# Введение

Вы почти закончили обучение и дошли до завершающей части курса: большая часть теории и практики осталась позади. Мы гордимся вашими трудолюбием и упорством! Теперь осталось совсем немного — выполнить и защитить **итоговый дипломный проект.**

Некоторые считают диплом формальностью: «Я получил знания, а бумажка мне не важна». Но диплом — не просто свидетельство. Это ещё один кейс в ваше портфолио, который можно показывать работодателю. Это полезно, даже если прямо сейчас вы не ищете работу: например, для роста внутри компании и для закрепления ваших знаний, полученных в рамках курса.

Поэтому давайте соберёмся и сделаем это. Будет интересно!

# Анализ продаж розничной торговой сети

Перед вами подробное описание будущего проекта, который станет отличным дополнением вашего портфолио. Здесь есть всё, что вам нужно, чтобы справиться с поставленной задачей. А именно:

[Введение](#_td9o6twjefuq)

[Анализ продаж розничной торговой сети](#_r5vopnveppcs)

[Описание задачи](#_gjdgxs)

[ТЗ на дипломную работу](#_30j0zll)

[Обязательное задание](#_hzdt8zae132l)

[Предварительные замечания](#_1v3n2ftxwqba)

[1. Загрузка данных и их предварительный анализ.](#_sjvmsj8z3vxa)

[2. Исследовательский и статистический анализ данных с помощью библиотеки pandas.](#_3mvyo6ypgtra)

[3. Бизнес-анализ данных с помощью библиотеки pandas.](#_kaaqk14raeq2)

[4. Выводы и рекомендации по развитию бизнеса в виде Аналитической записки](#_52vy266jcn8a)

[Задание на выбор](#_e0o6xlrpvec8)

[5. Создание реляционной Базы данных и исследование данных с помощью SQL запросов](#_ciec1xe14qb1)

[Создание Базы Данных](#_1v3avha5e60g)

[Исследование данных с помощью библиотеки sqlite3 (SQL-запросы).](#_2eq7u7tgddvq)

[6. Создание дашборда](#_yqdrgbfy7hav)

[Дашборд с помощью python Matpoltlib](#_ci4p7svogye9)

[Дашборд в программе MS power BI Descktop](#_1j1ndmp2jqi7)

[7. Поиск зависимости в данных](#_vdj5e1m7acd4)

[Сдача и оценивание проекта](#_5konfhjz3lr7)

[Формат сдачи материалов](#_vijrw6eyctj7)

[Шаги сдачи проекта](#_n0tg6x88mntu)

[Оценивание Дипломной работы](#_h8osvvg9g44u)

[Приложения](#_til65s418mjn)

[Рекомендации и подсказки](#_12g3rrbg4124)

[Python и Pandas](#_yj34cue3dbxd)

[Сохранение и запись данных в csv формате](#_2epgo7aadgdr)

[Статистический анализ](#_9s9e3vpn6mtp)

[Типы данных](#_qymledyll5af)

[Выбор данных по значению](#_iw4ut1cyqb8o)

[Группировка (суммирование, подсчет значений, подсчет среднего) по отдельным полям](#_2pqybpl4odlf)

[Matplotlib](#_lgrzflmpmh8x)

[Sqlite3](#_oios5lmg8blr)

[Дополнительные материалы](#_17dp8vu)

[Рекомендации по выполнению](#_9pthi3e7tz81)

Последовательно изучите каждую часть. Периодически обращайтесь к актуальным для вас разделам в этом документе.

## Описание задачи

Вы - начинающий аналитик данных, вышли на работу в розничную торговую сеть (ритейл). Ваша первая задача - подготовить отчёт для Совета директоров по продажам компании за последние месяцы. Руководство заинтересовано в оценке дивизионов, выявлении наиболее доходных категорий товаров, хочет определить какие типы клиентов приносят наибольшую выручку и самое главное получить рекомендации по повышению рентабельности бизнеса на основе данных.

Также необходимо разработать и реализовать базу данных для этой задачи. Эту БД в дальнейшем предполагается пополнять по мере накопления данных о продажах.

Отчет и выводы нужно наглядно визуализировать и подготовить дашборд для регулярного отслеживания ситуации. Важно, чтобы из визуализаций было понятно почему предложена та или иная гипотеза и сформулированы именно такие рекомендации.

## ТЗ на дипломную работу

Задание дипломного проекта состоит из 4х заданий обязательной части и 3х заданий на выбор. Для успешного выполнения дипломной работы вам необходимо выполнить все пункты из обязательного и минимум один пункт из заданий на выбор. Конечно же, выполнение большего количества пунктов приветствуется и позволит вам лучше прокачать свои навыки.

Все задания, где необходимо программировать должны быть выполнены на языке Python и представлены на проверку в файлах с расширением .ipynb. В задании №5 из раздела “на выбор”, необходимо использовать СУБД SQLite, созданная БД должна быть с расширением .db “Аналитическую записку”, созданная в задании №4 обязательной части можно сделать в MS Word или в MS PowerPoint

Данные проекта находятся в файле [supermarket\_sales.csv](https://disk.yandex.ru/d/oOSX1kiUodaLuw), который доступен по ссылке [supermarket\_sales.csv](https://drive.google.com/file/d/1Z63TIf0rN6Oh0YXfhaQmqGKsgfJ7RBaj/view?usp=share_link).

### Обязательное задание

#### Предварительные замечания

Обязательное задание (кроме пункта 4) выполняется в одном файле **work.ipynb**

Код с необходимыми комментариями пишется в ячейках “Код”, ответы на вопросы, и выводы пишутся в ячейках “Текст” (в данном поле используется форматирование markdown, ознакомиться с инструкцией по работе с ним можно по ссылке: <https://www.markdownguide.org/cheat-sheet/> )

Перед началом работы над каждым пунктом задания должно быть поле с форматом “Текст” в котором нужно указать, к какой части задания и какому пункту (вопросу) относятся следующие ячейки с кодом и с выводами.

Питоновские скрипты пишутся в ячейках типа “Код”. Под ними (после вывода результатов выполнения), где это необходимо, добавьте ячейку типа “Текст” в котором сформулируйте выводы и ответы на вопросы.

Обязательное задание состоит из следующих частей:

#### Загрузка данных и их предварительный анализ.

1. Откройте юпитер ноутбук установленный локально или воспользуйтесь google colab.
2. Загрузите данные из .csv файла.
3. Проверьте типы данных. Если необходимо, поменяйте тип данных.
4. Проверьте есть ли дубликаты в данных. Если дубликаты встречаются, удалите их.

#### Исследовательский и статистический анализ данных с помощью библиотеки pandas.

1. Выведите информацию о категориальных типах данных, нужно вывести уникальные значения и количество записей с этим значением, ответить на вопросы:
   * Какие города представлены в данных (поле City)?;
   * Какие типы клиентов встречаются (Customer\_Type)?;
   * Какие категории товаров (Product\_Line)?;
   * Какие типы платежей (Payment).
2. Проанализируйте статистические характеристики датасета через метод .describe(). Примените этот метод ко всему датафрейму и ответьте на вопрос: нет ли аномалий[[1]](#footnote-0) среди минимального и максимального значений числовых признаков?.
3. Найдите и удалите строки с аномалиями (в комментарии укажите, какие значения вы отнесли к аномалиям и почему).
4. Проверьте, есть ли пропуски в данных. Если есть, то примите решение, удалить строки с ними или заполнить. Для числовых признаков заполните их средним по полю (после удаления аномалий, см. п.2.2), для категориальных[[2]](#footnote-1) наиболее частотным значением.

После выполнения пункта 1 и 2 Обязательного задания, сохраните полученный датафрейм в отдельный файл **clear\_data.csv** и вся дальнейшая работа должны вестись с использованием данных из этого файла.

#### Бизнес-анализ данных с помощью библиотеки pandas.

Проведите анализ очищенного в п. 1 и 2 датасета и ответьте на приведенные ниже вопросы. Сопровождайте ответы на вопросы числовыми значениями, полученными из данных.

1. Сравните выручку (Total) и маржинальную прибыль[[3]](#footnote-2) (Gross\_income) дивизионов (Branch) или по городам (City). Выведите суммы выручки и маржинальной прибыли в разрезе дивизионов, сравните между собой, посчитайте маржинальную рентабельность[[4]](#footnote-3) каждого дивизиона, напишите выводы (есть ли среди дивизионов отстающие и дивизионы-лидеры)?
2. Какая категория товара лидирует по числу продаж в ед. товара, а какая по суммарным продажам?
3. Женщины или мужчины приносят больше выручки, прибыли?
4. Отличаются ли мужской средний чек от женского? (\*”средний чек” - сумма средней покупки в ритейле)
5. Сравните выручку, маржинальную прибыль и маржинальную рентабельность по типам клиентов (по тем, кто участвует в программе лояльности и тем, на кого программа лояльности не распространяется, поле Customer\_type). Какие клиенты приносят больше прибыли?
6. Какие товарные группы (Product\_line) больше интересуют женщин, а какие мужчин?
7. Сравните выручку, маржинальную прибыль и маржинальную рентабельность разных товарных групп. Какие товарные группы самые прибыльные?
8. Оцените влияние расходов на осуществление платежей (Payments\_fee) на прибыльность и рентабельность товарных групп:

посчитай по каждой товарной группе расходы на прием платежей и оцените рентабельность с учетом этих расходов.

Ответ выглядит примерно следующим образом:

Расходы для приема платежей:

- для кредитных карт Х% от оборота

- для электронных кошельков Х% от оборота

- для кассы - Х% от оборота, но …

Расходы по приему платежей для товарной группы 1: ХХ, что уменьшает ее маржинальную рентабельность на Х%

Рентабельность по товарной группе 1 с учетом расходов на сбор денег: хх%

#### Выводы и рекомендации по развитию бизнеса в виде Аналитической записки

Выводы и рекомендации нужно сформулировать в виде “Аналитической записки”, адресатами которой будут члены Совета директоров компании.

Постарайтесь сделать так, чтобы ваша аналитическая записка была аккуратно оформлена, хорошо структурирована, логична, выводы подкреплены цифровыми данными. При знакомстве с вашим документом у адресата должно появиться доверие к приводимым данными и выводам.

“Аналитическую записку” необходимо предоставить в виде отдельного файла, допускаются форматы:

* MS Word
* MS Powerpoint

Все ответы должны быть подтверждены цифрами (таблицами) и графиками.

В аналитической записке обязательно должны присутствовать ответы на вопросы:

1. Положительная, отрицательная или стабильная динамика продаж по дивизионам?
2. Какие категории товаров приносят наибольшую прибыль и их следует активно продвигать?
3. нужно ли прикладывать усилия для привлечения в магазины больше женщин (мужчин)?
4. Нужно ли прикладывать усилия для расширения программы лояльности?
5. Отличаются ли мужской средний чек от женского? (\*”средний чек” - сумма средней покупки в ритейле)
6. \* если вы выполнили п. 7 из дополнительного задания, то обязательно отразите его результаты в вашей аналитической записке.

Аналитическая записка является результатом творческого труда аналитика, вы можете добавить собственные выводы, которые по вашему мнению будут интересны руководству компании.

В Аналитической записке необходимо сформулировать рекомендации по повышению рентабельности продаж.

### Задание на выбор

#### Создание реляционной Базы данных и исследование данных с помощью SQL запросов

##### **Создание** Базы Данных

Для целей дальнейшего развития проекта и хранения исторических данных, вам необходимо создать реляционную БД для хранения данных о продажах.

БД должна состоять из следующих таблиц::

* Справочники для категориальных признаков с ключевыми полями (primary key):
  + Branch (Дивизион), тип **TEXT**
  + City (Город), тип **TEXT**
  + Product\_Line (Продуктовая группа), тип **TEXT**
  + Customer\_Type (Тип пользователя), тип **TEXT**
  + Gender (Пол), тип **TEXT**
  + Payment (Тип платежа)
* Основная таблица Sales с полями (указываются типы данных для каждого поля[[5]](#footnote-4)):
  + Invoice\_ID, тип **TEXT**
  + Branch, тип **TEXT** (Дивизион: зависимое поле от справочника “Branch”)
  + City, тип **TEXT** (Город: зависимое поле от справочника “City”)
  + Customer\_Type, тип **TEXT** (Тип пользователя: зависимое поле от справочника “Customer\_Type”)
  + Gender, тип **TEXT** (Пол: зависимое поле от справочника “Gender”)
  + Product\_Line, тип **TEXT** (Продуктовая группа: зависимое поле от справочника “Product\_Line”)
  + Unit\_Price, тип float
  + Quantity, тип **INTEGER**
  + Total, тип **REAL**
  + Date, тип **TEXT**
  + Time, тип **TEXT**
  + Payment, тип **REAL** (Тип платежа: зависимое поле от справочника “Payment”)
  + Payments\_fee,тип **REAL** (Стоимость осуществления платежей)
  + Gross\_income, тип **REAL** (Маржинальный доход)
  + Cost\_price, тип **REAL** (Себестоимость товара)
  + Unit\_cost, тип **REAL** (Себестоимость единицы товара)

После создания БД и заполнения справочников, заполните основную таблицу Sales данными из файла clear\_data.csv, т. е. данными, полученными после выполнения пунктов 1 и 2 обязательного задания (т. е. в таблице Sales должны оказаться “чистые данные” без аномалий и без пустых значений).

Файл с базой данных должен называться sales.db Для создания и заполнения базы данных используйте SQLite и питон (библиотека [sqlite3](https://docs.python.org/3/library/sqlite3.html)).

##### Исследование данных с помощью библиотеки sqlite3 (SQL-запросы).

Запросы пишутся и выполняются в файле **sqlite\_queries.ipynb**. Обращение к БД (к таблице Sales) нужно производить с помощью библиотеки Python **sqlite3**. В качестве ответа нужно должен быть SQL запрос по соответствующему вопросу и результат выполнения запроса (все в вашем рабочем файле **sqlite\_queries.ipynb**). Комментарии и пояснения писать в ячейках “Текст” вашего файла **sqlite\_queries.ipynb** (требования к оформлению см. в “Предварительных замечаниях”)

1. В каких городах были продажи товара за рассматриваемый период?
2. Какая категория товара лидирует по числу продаж в ед. товара, а какая по суммарным продажам?
3. В какой день были минимальные, а в какой максимальные продажи по суммарной выручке за день?
4. Женщины или мужчины приносят больше выручки?
5. Отличаются ли мужской средний чек от женского? (\*”средний чек” - сумма средней покупки в ритейле)
6. Вывести выручку и маржинальную прибыль в разрезе товарных категорий

#### Создание дашборда

Дашборд или “Панель индикаторов” - это визуализация ключевых показателей, которые необходимы менеджменту для того, чтобы “держать руку на пульсе” бизнеса. Эти показатели нужно регулярно предоставлять руководителям компании для принятия оперативных и стратегических решений.

Вам необходимо разработать дашборд для визуализации ключевых показателей, которые помогут менеджменту компании анализировать продажи. Вы можете сделать это либо с помощью библиотеки **Matplotlib** либо в программе **MS Power BI**

##### **Дашборд с помощью python** M**atpoltlib**

Напишите один скрипт, после запуска которого будут считаны данные из подготовленного вами файла clear\_data.csv с данными или из файла вашей БД SQLite и построены следующие графики и гистограммы:

1. График динамики продаж по дням для каждого дивизиона (3 отдельных графика)
2. Гистограммы суммарных продаж за рассматриваемый период:
   1. по каждому дивизиону
   2. по каждой категории товаров
3. Круговую гистограмму показывающую, доли выручки от продаж мужчинам и женщинам
4. Круговую гистограмму показывающую, доли прибыли по каждой товарной категории
5. Таблицу с выручкой, прибылью и рентабельностью рентабельностью товарных групп по месяцам

Добавьте свои гистограммы, графики и таблицы, которые помогут подкрепить ваши выводы и рекомендации.

##### Дашборд в программе MS power BI Descktop

Разработайте в программе **MS power BI Desktop** дашборд, который будет обращаться за данными в подготовленный вами файл clear\_data.csv или в созданную вами БД SQLite и постройте следующие графики и гистограммы:

1. График динамики продаж с возможностью выбора города расположения магазинов сети.
2. Круговую гистограмму показывающую долю продаж по категориям товаров за рассматриваемый период.
3. График динамики продаж в разрезе категорий товаров с возможностью выбора категории .
4. Таблицу с минимальными, максимальными, средними, медианными продажами в разрезе категорий товара
5. Таблицу с выручкой, прибылью и рентабельностью товарных категорий

Добавьте свои гистограммы, графики и таблицы, которые помогут подкрепить ваши выводы и рекомендации.

#### Поиск зависимости в данных

В результате анализа продаж, молодой аналитик обнаружил, что одна из категорий товара имеет существенно более низкую маржинальную рентабельность, чем остальные. С учетом постоянных издержек, логистических затрат, стоимости услуг платежных систем, возможно, что данная категория приносит компании убытки. Прежде чем вынести на Совет директоров предложение отказаться от этой товарной группы, молодой аналитик, обратился за советом к более опытному коллеге. Коллега посоветовал не торопиться и поискать связь продаж низкорентабельной товарной группы с другими. Так как в случае наличия такой связи, избавление от низкомаржинальной товарной группы может повлиять (привести к падению продаж) на другие товарные группы и рекомендация по избавлению от низкомаржинальной товарной группы может привести к финансовым потерям для компании.

**Задание**: Проведите исследование на возможную взаимозависимость между продажами (выручкой) различных товарных групп. Необходимо суммировать выручку от продаж по дням (отсортированным от наименьшей даты к наибольшей) в разрезе товарных групп и посчитать коэффициент корреляции между полученными числовыми рядами. Для товарных групп, с самым высоким коэффициентом корреляции отдельно посчитать коэффициент корреляции для покупок, осуществляемых женщинами и отдельно мужчинами. Ответить на вопросы:

1. Какие коэффициенты корреляции между продажами различных категорий товара? Постройте таблицу корреляции между товарными группами.
2. Для товарных групп с самой большой корреляцией продаж найдите отдельно коэффициент корреляции между покупками мужчин и покупками женщин.

Сформулируйте выводы, в которых укажите, к каким последствиям может привести отказ от самой низкорентабельной товарной группы. Все скрипты и выводы нужно сделать в отдельном файле dependency\_analysis.ipynb (информацию об исследовании также добавить в “Аналитическую записку” п. 4 Обязательного задания), данные для поиска зависимостей взять из файла clear\_data.csv, в котором должны быть сохранены “чистые данные” после выполнения пунктов 1 и 2 обязательного задания или из БД SQLite (если выполнено задание 5 из “Задания на выбор”)

## Сдача и оценивание проекта

### Формат сдачи материалов

Данные проекта передаются в виде ссылки на папку в облачном хранилище (Яндекс.Диск, Гугл.Диск, Дропбокс, …) с файлами проекта, либо передачей преподавателю zip файла с файлами проекта.

В папке проекта (в zip файле) должны быть:

1. файл **work.ipynb** с кодом Python и необходимыми комментариями к этому коду при выполнении “Основного задания”
2. файл **clear\_data.csv** с данными, с которыми студент должен проделать всю работу после очистки полученного датасета (после выполнения пунктов 1 и 2 основного задания)
3. Аналитическая записка в файле: **Анализ\_продаж.docx** или **Анализ\_продаж.pptx** для пункта 4 “Основного задания”
4. база данных SQLite с названием: **sales.db** и файл **sqlite\_queries.ipynb** со скриптами, если студент выполняет пункт 5 из “Задания на выбор”
5. файл Дашборда **dashboard.pbix** или **dashboard.ipynb**, если студент выполняет задание 6 из “Задания на выбор”
6. Файл **dependency\_analysis.ipynb**, если студент выполняет пункт 7 из задания на выбор

### Шаги сдачи проекта

На этапе «Сдача проекта» вы

1. сообщите преподавателю о завершении первых трех пунктов “Основного задания”. Передаете файлы в соответствии с п. 1 и 2 “Формата сдачи материалов”
2. Информируете и согласовываете с преподавателем, какой пункт из “Дополнительного задания” вы будете выполнять.
3. Получаете фидбек от преподавателя по результатам выполнения п. 1-3 “Обязательного задания”, устраняете недостатки (если такие выявлены), на которые указал преподаватель.
4. Выполняете выбранный пункт из “Дополнительного задания” и сдаете результат преподавателю. Получаете фидбек, исправляете недостатки (если такие выявлены)
5. Приступаете к “Аналитической записке” (4-ый пункт Обязательного задания). Готовите “Аналитическую записку” и сдаете ее преподавателю. Получаете фидбек, вносите изменения, устраняете недостатки.
6. После сдачи (приема преподавателем) “Аналитической записки” (выполнения пункта 4 “Основного задания”) Дипломный проект считается выполненным.

### Оценивание Дипломной работы

Готовый проект проверяется по следующим критериям:

1. На каждый из вопросов должен быть получен ответ в явном виде.
2. Каждый питоновский скрипт можно запустить и получить тот же результат, что у студента.
3. Пояснения и ответы на вопросы аккуратно отформатированы
4. Выводы и предложения сформулированы и оформлены в виде “Аналитической записки”, которую можно предоставить руководству как законченный документ.

Работа **возвращается на доработку (не принимается)**, если:

1. Не выполнен хотя бы один из первых четырех пунктов основного задания
2. Выполнены меньше 5 пунктов из основного и задания на выбор
3. при запуске скриптов результаты не соответствуют тем, которые получены у студента или скрипты не запускаются (выдается ошибка)
4. плохо оформлена “аналитическая записка” (п. 4 основного задания)

# Приложения

## Рекомендации и подсказки

### Python и Pandas

#### Сохранение и запись данных в csv формате

* Основные сведения с примерами: <https://pythonru.com/biblioteki/chtenie-i-zapis-dannyh-cvs-txt-html-xml-pd-7>
* Отдельные методы:
* pd.read\_csv() - загрузка данных из файла с расширением .csv, описание метода по ссылке: <https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html>
* pd.to\_csv() - загрузка данных из файла с расширением .csv, описание метода по ссылке: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_csv.html>

#### Статистический анализ

* Основные сведения с примерами: <https://pythonru.com/primery/primer-analiza-dannyh-na-python-koronavirus>
* Отдельные методы:
  + df.describe() - вывод статистической информации по числовым столбцам датафрейма с названием df, описание метода по ссылке: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>
  + df['Product line'].value\_counts() - подсчёт числа категорий каждого типа по столбцу ‘Product line’, описание метода по ссылке: <https://pandas.pydata.org/docs/reference/api/pandas.Series.value_counts.html>

#### Типы данных

* df['Product line'].astype('category') - изменение типа поля ‘Product line’ на категоиальный тип, описание метода по ссылке: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.astype.html>

#### Выбор данных по значению

#### Группировка (суммирование, подсчет значений, подсчет среднего) по отдельным полям

### Matplotlib

### Sqlite3

## Дополнительные материалы

Pandas:

* ключевые возможности (кратко) <https://habr.com/ru/companies/ruvds/articles/494720/>
* примеры использования Pandas для анализа данных: <https://habr.com/ru/articles/353050/>
* учебник по Pandas <https://pythonru.com/tag/pandas>:

Основные сведения по работе с языком разметки Markdown

<https://www.markdownguide.org/cheat-sheet/>

Поиск корреляция и регрессии в данных с помощью Python

<https://habr.com/ru/articles/557998/>

Поиск аномалий в данных:

<https://habr.com/ru/articles/251225/>

<https://exceltable.com/formuly/raschet-statisticheskih-vybrosov?ysclid=lnvfc6mx9p501993007>

## Рекомендации по выполнению

* Чтобы прийти к результату и выполнить проект целиком, рекомендуем разбить работу над ним на части. Например, первая часть — реализация первой команды и получение результата в самом скрипте. Вторая часть — вывод этого результата по команде бота и так далее. Иными словами, готовая часть — это то, что можно показать заказчику как выполненную работу. В вашем случае роль заказчика исполняет преподаватель.
* Составьте план работы по датам. Исходя из нашего опыта, продуктивнее выделять на дипломный проект по 2-3 часа несколько дней в неделю, чем делать этот же объём за один подход. Рекомендуем придерживаться этого графика и обязательно выделять время для отдыха.
* Отмечайте свой прогресс по мере выполнения плана. Это полезно по нескольким причинам: во-первых, вы будете держать ритм, во-вторых — контролировать ситуацию. И самое главное: каждый выполненный этап — это ваша маленькая победа: чем больше таких побед вы будете замечать, тем большее удовольствие от выполненного проекта получите в итоге.
* Прежде чем отправлять каждую часть проекта на проверку преподавателю (а в реальной работе — тестировщику), всегда уделяйте время проверке за собой (за своим кодом и результатом). Идеально, если между завершением работы и проверкой за собой пройдёт некоторое время — это позволит вам отключиться и в результате посмотреть на работу свежим незамыленным взглядом.
* Качество продукта — это не метрика, а личное отношение разработчика. Попросите нескольких друзей или коллег протестировать проект вместе с вами. Обращайте внимание на поведение, просите вслух проговаривать мысли, действия и ощущения. Это поможет вам понять, где у пользователя могут возникнуть проблемы, и сделать действительно удобный интерфейс. Задавайте наводящие вопросы. Вы можете заранее подготовить для этого чек-лист, отталкиваясь от функциональных требований.

1. Под аномалией в данных подразумеваются значения, которые признак не может принимать. Например, отрицательная стоимость товара или отрицательный рост человека. Или же слишком малые или наоборот слишком большие значения, например, выручка от продаж компании превышает ВВП страны, или сумма одной продажи соизмерима с продажами всей компании за день, неделю, месяц и т. п.. Самый простой способ анализа - посмотреть максимальные и минимальные значения каждого числового признака. Более подробно про анализ аномалий с помощью MS Excel можно почитать [здесь](https://habr.com/ru/articles/251225/) или с помощью Python [здесь](https://habr.com/ru/articles/251225/). [↑](#footnote-ref-0)
2. Категориальный признак – это признак, значения которого обозначают принадлежность объекта к какой-то категории. Примеры таких признаков: Дивизион, Город, Пол, Тип товара, Должность, номер группы студента, тарифный план, издательство, область науки и т.п [↑](#footnote-ref-1)
3. [Маржинальная прибыль] = [Выручка] - [Себестоимость (Cost\_price)] [↑](#footnote-ref-2)
4. [Маржинальная рентабельность] = [Маржинальная прибыль]/[Выручка] [↑](#footnote-ref-3)
5. С типами данных SQLite можно ознакомиться тут: [Datatypes In SQLite](https://www.sqlite.org/datatype3.html) [↑](#footnote-ref-4)