|  |  |
| --- | --- |
| *voenmeh* | МИНОБРНАУКИ РОССИИ  федеральное государственное бюджетное образовательное учреждение  высшего образования  **«Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова»**  **(БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова»)** |
| БГТУ.СМК-Ф-4.2-К5-02 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Факультет |  | О |  | “Естественнонаучный” |
|  |  | шифр |  | наименование |
| Кафедра |  | О7 |  | “Информационные системы и программная инженерия” |
|  |  | шифр |  | наименование |
| Дисциплина |  | “Системы искусственного интеллекта” | | |

|  |
| --- |
| Практическая работа №1 |
| Введение в Python |
| Вариант 4 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Выполнил студент группы | | | |  |  |
|  | | | | | |
| Фамилия И.О. | | | | | |
| **РУКОВОДИТЕЛЬ** | | | | | |
|  | |  |  | | |
| Фамилия И.О. Подпись | | | | | |
| Оценка |  | | | |  |
| «\_\_\_\_\_» |  | | | | 202 г. |

САНКТ-ПЕТЕРБУРГ

202 г.

**СОДЕРЖАНИЕ**

[Общая постановка задачи 3](#_Toc218973827)

[Часть 1 – подключение датасета 4](#_Toc218973828)

[Часть 2 – Визуализация данных 5](#_Toc218973829)

[Часть 3 – Визуализация дополнительных графиков 7](#_Toc218973830)

[ВЫВОД 10](#_Toc218973831)

# **Общая постановка задачи**

Для выполнения работы необходимо установить Python 3 и Jupyter Notebook.

После этого следует настроить виртуальное окружение и открыть в Jupyter Notebook приложенный файл PR1.ipynb.

Согласно инструкциям в файле открыть датасет, советующий варианту и визуализировать его параметры.

В отчете должны быть приведены полученные графики, к отчету должен быть приложен файл ipynb с результатами.

Вариант 4. Чужая зарплата

– датасет 4 Salary.csv;

– поле для прогнозирования: MonthlyIncome;

– поле для классификации: JobLevel.

# **Часть 1 – подключение датасета**

На рисунке 1 представлено подключение библиотек для загрузки датасета 4 salary.csv и дальнейшей работы с ним.

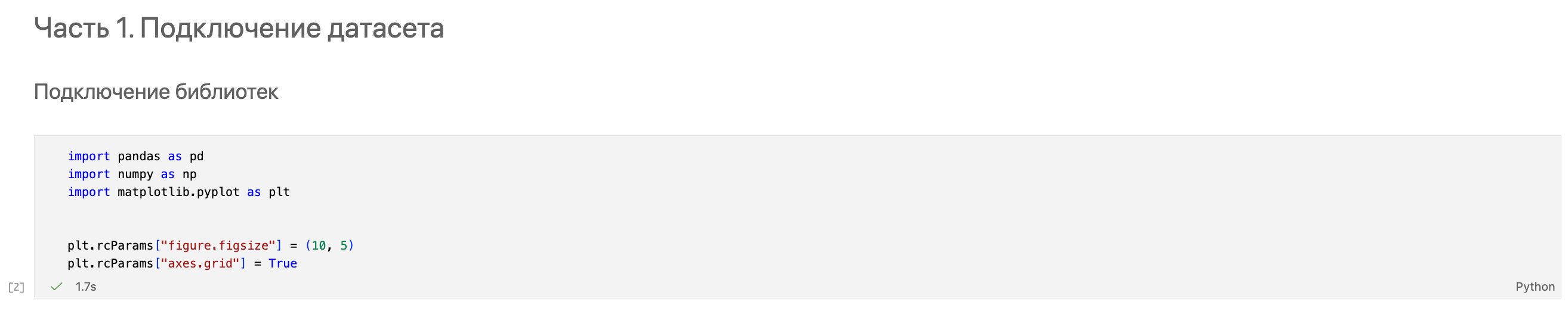


Рисунок 1 – Подключение библиотек

На рисунке 2 изображены загрузка и вывод данных датасета.

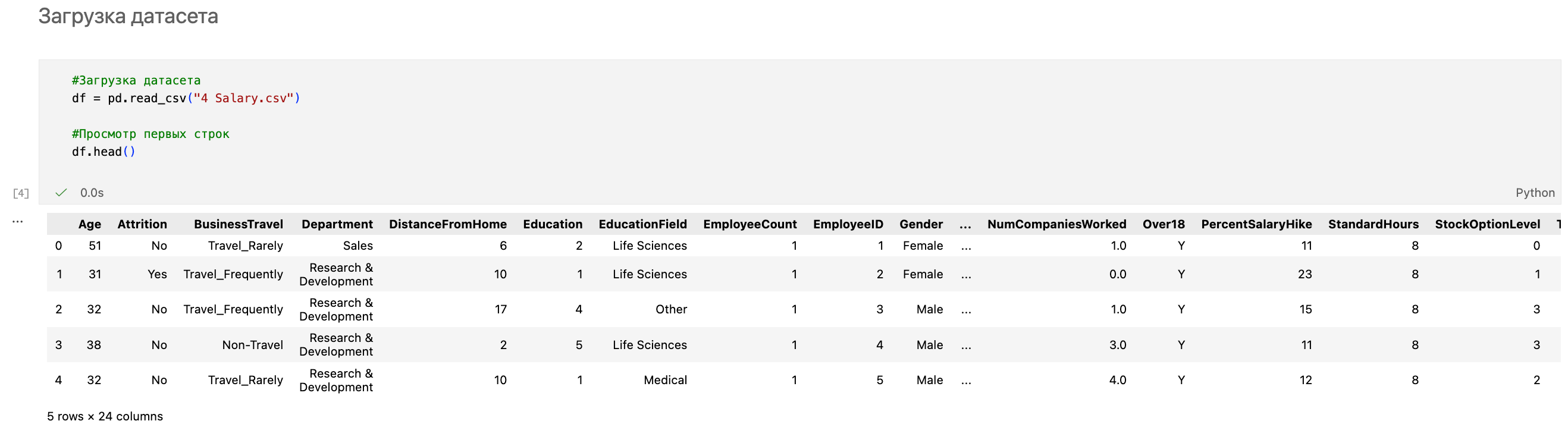


Рисунок 2 – Загрузка и вывод датасета

На этапе предобработки данных была выполнена проверка набора данных на наличие пропусков и корректность типов данных. Были выявлены и обработаны отсутствующие значения, а также приведены типы данных к корректному формату для обеспечения корректного проведения дальнейшего анализа и визуализации результатов, представлено на рисунке 3.

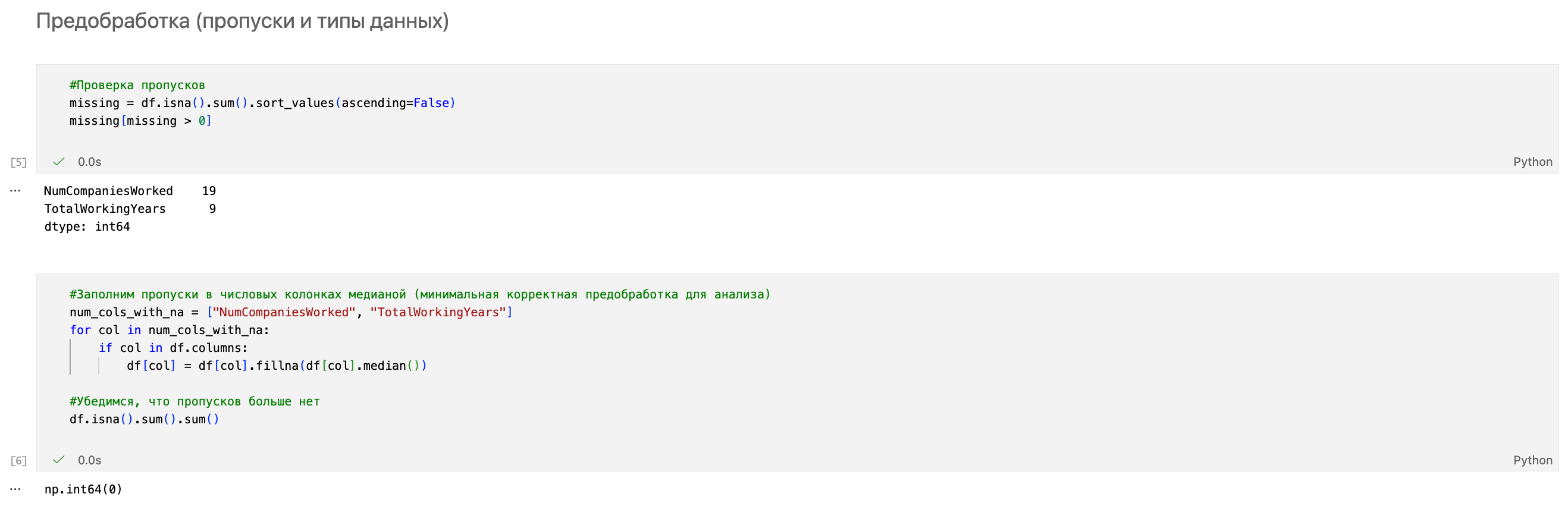


Рисунок 3 – Предобработка данных датасета

На рисунках 4-6 представлена полная информация о датасете.

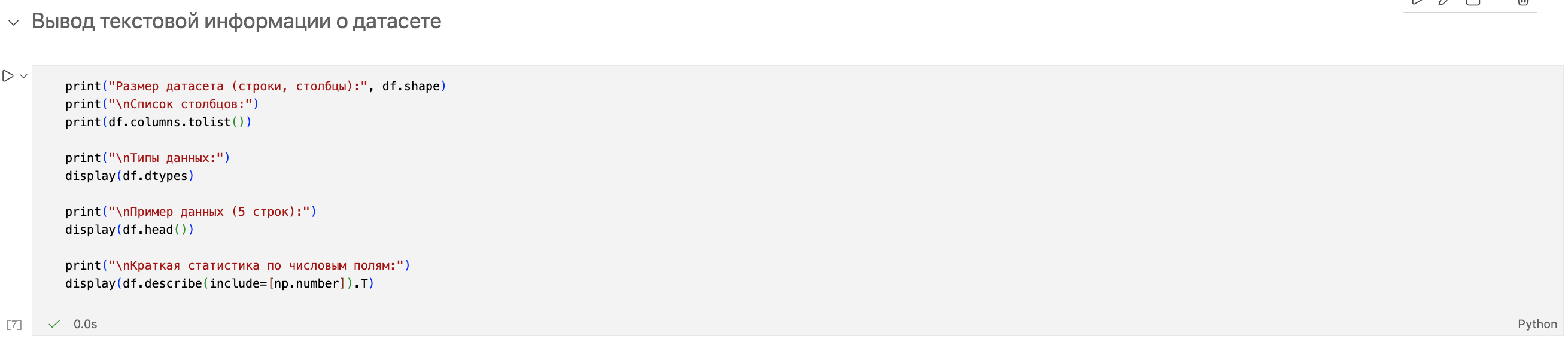


Рисунок 4 – Вывод текстовой информации и описательной статистики датасета



Рисунок 5 – Типы данных в датасете и список столбцов

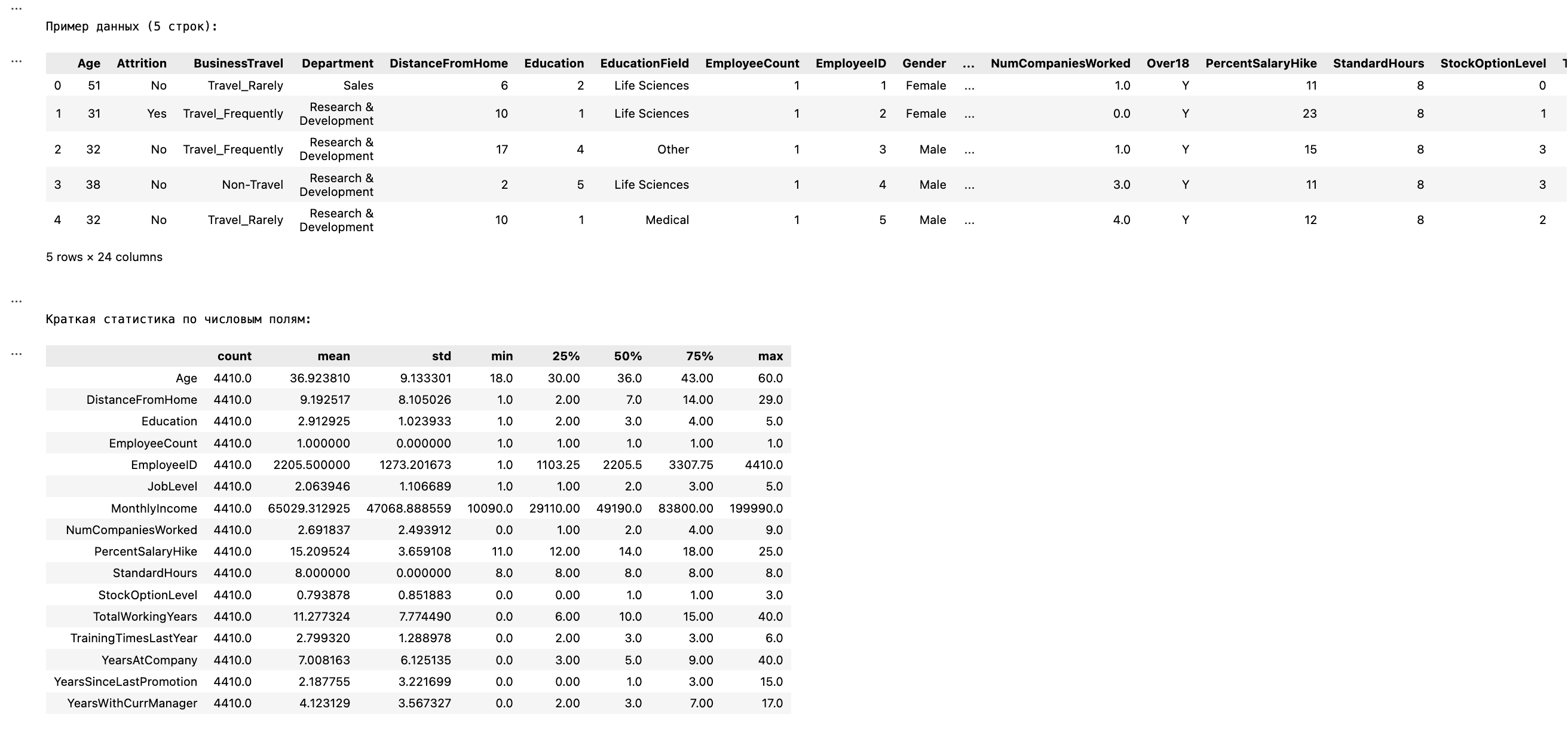


Рисунок 6 – Статистика по полям датасета и набор первых 5 строк

# **Часть 2 – Визуализация данных**

Для анализа целевого признака месячного дохода (MonthlyIncome) был построен линейный график по первым 100 наблюдениям датасета. Визуализация позволяет оценить вариативность доходов сотрудников и выявить возможные аномальные значения, представлено на рисунках 7-8.

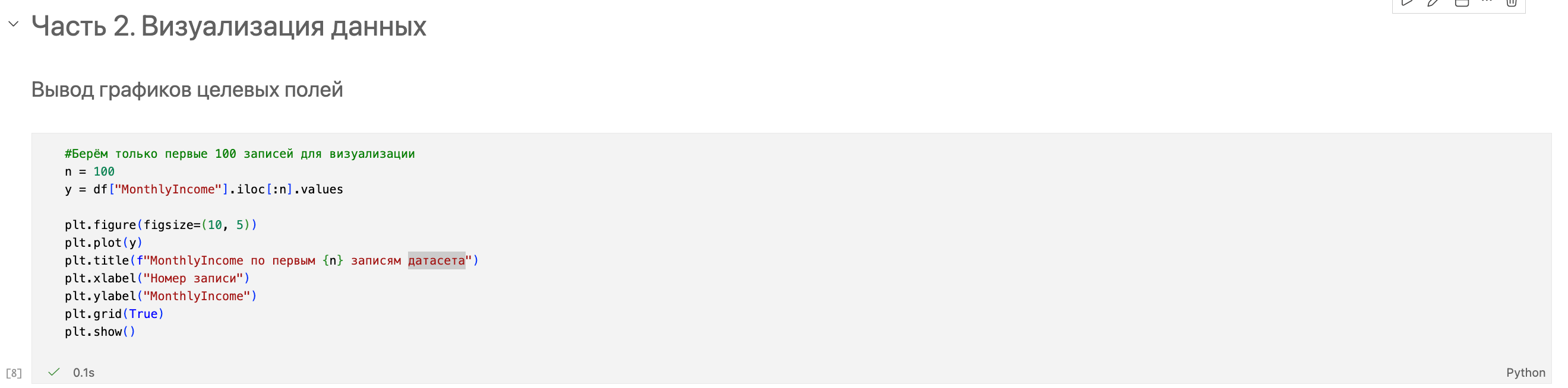


Рисунок 7 – Построение графика



Рисунок 8 – Линейный график значений MonthlyIncome

Для категориального признака должности сотрудника (JobLevel) была построена столбчатая диаграмма распределения классов, отражающая количество сотрудников на каждом уровне должности, представлено на рисунках 9-10.

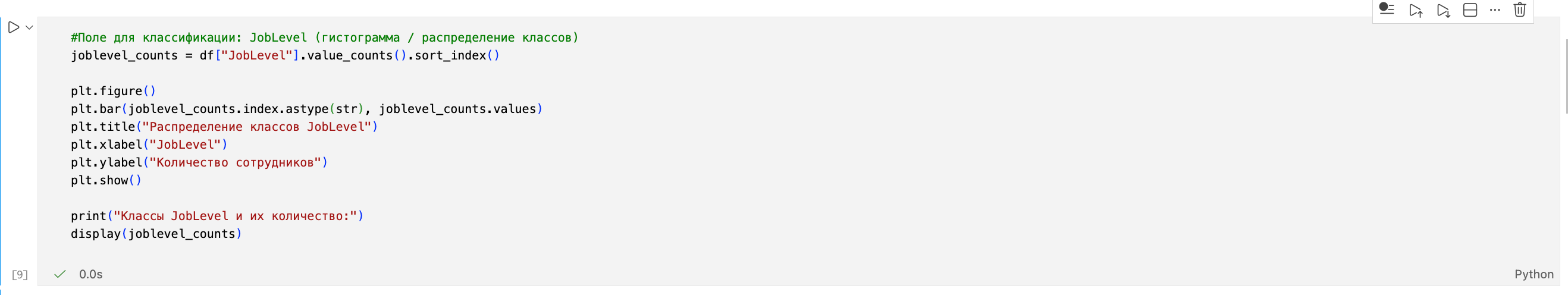


Рисунок 9 – Распределение классов классификационного признака JobLevel

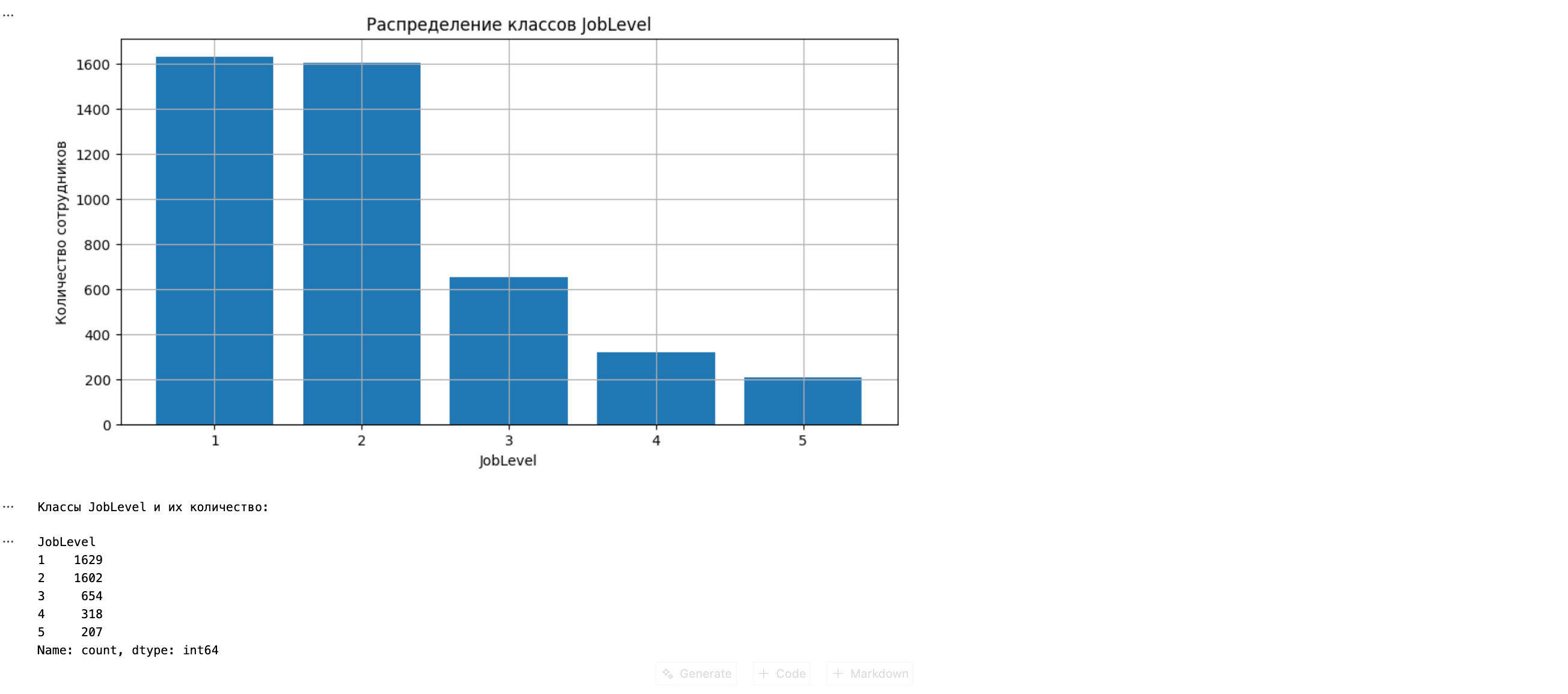


Рисунок 10 – Столбчатая диаграмма распределения классов признака JobLevel

# **Часть 3 – Визуализация дополнительных графиков**

В рамках дополнительного анализа данных были построены диаграммы рассеяния для сравнения числовых признаков с целевой переменной MonthlyIncome. Цветом были выделены классы JobLevel, что позволило визуально оценить различия и взаимосвязи признаков для разных уровней должностей, представлено на рисунке 11.

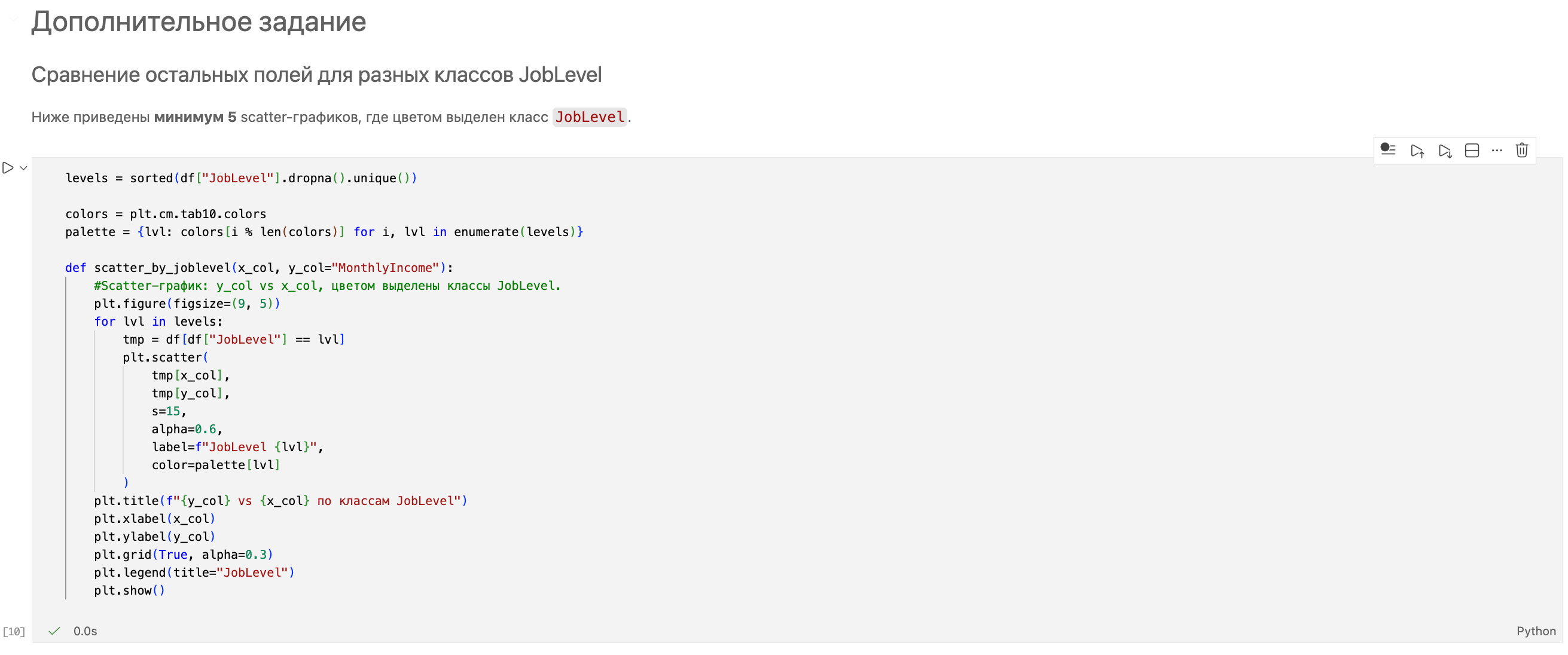


Рисунок 11 – Сравнение признаков по классам JobLevel

Для оценки зависимости уровня дохода от возраста сотрудников была построена диаграмма рассеяния MonthlyIncome от возраста (Age) с цветовым выделением JobLevel. Визуализация позволяет проанализировать, как изменяется уровень заработной платы в различных возрастных группах, а также показать, что влияние возраста на доход носит вторичный характер по сравнению с уровнем должности, представлено на рисунке 12.



Рисунок 12 – Диаграмма рассеяния MonthlyIncome от Age по классам JobLevel

Для анализа влияния общего трудового стажа на уровень дохода сотрудников была построена диаграмма рассеяния MonthlyIncome от общего стажа работника (TotalWorkingYears) с цветовым выделением JobLevel. Визуализация показывает выраженную положительную зависимость между стажем работы и уровнем дохода: с увеличением общего трудового стажа средние значения заработной платы возрастают. При этом внутри каждого уровня должности сохраняется значительный разброс доходов, что указывает на влияние дополнительных факторов, представлено на рисунке 13.

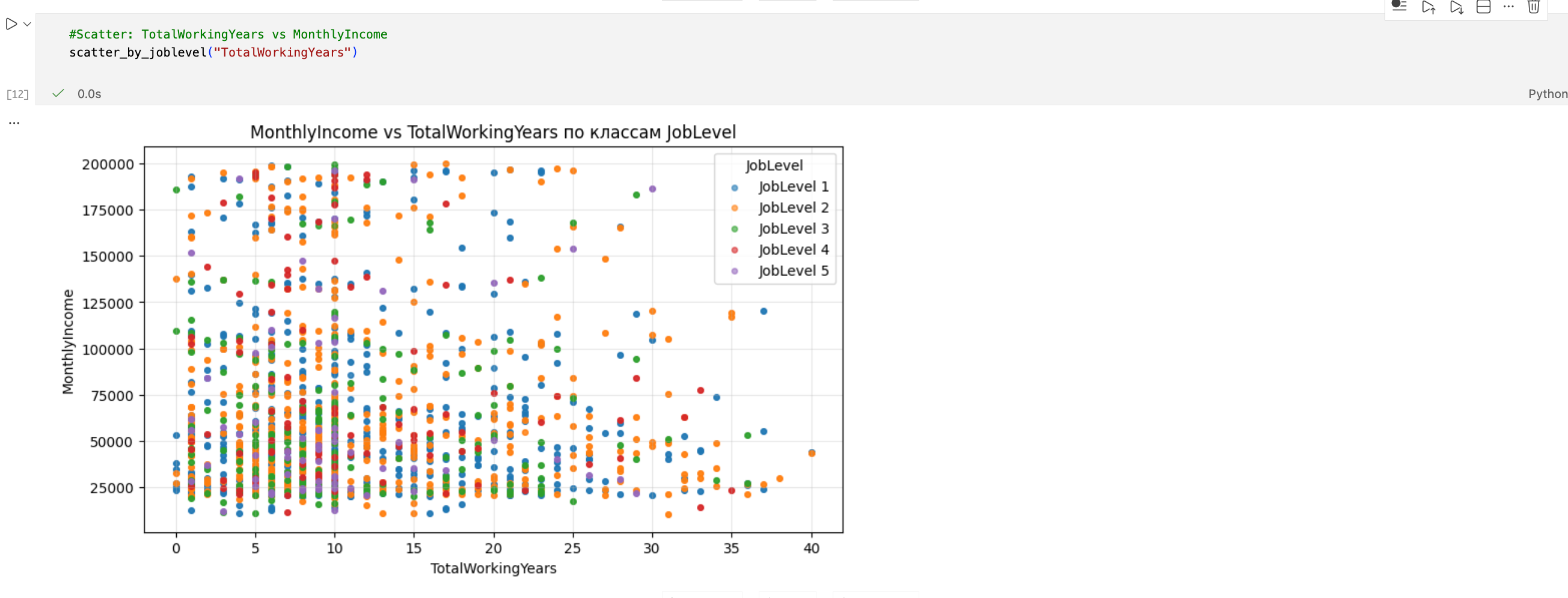


Рисунок 13 – Диаграмма стажа и зарплаты

Для анализа влияния стажа работы в текущей компании на уровень дохода сотрудников была построена диаграмма рассеяния MonthlyIncome от YearsAtCompany с цветовым выделением JobLevel. Визуализация показывает, что с увеличением стажа в компании доход в среднем возрастает, однако зависимость выражена слабее, чем для общего трудового стажа, представлено на рисунке 14.

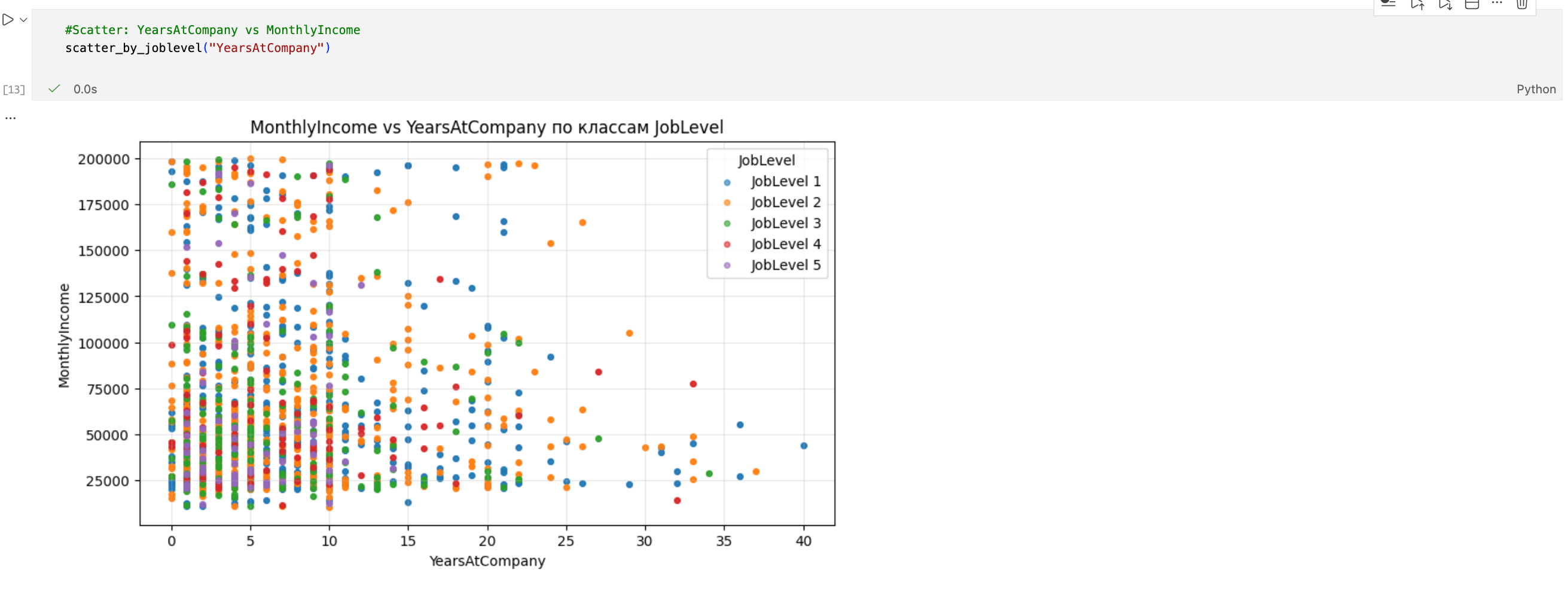


Рисунок 14 – Диаграмма уровня дохода от стажа работы в компании

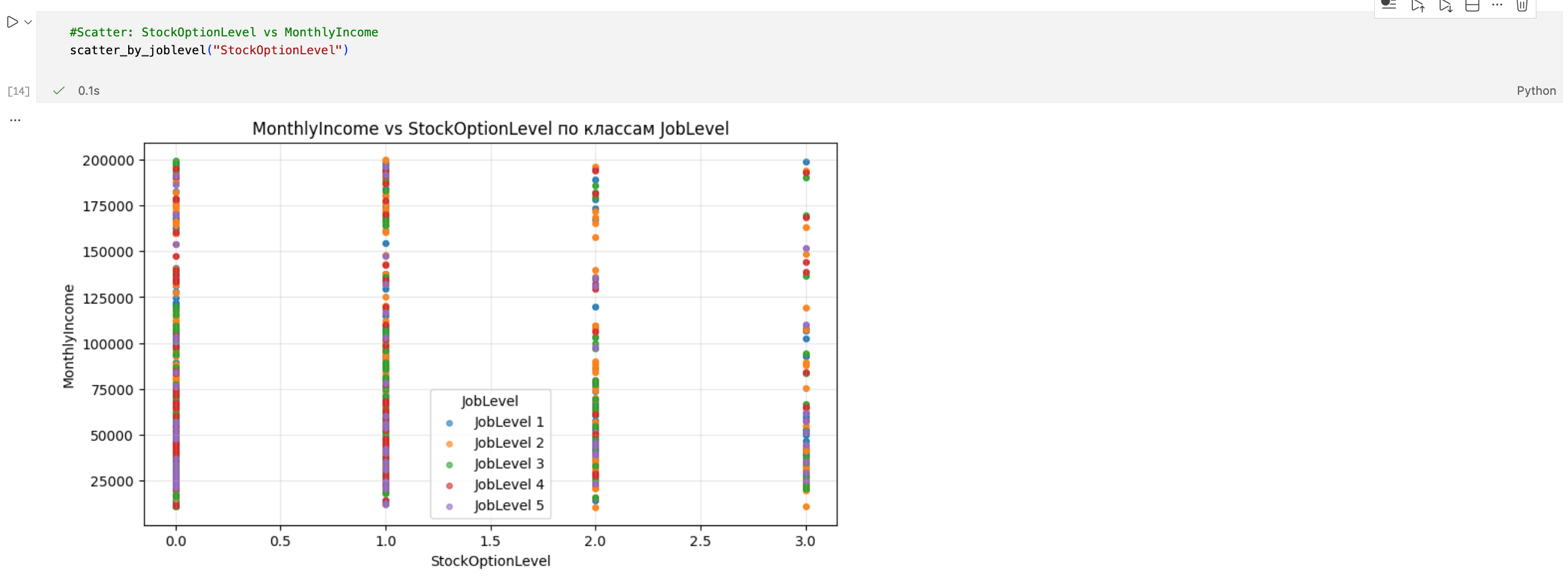
Для анализа зависимости между уровнем опционов на акции и ежемесячным доходом сотрудников была построена диаграмма рассеяния MonthlyIncome от StockOptionLevel. Визуализация позволяет сравнить распределение доходов при различных уровнях опционов и оценить влияние должностного уровня на доход, представлено на рисунке 15.

Рисунок 15 – Графики с аномалиями

Диаграмма рассеяния показывающая зависимость ежемесячного дохода от расстояния от дома до работы, представлено на рисунке 16.

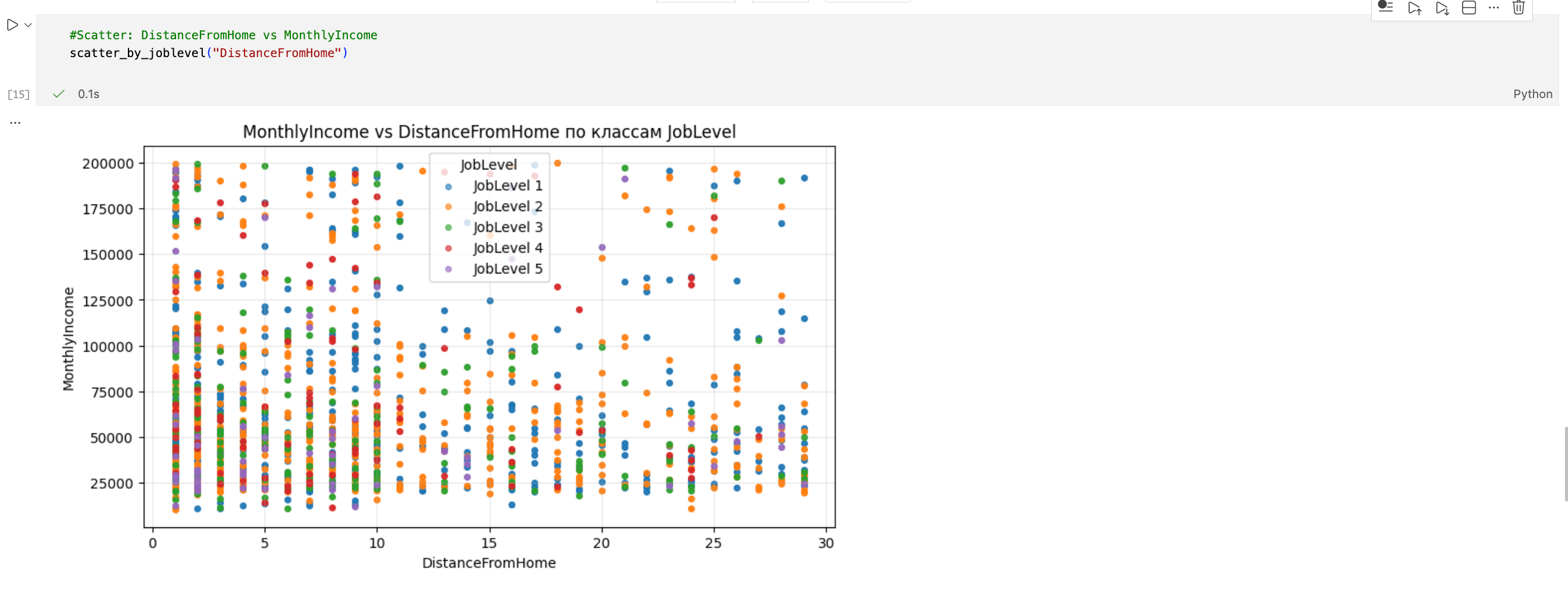


Рисунок 16 – Диаграмма по удаленности от дома

В рамках дополнительного анализа была построена диаграмма рассеяния MonthlyIncome от NumCompaniesWorked. Визуализация показала отсутствие выраженной зависимости уровня дохода от количества ранее сменённых компаний, представлено на рисунке 17

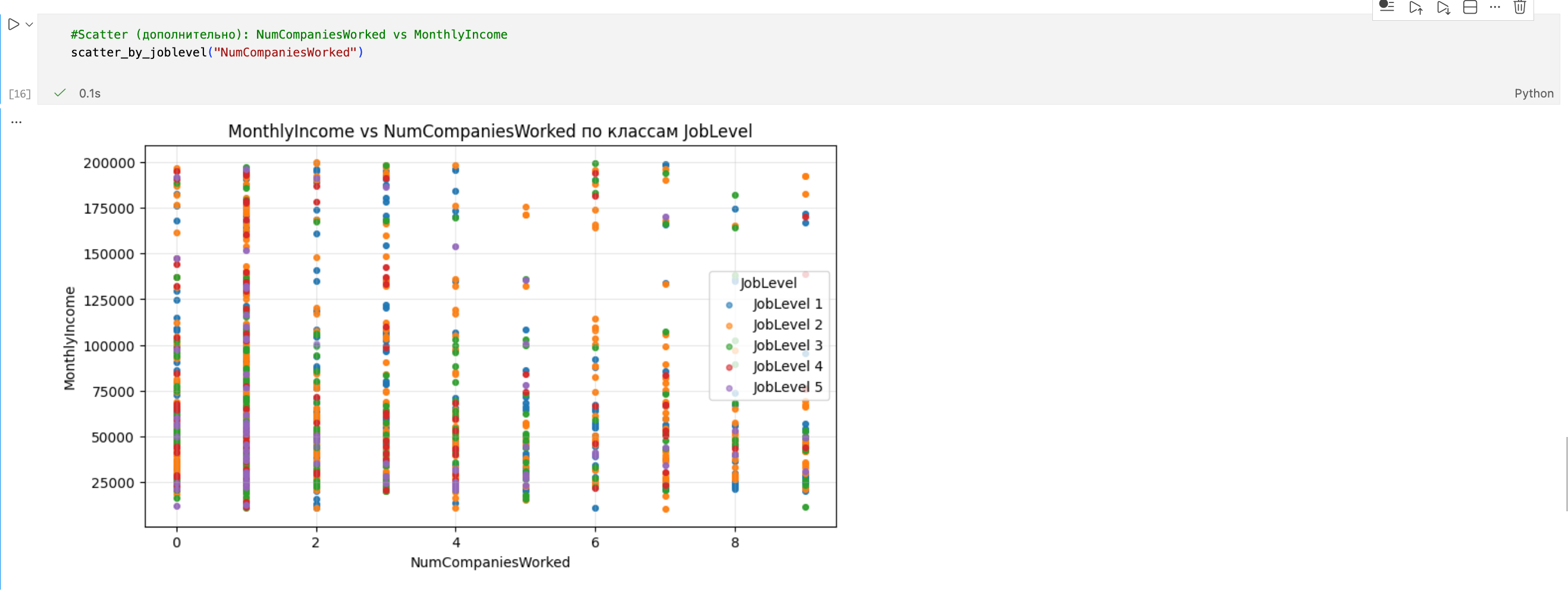


Рисунок 17 – Диаграмма уровня зарплат в зависимости от номера работы

# **ВЫВОД**

В ходе выполнения работы был подключён и проанализирован датасет Salary, содержащий 4410 наблюдений и 24 признака, характеризующих сотрудников компании. В качестве целевой переменной для задачи регрессии был выбран показатель MonthlyIncome, а для задачи классификации — признак JobLevel, отражающий уровень должности сотрудника.

На этапе предобработки данных была выполнена проверка корректности типов данных и наличия пропусков, что обеспечило корректность дальнейшего статистического анализа и визуализации.

В рамках визуального анализа были построены графики распределения целевых признаков. Анализ распределения JobLevel показал, что признак является дискретным и характеризуется выраженным дисбалансом классов: наибольшее количество сотрудников относится к начальным уровням должностей, тогда как высокие уровни представлены существенно реже.

Анализ целевого признака MonthlyIncome выявил значительную вариативность доходов сотрудников, а также наличие выбросов. При этом визуализации (boxplot и scatter-графики) продемонстрировали устойчивую тенденцию роста дохода с увеличением уровня должности, что подтверждает ключевую роль признака JobLevel в формировании заработной платы.

В ходе дополнительного сравнительного анализа были исследованы взаимосвязи между MonthlyIncome и рядом числовых признаков с выделением классов JobLevel. Наиболее выраженная положительная связь с доходом наблюдается у показателей, связанных с профессиональным опытом и стажем работы, таких как TotalWorkingYears и YearsAtCompany. Менее выраженное, но заметное влияние оказывает показатель StockOptionLevel. Признак Age также демонстрирует положительную связь с доходом, однако её сила уступает показателям, напрямую отражающим профессиональный опыт. В то же время признаки, связанные с расстоянием до работы и количеством ранее сменённых компаний, не показали выраженной зависимости с уровнем дохода.