



Что ты должен знать и уметь объяснить:



1. Что за данные мы использовали?

Ответ:

Мы использовали открытый датасет Airbnb, в котором содержится информация об объявлениях по аренде жилья (цены, отзывы, минимальное количество ночей, и т.д.).
Всего там более 100 000 записей.

♦ 2. Что было нашей целью?

Ответ:

Наша цель — **уменьшить размерность данных**, то есть перевести многомерные данные (7 признаков) в 2 измерения, чтобы упростить анализ и визуализацию.

♦ 3. Какие признаки мы выбрали и почему?

Ответ:

Мы выбрали числовые признаки:

- price
- service fee
- minimum nights
- number of reviews
- reviews per month
- review rate number
- availability 365

Потому что только числовые данные подходят для методов PCA и FA.

♦ 4. Что было целевым классом (y)?

Ответ:

Мы выбрали `room type` — это категориальный признак, показывающий тип жилья (например, Private room, Entire home/apt).

Мы использовали его, чтобы сравнивать, как объекты разных типов распределяются в новых координатах.

♦ 5. Почему мы сделали подвыборку из 3000 строк?

Ответ:

Потому что факторный анализ — тяжёлая операция для большого объема данных, и без этого он бы не завершился.

Подвыборка позволяет сохранить представительность, но ускорить работу.

♦ 6. Что такое MinMaxScaler и зачем он нужен?

Ответ:

MinMaxScaler приводит все значения к диапазону от 0 до 1. Это важно, потому что без этого признаки с большими значениями (например, цена) могли бы "подавить" другие признаки.

♦ 7. Что делает PCA?

Ответ:

PCA — метод, который находит такие направления (компоненты), в которых разброс (дисперсия) данных максимален.

Это позволяет **сжать** данные, сохраняя при этом основную информацию.

♦ 8. Что делает Factor Analysis?

Ответ:

Factor Analysis пытается найти **скрытые факторы**, которые якобы влияют на все признаки.

В отличие от PCA, он строит модель, а не просто делает преобразование.

♦ 9. Что показал результат PCA?

Ответ:

- Первая компонента объясняет 45.4% дисперсии,
- Вторая — 28.0%,
- Всего две компоненты — около 73.4% информации.


♦ 10. Что показывают графики?

Ответ:

- PCA-график показывает, как данные распределяются по новым координатам.
- Можно видеть, как классы (типы жилья) отделяются друг от друга.
- FA-график показывает похожую картину, но с другим распределением, потому что он фокусируется на скрытых связях.

♦ 11. Чем отличаются PCA и Factor Analysis?

Характеристика	PCA	Factor Analysis
Цель	сохранить как можно больше дисперсии	найти скрытые факторы
Тип	геометрическое преобразование	статистическая модель
Простота	проще	сложнее
Интерпретация	легко понять	труднее интерпретировать

Если ты расскажешь это уверенно — ты точно получишь  оценку.

Хочешь, могу сделать для тебя шпаргалку в виде таблицы или карточек для учёбы?

