РАБОЧАЯ ТЕТРАДЬ

слушателя курса Карапетяна Хачатура

**«DATA-АНАЛИТИК: СТАРТ КАРЬЕРЫ»**

****

**22.07.2022—07.11.2022**

**2022**

# Оглавление

[Оглавление 2](#_Toc109743140)

[Как работать с этой тетрадью? 3](#_Toc109743141)

[Постановка целей 4](#_Toc109743142)

[Введение в анализ данных 5](#_Toc109743143)

[Раздел 1. Проектирование дизайна исследования в Big Data 6](#_Toc109743144)

[Как вытащить из данных максимум 10](#_Toc109743145)

[Уровень 1: дескриптивные (описательные) аналитические методы 10](#_Toc109743146)

[Уровень 2: диагностические аналитические методы 10](#_Toc109743147)

[Уровень 3: предиктивные (предсказательные) аналитические методы 10](#_Toc109743148)

[Уровень 4: прескриптивные (предписывающие) аналитические методы 10](#_Toc109743149)

[Определение дата-грамотности 10](#_Toc109743150)

[Элемент 1: чтение данных 10](#_Toc109743151)

[ЧТЕНИЕ ДАННЫХ ОТДЕЛОМ ИССЛЕДОВАНИЙ И РАЗРАБОТОК 10](#_Toc109743152)

[ЧТЕНИЕ ДАННЫХ ОТДЕЛОМ МАРКЕТИНГА 11](#_Toc109743153)

[ЧТЕНИЕ ДАННЫХ ТОП-МЕНЕДЖМЕНТОМ 11](#_Toc109743154)

[Элемент 2: работа с данными 11](#_Toc109743155)

[РАБОТА С ДАННЫМИ: IT-ОТДЕЛ 11](#_Toc109743156)

[РАБОТА С ДАННЫМИ: ОТДЕЛ МАРКЕТИНГА 11](#_Toc109743157)

[РАБОТА С ДАННЫМИ: ОТДЕЛ ПРОДАЖ 11](#_Toc109743158)

[РАБОТА С ДАННЫМИ: ТОП-МЕНЕДЖМЕНТ 11](#_Toc109743159)

[Элемент 3: анализ данных 12](#_Toc109743160)

[АНАЛИЗ ДАННЫХ: ОТДЕЛ ИССЛЕДОВАНИЙ И РАЗРАБОТОК 12](#_Toc109743161)

[АНАЛИЗ ДАННЫХ: ОТДЕЛ ПРОДАЖ 12](#_Toc109743162)

[АНАЛИЗ ДАННЫХ: ТОП-МЕНЕДЖМЕНТ 12](#_Toc109743163)

[Элемент 4: общение на языке данных 12](#_Toc109743164)

[ОБЩЕНИЕ НА ЯЗЫКЕ ДАННЫХ: ФИНАНСОВЫЙ ОТДЕЛ 12](#_Toc109743165)

[ОБЩЕНИЕ НА ЯЗЫКЕ ДАННЫХ: ДАТА-АНАЛИТИКИ 12](#_Toc109743166)

[ОБЩЕНИЕ НА ЯЗЫКЕ ДАННЫХ: ТОП-МЕНЕДЖМЕНТ 12](#_Toc109743167)

[Зонтик дата-грамотности 13](#_Toc109743168)

[Стратегия в сфере данных и аналитики 13](#_Toc109743169)

[Дата-грамотность и обработка данных 13](#_Toc109743170)

[Дата-грамотность и визуализация данных 13](#_Toc109743171)

# Как работать с этой тетрадью?

Уважаемые Слушатели!

Рабочая тетрадь — это пространство, в котором вы сохраняете все результаты своей интеллектуальной и практической деятельности. После окончания обучения, мы надеемся, она станет вашим главным гидом в профессии data-аналитика.

Когда вы проходите обучение в традиционном очном формате, вы изначально рассчитываете на то, чтобы структурировать и зафиксировать все важное содержание «здесь и сейчас». Ваше общение с экспертом-преподавателем ограничено. Если у вас есть вопросы — нужно задавать их сейчас, если вы не хотите потерять какую-то важную мысль — нужно записать её сразу, а не откладывая на потом.

В онлайн-пространстве мы часто полагается на то, что материалы — видеолекции, презентации — будут нам доступны в любое время. Но вспомните, как часто вы находите возможность, чтобы пересмотреть что-то ещё раз? Чаще всего учебные материалы скачиваются с намерением «обязательно посмотреть их когда-нибудь потом, когда будет время», хранятся, а потом теряются при переустановке операционной системы или поломке компьютера.

Поэтому мы хотим мотивировать вас работать в настоящий момент, когда вы приняли решение выделить время, чтобы обучиться новой профессии, активно создавать свои собственные, индивидуальные материалы, делать их полезными для себя самого в будущем.

В конце обучения — при планомерной работе — в вашем распоряжении останутся:

* конспекты видеолекций, в которых вы зафиксируете важное именно вам содержание,
* выводы и рекомендации, которые вы сформулируете в конце каждого раздела,
* подсказки и инструкции, которые помогут вам в дальнейшей работе,
* ссылки на блокноты Colab (облачную среду для работы с кодом), в которых прописаны шаблоны решений, которые вы сможете использовать для решения аналогичных профессиональных задач.

У вас в руках останется руководство, которое поможет освежить в памяти ваши знания и будет содержать достаточное количество рабочих шаблонов для того, чтобы приступить к работе, не открывая большое количество ссылок и не расходуя время на поиски и повторный просмотр нужного текстового или видеоматериала.

# Постановка целей

*«Кто не знает, куда направляется, очень удивится, попав не туда.»*

*Марк Твен*

Начало обучения — хорошее время, чтобы задуматься о своих целях, отделить их от искусственных и навязанных, расставить приоритеты и вспомнить о том, что вдохновляет вас начинать что-то новое.

Осознание личной цели поможет фокусироваться на том содержании, которое будет для вас особенно ценно, а также сохранять мотивацию на протяжении всего пути.

Сделайте это прямо сейчас — напишите о том, какие изменения ожидаете в своей жизни: профессиональные, карьерные, личностные, социальные.

|  |
| --- |
| **Я пришел на программу, потому что …** |
| **Сменить профессию, на более высокооплачиваемою и интересную.** |
|  |
| **Меня вдохновляет мысль о том, что …** |
| **Новая профессия даст мне больше свободы передвижения.** |
|  |
|  |
| **После обучения я смогу / буду уметь** |
| **Программировать и анализировать данные.** |
|  |
|  |
| **Это поможет мне …** |
| **Найти новую работу и трудоустроится.** |
|  |
|  |
| **Я буду доволен результатами, если …** |
| **Я найду удаленную высокооплачиваемую работу.** |
|  |
|  |

# Введение в анализ данных

1. Как изменилось количество данных в мире за последнее время?

2. Что обусловило такой рост количества данных?

3. Что отличает аналитический тип мышления?

4. Что такое data-driven подход?

5. Что характерно для data-driven подхода?

6. Какие задачи решает аналитик данных?

7. Какие виды аналитиков востребованы на рынке труда?

А. Бизнес аналитик:

Б. Системный аналитик:

В. Продуктовый аналитик:

Г. Маркетинговый аналитик или маркетолог:

Д. UX-аналитик user experience:

Е. BI аналитик:

8. В каких сферах активно используется аналитика данных?

# Раздел 1. Проектирование дизайна исследования в Big Data

**Вы научитесь** составлять план исследования, включающий:

* выбор методологии Data Mining
* определение бизнес-цели
* определение цели анализа данных
* формулировку требований к результату исследования
* определение типов, источников и способов получения данных
* выбор методов и инструментов анализа
* оценку рисков и условий реализации проекта

**Кейс**

В Центр прикладного анализа больших данных обратился Клиент N. Он занимается строительством частных домов в городе Т. более 10 лет. В год он строит в среднем 30 частных домов. Под частными дома подразумеваются, как дачные дома в формате «все включено», так и небольшие коттеджи, рассчитанные на круглогодичное проживание. Заказчик начинает работу от фундамента до готового решения, включая внутреннюю отделку. Средний срок сдачи проекта от 6 до 12 месяцев. Заказчик занимается строительством на участках своих клиентов, а не продает дома вместе с участком. Среди продуктовой линейки преобладают 1-2 этажные дома в среднем ценовом сегменте. Дома являются типовыми, клиенты выбирают из нескольких готовых решений и вносят индивидуальные корректировки.

Текущая целевая аудитория — мужчины и женщины старше 35 лет, средний уровень дохода относительно города присутствия. Семейный статус — женаты, обычно 1-2 ребенка. Есть автомобиль среднего класса. Ключевые ценности — семья и здоровье. Один из частых сценариев переезда за город - лучшая экологическая обстановка для здоровья детей. Для клиентов важно относительно небольшая удаленность от города для того, чтобы было удобно возить детей в школу или детский сад. Сами клиенты также работают в городе.

Наш Заказчик делает ставку на качественные материалы, соблюдение сроков и прозрачную систему оплаты. Также имеется возможность получения рассрочки, Заказчик сотрудничает с банками города, которые предоставляют низкую % ставку для его клиентов. У заказчика есть 5 основных конкурентов с точки зрения схожей продуктовой линейки (ассортимент) и ценового сегмента. Заказчик знает своих конкурентов и отдел маркетинга ежемесячно проводит конкурентный анализ их предложений.

В структуре компании есть финансовый отдел, отдел маркетинга, отдел продаж, производственный (строительный), инженерный и управленческий отделы. Основной канал коммуникации с аудиторией (маркетинг) на данный момент — это Интернет (социальные сети, контекстная реклама, сайт). Отдел маркетинга заточен именно под работу в цифровом пространстве для привлечения новых клиентов.

Наш Клиент уверен в своем опыте работы, его команда готова работать с дорогостоящими материалами. Кроме того, в город T стало переезжать больше состоятельных людей по сравнению с 5 годами раннее и его партнёр по бизнесу начал продажу земель в привлекательном месте. В связи с развитием города T и ростом уровня благополучия его горожан, наш клиент собирается развивать новую нишу бизнеса — строительство домов премиум-сегмента, цена которых в разы превосходит его имеющийся ассортимент. Заказчик знает, как работать с товарами среднего класса, однако совершенно не понимает потребности покупателей, которые могли бы заинтересоваться товаром в новой нише.

ПЛАН ИССЛЕДОВАНИЯ

**Основные задачи**1.Определить цель анализа данных  
2. Определить требования к результату исследования — то есть тот результат, который хотим получить мы и наш заказчик  
3. Понять — какие типы данных нам нужны  
4. Определиться с источниками и способами их получения  
5. Выбрать методы и инструменты анализа данных  
6. Оценить риски и условия реализации проекта

**Data Mining** — процесс обнаружения в данных ранее **неизвестных,  
нетривиальных, практически полезных и доступных интерпретации** знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

**Популярные методологии Data Mining** CRISP-DM KDD Process SEMMA

**Как выбрать?**• Условия работы  
• Зона ответственности  
• Требуемый результат

**Для чего?**• Сохранение опыта и воспроизводимость проектов  
• Упрощение процесса планирования и управления  
• Простота включения новых членов команды  
• Уменьшение зависимости от «лидеров»

**Методология SEMMA  
1. Sample** (отбор данных, т. е. создание выборки)  
**2. Explore** (исследование отношений в данных)  
**3. Modify** (модификация данных)  
**4. Model** (моделирование взаимозависимостей)  
**5. Assess** (оценка полученных моделей и результатов)

**Методология KDD (Knowledge Discovery in Databases)**1. Отбор (Selection)  
2. Предварительная обработка (Preprocessing)  
3. Преобразование (Transformation)  
4. Data Mining  
5. Интерпретация (Intepretation | Evaluation)

**Методология Crisp-DM**  
1. Понимание бизнеса (business understanding)  
2. Понимание данных (data understanding)  
3. Подготовка данных (data preparation)  
4. Моделирование (modeling)  
5. Оценка результатов (evaluation)  
6. Внедрение (deployment)  
7. Контроль

**Основные процессы в методологиях  
Шаг 1.** Работа с заказчиком  
**Шаг 2.** Работа с данными  
**Шаг 3.** Разработка аналитического решения

**Практика: выбираем методологию для нашего кейса**  
1. Какая наша роль в проекте? Мы работаем с клиентом от начала и до конца  
2. Это новая отрасль для клиента. Он еще не понимает и не имеет наработок по интересующему нас вопросу  
3. В качестве решения в первую очередь нам нужен аналитический отчет. Исследование проводится  
впервые по данному направлению, соответственно, разрабатывать какие-то модели еще рано

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Объем глобальных данных достигнет 175 миллионов петабайт в течение следующих пяти лет

**Озера данных** объединяют и защищают данные из разных источников и форматов в единое хранилище для легкого доступа и анализа.

В зависимости от варианта использования озера данных могут быть построены на базе объектного хранилища или Hadoop. Оба они могут масштабироваться и легко интегрироваться с существующими корпоративными данными и инструментами.

Все эти данные в различных форматах должны быть усвоены и подготовлены для использования пользователями.

**Озеро данных** — это гибкий механизм хранения необработанных данных.

К 2025 году почти 60% существующих данных будут создаваться и управляться корпоративными организациями (по сравнению с 30% в 2015 году).

Каталоги данных необходимы для того, чтобы пользователи могли осуществлять поиск и находить данные для анализа.

Каталоги данных с машинным обучением представляют текущие наборы данных и соответствующие данные благодаря оптимизированному управлению метаданными.

99,5% собранных данных остаются неиспользованными, в первую очередь из-за отсутствия инфраструктуры, ресурсов и управления.

**Анализ всех данных**

Анализ данных из одного источника ограничивает понимание. Озера данных объединяют данные из нескольких источников, включая данные со всего бизнеса, позволяя пользователям анализировать все доступные данные.

В среднем для анализа и принятия решения, основанного на данных, проводится консультация с 5 источниками данных.

**Ускорение времени принятия решений**

Используя озера данных, возможности машинного обучения и аналитики, пользователи получают интеллектуальную информацию, которая помогает принимать бизнес-решения, основанные на данных.  
Заметки из книги Морроу Джордан

# [Как вытащить из данных максимум](#_Как_вытащить_из)

[Навыки аналитики для неспециалистов-2022](#_Как_вытащить_из)

Основа мира аналитики — четыре уровня: дескриптивный (описательный), диагностический, предиктивный (предсказательный) и прескриптивный (предписывающий).

## Уровень 1: дескриптивные (описательные) аналитические методы

**Дескриптивный анализ** позволяет нам оглянуться на то, что уже произошло в бизнесе, и изучить  
это при помощи данных.

**Методы** — отчеты, работа со сводками или информационными панелями, наблюдения.

**KPI** (ключевые показатели эффективности)

## Уровень 2: диагностические аналитические методы

Понятие, неразрывно связанное с диагностическим анализом, — это *инсайт*, проникновение в суть  
данных, понимание движущих сил и причин происходящего.

Ключевое слово здесь — **«инсайт».** Именно проникновение в суть помогает понять, как первый и второй уровни аналитики действуют вместе, и объединить их. Первый уровень описывает, ***что* произошло**, — и это ведет ко второму уровню, к попыткам разобраться, ***почему* это произошло**.

## Уровень 3: предиктивные (предсказательные) аналитические методы

Итак, **спрогнозировать (или предсказать)** — это «заявить или предположить, что некое событие произойдет в будущем само по себе или как следствие чего-либо».

«обработка данных», «статистика», «машинное обучение», «алгоритм», «большие данные» и т. д.? Именно эти понятия и составляют третий уровня аналитики (и часть четвертого).

## Уровень 4: прескриптивные (предписывающие) аналитические методы

Данные или технологии *предписывают* нам (или советуют), как поступить.

Есть много компьютерных программ и сервисов — от более простых и понятных Domo и Alteryx до таких продвинутых, как SAS или SAP Predictive Analytics.

Эти инструменты способны стать прекрасным подспорьем в прескриптивном анализе, но, если у вас нет сотрудников, которые умеют интерпретировать результаты работы программ и принимать решения самостоятельно, вложения в ПО могут оказаться бессмысленными.

## Определение дата-грамотности

Data science в чистом виде — это применение к данным научного метода.

Итак, **дата-грамотность** — это способность читать данные, работать с ними, анализировать их и использовать как аргумент.

**Дата-грамотность** — это способность читать данные, работать с ними, анализировать их и общаться на языке данных.

# Элемент 1: чтение данных

Первая составляющая дата-грамотности — это **чтение данных**.

## ЧТЕНИЕ ДАННЫХ ОТДЕЛОМ ИССЛЕДОВАНИЙ И РАЗРАБОТОК

Сотрудникам необходимо читать, понимать и использовать очень много данных. В нашем случае команда потратила много времени и сил на сбор внутренних и внешних данных — и в итоге,  
изучив результаты опросов, а также сведения о конкурентах и рыночной ситуации, смогла оценить жизнеспособность нового продукта и другие факторы. Вы и сами понимаете, что при  
чтении данных и поиске информации, необходимой для принятия решения, сотрудники использовали как дескриптивные, так и диагностические аналитические методы.

## ЧТЕНИЕ ДАННЫХ ОТДЕЛОМ МАРКЕТИНГА

От сотрудников требовалось разработать маркетинговую стратегию и систему распространения информации о новом продукте. Им пришлось изучить и понять горы данных, полученных от самой компании, а также проанализировать внешние тенденции, связанные с продуктом. Кампании какого рода были успешными в прошлом? А какие провалились? Какие внешние обстоятельства могут повлиять на запуск продукта? Умение использовать как дескриптивные, так и диагностические методы помогло отделу маркетинга сформировать стратегию для успешного запуска продукта.

## ЧТЕНИЕ ДАННЫХ ТОП-МЕНЕДЖМЕНТОМ

Для принятия серьезных решений умение читать данные критически важно. Всем известно, что у руководства мало свободного времени, а данных, требующих изучения, очень много. Топ-менеджеры должны уметь быстро читать и оценивать данные, чтобы принимать на их основе правильные решения. В нашем случае команда топ-менеджеров смогла быстро прочесть и проанализировать информацию о новом продукте, чтобы принять обоснованное решение, подкрепленное данными.

# Элемент 2: работа с данными

Работа с данными — это действия с данными ради достижения цели или результата.

В работе с данными на четырех уровнях аналитики есть много общего, но вместе с тем каждый уровень обладает своими уникальными особенностями.

## РАБОТА С ДАННЫМИ: IT-ОТДЕЛ

Для начала возьмем IT-специалистов. Нужно ли им работать с данными, чтобы помочь в запуске подобной кампании? Разумеется! В рамках этой конкретной кампании IT-отдел должен был и получить, и сделать доступными данные, необходимые для принятия обоснованных решений. Команда работала с данными самыми различными способами. В результате конечные пользователи получили возможность анализировать и использовать данные для обеспечения успеха кампании.

## РАБОТА С ДАННЫМИ: ОТДЕЛ МАРКЕТИНГА

Далее рассмотрим саму команду маркетологов. Требуется ли от них работа с данными? Конечно! Маркетологи должны работать с данными, чтобы провести описательный анализ. Им нужно определить тенденции и закономерности в данных, касающихся как самой организации, так и рынка в целом. Именно данные помогут им составить прогноз успешности кампании.

## РАБОТА С ДАННЫМИ: ОТДЕЛ ПРОДАЖ

Теперь обратимся к отделу продаж. Команда «продажников» находится в авангарде: именно она непосредственно взаимодействует с существующими и потенциальными клиентами. Отдел продаж отвечает на вопросы и проводит исследования для кампании, изучает новые продукты, доступные для клиентов организации, и решает, как использовать всю эту информацию 62 КАК ВЫТАЩИТЬ ИЗ ДАННЫХ МАКСИМУМ для достижения успеха. «Продажники» должны уметь эффективно обращаться с данными, лежащими в основе маркетинговой кампании, а другие подразделения организации должны делиться с ними теми данными, которые есть у них.

## РАБОТА С ДАННЫМИ: ТОП-МЕНЕДЖМЕНТ

И наконец, топ-менеджеры. Им непременно нужно работать с данными при запуске новой кампании: они никогда не занимались ничем подобным, это беспрецедентная кампания. Топ-менеджеры получают отчеты и другую информацию, что помогает им принимать решения, подкрепленные данными (обратите внимание: наша мозаика — четыре уровня аналитических методов — успешно складывается!). Получив данные, руководство начинает с ними работать. Естественно, это необходимо для успешного запуска кампании.

# Элемент 3: анализ данных

**Анализ** помогает нам отделять нужное от ненужного в потоке информации.

**Анализ** — это подробное изучение элементов или структуры  
чего-либо.

## АНАЛИЗ ДАННЫХ: ОТДЕЛ ИССЛЕДОВАНИЙ И РАЗРАБОТОК

Сотрудники анализируют не только внутреннюю информацию, поступающую из других подразделений компании, но и внешние данные. Например, представьте, что вы запустили новый продукт именно тогда, когда в экономике начался спад. Кто-то скажет, что это была изначально плохая идея и что продукт не мог не провалиться, но так ли это? Если внешние данные свидетельствуют, что ситуация на рынке в целом неблагоприятна, это действительно могло повредить запуску. Однако отдел исследований и разработок должен поставить правильные вопросы и проанализировать информацию, чтобы составить заключение  
об успешности запуска.

## АНАЛИЗ ДАННЫХ: ОТДЕЛ ПРОДАЖ

Они должны задавать вопросы, исследовать различные факторы и изучать огромное количество самых разных данных, чтобы определить, был ли запуск успешным.

## АНАЛИЗ ДАННЫХ: ТОП-МЕНЕДЖМЕНТ

Топ-менеджеры управляют компанией: если они не анализируют успешность запуска каждого нового продукта, то чем они вообще занимаются? Стреляют наугад?  
Топ-менеджерам необходимо проанализировать большие объемы данных, чтобы определить степень успеха. Как запуск последнего продукта повлиял на чистую прибыль компании (и повлиял ли вообще)? Каковы объемы продаж нового продукта и насколько эффективно сработал отдел продаж? Удалось ли маркетологам повысить интерес к продукту? Столько вопросов — а это лишь  
верхушка айсберга

# Элемент 4: общение на языке данных

Нам нужно делиться или обмениваться информацией, чтобы описать, что случилось в прошлом, знаниями, полученными в результате диагностического анализа, прогнозами и указаниями на основании предиктивного и прескриптивного анализа.

## ОБЩЕНИЕ НА ЯЗЫКЕ ДАННЫХ: ФИНАНСОВЫЙ ОТДЕЛ

Финансовый отдел и должен рассказать остальным результаты общей работы. В данном случае команда финансистов может поделиться только результатами дескриптивного анализа (а это  
первый уровень). Финансовый отдел сообщает топ-менеджерам и другим сотрудникам компании о численных показателях, чтобы у всех сложилась верная картина происходящего.

## ОБЩЕНИЕ НА ЯЗЫКЕ ДАННЫХ: ДАТА-АНАЛИТИКИ

Они способны находить, анализировать и раскрывать то, что могли пропустить другие. Дата-аналитикам, как правило, нужно совершенствовать свои навыки дата-грамотности именно в сфере общения на языке данных. Вопрос в другом: нужно ли им это? Конечно же, да. Раньше общение не было приоритетом для тех, кто занимается обработкой данных. Однако в новом мире дата-грамотности все иначе, и дата-аналитикам нужно учиться общаться на языке данных со всеми сотрудниками своей организации.

## ОБЩЕНИЕ НА ЯЗЫКЕ ДАННЫХ: ТОП-МЕНЕДЖМЕНТ

Топ-менеджерам нужно уметь донести до всех сотрудников организации результаты различных типов анализа. Затем они должны рассказать о причинах успеха, о том, что планируется делать для сохранения положительной динамики, и т. д.

# Зонтик дата-грамотности

**Дата-грамотность** — это способность читать данные, работать с ними, анализировать их и общаться на языке данных.

Важнейшие области дата-грамотности:

* стратегия в сфере данных и аналитики;
* обработка данных;
* визуализация данных;
* топ-менеджмент;
* культура;
* качество данных;
* управление данными;
* этика и законодательство в сфере данных.

## Стратегия в сфере данных и аналитики

**Конечная цель** — это цели и задачи организации, а данные и аналитика — это инструменты и материалы, обеспечивающие успешную деятельность организации.

Стратегия обязательно должна включать в себя человеческий фактор дата-грамотности.

## Дата-грамотность и обработка данных

**STEM** (**Science** — естественные науки, **Technology** — научно-технические дисциплины, **Engineering** — инженерия, **Mathematics** — математика)

## Дата-грамотность и визуализация данных

Инструмент работы с данными — визуализация. Такие компании, как Qlik, Tableau Software и ThoughtSpot

**Тема 1: Методологии Data Mining**

1. Зачем нужны стандарты изучения данных?

2. Какие эффективные методологии (стандарты) изучения данных существуют на сегодняшний день?

А. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ :

Б. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ :

В. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ :

3. В чем заключаются их основные отличия:

4. Как выглядит общий алгоритм анализа данных:

5. На что нужно ориентироваться аналитику данных при выборе подходящей для конкретного исследования методологии?

6. Топ-5 выводов:

**Тема 2: Определение бизнес-цели**

1. Почему этап определения бизнес-цели заказчика так важен для планирования дизайна исследования?

2. Какую информацию — от заказчика или о заказчике — необходимо получить?

3. Как правильно сформулировать бизнес-цель?

4. Топ-5 выводов:

**Тема 3: Определение цели анализа данных**

1. Как перейти от бизнес-цели к цели анализа данных?

2. Как сформулировать цель анализа данных?

3. На какие вопросы необходимо ответить для постановки цели исследования?

4. Топ-5 выводов:

**Тема 4: Требования к результату исследования**

1. Почему в процессе разработки дизайна исследования необходимо утвердить требования к результату аналитического исследования?

2. Какие типы взаимодействия с заказчиком и предоставления результата можно выделить?

3. Какие вопросы необходимо обсудить, когда договариваешься о формате предоставления результатов исследования?

4. Топ-5 выводов:

**Тема 5: Типы данных**

1. Какие основные требования к данным мы должны сформулировать в начале проекта?

2. Что такое данные и чем они отличаются от информации?

3. Какие бывают типы данных?

4. Чем отличаются типы данных?

5. Что такое большие данные?

6. Топ-5 выводов:

**Тема 5.1: Форматы данных**

1. В каких основных форматах могут быть представлены данные?

**Тема 6: Способы получения и источники данных**

1. Какие бывают данные с точки зрения их получения?

2. Какие существуют способы получения данных?

3. Какие популярные источники данных существуют?

4. Что такое открытые данные?

5. Что необходимо учитывать при использовании открытых данных?

6. Топ-5 выводов:

**Тема 7: Выбор методов и инструментов анализа**

1. Какие методы анализа данных существуют и для решения каких задач они подходят?

1.1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_: применяется для …

1.2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_: применяется для …

1.3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_: применяется для …

1.4 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_: применяется для …

1.5. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_: применяется для …

1.6. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_: применяется для …

…?

2. На какие две группы можно условно разделить инструменты, применяемые для анализа данных?

3. Какие инструменты относятся к первой группе?

4. Какие инструменты можно отнести ко второй группе?

**Тема 8: Оценка рисков и условий реализации проекта**

1. Как оценить текущую ситуацию проекта?

Ресурсы:

Персонал:

Техника:

Данные:

2. Как оценить риски?

Риски:

Выявление:

Оценка:

Управление:

3. Почему важна единая терминология проекта?

4. Топ-5 выводов:

Поздравляем! Вы закончили изучение первого раздела курса :)

Теперь Вы можете приступить к выполнению первой части задания по стажировке — **Составлению плана исследования**.

*P. S. Подробную информацию можно найти в тетради по стажировке (Дисциплина «Стажировка» в ODIN).*