**Вопросы**

Ответьте на следующие вопросы:

* Как можно заполнить пропущенные значения в данных?

Медианой, модой, спец.меткой, с помощью модели предсказания данных.

* Что делать, если таргет несбалансированный?

Синтезировать доп. данные недостающего класса (пересемплирование). Использовать подходящие метрики, например, F-меру, ROC AUC. Подбирать пороги.

* Как бороться с переобучением?

Кросс-валидацией, регуляризацией (штрафы).

* Найдите потенциальные причины аномалий в наборе данных.

Причины аномалий — ошибки фиксирования данных в массив. Это может быть и чел. фактор, и неисправность датчиков, передающих неверные данные, и т. п.

* Что делать, если данных не хватает или они плохого качества?

Собирать больше данных, очищать данные (исправлять ошибки, восстанавливать, убирать выбросы и т.п.)

* Что можно сделать чтобы повысить качество?

Удалить неинформативные признаки, повторы. Данные в разных регистрах привести к одному. Данные в разных форматах (временные, например) — сделать тоже самое.