

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«АЛТАЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Институт математики и информационных технологий  
Кафедра информатики

**АНАЛИЗ, ОЦЕНКА И ВИЗУАЛИЗАЦИЯ ДАННЫХ  
ВСЕРОССИЙСКИХ ПРОВЕРОЧНЫХ РАБОТ  
(НА ОСНОВЕ ОФИЦИАЛЬНЫХ ДАННЫХ АЛТАЙСКОГО КРАЯ)  
(выпускная квалификационная работа)**

Выполнила:  
студент 474а группы,  
Дранева Софья Сергеевна

\_\_\_\_\_  
(подпись)

Научные руководители:  
к.ф.-м.н., доцент  
Рязанов Михаил Анатольевич

\_\_\_\_\_  
(подпись)

Работа защищена:  
«\_\_»\_\_\_\_\_2021 г.

Оценка:\_\_\_\_\_

Председатель ГЭК:

\_\_\_\_\_  
(подпись)

Допустить к защите:  
Зав. кафедрой: к.ф.-  
м.н., доцент,  
Козлов Денис Юрьевич

\_\_\_\_\_  
(подпись)

«\_\_»\_\_\_\_\_2021 г.

Барнаул 2021

## РЕФЕРАТ

**Тема выпускной квалификационной работы:** Анализ, оценка и визуализация данных Всероссийских проверочных работ (на основе официальных данных Алтайского края).

**Цель работы** – провести анализ данных, позволяющий представить оценку уровня образовательных достижений по русскому языку и математике среди 5-7 классов для принятия управленческих решений в рамках региональной системы оценки качества образования в Алтайском Крае.

**Объект исследования** – данные Всероссийских проверочных работ в Алтайском крае.

**Предмет исследования** – анализ, оценка и визуализация данных Всероссийских проверочных работ по математике и русскому языку в 5-7 классах.

В результате исследования были решены следующие задачи: выполнена оценка качества образования и объективности результатов образовательных организаций; выявлены образовательные организации с завышенными и/или необъективными результатами в Алтайском крае; составлен технический отчет по проведению ВПР в 2020 году; разработаны вычислительные и графические методы для визуализации данных на языке программирования python.

Объем работы – 52 страниц печатного текста, количество рисунков – 9, таблиц – 10, приложений – 1, 18 использованных источников литературы.

Ключевые слова: Всероссийские проверочные работы, ВПР, анализ данных, визуализация данных.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1. Постановка задачи анализа результатов Всероссийских проверочных работ .....	7
1.1. Основные понятия и практическая значимость Всероссийских проверочных работ .....	7
1.2. Описание процесса сбора и обработки данных .....	9
1.3. Постановка задачи .....	10
2. Предварительная обработка исходных данных.....	14
2.1. Методы предобработки данных .....	14
2.2. Описание и предобработка исходных данных.....	16
2.3. Сравнительный анализ и визуализация данных .....	20
3. Оценка качества образования и объективности результатов образовательных организаций .....	27
3.1. Применение статистических методов.....	27
3.2. Практическое обоснование результатов исследования в предметной области .....	29
Заключение .....	32
Литература .....	33
Приложение .....	35

## ВВЕДЕНИЕ

Ежегодно в школах России проводится процедура в области оценки качества образования – Всероссийские проверочные работы (ВПР). Они необходимы для формирования единых ориентиров в оценке результатов обучения и подходов к оцениванию образовательных достижений обучающихся. Эти работы не влияют на достижения учеников, они необходимы для анализа данных о состоянии региональных и муниципальных систем образования.

Результаты ВПР используются для мониторинга ведения образовательных стандартов, продвижения единых ориентиров в оценивании уровня подготовки школьников, самодиагностики школ. Целью таких исследований является повышение эффективности системы оценки качества образования, что включает в себя улучшение качества развития образования, совершенствование образовательных программ и выявление проблем в знаниях учеников.

Данная система оценивания была введена в 2015 [17] году и совершенствуется до сих пор, но также она имеет ряд таких проблем, как отсутствие полного контроля процесса проведения тестирования со стороны общественных наблюдателей, систем видеонаблюдения, а также не исключена помощь со стороны организаторов при проведении работ. Таким образом, нарушается объективность результатов многих школ – результаты тестирования завышены, либо не соответствуют школьным отметкам.

Помимо этого, последние годы проблема многообразия учебных программ привела к тому, что в образовательных организациях (ОО) по-разному преподают материал и упускают требования государственных образовательных стандартов. Кроме того, по результатам проведения Национального исследования качества образования фиксируется падение уровня подготовки школьников после 4 класса.

**Цель работы** – провести анализ данных, позволяющий представить оценку уровня образовательных достижений по русскому языку и математике среди 5-7 классов для принятия управленческих решений в рамках региональной системы оценки качества образования в Алтайском Крае.

**Объект исследования** – результаты Всероссийских проверочных работ в Алтайском крае.

**Предмет исследования** – анализ, оценка и визуализация данных Всероссийских проверочных работ по математике и русскому языку в 5-7 классах.

**Практическая значимость** исследования состоит в выявлении образовательных организаций с завышенными результатами относительно оценок учеников, что дает возможность провести в следующем году тщательный контроль школ, а также предоставить методологические материалы для выявленных образовательных организаций.

Полученные данные позволяют решать следующие **задачи**:

- проведение оценки качества образования и объективности результатов образовательных организаций;
- выявление образовательных организаций с завышенными и/или необъективными результатами;
- описание на уровне региона предметных результатов, а также универсальных действий для подготовки технических отчетов;
- разработка вычислительных и графических методов для визуализации данных.

**Информационная база** исследования: результаты Всероссийских проверочных работ по русскому языку и математике в Алтайском крае за 2020 год. Данные содержат в себе информацию об образовательной организации,

числе учеников, отметке по школе и количестве баллов за каждое выполненное задание.

**Методы исследования:** методы и средства интеллектуального анализа данных, методы статистической обработки и визуализации данных.

В рамках исследования проверяется **гипотезы:**

- Гипотеза о заметном снижении качества подготовки в образовательных организациях Алтайского края от 5 к 7 классу по основным предметам (русскому языку и математике).
- Гипотеза о завышении оценок по ВПР относительно оценок по школе образовательными организациями (Результаты выполнения ВПР существенно выше средних результатов по региону, при этом образовательная организация не является лицеем или гимназией с углубленным изучением данных предметов).

# **1. Постановка задачи анализа результатов Всероссийских проверочных работ**

## **1.1. Основные понятия и практическая значимость Всероссийских проверочных работ**

Всероссийские проверочные работы (ВПР) – это диагностические работы, которые применяются для оценки индивидуальных достижений обучающихся и проверки соответствия знаний школьников требованиям Федеральным государственным образовательным стандартами (ФГОС) [2]. ВПР необходимы для выявления регионов и школ, в которых предоставляют качественное образование, либо другой уровень подготовки. Помимо этого, проводя анализ данной процедуры, можно получить общие итоги качества образования не только после окончания основных ступеней образования, но и на промежуточных этапах, а также проводить своевременную корректировку образовательных программ [3].

Отличительной чертой ВПР от других систем оценки качества образования является то, что они представляют собой итоговые контрольные работы, целью которых является мониторинг системы образования в Российской Федерации, который проводится для того, чтобы определить уровень подготовки обучающихся во всех регионах страны.

Проведение Всероссийских проверочных работ в Алтайском Крае осуществлялось во исполнение приказа Министерства образования и науки Алтайского края «О проведении мониторинга качества подготовки обучающихся общеобразовательных организаций Алтайского края в 2020 году» [7].

Процедура проведения ВПР не отличается от обычных контрольных работ в школе – наблюдателями и проверяющими являются сами учителя школы. Работы проверяются в день проведения, а результаты не влияют на достижения

учеников. Результаты вносятся в единую информационную систему, с данными которой могут работать эксперты.

Результаты ВПР используются:

- Методическими службами для разработки методических рекомендаций преподавания соответствующего учебного предмета в помощь педагогу.
- На региональном уровне – для повышения качества образования в муниципальных образовательных организациях, принятие управленческих решений, адресная помощь муниципалитетами, у которых были отмечены низкие результаты, а также совершенствование преподавания учебных предметов, повышение квалификации учителей.
- На уровне Российской Федерации – для формирования базы результатов обучающихся, развития модели единой системы оценки качества образования и для принятия эффективных управленческих решений по формированию программы повышения качества образования [4].

Последние годы проблема многообразия программ, учебников и учебных пособий привела к тому, что в российских образовательных организациях по-разному преподают материал и трактуют требования государственных образовательных стандартов, написанных в весьма обобщенном виде. Кроме того, результаты ВПР фиксируют падение уровня подготовки школьников после 4 класса. Между 5-м и 9-м классами проходит длительный период времени. Дети начинают взрослеть, некоторые из них в силу множества причин теряют мотивацию.

Средняя школа – очень сложный для детей период, когда нужно контролировать процесс обучения не только учителям, но и государству. Оно должно предлагать единые измерители качества образования на основе Федерального государственного образовательного стандарта, в котором содержаться контролируемые элементы содержания образования.



## 1.2. Описание процесса сбора и обработки данных

Вся информация о результатах ВПР находится в закрытом доступе в Федеральной информационной системе оценки качества образования (ФИС ОКО) [18], и открыта только зарегистрированным пользователям на официальном сайте. ФИС ОКО автоматизирует такие виды деятельности, как сбор, хранение и предоставление пользователю информации об проведенных работах.

Доступ пользователям системы на региональном уровне предоставляется:

- Специалистам органов исполнительной власти субъектов Российской Федерации, осуществляющих управление в сфере образования (ОИВ), курирующие вопросы государственного контроля качества образования и другие.
- Специалистам ОИВ и региональных учреждений повышения квалификации в системе образования, осуществляющие анализ качества образования на региональном уровне с целью выявления факторов, влияющих на качество образования, тенденций в системе образования, а также подготовки рекомендаций для специалистов системы образования.

Результаты проведенных работ загружаются каждой школой по отдельности. Работа с отчетностью осуществляется через веб-интерфейс. Муниципальный координатор контролирует своевременную загрузку форм сбора результатов общеобразовательных организаций своего муниципального образования. После окончания периода написания и загрузки работ в систему, появляется возможность просмотр отёчности по ВПР.

В общеобразовательные учреждения Алтайского края передается полная информация о результатах Всероссийских проверочных работах каждого ученика: о выполнении/невыполнении каждого из заданий, о количестве баллов, полученных за выполнение каждого из заданий по каждому предмету ВПР и код

школы. Протоколы результатов ВПР являются документом, который в обязательном порядке хранится в общеобразовательном учреждении.

Число заданий варьируется из года в год в зависимости от предмета и класса проведения работ. Для дальнейшей обработки данных необходимо иметь контрольно-измерительный материал (КИМ), который представляет собой систему оценивания выполнения всей работы и содержит максимальный балл за выполнение работы с рекомендациями по переводу первичных баллов в отметки по пятибалльной шкале.

### **1.3. Постановка задачи**

В Алтайском крае ежегодно составляется отчет по анализу результатов Всероссийских проверочных работ и имеется единый вариант использования результатов такого анализа. Традиционно в качестве интегрального показателя при оценке результатов итоговых работ в районе используются средние отметки, полученные учениками общеобразовательных учреждений по каждому предмету, переведенные по соответствующим критериям оценивания всей работы. Далее эти отметки сравниваются со средними отметками по краю; а средние отметки учеников образовательных учреждений сравниваются друг с другом. Эти данные используются для составления разнообразных отчетов для представления общей информации об проведенных работах разным кругам лиц.

Однако для понимания структуры образовательных результатов явно недостаточно информации о средних тестовых отметок учеников, эти данные могут являться лишь отправной точкой для дальнейшего анализа. Используя результаты более глубокого анализа данных можно обосновать выбор направления дальнейшего развития системы образования в крае, понять причины заниженных и завышенных результатов в образовательных организациях, определить уровень подготовки обучающихся и его качество образования в

регион.

В связи со спецификой районов Алтайского края, имеющих в районе структурой разных типов образовательных учреждений и результатов, получаемых в ходе анализа, сам такой анализ на уровне муниципалитетов края является сугубо «индивидуальным», сохраняя общность целей и подходов. В рассматриваемом ниже варианте анализ результатов ВПР на уровне края направлен на:

- определение направлений, тактики и стратегии развития системы образования в крае, создание основы для развития структуры образовательных учреждений района;
- удовлетворение потребностей методической службы края в построении системы повышения квалификации;
- возможность оценить уровень предметной подготовки в общеобразовательных учреждениях района на фоне Алтайского края.

При проведении анализа в первую очередь необходимо определить обобщенную цель проведения анализа результатов ВПР – определение степени достижения качества образования в образовательных учреждениях и в регионе в целом. Для достижения этой цели в процессе анализа должны быть решены следующие задачи:

1. Рассмотреть распределение первичных баллов и их перевод в соответствующие отметки по шкале перевода.
2. Оценить динамику изменения результатов учеников с 5 по 7 класс.
3. Определение факторов, оказавших наибольшее влияние на результат предметной подготовки.
4. Выявление образовательных организаций, в которых результаты были завышены или необъективными.
5. Провести оценку качества образования в Алтайском крае.

б. Разработать вычислительные методы для проведения анализа и визуализации данных.

В качестве объекта анализа будут рассматриваться результаты ВПР в Алтайском крае, а в качестве предмета анализа - оценка и визуализация данных ВПР по математике и русскому языку в 5 – 7 классах.

Анализ качества образования в 5 – 7 классах направлен на выявление системных тенденций и факторов, затрудняющих проведение мониторинга результатов ФГОС и развитию единого образовательного пространства в России. Выбор классов – с 5 по 7 – обусловлен фиксируемыми в последние годы существенными различиями в результатах внешней оценки успешности изучения предметов обучающимися начальной школы, с одной стороны, и обучающимися/выпускниками основной и средней школы – с другой [8].

Таким образом, необходимо проверить гипотезу о заметном снижении качества подготовки в образовательных организациях Алтайского края от 5 к 7 классу по основным предметам (русскому языку и математике).

В массив исходных данных для анализа входят результаты ВПР учеников общеобразовательных учреждений всех типов: средних общеобразовательных школ с углубленным изучением отдельных предметов или без, гимназий, лицеев. Министерством образования Алтайского края предоставляется список школ, в которых могут быть результаты итоговых работ выше среднего балла по краю из года в год. Такая ситуация обусловлена наличием школ, гимназий или лицеев, осуществляющих образовательный процесс, реализующий учебные программы углубленного изучения одного или нескольких предметов. Это необходимо учитывать при дальнейшем анализе данных.

Задания по ВПР сформированы таким способом, что степени преподаваемых знаний в школе будет достаточно для получения базового уровня результатов. При большом количестве учеников, написавших работы выше

базового уровня можно судить о том, что возможны ситуации, когда образовательная организация завышает оценки, либо учеников предварительно подготавливают к написанию ВПР. В связи с этим результаты могут считаться необъективными, что мешает своевременной оценке результатов обучения и улучшению развития образования в регионах страны.

Таким образом, в рамках исследования необходимо проверить гипотезу о завышении оценок по ВПР школами, где результаты выполнения ВПР существенно выше средних результатов по краю, при этом образовательная организация не является лицеем или гимназией с углубленным изучением данных предметов. В случае подтверждения гипотезы – выявить данные образовательные организации и предоставить полученные данные соответствующие организациям, которые предоставят методические указания данным образовательным организациям.

## **2. Предварительная обработка исходных данных**

### **2.1. Методы предобработки данных**

Для дальнейшего анализа и оценки необходимо провести предобработку исходных (необработанных) данных. Это процесс позволит привести необработанные данные к соответствующим требованиям для решения исходной задачи.

Предварительная обработка данных – это метод интеллектуального анализа данных, который является важнейшим этапом, от качества выполнения которого зависит возможность получения качественных результатов всего процесса [1]. Поскольку данные поступают из нескольких источников, исходные данные часто бывают неполными, непоследовательными или могут возникнуть проблемы из-за человеческих ошибок.

Одним из первых этапов предобработки данных является обезличивание персональных данных. Для проведения статистических, социологических, исторических, медицинских и других научных и практических исследований держатель (обладатель) массива персональных данных обезличивает используемые данные, придавая им форму анонимных сведений. Режим конфиденциальности, установленный для персональных данных, снимается. Обезличивание должно исключать возможность идентификации субъекта персональных данных [5].

К методам обезличивания персональных данных относятся:

- метод введения идентификаторов – замена части сведений (значений персональных данных) идентификаторами с созданием таблицы соответствия идентификаторов исходным данным;

- метод изменения состава или семантики – изменение состава или семантики персональных данных путем замены результатами статистической обработки, обобщения или удаления части сведений;
- метод декомпозиции – разбиение множества (массива) персональных данных на несколько подмножеств (частей) с последующим отдельным хранением подмножеств;
- метод перемешивания – перестановка отдельных записей, а также групп записей в массиве персональных данных [6].

Очень часто в наборе данных есть пропущенные значения. Это могло произойти во время сбора данных или объединения из нескольких наборов данных. Чтобы справиться с отсутствующими данными, можно использовать несколько подходов:

- устранить с помощью удаления полей с недостающими данными;
- заменить пропущенное значение вручную;
- игнорировать пропуски;
- заполнение их средним или медианным значением.
- подстановка по регрессии — использование регрессионной модели для замены пропущенных значений регрессионными данными.

Большой объем дополнительных бессмысленных данных называется шумом. Это могут быть:

- дубликаты записей данных;
- сегменты данных, не представляющие ценности для конкретного исследования;
- ненужные информационные поля для каждой из переменных.

Для решения этой проблемы можно применить один из следующих методов:

- регрессионный анализ – помогает решить, какие переменные действительно имеют влияние. Он необходим для сглаживания больших объемов данных.
- применение алгоритмов кластеризации для группировки данных. При этом необходимо уделить особое внимание выбросам. Выбросы – это особые точки данных, не похожие на остальную часть домена. Важно не подменять выбросы, воспринимая их как шум.

Данные могут содержать противоречивые значения. Это может быть связано с человеческой ошибкой или информация была неправильно прочитана при сканировании с рукописного бланка. Поэтому всегда рекомендуется выполнять оценку данных, знать, каким должен быть тип данных и является ли он одинаковым для всех объектов данных.

Помимо всего, набор данных может включать в себя объекты данных, которые дублируют друг друга. В большинстве случаев дубликаты удаляются, чтобы не дать этому конкретному объекту данных преимущества и повлиять на конечный результат.

Предобработка исходных данных – одна из наиболее актуальных задач интеллектуального анализа. Плохое качество данных является одной из самых больших проблем при построении аналитических решений, так как на основе некорректной информации делаются неверные выводы. Поэтому стоит уделить особое внимание данному этапу.

## **2.2. Описание и предобработка исходных данных**

На начальном этапе был произведен сбор информации об учащих, написавших работу по ВПР за 2020 год в Алтайском крае с 5 по 7 класс по математике и русскому языку. Массив исходных данных задан в формате CSV, где заголовок содержит информацию для каждого поля. Указаны:



- user – код школы с созданной таблицей соответствия идентификаторов исходных данных;
- id, code – номер ученика в классе
- otc – отметка по школе;
- gender – пол ученика;
- klass – литер класса;
- var-1, var-2 – вариант задания;
- sum – сумма баллов за выполнение всей работы;
- a-1-grade и т.д. – номера заданий.

user	id	code	otc	gender	klass	var-1	var-2	sum	a-1-grade	a-2-grade	a-3-grade	a-4-grade	a-5-grade	a-6-grade	a-7-grade	a-8-grade	a-9-grade	a-10-grade	a-11-grade	a-12-grade	a-13-grade	a-14-grade
sch223080	1	60001	3	м	а	2	NaN	11.0	1	0	1	0	2	X	1	0	0	X	1	X	0	X
sch223080	2	60002	4	ж	а	1	NaN	23.0	2	3	2	3	3	1	1	2	0	0	0	X	1	1
sch223080	3	60003	NaN	NaN	NaN	отсутствовал	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
sch223080	4	60004	3	ж	а	1	NaN	11.0	0	2	2	0	3	0	0	1	0	X	0	X	X	X

Рисунок 2.1. Пример исходных данных

Для анализа и визуализации данных был выбран язык программирования Python [11]. Python является отличным инструментом для работы с данными, их обработке и визуализации. Он прост в изучении, достаточно мощный и гибкий, позволяет работать с большими объемами данных. В нем уже существует огромное количество готовых библиотек для визуализации, машинного обучения и интеллектуального анализа данных. Так же его использует большое число людей и организаций, что способствует его быстрому развитию. Язык является кроссплатформенным и поддерживает почти все современные системы.

Также одним из популярных средств для статистического анализа является EXCEL [16]. Но при обработке сотни или тысяч строк данных, анализировать и визуализировать их в EXCEL довольно затруднительно. Против удобства

начинает играть то, что считается преимуществом в других ситуациях – EXCEL показывает таблицу, какой бы большой она ни была. Её трудно куда-то спрятать и рассматривать только ее часть. Что-то из этого можно сделать с помощью фильтров, но их функционал довольно ограничен. Также при разработке и применении новых методов анализа становится затруднительно составлять отчеты и выполнять одну и ту же обработку данных каждый раз на новом листе. Написание макросов на встроенном в MS Office языке помогает решить проблему, но ненадолго.

Таким образом Python предпочтительнее EXCEL, он дает возможность автоматизировать сбор и анализ данных, а при работе с большими объемами данных, Python сократит их обработку с часов до минут.

На языке Python реализованы библиотека NumPy [15], необходимая для обработки массивов, предназначенная для эффективной работы с большими многомерными массивами произвольных записей без потери слишком большой скорости и pandas [12] – программная библиотека для обработки и анализа данных. Работа pandas с данными строится поверх библиотеки NumPy. Для построения графиков использовалась библиотека pyplot [14], которая предоставляет собой процедурный интерфейс к объектно-ориентированной библиотеке построения графиков matplotlib [13].

Первым этапом для дальнейшей обработки данных является предобработка данных. Поскольку большинство признаков не несут в себе необходимой информации для дальнейшей обработки, а именно номер ученика, вариант работы и литер класса, их необходимо исключить. Помимо всего, необходимо заменить категориальные переменные на числовые, поскольку в номерах заданий встречается символ «X» обозначающий, что ученик не приступил к выполнению задания, что соответствует 0 баллов. Также стоит учитывать то, что не все ученики приняли участие в контрольных работах, такие ученики отмечены как

«отсутствовал», и их необходимо исключить, так как они могут сильно повлиять на результаты анализа.

Для оценки всей работы были предоставлены КИМы со шкалой перевода и возможным количеством баллов для каждого задания по каждому предмету. Для того, чтобы перевести первичные баллы в оценку, необходимо проверить, действительно ли были набраны соответствующие баллы. Поскольку информация об учениках заносилась в систему ФИС ОКО вручную учителями, возможны ошибки, когда данные имеют противоречивые значения. Проверку можно провести с помощью функции `value_counts()`, которая возвращает число уникальных значений в массиве. Если в результате обработки нет значений, которые выходят за предел предполагаемой шкалы, то данные верны.

Также может возникнуть ситуация, когда учителя не указывали балл, а оставляли пустое поле. Отсутствующие значения отмечены как NaN, их необходимо заменить с помощью функции `fillna()`, например, медианным значением: `fillna(D.median(axis=0), axis=0)`. Если рассматривать категориальные признаки, такие, как пол, их пропущенные значения можно заменить самыми часто встречаемыми значениями в столбце. Таким образом, в результате предобработки данных, все значения таблицы заполнены и получены следующие результаты, представленные в таблице 2.1.

Таблица 2.1

Число наблюдений в разрезе каждого класса и предмета до и после обработки данных

	Математика		Русский язык	
	До обработки	После обработки	До обработки	После обработки
5 класс	26626	23324	26589	22620
6 класс	25907	22031	25708	21790
7 класс	23857	20236	23801	20235

Снижение числа учащихся с 5 по 7 класс обусловлено демографической

ситуацией в регионе, а снижение размерности данных после обработки – в связи с отсутствием некоторого числа учеников на итоговых работах из-за возможной болезни либо другой уважительной причины.

### 2.3. Сравнительный анализ и визуализация данных

Визуализация аналитических данных – представление информации в виде рисунков, диаграмм, с использованием интерактивных возможностей как для получения результатов, так и для использования в качестве исходных данных для дальнейшего анализа.

Информативная визуализация – одна из важнейших задач анализа данных. Она может быть частью процесса исследования, например, применяться для выявления выбросов, определения необходимых преобразований данных или поиска идей для построения моделей [9].

Всероссийские проверочные работы, предлагавшиеся ученикам в каждом классе, были составлены по единой для каждого из классов спецификации, имели стандартизированную систему оценивания (таблица 2.2).

Таблица 2.2

Средние баллы участников ВПР

	Математика		Русский язык	
	Средний балл	Максимальный балл	Средний балл	Максимальный балл
5 класс	10.65	20	21.57	45
6 класс	9.23	16	24	51
7 класс	7.29	19	23.33	47

С учетом того, что максимальный балл в работах был разным, представленные значения свидетельствуют о более низких результатах

выполнения работ в 7 классе по сравнению с 5 и 6 классами. Ниже дано более подробное представление результатов.

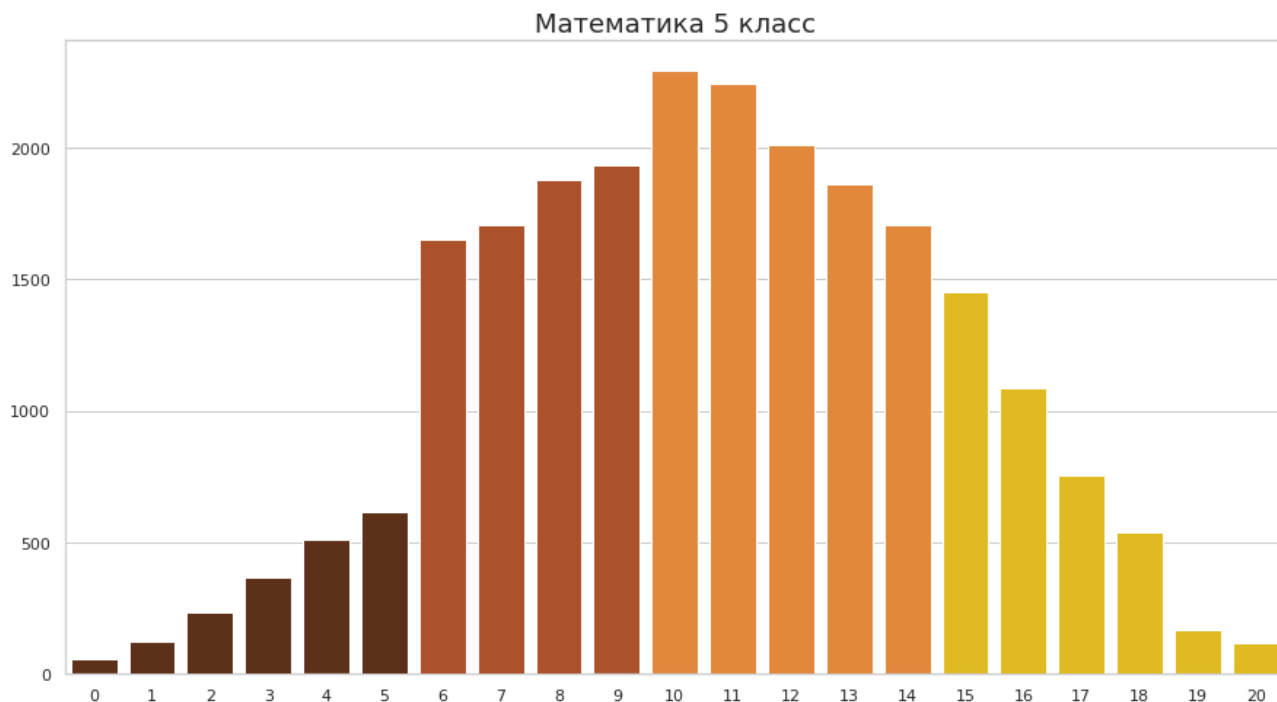


Рисунок 2.2. Распределение баллов ВПР в 5 классе по математике

Данное распределение баллов свидетельствует о том, что сложность работы соответствовала познавательным возможностям пятиклассников. Этот вывод подтверждается также данными, представленными в таблице 2.3: большинство участников показали удовлетворительные и хорошие результаты.

Таблица 2.3

Границы отметок ВПР по математике в 5 классе

Отметка ВПР	«2»	«3»	«4»	«5»
Первичные баллы	0-5	6-9	10-14	15-20
Доля участников (в %)	8,20	30,73	43,35	17,72

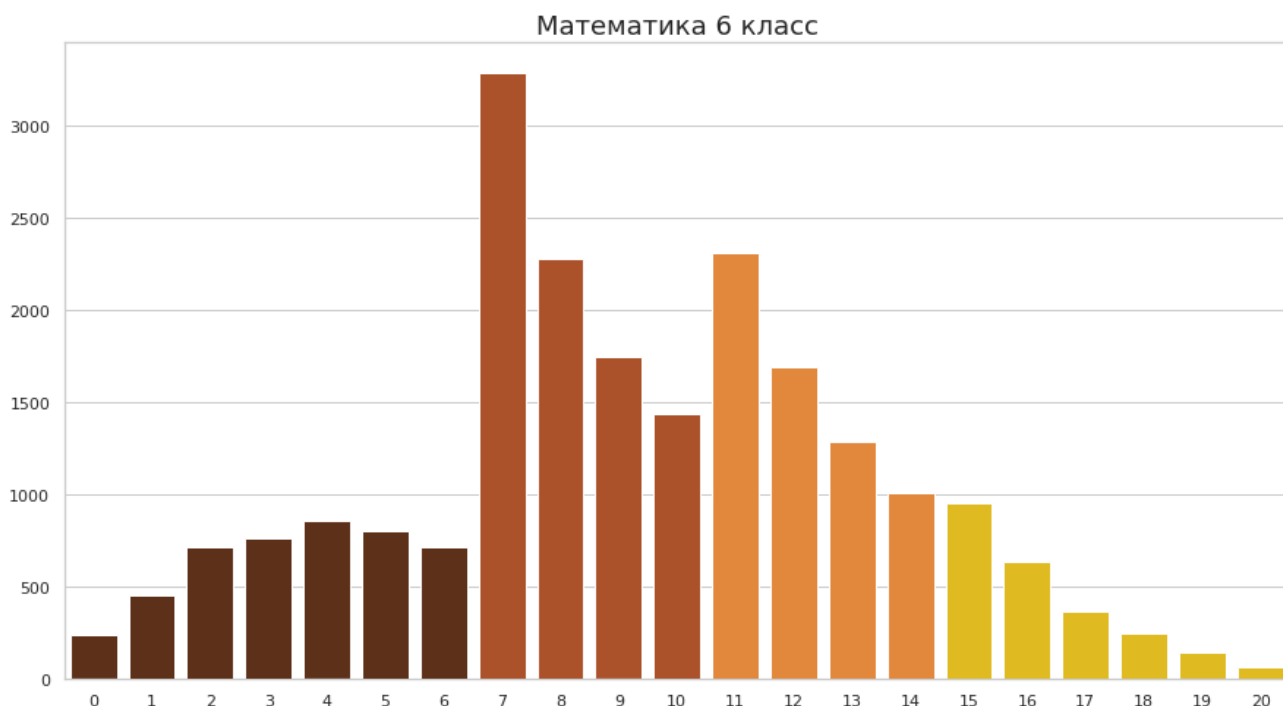


Рисунок 2.3. Распределение баллов ВПР в 6 классе по математике

Характер распределения баллов в 6 классе имеет значительные изменения в сравнении с 5 классом – график приобретает два выделяющихся столбца, которые стоят на границе с неудовлетворительной оценкой в пользу удовлетворительной. Таким образом, распределение баллов может свидетельствовать о необъективности результатов.

Таблица 2.4

Границы отметок ВПР по математике в 6 классе

Отметка ВПР	«2»	«3»	«4»	«5»
Первичные баллы	0-6	7-10	11-14	15-20
Доля участников (в %)	20,68	39,72	28,60	11

Большинство участников исследования в 6 классе показали удовлетворительные и хорошие результаты, однако весьма высока доля участников (более 20%), набравших от 0 до 6 баллов, что существенно выше аналогичного показателя для 5 класса.

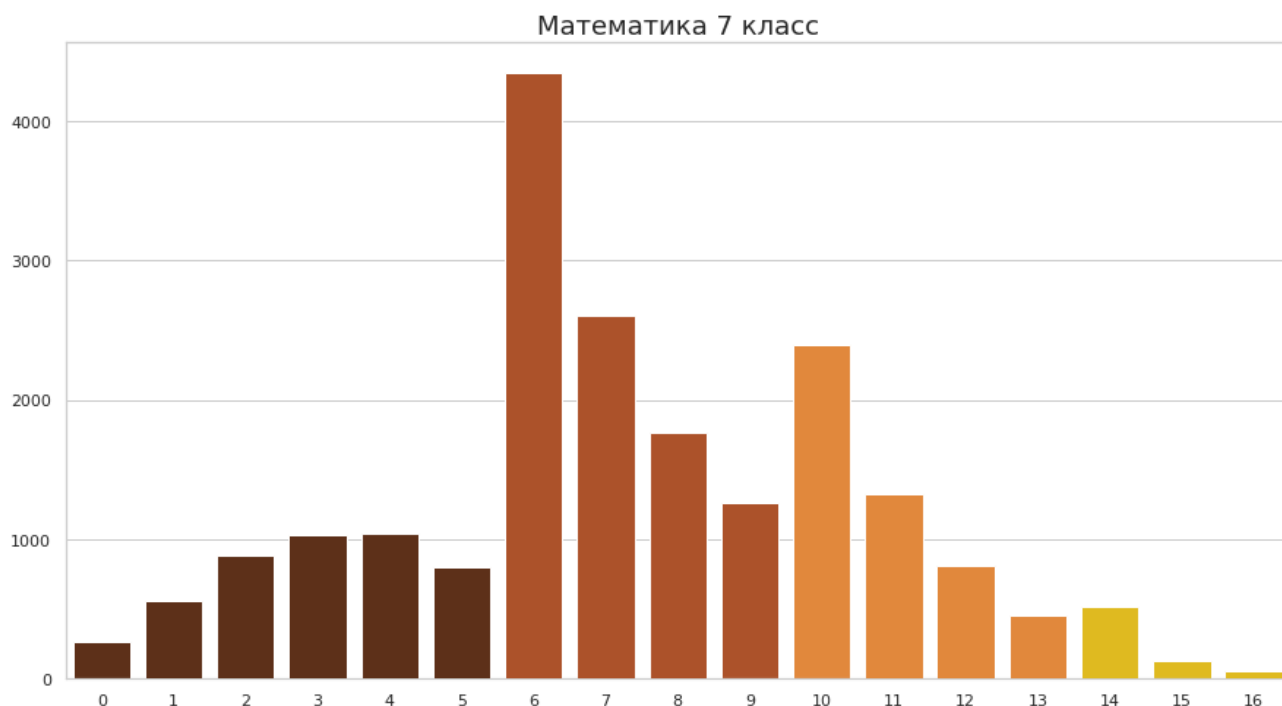


Рисунок 2.4. Распределение баллов ВПР в 7 классе по математике

Кривая распределения баллов несколько смещена влево, что свидетельствует о наличии затруднений участников при выполнении заданий. Это наблюдение подтверждается и данными таблицы 2.5.

Таблица 2.5

Границы отметок ВПР по математике в 7 классе

Отметка ВПР	«2»	«3»	«4»	«5»
Первичные баллы	0-5	6-9	10-13	14-16
Доля участников (в %)	22,63	49,22	24,62	3,52

Среди обучающихся 7 класса доля набравших высокие баллы ВПР меньше, чем среди обучающихся 5 и 6 классов, а доля набравших балл, соответствующий неудовлетворительной отметке, существенно выше. Также отмечены весьма высокие показатели числа учеников, которые выполнили работу буквально на границе с неудовлетворительной оценкой – столбцы диаграммы сильно выделяются среди всей работы.

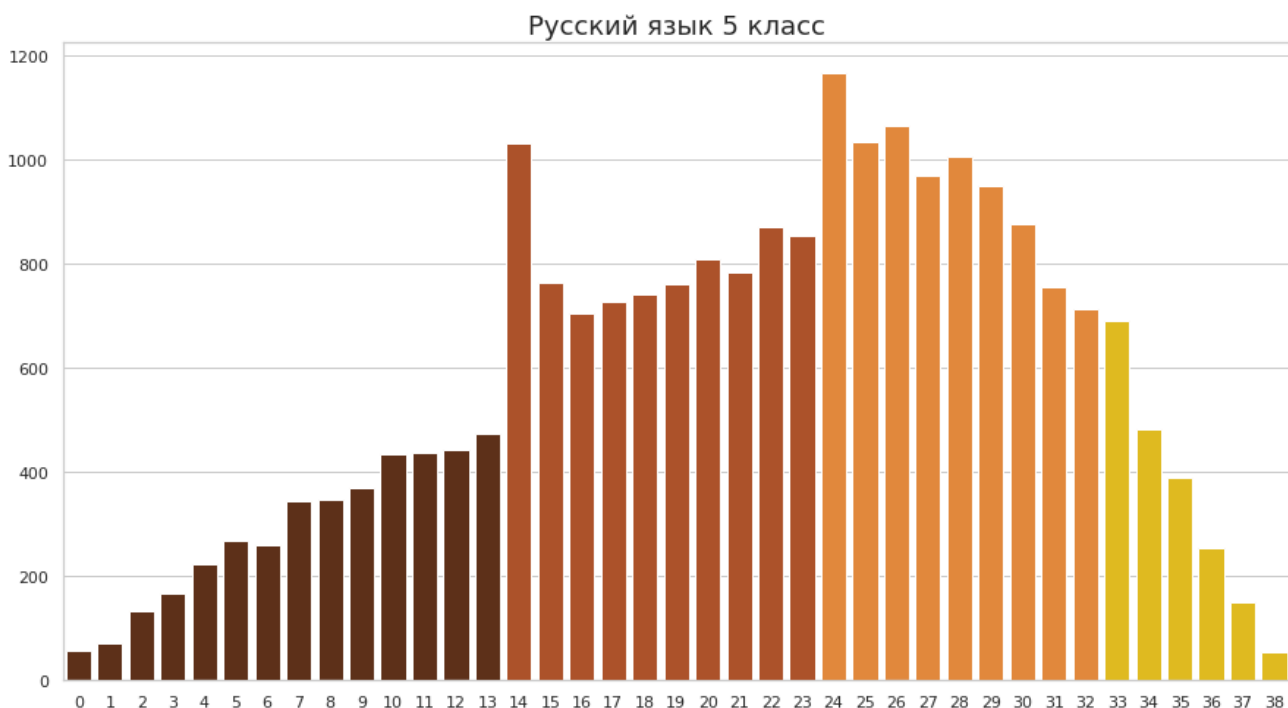


Рисунок 2.5. Распределение баллов ВПР в 5 классе по русскому языку

Рассмотрим результаты по русскому языку. На данном рисунке 2.5 видно, что ученики пишут работы на удовлетворительные и хорошие результаты в большинстве случаев. Этот вывод подтверждается также данными, представленными в таблице 2.6. Такую же ситуацию мы наблюдали и у 5 класса с результатами по математике.

Таблица 2.6

Границы отметок ВПР по русскому языку в 5 классе

Отметка ВПР	«2»	«3»	«4»	«5»
Первичные баллы	0-13	14-23	24-32	33-38
Доля участников (в %)	15,75	36,45	38,66	9,14





Рисунок 2.6. Распределение баллов ВПР в 6 классе по русскому языку

Характер распределения баллов в 6 классе имеет явные границы с завышенными шкалами относительно других – на границах отметок «3», «4» и «5». Данную ситуацию наблюдалась также в результатах по математике среди 6 классов.

Таблица 2.7

Границы отметок ВПР по русскому языку в 6 классе

Отметка ВПР	«2»	«3»	«4»	«5»
Первичные баллы	0-17	18-28	29-38	39-45
Доля участников (в %)	15,75	36,45	38,66	9,14

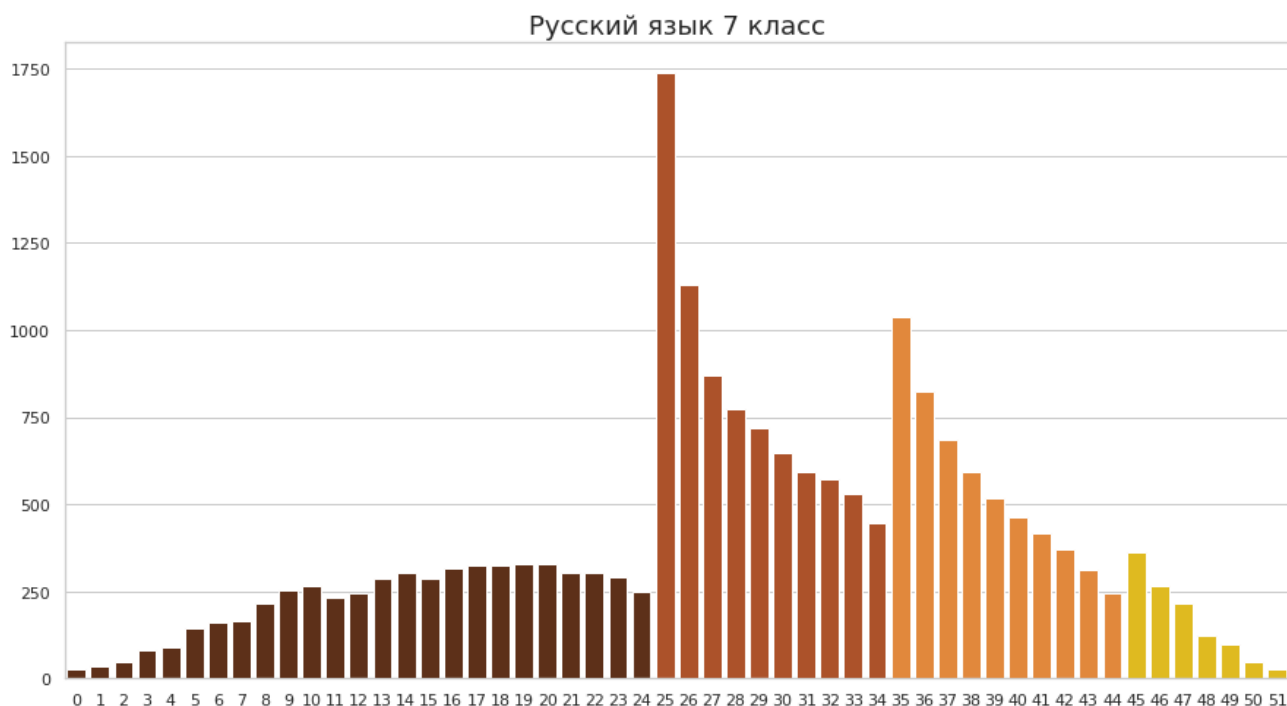


Рисунок 2.7. Распределение баллов ВПР в 7 классе по русскому языку

Аналогично предыдущему классу, распределение содержит явно выраженные пики. Также количество неудовлетворительных оценок увеличивается. Это наблюдение подтверждается и данными таблицы 2.8.

Таблица 2.8

Границы отметок ВПР по русскому языку в 7 классе

Отметка ВПР	«2»	«3»	«4»	«5»
Первичные баллы	0-24	25-34	35-44	45-51
Доля участников (в %)	27,75	39,62	27	5,63

В результате сравнительного анализа было выявлено, что в Алтайском крае наблюдаются ситуации снижения уровня подготовки учеников с 5 по 7 класс. Также можно отметить то, что заявленные распределения первичных баллов имеют явно выраженные «пики», которые могут свидетельствовать о возможных ситуациях, когда ученикам завышают оценки. Необходимо провести оценку объективности результатов образовательных организаций.

### **3. Оценка качества образования и объективности результатов образовательных организаций**

#### **3.1. Применение статистических методов**

Методика расчета показателя «Уровень объективности оценки образовательных результатов в субъекте Российской Федерации» предоставлена Федеральным институтом оценки качества образования. Она представляет собой математический алгоритм маркировки образовательных организаций: завышенные значения среднего балла по русскому языку и математике среди 5-7 классов [10].

Данная методика предполагает выявление школ с завышенными значениями для получения объективной обстановки в регионах страны. Полученные данные будут направлены в региональные органы управления для принятия управленческих решений и контроля выявленных образовательных организаций.

Для того, чтобы провести данный анализ, необходимо учитывать то, что в Алтайском крае есть школы с углубленным изучением предметов. В данных образовательных организациях возможны ситуации, когда ученики более подготовлены к выполнению работ. Данные образовательные организации необходимо исключить из выборки.

В результате обработки данных получилось, что в 2020 году приняли участие 390 общеобразовательных организаций. Данные о числе учеников представлены в таблице 3.1.

Число учеников, написавших ВПР в общеобразовательных школах Алтайского края

	5 класс	6 класс	7 класс
Математика	11864	11329	10335
Русский язык	11555	11201	10321

Методика выявления завышенных значений среднего балла по каждой процедуре ВПР по русскому языку и математике:

**1 этап:** Строятся доверительные интервалы среднего балла по регионам:

Доверительный интервал (для выборочного среднего балла):

$$\left[ \bar{X} - t_a * \frac{S}{\sqrt{n}}; \bar{X} + t_a * \frac{S}{\sqrt{n}} \right]$$

$t_a = 1.96$  – (на уровне достоверности 95%);

$n$  – количество участников в регионе;

$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$  – выборочное среднее итогового балла,  $X_i$  – балл  $i$ -го участника;

$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$  – выборочное стандартное отклонение балла по региону.

**2 этап:** Строятся доверительные интервалы среднего балла по ОО:

$$\left[ \bar{x} - t_a * \frac{S}{\sqrt{m}}; \bar{x} + t_a * \frac{S}{\sqrt{m}} \right]$$

$m$  – количество участников в ОО;

$\bar{x} = \frac{\sum_{i=1}^m x_i}{m}$  – выборочное среднее итогового балла по ОО,  $x_i$  – балл  $i$ -го участника;

$S = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2}{m-1}}$  – выборочное стандартное отклонение балла по ОО;

**3 этап:** Выявление ОО, левая (нижняя) граница доверительных интервалов которых находится правее, чем правая (верхняя) граница доверительного интервала среднего балла по региону.

**4 этап:** Результат представляется в виде списка ОО. Число ОО представлено в таблице 3.2.

Таблица 3.2

Число ОО с необъективными результатами.

	5 класс	6 класс	7 класс
Математика	49	78	82
Русский язык	75	91	88

Таким образом выявлено, что в Алтайском крае действительно наблюдаются ОО с завышенными и/или необъективными результатами.

### **3.2. Практическое обоснование результатов исследования в предметной области**

После завершения процедуры написания ВПР Министерство Образования Алтайского края составляет базовый отчет, который содержит основные сведения о проведении работ. В том числе поступает список ОО, в которых замечены завышенные результаты среди 5-6 классах. В ходе обработки результатов было выявлено, что не все ОО имеют завышенные значения – распределение баллов без явных перевесов в сторону более хорошей отметки, либо отметки по ОО имеют наоборот заниженные значения – что не входит в категорию завышенных результатов (рисунок 3.1).

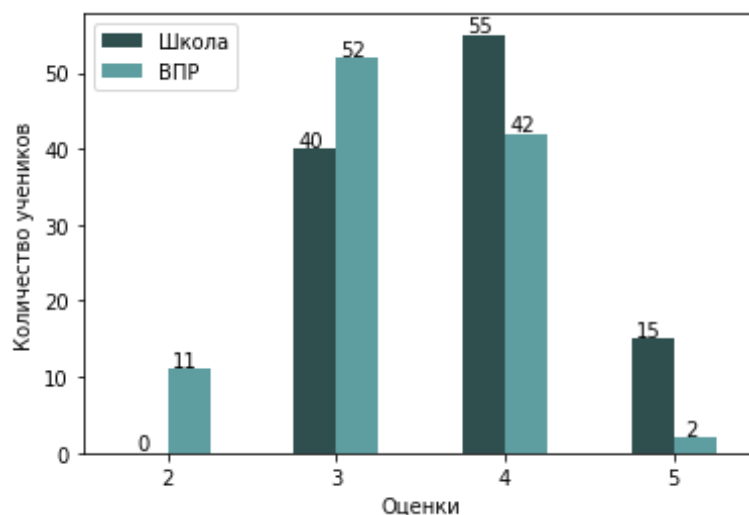


Рисунок 3.1. Пример школы, которая вошла в категорию завышенных, но при этом баллы в ней явно занижены относительно отметок по школе

Также отмечается, что в данном отчете были отмечены не все ОО, в которых действительно были завышенные результаты относительно оценок по школе. ОО, выявленные предложенным методом построения доверительных интервалов, были рассмотрены отдельно, и в сравнении с отметками по школе в них наблюдаются ситуации, когда число «отличных» отметок по школе ниже числа «отличных» отметок за выполнение ВПР (рисунок 3.2).

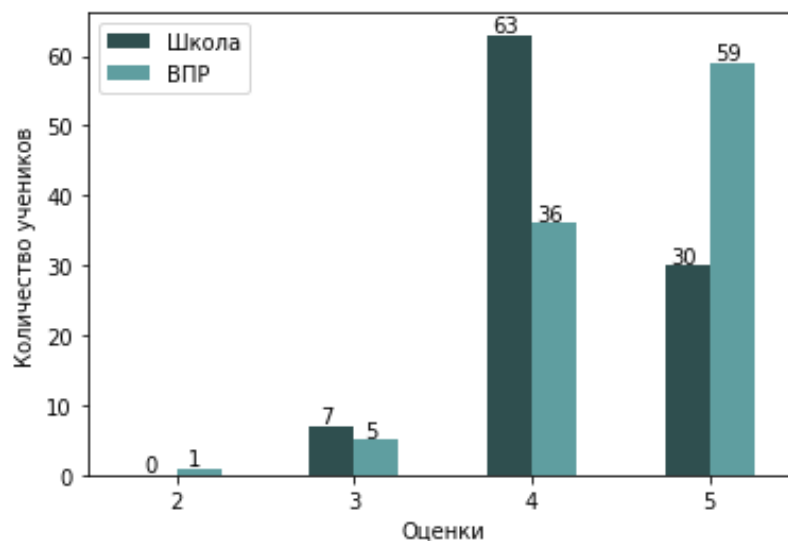


Рисунок 3.2. Пример школы, которая выявлена методикой

Таким образом, были выявлены ОО, в которых результаты выполнения ВПР существенно выше средних результатов по региону, при этом ОО не является лицеем или гимназией с углубленным изучением данных предметов. Данная ситуация возможна в таких случаях, когда проверяющие намеренно завышают отметки ученикам, чтобы повысить средний балл по школе, тем самым испортив объективную оценку качества образования в Алтайском крае в целом. Такие случаи необходимо устранять путем введения полного контроля проведения ВПР в пунктах ОО – независимые эксперты, системы видеонаблюдения.

Полученные в данной работе результаты позволяют провести более обширную оценку качества образования в Алтайском крае, а также принять управленческие решения для повышения объективности результатов в муниципальных образовательных организациях.

Созданная в данной работе программа позволит аналитику получать результаты проведения работ в полуавтоматическом режиме за определенный класс и предмет, что в дальнейшем может помочь ускорить процесс получения результатов для составления итоговых отчетов проведения процедуры ВПР.

## Заключение

Итоговый анализ данных позволяет предоставить содержательную информацию о результатах проведения Всероссийских проверочных работ как в целом по Алтайскому краю, так и по отдельным образовательным организациям. Результаты анализа могут быть использованы специалистами отдела мониторинга и оценки качества образования в Алтайском крае для составления итоговых отчетов.

При написании работы были изучены различные методы и средства интеллектуального анализа данных, методы статистической обработки и визуализации данных.

В ходе работы были выполнены следующие задачи:

- рассмотрены распределения первичных баллов и их перевод в соответствующие отметки по шкале перевода;
- проведена оценка динамики изменения результатов учеников с 5 по 7 класс;
- определение факторы, оказавшие наибольшее влияние на результат предметной подготовки выпускников;
- выявлены образовательные организации, в которых результаты были завышены или необъективны;
- проведена оценка качества образования в Алтайском крае;
- разработаны вычислительные методы для проведения анализа и визуализации данных.

В перспективе развития данной работы планируется разработка собственных методов оценки качества образования, которые могут предоставить более детальную информацию о проведенных работах в последующих годах.



## Литература

1. Чубукова И.А. Data Mining: учебное пособие. — М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. — С. 208.
2. Баталова Ю. А. Анализ результатов оценочных процедур как способ совершенствования профессиональных компетенций педагогов // Наука и школа, 2019. — № 2 — С. 76.
3. Лях Ю. А. Совершенствование системы оценки качества общего образования // Вестн. Костромского гос. Ун-та. Сер.: Педагогика. Психология. Социокинетика. — 2016. — № 2. — С. 11–13.
4. А. В. Мендель Всероссийские проверочные работы как инструмент независимой оценки и управления качеством образования социологический аспект // Власть и управление на Востоке России. — 2018 — № 2 — С.103 – 104.
5. Правовое регулирование информационных отношений в области государственной и коммерческой тайны, персональных данных : учебное пособие / О. В. Ахрамеева, И. Ф. Дедюхина, О. В. Жданова, Н. В. Мирошниченко. — Ставрополь : СтГАУ, 2015. — 59 с. — С. 52.
6. Приказ Роскомнадзора от 5.09.2013 № 996 «Об утверждении требований и методов по обезличиванию персональных данных».
7. Приказ Министерства образования и науки Алтайского края от 26.02.2020 № 312 «О проведении мониторинга качества подготовки обучающихся общеобразовательных организаций Алтайского края в 2020 году».
8. Аналитические материалы по результатам проведения Национального исследования качества математического образования в 5-7 классах // Федеральная служба по надзору в сфере образования и науки — 2014-2015 — С. 4.

9. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — С. 273.
10. ФИОКО: Методика расчета показателя «Уровень объективности оценки образовательных результатов в субъекте Российской Федерации — ФГБУ Федеральный институт оценки качества образования, 2019.
11. Документация языка программирования Python [электронный ресурс] URL: <https://docs.python.org/3/>
12. Документация библиотеки pandas [электронный ресурс] URL: <https://pandas.pydata.org/docs/>
13. Документация библиотеки matplotlib [электронный ресурс] URL: <https://matplotlib.org/>
14. Документация библиотеки pyplot [электронный ресурс] URL: [https://matplotlib.org/stable/api/\\_as\\_gen/matplotlib.pyplot.html](https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.html)
15. Документация библиотеки NumPy [электронный ресурс] URL: <https://numpy.org/doc/>
16. Документация Excel [электронный ресурс] URL: <https://support.microsoft.com/en-us/excel>
17. Приказ Министерства образования и науки Российской Федерации (Минобрнауки России) [электронный ресурс] URL: <https://xn--80aakec5bilkue.xn--33-6kcadhwnl3cfdx.xn--p1ai/upload/iblock/4a5/4a51f06453982b3592a84196a635985c.pdf>
18. Сайт Федеральной информационной системы оценки качества образования [электронный ресурс] URL: <https://lk-fisoko.obrnadzor.gov.ru/>

## Приложение

Ссылка на проект: <https://github.com/Helintaine/Diplom.git>

```
# Загружаем данные и редактируем их

import numpy as np
import pandas as pd
import scipy.stats as st
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
from math import *
pd.set_option('display.max_columns', 50)
pd.set_option('display.max_rows', 100)

from google.colab import drive
drive.mount('/content/drive')

# Заполнение отсутствующих значений
def editToIntVPR(df):
    df['Unnamed: 5'] = df['Unnamed: 5'].fillna(0)
    df['Unnamed: 6'] = df['Unnamed: 6'].fillna(0)
    df['Unnamed: 7'] = df['Unnamed: 7'].fillna(0)
    df['Unnamed: 8'] = df['Unnamed: 8'].fillna(0)

def coerce_df_columns_to_numeric(df, column_list):
    df[column_list] = df[column_list].apply(pd.to_numeric, errors='coerce')
    df[column_list] = df[column_list].fillna(0)

# Файл с несоответствующими оценками в школах от министерства образования
vpr = pd.read_excel('/content/drive/MyDrive/Работа/ВІР2020/vpr2020.xls', header=3)
```

```

editToIntVPR(vpr)
vpr = vpr[["Unnamed: 3", "Unnamed: 4",
         "Unnamed: 7", "Unnamed: 8"]].rename(columns = {
         "Unnamed: 3" : "Школа",
         "Unnamed: 4" : "Код школы",
         "Unnamed: 5" : "Русский язык 5(4) класс",
         "Unnamed: 6" : "Русский язык 6(5) класс",
         "Unnamed: 7" : "Математика 5(4) класс",
         "Unnamed: 8" : "Математика 6(5) класс"
         })

# Вывод числа школ
def printAll(df):
    bottom, upper = mean_confidence_interval(df)
    result = editColumns(df)
    a, listSchool = region(result, upper)
    print("Кол-во школ:", a)
    # print(listSchool)
    return listSchool

vpr_hight = pd.read_excel('/content/drive/MyDrive/Работа/ВПР2020/edit.xlsx')
vpr_hight = vpr_hight[vpr_hight['yes'] == 1.0]
columns_school = vpr_hight["sch"].reset_index()
del columns_school['index']

def schoolEdit(columns_school,df):
    for row in columns_school.iterrows():
        df = df.loc[~df['user'].isin([row[1][0])]
    return df

""""# 1 этап""""

```

```

def mean_confidence_interval(df, confidence=0.95):
    df = df['Сумма'].values
    n = len(df)
    m, se = np.mean(df), st.sem(df)
    h = se * st.t.ppf((1+confidence) / 2., n-1)
    return m-h, m+h

""""# 2 этап""""

def editColumns(df):
    columns_number = df[["Код школы", "Сумма"]] # выбираем два столбца с номером школы и
суммой баллов по ВПР

    # СРЕДНИЙ БАЛЛ
    res = columns_number.groupby('Код школы').mean()[['Сумма']].reset_index() # подсчет
среднего балла по школам
    res = res[["Код школы", "Сумма"]].rename(columns = {"Сумма" : "x"}) # Переименование
столбцов

    # СТАНДАРТНОЕ ОТКЛОНЕНИЕ
    res2 = columns_number.groupby('Код школы').std()[['Сумма']].reset_index() # подсчет
стандартного отклонения по школам
    res2 = res2[["Код школы", "Сумма"]].rename(columns = {"Сумма" : "S"}) # переименование
столбцов

    result1 = pd.merge(res, res2, left_on='Код школы', right_on='Код школы') # объединение двух
таблиц

    # ЧИСЛО УЧЕНИКОВ В ШКОЛЕ
    column = df['Код школы'].value_counts().reset_index() # подсчет числа учеников в каждой
школе
    column = column[["Код школы", "index"]].rename(columns = {"Код школы" : "m", "index" :

```

```

"number"}) # переименование столбцов

result = pd.merge(result1, column, left_on='Код школы', right_on='number') # объединение двух
таблиц
del result['number'] # удалить лишний столбец

return result

"""# 3 этап"""

def region(df, upper):
    count = 0
    listSchool = []
    for i,row in df.iterrows():
        m, x, S = row["m"], row["x"], row["S"]
        result = x - 1.96 * (S/sqrt(m))
        if (result > upper):
            listSchool.append(row["Код школы"])
            count += 1
    return count, listSchool

"""# Математика 5 класс"""

df = pd.read_csv('/content/drive/MyDrive/Работа/ВПР2020/По
классам/5class/3228_VPR_2020_Mat5_reg22.csv', delimiter=';',
                low_memory=False, index_col=0)

df = schoolEdit(columns_school,df)

# Удаляем данные о учениках, которые отсутствовали на сдаче ВПР
df = df[df['var-1'] != 'отсутствовал']
df = df[df['otc'] != 'нет отметки']

```

```

# Проверка на отсутствующим данные в столбцах
nan_cols = [i for i in df.columns if df[i].isnull().any()]

# Переименовываем столбцы по номерам заданий
df = df[["user", "a-1-grade", "a-2-grade", "a-3-grade", "a-4-grade", "a-5-grade",
        "a-6-grade", "a-7-grade", "a-8-grade", "a-9-grade", "a-10-grade",
        "a-11-grade", "a-12-grade", "a-13-grade", "a-14-grade", "a-15-grade"
        ]].rename(columns = {"user": "Код школы",
                             "a-1-grade" : "Задание 1",
                             "a-2-grade" : "Задание 2",
                             "a-3-grade" : "Задание 3",
                             "a-4-grade" : "Задание 4",
                             "a-5-grade" : "Задание 5",
                             "a-6-grade" : "Задание 6",
                             "a-7-grade" : "Задание 7",
                             "a-8-grade" : "Задание 8",
                             "a-9-grade" : "Задание 9",
                             "a-10-grade" : "Задание 10",
                             "a-11-grade" : "Задание 11",
                             "a-12-grade" : "Задание 12",
                             "a-13-grade" : "Задание 13",
                             "a-14-grade" : "Задание 14",
                             "a-15-grade" : "Задание 15"})

# Предобработка исходных данных (отсутствующим значениям присваиваем нули)
coerce_df_columns_to_numeric(df, ["Задание 1", "Задание 2", "Задание 3",
                                   "Задание 4", "Задание 5", "Задание 6",
                                   "Задание 7", "Задание 8", "Задание 9",
                                   "Задание 10", "Задание 11", "Задание 12",
                                   "Задание 13", "Задание 14", "Задание 15"])

```

```

# Добавление столбца с суммы баллов для каждого ученика
df = df.assign(Сумма = df.sum(axis=1))

listSchool5 = printAll(df)
listSchool5

""""# Математика 6 класс""""

df = pd.read_csv('mat6_reg22.csv', delimiter=';', low_memory=False, index_col=0)

df = schoolEdit(columns_school,df)

# Удаляем данные о учениках, которые отсутствовали на сдаче ВПР
df = df[df['var-1'] != 'отсутствовал']
df = df[df['otc'] != 'нет отметки']

# Проверка на отсутствующим данные в столбцах
nan_cols = [i for i in df.columns if df[i].isnull().any()]

# Переименовываем столбцы по номерам заданий
df = df[["user", "a-1-grade", "a-2-grade", "a-3-grade", "a-4-grade", "a-5-grade",
        "a-6-grade", "a-7-grade", "a-8-grade", "a-9-grade", "a-10-grade",
        "a-11-grade", "a-12-grade", "a-13-grade", "a-14-grade", "a-15-grade",
        "a-16-grade"]].rename(columns = {"user": "Код школы",
        "a-1-grade" : "Задание 1",
        "a-2-grade" : "Задание 2",
        "a-3-grade" : "Задание 3",
        "a-4-grade" : "Задание 4",
        "a-5-grade" : "Задание 5",
        "a-6-grade" : "Задание 6",
        "a-7-grade" : "Задание 7",

```



```
"a-8-grade" : "Задание 8",  
"a-9-grade" : "Задание 9",  
"a-10-grade" : "Задание 10",  
"a-11-grade" : "Задание 11",  
"a-12-grade" : "Задание 12",  
"a-13-grade" : "Задание 13",  
"a-14-grade" : "Задание 14",  
"a-15-grade" : "Задание 15",  
"a-16-grade" : "Задание 16"}))
```

```
# Предобработка исходных данных (отсутствующим значениям присваиваем нули)
```

```
coerce_df_columns_to_numeric(df, ["Задание 1", "Задание 2", "Задание 3",  
    "Задание 4", "Задание 5", "Задание 6",  
    "Задание 7", "Задание 8", "Задание 9",  
    "Задание 10", "Задание 11", "Задание 12",  
    "Задание 13", "Задание 14", "Задание 15",  
    "Задание 16"])
```

```
# Добавление столбца с суммы баллов для каждого ученика
```

```
df = df.assign(Сумма = df.sum(axis=1))
```

```
listSchool6 = printAll(df)
```

```
""""# Математика 7 класс""""
```

```
df = pd.read_csv('3226_VPR_2020_Mat7_reg22.csv', delimiter=';',  
    low_memory=False, index_col=0)
```

```
df = schoolEdit(columns_school, df)
```

```

# Удаляем данные о учениках, которые отсутствовали на сдаче ВПР
df = df[df['var-1'] != 'отсутствовал']
df = df[df['otc'] != 'нет отметки']

# Проверка на отсутствующим данные в столбцах
nan_cols = [i for i in df.columns if df[i].isnull().any()]

# Переименовываем столбцы по номерам заданий
df = df[["user", "a-1-grade", "a-2-grade", "a-3-grade", "a-4-grade", "a-5-grade",
        "a-6-grade", "a-7-grade", "a-8-grade", "a-9-grade", "a-10-grade",
        "a-11-grade", "a-12-grade", "a-13-grade"]].rename(columns = {
    "user": "Код школы",
    "a-1-grade": "Задание 1",
    "a-2-grade": "Задание 2",
    "a-3-grade": "Задание 3",
    "a-4-grade": "Задание 4",
    "a-5-grade": "Задание 5",
    "a-6-grade": "Задание 6",
    "a-7-grade": "Задание 7",
    "a-8-grade": "Задание 8",
    "a-9-grade": "Задание 9",
    "a-10-grade": "Задание 10",
    "a-11-grade": "Задание 11",
    "a-12-grade": "Задание 12",
    "a-13-grade": "Задание 13"
})

# Предобработка исходных данных (отсутствующим значениям присваиваем нули)
coerce_df_columns_to_numeric(df, ["Задание 1", "Задание 2", "Задание 3",
    "Задание 4", "Задание 5", "Задание 6",
    "Задание 7", "Задание 8", "Задание 9",
    "Задание 10", "Задание 11", "Задание 12",

```

"Задание 13")

```
# Добавление столбца с суммы баллов для каждого ученика
```

```
df = df.assign(Сумма = df.sum(axis=1))
```

```
listSchool7 = printAll(df)
```

```
for i in listSchool5:
```

```
    if i in listSchool6:
```

```
        if i in listSchool7:
```

```
            print(i)
```

```
""""# Русский язык 5 класс""""
```

```
df = pd.read_csv('/content/drive/MyDrive/Работа/ВПР2020/По
```

```
классам/5class/3240_VPR_2020_Rus5_reg22.csv', delimiter=';',
```

```
            low_memory=False, index_col=0)
```

```
df = schoolEdit(columns_school,df)
```

```
# Удаляем данные о учениках, которые отсутствовали на сдаче ВПР
```

```
df = df[df['var-1'] != 'отсутствовал']
```

```
df = df[df['otc'] != 'нет отметки']
```

```
# Проверка на отсутствующим данные в столбцах
```

```
nan_cols = [i for i in df.columns if df[i].isnull().any()]
```

```
# Переименовываем столбцы по номерам заданий
```

```
df = df[["user", "a-1-grade", "a-2-grade", "a-3-grade", "a-4-grade", "a-5-grade",
```

```
        "a-6-grade", "a-7-grade", "a-8-grade", "a-9-grade", "a-10-grade",
```

```
        "a-11-grade", "a-12-grade", "a-13-grade", "a-14-grade", "a-15-grade",
```

```
        "a-16-grade", "a-17-grade", "a-18-grade", "a-19-grade", "a-20-grade"]
```

```
]].rename(columns = {"user": "Код школы",  
                    "a-1-grade" : "Задание 1",  
                    "a-2-grade" : "Задание 2",  
                    "a-3-grade" : "Задание 3",  
                    "a-4-grade" : "Задание 4",  
                    "a-5-grade" : "Задание 5",  
                    "a-6-grade" : "Задание 6",  
                    "a-7-grade" : "Задание 7",  
                    "a-8-grade" : "Задание 8",  
                    "a-9-grade" : "Задание 9",  
                    "a-10-grade" : "Задание 10",  
                    "a-11-grade" : "Задание 11",  
                    "a-12-grade" : "Задание 12",  
                    "a-13-grade" : "Задание 13",  
                    "a-14-grade" : "Задание 14",  
                    "a-15-grade" : "Задание 15",  
                    "a-16-grade" : "Задание 16",  
                    "a-17-grade" : "Задание 17",  
                    "a-18-grade" : "Задание 18",  
                    "a-19-grade" : "Задание 19",  
                    "a-20-grade" : "Задание 20"})
```

```
# Предобработка исходных данных (отсутствующим значениям присваиваем нули)
```

```
coerce_df_columns_to_numeric(df, ["Задание 1", "Задание 2", "Задание 3",  
                                "Задание 4", "Задание 5", "Задание 6",  
                                "Задание 7", "Задание 8", "Задание 9",  
                                "Задание 10", "Задание 11", "Задание 12",  
                                "Задание 13", "Задание 14", "Задание 15",  
                                "Задание 16", "Задание 17", "Задание 18",  
                                "Задание 19", "Задание 20"])
```

```
# Добавление столбца с суммы баллов для каждого ученика
```

```

df = df.assign(Сумма = df.sum(axis=1))
df
# listSchool5 = printAll(df)

""""# Русский язык 6 класс""""

df = pd.read_csv('3239_VPR_2020_Rus6_reg22.csv', delimiter=';',
                low_memory=False, index_col=0)

df = schoolEdit(columns_school,df)

# Удаляем данные о учениках, которые отсутствовали на сдаче ВПР
df = df[df['var-1'] != 'отсутствовал']
df = df[df['otc'] != 'нет отметки']

# Проверка на отсутствующим данные в столбцах
nan_cols = [i for i in df.columns if df[i].isnull().any()]

# Переименовываем столбцы по номерам заданий
df = df[["user", "a-1-grade", "a-2-grade", "a-3-grade", "a-4-grade", "a-5-grade",
        "a-6-grade", "a-7-grade", "a-8-grade", "a-9-grade", "a-10-grade",
        "a-11-grade", "a-12-grade", "a-13-grade", "a-14-grade", "a-15-grade",
        "a-16-grade", "a-17-grade", "a-18-grade", "a-19-grade", "a-20-grade",
        "a-21-grade"]].rename(columns = {"user": "Код школы",
        "a-1-grade" : "Задание 1",
        "a-2-grade" : "Задание 2",
        "a-3-grade" : "Задание 3",
        "a-4-grade" : "Задание 4",
        "a-5-grade" : "Задание 5",
        "a-6-grade" : "Задание 6",
        "a-7-grade" : "Задание 7",
        "a-8-grade" : "Задание 8",

```

```
"a-9-grade" : "Задание 9",  
"a-10-grade" : "Задание 10",  
"a-11-grade" : "Задание 11",  
"a-12-grade" : "Задание 12",  
"a-13-grade" : "Задание 13",  
"a-14-grade" : "Задание 14",  
"a-15-grade" : "Задание 15",  
"a-16-grade" : "Задание 16",  
"a-17-grade" : "Задание 17",  
"a-18-grade" : "Задание 18",  
"a-19-grade" : "Задание 19",  
"a-20-grade" : "Задание 20",  
"a-21-grade" : "Задание 21"}))
```

```
# Предобработка исходных данных (отсутствующим значениям присваиваем нули)
```

```
coerce_df_columns_to_numeric(df, ["Задание 1", "Задание 2", "Задание 3",  
"Задание 4", "Задание 5", "Задание 6",  
"Задание 7", "Задание 8", "Задание 9",  
"Задание 10", "Задание 11", "Задание 12",  
"Задание 13", "Задание 14", "Задание 15",  
"Задание 16", "Задание 17", "Задание 18",  
"Задание 19", "Задание 20", "Задание 21"])
```

```
# Добавление столбца с суммы баллов для каждого ученика
```

```
df = df.assign(Сумма = df.sum(axis=1))
```

```
listSchool6 = printAll(df)
```

```
""""# Русский язык 7 класс""""
```

```
df = pd.read_csv('3238_VPR_2020_Rus7_reg22.csv', delimiter=';',  
low_memory=False, index_col=0)
```

```

df = schoolEdit(columns_school,df)

# Удаляем данные о учениках, которые отсутствовали на сдаче ВПР
df = df[df['var-1'] != 'отсутствовал']
df = df[df['otc'] != 'нет отметки']

# Переименовываем столбцы по номерам заданий
df = df[["user", "a-1-grade", "a-2-grade", "a-3-grade", "a-4-grade", "a-5-grade",
        "a-6-grade", "a-7-grade", "a-8-grade", "a-9-grade", "a-10-grade",
        "a-11-grade", "a-12-grade", "a-13-grade", "a-14-grade", "a-15-grade",
        "a-16-grade", "a-17-grade", "a-18-grade", "a-19-grade", "a-20-grade",
        "a-21-grade", "a-22-grade", "a-23-grade", "a-24-grade", "a-25-grade"
        ]].rename(columns = {"user": "Код школы",
                             "a-1-grade" : "Задание 1",
                             "a-2-grade" : "Задание 2",
                             "a-3-grade" : "Задание 3",
                             "a-4-grade" : "Задание 4",
                             "a-5-grade" : "Задание 5",
                             "a-6-grade" : "Задание 6",
                             "a-7-grade" : "Задание 7",
                             "a-8-grade" : "Задание 8",
                             "a-9-grade" : "Задание 9",
                             "a-10-grade" : "Задание 10",
                             "a-11-grade" : "Задание 11",
                             "a-12-grade" : "Задание 12",
                             "a-13-grade" : "Задание 13",
                             "a-14-grade" : "Задание 14",
                             "a-15-grade" : "Задание 15",
                             "a-16-grade" : "Задание 16",
                             "a-17-grade" : "Задание 17",
                             "a-18-grade" : "Задание 18",

```

```
"a-19-grade" : "Задание 19",  
"a-20-grade" : "Задание 20",  
"a-21-grade" : "Задание 21",  
"a-22-grade" : "Задание 22",  
"a-23-grade" : "Задание 23",  
"a-24-grade" : "Задание 24",  
"a-25-grade" : "Задание 25"}))
```

```
# Предобработка исходных данных (отсутствующим значениям присваиваем нули)  
coerce_df_columns_to_numeric(df, ["Задание 1", "Задание 2", "Задание 3",  
    "Задание 4", "Задание 5", "Задание 6",  
    "Задание 7", "Задание 8", "Задание 9",  
    "Задание 10", "Задание 11", "Задание 12",  
    "Задание 13", "Задание 14", "Задание 15",  
    "Задание 16", "Задание 17", "Задание 18",  
    "Задание 19", "Задание 20", "Задание 21",  
    "Задание 22", "Задание 23", "Задание 24",  
    "Задание 25"])
```

```
# Добавление столбца с суммы баллов для каждого ученика  
df = df.assign(Сумма = df.sum(axis=1))
```

```
listSchool7 = printAll(df)
```

```
listSchoolRus = []  
for i in listSchool5:  
    if i in listSchool6:  
        if i in listSchool7:  
            listSchoolRus.append(i)
```

```
for i in listSchool5:  
    if i in listSchool6:
```



```

if i in listSchool7:
    print(i)

# Графики

df = pd.read_csv('3240_VPR_2020_Rus5_reg22.csv', delimiter=';',
                skiprows=0, low_memory=False, index_col=0)
# Вывод списка школ, которые вошли в список школ с завышенными оценками
def printSchool(df, listSchool):
    col = df.columns
    test = pd.DataFrame(columns=col)
    for i in listSchool:
        df1 = df[df['user'] == i]
        test = test.append(df1)
    return test
test = printSchool(df, listSchoolRus)

math5 = [5, 6, 9, 10, 14, 15]
math6 = [6, 7, 10, 11, 14, 15]
math7 = [5, 6, 9, 10, 13, 14]
rus5 = [13, 14, 23, 24, 32, 33]
rus6 = [17, 18, 28, 29, 38, 39]
rus7 = [24, 25, 34, 35, 44, 45]

def lenght(df):
    vvv = len(df)
    return int(vvv)-5

def countMarks(df, marks):
    for col in ['sum']:

```

```

df.loc['2'] = pd.Series((df[col] <= marks[0]).sum(), index = [col])
df.loc['3'] = pd.Series(((df[col] >= marks[1]) & (df[col] <= marks[2])).sum(), index = [col])
df.loc['4'] = pd.Series(((df[col] >= marks[3]) & (df[col] <= marks[4])).sum(), index = [col])
df.loc['5'] = pd.Series((df[col] >= marks[5]).sum(), index = [col])

df = pd.read_csv('/content/drive/MyDrive/Работа/ВПР2020/По
классам/7class/3238_VPR_2020_Rus7_reg22.csv', delimiter=';',
                skiprows=0, low_memory=False, index_col=0)
df = df[df['var-1'] != 'отсутствовал']
df = df[df['otc'] != 'нет отметки']
df = df[df['otc'] != '12']
df = df[df['otc'] != '6'] # для 7 класса

Total = df['sum'].sum()
print ("Column 1 sum:", Total / lenght(df))

df = pd.read_csv('/content/drive/MyDrive/Работа/ВПР2020/По
классам/7class/3226_VPR_2020_Mat7_reg22.csv', delimiter=';',
                skiprows=0, low_memory=False, index_col=0)
df = df[df['var-1'] != 'отсутствовал']
df = df[df['otc'] != 'нет отметки']
df = df[df['otc'] != '12']
df = df[df['otc'] != '6'] # для 7 класса

# df = df[df['user'] == 'sch220206'] # Выбираем одну школу из списка
a = df[["otc"]].rename(columns = {"otc" : "Школа"}) # Переименовываем колонки
a = a['Школа'].value_counts().reset_index() # Подсчет оценок
b = df['sum'].value_counts().reset_index()

countMarks(df, math7)
lenn = lenght(df)

```

```

df = df[["sum"]].rename(columns = {"sum" : "ВПР"}) # Сумма баллов по ВПР
df = df.reset_index()

a = a.sort_values(by='index').astype(np.int64) # Сортируем оценки по возрастанию

df = df[df.index > lenn] # оставляем четыре последние строчки с оценками
df = df.reset_index().astype(np.int64)

result = pd.merge(a, df, left_on='index', right_on='timestamp') # Объединение двух таблиц с
оценками по школе и за ВПР
del result['timestamp']
del result['index_y']

result = result.rename(index={0: '2', 1: '3', 2: '4', 3: '5'})
del result['index_x']

math7pro = ((result['Школа'].values * 100 / lenn).round()).astype(np.int)

math6pro = ((result['Школа'].values * 100 / lenn).round()).astype(np.int)

math5pro = ((result['ВПР'].values * 100 / lenn).round(2))

vpr_columns = result['ВПР'].values
school_columns = result['Школа'].values

school_columns = ((result['Школа'].values * 100 / lenn).round()).astype(np.int)
vpr_columns = ((result['ВПР'].values * 100 / lenn).round()).astype(np.int)

```

```

group = ['2', '3', '4', '5']
lenn = len(df)

mpl.rcParams['axes.prop_cycle'] = mpl.cycler(color=['#fc8621', '#ffcd00', '#682c0e', '#c24914'])
def bar_group(classes, values, width=0.8):
    plt.title('Русский язык, оценки по школе')
    plt.xlabel("Оценки")
    plt.ylabel("Процент выполнения работы")
    total_data = len(values)
    classes_num = np.arange(len(classes))
    for i in range(total_data):
        bars = plt.bar(classes_num - width / 2. + i / total_data * width,
                        values[i], width=width / total_data, align="edge",
                        animated=0.4)

        plt.xticks(classes_num, classes, size=11)
        plt.legend(['5 класс', '6 класс', '7 класс'])
        # plt.legend(['5 класс', '6 класс', '7 класс'])
        for rect in bars:
            height = rect.get_height()
            plt.text(rect.get_x() + rect.get_width()/2.0, height,
                     '%d' % int(height), ha='center', va='bottom')

# bar_group(group, [vpr_columns, school_columns])
bar_group(group, [math5pro, math6pro, math7pro])

plt.show()

```

## ПОСЛЕДНИЙ ЛИСТ ВКР

Выпускная квалификационная работа выполнена мной совершенно самостоятельно. Все использованные в работе материалы и концепции из опубликованной научной литературы и других источников имеют ссылки на них.

«\_\_» \_\_\_\_\_ 2021 г.

\_\_\_\_\_ (\_\_\_\_\_)

*(подпись)*

*(Ф.И.О)*