



НЕТОЛОГИЯ
групп

Компетенции и инструменты аналитика. Роль аналитики в принятии решений



Максим Чикуров

Data Scientist и руководитель команды
аналитики

Работал в компаниях Citibank, BNP Paribas,
Barclays Bank, Teradata



maxim.chikurov@gmail.com

Что такое анализ данных?

Процесс изучения, очистки, трансформирования и моделирования данных с целью извлечение знаний, формулирования выводов и поддержки процесса принятия решений.

- Впервые использовано в 1961 году статистиком Джоном Тьюки
- Используется в Business Intelligence, Data Science и Data Mining
- Важнейшая часть культуры принятия решений

ГДЕ?

Data Driven подход в принятии решений

Что такое
управленческие
решения

Стили
управленческих
решений

Зачем нужно
принимать
управленческие
решения?

Data Driven vs
Data Informed

Data



Reporting



Analysis



Action



Value



Data Driven vs Data Informed

Численные данные очень мощественны, так как очень конкретны.

Для предпринимателей и руководителей с техническим образованием, эмпирические данные важнее большинства других факторов (опыт, советы консультантов) при принятии решений.

При подходе **data-driven** вы откидываете все кроме численных данных. Ответы на вопросы «Почему?», «Кто?» и «Что дальше?» основаны на результатах статистического анализа, проверки гипотез, интерпретации результатов измеряемых экспериментов.

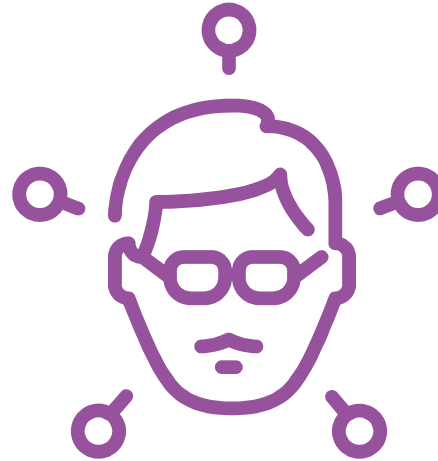
При подходе **data-informed** имеющиеся у вас данные являются лишь одним из факторов для принятия решений.

КТО?

Кто такой Аналитик данных

Аналитик данных переводит «язык чисел» на «человеческий язык».

Делает полезными
огромное количество
данных



«Вертикальный мост»
передачи информации
для принятия решений

Задачами могут быть
отчетность, анализ данных, а
также построение
технических решений

Зачастую для выполнения
работы требуется
специфические технические
знания

Компетенции аналитика

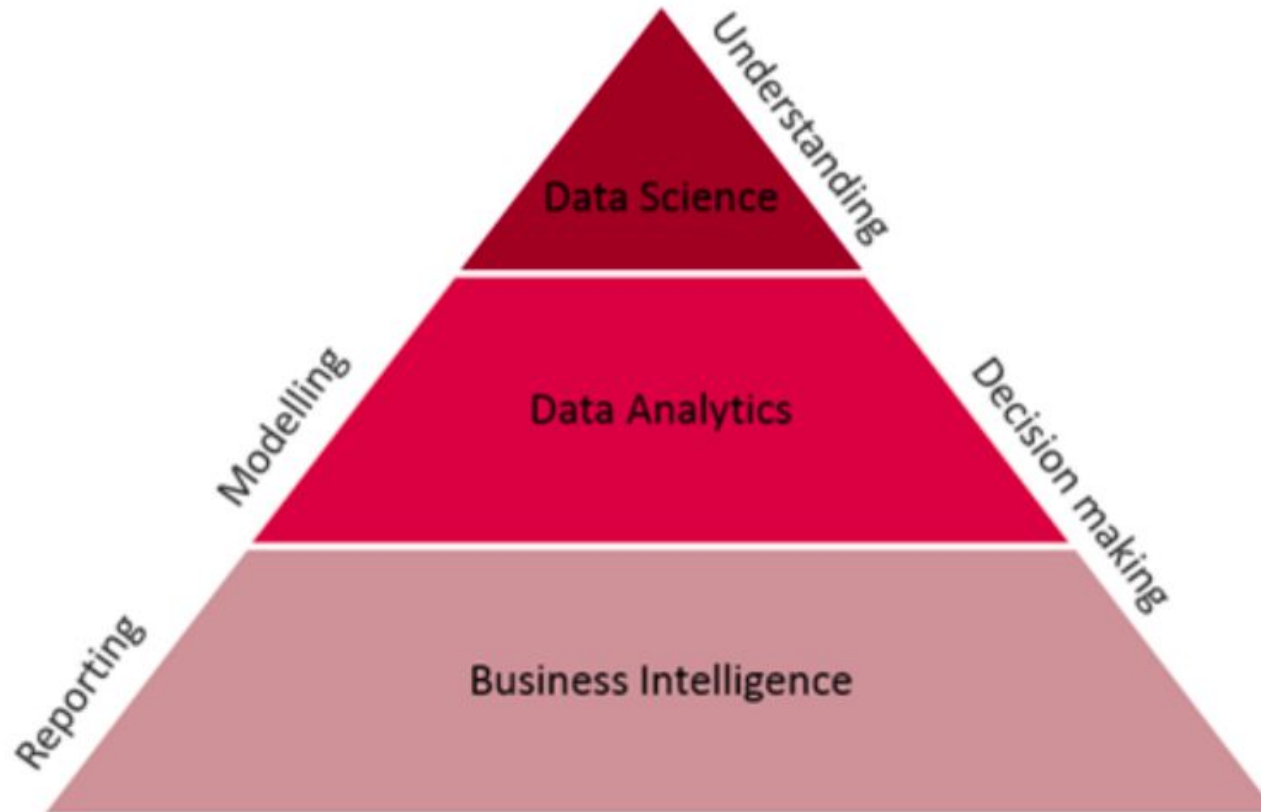
СОФТ-СКИЛЛС

- Знать ключевые математические методы и основы статистики
- Мыслить абстрактно
- Скетчить и прототипировать
- Уметь видеть за метриками физический смысл и находить причины и взаимосвязи событий
- Обладать развитым эмоциональным интеллектом
- Быть способным дать рекомендации бизнесу

ХАРД-СКИЛЛС

- Проводить сбор и анализ требований заказчиков к отчетности
- Получать, очищать и преобразовывать, визуализировать и представлять данные
- Интерпретировать данные и делать обоснованные выводы на их основе
- Разрабатывать требования к аналитическим инструментам и сопровождать их внедрение
- Проводить исследования и А/Б тесты. Готовить аналитические материалы для выработки стратегических решений

Компетенции аналитика



Инструменты аналитика

- Электронные таблицы
- Инструменты доступа к данным и хранилища данных
- Инструменты визуализации данных и построения отчетов
- OLAP-инструменты
- Инструменты дата-майнинга (анализа данных): SAS, Python, R, Weka, Knime, RapidMiner и т.д.

Microsoft лидер рынка?



КЕЙСЫ

Бизнес-кейс

Компания занимается автомобильными грузоперевозками между городами России и Стран Прибалтики.

Заказы составные: в заказ включается некоторое количество различных коробок имеющих разный вес.

Один заказ доставляется одним автомобилем.

Имеем данные о заказах за месяц.

Основные характеристики заказа:

- день
- количество коробок
- общий вес
- пункт отправки
- пункт доставки
- вес каждой коробки включенной в заказ

Разные меры имеют различное распределение, кое-где есть аномалии. На этих данных мы проиллюстрируем все статистические понятия, включенные в занятие.

Набор данных

- Наборы данных и таблицы фактов
- Дискретные и непрерывные величины
- Измерения и меры
- Гранулярность

[ссылка на датасет](#)

НАБОР ДАННЫХ

	A	B	C	D	E	F
1	Номер заказа	Вес груза	День	Количество коробок	Город отправки	Город доставки
2	22	1 122,21	1	157	Таллин	Санкт-Петербург
3	37	1 110,50	1	148	Рига	Санкт-Петербург
4	64	1 165,27	1	153	Москва	Рига
5	95	1 157,63	1	153	Таллин	Санкт-Петербург
6	105	1 000,52	1	145	Вильнюс	Санкт-Петербург
7	184	1 110,26	1	154	Рига	Санкт-Петербург
8	194	1 030,81	1	40	Рига	Санкт-Петербург
9	277	1 093,69	1	139	Хельсинки	Санкт-Петербург
10	473	1 025,65	1	129	Таллин	Санкт-Петербург
11	586	1 001,96	1	140	Хельсинки	Москва
12	592	1 297,76	1	162	Рига	Санкт-Петербург
13	660	1 110,30	1	149	Рига	Санкт-Петербург
14	669	1 180,50	1	154	Хельсинки	Калининград
15	675	1 005,16	1	142	Таллин	Санкт-Петербург
16	680	975,61	1	127	Калининград	Вильнюс

АГРЕГАЦИЯ

День	Город отправки	Город доставки	Кол-во поездок	Суммарное кол-во коробок	Общий вес груза
+ Всего(1)			19	2 791	20 877,61
+ Всего(2)			53	7 993	60 484,58
+ Всего(3)			46	6 980	52 247,32
+ Всего(4)			42	6 280	47 102,55
+ Всего(5)			52	7 756	57 398,69
+ Всего(6)			42	6 314	47 179,19
+ Всего(7)			23	5 641	41 938,29
+ Всего(8)			27	5 580	41 985,78
+ Всего(9)			47	7 097	53 460,71
+ Всего(10)			56	8 520	63 866,80
+ Всего(11)			40	6 021	45 119,20
+ Всего(12)			42	6 308	47 767,00
+ Всего(13)			46	7 037	52 372,08



Основные понятия статистики

Основные понятия статистики. Примеры

- Минимум/Максимум
- Среднее
- Медиана
- Мода
- Дисперсия
- Стандартное отклонение

Основные понятия статистики. Определения

Вес груза основные характеристики	
Минимум	823,52
Максимум	1 375,52
Среднее	1 132,86
Дисперсия	8 433,80
Стандартное отклонение	91,84
Медиана	1 136,45
Мода	1 157,63

Квартили —

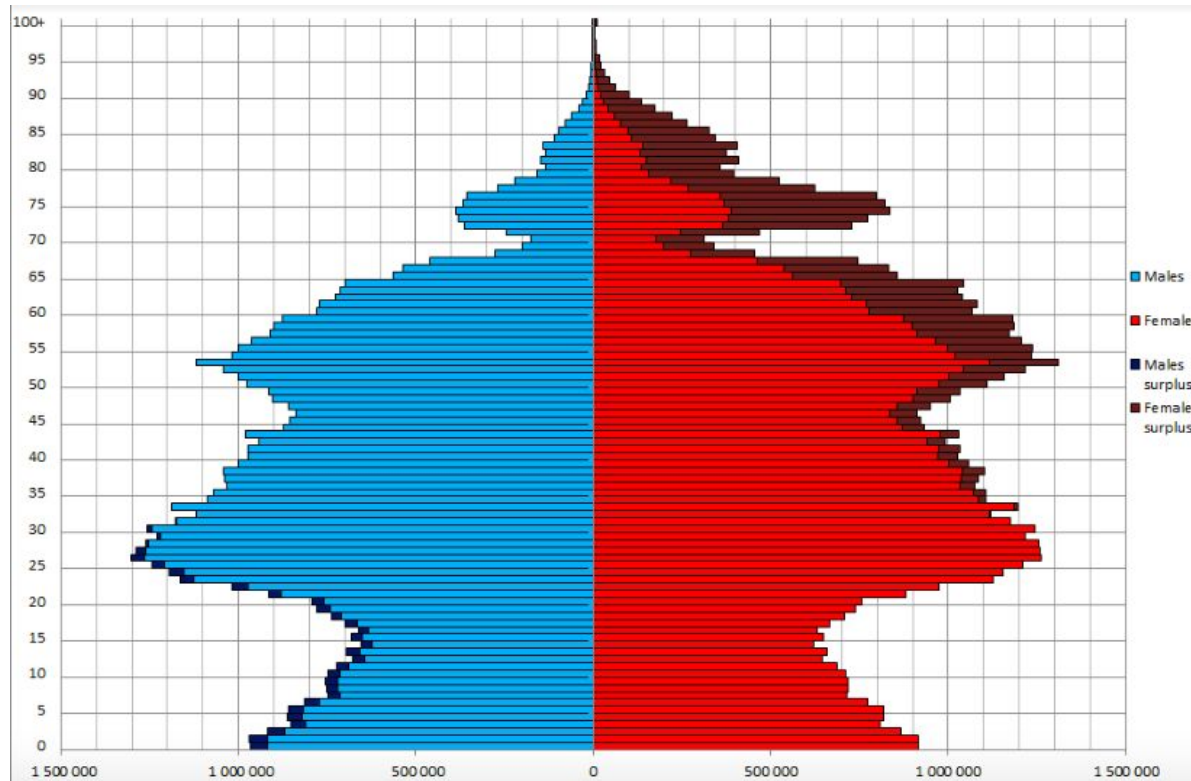


значения, которые делят таблицу данных или ее часть на четыре группы, содержащие приблизительно равное количество наблюдений.

Общий объем делится на четыре равные части: 25%, 50%, 75% 100%.

Гистограмма и распределение

Понятие гистограммы и распределения на примере возрастно-гендерной пирамиды

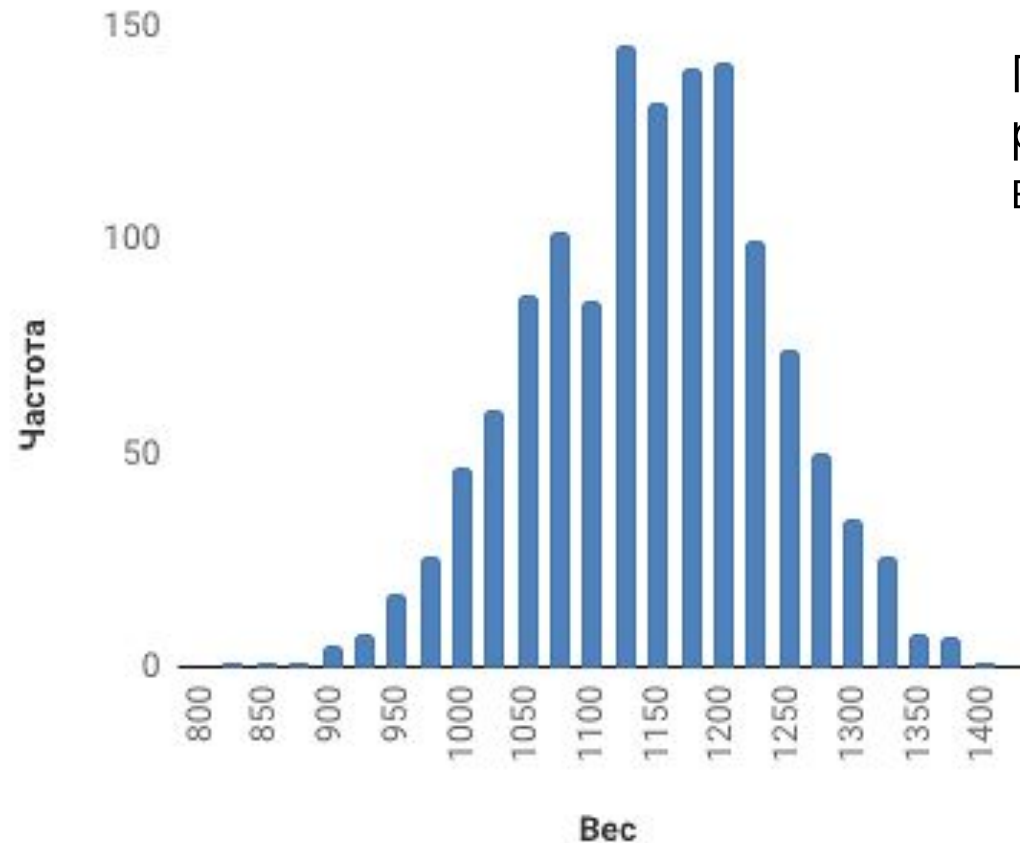


Картинка
возрастно-
гендерной
пирамиды
за 2014 год

Примеры

Гистограмма позволяет понять:

- Существуют ли пробелы в данных.
- Симметричны ли данные.
- Насколько плотно сгруппированы данные.
- Смещены ли данные и, если да, то в каком направлении.

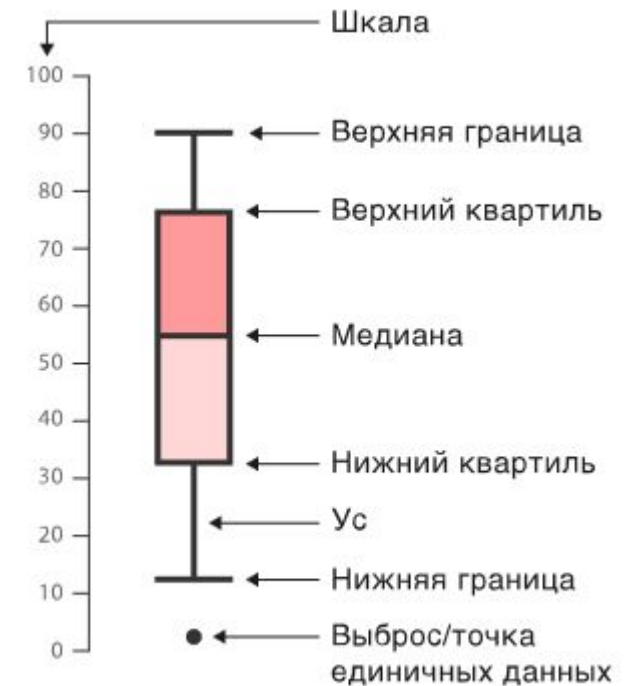


Гистограмма
распределения
веса груза

Box Plot и «выбросы»

График, использующийся в описательной статистике, компактно изображающий **одномерное** распределение:

- в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы
- быстрый способ изучения одного или нескольких наборов данных в графическом виде.
- **Выброс** (Outlier) – значение, выделяющееся из общей выборки.



Процентили

Процентиль, %	Общий вес груза
0	823,52
5	977,15
10	1 011,21
15	1 033,69
20	1 053,08
25	1 067,89
30	1 085,20
35	1 103,10
40	1 112,71
45	1 124,93

Процентиль, %	Общий вес груза
50	1 136,45
55	1 149,24
60	1 159,61
65	1 172,70
70	1 183,03
75	1 195,81
80	1 211,74
85	1 226,46
90	1 246,62
95	1 282,64
100	1 375,52



Практика

Бизнес задача

Руководство компании задумалось о выборе новой модель автомобиля для перевозок. Владелец компании считает, что нас ограничивает грузоподъемность имеющегося парка автомобилей. Директор по развитию хочет приобрести часть парка автомобилей разорившегося конкурента.

Необходимо ответить на вопросы:

- Будем ли мы перевозить больше заказов, если купим автомобили большей грузоподъемности?
- Подойдут ли нам автомобили грузоподъемностью 1,2 тонны для увеличения мощности (количества) перевезенных заказов?



СПАСИБО ЗА ВНИМАНИЕ