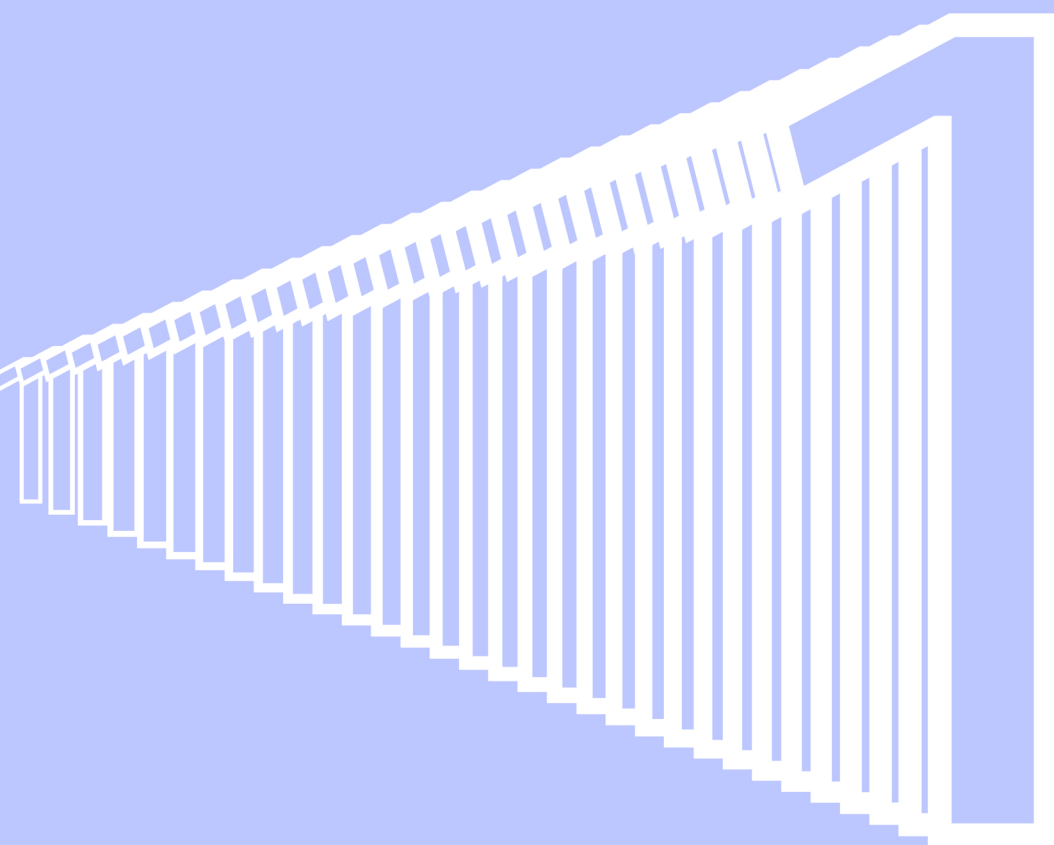




# КЕЙС «ВЛИЯНИЕ СОЦИАЛЬНЫХ СЕТЕЙ НА УСПЕВАЕМОСТЬ»



**IT, математика,  
анализ данных**

# **Методическая разработка**

## **Кейс «Влияние социальных сетей на успеваемость»**

Направление: информационные технологии,  
математика, анализ данных

Автор: Макаренко Мария Дмитриевна,  
наставник ЯндексЛицея

Организация: ГБОУ «Гимназия «Диалог»,  
центр цифрового образования детей «IT-Куб»

2021

Общее описание кейса.....	4
Категория кейса .....	4
Место в структуре программы.....	4
Количество академических часов .....	4
Учебно-тематическое планирование.....	5
Предполагаемые результаты обучающихся.....	13
Руководство наставника .....	13
Мотивационный блок .....	14
Блок формирования команд .....	15
Блок сбора информации.....	16
Предобработка данных .....	18
Первичный анализ данных средствами Excel.....	20
Анализ данных графическими методами .....	21
Анализ данных средствами машинного обучения .....	23
Формирование выводов и анализ их обоснованности.....	24
Материалы по кейсу .....	27
Список литературы .....	28

## **Общее описание кейса**

Провокационное название кейса с одной стороны заставляет учащихся защищаться, потому что они часто слышат о том, что “не фиг сидеть в соцсетях, лучше делом заняться”. На самом деле это предложение к разговору и анализу “голословных утверждений”, на уровне “британских ученых”. Задача кейса провести исследование, которое подтвердит или опровергнет гипотезы участников группы. Ведь само название кейса является абсолютно нейтральным. Задача кейса пройти все этапы по постановке гипотезы, сбору данных, их анализу, формированию выводов, оценке достоверности этих выводов.

## **Категория кейса**

Кейс в описанном варианте можно использовать во вводном модуле, но продолжительность полного кейса большая, поэтому лучше ориентироваться на подготовленных участников, которые могут несколько занятий оставаться в одной тематике. Необходимые цифровые навыки:

- работа с файлами
- простейшие навыки работы с данными в Excel
- построение графиков средствами Excel (не обязательно)
- программирование в Python
- работа с пакетом matplotlib (необязательно)
- работа с пакетом sklearn (необязательно)

Примерный возраст обучающихся: от 14 лет

## **Место в структуре программы**

«Автономный» кейс, который можно провести после серии маленьких кейсов с использованием Excel, или после небольшого введения в Python.

## **Количество академических часов**

от 2 до 6 уроков по 40 минут, 1–3 занятия по 90 минут в зависимости от подготовленности участников. Кейс можно проводить не полностью, взяв отдельные модули.

## Учебно-тематическое планирование

<b>Блок 1. Мотивационный</b>	
<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
10–15 мин	<i>организовать обсуждение вопроса, поиск информации в интернете, обмен мнениями</i>
<b>Что делаем:</b> Прошу ответить на вопросы: <ol style="list-style-type: none"> <li>1. какие и от кого высказывания слышали о влиянии соцсетей на успеваемость</li> <li>2. что “говорит” об этом интернет</li> <li>3. согласны или не согласны учащиеся с мнениями, которые прочитали или услышали</li> <li>4. насколько доказательны высказывания людей и статей в интернете</li> </ol> Подвожу к выводу, необходимо хоть как-то фактически подтверждать свою точку зрения, либо читать исследования, которые основаны на анализе данных.	
<b>Итог</b>	Будем сами анализировать данные, чтобы не быть голословными
<b>Риски</b>	Дети не готовы к открытому диалогу, боятся, что их будут ругать за личное мнение

<b>Блок 2. Формирование команд</b>	
<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
10–20 минут	<i>формирование групп единомышленников по высказанным гипотезам</i>
<b>Что делаем:</b>	

Перед началом исследования ученые выдвигают гипотезы, которые затем либо подтверждают, либо опровергают.

Каждый участник группы высказывает гипотезу:

- соцсети не влияют на успеваемость
- соцсети положительно влияют на успеваемость
- соцсети отрицательно влияют на успеваемость (может быть, будут выдвинуты и какие-то другие гипотезы)

Организуем блиц-выступления каждого учащегося по выдвигаемой гипотезе, и учащиеся с общими взглядами объединяются в группы.

<b>Итог</b>	Объединяемся в команды по выдвинутой гипотезе. Например, команда, которая считает, что соцсети плохо влияют на успеваемость. Далее вся работа идет в сформированных командах.
<b>Риски</b>	Команды очень неравнозначны по количеству человек, надо попросить сомневающихся пересест в другую команду

### **Блок 3. Сбор информации**

<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
15–30 минут	<i>сформировать данные, которые будем анализировать</i>

#### **Что делаем:**

1. Обсудить требования к данным
  - a. легко измеримые - числовые
  - b. легко получить эти данные
  - c. выбираем небольшой объем данных, так как у нас нет пока опыта работы с ними
  - d. выбираем небольшое количество параметров
  - e. ограничиться одной социальной сетью
  - f. ограничиться одним предметом
2. Обсуждаем критерии/параметры, которые будем использовать

**(мозговой штурм по командам с публичной защитой предложений):**

Обсуждаем все предложенные параметры и подвожу к мысли, что в первом приближении можно остановиться на:

- a. количество друзей во «Вконтакте»
- b. количество музыкальных записей
- c. оценка по математике
- d. баллы ЕГЭ по математике

Зафиксировать все предложенные параметры командами (гугл документ), для дальнейшего расширения работы над кейсом.

3. Используем готовый опрос, в котором уже есть данные, но добавляем каждый свои

<https://docs.google.com/forms/d/e/1FAIpQLScCQ-PuGaP1ivXWqubrQmVAmrCmq0UF8H7xRta7qQDg3oo6w/viewform>

**Если какая-то группа сможет найти готовые данные, относящиеся к теме, они могут использовать их**

<b>Итог</b>	Дети получают данные в формате таблицы по результатам опроса
<b>Риски</b>	Дети не могут предложить числовых данных, предлагаю пример и пытаюсь настроить детей на генерацию числовых параметров и оценки возможности их получить

**Блок 4. Предподготовка данных**

<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
20–30 мин	подготовка неоднородных данных к анализу

**Что делаем:**

1. Рассматриваем данные, скачанные после опроса (формат Excel)  
<https://docs.google.com/spreadsheets/d/1mG1nZQ1iYv6Jk-WBioZYjhB4oOXTPD4ULyhnKEEfkB0/edit?usp=sharing>
2. Каждая команда ищет ошибки в данных и пытается сформировать идею их исправления (формируем список ошибок в документе, выделяя цветом ячейки)

3. Вырабатываем общие подходы к исправлению ошибок (возможно, разные команды сделаю исправления ошибок по-разному)
4. Обращаю внимание на типы ошибок:
  - a. текстовые данные вместо числовых
  - b. текстовые и числовые данные в одном столбце
  - c. отбрасывание строк, не несущих полезной информации
  - d. замена неподходящих данных (например, текст)
  - e. наличие выбросов

В формате ЧГК (каждая команда формирует на листе презентации ответ обработки ошибки в течение 1 минуты) - потом каждая команда обосновывает такой выбор исправления

Инструменты, о которых можно поговорить и использовать:

- условное форматирование
- поиск и замена
- сортировка данных
- фильтры

<b>Итог</b>	Получить таблицу с числовыми данными для дальнейшей обработки
<b>Риски</b>	Удалили не те данные, произвели неправильные замены. Всегда держу в доступности исходный файл данных

## **Блок 5. Первичный анализ данных средствами Excel**

<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
20–30 минут	перебираем варианты анализа данных и пробуем их

**Что делаем:**

1. На онлайн доске команды пишут предложения по анализу данных. Возможные варианты:
  - a. среднее значение по музыке и друзьям, по оценкам по математике (ручками)
  - b. медианные значения по музыке и друзьям для разных оценок (инструмент "промежуточные итоги")
  - c. вложенные сортировки (сначала по оценке, потом по музыке, а потом по друзьям, можно два последних параметра наоборот)
  - d. формирование интервалов значений музыка и друзья для каждого вида оценки
  - e. проверка соответствия оценки в школе баллам ЕГЭ
  - f. формирование интервалов ЕГЭ
  - g. введение категориальных признаков (например: мало музыки  
-  $< 50$ , средне -  $> 50$  и  $< 100$ , много  $> 100$ )
  - h. и все остальное, что предложат учащиеся
2. Обсуждаем каждый метод. Потом каждая команда выбирает стек методов и пытается их применить
3. Презентация анализа каждой команды. Вопросы других команд

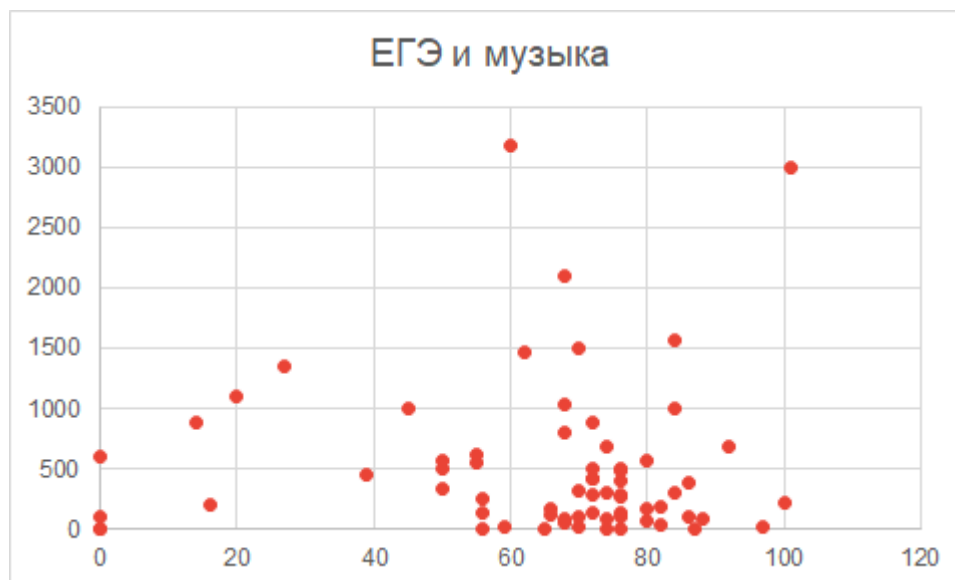
<b>Итог</b>	Проверяем получили ли мы доказательство нашей гипотезы, или нет
<b>Риски</b>	Дети очень хотят доказать свою, поэтому малейшее смещение данных в сторону их гипотезы вызывает бурю эмоций. Обсуждаем достоверность полученных результатов.

**Блок 6. Анализ данных графическими средствами**

<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
20–30 минут	получить дополнительную информацию о структуре данных, используя графические средства Excel и matplotlib

### Что делаем:

1. обсуждаем разницу между графическим представлением данных, устным и цифровым, т.е понятие наглядности
2. ищем примеры в интернете, узнаем, что такое инфографика
3. показываю пару примеров графиков на основе наших данных
  - а. распределение баллов ЕГЭ и количества аудиозаписей



б. Процентное соотношение троечников, хорошистов и отличников



<https://drive.google.com/file/d/1xoVNTfeltE70ufJFaQZyPKGQ7vHULg6J/view?usp=sharing>

4. командам дается время на создание собственных, как минимум трех различных графиков по данным
5. демонстрация своей работы другим командам
6. С наставником строим карту рассеяния с помощью matplotlib
  - а. подготовка и сохранение данных в формате csv для загрузки в CoLab (убрать столбик с датами, заголовки столбцов по-английски, удалить строки с пропущенным ЕГЭ)
  - б. создания блокнота в CoLab
  - в. загрузка данных
  - д. рисование графика рассеяния

<https://colab.research.google.com/drive/1fhJt8KFJ4GPaUDaT2sH5mX>

[DSEN MP75RY?usp=sharing](https://colab.research.google.com/drive/1of7a5hVPG4QA9TDKTVUfkzj3DfjP_6nUG?usp=sharing)

7. Каждая команда средствами matplotlib строит еще один график
- совместная корреляция параметров
  - коробка с усами
  - и др.

<b>Итог</b>	Дети получают графики и должны попытаться их интерпретировать
<b>Риски</b>	Графики не воспринимаются детьми как более понятные. Нужно чаще показывать и обсуждать графики с учениками.

### **Блок 7. Анализ данных алгоритмами машинного обучения**

<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
20–30 мин	<i>демонстрация других подходов к анализу данных, предсказание своего балла ЕГЭ</i>
<b>Что делаем:</b> Наставник рассказывает: <ol style="list-style-type: none"><li>Понятие искусственного интеллекта</li><li>Понятие машинного обучения, как части ИИ</li><li>Простейший метод KNN (к ближайших соседей)</li><li>Разбор “ручками” метода на примере</li><li>Создание CoLab блокнота для демонстрации метода<ol style="list-style-type: none"><li>просмотр 3-х ближайших соседей (NearestNeighbors)</li><li>предсказание значений на основании 3-х соседей (KNeighborsRegressor)</li></ol></li></ol> <a href="https://colab.research.google.com/drive/1of7a5hVPG4QA9TDKTVUfkzj3DfjP_6nUG?usp=sharing">https://colab.research.google.com/drive/1of7a5hVPG4QA9TDKTVUfkzj3DfjP_6nUG?usp=sharing</a>	
<b>Итог</b>	Получение прогноза по ЕГЭ для себя
<b>Риски</b>	Прогноз с низкими баллами вызывает эмоциональное напряжение. Обсуждаем понимание статистики и вероятности.

<b>Блок 8. Формирование выводов и анализ их обоснованности</b>	
<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
10–20 мин	<i>показать, что наше исследование околонуучное, но не голословное</i>
<b>Что делаем:</b> <ol style="list-style-type: none"> <li>Каждая команда готовить защиту на 2–3 минуты по результатам исследования <ol style="list-style-type: none"> <li>наша гипотеза</li> <li>средства, которые мы использовали</li> <li>демонстрация результатов</li> <li>выводы (подтвердилась гипотеза или нет)</li> <li>насколько результаты анализа достоверны</li> </ol> </li> <li>Вопросы других команд</li> </ol>	
<b>Итог</b>	по результатам исследования скорее всего ничего доказать нельзя, поэтому важно говорить об этапе в исследовании, которое можно и нужно продолжать
<b>Риски</b>	дети не приводят числовые доказательства, а говорят вновь только о своем восприятии проблемы

<b>Блок 9. Рефлексия</b>	
<b>Предполагаемая продолжительность</b>	<b>Цель блока</b>
5–10 минут	<i>подвести итоги как командные, так и личные</i>
<b>Что делаем:</b> <ol style="list-style-type: none"> <li>Командная <ol style="list-style-type: none"> <li>нарисовать мем - как это было</li> </ol> </li> <li>Личная <ol style="list-style-type: none"> <li>что было самым неожиданным в кейсе</li> <li>чему научился</li> <li>что возьму на будущее. Можно использовать <a href="https://mentimeter.com">mentimeter.com</a></li> </ol> </li> </ol>	

<b>Итог</b>	Мем прекрасная память о прошедшем исследовании как эмоциональное закрепление
<b>Риски</b>	Дети бояться высказывать свои мысли и повторяют за окружающими

### **Предполагаемые результаты обучающихся**

#### **Артефакты:**

- *электронные материалы команд (заметки, презентации, документы Excel и CoLab)*
- *MEM*

#### **Soft skills:**

- Коммуникационные навыки
- Компьютерная и техническая грамотность
- Адаптивность
- Навыки исследования
- Навыки управления проектами
- Умение решать проблемы

#### **Hard skills:**

- анализ данных
- опыт работы в Excel (фильтры, сортировки, поиск и замена, функции, диаграммы)
- опыт работы в Colab (matplotlib, sklearn)

## Руководство наставника

Название кажется провокационным, однако это предложение к анализу голословных утверждений – так называются «утверждения без приведения доказательств, фактов» [6]. Выявление неподтвержденной информации как в интернет-публикациях, так и в официальных СМИ требует особого навыка, который сегодня называется грамотностью в работе с данными и является составляющей информационной грамотности. Многие современные учебники включают разделы, связанные с анализом данных, но чаще всего изучение этой темы происходит на основе теоретического материала. Автор же предлагает обсудить вопросы достоверности информации и возможности манипуляции данными на практике, в ходе построения простейшей, доступной для школьника математической модели влияния социальных сетей на успеваемость.

Цель кейса – подтвердить или опровергнуть гипотезы участников группы о влиянии социальных сетей на успеваемость, пройдя через все этапы исследования: постановка гипотезы, сбор и анализ данных, формирование выводов и оценка их достоверности. Итогом работы над кейсом должна стать система предсказания количества баллов на ЕГЭ по математике на основе данных аккаунтов в соцсетях и текущей успеваемости школьников.

В ходе работы над кейсом от учащихся потребуются следующие навыки: работа с файлами, простейшие навыки обработки данных и построение графиков в Excel, программирование на Python, работа с пакетом `matplotlib` (факультативно), работа с пакетом `sklearn` (факультативно). Часть из них может быть уже изучена, а часть будет вводиться в качестве нового материала. Кейс предназначен для школьников старше 13 лет и может представлять собой серию занятий общей продолжительностью от 2 до 6 академических часов. Кейс разбит на следующие блоки:

1. Мотивационный, целью которого является обсуждение вопроса, поиск информации в интернете, обмен мнениями.

2. Формирование команд на основе объединения единомышленников по заданному вопросу.

3. Сбор информации – создание таблицы данных с использованием легко измеримых параметров для социальных сетей и успеваемости. Для сбора данных рекомендуется использование опросов, например, с помощью Google Forms или «Яндекс.Форм».

4. Предварительная обработка данных, как важнейший этап любого исследования.

5. Первичный анализ данных простыми инструментами Excel, которые должны быть знакомы учащимся по школьной программе:

среднее арифметическое, медиана, сортировки, фильтры, интервалы.

6. Анализ данных графическими методами как возможность увидеть наиболее общие тенденции в данных. Для этого этапа предлагается использование Excel, для более подготовленной аудитории — Matplotlib, а также других популярных цифровых инструментов.

7. Анализ данных алгоритмами машинного обучения, в частности, простейшим методом k ближайших соседей.

8. Формирование выводов и анализ их обоснованности.

9. Рефлексия, как метод подведения итогов и осмысления всех этапов кейса.

### **Мотивационный блок**

Исследовательская цель блока: формирование цели исследования

Образовательная цель блока: развить навыки поиска информации, определения ее достоверности, познакомить с понятием репрезентативности выборки

Мотивационный блок является одним из важнейших, потому что время, потраченное на формирование стойкого интереса к теме кейса, окупится на дальнейших этапах работы, когда учащимся придется прикладывать усилия для выполнения сложных объемных вычислительных и интеллектуальных действий. Понимание четкой цели всего процесса и каждого отдельного блока формирует самостоятельный подход к работе и умение ставить и достигать цели. Поэтому главным итогом мотивационного блока должно стать формирование цели и постановка задач, которые необходимо выполнить для ее достижения. Каким образом этого достичь?

Сама тема уже является актуальной для школьников, поскольку они очень часто слышат от взрослых высказывания о негативном влиянии социальных сетей на успеваемость. Однако насколько они обоснованы? Необходимо дать учащимся привести примеры таких высказываний из своего опыта, выяснить их предварительное мнение. Однако все это — лишь мнения так называемой нерепрезентативной выборки (это одно из понятий, которое учащиеся должны уяснить в ходе работы над кейсом).

Следующим этапом станет поиск большего числа высказываний. Их источником может выступить интернет. При поиске в сети необходимо обращать внимание учащихся не только на окраску высказываний, но и на источники данных. Очень важно обращать внимание не только на количество высказываний, но и на их оригинальность. Например, в ходе работы с учащимися была найдена информация о работе “профессора Пола Киршнера” [7]. Статья

многokrатно перепечатывалась различными изданиями и сайтами, однако практически ни в одном не указано, где, когда и как проводились исследования.

Одним из способов формирования собственного мнения является личный опыт, подкрепленный фактами. Поэтому исследование будет строиться на данных участников кейса, что сформирует устойчивый интерес аудитории.

### **Блок формирования команд**

Исследовательская цель блока: формирование команд

Образовательная цель блока: сформировать умения аргументировать свои высказывания, познакомиться с этапами исследования

После обсуждения мнений из личного опыта и информации, найденной в интернете, выведем цель: выяснить, как влияют социальные сети на успеваемость ближайшего окружения — на нас, наших одноклассников и друзей.

В научных исследованиях принято сначала формулировать гипотезу исследования, а затем собирать данные и, анализируя их, подтверждать или опровергать изначальную гипотезу. Сформируем как минимум три возможные гипотезы, которые может предложить нам аудитория:

- соцсети положительно влияют на успеваемость,
- соцсети отрицательно влияют на успеваемость,
- соцсети не влияют на успеваемость.

Возможно, будут предложены и другие варианты. Теперь можно предложить каждому учащемуся в нескольких фразах сформулировать свою гипотезу и публично выступить с ее обоснованием (пока голословным). Такой подход позволит разделить учащихся по группам с общими взглядами на поставленную задачу.

Одним из важных навыков в ходе любой деятельности является планирование своих действий. Однако учащимся редко достается возможность выполнить этот этап работы самостоятельно: на уроке за них уже все распланировали. Мы составим общий план действий на доске с помощью одного из активных «бойцов», а остальная аудитория поможет с выявлением необходимых шагов. Задача учителя — подсказать шаги, если учащиеся не смогли найти их самостоятельно: выдвижение гипотезы, поиск данных, анализ данных, построение выводов и проверка полученных выводов на достоверность. После этого, возможно, потребуется повторное исследование, поскольку для получения адекватной математической модели обычно проходит

несколько итераций.

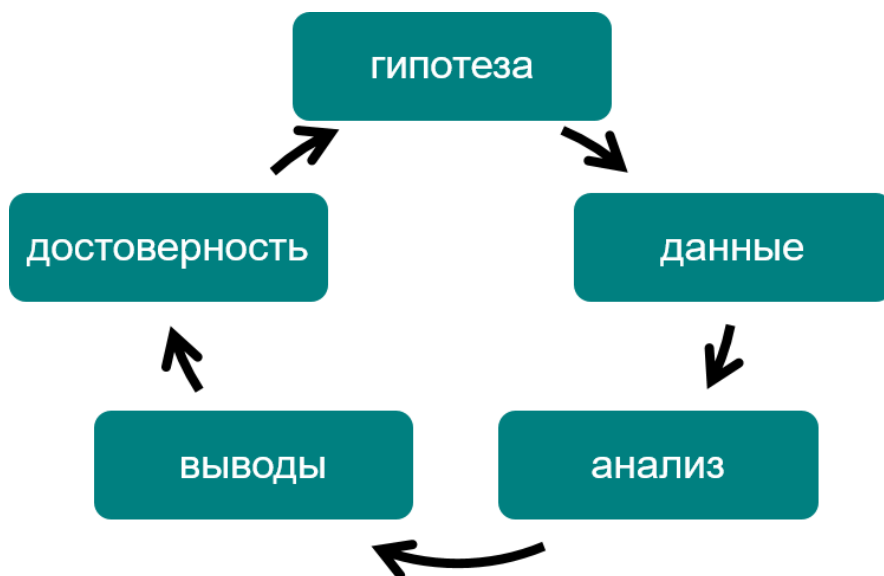


Рисунок 1. План исследования

### Блок сбора информации

Исследовательская цель блока: сбор данных

Образовательная цель блока: сформировать понимание, как и откуда берутся данные, как происходит их отбор, познакомить с требованиями к данным

Итак, исследовательские команды сформированы, а гипотезы выдвинуты. Теперь необходимо понять, на основе каких данных будет проводиться анализ и доказательство или опровержение гипотез. Подбор данных — это ключевой момент любого исследования, от его качества во многом зависят результаты. С другой стороны, поскольку это учебный кейс, мы не можем позволить себе потратить большое количество времени на поиск информации.

Впервые в кейсе мы сталкиваемся с проблемами математического моделирования: с одной стороны, необходимо взять как можно больше разнообразных данных, с разных сторон описывающих исследуемый процесс; с другой стороны, моделирование состоит именно в выделении существенных деталей, выявлении важного и неважного для решаемой задачи. Все это является личной ответственностью автора модели, и главное, чтобы он осознавал ограниченность своих предположений.

Остановимся на важных свойствах данных, которые мы должны учитывать:

- Данные должны быть легко измеримыми. Лучше, чтобы это были

числовые данные, которые легко использовать при математических вычислениях.

- Эти данные должны быть легко получаемыми, не должны содержать персональных данных, защищаемых законом.
- Нет смысла для первого эксперимента брать огромное число данных. Одним из принципов математического моделирования является постепенное наращивание сложности модели. Поэтому можно обойтись десятками записей об учащихся.
- Информация об учащихся должна быть выражена в 3–5 параметрах о каждом, чтобы не нагружать первое приближение в исследовании.
- Социальных сетей много, и не все участники являются их пользователями, поэтому достаточно остановиться на одной из популярных. Выбор социальной сети, из которой будут браться данные, может варьироваться, но чаще всего выбирается социальная сеть “ВКонтакте”.
- Еще один сложный параметр нашей задачи — успеваемость. Необходимо обсудить с учащимися, что можно подразумевать под этим термином. Для простоты остановимся на оценке по одному предмету. Чаще всего выбор падает на математику, так как ее подсказывает тематика кейса.

Обсуждение требований к данным должно быть зафиксировано в презентации или на доске, чтобы они были перед глазами учащихся для выполнения следующего этапа.

Следующим этапом является командная работа: учащиеся выдвигают предложения о том, какие конкретно данные необходимы для исследования, и как их найти. На обсуждение и поиск данных выделяется до 10 минут. Каждая команда может зафиксировать свои мысли на одном листе в любом сетевом инструменте: например, в Google Presentation или Padlet. С одной стороны, это знакомство с облачными сервисами и обучение их использованию, с другой — способ сохранить информацию в доступном для всех участников месте. По итогу этапа каждая команда должна представить свои предложения и обосновать их.

Далее следует выбор параметров, которые наиболее часто встречаются у команд. Важно зафиксировать как можно больше интересных предложений, но необходимо также не перегружать первую итерацию исследования и отобрать минимальное количество параметров.

Опыт показывает, что учащиеся обычно останавливают свой выбор на следующих параметрах:

- Количество друзей социальной сети как простой показатель

времени, которое в ней проводит участник. Чем больше друзей, тем больше школьник общается, а значит, тратит времени. Необходимо обратить внимание участников на то, что это суждение

— только предположение, которое в настоящем исследовании пришлось бы доказывать.

- Количество музыкальных записей в социальной сети. Если участник не любит общаться, то чаще всего он использует сеть именно как источник музыки (опять предположение). К тому же это легко измеримый числовой параметр.
- Оценка по математике чаще всего предлагается как легко измеримый параметр успеваемости. Математика считается одним из сложных предметов школьной программы, и, кажется, что именно она может показать дифференцируемость по влиянию социальных сетей. Опять предположение, которое необходимо исследовать.
- Оценка ЕГЭ/ОГЭ в процентах может стать еще одним показателем успеваемости. Однако учащиеся чаще всего не предлагают этот критерий: с одной стороны, они в силу возраста еще могут не участвовать в итоговых экзаменах, с другой — они не осознают его абсолютность, в отличие от оценки. Задача наставника — донести это понимание до учащихся и предложить по результатам анализа модели предсказать количество баллов, которые они смогут получить на ЕГЭ.

Еще один вопрос — как собрать такие данные? Можно попытаться поискать открытые источники, но всегда интересней использовать собственные данные. Таким образом учащиеся получают еще большую мотивацию для дальнейших исследований. Наиболее эффективным способом получения такой информации непосредственно от участников исследований представляются «Яндекс.Формы» или «Google Forms». Кстати, эти инструменты можно использовать неоднократно, таким образом наращивая количество данных для исследования. Ниже приведен вид такой формы, которую заполняли не только школьники, но и студенты.

**Влияние социальной сети на успеваемость**

\* Обязательно

Сколько у вас друзей в ВК? \*

Мой ответ

Сколько у вас аудиозаписей в разделе музыка в ВК? \*

Мой ответ

Какая у вас оценка по математике? \*

☐ 2

☐ 3

☐ 4

☐ 5

Сколько у вас баллов ЕГЭ/ОГЭ по математике в процентах (число от 0 до 100)

Мой ответ

Отправить

*Рисунок 2. Окно формы опроса*

После заполнения формы можно получить данные в формате Excel-таблицы.

### **Предобработка данных**

Исследовательская цель блока: подготовка данных для анализа

Образовательная цель блока: познакомить с понятиями грязных данных и ошибок пользовательского ввода, сформировать понимание, как выглядят заполненные вручную, то есть плохо форматированные данные, научить работе с инструментами Excel

Любые данные, собранные с помощью опросов, являются «грязными» и требуют предварительной обработки. Пример таких данных:

	A	B	C	D	E
1	Отметка времени	Сколько друзей в ВК?	Сколько аудиозаписей в	Оценка по математике?	Сколько баллов ЕГЭ/ОГЭ
34	12.17.2020 20:56:55	230	~100	3	0
35	12.17.2020 20:57:06	438	615	4	55
36	12.17.2020 20:57:15	66	3184	5	60
37	12.17.2020 20:57:29	50	179	5	82
38	12.17.2020 20:57:42	4	446	5	39
39	12.17.2020 20:58:46	1	600	4	0
40	12.17.2020 20:59:21	21	83	4	74
41	12.17.2020 21:00:27	68	1340	3	Не помню
42	12.17.2020 21:02:17	120	1435	5	
43	12.17.2020 21:07:35	13	800	3	Не помню
44	12.17.2020 21:08:28	Много	Много	5	Понятие не имею
45	12.17.2020 21:10:04	20	250	4	56
46	12.17.2020 21:14:25	10	1000+	3	45
47	12.17.2020 21:15:19	197	Точно не помню, где то !	3	72
48	12.17.2020 21:18:32	1360	688	5	92
49	12.17.2020 21:18:54	40	100	5	70
50	12.17.2020 21:18:57	70	500+	4	50
51	12.17.2020 21:44:22	26	0	4	0
52	12.17.2020 21:50:36	238	+140	4	56
53	12.17.2020 21:53:52	997	80	4	68
54	12.17.2020 21:56:44	90	568	5	80

*Рисунок 3. Пример данных в формате электронной таблицы*

Многие системы сбора данных не позволяют ограничить формат вводимых данных: например, только текст, только число или число в заданном диапазоне. Пользователи, плохо знакомые с автоматической обработкой данных, не понимают, что требование к формату — не прихоть исследователя, а способ минимизировать время на предобработку данных.

Типы ошибок, характерные для нашего примера:

- Ввод нечисловых данных, т.е. написание текста вместо чисел. Так, например, строка 44 в таблице данных (см. рисунок 3) абсолютно непригодна для использования и не несет никакой полезной информации. Скорее всего, такую строку проще удалить, чем прилагать усилия для ее заполнения.
- Ввод данных вместо пропуска для необязательных полей. Примером могут служить строки 41 и 43. В форме (см. рисунок 2) последнее поле является необязательным для заполнения. Однако пользователи этого не увидели и вписали в поле текст, который на этапе предобработки нужно просто удалить.
- Внесение, кроме числовых, дополнительных текстовых данных: например, в строках 34, 47, 50 и 52. Чтобы исправить такие данные, достаточно просто удалить лишние символы. Для небольшого числа данных это можно сделать вручную, а можно

применить средства Excel: например, условное форматирование для выделения таких данных или поиск нечисловых данных. Можно также написать макрос, но для начинающих это высший пилотаж. Если кто-то из школьников готов проделать это самостоятельно, необходимо дать ему такую возможность и показать остальным, как реализуется этот способ.

- Существуют ошибки не только формата, но и значения. Например, введение отрицательных значений для оценки по математике или значения за пределами интервала [2, 5], введение нуля в поле для баллов по ЕГЭ (даже если это реальная ситуация, она не поможет нам в дальнейшем). Такие данные нужно отыскать и только потом исправлять. Можно использовать фильтры, сортировки и условное форматирование, автозамены и большое число других инструментов.

Организационно этап предобработки данных можно провести в формате «Что? Где? Когда?» (можно назвать «Что? Где? Как?»). Каждая команда в течение 1 минуты пишет на листе презентации список замеченных ошибок и способы их исправления. Для этого необходимо подготовить на листе таблицу соответствующего формата. Советую проводить данный этап, не разрешая командам сразу вносить изменения в грязные данные, поскольку «коллективный разум» может предложить более интересные и верные способы предобработки.

Команда 1

	Какой тип ошибки ошибки	Пример из данных	Как исправить
1			
2			
3			
4			
5			

*Рисунок 4. Пример таблицы для игры «Что? Где? Как?»*

Команде, нашедшей наибольшее количество типов ошибок, предоставляется право выступить первой. Каждая команда по очереди оглашает по одному типу ошибок, который не был представлен ранее, и обосновывает необходимость исправления и его способы. После того, как выскажутся все, каждая команда приводит данные в порядок.

## **Первичный анализ данных средствами Excel**

Исследовательская цель блока: поиск зависимостей в данных

Образовательная цель блока: научиться применять инструменты Excel к анализу реальных данных

Первичный анализ данных — важный этап. Для его проведения будем использовать электронные таблицы. С одной стороны, это знакомый школьникам инструмент, с другой

— они редко его используют из-за отсутствия потребности в нем и слабых навыков. Даже если школьники раньше не использовали Excel или его аналоги, на наш взгляд, можно сразу предложить им поработать с данными, используя интернет как справочник. Таким образом мы создаем мотивацию для самостоятельного изучения инструмента.

На этот этап работы можно выделить от 20 до 30 минут. Чем более подготовлены участники, тем больше времени им дается на самостоятельную работу. Желательно, чтобы за отведенное время у учащихся не кончились идеи, как можно проверить зависимость успеваемости от имеющихся данных. В ходе работы задача наставника — перемещаться между командами, разрешать технические затруднения, подсказывать идеи тем, кто зашел в тупик.

После завершения этой работы каждой команде необходимо публично рассказать об идеях анализа и полученных результатах. Если методы, предложенные учащимися, оказались недостаточно разнообразны, наставник должен продолжить анализ в формате «делай как я», показывая технологии и подходы, которые не были использованы школьниками.

Какие методы можно применить:

1. Вычислить среднее значение по количеству музыкальных композиций и друзьям для каждого изменения оценки по математике. Это можно сделать в ручном режиме: сначала отсортировать данные по столбцу "Оценка", а затем подсчитать сумму и количество для каждого значения; после фильтрации можно применить функцию СРЗНАЧ для заданных интервалов, а можно воспользоваться инструментом ПРОМЕЖУТОЧНЫЕ ИТОГИ.
2. Вычислить значения медианы для каждого изменения оценки. Школьники редко используют этот показатель, поскольку он не является частью школьной программы. Сейчас самое время обсудить его, его преимущества и недостатки по сравнению со средним значением. Вычисление можно сделать как ручным способом, так и используя функции МЕДИАНА и ПРОМЕЖУТОЧНЫЕ ИТОГИ. Если позволяет уровень подготовки учащихся, можно поговорить о математическом ожидании

величины.

3. Сформировать интервалы значений для каждого вида оценки. Для этого можно использовать вложенные сортировки или фильтры. Интервальные значения могут рассказать об обособленностях или их отсутствии у значений параметров.
4. Вычислить корреляцию оценки по математике с баллами ЕГЭ. Это можно сделать простой вложенной сортировкой или с помощью специальной функции КОРРЕЛ.
5. Ввести категориальные признаки (например: мало музыки —  $< 50$ , средне —  $> 50$  и  $< 100$ , много —  $> 100$ ) и превратить их далее в числовые данные. Это позволит познакомить учащихся с подходами машинного обучения.

Завершением блока должно стать представление командами полученных результатов. Они должны показать, подтвердились ли их гипотезы в результате анализа данных, нашли ли они доказательства своим предположениям. Особенный интерес вызывает обсуждение полученных доказательств, поскольку иногда происходит «притягивание за уши» результатов. Если это сделали две команды, значит, мы получили противоречие, и можно говорить о степени достоверности выводов и оценок.

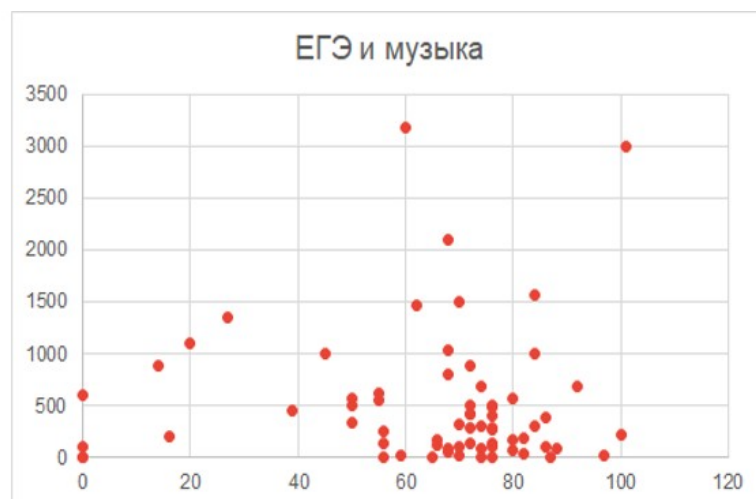
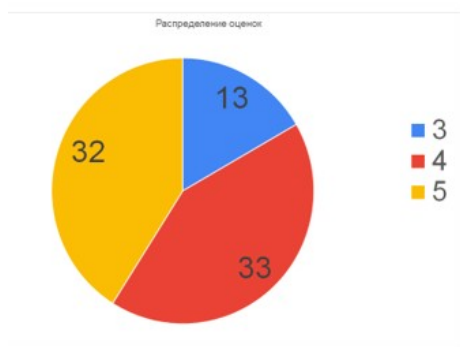
### **Анализ данных графическими методами**

Исследовательская цель блока: поиск скрытых зависимостей, качественный анализ данных

Образовательная цель блока: познакомить учащихся с графическими возможностями Excel и библиотеки Matplotlib

Обычно анализ любых данных начинают с качественного, т.е. визуального анализа. Но этот подход редко встречается в школьной программе, поэтому он абсолютно непривычен школьникам. Можно обсудить с учащимися разные способы представления информации: устное, текстовое, цифровое и графическое. Лучше, если учащиеся найдут примеры этих способов из реальной жизни. Это могут быть правила дорожного движения, которые можно услышать от мамы, прочитать в справочнике или увидеть в виде графических знаков на дороге; или информация о росте населения Земли в виде репортажа, статьи или графика.

Построим несколько графиков на основе наших данных вместе с учителем и позволим командам придумать еще 5 штук для подтверждения своей гипотезы (примеры графиков см. на рисунке 5).



*Рисунок 5. Примеры построения графиков.*

Обсуждая полученные графики, нужно обратить внимание учащихся на смещенность данных, полученных в результате наших опросов. Количество учащихся с тройками по математике у нас очень небольшое. Это, скорее всего, не отражает истинную картину оценок по математике, например, в пределах России (это тоже предположение, которое надо проверять). Отсюда следует, что к выводам, полученным на основе наших данных, нужно относиться с осторожностью.

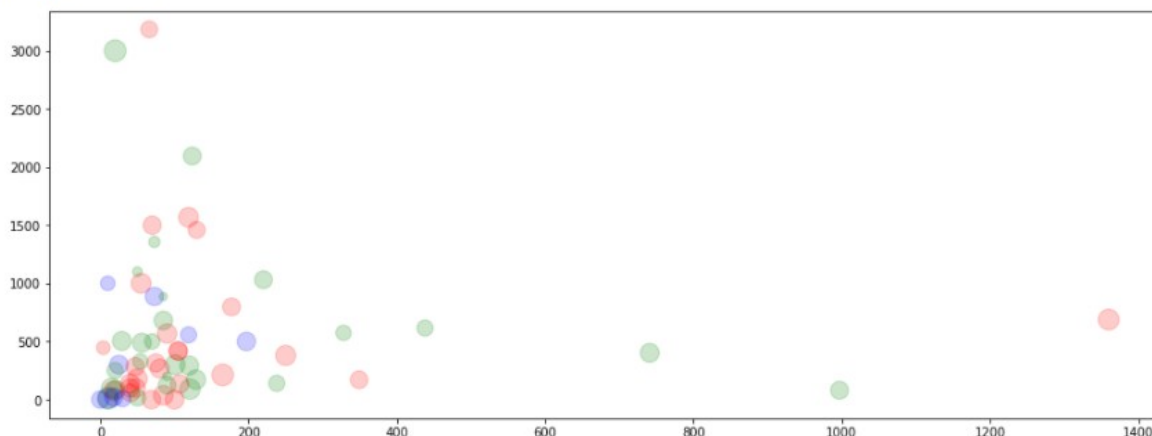
Однако возможности Excel, конечно, ограничены. Поэтому подготовленным школьникам можно предложить построение более нагруженных графиков с использованием библиотеки Matplotlib на языке Python. Самый простой способ — это воспользоваться инструментом Google Collab. Для этого необходим доступ в интернет и аккаунт Google у каждого ученика или хотя бы у каждой команды.

Для загрузки данных в Python необходимо сохранить данные Excel в формате csv, что можно сделать стандартными средствами Excel. Желательно предварительно дать столбцам английские названия. Пример кода для загрузки данных и построения диаграммы рассеяния можно посмотреть в материалах кейса.

Хочется отметить, что необходимо постепенно усложнять построения диаграммы рассеяния: сначала построить аналог Excel-диаграммы зависимости количества друзей от количества аудиозаписей; затем добавить размер точек в зависимости от значения баллов ЕГЭ; потом изменить цвет точек согласно оценке по математике.

Рассматривая с учащимися диаграмму, необходимо обсудить ее особенности и, возможно, ввести понятие выбросов. Затем можно построить новую диаграмму, исключив статистически незначимые данные.

```
fig = plt.pyplot.figure(figsize=(16, 6))
colors = ["", "", "", "blue", "green", "red"]
plt.pyplot.scatter(data['friend'], data['music'], color=data['math'].apply(lambda x: colors[x]), s=data['ege']*3, alpha = 0.2)
plt.pyplot.show()
```



*Рисунок 6. Диаграмма рассеяния с учетом всех имеющихся данных*

Одним из важных итогов блока должен стать вывод о том, что на основании наших данных нельзя утверждать о сколько-нибудь значимом влиянии социальных сетей на успеваемость школьников по математике. Это совсем не означает, что ее нет, это значит, что исследование можно продолжать на следующей итерации: можно искать новые данные и пробовать новые методы. Например, можно еще раз обратиться к литературе и узнать про другие подобные исследования: в частности, исследование С.Л. Таланова [3].

### **Анализ данных средствами машинного обучения**

Исследовательская цель блока: предсказать оценку своего ЕГЭ по математике

Образовательная цель блока: познакомить учащихся с простейшими методами машинного обучения

Необходимо расширять кругозор школьника и не ограничиваться простейшими статистическими методами. Погрузив учащихся так глубоко в задачу, можно предложить и более сложные подходы, такие как методы машинного обучения, которые активно развиваются и внедряются во все области нашей жизни: от наложения ушек на свое фото в смартфоне до прогнозирования успеваемости будущих студентов на основе их достижений в школе. Подходы методов машинного

обучения значительно отличаются от классических, но с простейшими из них все же можно познакомить школьников.

Автор предлагает использовать метод  $k$  ближайших соседей (kNN) для прогнозирования баллов ЕГЭ по данным участников кейса. Общая идея метода проста: построим точку по всем известным нам параметрам в пространстве этих параметров. В нашем случае их три: количество аудиозаписей, количество друзей и оценка по математике. Для простоты можно оставить любые два параметра, чтобы работать в двумерном пространстве. Найдем  $k$  ближайших точек в построенной нами плоскости и рассчитаем значение оценки ЕГЭ как среднее арифметическое значений в ближайших точках.

В методе ближайших соседей существует два гиперпараметра, которые необходимо обсудить и экспериментально подобрать: количество используемых соседей, т.е. значение величины  $k$ , и понятие расстояния между точками. В простейшем случае можно взять Евклидово расстояние, но можно обсудить и другие варианты.

$$\hat{y} = \frac{\sum_{k=1}^K y_k}{K}.$$

*Рисунок 7. Вычисление значения целевой функции как среднего арифметического ближайших соседей*

Первый подход к методу лучше провести в «ручном» режиме, т.е. посчитать для некоторого ученика все значения расстояний, используя калькулятор, Excel или Python, если у школьников имеются навыки программирования. Затем можно продемонстрировать работу функции из библиотеки `sklearn`.

```

# подключение функции из библиотеки
from sklearn.neighbors import KNeighborsRegressor

# создание модели с k = 3 (предсказываем по трем соседям)
neigh = KNeighborsRegressor(n_neighbors=3)
# выбираем три столбца как данные на основе которых прогнозируем
X = data[['friend', 'audio', 'math']]
# выбираем столбец ЕГЭ, как предсказываемый
y = data['ege']
# запускаем обучение модели
neigh.fit(X, y)

# будем предсказывать для пользователя с такими параметрами
# количество друзей, количество аудиозаписей, оценка по математике
user = [520, 109, 5]

# предсказание количества баллов на ЕГЭ по профилю в соцсети
neigh.predict([user])

```

*Рисунок 8. Пример кода по прогнозированию методом kNN*

Пусть код, предложенный в материалах к кейсу, пока выглядит для школьников как магия: это породит в них любопытство и желание разобраться в нем в дальнейшем.

### **Формирование выводов и анализ их обоснованности**

Исследовательская цель блока: подвести итоги, сформировать выводы о состоятельности гипотезы

Образовательная цель блока: сформировать понимание достоверности данных и выводов на их основе

Подводя итоги кейса, необходимо вспомнить о том, какие гипотезы мы выдвигали в начале работы, и подтверждают ли их наши исследования. Для этого предлагается каждой команде подготовить представление своих результатов в виде презентации со следующими разделами:

- наша гипотеза,
- методы и инструменты, которые мы использовали,
- демонстрация полученных результатов (диаграммы и цифры),
- выводы – подтвердилась гипотеза или нет,
- оценка достоверности результатов анализа,
- как можно продолжить исследование.

Это важнейший этап всего кейса. Новые навыки в работе с цифровыми инструментами, новые термины и красивые картинки не должны затмить главную идею проведенного исследования. Никакие исследования и модели не могут быть непогрешимыми, все выводы

должны опираться на данные, но выбор этих данных — это уже построение модели, а значит, упрощение существующей ситуации. Любые исследования можно продолжать бесконечно, бесконечно приближаясь к некоторому результату, но отнюдь не к истине.

### **Рефлексия**

Исследовательская цель блока: уметь анализировать собственные действия и достижения

Образовательная цель блока: выработать у учащихся привычку анализировать свои действия и достижения

Этап самооценки очень важен в любом действии, а особенно в таком большом для ребенка исследовании. Чаще всего учащиеся пытаются оценить материал, наставника, но не свои действия. Рефлексия — это один из способов учить учиться. «Рефлексия помогает ребенку не только осознать пройденный путь, но и выстроить логическую цепочку, систематизировать полученный опыт, сравнить свои успехи с успехами других учеников». [8].

Рефлексию можно проводить в командах, например, предложив нарисовать тематический мем «Как это было?». Популярный среди подростков формат позволяет анализировать самые яркие моменты в процессе работы. Однако личный формат представляется более предпочтительным, поскольку позволяет проявить себя каждому. А безличный формат облака тегов с помощью инструмента [mentimeter.com](https://www.mentimeter.com) позволит высказать свое мнение даже самым застенчивым.

## **Влияние социальных сетей, как это было?**

Mentimeter

исследование циклично  
python в гугле  
анализу данных  
познавательно  
собирать и анализировать  
евклид для xd измерений  
проверять исследования

*Рисунок 9. Пример рефлексии в виде облака тегов.*

Ребятам было предложено ответить на один из вопросов:

- что было самым неожиданным,
- чему я научился,
- что возьму на будущее.

Результаты, полученные на основе работы над данным кейсом, могут быть абсолютно разными даже на основе одних и тех же данных и очень сильно зависят от выбранных инструментов. Однако кажется полезным познакомить учащихся, завершивших кейс, с результатами реальных научных исследований в этом направлении. Так, С.Л. Таланов в своем исследовании [3] утверждает, что "для студентов, которые имеют развитые навыки, знания и умения в сфере поиска и обработки информации и осознанно выбрали направление обучения и вуз, социальные сети способствуют повышению успеваемости. Для студентов, поступивших по мотиву «все равно в какой вуз и направление обучения», социальные сети являются способом ухода от реальности (избегание неприятного, скучного в жизни)".

Использование кейсов позволяет усваивать знания и получать навыки их применения в результате активных самостоятельных действий в условиях противоречий, учит оценивать средства достижения цели и достоверность результатов.

## Материалы по кейсу

Все материалы по кейсу	<a href="https://clck.ru/UJrHB">https://clck.ru/UJrHB</a>
Опрос	<a href="https://docs.google.com/forms/d/e/1FAIpQLScCQ-PuGaP1ivXWqubrQmVAmrCmq0UF8H7xRta7qQDg3o6w/viewform">https://docs.google.com/forms/d/e/1FAIpQLScCQ-PuGaP1ivXWqubrQmVAmrCmq0UF8H7xRta7qQDg3o6w/viewform</a>
Результаты опроса (не обработанный файл)	<a href="https://docs.google.com/spreadsheets/d/1mG1nZQ1iYv6Jk-WBioZYjhB4oOXTPD4ULyhnKEEfKB0/edit?usp=sharing">https://docs.google.com/spreadsheets/d/1mG1nZQ1iYv6Jk-WBioZYjhB4oOXTPD4ULyhnKEEfKB0/edit?usp=sharing</a>
Результаты опросов элементами анализа	<a href="https://docs.google.com/spreadsheets/d/1RNv8VNzzy4g7JVmsO-dE-9tJUv2P8hJO/edit?usp=sharing&amp;ouid=117758292418578862242&amp;rtpof=true&amp;sd=true">https://docs.google.com/spreadsheets/d/1RNv8VNzzy4g7JVmsO-dE-9tJUv2P8hJO/edit?usp=sharing&amp;ouid=117758292418578862242&amp;rtpof=true&amp;sd=true</a>
Результаты опросов в формате csv для анализа с использованием Python	<a href="https://drive.google.com/file/d/1AY51oHQ6pWoq0UwZV7nOE7C5a0x4-cPy/view?usp=sharing">https://drive.google.com/file/d/1AY51oHQ6pWoq0UwZV7nOE7C5a0x4-cPy/view?usp=sharing</a>
Excel документ с данными и графиками	<a href="https://drive.google.com/file/d/1xoVNTfeltE70ufJFaQZyPKGQ7vHULg6J/view?usp=sharing">https://drive.google.com/file/d/1xoVNTfeltE70ufJFaQZyPKGQ7vHULg6J/view?usp=sharing</a> необходимо скачать для просмотра графиков
Colab блокнот с графиками	<a href="https://colab.research.google.com/drive/1fhJt8KFJ4GPaUDaT2sH5mXDSENP75RY?usp=sharing">https://colab.research.google.com/drive/1fhJt8KFJ4GPaUDaT2sH5mXDSENP75RY?usp=sharing</a>
Colab блокнот методом KNN	<a href="https://colab.research.google.com/drive/1of7a5hVPG4QA9TDKTVufkzj3DfjP6nUG?usp=sharing">https://colab.research.google.com/drive/1of7a5hVPG4QA9TDKTVufkzj3DfjP6nUG?usp=sharing</a>
Презентация	<a href="https://docs.google.com/presentation/d/1OT6iioRzPQEgae8Acp35xRIJzgifNLf1IkXh_J8t9cs/edit?usp=sharing">https://docs.google.com/presentation/d/1OT6iioRzPQEgae8Acp35xRIJzgifNLf1IkXh_J8t9cs/edit?usp=sharing</a>

## Список литературы

1. Калинина С.Н. Использование кейс-технологии в образовательном процессе.  
<https://dopobr.68edu.ru/archives/14376>
2. Введение в кейс-метод: что такое кейсы и зачем они нужны.  
<https://challengelenge.com/article/chto-takoe-keysy/>
3. Таланов С.Л. Влияние социальных сетей на успеваемость студентов. Социальнополитические исследования, 2009 №2(11), стр. 117–135
4. Журавлева Д. М. Вконтакте или В учебе? Влияние социальных сетей на успеваемость студентов факультета экономики НИУ ВШЭ. Аннотация ВКР ВШЭ, <https://www.hse.ru/edu/vkr/125363404>
5. Гладких И. В. Методические рекомендации по разработке учебных кейсов. Вестник Санкт-Петербургского университета Сер.8, Вып.2 (№16).
6. Малый академический словарь / Ред.: А. П. Евгеньева, М., 1984
7. Глебов С. Ученые заявили о вреде социальных сетей. Новостной портал YTRO, 08.09.2010  
<https://ytro.news/articles/2010/09/08/920744.shtml>
8. Сизикова, Т.Э. Кейс-метод как средство развития рефлексии / Т.Э. Сизикова, О. А. Дураченко // Психология. Психофизиология. - 2019. - Т. 12, № 2. - С. 18–28.