




ML for Business. Project

📅 Dates	@Mar 17, 2021 → Mar 21, 2021
☑ Finished	<input type="checkbox"/>
# Module	2
# Time spent, hours	
📌 Type	 Project

Project description

You work for the OilyGiant mining company. Your task is to find the best place for a new well.

Steps to choose the location:

- Collect the oil well parameters in the selected region: oil quality and volume of reserves;
- Build a model for predicting the volume of reserves in the new wells;
- Pick the oil wells with the highest estimated values;
- Pick the region with the highest total profit for the selected oil wells.

You have data on oil samples from three regions. Parameters of each oil well in the region are already known. Build a model that will help to pick the region with the highest profit margin. Analyze potential profit and risks using the *Bootstrapping* technique.

Project instructions

1. Download and prepare the data. Explain the procedure.
2. Train and test the model for each region:

- 2.1. Split the data into a training set and validation set at a ratio of 75:25.
- 2.2. Train the model and make predictions for the validation set.
- 2.3. Save the predictions and correct answers for the validation set.
- 2.4. Print the average volume of predicted reserves and model *RMSE*.
- 2.5. Analyze the results.
3. Prepare for profit calculation:
 - 3.1. Store all key values for calculations in separate variables.
 - 3.2. Calculate the volume of reserves sufficient for developing a new well without losses. Compare the obtained value with the average volume of reserves in each region.
 - 3.3. Provide the findings about the preparation for profit calculation step.
4. Write a function to calculate profit from a set of selected oil wells and model predictions:
 - 4.1. Pick the wells with the highest values of predictions.
 - 4.2. Summarize the target volume of reserves in accordance with these predictions
 - 4.3. Provide findings: suggest a region for oil wells' development and justify the choice. Calculate the profit for the obtained volume of reserves.
5. Calculate risks and profit for each region:
 - 5.1. Use the bootstrapping technique with 1000 samples to find the distribution of profit.
 - 5.2. Find average profit, 95% confidence interval and risk of losses. Loss is negative profit, calculate it as a probability and then express as a percentage.
 - 5.3. Provide findings: suggest a region for development of oil wells and justify the choice.

Data description

Geological exploration data for the three regions are stored in files:

- `geo_data_0.csv` . [download dataset](#)

- `geo_data_1.csv` . [download dataset](#)
- `geo_data_2.csv` . [download dataset](#)
- *id* — unique oil well identifier
- *f0, f1, f2* — three features of points (their specific meaning is unimportant, but the features themselves are significant)
- *product* — volume of reserves in the oil well (thousand barrels).

Conditions:

- Only linear regression is suitable for model training (the rest are not sufficiently predictable).
- When exploring the region, a study of 500 points is carried with picking the best 200 points for the profit calculation.
- The budget for development of 200 oil wells is 100 USD million.
- One barrel of raw materials brings 4.5 USD of revenue The revenue from one unit of product is 4,500 dollars (volume of reserves is in thousand barrels).
- After the risk evaluation, keep only the regions with the risk of losses lower than 2.5%. From the ones that fit the criteria, the region with the highest average profit should be selected.

The data is synthetic: contract details and well characteristics are not disclosed.

Project evaluation

We've put together the evaluation criteria for the project. Read this carefully before moving on to the task.

Here's what the reviewers will look at when reviewing your project:

- How did you prepare the data for training?
- Have you followed all the steps of the instructions?
- Have you taken into account all the business conditions?
- What are your findings about the task study?
- Have you applied the *Bootstrapping* technique correctly?

- Have you suggested the best region for well development? Is the choice justified?
- Did you avoid code duplication?
- Have you kept to the project structure and kept the code neat?

You have your takeaway sheets and chapter summaries, so you are ready to proceed to the project.

Good luck!