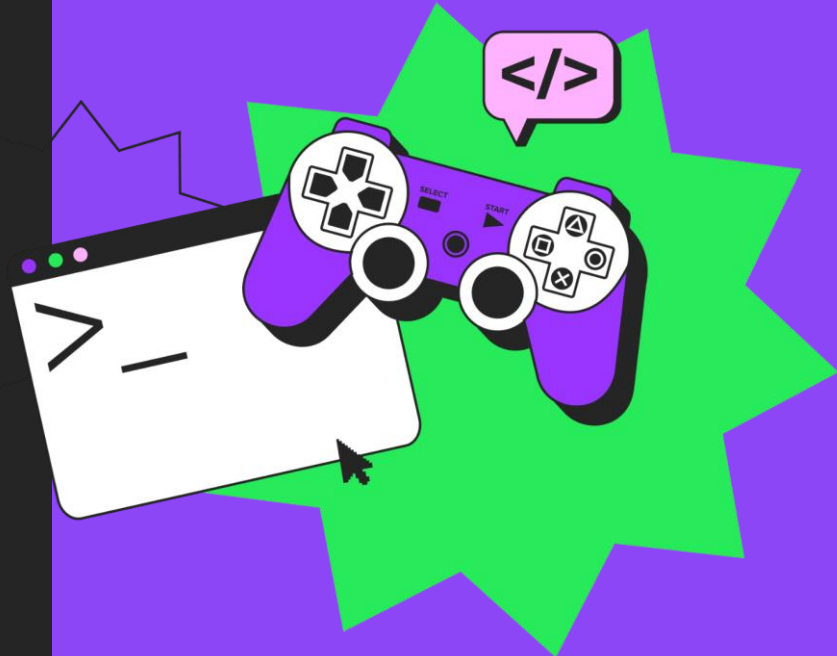


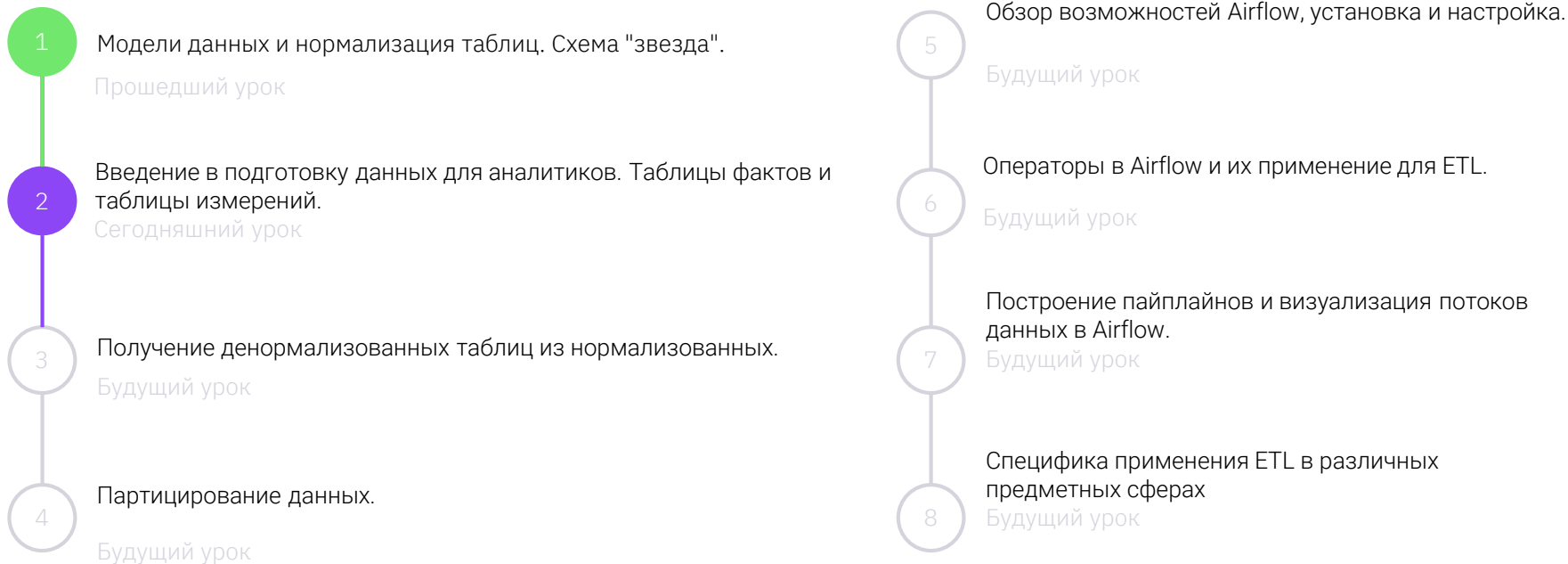
Введение в подготовку данных для аналитиков. Таблицы фактов и таблицы измерений.

Урок 2













План курса (вертикальный)





Что будет на уроке сегодня

-  Тенденции в бизнес аналитике
-  Понимание бизнеса и данных
-  Подготовка и преобразование данных
-  Исследование и визуализация данных
-  Моделирование
-  Оценка
-  Развертывание
-  Таблицы фактов и измерений



Викторина



Что такое BI?

1. Ключевые показатели эффективности
2. Бизнес аналитика
3. Индекс оценки бизнеса



Что такое BI?

1. Ключевые показатели эффективности
2. Бизнес аналитика
3. Индекс оценки бизнеса



Для чего нужна бизнес-аналитика?

1. Выявлять рыночные тенденции и повышать эффективность бизнеса
2. Установить критерии процессов внутри компании
3. Оба варианта верны



Для чего нужна бизнес-аналитика?

1. Выявлять рыночные тенденции и повышать эффективность бизнеса
2. Установить критерии процессов внутри компании
3. Оба варианта верны



Что входит в понятие анализ данных?

1. Извлечение, трансформация, загрузка
2. Извлечение, подготовка, моделирование



Что входит в понятие анализ данных?

1. Извлечение, трансформация, загрузка
2. Извлечение, подготовка, моделирование



Что такое сглаживание данных?

1. Процесс удаления избыточности
2. Процесс удаления шума из данных
3. Приведение данных к заданному диапазону
4. Все варианты верны



Что такое сглаживание данных?

1. Процесс удаления избыточности
2. Процесс удаления шума из данных
3. Приведение данных к заданному диапазону
4. Все варианты верны



Что такое нормализация данных?

1. Процесс удаления избыточности
2. Процесс удаления шума из данных
3. Приведение данных к заданному диапазону
4. Все варианты верны



Что такое сглаживание данных?

1. Процесс удаления избыточности
2. Процесс удаления шума из данных
3. Приведение данных к заданному диапазону
4. Все варианты верны



В какой таблице хранятся редко изменяемые данные?

1. Таблица фактов
2. Таблица измерений
3. В обеих



В какой таблице хранятся редко изменяемые данные?

1. Таблица фактов
2. Таблица измерений
3. В обеих



Вопросы?

Вопросы?



Вопросы?





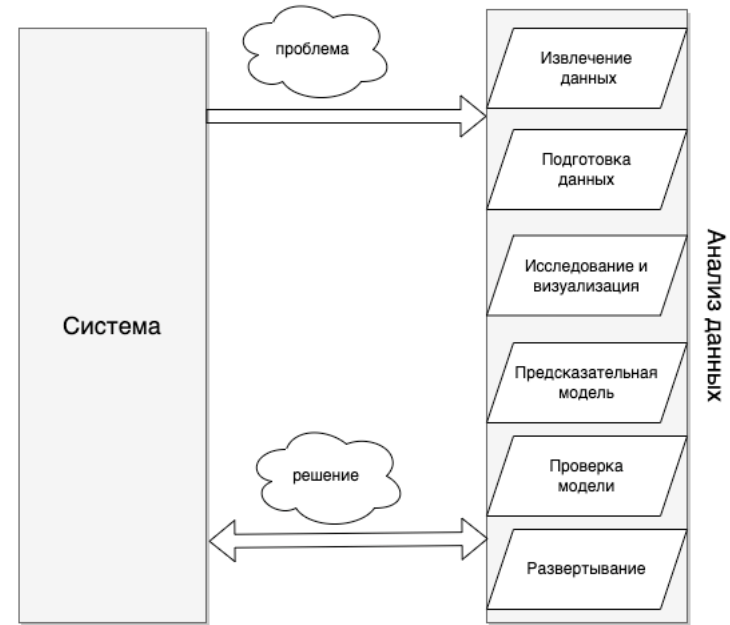
Практика



Анализ данных

Анализ данных — это всего лишь последовательность шагов, каждый из которых играет ключевую роль для последующих. Этот процесс похож на цепь последовательных, связанных между собой этапов:

- Определение проблемы;
- Извлечение данных;
- Подготовка данных — очистка данных;
- Подготовка данных — преобразование данных;
- Исследование и визуализация данных;
- Моделирование;
- Оценка (проверка) модели;
- Развертывание — визуализация и интерпретация результатов;
- Развертывание — развертывание решения.





Задание 1

Скачайте датасет. Проанализируйте его на наличие пропусков используя pandas.

Напишите в чат какие пропущенные значения вы обнаружили и причины их появления.



15 минут

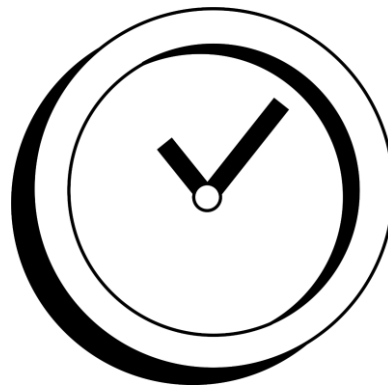


Задание 1

<<15:00->>

Скачайте датасет. Проанализируйте его на наличие пропусков и аномальных значений используя pandas.

Напишите в чат какие пропущенные значения вы обнаружили и причины их появления.





Задание 2

Найдите в датафрейме дубликаты. И удалите их. Значения могут быть одинаковыми но написаны по разному. Например может отличаться размер регистра(заглавные и строчные буквы)

Напишите в чат возможные причины появления дубликатов.



15 минут

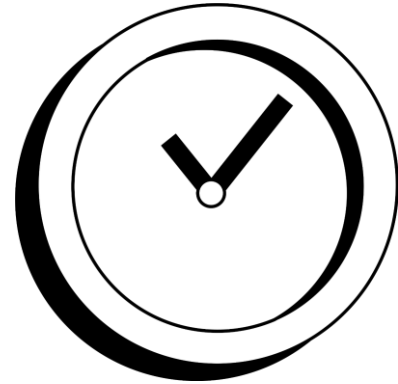


Задание 2

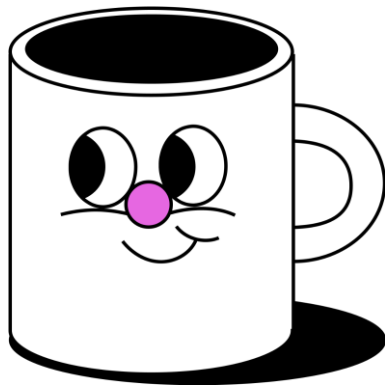
Найдите в датафрейме дубликаты. И удалите их. Значения могут быть одинаковыми но написаны по разному. Например может отличаться размер регистра(заглавные и строчные буквы)

Напишите в чат возможные причины появления дубликатов.

<<15:00->>



Перерыв



<<5:00->>



Задание 3

Сделайте колонку `purpose_category` в которую войдут следующие категории:

- операции с автомобилем,
- операции с недвижимостью,
- проведение свадьбы,
- получение образования

В чат напишите какое количество строк у вас получилось в каждой категории.



20 минут



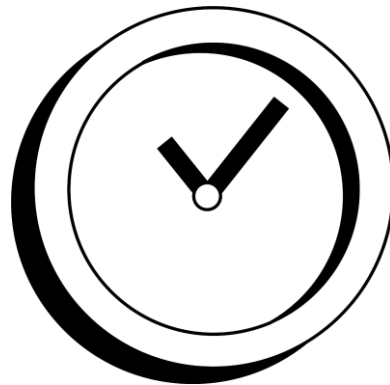
Задание 3

<<20:00->>

Сделайте колонку `purpose_category` в которую войдут следующие категории:

- операции с автомобилем,
- операции с недвижимостью,
- проведение свадьбы,
- получение образования

В чат напишите какое количество строк у вас получилось в каждой категории.

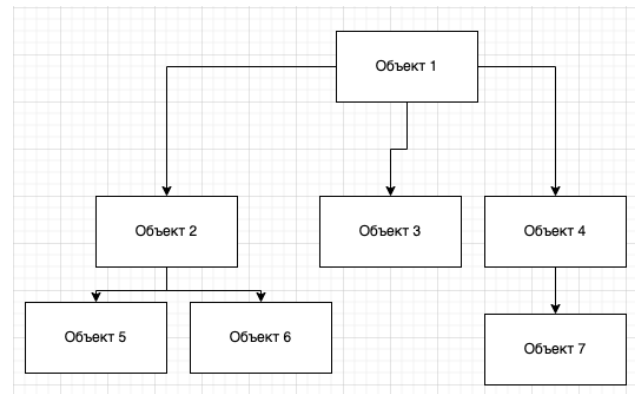




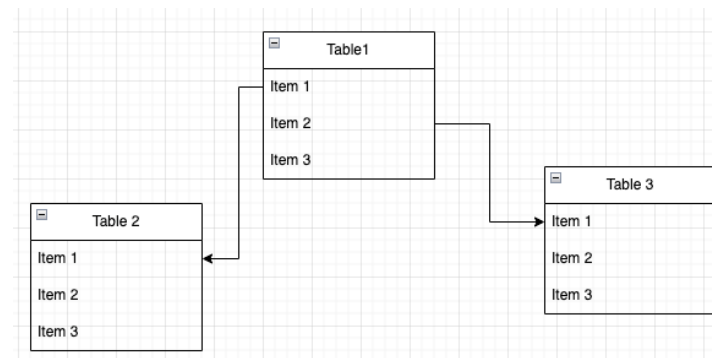
Иерархическая и реляционная модели

Иерархическая модель представляет собой совокупность элементов, расположенных в порядке их подчинения от общего к частному и образующих перевернутое по структуре дерево (граф).

Реляционная модель данных объекты и связи между ними представляет в виде таблиц, при этом связи тоже рассматриваются как объекты. Все строки, составляющие таблицу в реляционной базе данных, должны иметь первичный ключ. Все современные средства СУБД поддерживают реляционную модель данных.



Пример иерархической модели



Пример реляционной модели



Задание 3шщ

Постройте иерархическую и реляционную модели описывающие структуру предприятия состоящие из объектов
Отдел, Начальник, Сотрудник

Нарисуйте схему моделей используя app.diagrams.net и поделитесь картинкой в чате.



15 минут

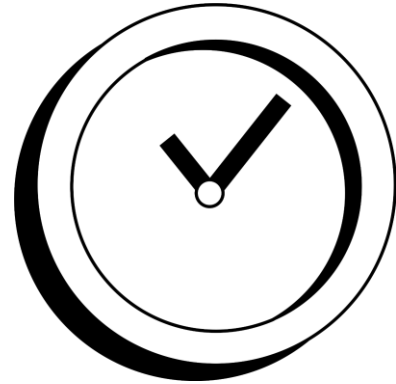


Задание 3

Постройте иерархическую и реляционную модели описывающие структуру предприятия состоящие из объектов
Отдел, Начальник, Сотрудник

Нарисуйте схему моделей используя app.diagrams.net и поделитесь картинкой в чате.

<<15:00->>





Вопросы?

Вопросы?



Вопросы?





Спасибо за внимание

