**Вопросы:**

1. Что представляет из себя разведочный анализ данных?
2. Какая основная библиотека Python используется для анализа данных этим способом?
3. Опишите своими словами процесс анализа данных в соответствии с тем, что рассказывает спикер.
4. Решите задачу, предложенную спикером.
5. Что представляет из себя процесс ETL?

**Ответ:**

1. Разведочный анализ данных – это предварительное исследование датасета с целью определения его основных характеристик, взаимосвязей между признаками

2. Ею является Pandas profiling

3. Сначала мы узнаем общую информацию по каждому параметру, такую как среднее значение , максимальное и минимальное значения, процентили. А далее более подробно выводи информацию с помощью библиотеки pandas\_profiling и соответственно анализируем полученные данные.

4. ссылка на коллаб с кодом для анализа

<https://colab.research.google.com/drive/1vvuG18H_ksIXtb2z4zFkyHoVfu28vASS?usp=sharing>

Полученные данные помогают в предсказании энергопотребления. Мы можем оперировать такими данными как температура воздуха и давление, так как они коррелируют между собой. Т.е. с понижение температуры и понижением давления энергопотребление тоже уменьшается. Так же мы можем заметить что при почти нулевых значениях скорости ветра , давление тоже понижается. И это мы видимо из графиков корреляции данных.

5. ETL (Extract, Transform, Load) — извлечение, преобразование и загрузка. То есть процесс, с помощью которого данные из нескольких систем объединяют в единое хранилище данных.