |

|  |  |
| --- | --- |
| **Студент группы ШАД 21**2 | **Гаганов А. А.** |
| **Научный руководитель** | **Зуев И. А.** |

**Москва 2023 г.**

# **Оглавление**

[**Оглавление** 2](#_Toc153555644)

[**Введение** 3](#_Toc153555645)

[**Исследование данных** 4](#_Toc153555646)

[**Заключение** 11](#_Toc153555647)

[**Список использованных источников и программ** 14](#_Toc153555648)

[**Приложение** 15](#_Toc153555649)

# **Введение**

Тема успеваемости учащихся всегда будет актуальной, ввиду её большого влияния на будущее как самих школьников, так и образовательных учреждений. Целью данной курсовой работы является проведение аналитического исследования данных об успеваемости учеников с целью выявления основных факторов, влияющих на неё, а также выработки рекомендаций по её улучшению.

Полученные выводы помогут понять, что именно влияет на успеваемость, и решить, как её можно улучшить. Это несёт пользу для 3 групп людей: учителя смогут улучшить статистику своих занятий, тем самым получая повышения зарплат, премии и повышения; родители смогут быть более спокойными за будущее своих детей; ученики смогут улучшить свои шансы на поступление в лучшее высшее учебное заведение, тем самым увеличив шансы найти лучшую работу. Государство же имеет смешанное влияние: с одной стороны в стране появится больше высококвалифицированных работников, с другой стороны работников, наоборот, низкой квалификации станет меньше, что, в зависимости от государства, может иметь негативные последствия. То же самое относится и к работодателям, влияние на которых зависит от их рода деятельности.

В ходе работы будут использованы соответствующие теоретические и практические материалы, чтобы достичь поставленных целей.

# **Исследование данных**

Для получения описательной статистики был взят набор данных, в котором было приведено исследование, отражающее успеваемость учащихся в системе среднего образования двух португальских школ. Атрибуты данных включают оценки учащихся, демографические, социальные и связанные со школой характеристики, и они были собраны с использованием школьных отчетов и анкет. Данные были обработаны и проанализированы, после начался этап построения графиков. Опишем эти графики:

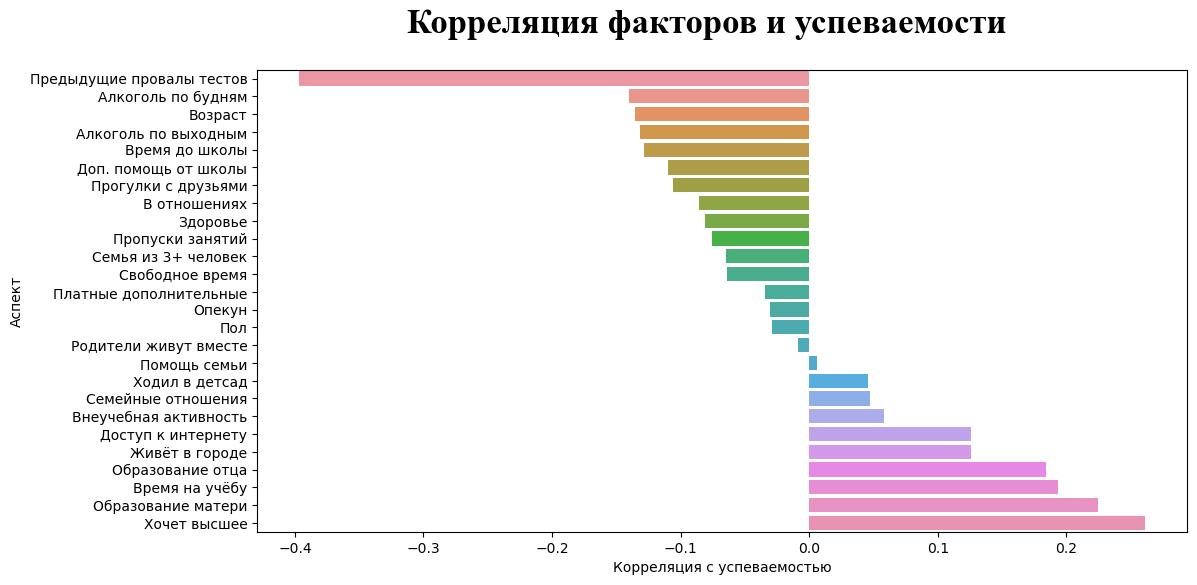


Рис. 1

В используемом датасете можно выделить 26 факторов, влияющих на успеваемость учеников. Наибольшее влияние имеют количество предыдущих провалов тестов, желание получит высшее образование, время, выделенное на учёбу в неделю, и образование родителей.

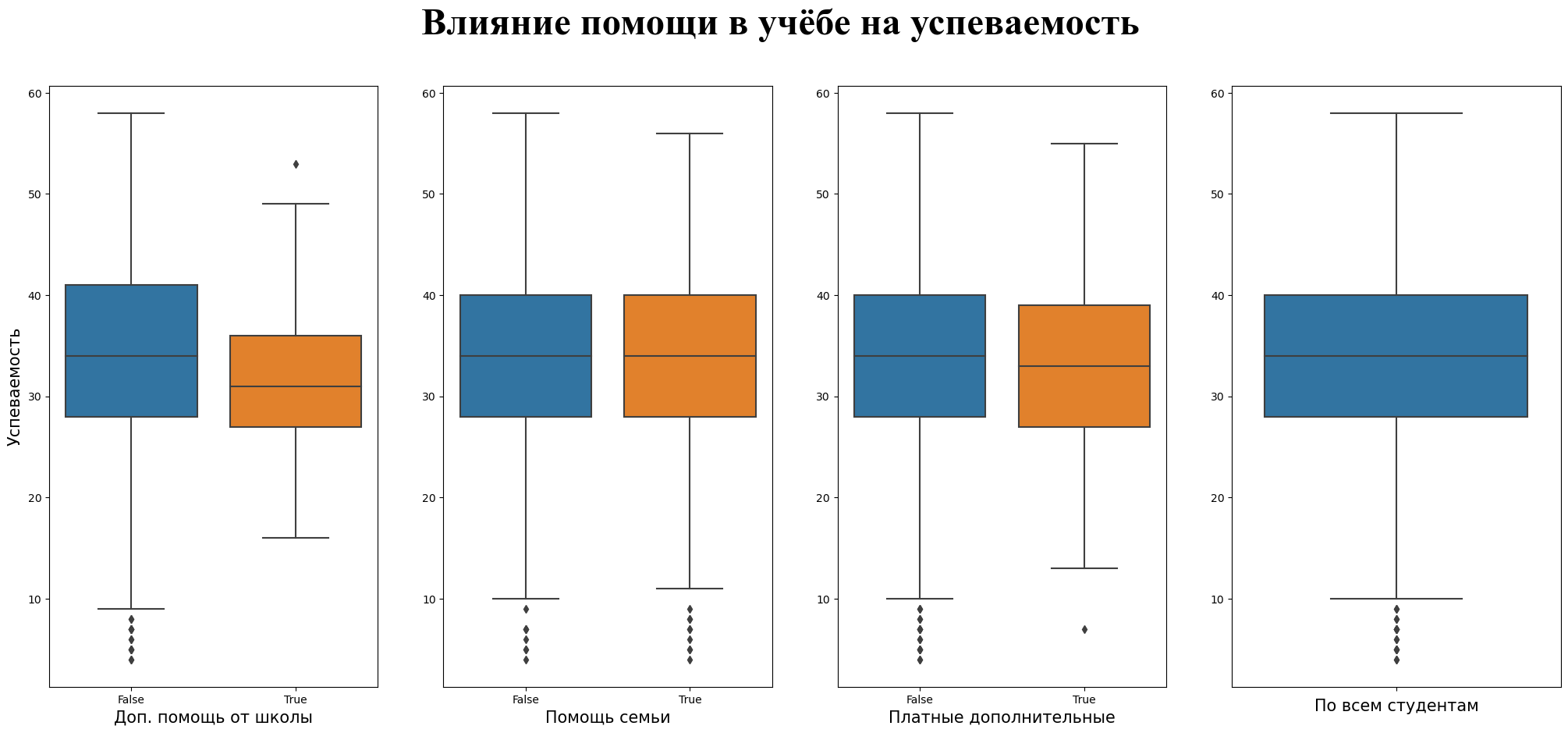


Рис. 2

Помощь в обучении действительно может иметь положительный эффект: учащиеся, получающие помощь имеют более высокий порог минимальной суммы балов, но также стоит заметить, что лучшие ученики находятся именно среди тех, кто помощь не получает.

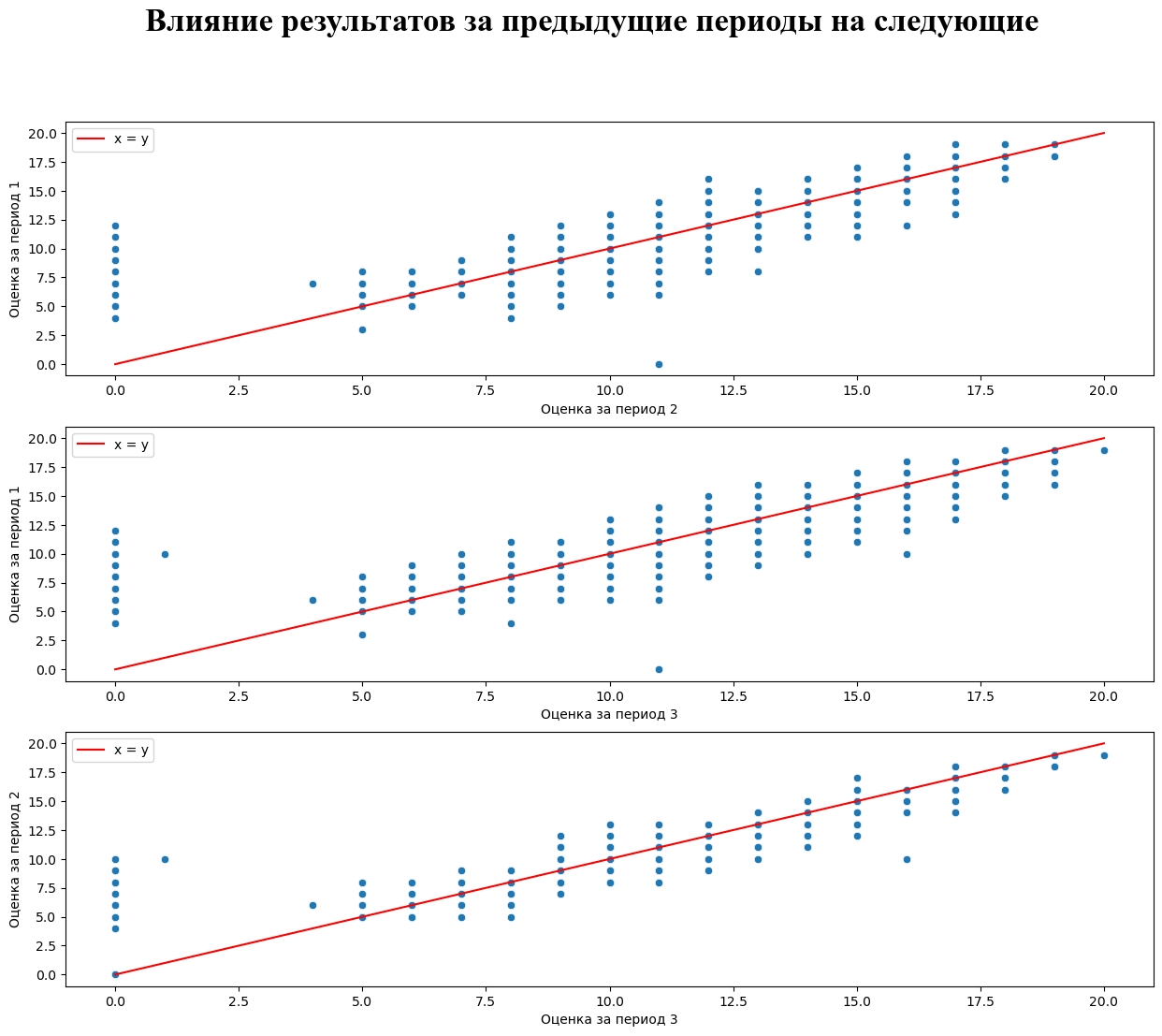


Рис. 3

Не имеет значения, какие 2 периода взять, результат один – как школьник учился во время предыдущего периода – примерно так же он будет учиться и в следующем. Нельзя сказать, что кто-то «решил взяться за ум» после провала за прошлый период (в таком случае большинство точек было бы ниже красной линии), так же, как и нельзя говорить об обратном, что ученики «скатываются» в плохие оценки (в таком случае большинство точек было бы выше красной линии).

Присутствуют так же выбросы, представленные учащимися, которые покинули школу после первого или второго периода.

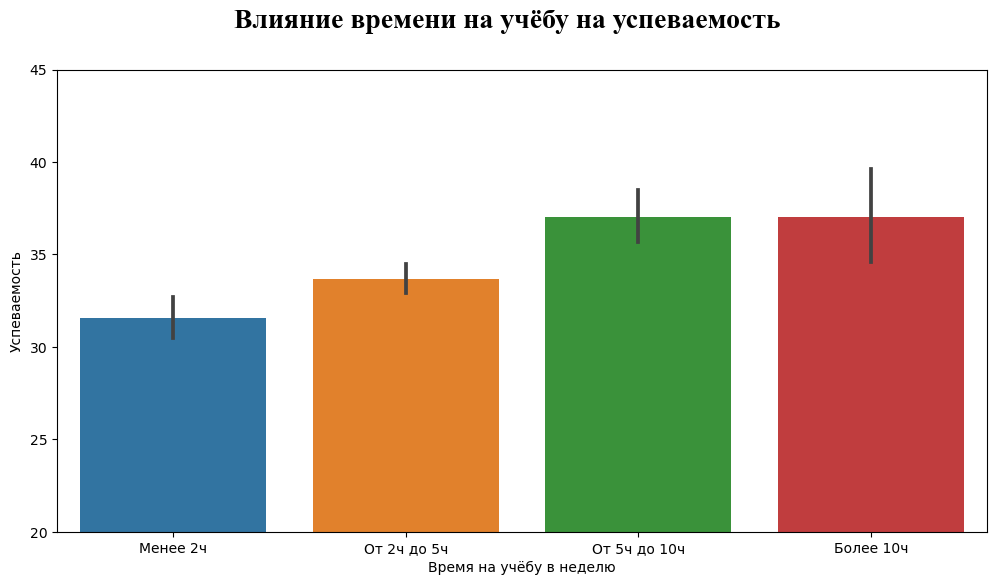


Рис. 4

Оптимальное время на учёбу лежит в районе от 5 до 10 часов, потому что средний бал практически не меняется, а прирост максимального становится несущественным, а что самое главное – минимальный бал заметно падает.

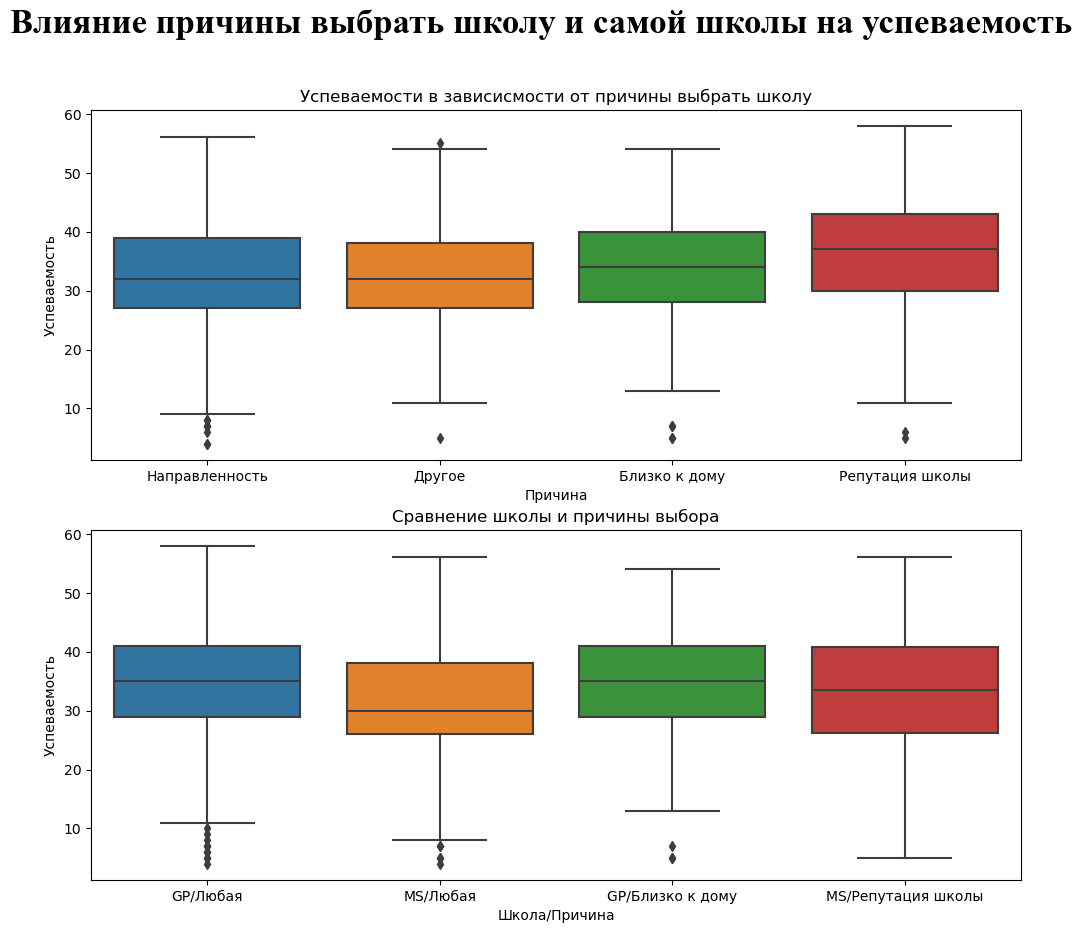


Рис. 5

Причина выбрать школу имеет немалый эффект на успеваемость, вероятно, из-за её прямой связи с мотивацией на учёбу, пример: выбрал школу из-за её репутации – стараешься соответствовать – учишься лучше. Эффект достаточно сильный, чтобы практически нивелировать разницу в силе преподавания в разных школах. Хоть школа GP и сильнее в целом, если сравнить её учеников с условно низкой мотивацией и учеников школы MS с условно высокой – то разница станет сильно меньше, а максимум успеваемости и того станет выше.

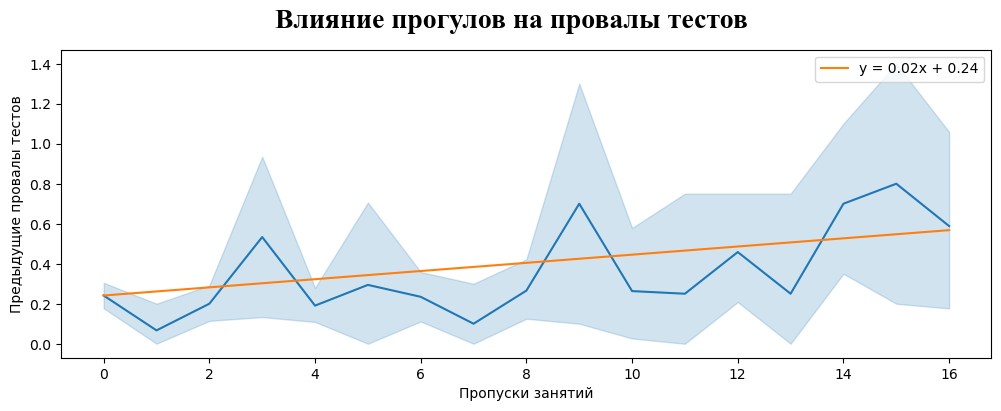


Рис. 6

Пропуск занятий имеет влияние на провалы тестов, а следовательно, и успеваемость. Хоть рост среднего провалов и не велик, но разброс становится сильно больше, а следовательно пропуски помимо очевидного ухудшения успеваемости влекут так же и её нестабильность.

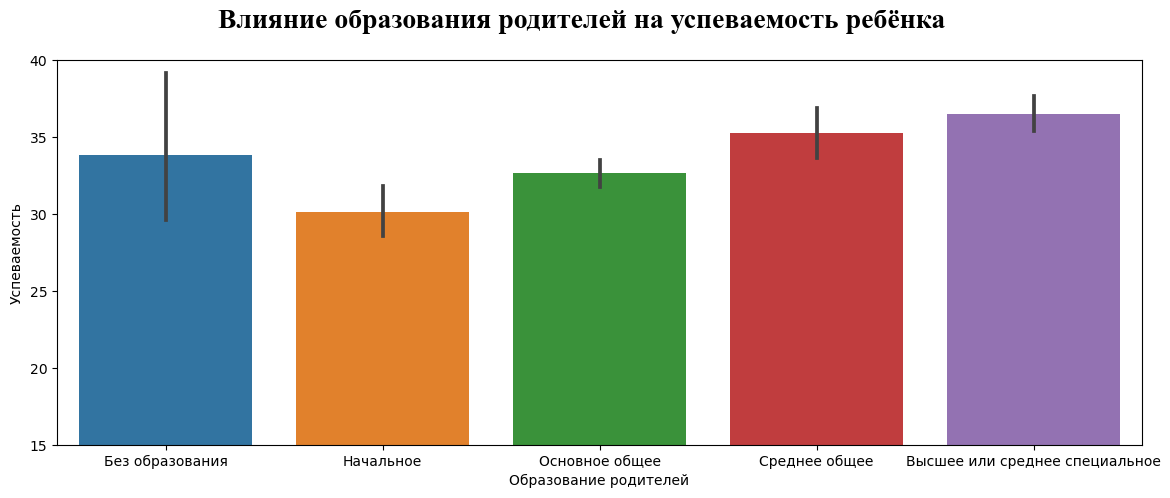


Рис. 7

Хоть и, смотря на рис. 1, можно было сделать вывод, что образование родителей имеет достаточное положительное влияние на успеваемость их ребёнка, на деле можно увидеть, что тенденция не такая однозначная. Если смотреть на значения с начального по высшее, то действительно видно, что чем лучше образование родителей – тем и выше успеваемость ребёнка, но отсутствие образования у обоих родителей показывает показатели почти на уровне законченного среднего, и это даже не смотря на максимальную успеваемость, которая у детей с необразованными родителями самая высокая.

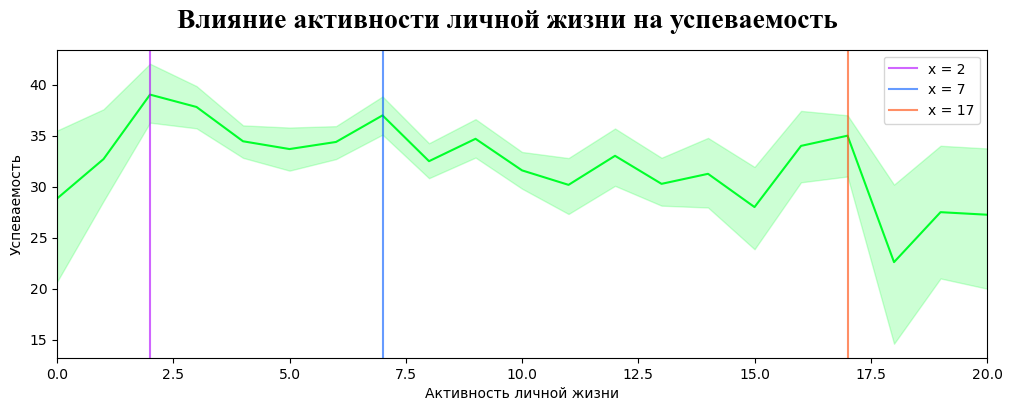


Рис. 8

Можно заметить, что ученики совсем без личной жизни имеют успеваемость ниже тех, кто имеет хоть какую-то, причём максимум достигается при значении 2 из 20, а дальше идёт плавное убывание. Пик на значении 7 объясняется переносом пика на 2 у людей, имеющих романтические отношения, так как этот параметр влияет не плавно, а резко (либо 0, либо сразу 4, а 4+2=6, где и начинается пик 6-7). Пик же на 17 сложно объяснить, вероятно повторяется история с переносом пика на 2, но это уже имеет мало значения, так как пик локальный, а в общем значение успеваемости всё равно низкое.

# **Заключение**

1. Можно выделить множество факторов жизни учеников, которые влияют на их успеваемость, тем не менее весомую корреляцию с ней имеют лишь несколько из них, а именно: положительно влияют желание получит высшее образование, время, выделенное на учёбу в неделю, и образование родителей; негативно же влияет количество предыдущих провалов тестов.
2. Помощь в обучении как от семьи, так и от школы в первую очередь помогает избежать полного провала в обучении, нежели достичь максимальных результатов. Ученики, способные набирать действительно хорошие балы делают это без помощи.
3. Школьники, в большинстве своём, не имеют тенденций менять своё отношение к учёбе после получения результатов тестов. Те, кто получил плохие оценки продолжат их получать, в то время как лидеры успеваемость продолжат поддерживать свой статус. Те, кто решил подтянуть знания после провала теста или кто по какой-то причине перестал учиться после хороших результатов, практически не встречаются.
4. На учёбу стоит выделять достаточно много времени в неделю: от 5 до 10 часов. Тем не менее последующее увеличение продолжительности самостоятельных занятий влечёт за собой лишь небольшое увеличение максимального порога балов, что не стоит того, особенно учитывая, что средние и минимальные балы начинают падать ввиду усталости учеников от постоянной учёбы.
5. Выбор школы имеет большое значение на успешность последующего обучения ученика, некоторые образовательные учреждения попросту имеют преподавание лучше, чем другие. Тем не менее этот фактор может быть несколько нивелирован мотивацией школьника к учёбе.
6. Прогулы уроков имеют влияние на успеваемость, тем не менее оно не такое значительное, как может изначально показаться, и те, кто прогулял много уроков имеют в среднем результаты не слишком отличные от тех, кто посещал школу исправно. Тем не менее пагубное влияние заключается в нестабильности успеваемости прогульщиков, что, вероятно, вызвано тем, что прогулять возможно как очень важный урок с объяснением новых тем, так и урок, на котором учитель отсутствовал из-за внеочередного совещания.
7. Чем лучше образование у родителей ученика, тем лучше он учится сам, что вполне логично, так как в таком случае они могут больше помогать ему в учёбе, а также это означает лучшее материальное положение семьи в целом. Однако семьи с родителями вообще без образования резко выделяются, имея лучшие максимальные результаты и средние на уровне с родителями, получившими среднее образование. Это может быть вызвано повышенной мотивацией как самих родителей, которые хотят обеспечить детям образование, которое они сами получить не смогли, так и детей, которые хотят жить лучше в будущем.
8. Школьники, совсем не имеющие личной жизни, имеют успеваемость резко ниже тех, кто выделяют на неё некоторое время, что может быть связано с выводом 4 и усталостью от учёбы. Однако последующее увеличение активности личной жизни влечёт плавное ухудшение успеваемости, так как на учёбу перестаёт хватать времени и/или приоритеты учеников смещаются в сторону развлечений.

В заключении можно дать следующие советы по улучшению успеваемости:

1. Как семье, так и школе стоит помогать отстающим ученикам.
2. Не стоит лишний раз беспокоить отличников, которые сами справляются со своим обучением.
3. Стоит приложить больше усилий к мотивации учащихся.
4. Стоит стараться обеспечить ученика 5-10ч самостоятельной учёбы в неделю.
5. Стоит выбирать школу в первую очередь по её репутации, но никак не исключительно по близости к дому или другим причинам.
6. Стоит следить за посещаемостью ученика и не только пресекать несоблюдение дисциплины, но и обеспечить ему простой доступ к школе (имеется в виду не обязательно близость к дому, а выделение денег на общественный транспорт до неё или вообще самостоятельное сопровождение на семейном транспорте).
7. Нельзя полностью запрещать ученикам вести личную жизнь, однако если их приоритеты стали явно перевешены в сторону развлечений, то стоит им напомнить, как это повлияет на их успеваемость и, следовательно, будущую жизнь.

И подводя итог в результате анализа успеваемости школьников были выполнены все поставленные задачи и выявлены ключевые факторы, влияющие на успешность учащихся в учёбе.

# **Список использованных источников и программ**

1. Использовался датасет с сайта Kaggle: <https://www.kaggle.com/datasets/whenamancodes/student-performance>
2. Библиотеки, использованные в изучении датасета: numpy, pandas, matplotlib, seaborn

# **Приложение**

Файл с кодом находится на GitHub по ссылке – <https://github.com/GaganovAlexander/BDST-course-paper>