**Лекция 1**

Можно обращаться, Тигран (23 года).

**Big Data делится на 3 раздела**

**Дата инжиниринг** – то, что мы будем делать.

Это просто подготовка данных, их чистка, формирование хранилищ и правильное формирование хранилищ данных.

**Дата аналитика** – аналитика данных.

**Датасайнс** – прогнозы по наличию данных.

Плоские файлы – файлы в виде документа

Структурированные данные – данные в таблицах с атрибутами и т.д.

Наша задача выстроить процесс загрузки данных в БД, с приведением их к единому формату.

**Мы будем использовать следующие инструменты:**

SQL Developer Oracle

Для курса необходимы знания:

SQL, Python (SQLite)

**В начале курса изучаем:**

SQL

Операторы

Объединения данных

Самостоятельная на оценку (4-5 задач на написание вопросов)

Строковые функции

Регулярные выражения

Аналитические функции

Оконные функции

Инкрементальная загрузка (Формирование хранилища и автоматизация)

В DS работают с данными, которые могут весит 50 ТБ, 100 ТБ и больше.

Файлы дробят и постепенно загружают в БД, чтобы избегать ошибок.

Посещение лекций + Практика на лекциях (сообщения в MS Teams) + Выполнение всех КТ

Сдача работ не вовремя => оценка не больше 6 баллов

При возникновении вопросов – задавай.

Запросы на курсе могут достигать длины 60-150 строк.