

Практическая работа к уроку «Что такое данные и как с ними работать?»

Данные нужны для получения информации, на основе которой принимается решение. Для этого данные надо обработать (проанализировать и визуализировать), верно интерпретировать результаты анализа и далее предпринять то или иное действие (например, принять решение на основе данных).

Наука о данных изучает проблемы анализа, обработки и представления данных в цифровой форме. На практическом занятии мы проанализируем данные из открытых источников и соберем социологические данные при помощи онлайн-опроса.

Практическая работа 1. Поиск данных в открытых источниках

Существуют открытые (доступные) источники данных, размещенные в сети интернет в виде систематизированных данных, организованных в формате, обеспечивающим их автоматическую обработку без предварительного изменения человеком, в целях неоднократного, свободного и бесплатного использования.

Познакомимся с открытыми ресурсами данных.

Работа с открытыми государственными данными

Задание 1.1 Поиск на портале открытых данных Министерства культуры РФ.

Ежегодно в апреле в рамках национального проекта «Культура» по всей России проводится акция «Всероссийский день заботы о памятниках истории и культуры», приуроченная к Международному дню защиты памятников и исторических мест.

Найдите на территории своего региона, города, района объекты культурного наследия, общественно-значимые, исторические и культурные памятники, которые нуждаются в облагораживании и благоустройстве. Примите участие в акции!



ТЕХНОЛОГИЯ РАБОТЫ

1. Для поиска данных о памятниках вашего региона зайдите на портал открытых данных Министерства культуры РФ <https://opendata.mkrf.ru/opendata>.
2. Выберите набор данных Памятники - Сведения из Единого государственного реестра объектов культурного наследия (памятников истории и культуры) народов Российской Федерации.

Сортировать по востребованности Показывать по: 50

Сведения из Единого государственного реестра объектов культурного наследия (памятников истории и культуры) народов Российской Федерации паспорт -

1 - Памятники На карте Археология

Реестр лицензий на осуществление деятельности по сохранению объектов культурного наследия паспорт -

1 - Памятники Министерство

Реестр аттестованных специалистов в области сохранения объектов культурного наследия паспорт -

1 - Памятники

Реестр аттестованных экспертов по проведению государственной историко-культурной экспертизы паспорт -

1 - Памятники

Все наборы

- Археология
- Библиотеки
- ДК и клубы
- Кино
- Министерство
- Музеи
- Музыка
- На карте
- Образование
- Памятники**
- Парки
- Статистика
- Театры
- Цирки

3. Во вкладке Таблица выберите из списка Регион название вашего региона - Применить. Данные можно вывести в виде таблицы, а можно отобразить их на карте (вкладка Карта).

ПАСПОРТ ТАБЛИЦА КАРТА

Фильтр Поиск по таблице НЕ НАШЛИ ЗА

Объект Содержит

Номер в реестре Содержит

Регион Содержит

Полный адрес Содержит

АЛТАЙСКИЙ КРАЙ

ПРИМЕНИТЬ



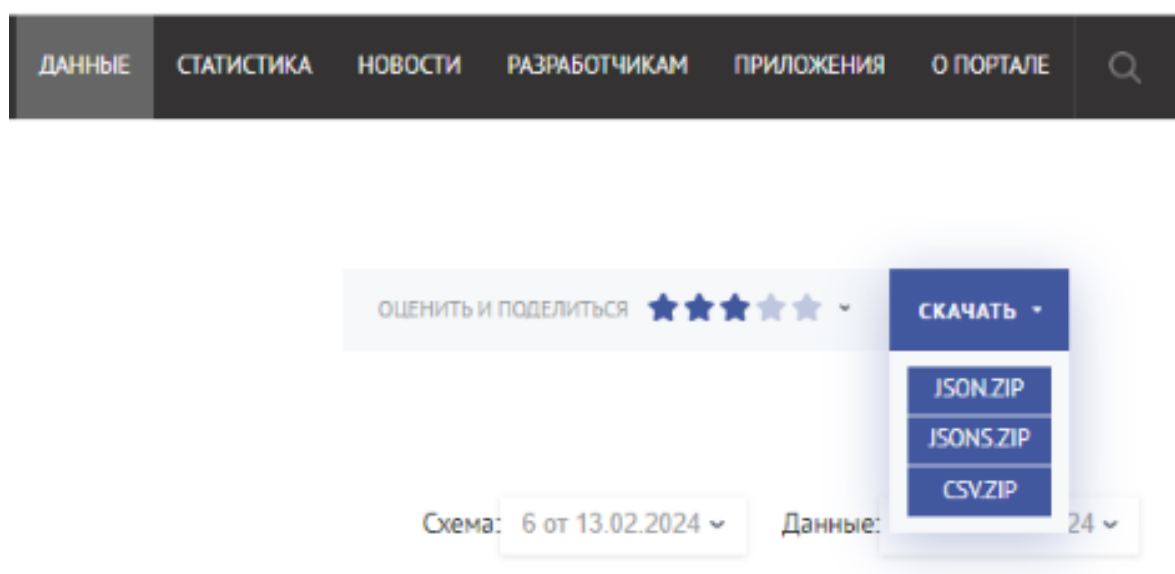
4. Изучите список памятников выбранного региона. На рисунке представлен список памятников Алтайского края. На странице отображается по 25 наименований. Общее количество представлено внизу страницы, для Алтайского края это 4 512 наименований. В таблице представлены сведения о памятнике и его географическое положение. Выберите памятник, который находится ближе всего к вашему образовательному учреждению и примите участие в акции «Всероссийский день заботы о памятниках истории и культуры».

Сведения из Единого государственного реестра объектов культурного наследия (памятников истории и культуры) народов Российской Федерации

Объект	Номер в реестре	Регион	Полный адрес	Категория историко-культурного значения	Вид объекта	Принадлежность к Юнеско	Особо ценный объект	На карте
Дом жилой	021410014280005	Алтайский край	Алтайский край, г.Барнаул, район Центральный, улица Пролетарская, 74	Регионального значения	Памятник	нет	нет	
Барнаульский строительный техникум	221210002970005	Алтайский край	Алтайский край, г.Барнаул, район Октябрьский, проспект Ленина, д. 68 (Алтайский край, г.Барнаул, пр. Ленина, д. 68)	Регионального значения	Памятник	нет	нет	
Бывший дом купца Шестакова	021410019790005	Алтайский край	Алтайский край, Усть-Пристанский район, с. Усть-Чарышская Пристань, улица 1 Мая, д. 19	Регионального значения	Памятник	нет	нет	
Прогимназия мужская	221210003040005	Алтайский край	Алтайский край, г.Барнаул, район Центральный, ул. Чкалова, д. 194	Регионального значения	Памятник	нет	нет	
Дом жилой	221310007810005	Алтайский край	Алтайский край, г.Барнаул, площадь им. В.Н.Баварина, д.17(Алтайский край, г.Барнаул, ул. Гудяева, д.5)	Регионального значения	Памятник	нет	нет	
Торговый дом Гилева	231410037210005	Алтайский край	Алтайский край, г.Бийск, ул. Советская, 4	Регионального значения	Памятник	нет	нет	
Высшая начальная школа им. Тургенева И.С.	221210003050005	Алтайский край	Алтайский край, г.Барнаул, ул. Аванесова, д. 152, 6 (Алтайский край, г.Барнаул, ул. Аванесова, д. 152)	Регионального значения	Памятник	нет	нет	
Енисейское 4	221440156410006	Алтайский край		Федерального значения	Памятник	нет	нет	

Показаны с 1 по 25 из 4512 записей

5. Для работы с данными в режиме офлайн, то есть без доступа к интернету, нужно скачать файл в одном из предоставляемых форматов. Например, в формате CSV. Скачайте файл CSV.ZIP, разархивируйте и откройте. Внимание: файл содержит большое количество данных и его открытие может вызвать переполнение памяти компьютера.



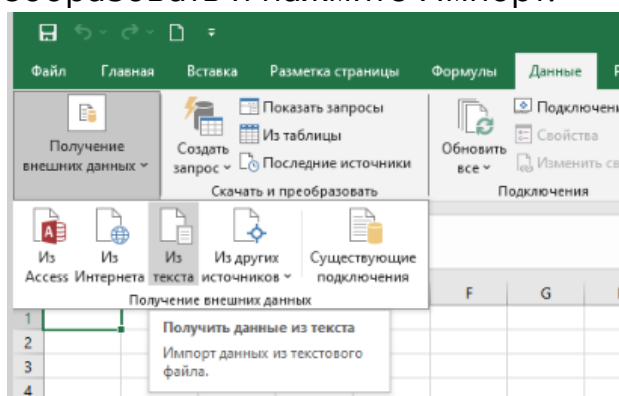
Задание 1.2 Поиск на портале открытых данных Министерство просвещения Российской Федерации

Во всех регионах РФ работают детские технопарки – учреждения дополнительного образования для поддержки одаренных учащихся, развития у них интереса к научно-технической и инновационной деятельности, стремления к личностным научным достижениям.

Найдите на территории своего региона детский технопарк, зайдите на его сайт, познакомьтесь с информацией о предлагаемых направлениях.

ТЕХНОЛОГИЯ РАБОТЫ

1. Для поиска данных о детских технопарках своего региона зайдите на портал открытых данных Министерства просвещения Российской Федерации <https://opendata.edu.gov.ru/opendata/>.
2. В разделе Реестр архивных наборов открытых данных выберите пункт Перечень детских технопарков (номер 53, но может измениться) и перейдите на страницу данных.
3. Посмотрите сведения о данных, дату их обновления. Скачайте данные. Гиперссылка (URL) на открытые данные <https://opendata.edu.gov.ru/opendata/7710539135-DT/data-20230907T155035-structure-20220401.csv>
Обратите внимание, что файл имеет формат CSV. Файл автоматически сохраняется в папке Загрузки.
4. Откройте файл data-20220404T135158-structure-20220401.csv. Файл открывается в табличном редакторе, например Эксел и имеет нечитабельный вид.
5. Откройте файл с конвертацией его в формат Эксел. Вкладка Данные – Получение внешних данных – Из текста. Выберите файл, который нужно преобразовать и нажмите Импорт.



6. В диалоговом окне выберите формат данных с разделителями – Далее.



Мастер текстов (импорт) - шаг 1 из 3

Данные восприняты как список значений фиксированной ширины.
Если это верно, нажмите кнопку "Далее >", в противном случае укажите формат данных.

Формат исходных данных

Укажите формат данных:

☒ с разделителями — значения полей отделяются знаками-разделителями

☐ фиксированной ширины — поля имеют заданную ширину

Начать импорт со строки: 1 Формат файла: 65001 : Юникод (UTF-8)

☐ Мои данные содержат заголовки

Предварительный просмотр файла C:\Users\KRUZHOK36\Downloads\data-20230907T155035-structure-20220401

1	"shortname", "name", "region", "town", "site", "edudirection"
2	"Детский технопарк «Кванториум», Барнаул", "КТБУ ДО «Детский технопарк Алтайского к
3	"Детский технопарк «Кванториум», Бийск", "МБУ «Молодёжный центр «Родина», "Алтайски
4	"Детский технопарк «Кванториум», Благовещенск", "ГАУ ДПО «Амурский областной инстит
5	"Детский технопарк «Кванториум», Свободный", "ГАУ ДПО «Амурский областной инстит

< >

Отмена < Назад **Далее >** Готово

7. Выберите разделители текста запятая и табуляция и нажмите Готово.

Мастер текстов (импорт) - шаг 2 из 3

В этом диалоговом окне можно установить разделители для текстовых данных. Результат выводится в окне образца разбора.

Символом-разделителем является:

☒ знак табуляции

☐ точка с запятой

☒ запятая

☐ пробел

☐ другой:

☐ Считать последовательные разделители одним

Ограничитель строк:

Образец разбора данных

shortname	name
Детский технопарк «Кванториум», Барнаул	КТБУ ДО «Детский технопарк Алтайского к
Детский технопарк «Кванториум», Бийск	МБУ «Молодёжный центр «Родина»
Детский технопарк «Кванториум», Благовещенск	ГАУ ДПО «Амурский областной институт ра
Детский технопарк «Кванториум», Свободный	ГАУ ДПО «Амурский областной институт ра

< >

Отмена < Назад **Далее >** Готово

8. На листе отобразятся данные по всем технопаркам РФ. Найдите свой регион, изучите данные о технопарке.

Задание 1.3 Поиск открытых данных на портале Федеральной службы государственной статистики

Федеральная служба государственной статистики (Росстат) является федеральным органом исполнительной власти, осуществляющим функции по формированию официальной статистической информации о социальных, экономических, демографических, экологических и других общественных процессах в Российской Федерации, а также в порядке и случаях, установленных законодательством Российской Федерации по контролю в сфере официального статистического учета.

Познакомимся с данными Росстата по переписи населения.

ТЕХНОЛОГИЯ РАБОТЫ

1. Зайдите на сайт Росстата раздел Статистика <https://rosstat.gov.ru/statistic>
2. Выберите пункт Переписи населения - Всероссийская перепись населения 2020 года.
3. Выберите пункт **Итоги ВПН-2020. Том 1 Численность и размещение населения**. Скачайте и откройте файл *6. Группировка городских округов по численности населения по субъектам российской федерации*.
4. Найдите данные по вашему региону. Ответьте на вопросы:
 - ◆ Сколько в вашем регионе городских округов с числом жителей 1 млн и более?
 - ◆ Сколько в вашем регионе городских округов всего?
 - ◆ Кого в регионе проживает больше мужчин или женщин?
 - ◆ Городских округов с каким числом жителей более всего в регионе?
 - ◆ Какого населения больше – городского или сельского?

Работа с открытыми данными компаний

Задание 1.4 Поиск на портале открытых данных Домклик

Домклик это онлайн-сервис от Сбербанка (сайт и мобильное приложение) для работы с недвижимостью - поиск, покупка, продажа и аренда недвижимости, проведение сделок по недвижимости.

Познакомимся с данными по недвижимости вашего региона.

ТЕХНОЛОГИЯ РАБОТЫ

1. Зайдите на сайт открытых данных Домклик <https://opendata.domclick.ru/>
2. Посмотрите рейтинг регионов по количеству объявлений о продаже: Россия и ответьте на вопросы.
 - ♦ Найдите ваш регион и сравните данные по продаже недвижимости с соседними регионами.
 - ♦ При помощи сортировки выведите 10 самых популярных и 10 менее востребованных регионов по полю Активных объявлений о продаже. Как можно объяснить полученные данные?

№	Регион	Активных объявлений о продаже	Активных объявлений, вторичка	Активных объявлений, новостройки
85 ^{↓5}	Ненецкий Автономный округ	81	56	—
84 ^{↓4}	Чукотский Автономный округ	135	123	—

- ♦ В каком регионе быстрее всего продаются квартиры, а в каком дольше всего? Сортировка по полю Средний срок экспозиции, вторичка.
- ♦ Посмотрите количество активных объявлений, цены на жилье год назад.

Рейтинг регионов по количеству объявлений о продаже: **Россия**

 Март 2023

Где еще искать данные



1. Каталог данных <https://www.datacatalogs.ru/>. Проект от Инфокультура. В каталоге систематизированы государственные источники открытых данных, а также собраны альтернативные базы, каталоги, репозитории, порталы, наборы и иные источники данных, потенциально интересные для всех, кто работает с данными.
2. ЕАЭС <https://opendata.eaeunion.org/opendata/ru>. Портал общих информационных ресурсов и открытых данных. Инструмент для обеспечения доступа к актуальным данным общих информационных ресурсов ЕАЭС, а также информирования граждан, бизнес-сообщества, некоммерческих и международных организаций о преимуществах и практических возможностях Евразийского экономического союза.
3. RAEX («РАЭК-Аналитика») <https://raex-rr.com/> — подготовка рейтингов в сфере образования, некоммерческого сектора, устойчивого развития, а также инфраструктуры развития бизнеса.

Практическая работа 2. Работа с данными в концепции Data Science (сбор, подготовка, обработка, анализ, коммуникация) на примере организации социологического опроса.

Практическое занятие можно провести совместно с учителем обществознания.

Порядок работы с данными в Data Science разделен на пять этапов: сбор, подготовка, обработка, анализ, коммуникация, что идеально вписывается в технологию проведения социологического исследования.

Сбор. Определение объема и структуры данных и их сбор в удобной для обработки форме. Определение целей анализа данных и подбор нужных методов для следующих этапов.

Подготовка. Проверка и очистка от нерелевантных сведений. К нерелевантным данным относятся отсутствующие значения, посторонние данные и другие данные, которые выбиваются из общего набора.

Обработка. Классификация (разделение на выборки и группы), применение математических моделей и программных методов, например машинного обучения.

Анализ. Проведение качественного и количественного анализа, подтверждение или опровержение поставленных изначально гипотез.

Коммуникация. Визуализация данных, создание аналитических отчетов и предложений для последующего принятия решений.

Проведем в классе социологический опрос, посвященный мероприятию «Неделя высоких технологий и технопредпринимательства (НВТиТ)». Гипотеза исследования «Учащиеся, участвующие во внеклассных мероприятиях по научной и технологической тематикам, более успешны в учёбе».

Для сбора социологических данных применим опрос, так как он реализуется через заполнение анкеты (ручной сбор данных), которую можно разместить онлайн, а значит в опросе могут принять респонденты из разных уголков страны и мира.

Учитель демонстрирует работу анкеты опроса по теме «Влияние внеклассных мероприятий по научной и технологической тематикам на уровень знаний школьников», созданную ранее, и просит учащихся пройти опрос. После прохождения опроса учитель демонстрирует его результаты и просит учащихся сделать выводы о том, есть ли зависимость между участием и высокими оценками или нет, подтвердить или опровергнуть гипотезу. Для каждой целевой аудитории (профильный класс или обычный, учащийся 8 класса или 10 класса) результаты исследования могут быть различными.

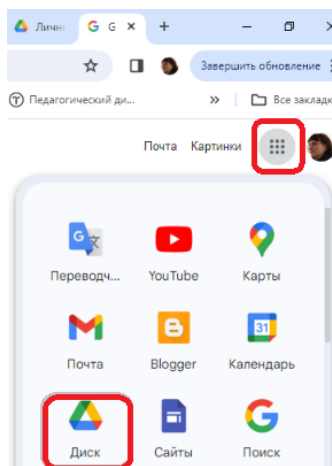
Далее учитель предлагает учащимся придумать тему собственного социологического исследования, сформулировать его гипотезу и создать форму для анкетирования респондентов, провести опрос и сделать выводы.

Задание 2.1. Сбор первичных данных при помощи Яндекс или Гугл формы. Ручной ввод данных.

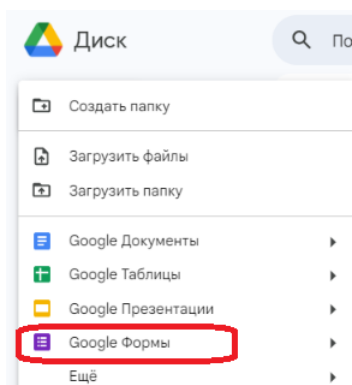
Создание онлайн-анкеты для первичного сбора данных средствами онлайн-сервисов. Пример анкеты <https://forms.gle/xvUF8yz3XDB4Z2iSA>
Для создания онлайн опроса можно использовать любой инструмент. Воспользуемся Гугл формой и создадим анкету для ручного ввода данных об опрашиваемых (респондентах).

ТЕХНОЛОГИЯ РАБОТЫ

1. Зайдите на сайт
2. Авторизуйтесь и перейдите в Приложения Google. Выберите приложение Диск.



3. Выберите команду Создать – Google Формы.



4. В новом окне введите название опроса, его краткое описание, информацию о возрасте опрашиваемых, иную полезную информацию.
5. Введите вопросы анкеты. Можно набрать вопросы анкеты, приведенной в примере. В опросе всего 4 вопроса, три вопроса имеют тип **Один из списка** и один вопрос подразумевает ввод текста – тип **Текст (абзац)**. Включите опцию **Обязательный вопрос**.



Влияние внеклассных мероприятий по научной и технологической тематикам на уровень знаний школьников

Социологический опрос проводится в рамках мероприятия «Неделя высоких технологий и технопредпринимательства (НВТиТ)». Приглашаются все заинтересованные школьники любого возраста.

Укажите уровень образования

Один из списка

☐ Начальное общее образование (1-4 классы) X

☐ Основное общее образование (5-9 классы) X

☐ Среднее общее образование (10-11 классы) X

☐ Добавить вариант или [добавить вариант "Другое"](#)

Обязательный вопрос

6. Выбор типа вопроса в раскрывающемся списке

Влияние внеклассных мероприятий по научной и технологической тематикам на уровень знаний школьников

Социологический опрос проводится в рамках мероприятия «Неделя высоких технологий и технопредпринимательства (НВТиТ)». Приглашаются все заинтересованные школьники любого возраста.

Укажите уровень образования

Один из списка

☐ Начальное общее образование (1-4 классы)

☐ Основное общее образование (5-9 классы)

☐ Среднее общее образование (10-11 классы)

☐ Добавить вариант или [добавить вариант "Другое"](#)

Текст (строка)

Текст (абзац)

☒ Один из списка

☒ Несколько из списка

☒ Раскрывающийся список

Загрузка файлов

Шкала

7. Вопросы в анкету добавляют кнопкой Добавить вопрос

Добавить вопрос

8. Настройте сообщение, которое респондент будет видеть после завершения опроса. Вкладка Настройки – Презентация – Текст подтверждения – Изменить.



Вопросы Ответы **Настройки**

показывать ход выполнения

Перемешать вопросы ☐

после сдачи

Текст подтверждения **Изменить**

Спасибо за участие в опросе! После его завершения мы поделимся результатами исследования!

Показывать ссылку для повторного заполнения формы ☒

Посмотреть ответы ☐

Поделиться **обобщенными результатами** с респондентами (важная информация)

9. Иконка с палитрой позволяет настроить внешний вид формы – добавить изображения, изменить цвет фона. Иконка с глазом позволяет посмотреть, как будет выглядеть форма для тех, кто будет отвечать на вопросы. Если все готово, то нажмите на Отправить.



10. В окне **Отправить** выберите значок **Ссылка**, можно указать что нужна **Короткая ссылка**. Google сгенерирует URL ссылки, которую можно скопировать и отправить всем желающим пройти опрос.

Отправить ×

Собирать адреса электронной почты Отключить сбор адресов ▾

Как отправить: ☒ ☒ ☐

Ссылка

<https://forms.gle/v3LronUtB3Dzd5gP7>

☒ Короткий URL

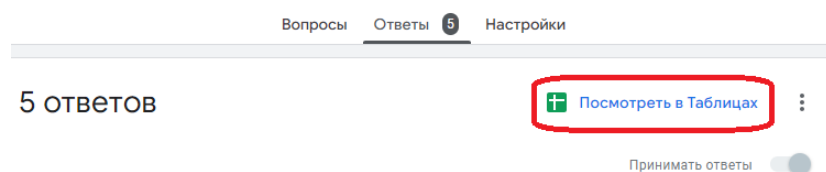
Отмена Копировать

11. Отправьте ссылку как можно большему числу респондентов, чтобы получить больше данных для анализа.

Задание 2.2. Проверка данных и их очистка от нерелевантных сведений.

После завершения анкетирования и получения сырых данных необходимо перейти к их обработке и анализу. Проверим собранные данные на нерелевантность, то есть на отсутствующие значения, посторонние данные и другие данные, которые выбиваются из общего набора.

1. Данные опроса доступны в виде диаграмм, также их можно скачать в виде Эксел таблицы – Посмотреть в Таблицах.



2. Таблица откроется на Диске, но ее можно скачать и работать с данными офлайн.
3. В опросе только одно поле содержало ввод данных с клавиатуры, все остальные значения выбирались из списка, а значит не могут содержать нерелевантных значений.
4. Посмотрите в Таблице поле с письменными ответами, удалите те, которые не относятся к теме опроса.

Задание 2.3 Обработка данных. Классификация (разделение на выборки и группы), применение математических моделей.

Данные опроса можно классифицировать по разным группам, например по уровню образования, или по участию во внеурочных мероприятиях. И уже внутри каждой группы провести обработку и анализ. Пример модели обработки полученных данных опроса приведен в таблице.

Уровень образования	Уровень включенности в мероприятия	Успеваемость
Начальное общее образование (1-4 классы)	Участвует	5 и 4 по всем предметам
		Учусь на 5 и 4 по естественным наукам, математике и информатике
		4 и 3 по всем предметам
		Учусь на 4 и 3 по естественным наукам, математике и информатике
	Не участвует	5 и 4 по всем предметам
		Учусь на 5 и 4 по естественным наукам, математике и информатике
		4 и 3 по всем предметам

		Учусь на 4 и 3 по естественным наукам, математике и информатике
Основное общее образование (5-9 классы)	Аналогично	Аналогично
Среднее общее образование (10-11 классы)	Аналогично	Аналогично

Задание 2.4 Проведение качественного и количественного анализа, подтверждение или опровержение поставленных изначально гипотез.

В уроке 2 будем учиться обрабатывать и анализировать сложные данные. Но даже сейчас можем подсчитать количество ответов по каждому варианту и принять решение о подтверждении гипотезы исследования на уровне вашего класса.

Качественный анализ данных делается по вопросу, где требовалось перечислить названия мероприятий, где принимали участие или причины неучастия. Из него можно сформировать список мероприятий и ознакомить с ним всех учеников, тем более, если в списке есть ответы о незнании о подобных мероприятиях.

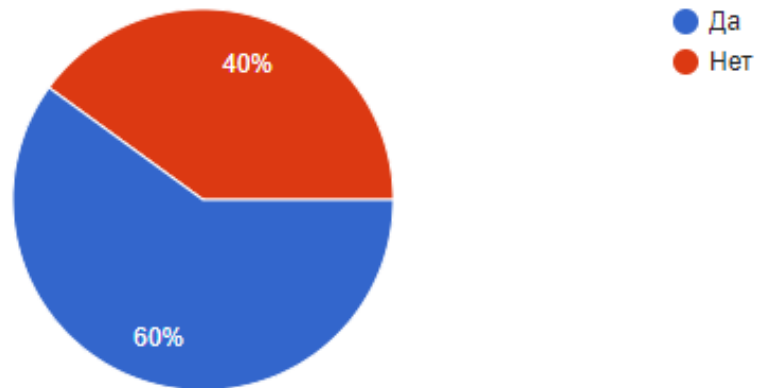
Задание 2.5. Визуализация данных, создание аналитических отчетов и предложений для последующего принятия решений.

В уроке 3 будем учиться визуализировать данные, строить интерактивные графики. Но даже сейчас можем увидеть диаграмм, которые построил Гугл по собранным данным. Круговые диаграммы не всегда удачно визуализируют данные, но для части данных они информативны, например показывают долю участия/неучастия в мероприятиях.



Участвуете ли вы в дополнительных мероприятиях по научной и технологической тематикам (олимпиады, хакатоны, фестивали, дни науки и т.п)?

5 ответов



Работа с данными в концепции Data Science (сбор, подготовка, обработка, анализ, коммуникация) на примере организации социологического опроса завершена.

Спасибо за работу!

