

Конспект по теме "Подготовка презентации"

Как рассказать о своём исследовании

Профессиональный аналитик из чисел и графиков создаёт историю. История объясняет, почему данные выглядят именно так. Умение её сформулировать означает, что аналитик понимает, что происходит с бизнесом сейчас. А ещё способен прогнозировать и давать рекомендации на будущее.

За любой историей должны стоять точные данные и качественная визуализация. Помните: история следует из данных, а не наоборот.

Когда вы делаете выводы на основе данных, не ограничивайтесь сухими цифрами: изучите происходящее вокруг.

Источники сюжета истории:

- общая информации о компании и бизнес-процессах.
- финансовые, аналитические и другие метрики.
- внутренние и внешние факторы, которые влияют на компанию.

Кому рассказывать истории

Важно понимать, кто будет изучать результаты вашего исследования. Например, если это:

- аналитики
Подготовьте подробный отчёт с кодом, комментариями к нему и ссылкой на данные. Лучше собрать его в *Jupyter Notebook*. Уточните, что именно хотят увидеть коллеги: методику расчёта или результаты с объяснениями. Это поможет скорректировать ваш отчёт.
- менеджеры продукта
Подготовьте подробный отчёт в формате презентации: *pptx* или *pdf*. Отрадите основные результаты, выводы по ним и ваши рекомендации.

- руководители высшего звена
Отчёт должен содержать ответы на конкретные вопросы руководителя.

Чтобы определиться, какую информацию добавить, а какую — отсечь, руководствуйтесь правилом: *чем дальше от анализа данных пользователь презентации, тем проще она должна быть.*

Сезонность и внешние факторы

Когда соотнесли внешние события с данными, время делать выводы.

Популярность того или иного товара или услуги может иметь сезонный характер. В аналитике сезонность проверяют на исторических данных. Если вы наблюдаете резкий всплеск в данных — посмотрите, было ли что-то похожее ранее? Изучите историю запросов, например на сайте Яндекс.Вордстат — <https://wordstat.yandex.ru/>. Это поможет выдвинуть гипотезы о наличии сезонности, которые нужно проверить.

Чтобы отследить сезонность товара, узнайте, есть ли значимые отличия в пользовательском поведении в разные временные промежутки? Когда найдёте изменения в одном из месяцев, которые так и тянет назвать сезонностью, остановитесь. Сформулируйте гипотезу и проверьте статистическую значимость различий.

Окружающая среда и её изменения — ещё один важный фактор для выводов.

Абсолютные и относительные величины

Обычно люди склонны оперировать абсолютными величинами — объёмом, величиной, размером наблюдаемого явления или события. Однако руководствоваться в работе только ими — нельзя.

Относительная величина — показатель, который даёт числовую меру соотношения двух сопоставимых величин. Относительные величины помогут правильно объяснить изменения абсолютных величин.

Парадокс Симпсона

Иногда исследователь всё делает правильно, а результаты получает **контринтуитивные** —выводы, которые на первый взгляд не соответствуют здравому смыслу. Такие явления называются **парадоксами**.

Пример контринтуитивного явления — **парадокс Симпсона** (**парадокс Юла-Симпсона**, или **парадокс объединения**). Например, этот эффект наблюдали при приёме абитуриентов в университет Беркли в 1973 году. Отношение числа принятых на учёбу к числу поданных заявлений показало, что у мужчин шансов поступить в Беркли больше, чем у женщин. Если изучить статистику по каждому факультету, результаты будут неожиданными. Вероятность поступить в университет одинакова для мужчин и женщин.

При объединении двух групп с одинаковой зависимостью зависимость или исчезает, или становится обратной. Чаще всего такое происходит, когда не учитываются скрытые данные или контрольные группы не репрезентативны.

Чтобы найти причины явления, нужно:

- подвергнуть сомнению полученные данные, сравнить их с действительностью;
- посмотреть на данные в других разрезах или выяснить, есть ли скрытые данные, влияющие на результат;
- расследовать до тех пор, пока не будет найден ответ о реальной зависимости в данных.

Когда график — ваш враг

Графики бывают неправильными из-за ошибок в расчётах, либо неверно выбранного вида визуализации.

- следите за точностью расчётов и исходных данных;
- всегда проверяйте реалистичность визуализации;
- правильно выбирайте шкалы для осей;

- сделайте визуализацию в разных разрезах данных;
- грамотно подбирайте цвет визуализации.

Для графиков работает закон: «чем меньше, тем лучше». Вы всегда можете разнести информацию на несколько визуализаций и представить её в разных разрезах.

Виды графиков для различной информации

Практические советы по выбору графиков.

Сравнение во времени

Сколько временных периодов нужно отразить?

- Если много, рисуйте линейный график
- Если несколько (например, нескольких месяцев), нагляднее столбчатая гистограмма

Распределение набора данных

Сколько переменных нужно показать?

- Одну. Подойдёт столбчатая гистограмма. Также можно добавить наложение линейного графика, который отображает, например, нормальное распределение.
- Две. Диаграмма рассеяния. Она также подойдёт, чтобы показать взаимосвязь данных:

Распределение данных по категориям

Если важно наблюдать изменения во времени — стройте столбчатую гистограмму. Если нет, то круговую диаграмму, но будьте внимательны с отображением долей.

Воронки и другие специальные графики

Эти графики не классические, но их часто нужно строить. К специальным графикам относится, например, тепловая карта.

Убедитесь, что:

- **цвет**
 - подобран согласно цветовой гамме;
 - не мешает восприятию в случае чёрно-белой печати;
 - не вызовет проблем у людей с нарушениями цветового зрения;
 - контрастирует с фоном.
- **текст**
 - читаемый;
 - расположен горизонтально;
 - данные на графике подписаны;
 - есть легенда.
- **линии**
 - линии сетки (если есть) скрыты;
 - у графика нет рамки;
 - на осях нет ненужных отметок;
 - на графике одна горизонтальная и одна вертикальная ось.
- **общий вид**
 - пропорции соблюдены;
 - расстояния между осями равноудаленные;
 - график двухмерный;
 - минимум украшательств.

Вот памятка по методам, которые помогут построить информативный график в *matplotlib*.

1. Заголовок графика добавляют методом **set_title()**:

```
import matplotlib.pyplot as plt
fig, test = plt.subplots()

test.set_title('Нормальное распределение')
```

2. Подписи к осям — методами **set_xlabel()** и **set_ylabel()**:

```
import matplotlib.pyplot as plt
fig, test = plt.subplots()

test.set_xlabel('Подпись для оси X')
test.set_ylabel('Подпись для оси Y')
```

3. Легенду — методом **legend()**:

```
import matplotlib.pyplot as plt
fig, test = plt.subplots()

test.plot(x, y1, label = 'Подпись 1')
test.plot(x, y2, label = 'Подпись 2')
test.plot(x, y3, label = 'Подпись 3')

test.legend()
```

4. Изменяют размеры графика методами **set_figheight()** и **set_figwidth()**:

```
import matplotlib.pyplot as plt
fig, test = plt.subplots()

fig.set_figheight(10) # высота
fig.set_figwidth(8) # ширина
```

Основные принципы построения презентации

Когда вы провели исследование, сделали выводы и построили графики — можно собирать финальную презентацию.

Инструменты

Пользователи *Windows* создают презентации в *Power Point*. Это платное программное обеспечение из пакета программ *Microsoft Office*.

Альтернативой может стать *Power Point Online* или *Libre Office*. У владельцев *MacOS* есть *Key Note*.

Презентацию делают в двух основных форматах: **pptx** (формат Power Point) и **pdf**.

Если над ней работают в команде или она нужна для внутреннего пользования, сохраните презентацию как *pptx*. Если вы готовитесь к выступлению, сохраните файл в формате *pdf*. Тогда стили, шрифты и картинки отобразятся как задумано. Ещё *pdf* можно открыть почти с любого устройства, а с *pptx* такое получается не всегда.

Структура презентации

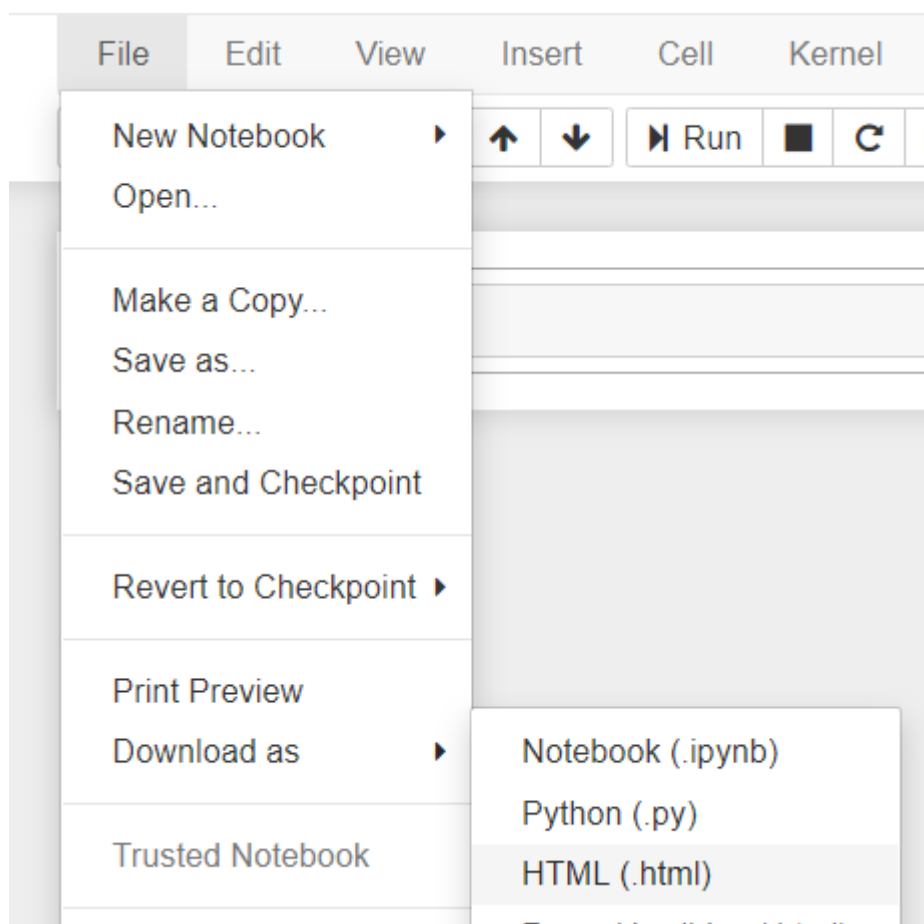
1. На вводном слайде укажите предмет и период исследования.
2. Общие выводы — в начало.
3. Один слайд — одна мысль. Всё, что непонятно, смело убирайте.
4. В заголовке слайда — ключевой вывод. Укажите в заголовке вывод, который укажет, на что нужно обратить внимание и ответ на какой вопрос искать на слайде.
5. Укажите выводы исследования.
6. Поддерживайте единую историю. Решайте задачи исследования постепенно. Завершайте каждый блок слайдов так, чтобы пользователи презентации делали нужные выводы.
7. Убедитесь, что данные корректны. Проверьте, что графики не противоречат друг другу.
8. Указывайте источники данных.
9. В конце презентации — рекомендации. Не обязательно подробные, но отражающие суть.

В компании скорее всего есть **гайдлайн** (шаблон) по оформлению презентаций, где указаны все нюансы и требования. Уточните это у своих коллег.

Отчёт в Jupyter Notebook

Вы уже разобрали случаи, когда вашим коллегам пригодится отчёт в *Jupyter Notebook*. Есть два формата, в которых можно его сохранить:

- **ipynb** — стандартный формат сохранения файла в *Jupyter Notebook*. Его часто называют «тетрадка» или «ноутбук». Тогда у вашего коллеги под рукой должен быть компьютер, чтобы изучить отчёт.
- **HTML** подойдёт, когда *Jupyter Notebook* не открыть (например, с телефона). Нужная разметка генерируется автоматически. Сохранить тетрадку в нужном формате просто: `File -> Download as... -> HTML`



Вот чек-лист подготовки отчёта в *Jupyter Notebook*:

1. Если вы отправляете ipynb-файл, обязательно приложите исходные данные.
2. Если вы отправляете HTML-файл, убедитесь, что ячейки выполнены корректно: появились все таблицы и графики.

3. В ячейке типа *markdown* обязательно обозначьте, какая задача решается в этом файле.
4. Разметьте основные этапы. Обозначьте, где идёт предобработка данных, где ставятся вопросы, а где сформулированы основные выводы.
5. В начале файла сделайте оглавление.
6. Обозначьте основные результаты и выводы. Это поможет ответить на вопросы быстро.
7. Оставьте комментарии в коде. Укажите, что значит переменная и выполняет функция.

Разберём, как оформить оглавление в тетрадке. Оно поможет организовать навигацию по документу и покажет ход исследования.

Выберите тип ячейки *markdown* и запишите там все основные пункты вашего исследования так:

```
1. [Открытие данных]
2. [Предобработка данных]
  * [Обработка пропущенных значений]
  * [Обработка дубликатов]
```

В каждую *markdown*-ячейку с названием соответствующего пункта добавьте ссылку:

```
<a id="some_id"></a>
```

Укажите уникальный *id*, по которому ячейку можно будет найти. К созданному списку добавьте ссылки:

```
1. [Открытие данных](#start)
2. [Предобработка данных](#preprocessing)
  * [Обработка пропущенных значений](#null)
  * [Обработка дубликатов](#duplicates)
```

Нажмите **Shift + Enter** и вы увидите список со ссылками.

Хорошая практика — не оставлять путь к файлу на вашем компьютере. Чтобы избежать недоразумений, создайте переменную `path`, в которой сохраните уникальный путь. Затем добавьте имя файла:

```
import pandas as pd

path = # путь к папке на локальном компьютере
df = pd.read_csv(path + 'auto.csv')
```