**Project Description**

We need to protect the data of customers of the insurance company "Though the Flood". Develop a data transformation method that makes it difficult to recover personal information from it. Justify the correctness of his work.

We need to protect the data so that the quality of the machine learning models does not deteriorate during the transformation. There is no need to select the best model.

Project implementation instructions

1. Download and explore data.

2. Answer the question and justify the decision.

The features are multiplied by an invertible matrix. Will the quality of linear regression change? (It can be retrained.)

a. Will change. Give examples of matrices.

b. Will not change. Specify how the linear regression parameters are related in the original problem and in the transformed one.

3. Suggest a data transformation algorithm for solving the problem. Justify why the quality of the linear regression will not change.

4. Program this algorithm by applying matrix operations. Check that the quality of the linear regression from sklearn is not different before and after the transformation. Apply the R2 metric.

Data Description

The dataset is located in the /datasets/insurance.csv file. Download dataset.

• Characteristics: gender, age and salary of the insured person, number of family members.

• Target attribute: the number of insurance payments to the client over the past 5 years.