Question 1

What can be an indicator of usefulness of mean encodings?

Correct answers:

- Categorical variables with lots of levels.

Incorrect answers:

- A lot of binary variables. This is not an indicator because the majority of ML models deal with binary variables just fine. But keep in mind that there could be special cases when encoding binary variables with target mean may be useful. For example, KNN models might work better on mean-encoded binary features.
- Learning to rank task. There is no connection between mean encodings and learning to rank tasks.

Question 2

What is the purpose of regularization in case of mean encodings?

Correct answers:

- Regularization reduces target variable leakage during the construction of mean encodings.
- Regularization allows us to better utilize mean encodings. Only with regularization we can use mean encodings to the fullest.

Incorrect answers:

- Regularization allows to make feature space more sparse. Don't mix it up with L1 penalty.

Question 3

What is the correct way of validation when doing mean