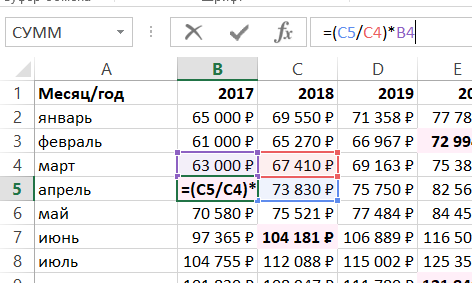
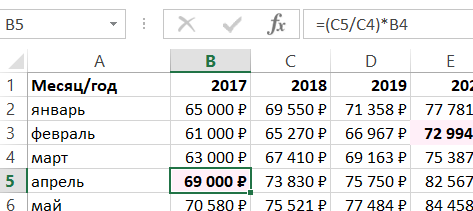
**ДОКУМЕНТАЦИЯ ВЫПОЛНЕННОЙ РАБОТЫ**

**Цели:** заполнить пропущенные месяца, сделать прогноз на 2023 и 2024 год

Для того, чтобы заполнить пропуска (апрель 2017, июнь 2018, ноябрь 2019, февраль и август 2020) была применена формула коэффициента сезонности, которая работает так: апрель 2018 года нужно разделить на март 2018 года, а затем получившееся число умножить на март 2017. Таким образом мы получим конечный ответ и заполним пропуск.



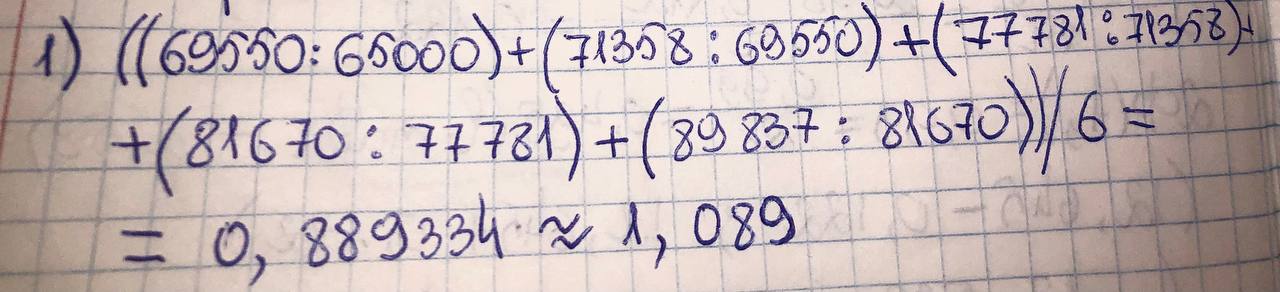


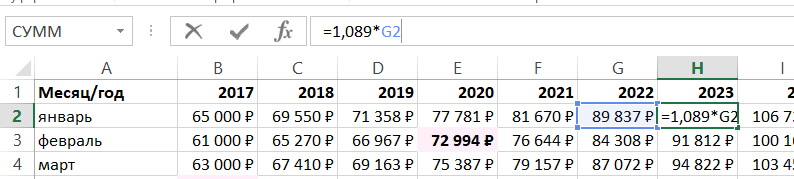
**Математическое ожидание** – понятие в теории вероятности, означающее среднее значение случайной величины.

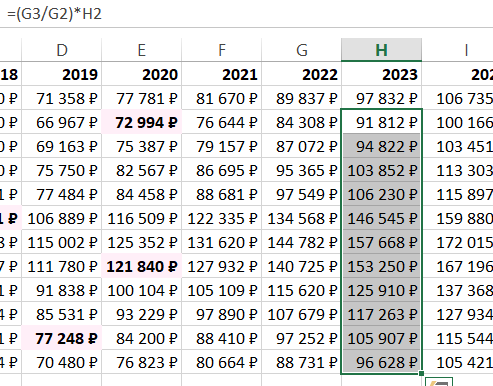
**Коэффициент сезонности** - величина, на которую уменьшаются/увеличиваются продажи в определенный период времени по сравнению со средним.

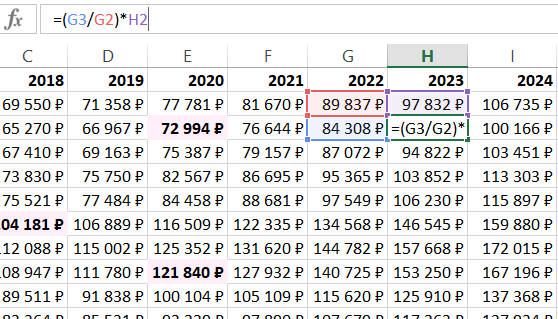
После того, как мы заполнили все пропуски, надо найти прогноз на 2023 и 2024 год. Чтобы это сделать сначала я нашла январь 2023 года по коэффициенту математического ожидания, работает это следующим образом:

(((2018 год / 2017 год)+(2019 год / 2018 год)+(2020 год / 2019 год)+(2021 год / 2020 год)+(2022 год / 2021 год)) / 6(общее число годов) \* 2022 год



Тем самым мы узнаем какая сумма доходов будет в январе 2023 года

А для того, чтобы узнать все оставшиеся месяца 2023 года я воспользовалась формулой **коэффициента сезонности**

2024 год я рассчитывала точно также, как и 2023, пользовалась абсолютно такими же формулами и получила нужный результат.

Далее по целям нашей работы нужно определить оптимистичный и пессимистичный прогнозы 2023 и 2024 года. Чтобы это сделать нам нужно найти отклонение.

Для нахождения отклонения была использована функция ДОВЕРИТ() в Excel. Функция ДОВЕРИТ в Excel предназначена для определения доверительного интервала для среднего значения, найденного для генеральной совокупности, которая имеет нормальное распределение.

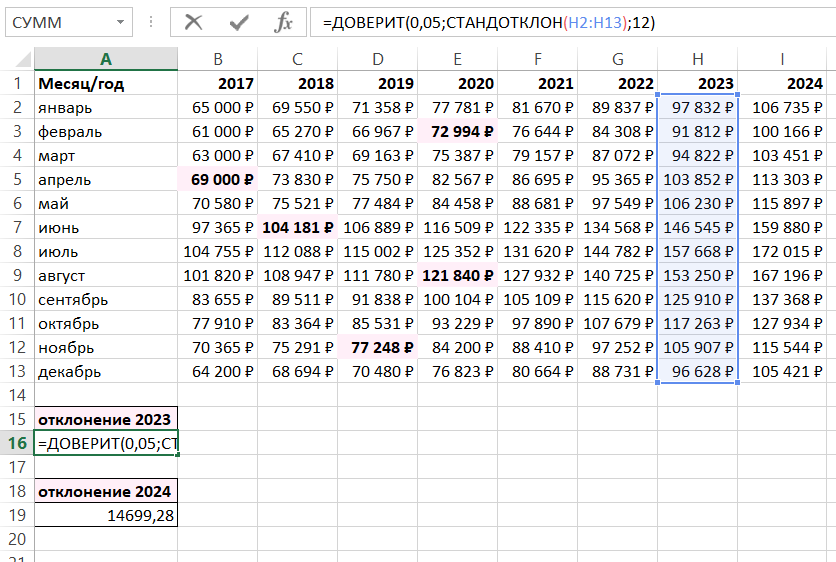
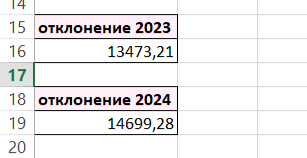
**Доверительный интервал** — это диапазон значений. Выборка "x" находится в центре этого диапазона, а диапазон — x ± ДОВЕРИТ. Например, если x — это пример времени доставки продуктов, заказаных по почте, то x ± ДОВЕРИТ — это диапазон средств численности населения.

**Синтаксис:**

ДОВЕРИТ(альфа;стандартное\_откл;размер)

Аргументы функции ДОВЕРИТ():

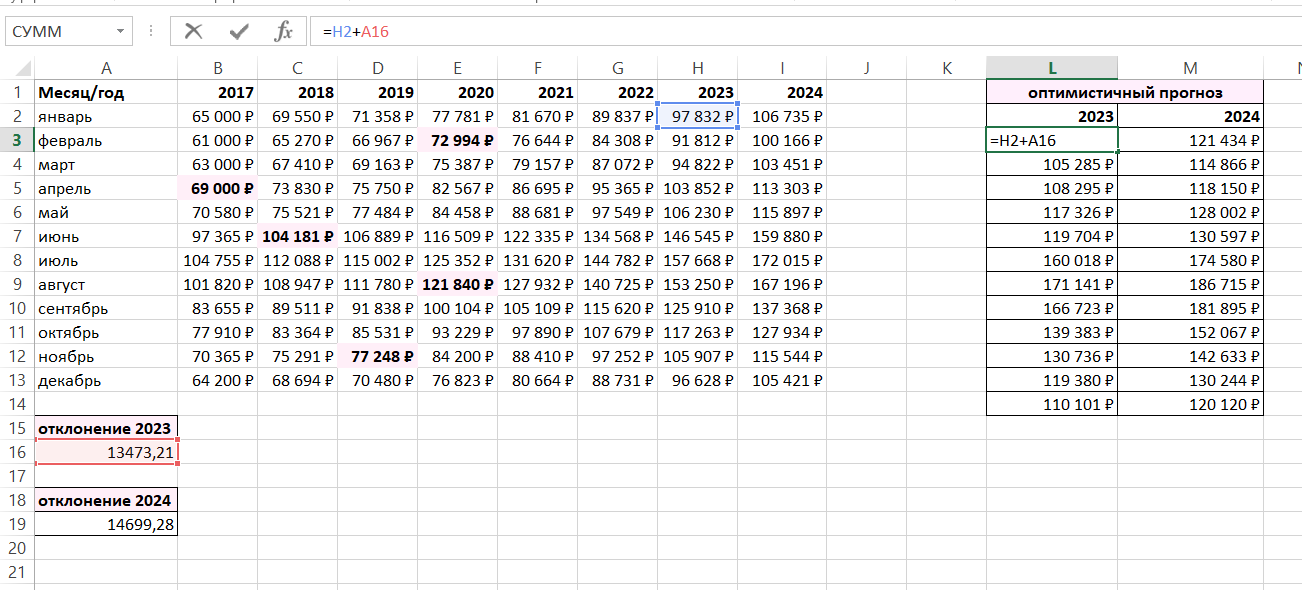
* **Альфа**     — обязательный аргумент. Уровень значимости, используемый для вычисления доверительного уровня. Доверительный уровень равен 100\*(1 - альфа) процентам или, иными словами, значение аргумента "альфа", равное 0,05, означает 95-процентный доверительный уровень.
* **Стандартное\_откл**     — обязательный аргумент. Стандартное отклонение генеральной совокупности для диапазона данных, предполагается известным.
* **Размер**     — обязательный аргумент. Размер выборки.



После того как мы узнали отклонение, мы спокойно можем посчитать пессимистичны и оптимистичный прогнозы на 2023 и 2024 год.

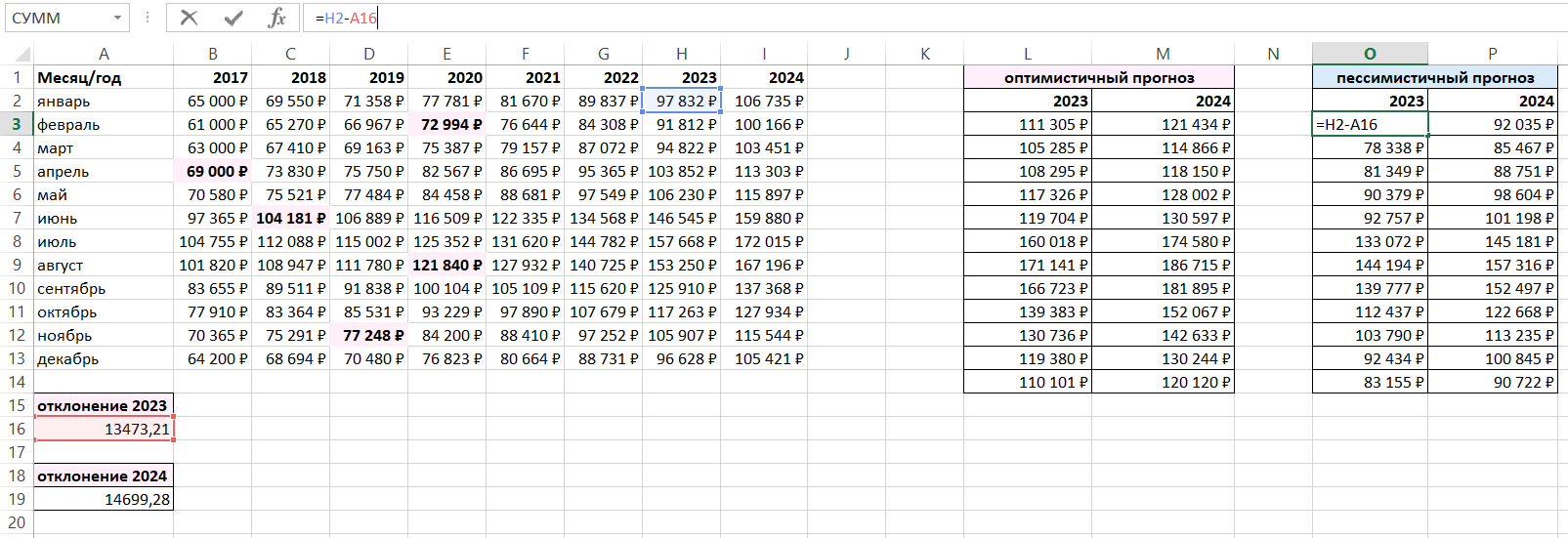
Оптимистичный прогноз был рассчитан следующим образом:

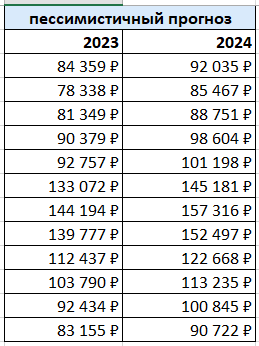
Получившееся отклонение нужного года складывается с прогнозом предыдущего года.



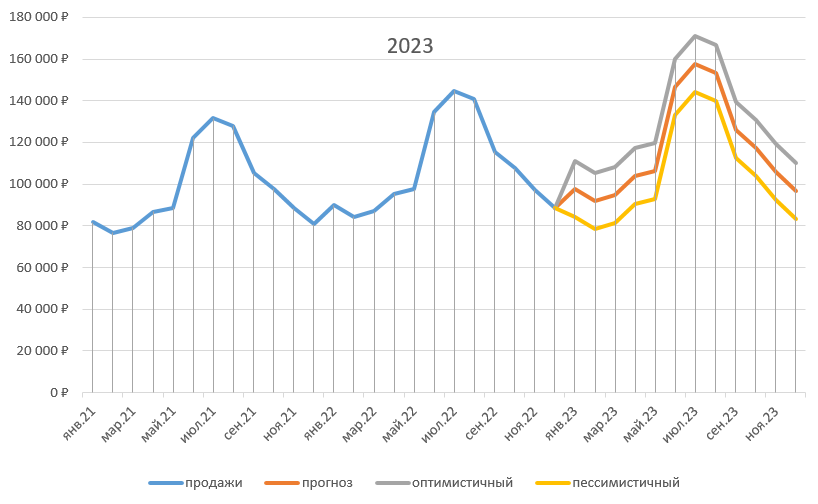


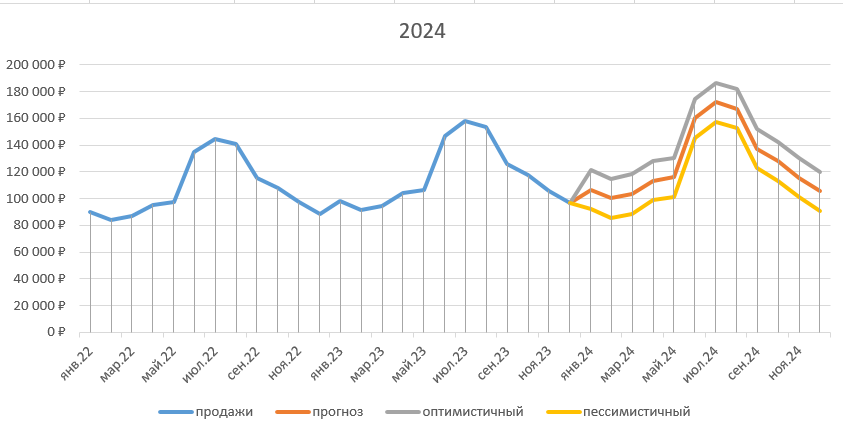
Пессимистичный прогноз был рассчитан следующим образом:

Из прогноза предыдущего года вычитается получившееся отклонение нужного года:



А также были созданы графики, по которым можно посмотреть прогнозы:





По итогу сделанной работы были выполнены все поставленные цели, а именно мы заполнили пропуски некоторых месяцев, рассчитали прогноз на 2023 и 2024 год и вычислили оптимистичный и пессимистичный прогнозы 2023 и 2024 года.

**PYTHON CODE**

Линейная регрессия. Регрессия ищет отношения между переменными. Зависимые данные называются **зависимыми переменными**, **выходами** или **ответами**.

Независимые данные называются **независимыми переменными, входами или предсказателями**.

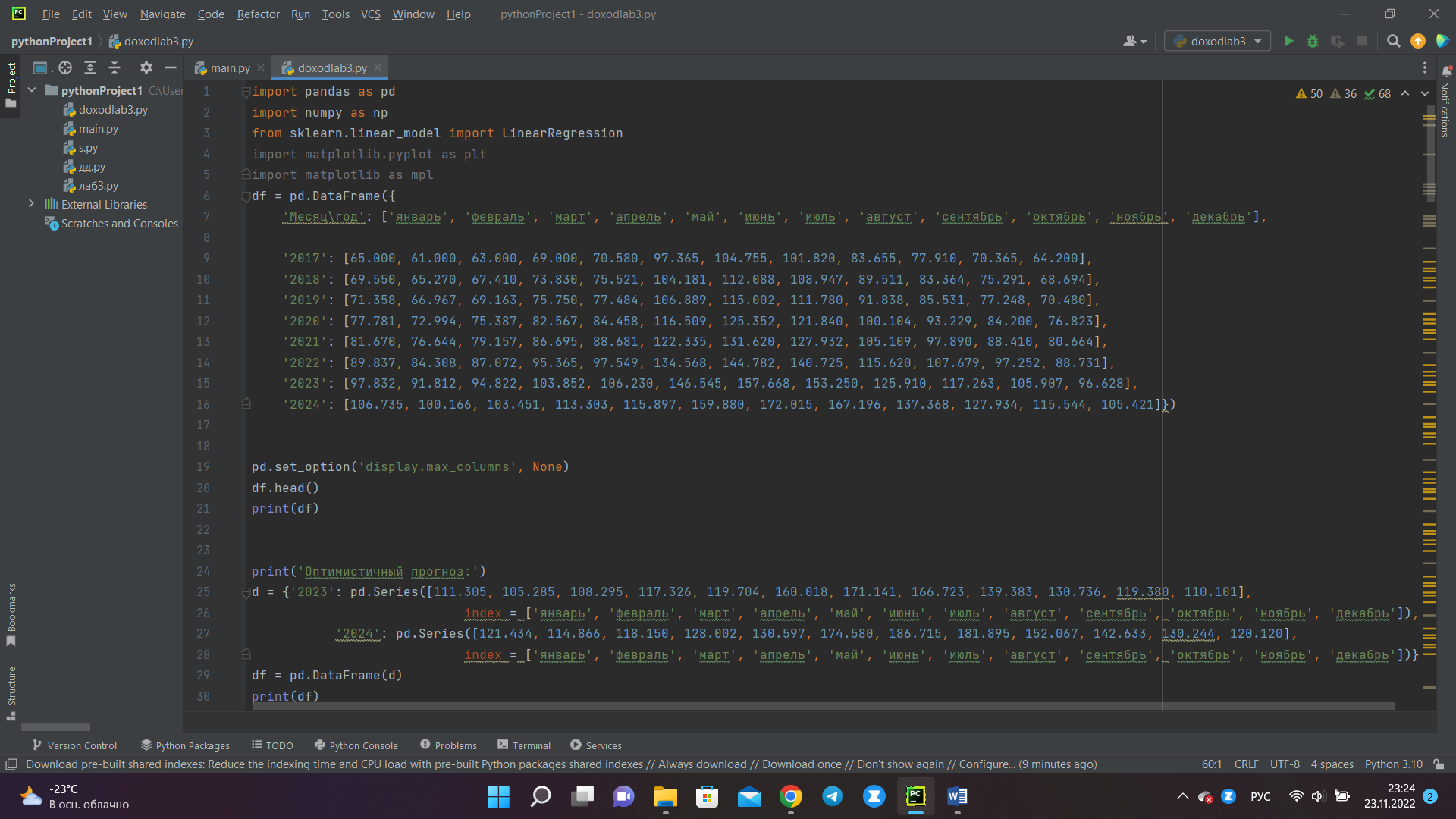
Обычно в регрессии присутствует одна непрерывная и неограниченная зависимая переменная. Входные переменные могут быть неограниченными, дискретными или категорическими данными, такими как пол, национальность, бренд, etc.

Регрессия полезна **для** **прогнозирования ответа** на новые условия. Можно угадать потребление электроэнергии в жилом доме из данных температуры, времени суток и количества жильцов.

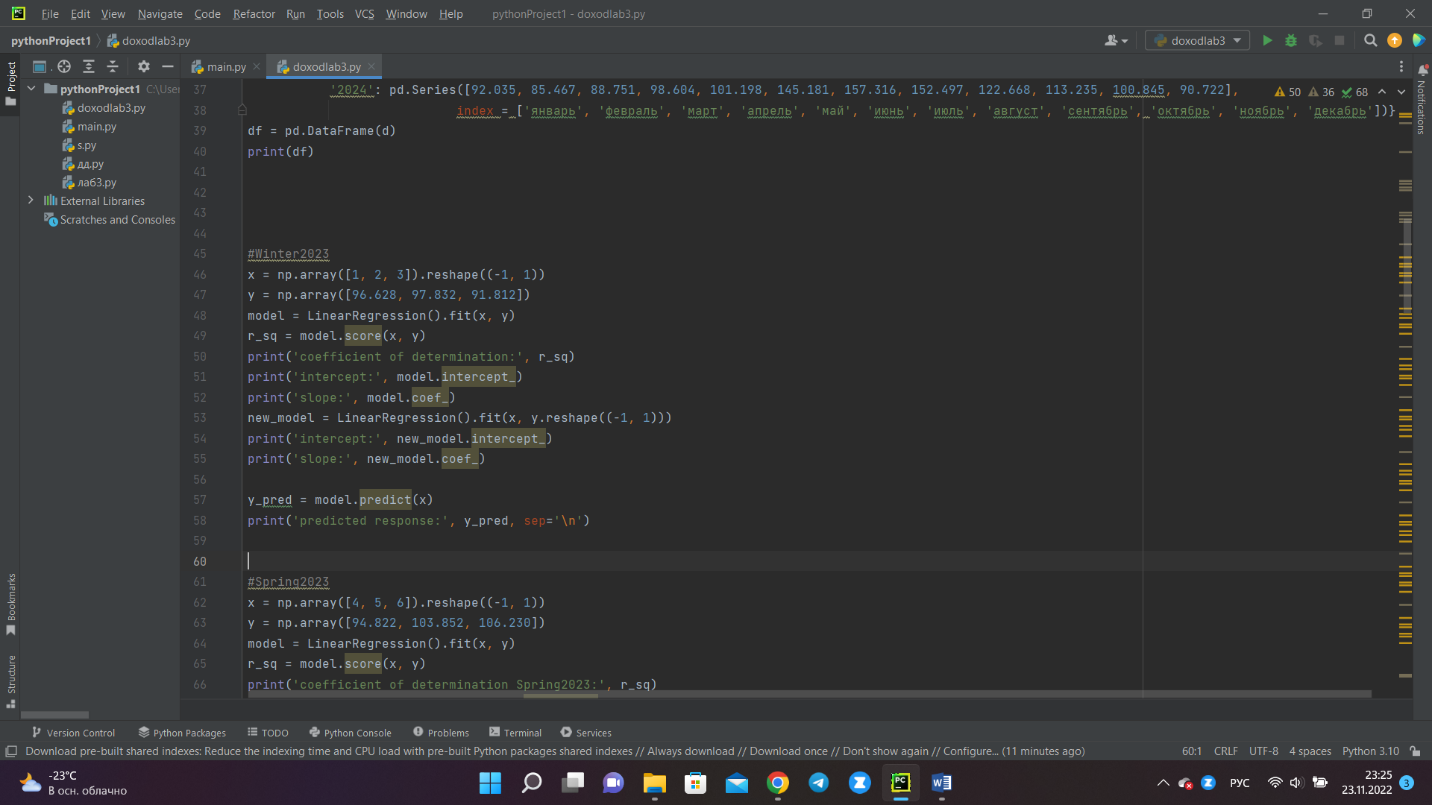
**NumPy** – фундаментальный научный пакет для быстрых операций над одномерными и многомерными массивами.

Пакет **scikit-learn** – это библиотека, широко используемая в машинном обучении. Она предоставляет значения для данных предварительной обработки, уменьшает размерность, реализует регрессию, классификацию, кластеризацию и т. д.

Первым шагом импортируем пакет NumPy и класс LinearRegressionиз sklearn.linear\_model:

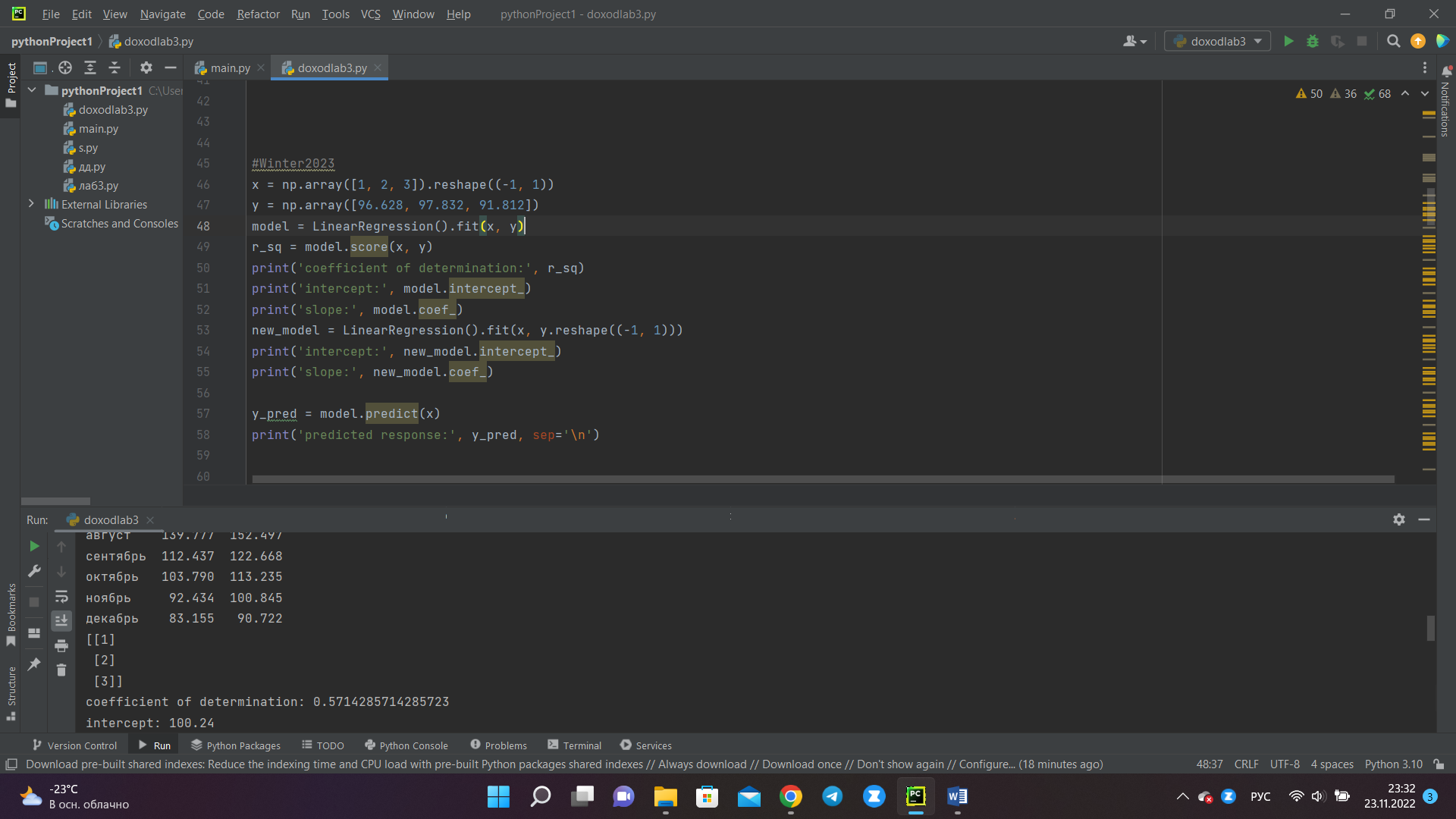


Вторым шагом определяем данные, с которыми предстоит работать. Входы (x) и выход (y) должны быть массивами или похожими объектами:



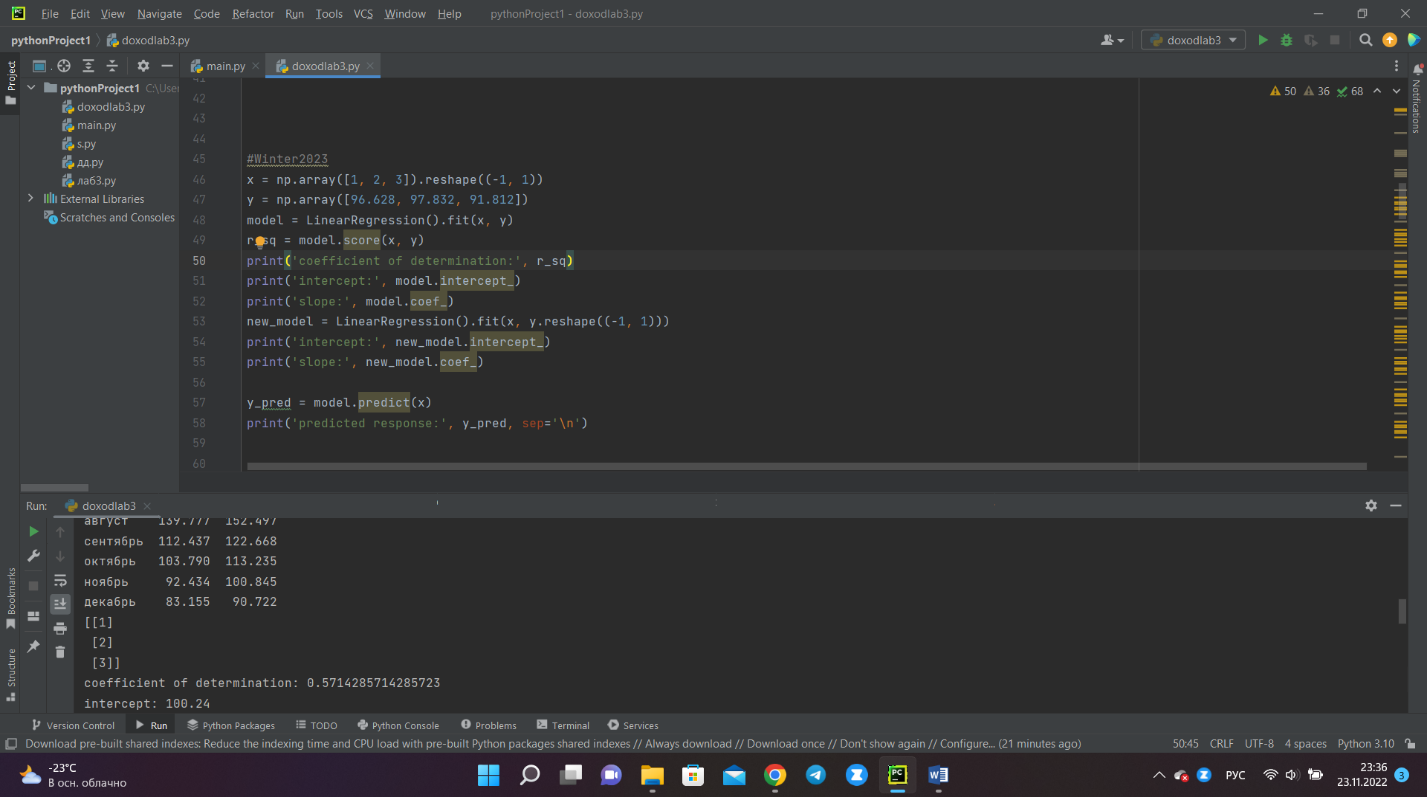
Теперь у нас два массива: вход **x** и выход **y**. Мы вызываем .reshape()на **x**, потому что этот массив должен быть **двумерным**или более точным – иметь **одну колонку и необходимое количество рядов**. Это то, что определяет аргумент **(-1, 1)**

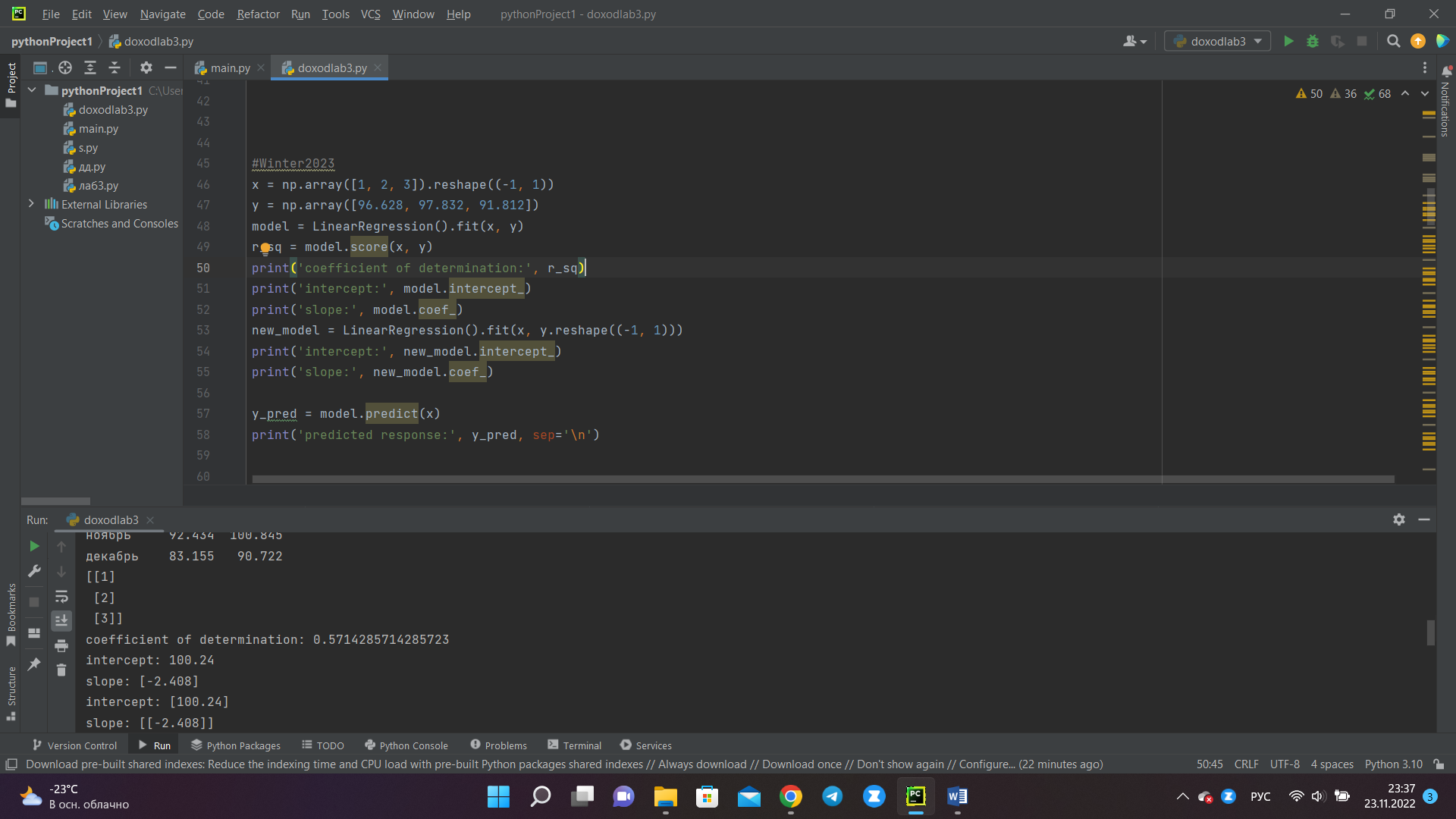
Следующим шагом мы сделаем экземпляр класса LinearRegression, который представит нам модель регрессии:



После совмещения модели нам нужно убедиться в удовлетворительности результатов для интерпретации.

Мы получаем определение (R²) с помощью .score(), вызванной на model:



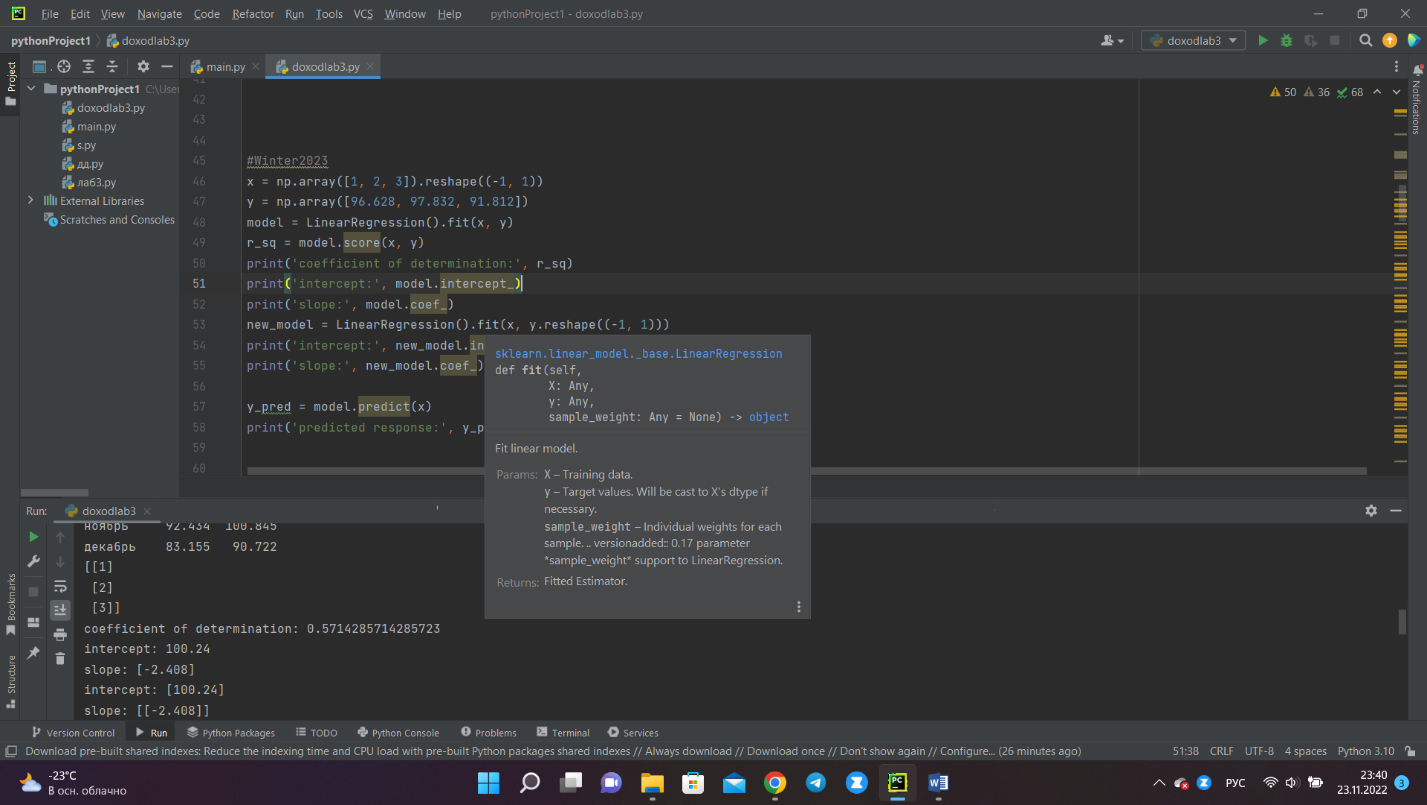


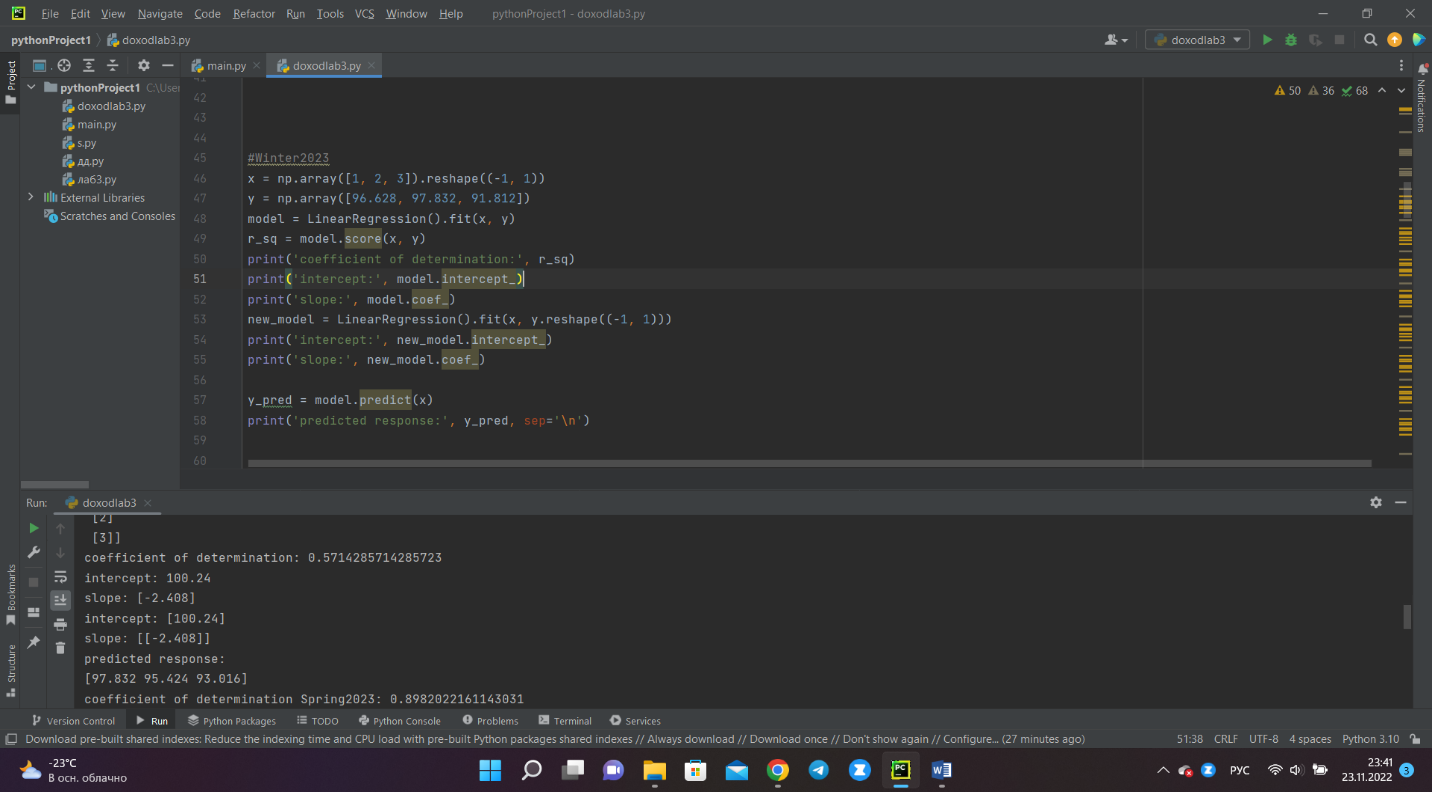
(R²) представляет собой долю дисперсии переменной отклика, которая может быть объяснена переменными-предикторами в модели линейной регрессии.

.score() принимает в качестве аргументов предсказатель **x** и регрессор **y**, и возвращает значение R².

model содержит атрибуты .intercept\_, который представляет собой коэффициент, и b₀ с .coef\_, которые представляют b₁:

Так мы можем получить b₀ и b₁:





Чтобы получить предсказанный ответ, используется функция .predict():

