**Домашнее задание № 12**

**Лабораторная работа по теме «Решение задач Классического машинного обучения»**

**Автор: Сорокин Максим Евгеньевич**

**Формулировка задания:**

1. Примените алгоритм KNN (k-ближайших соседей) для данных по заболеваемости сахарным диабетом. Датасет представлен по ссылке.

Выходной столбец “outcome” принимает значения 0 или 1 и является классификатором текущего набора данных.

- имеется диабет (значение 1)

- заболевание отсутствует (значение 0)

Используйте все остальные столбцы как data features (входные числовые данные) для адаптации модели kNN.

Рассчитайте модель kNN для k = 5, 10, 15, 20, 25. Укажите объем тестового набора данных равным 25% от общего количества данных. Укажите score модели и постройте визуализацию обученной классификации для указанных k.

2. Постройте модель множественной линейной регрессии для атрибутов X = {SkinThickness, BMI} и Y = {Insulin} из датасета предыдущего задания.

Укажите score модели и предоставьте значения b, a1, a2 формулы:

y = b + a1\*x1 + a2\*x2

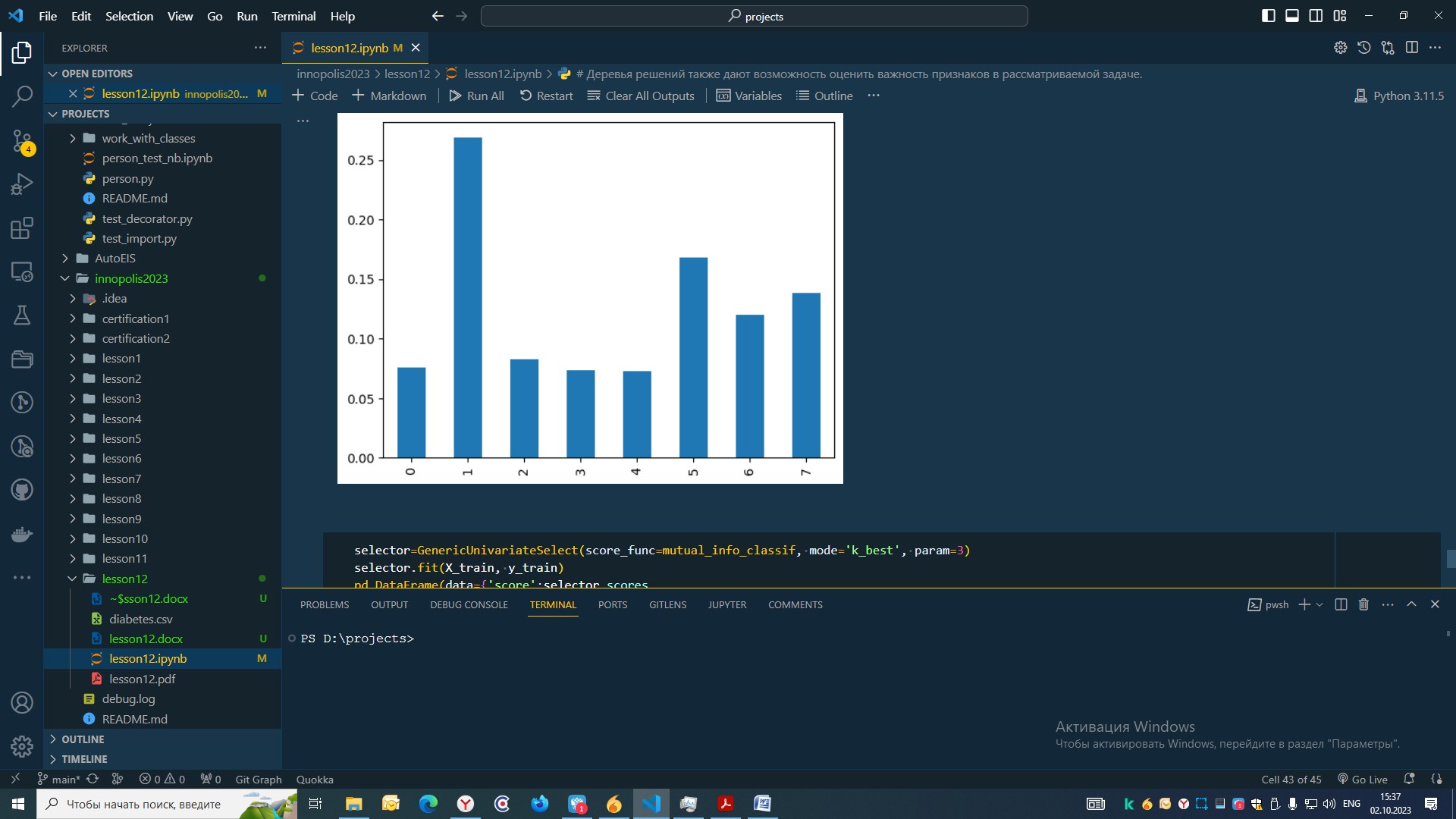
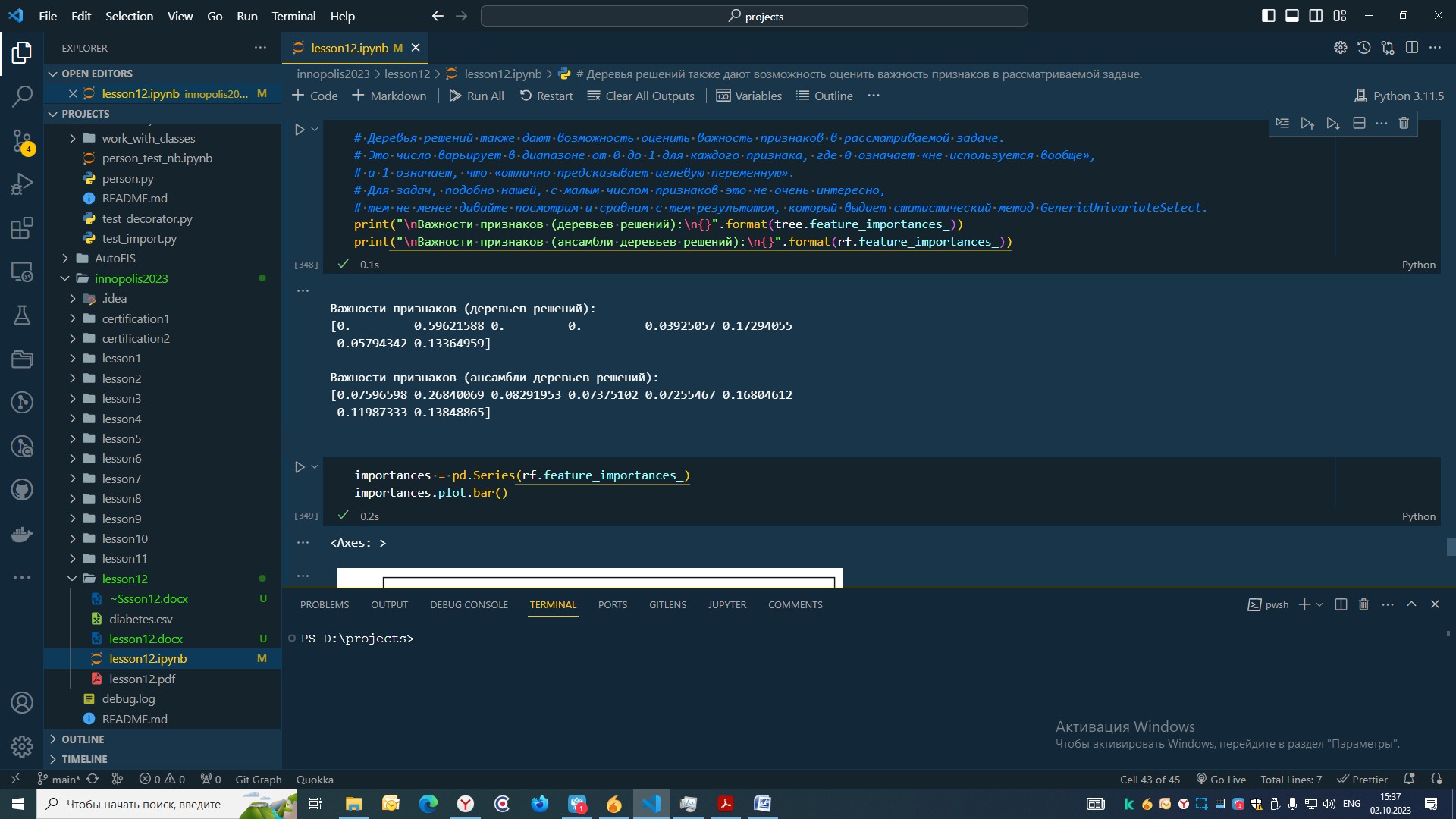
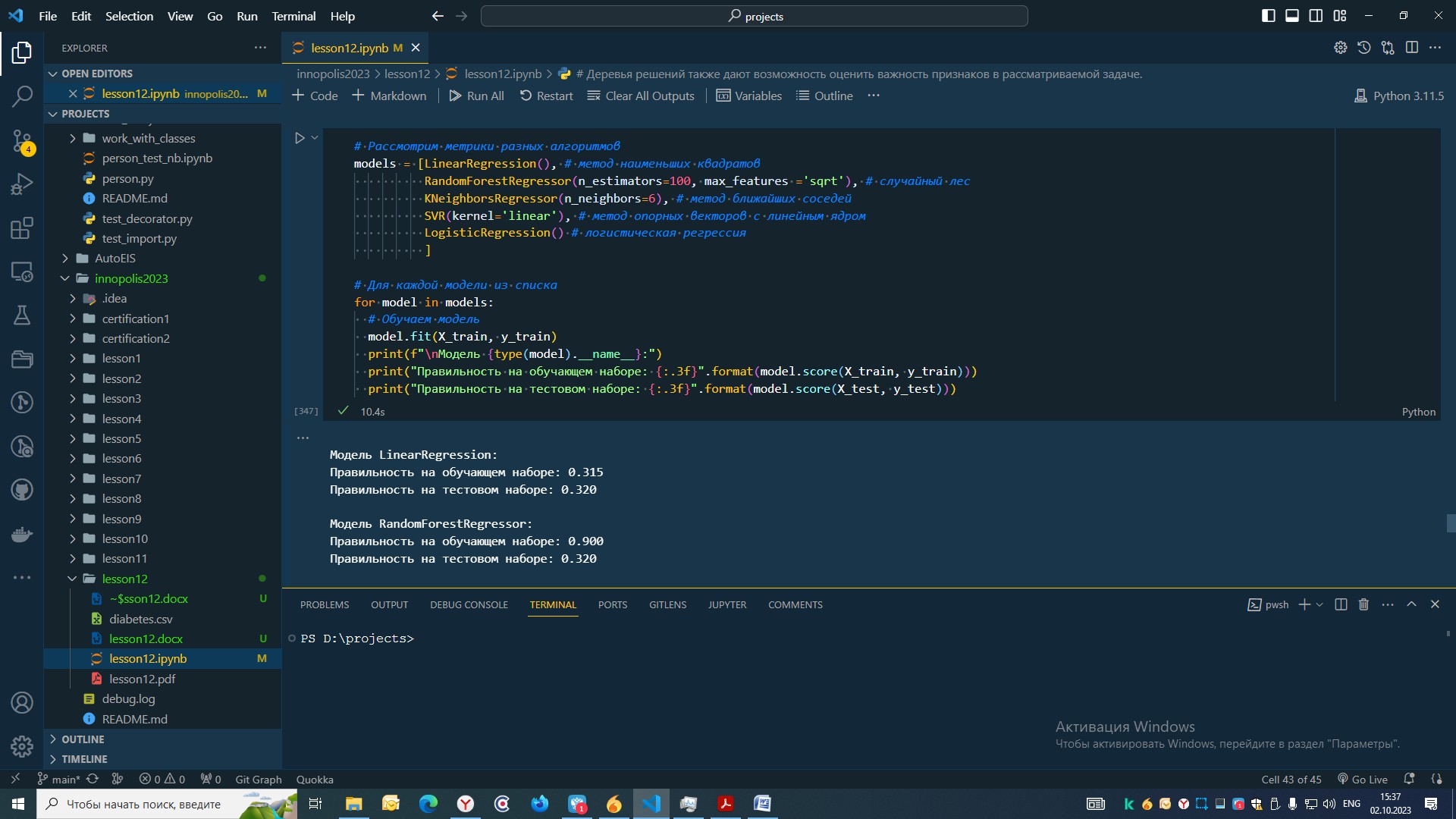
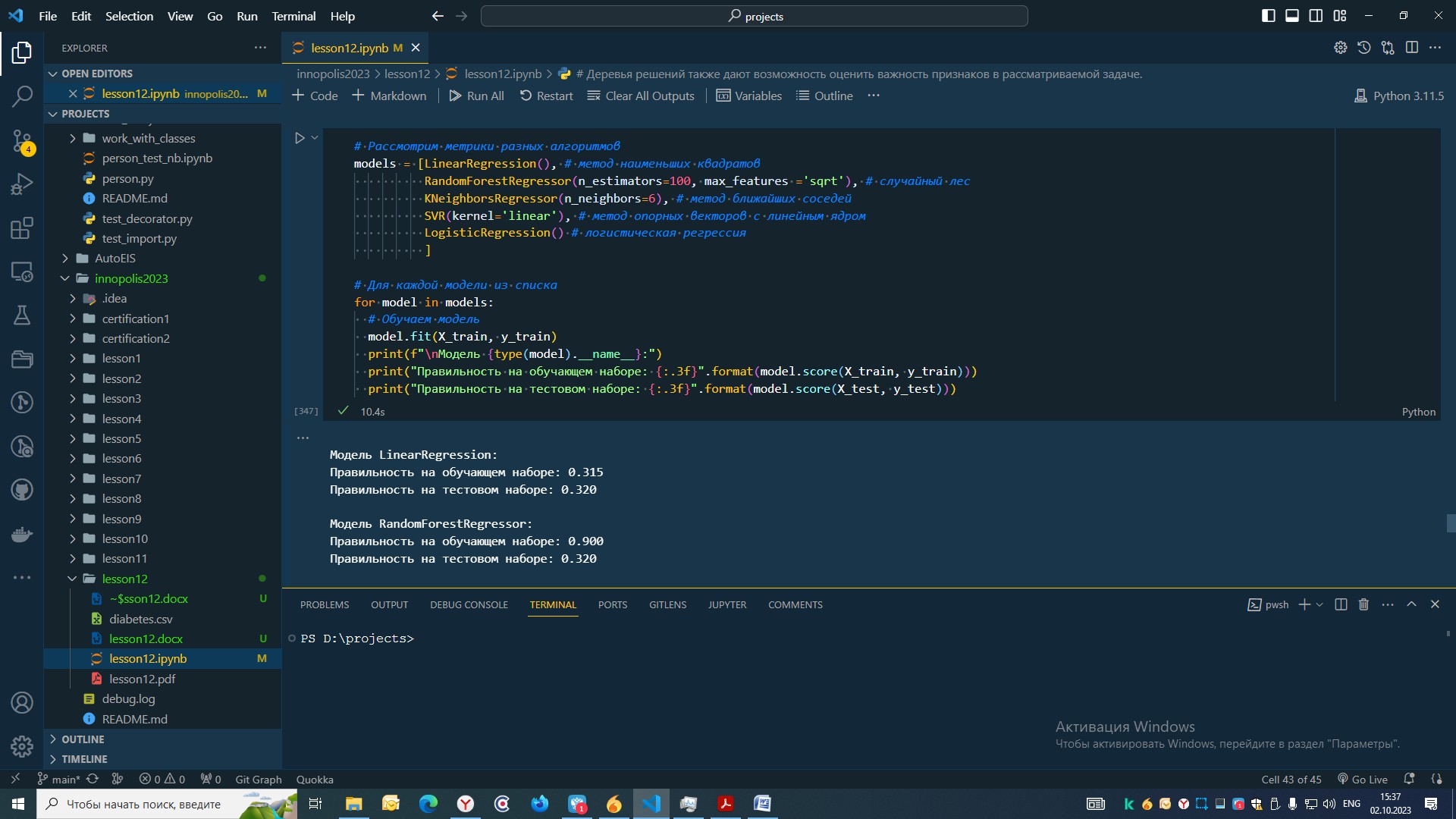
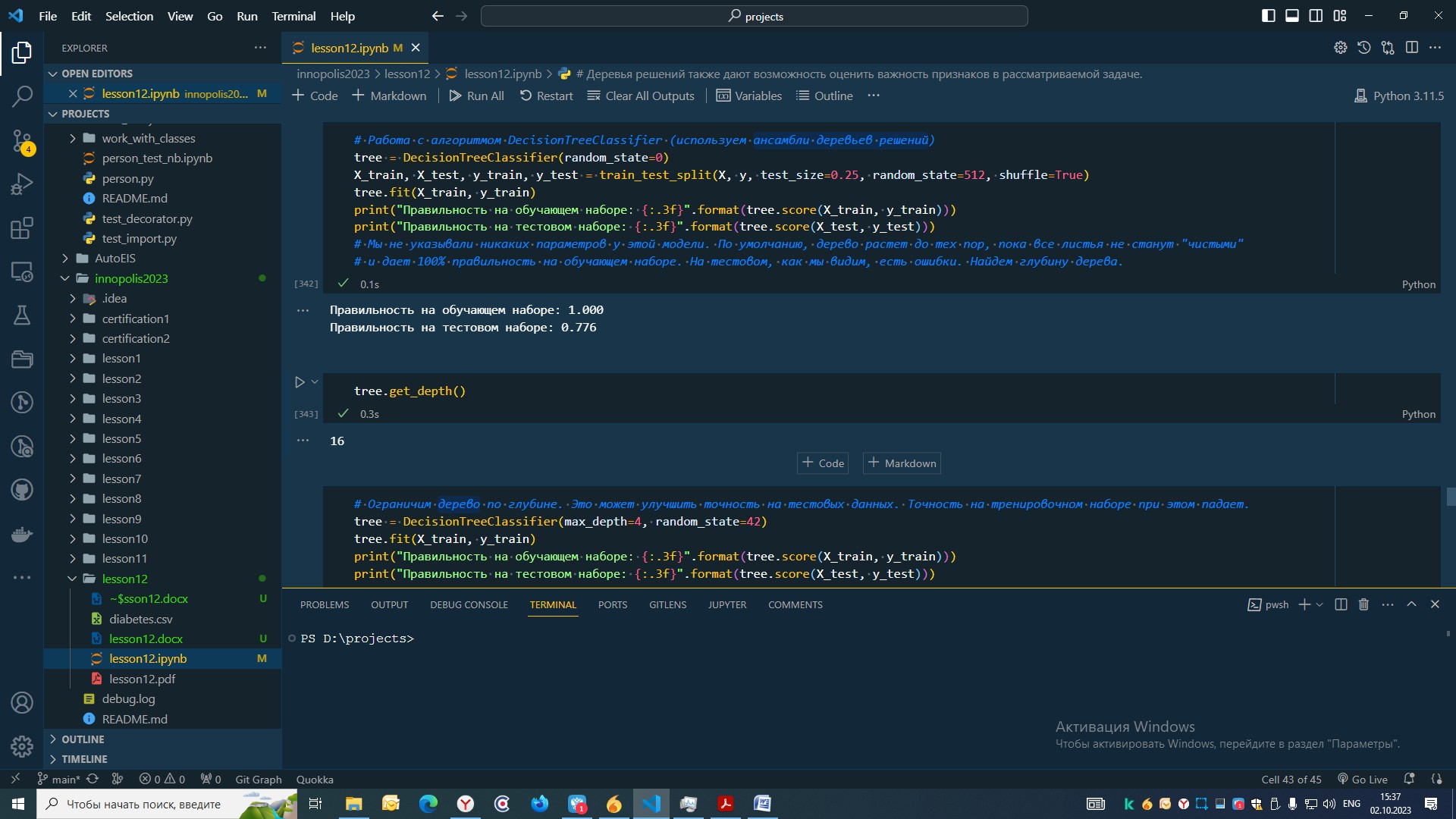
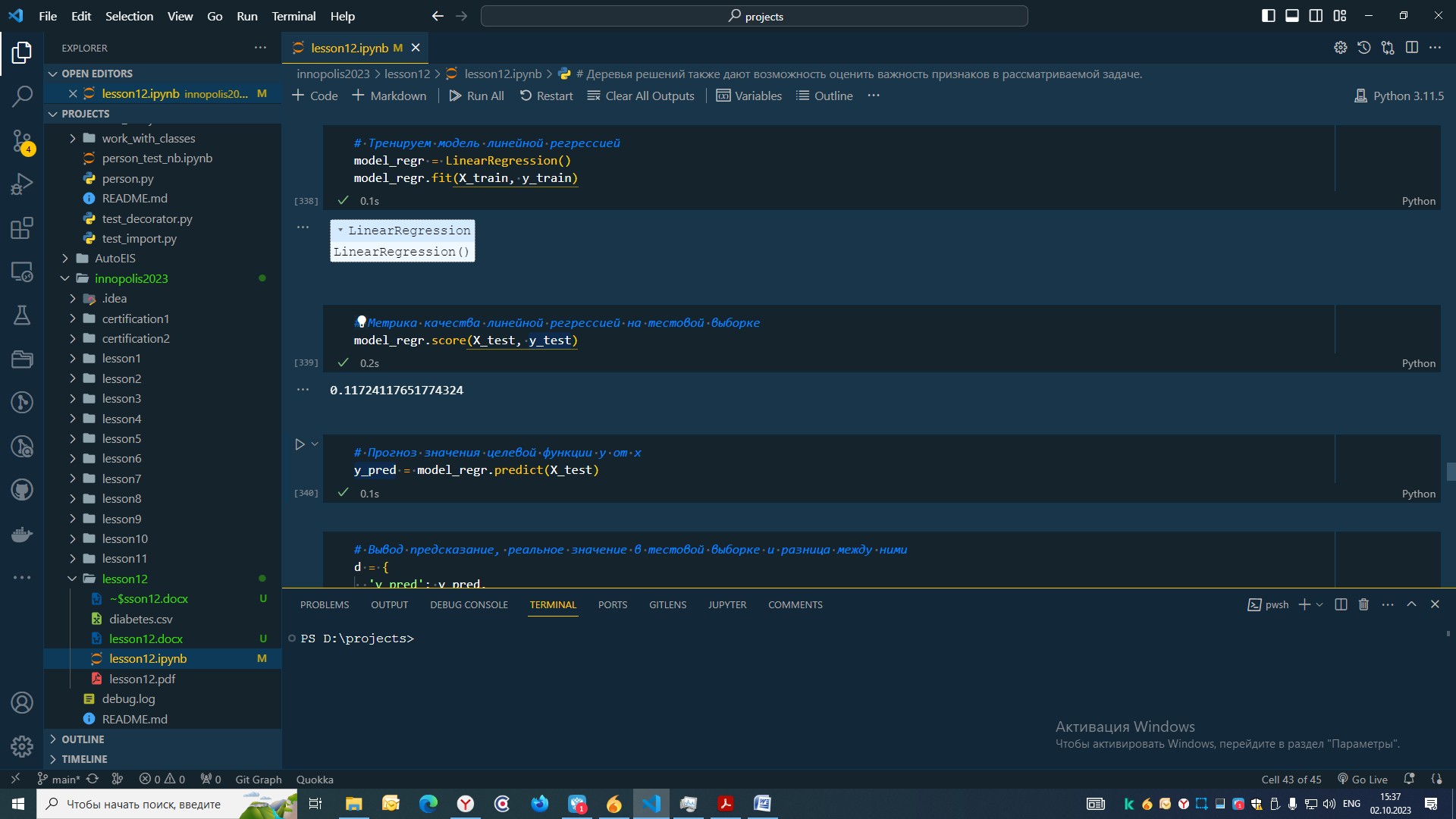
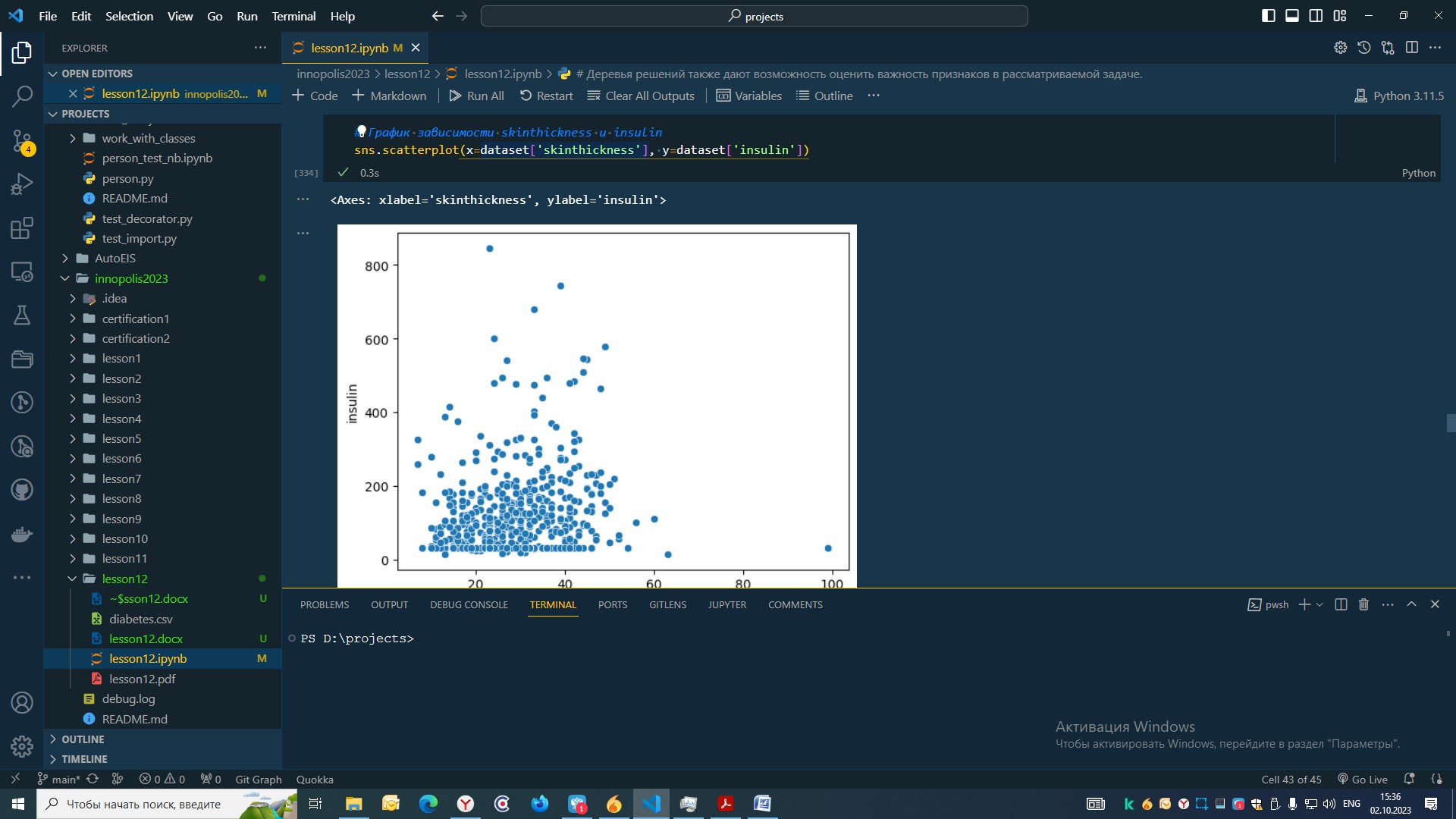
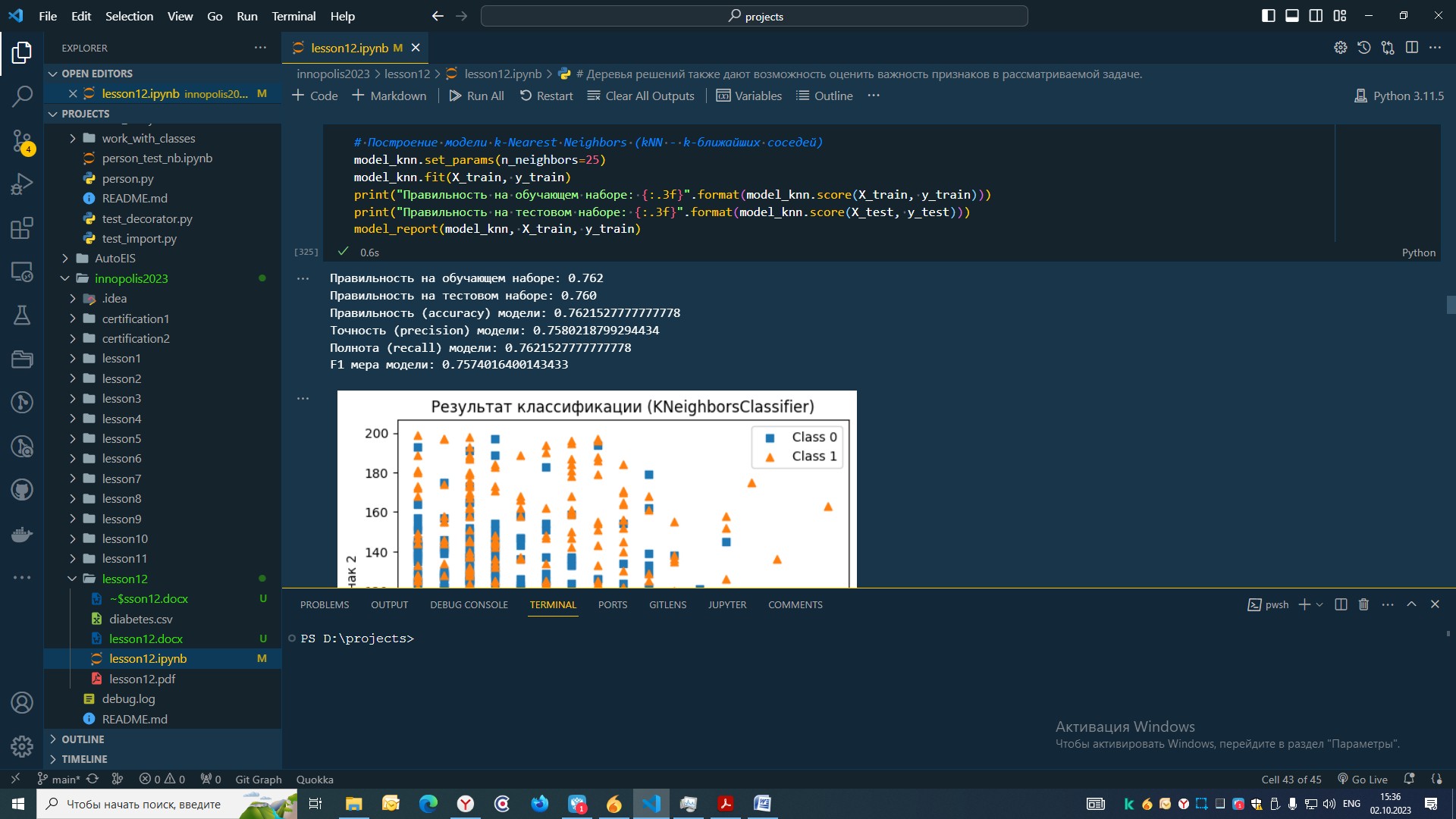
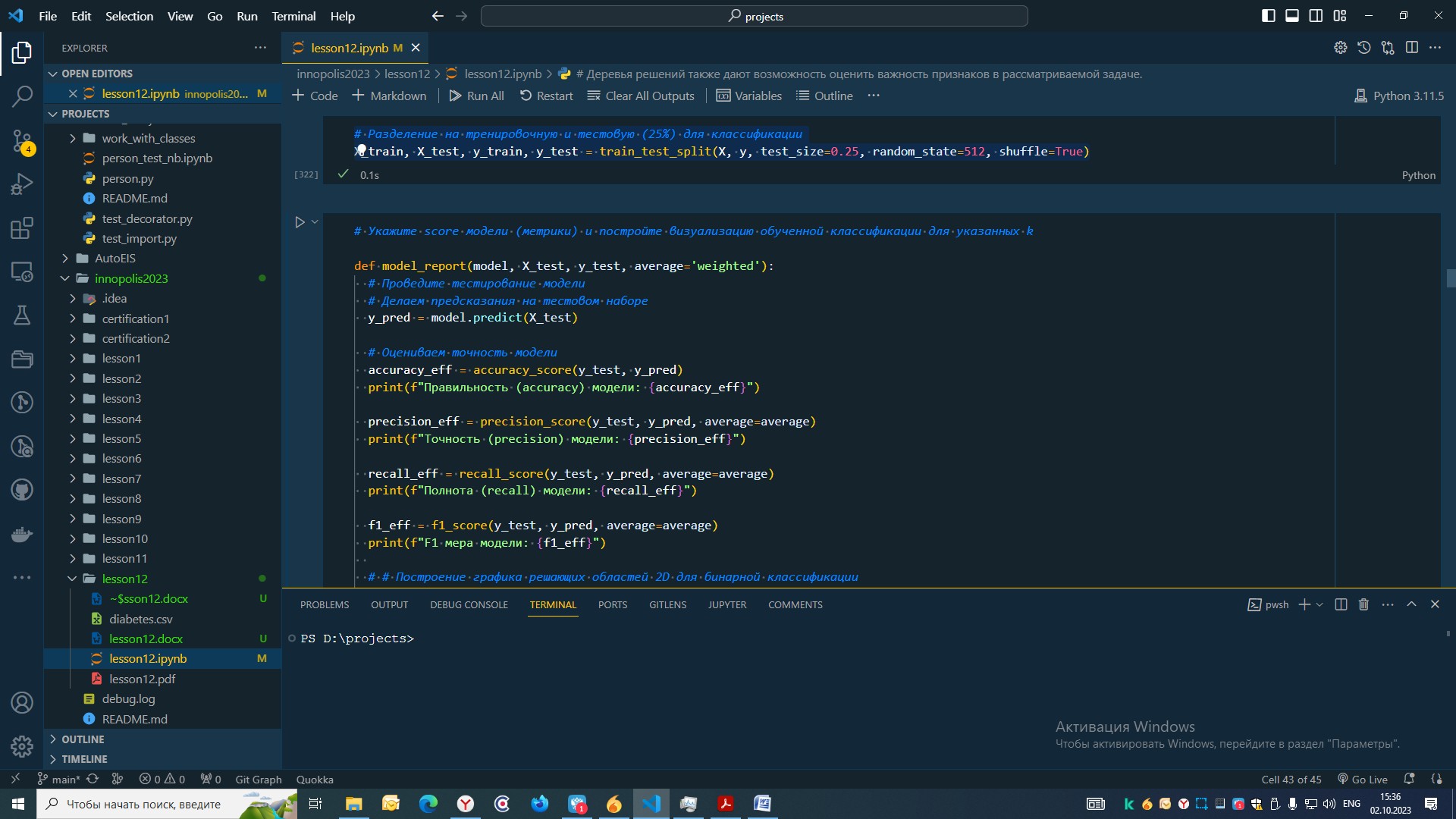
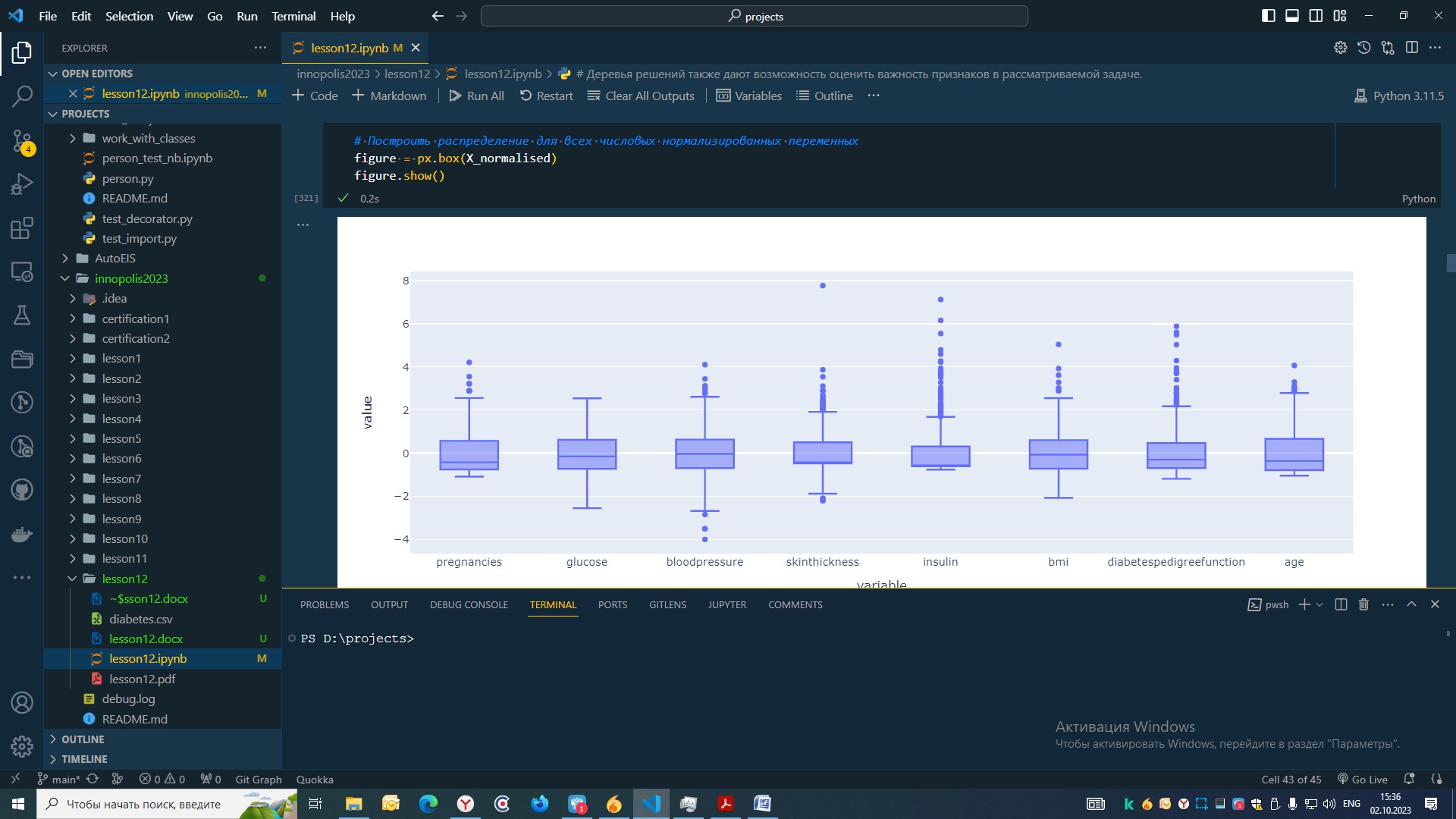
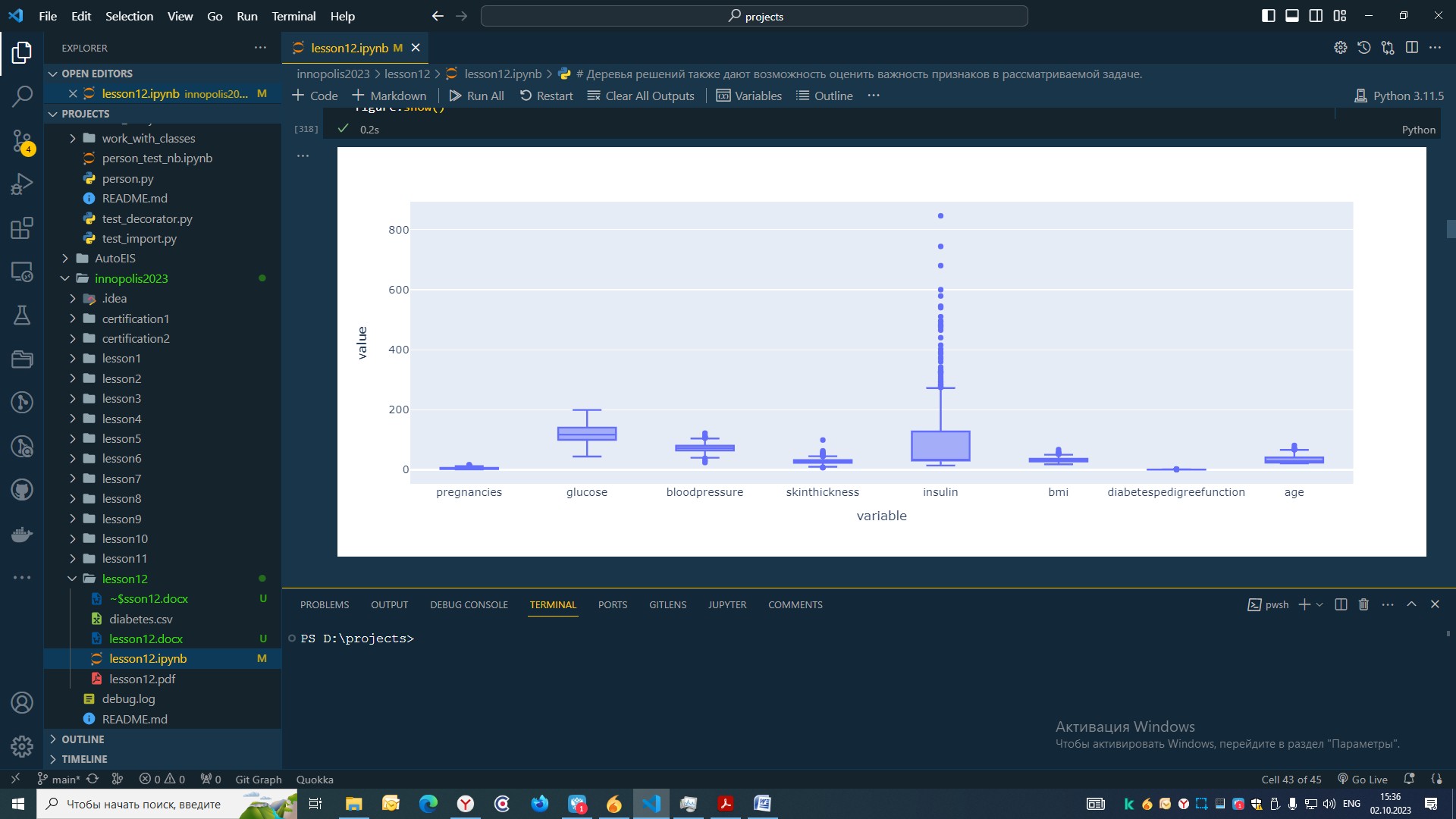
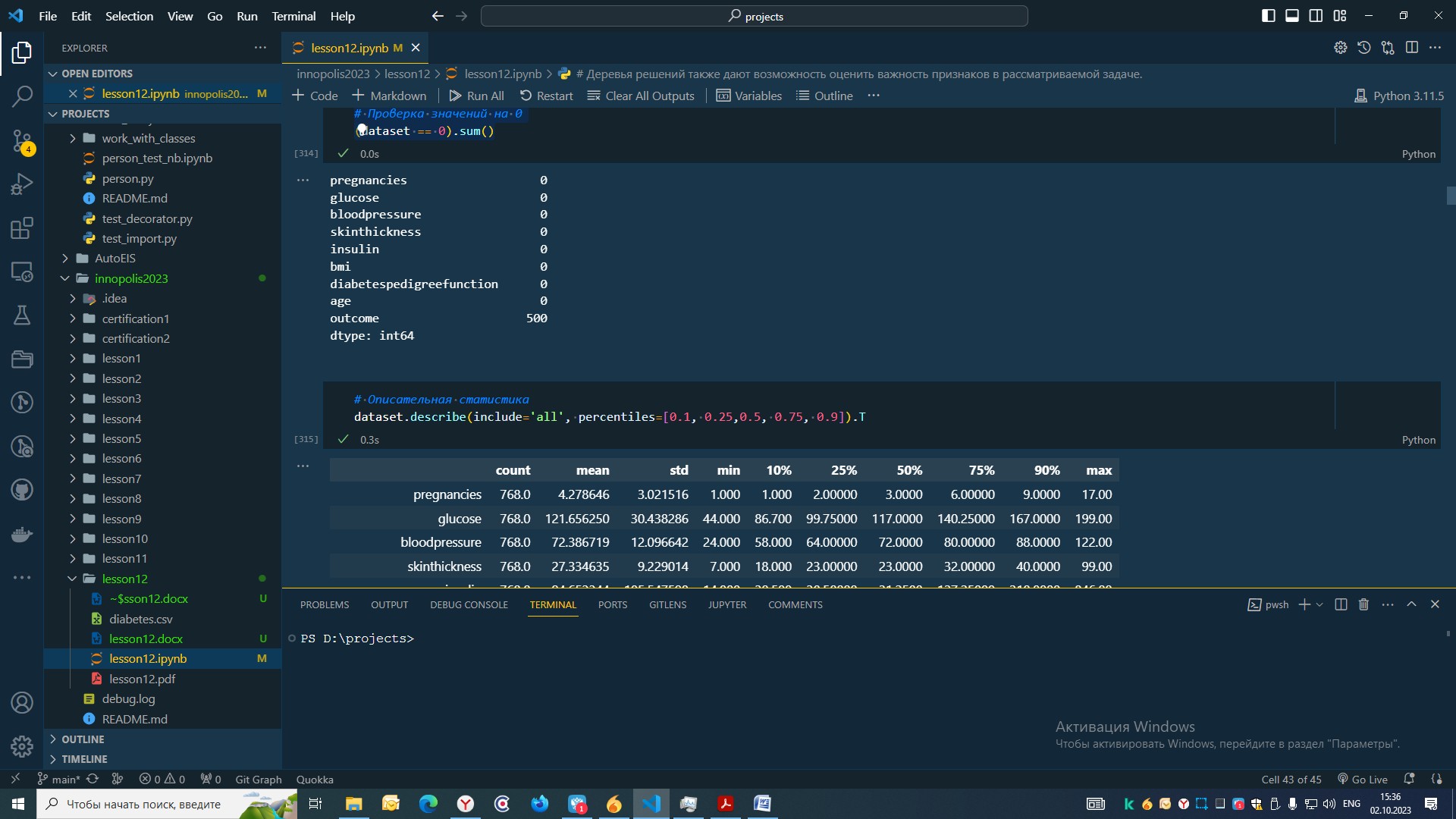
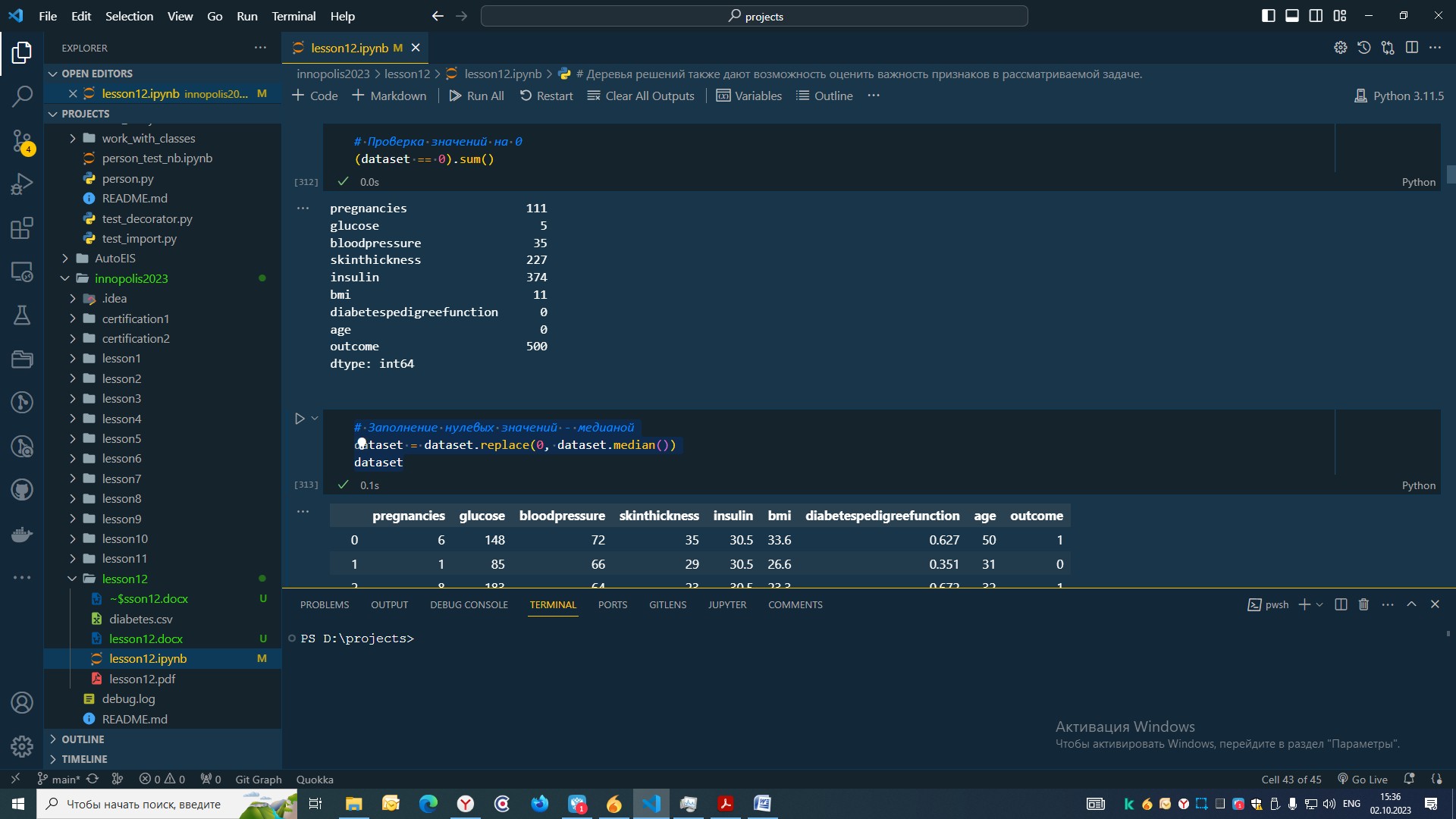
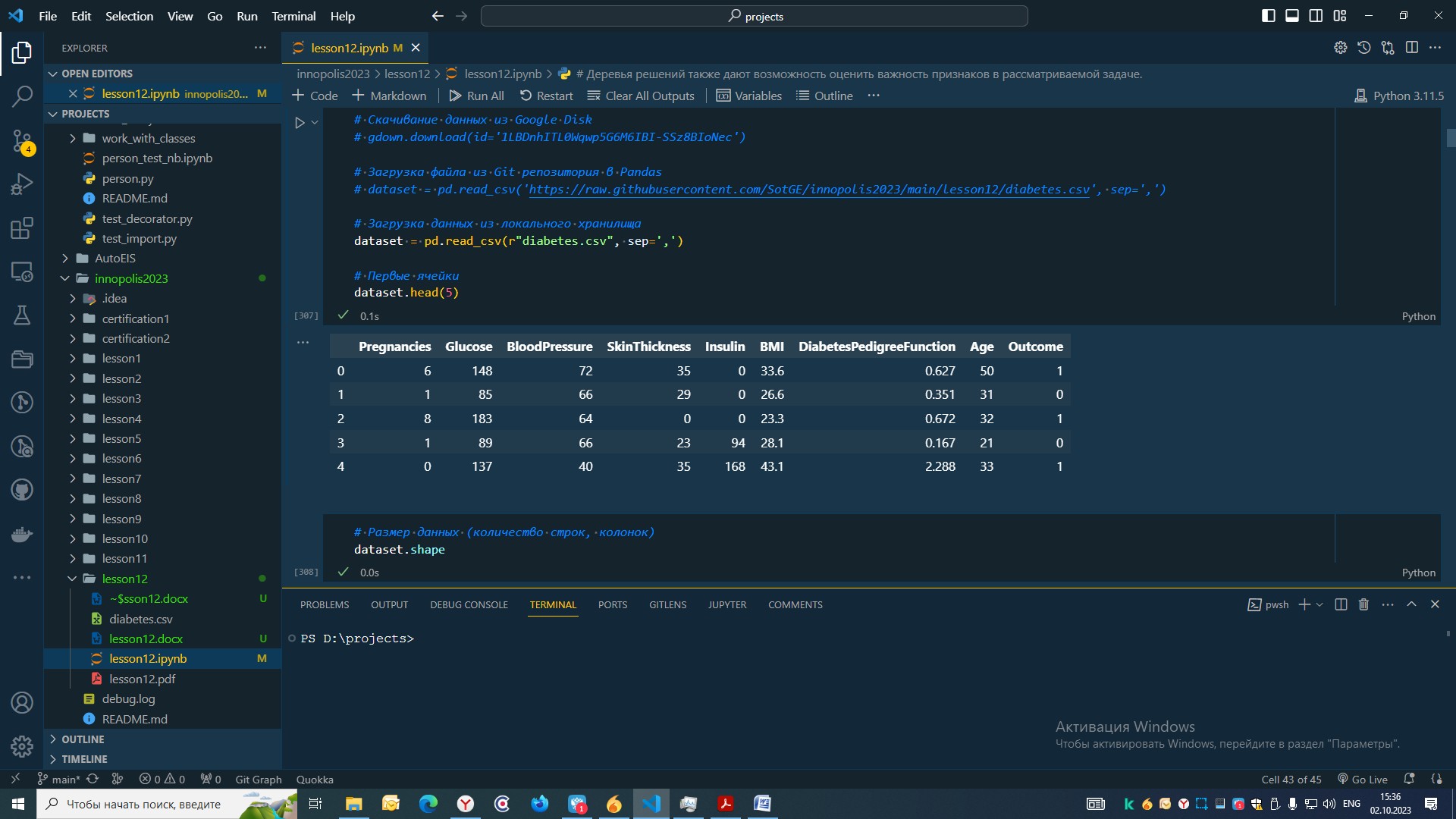
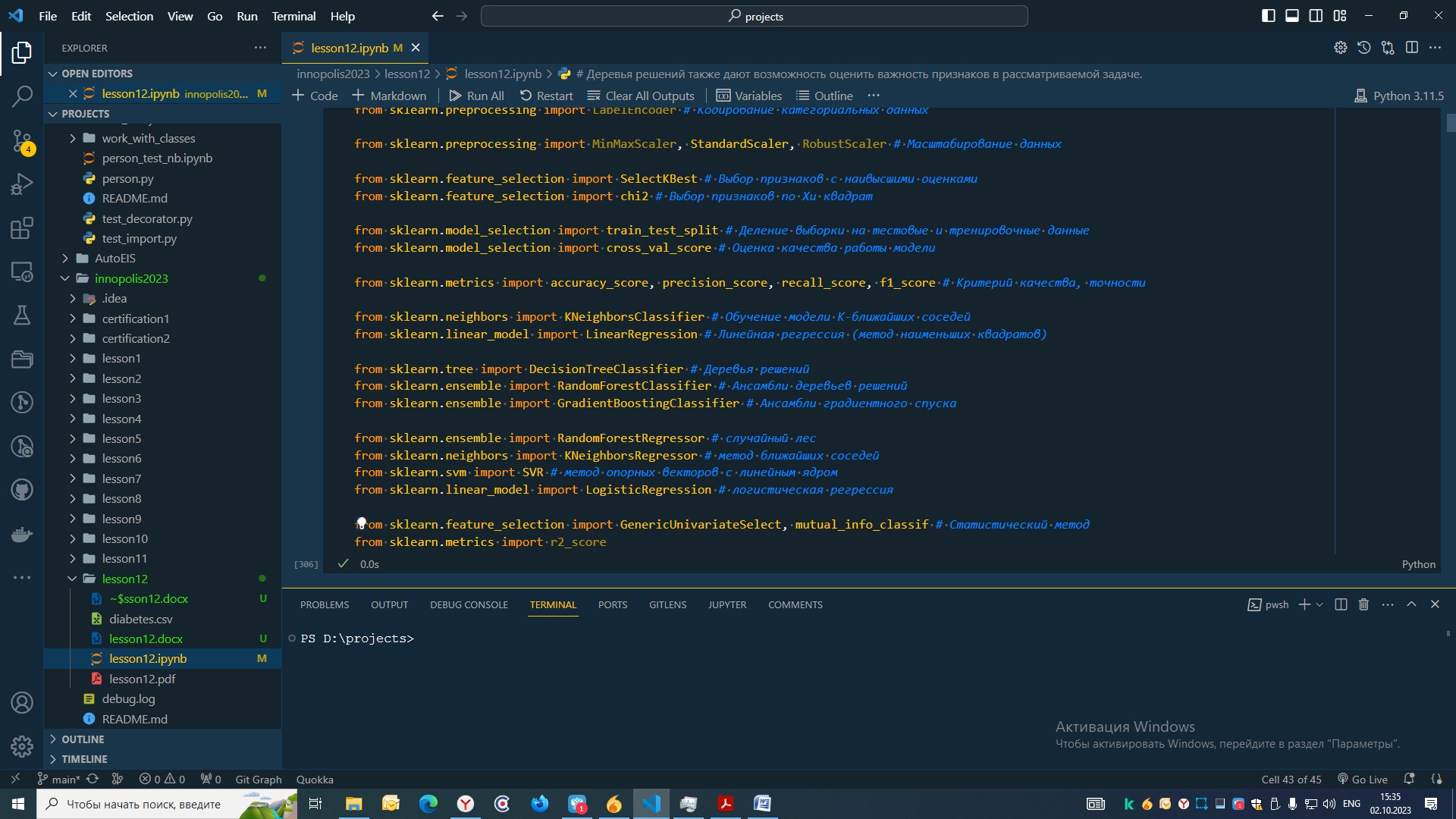
3. Примените к указанным данным любую одну модель (решите сами - для задачи классификации или задачи регрессии) из списка ниже

- SVM

- DecisionTree

- Полиномиальная регрессия

**Решение:**



**К решению приложены файлы:**

«lesson12.ipynb» – для Google Colab

«lesson12.py» – исходный код для Python

«diabetes.csv» –данные

Ссылка на Notebook Colab:

<https://colab.research.google.com/drive/1FfYvIcRzEySXFuF3a2b5W84JV0wIyYTM?usp=sharing>

Ссылка на GitHub:

<https://github.com/SotGE/innopolis2023/tree/main/lesson12>