

Анализ влияния возраста, уровня образования и профессионального опыта на зарплаты сотрудников с использованием Google Colaboratory

Ран Максим Александрович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Целью исследования является анализ данных о зарплатах сотрудников с учетом факторов, таких как возраст, уровень образования, опыт работы и должность, с использованием облачного сервиса Google Colaboratory. Анализ позволяет выявить ключевые закономерности, влияющие на уровень зарплаты, и может быть использован для улучшения кадровой политики и управления персоналом.

Ключевые слова: анализ данных, зарплата, Google Colaboratory, визуализация данных.

Анализ влияния возраста, уровня образования и профессионального опыта на зарплаты сотрудников с использованием Google Colaboratory

Ran Maksim Aleksandrovich

Sholom-Aleichem Priamursky State University

Student

Abstract

The purpose of the study is to analyze employee salary data, taking into account factors such as age, education level, work experience, and position, using the Google Colaboratory cloud service. The analysis allows us to identify key patterns affecting the salary level, and can be used to improve HR policy and personnel management.

Keywords: data analysis, salary, Google Colaboratory, data visualization.

1 Введение

1.1 Актуальность

В условиях современного рынка труда анализ данных о зарплатах сотрудников становится важным инструментом для компаний и специалистов в области управления персоналом. Понимание факторов, влияющих на размер заработной платы, таких как возраст, уровень образования, профессиональный опыт и занимаемая должность, позволяет не только формировать справедливую систему оплаты труда, но и привлекать высококвалифицированных сотрудников. Использование современных технологий обработки данных, таких как Google Colaboratory, дает возможность автоматизировать процесс анализа и визуализации данных, делая

его доступным и эффективным. Это особенно важно в условиях возрастающей цифровизации и необходимости быстрого принятия управленческих решений.

1.2 Обзор исследований

Современные подходы к анализу данных активно исследуются и применяются в различных областях. С.А. Агалаков в своей работе рассматривает использование ресурсов Google Colaboratory и языка R в образовательных целях для обучения методам анализа данных, что подчеркивает значимость облачных технологий в современном образовательном процессе [1]. Д.А. Романов провел исследование первичного анализа оттока клиентов с помощью Python, что акцентирует внимание на прикладных аспектах анализа данных и визуализации результатов [2]. А.А. Локтионов в своей работе "Облачные технологии в образовании" освещает преимущества и особенности применения облачных технологий, таких как Google Colaboratory, в образовательной сфере, подчеркивая их доступность и удобство для работы с большими объемами данных [3]. Методология классификации объектов с использованием искусственного интеллекта была изучена А.В. Овсянниковым, что указывает на универсальность и адаптивность современных технологий анализа данных [4]. Т.С. Волокитина в своем исследовании анализировала возможности Google Colaboratory, отметив его удобство для проведения анализа и построения моделей, что делает эту платформу особенно ценной для начинающих и опытных исследователей [5].

1.3 Цель исследования

Целью данного исследования является выявление закономерностей в распределении зарплат сотрудников на основе факторов, таких как возраст, уровень образования, опыт работы и занимаемая должность. Для анализа данных используется облачный сервис Google Colaboratory, который позволяет эффективно обрабатывать и визуализировать данные. Полученные результаты могут быть применены для оптимизации кадровой политики и разработки стратегий управления персоналом.

2 Материалы и методы

Для выполнения исследования был использован датасет, содержащий информацию о сотрудниках: возраст, пол, уровень образования, должность, стаж работы и размер зарплаты. Данные были обработаны и проанализированы с использованием облачного сервиса Google Colaboratory, который предоставляет удобную платформу для работы с большими массивами данных и визуализацией.

3 Результаты и обсуждения

Для анализа данных и визуализации результатов был разработан код, выполняющий несколько последовательных этапов обработки данных: предварительная обработка данных и импорт данных из CSV-файла с помощью библиотеки Pandas (рис. 1).

```
[8] # Импорт необходимых библиотек
import pandas as pd

[9] # Загрузка файла
file_path = "/content/Salary Data.csv" # Укажите путь к файлу
data = pd.read_csv(file_path)
```

Рисунок 1 – Загрузка файла

Также было сделано выполнение базовых операций: отображение первых строк, проверка на наличие пропущенных значений и получение статистики по числовым столбцам. Код проверяет структуру данных и позволяет выявить потенциальные проблемы на начальных этапах анализа (рис. 2, 3).

Выполнения базовых операций, таких как отображение первых строк, проверка наличия пропущенных значений и получение базовой статистики.

```
# Отображение первых 5 строк
print("Первые 5 строк датасета:")
print(data.head())
```

Первые 5 строк датасета:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32	Male	Bachelor's	Software Engineer	5.0	90000
1	28	Female	Master's	Data Analyst	3.0	65000
2	45	Male	PhD	Senior Manager	15.0	150000
3	36	Female	Bachelor's	Sales Associate	7.0	60000
4	52	Male	Master's	Director	20.0	200000

```
[11] # Общая информация о датасете
print("\nОбщая информация о данных:")
print(data.info())
```

Общая информация о данных:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 373 entries, 0 to 372
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Age                   373 non-null   int64
1   Gender                373 non-null   object
2   Education Level       373 non-null   object
3   Job Title             373 non-null   object
4   Years of Experience    373 non-null   float64
5   Salary                373 non-null   int64
dtypes: float64(1), int64(2), object(3)
memory usage: 17.6+ KB
None
```

Рисунок 2 – Выполнения базовых операций

```
[12] # Проверка на пропущенные значения
print("\nКоличество пропущенных значений в каждом столбце:")
print(data.isnull().sum())
```

```
Количество пропущенных значений в каждом столбце:
Age          0
Gender        0
Education Level  0
Job Title     0
Years of Experience  0
Salary        0
dtype: int64
```

```
# Описание числовых данных
print("\nОсновные статистические характеристики числовых данных:")
print(data.describe())
```

```
Основные статистические характеристики числовых данных:
               Age  Years of Experience      Salary
count  373.000000      373.000000      373.000000
mean    37.431635       10.030831  100577.345845
std      7.069073       6.557007   48240.013482
min     23.000000       0.000000    350.000000
25%     31.000000       4.000000   55000.000000
50%     36.000000       9.000000   95000.000000
75%     44.000000      15.000000  140000.000000
max     53.000000      25.000000  250000.000000
```

Рисунок 3 – Выполнения базовых операций

На следующем этапе выполняется группировка данных по возрасту сотрудников. Для каждого возраста вычисляется среднее значение заработной платы. Это позволяет оценить, как возраст влияет на уровень дохода. Результаты визуализируются с помощью линейного графика, где ось X представляет возраст, а ось Y — среднюю зарплату. `data.groupby('Age')` группирует строки датасета по уникальным значениям в столбце Age (возраст). `['Salary'].mean()` для каждой группы вычисляет среднее значение столбца Salary. В результате получается новый объект, содержащий возраст как индекс и соответствующую ему среднюю зарплату. Далее создается визуализация (рис. 4).

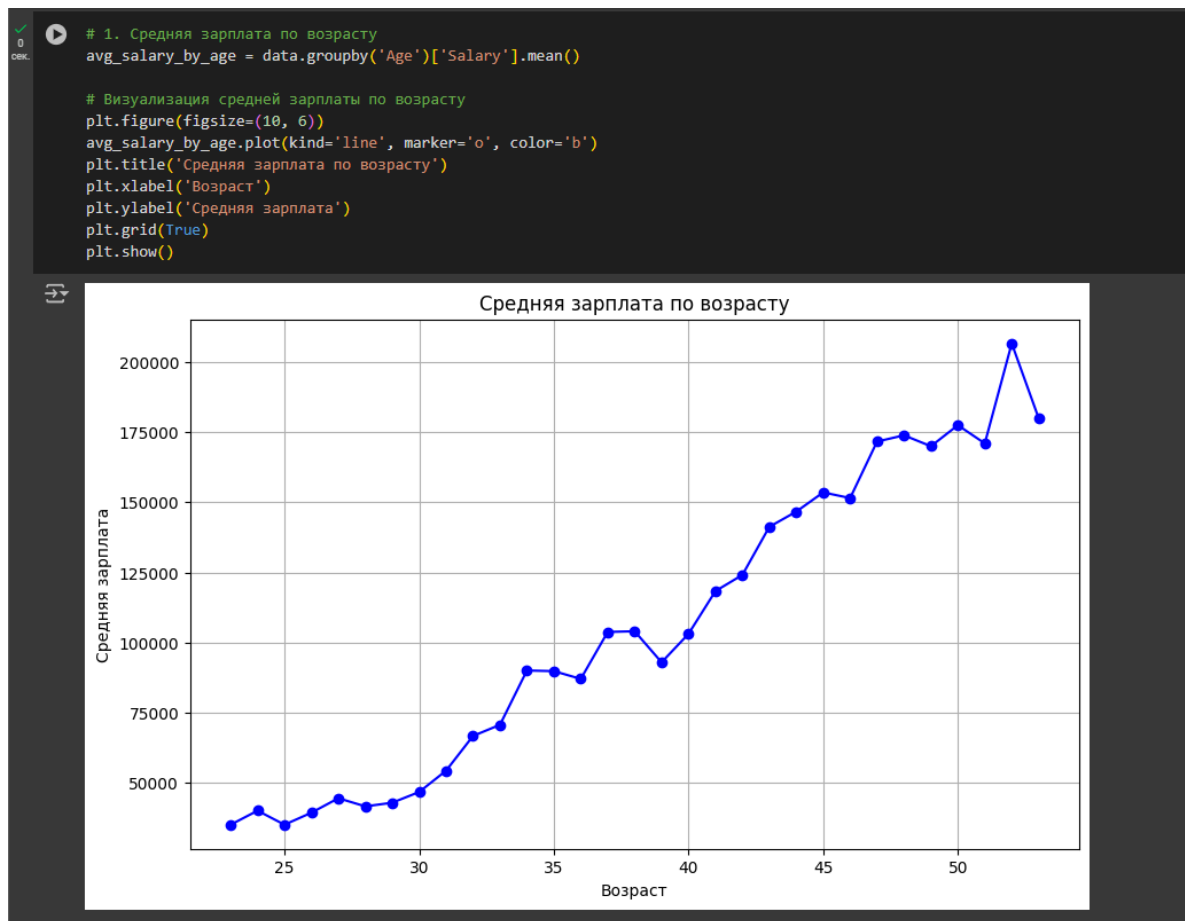


Рисунок 4 – Визуализация средней зарплаты по возрасту

Линейный график показал, что с возрастом заработная плата увеличивается, достигая максимума у старших возрастных категорий. Это свидетельствует о влиянии опыта работы и накопленных знаний на уровень дохода.

Далее анализируется связь между уровнем образования сотрудников и их зарплатами. Группировка данных проводится по уровню образования, а результаты визуализируются в виде столбчатой диаграммы (рис. 5).

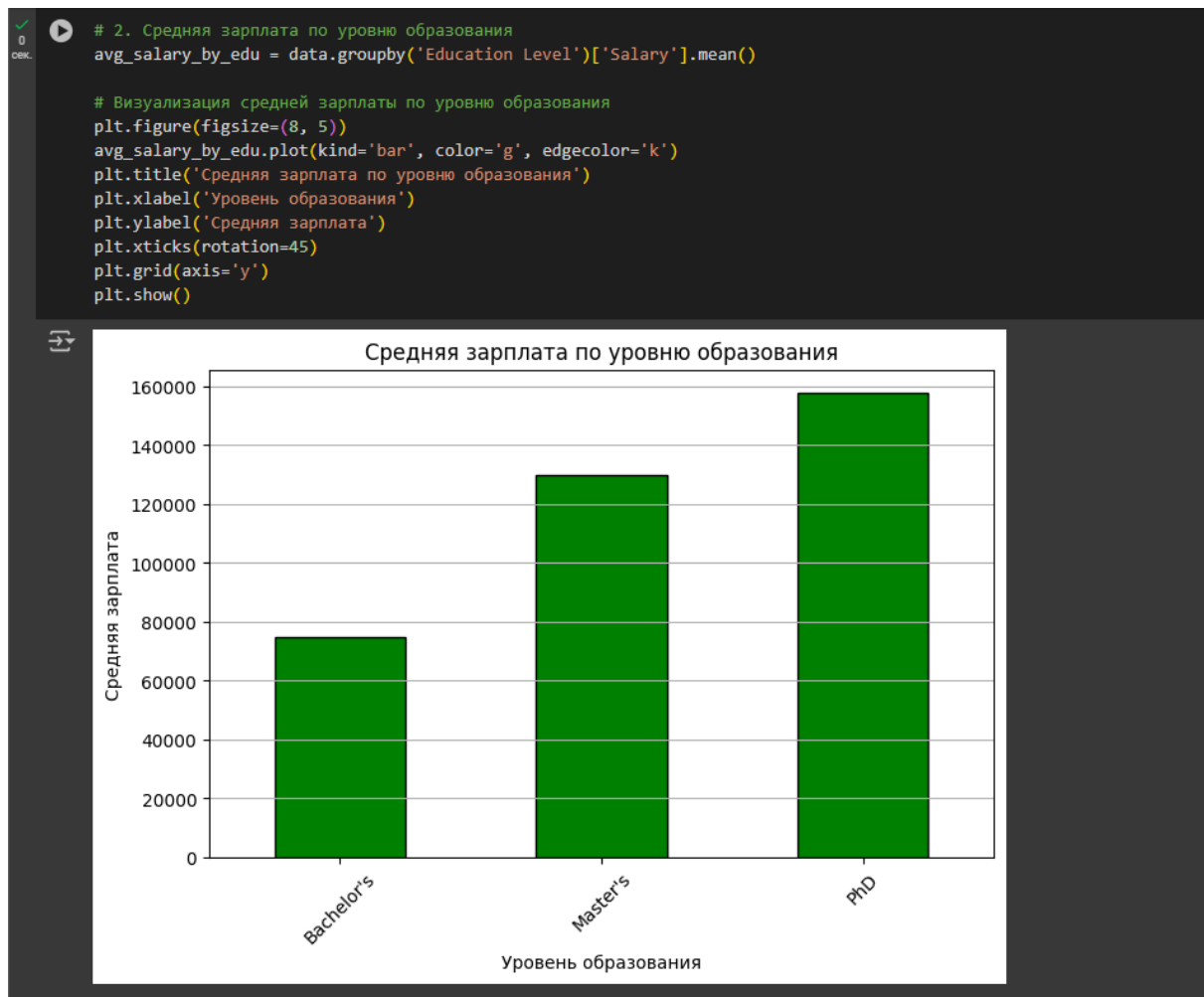


Рисунок 5 – Средняя зарплата по уровню образования

Столбчатая диаграмма показала, что сотрудники с высшим уровнем образования (например, PhD) зарабатывают значительно больше. Это подтверждает гипотезу о влиянии квалификации на заработную плату.

Следующий этап исследования направлен на определение наиболее высокооплачиваемых должностей. Для анализа данные группируются по столбцу Job Title (должность), после чего для каждой должности вычисляется средняя заработная плата. Затем результаты сортируются по убыванию, и на графике отображаются только 10 должностей с наивысшими зарплатами (рис. 6).

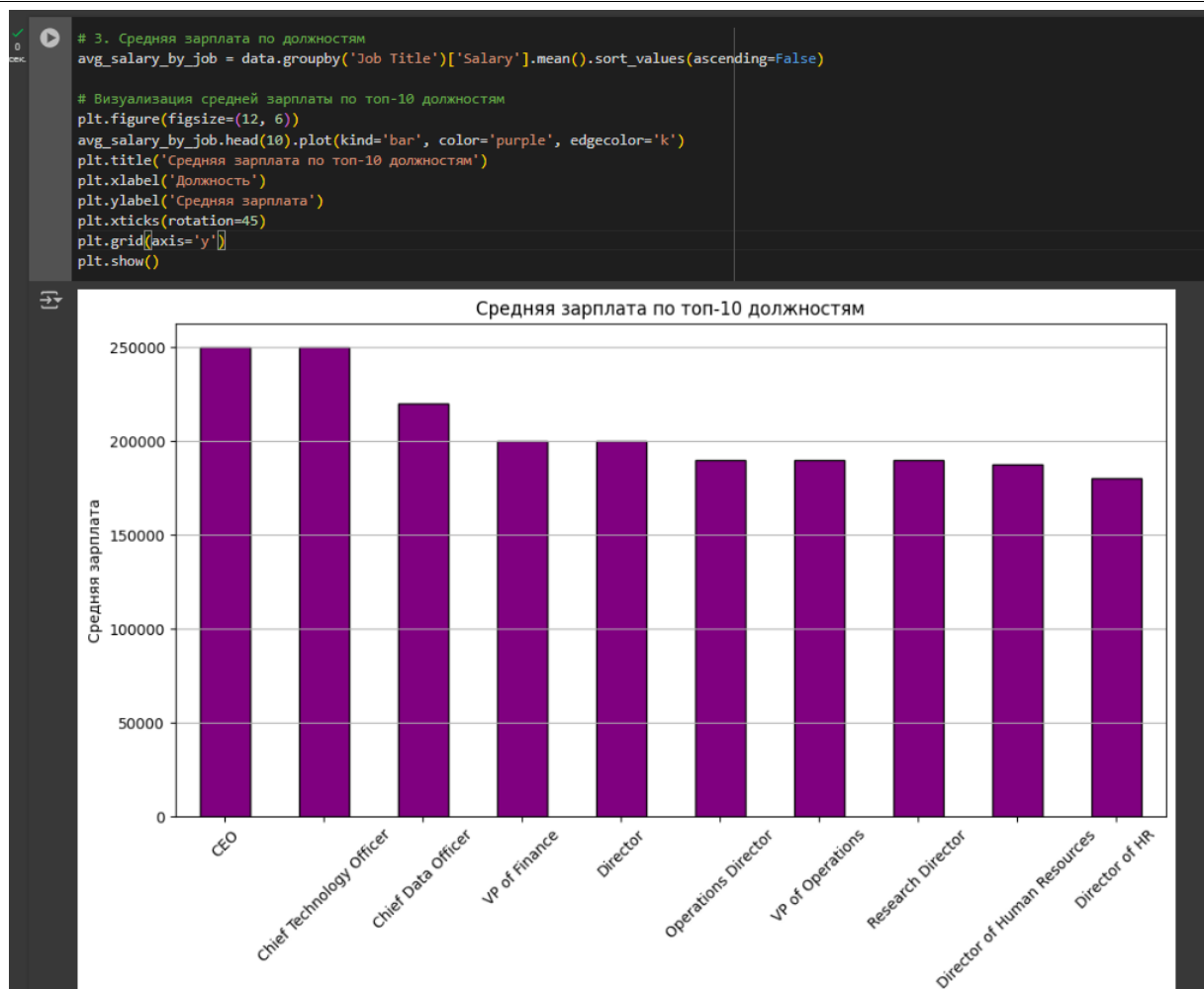


Рисунок 5 – Средняя зарплата по топ-10 должностям

График наглядно демонстрирует, что управленческие должности, такие как "Director" и "Senior Manager", входят в число самых высокооплачиваемых. Значительно ниже находятся профессии, связанные с технической поддержкой или административными обязанностями. Этот анализ подтверждает, что уровень зарплаты напрямую связан с ответственностью и ролью сотрудника в компании, что соответствует общепринятым тенденциям в управлении персоналом.

4 Выводы

Анализ данных выявил, что заработная плата сотрудников зависит от возраста, уровня образования и занимаемой должности. Старшие возрастные группы и сотрудники с высоким уровнем образования получают больше, а наибольшие зарплаты наблюдаются у управленческих позиций. Использование Google Colaboratory подтвердило свою эффективность в обработке данных и визуализации результатов, что делает этот инструмент полезным для анализа данных и разработки кадровых стратегий.

Библиографический список

1. Агалаков С.А. О дистанционном обучении методам анализа данных с использованием ресурсов интернет: Google Colab и язык R // В сб.: Методика преподавания математических и естественно-научных дисциплин: современные проблемы и тенденции развития. Материалы VIII Всероссийской научно-практической конференции. Отв. редактор Ю.В. Коваленко. Омск, 2021. С. 214-218.
2. Романов Д.А. Первичный анализ оттока клиентов с помощью Python // Постулат. 2023. № 1 (87).
3. Локтионов А.А. Облачные технологии в образовании // В сб.: Наука, образование, транспорт: актуальные вопросы, приоритеты, векторы взаимодействия. Материалы II Международной научно-методической конференции. В 3-х частях. Оренбург, 2023. С. 213-216.
4. Овсянников А.В. Методология классификации пород деревьев бореальных лесов с использованием архитектуры EfficientNet искусственного интеллекта.
5. Волокитина Т.С. Анализ возможностей Google Colab // Современные научные исследования и инновации. 2020. № 12 (116). С. 1.