**Вопросы**

* Как можно заполнить пропущенные значения в данных?

Средним — подходит для данных без выбросов.

Медианой — лучше, чем среднее, если есть выбросы.

Интерполяцией — подходит для временных рядов, чтобы заполнить пробелы между соседними значениями.

* Что делать, если таргет несбалансированный?

Можно использовать модели учитывающие веса, где у меньшего класса будет больший вес и аналогично наоборот.

* Как бороться с переобучением?

Кросс-валидация: модель обучается на разных частях данных, что помогает выявить её устойчивость и избежать переобучения.

Регуляризация: добавление штрафов за сложные модели, чтобы снизить их склонность к запоминанию данных.

* Найдите потенциальные причины аномалий в наборе данных.

Во многих данных присутствуют выбросы, особенно в признаках 80-122.

Так же в признаках 102-106 есть пропущенные значения, которые надо учитывать.

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

Удалить или заполнить пропуски, исправить выбросы и ошибки, создать новые признаки или найти дополнительные данные из внешних источников.

* Что можно сделать чтобы повысить качество?

Улучшить качество данных, подобрать наилучшие гиперпараметры, использовать кросс-валидацию, создать новые признаки, которые помогут модели лучше понять данные.