СОВРЕМЕННЫЕ МЕТОДЫ СБОРА И АНАЛИЗА ДАННЫХ

Организационный раздел

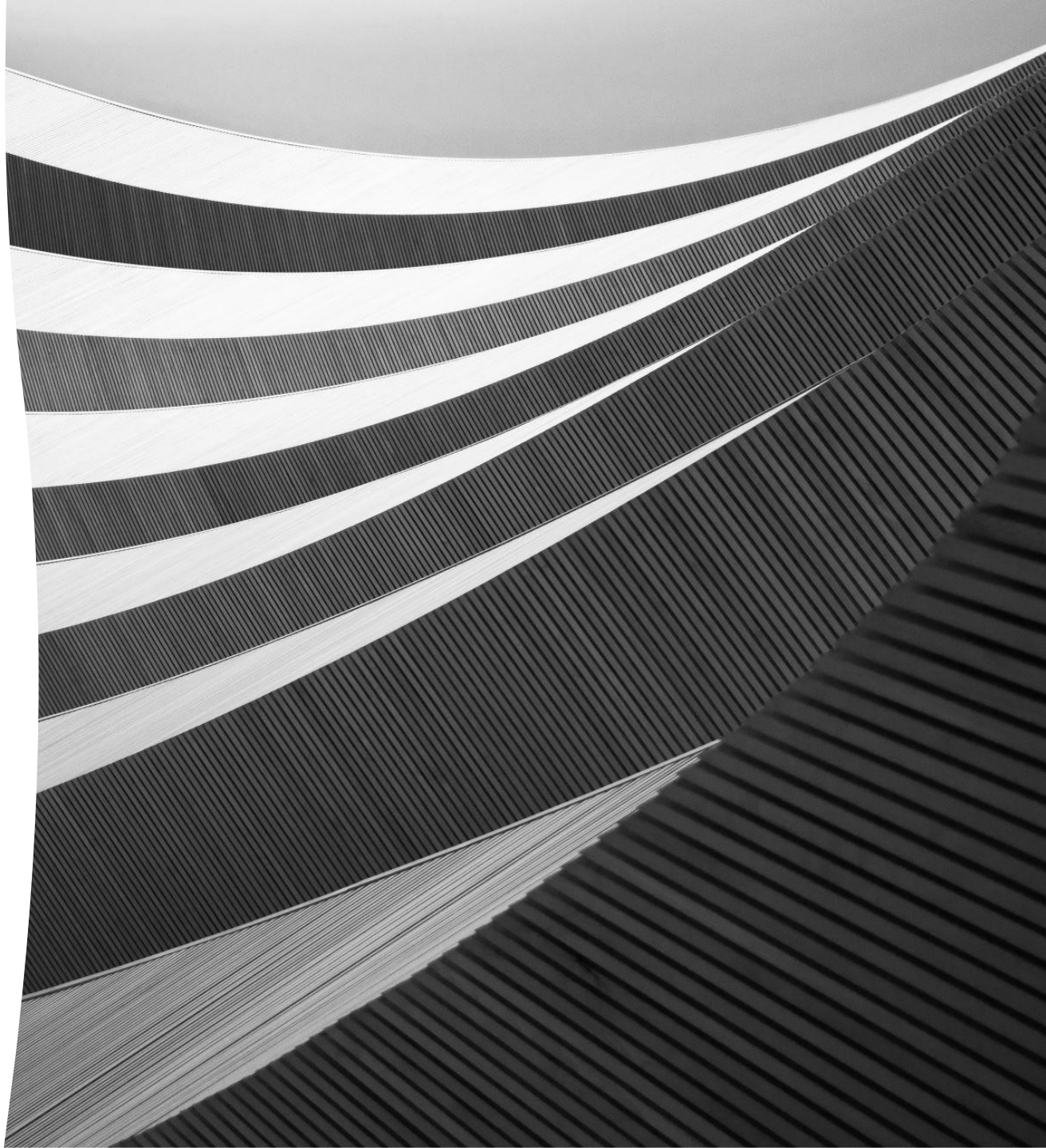
Форма контроля: зачет

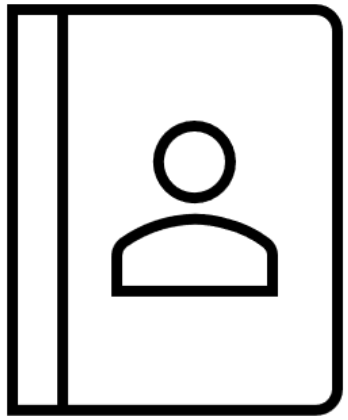
Общее количество часов: 34. Количество занятий: 12

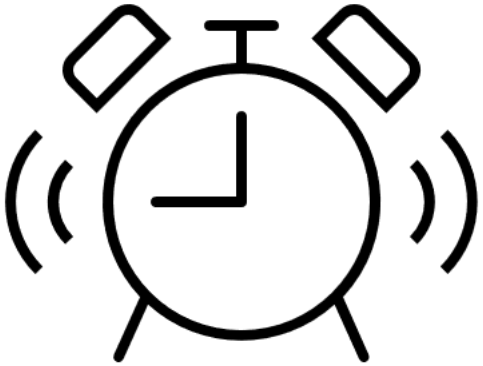
Итоговое занятие: защита проделанной работы (доклад с презентацией).

Формат представления практических работ: визуализация (видео-файл)

Цифровая площадка для прикрепления выполненных работ: группа в ВК https://vk.com/club207632290







ЧЕК-ЛИСТ НА РАБОТЫ СОКУРСНИКОВ (ОТ 0 ДО 5 БАЛЛОВ)

1. Содержание
2. Визуализация
3. Выступление

Основные термины курса

- В широком смысле *данные* – факты, текст, графики, картинки, звуки, аналоговые или цифровые видео-сегменты и т.д.
- Они могут быть **получены в результате:** измерений, экспериментов, арифметических и логических операций и т. д.
- **Данные должны быть представлены в форме, пригодной для хранения, передачи и обработки.**
- ***Данные*** – это необработанный материал, предоставляемый поставщиками *данных* и используемый потребителями для формирования **информации** на основе *данных*.

Информация ?

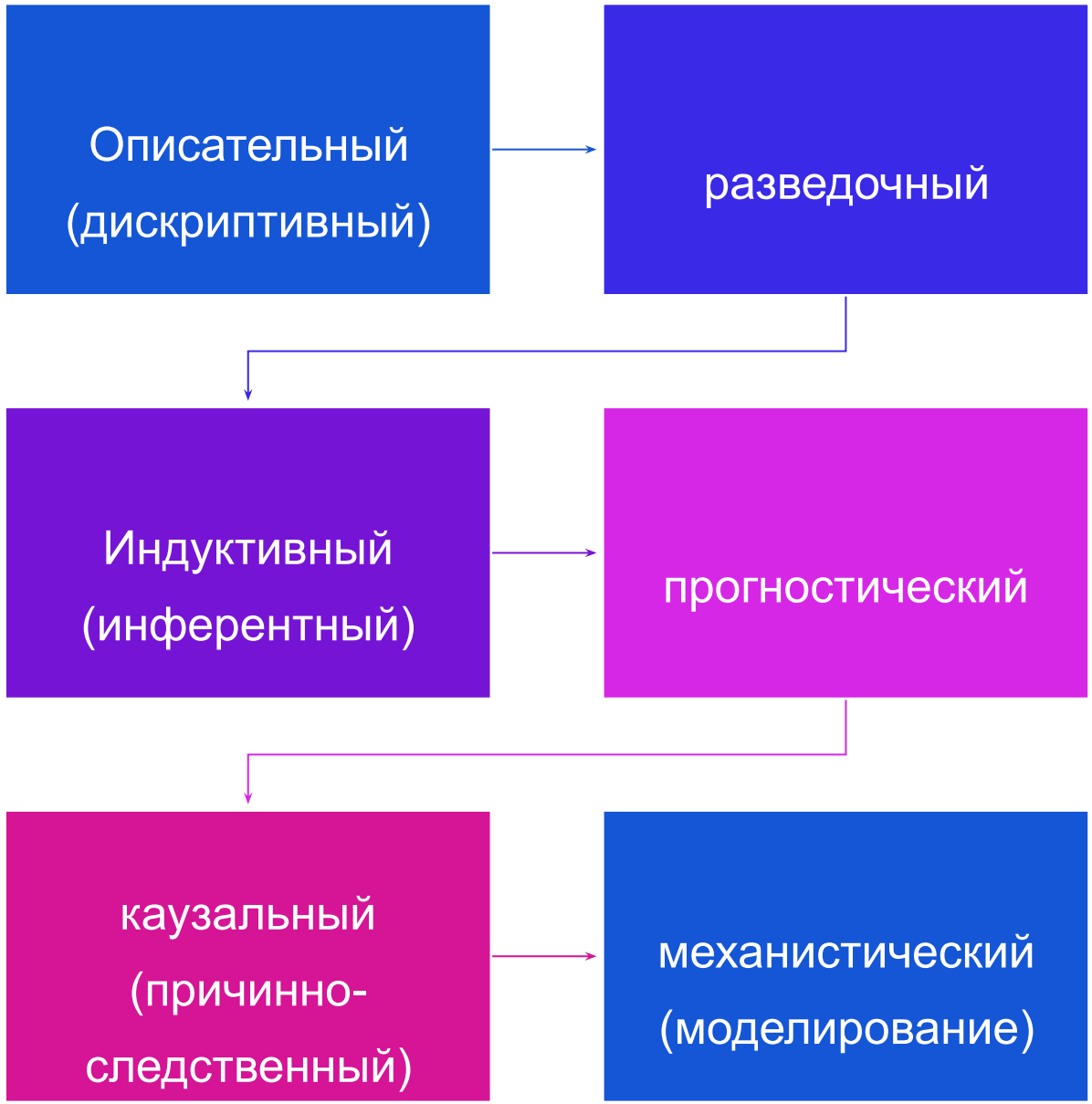
Основные термины курса

- **Объект** – сущность, о которой данных хранятся в какой-либо форме
- **Атрибут** – измеримое свойство объекта, его характеристика.
- **Анализ** — преобразование данных в выводы, на основе которых будут приниматься решения и строиться действия с помощью людей, процессов и технологий

- При анализе *данных*, как правило, нет возможности рассмотреть всю совокупность *объектов*.
- Изучение очень больших объемов *данных* является дорогостоящим процессом, требующим больших временных затрат, а также неизбежно приводит к ошибкам, связанным с человеческим фактором.
- Вполне достаточно рассмотреть некоторую часть всей совокупности, то есть *выборку*, и получить интересующую информацию на ее основании.

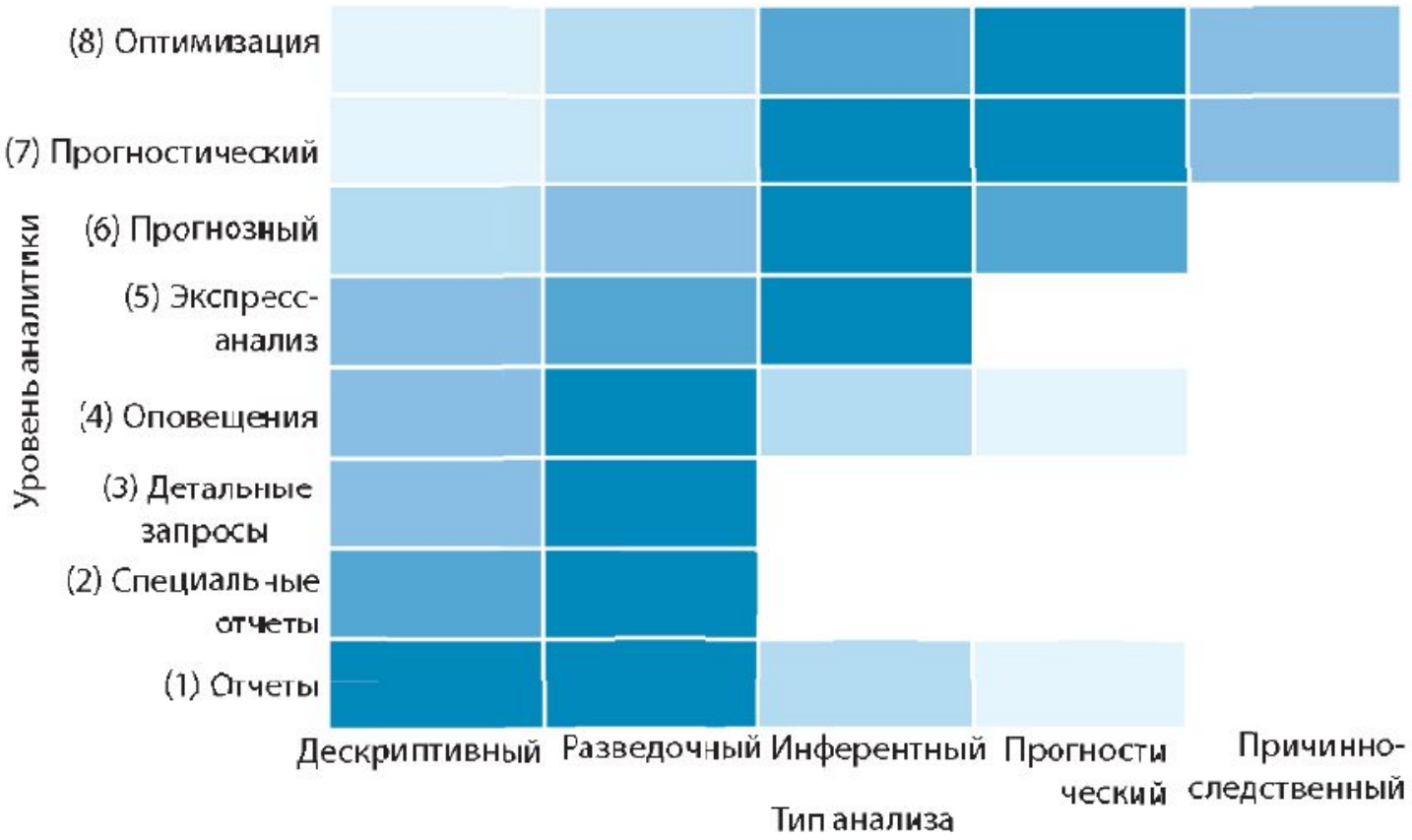
Виды анализа

(от простого к сложному)



Привести пример каждого вида анализа ?

Уровень аналитики и вид анализа



Интенсивность цвета каждой ячейки обозначает примерную оценку усилий или времени, затраченных на проведение этого типа анализа.

Например, подготовка стандартных отчетов обычно осуществляется на основе описательного и разведочного типов анализа, при этом крайне маловероятно использование причинно-следственных моделей.

Аналитика оптимизации строится на описательном и разведочном анализе, но в первую очередь сосредоточена на прогностическом и, возможно, причинно-следственном анализе.

Измерения

- **Измерение** – процесс присвоения чисел характеристикам изучаемых *объектов* согласно определенному правилу.
- В процессе подготовки *данных* измеряется не сам *объект*, а его характеристики.
- ***Шкала*** – правило, в соответствии с которым объектам присваиваются числа.
- **Переменная** – свойство или характеристика, общая для всех изучаемых *объектов*, проявление которой может изменяться от *объекта* к *объекту*.
- **Значение переменной** является проявлением признака объекта.

Измерения

- Переменные могут являться **числовыми** данными либо **символьными.**
- Числовые данные, в свою *очередь*, могут быть **дискретными и непрерывными.**
- **Дискретные данные** являются значениями признака, общее число которых конечно либо бесконечно, но может быть подсчитано при помощи натуральных чисел от одного до бесконечности.
- Пример дискретных *данных*. Продолжительность маршрута автобуса (количество вариантов продолжительности конечно): 15, 20, 30 мин.
- **Непрерывные данные** – данные, значения которых могут принимать какое угодно *значение* в некотором интервале.

Измерение непрерывных *данных* предполагает большую *точность*.

- Пример непрерывных *данных*: температура, *высота*, *вес*, *длина* и т.д.

Шкалы

Существует пять типов шкал измерений:

- номинальная
- порядковая
- интервальная
- относительная
- дихотомическая

Привести пример каждого типа шкал измерений?

Задания

Типы наборов данных

Форматы хранения данных

Классификация видов данных

Индивидуальная работа (визуализация, доклад)

- 06.11 Данные в цифровой экономике, их визуализация и предварительная обработка
- 13.11. Технологии анализа данных. Инструменты описательной статистики
- 20.11. Агрегирование данных. Методы одномерного и двумерного анализа данных
- 27.11. Методы многомерного анализа данных. Методы анализа временных рядов.
- 04.12 Программные средства анализа данных
- 11.12. Гипотезы. Проверка статистических гипотез.
- 18.12 Дисперсионный анализ. Основы непараметрической статистики.
- 25.12. Итоговая комплексная работа

Самостоятельная работа

- На портале intuit.ru записаться на бесплатное изучение курса **Data Mining - https://intuit.ru/studies/courses/6/6/info**
- Пройти итоговое тестирование
- Сделать визуализацию материала (2-3 кадра/слайда на каждую тему лекции)
- Завершающие слайды – оценка по итоговому тесту и бесплатный сертификат