Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Южно-Уральский государственный университет»
(национальный исследовательский университет)
Высшая школа экономики и управления
Кафедра «Цифровая экономика и информационные технологии»

ОТЧЕТ
по учебной практике
ЮУрГУ–09.03.02.2025

сроки прохождения практики: с 30.06.2025 по 27.07.2025

Руководитель практики, ст. преподаватель
_____С. С. Аверьянова
_____2025 г.

Автор отчета,
студент группы ЭУ-140
_____Н. С. Поспелов
_____2025 г.

Нормоконтролер, ст. преподаватель
_____С. С. Аверьянова
_____2025 г.

Отчет защищен с оценкой
_____
_____2025 г.

Челябинск 2025

«Южно-Уральский государственный университет»
(национальный исследовательский университет)
Высшая школа экономики и управления
Кафедра «Цифровая экономика и информационные технологии»

ЗАДАНИЕ
на семестровую работу студента
Поспелова Никиты Сергеевича

Группа ЭУ–140

Тема: Прогнозирование и анализ зарплатных ожиданий в различных сферах на основе большого набора входных данных

Цели работы:

Изучение возможностей современного инструментария анализа данных на языке Python с использованием таких библиотек, как pandas, matplotlib, seaborn.

Применение методов предобработки и исследовательского анализа данных для решения поставленной бизнес-задачи.

Определить ключевые факторы, влияющие на уровень заработной платы, выявить зависимости между характеристиками вакансий и зарплатными предложениями, а также сформировать аналитические рекомендации для HR-специалистов и соискателей.

Содержание работы:

Постановка бизнес-задачи и описание предметной области Современный рынок труда характеризуется высокой конкуренцией как среди кандидатов, так и среди работодателей. Возникает необходимость объективно оценивать уровень заработной платы в зависимости от компании, профессии, региона, требуемых навыков и условий работы. Неправильная оценка зарплатных ожиданий приводит к снижению эффективности найма, увеличению сроков закрытия вакансий и потере конкурентоспособности работодателя.

Предобработка данных Первичный обзор датафрейма; проверка и корректировка названий столбцов; поиск и устранение пропусков; изменение типов данных при необходимости; выявление и удаление дубликатов; формулировка промежуточных выводов.

Исследовательский анализ данных (EDA) Выполнение индексации и логической выборки (не менее 5 примеров каждого вида); сортировка данных и

формулировка выводов; фильтрация с помощью методов query и where (не менее 5 примеров каждого вида); построение не менее 3 сводных таблиц и групповых агрегатов; выявление ключевых закономерностей; промежуточные выводы.

Графический анализ данных Построение не менее 3 различных диаграмм с использованием Matplotlib и не менее 2 диаграмм с использованием Seaborn; настройка подписей, осей, сетки, легенд.

Выводы и рекомендации Краткое резюме проведённого анализа; ответ на поставленную бизнес-задачу; формулировка 2–3 практических рекомендаций для бизнеса; определение направлений дальнейшей работы.

Разработка мини-приложения на Python (tkinter) Создание графического интерфейса (GUI) для загрузки очищенного CSV-файла, отображения структуры данных, выполнения базовой фильтрации и генерации текстовых рекомендаций на основе проведённого анализа.

Задание выдал: Ст. преподаватель каф. ЦЭиИТ С.С. Аверьянова

Задание получил: Студент группы ЭУ-140 Н.С. Поспелов

ОГЛАВЛЕНИЕ

Оглавление

ВВЕДЕНИЕ

Современный рынок труда характеризуется высокой динамичностью и высокой конкуренцией как среди соискателей, так и среди работодателей. Компании стремятся привлечь квалифицированных специалистов, а кандидаты — получить справедливый уровень заработной платы в зависимости от опыта, навыков, отрасли и региона. В этих условиях объективный анализ зарплатных ожиданий становится важным инструментом для стратегического планирования, оптимизации HR-процессов и повышения эффективности найма.

Одной из ключевых проблем рынка труда является отсутствие прозрачности в вопросах ценообразования труда. Неверная оценка уровня зарплат приводит к увеличению сроков подбора персонала, снижению привлекательности вакансий и уменьшению конкурентоспособности компаний. Для соискателей это выражается в сложностях планирования карьеры и выборе направлений профессионального развития.

В рамках данной практики проводится анализ большого набора данных, содержащего информацию о вакансиях, компаниях, требованиях к кандидатам и диапазонах заработных плат. Используемый датасет основан на данных с платформы Kaggle и позволяет исследовать факторы, влияющие на уровень оплаты труда, а также формировать аналитические рекомендации для работодателей и соискателей.

Цель практики: Освоение методов анализа данных на языке Python и закрепление навыков предобработки, исследовательского и графического анализа, а также построение простого аналитического приложения для визуализации результатов на примере датасета о зарплатах и вакансиях.

Задачи практики:

1. Изучить структуру и содержание датасета, определить его потенциал для решения бизнес-задач рынка труда.
2. Выполнить предварительную обработку данных, включающую:
   2.1. удаление дубликатов и проверка на пропуски;
   2.2. нормализация названий столбцов;
   2.3. приведение типов данных к оптимальным форматам.

3. Провести исследовательский анализ данных (EDA), включающий:
   3.1. индексацию, фильтрацию и сортировку записей по ключевым критериям;
   3.2. построение сводных таблиц и группировок;
   3.3. расчёт дополнительных метрик (минимальная, максимальная и средняя зарплаты, возраст компании, наличие навыков и др.).
4. Выполнить графический анализ с использованием библиотек Matplotlib и Seaborn для наглядного представления закономерностей и взаимосвязей в данных.
5. Сформулировать выводы и рекомендации для работодателей и соискателей на основе выявленных закономерностей.
6. Разработать мини-приложение на Python с использованием tkinter, позволяющее:
   6.1. загружать очищенный CSV-файл;
   6.2. отображать структуру данных;
   6.3. выполнять фильтрацию по ключевым параметрам;
   6.4. формировать текстовые рекомендации на основе проведённого анализа.

Ожидаемые результаты

в ходе выполнения практики планируется:

1. Получить очищенный и структурированный датасет без пропусков и дубликатов, с корректными типами данных.
2. Построить аналитические таблицы, группировки и визуализации, отражающие ключевые закономерности в формировании уровня заработной платы.
3. Сформулировать выводы и подготовить набор практических рекомендаций для работодателей и соискателей.
4. Продемонстрировать навыки применения современных инструментов анализа данных Python в реальной предметной области рынка труда.

ГЛАВА 1 АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

1.1 ПОДБОР ДАТАСЕТА

Для исследования выбран датасет, основанный на материалах платформы Kaggle, содержащий данные о вакансиях, уровне заработной платы, характеристиках компаний и требованиях к кандидатам. Используемый файл salary_data_cleaned.csv включает не менее 1000 записей и 28 столбцов, описывающих разнообразные параметры рынка труда.

Датасет содержит как числовую информацию (минимальная, максимальная и средняя зарплаты, возраст компании), так и категориальные признаки (должность, местоположение, тип собственности компании, отрасль, наличие навыков Python, AWS, Spark и др.). Такой набор данных позволяет выполнить всесторонний анализ факторов, влияющих на зарплатные ожидания, и выявить ключевые закономерности.

1.2 ПРЕДМЕТНАЯ ОБЛАСТЬ

Рынок труда в сфере технологий и аналитики развивается крайне стремительно. Спрос на квалифицированных специалистов растёт, при этом зарплаты могут значительно различаться между штатами, компаниями и профессиями. Компании стремятся предлагать конкурентные условия, чтобы привлекать ключевые кадры, а соискателям важно ориентироваться в реальных уровнях компенсаций, чтобы адекватно оценивать свою ценность на рынке.

1.3 КОНТЕКСТ БИЗНЕСА

Анализ данных о вакансиях и зарплатных предложениях имеет большое значение для:

Работодателей

- корректирования зарплатных вилок в зависимости от региона и профессии;
- оптимизации требований к кандидатам;
- повышения конкурентоспособности вакансий.

Соискателей

- определения востребованных навыков;
- понимания, какие компетенции повышают стоимость специалиста;
- ориентирования в уровне зарплат по должностям и регионам.

HR-аналитиков

- определения ключевых факторов, влияющих на зарплаты;
- анализа структуры рынка труда;
- прогнозирования тенденций развития профессий.

Таким образом, исследование датасета позволяет решать реальные бизнес-задачи и помогает участникам рынка принимать более обоснованные решения.

1.4 БИЗНЕС-ЗАДАЧА

На основе данных о вакансиях необходимо:

1. Определить факторы, влияющие на размер заработной платы.
   (регион, профессия, рейтинг компании, её возраст, требования к навыкам)
2. Провести сравнение зарплат по должностям, компаниям и регионам.
3. Выявить влияние конкретных навыков (Python, AWS, Spark, Excel) на уровень оплаты.
4. Сформировать аналитические рекомендации для работодателей и соискателей.
5. Создать мини-приложение, позволяющее:
   - загружать датасет,
   - выполнять фильтрацию,
   - анализировать параметры вакансий,
   - получать рекомендации.

1.5 ОСНОВНЫЕ ПРОБЛЕМЫ И ВЫЗОВЫ

При работе с подобными данными возникает ряд проблем:

- Нестандартизированные форматы подачи информации в оригинальных вакансиях. (разные работодатели указывают данные по-разному)
- Неполнота данных - некоторые вакансии не содержат сведения о навыках или полной зарплатной информации.
- Шум в текстовых данных - названия должностей часто содержат дубли, сокращения или вариации формата.
- Влияние региональных факторов - зарплаты значительно отличаются в зависимости от штата или города.
- Разнородность компаний - различия по размеру, возрасту, формам собственности усложняют сравнение.

1.6 ПРИЧИНЫ АКТУАЛЬНЫХ ПРОБЛЕМ

- Отсутствие единой стандартизации описания вакансий при публикации.
- Субъективность требований работодателей и формулировки обязанностей.
- Различия в рыночной стоимости навыков по регионам.
- Неполная и неконсистентная информация в исходных данных.

Эти факторы требуют предобработки, нормализации и очистки данных перед проведением анализа.

1.7 СТРУКТУРА ДАННЫХ

№	Название столбца	Тип данных	Описание
1	job_title	object	Название должности
2	salary_estimate	object	Исходная оценка зарплаты
3	rating	float64	Рейтинг компании
4	company_name	object	Название компании
5	location	object	Местоположение
6	size	object	Размер компании
7	founded	int64	Год основания
8	type_of_ownership	object	Тип собственности
9	industry	object	Отрасль
10	sector	object	Сектор экономики
11	revenue	object	Годовая выручка
12	competitors	object	Конкуренты
13	min_salary	int64	Минимальная зарплата
14	max_salary	int64	Максимальная зарплата
15	avg_salary	int64	Средняя зарплата
16	job_state	object	Штат, где открыта вакансия
17	age	int64	Возраст компании
18	python_yn	int64	Требуется ли навык Python
19	aws	int64	Требуется AWS
20	spark	int64	Требуется Spark
21	excel	int64	Требуется Excel

1.8 ВЫВОДЫ ГЛАВЫ

В используемом датасете содержится более тысячи записей и около тридцати столбцов, включая числовые, категориальные и бинарные признаки, связанные с навыками. Анализ структуры данных показывает, что в наборе содержится полный набор факторов, необходимых для анализа рынка труда: должности, компании, зарплаты, отрасли, навыки, региональные особенности.

Данные позволяют выделить ключевые параметры, влияющие на зарплату, такие как:

- профессия и уровень позиции;

- местоположение вакансии;
- набор требуемых навыков;
- рейтинг и возраст компании;
- отрасль и тип собственности.

Таким образом, датасет предоставляет широкие возможности для исследования рынка труда, построения аналитических выводов и разработки рекомендаций для участников рынка.

ГЛАВА 2 РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

2.1 ПРЕДОБРАБОТКА ДАННЫХ

2.1.1 Импорт библиотек

Цель. Подключить инструменты для табличной работы, численных расчётов, визуализации и будущего GUI.

```
# Импорт библиотек
import tkinter
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Комментарий.

pandas — основа обработки таблиц;

numpy — численные операции;

matplotlib/seaborn — графики;

tkinter — построение графического интерфейса приложения.

2.1.2 Загрузка датасета

Цель. Считать исходный CSV в DataFrame.

```
# Загружаем датасет
dataframe = pd.read_csv('salary_data_cleaned.csv')
```

Комментарий. Создаётся рабочий DataFrame. Файл содержит более 1000 строк и около 28 признаков, описывающих рынок труда: должность, компания, регион, навыки, зарплаты и т.д.

2.1.3 Первичный осмотр

Цель. Быстро убедиться, что прочитан именно тот файл, и структура/значения выглядят корректно.

```
# Первичный осмотр
print(dataframe.head())
 # первые 5 строк
print(dataframe.tail())
```

```
# последние 5 строк
print(dataframe.sample(5))

    # случайная выборка из 5 строк
```

Комментарий. Создаётся рабочий DataFrame. На этом этапе выявляются очевидные артефакты: странные значения зарплат, пустые должности, шум в названиях компаний, некорректные возрастные данные и др.

2.1.4 Структура и типы данных

Цель. Проверить размерность, типы столбцов и возможные пропуски.# # Базовая информация о таблице

```
dataframe.info()        # типы и пропуски
print(dataframe.shape)  # строки и столбцы
```

Комментарий. Метод info() показывает, какие столбцы object, int, float, а также наличие пропусков. Ожидаемый формат: (1000+, 28)

2.1.5 Промежуточные выводы

Цель. Подтвердить корректность набора данных перед EDA.

```
print('\nПромежуточные выводы:')

print(f'Итоговая форма датафрейма: {dataframe.shape}')

print('Типы данных после обработки:')

print(dataframe.dtypes)

print('Количество пропусков после обработки:')

print(dataframe.isnull().sum().sum())

print('Предобработка завершена.')
```

Комментарий. Проверяется размерность, пропуски и соответствие типов исходному набору. Вывод «Предобработка завершена» завершает вводную часть анализа.

2.2 ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ ДАННЫХ

В ходе EDA изучается структура датасета: статистики, уникальные значения, индексация, сортировки и фильтрации по ключевым признакам (зарплата, навыки, локация, компания). Построены сводные таблицы и агрегаты для выявления закономерностей рынка труда.

Результаты позволяют сформировать гипотезы о влиянии навыков, рейтинга компаний и региональных особенностей на уровень зарплат.

2.2.1 Индексация по координатам

```
# Индексация по координатам

dataframe.iloc[:5]  # первые строки

dataframe.loc[7, ['job_title', 'avg_salary']]

dataframe.iloc[:, 0:3]  # первые три столбца

dataframe.loc[230:240, ['company_name', 'rating']]

dataframe.iloc[-10:, -5:]  # последние строки и столбцы
```

Комментарий. Используются позиционная (iloc) и меточная (loc) индексации. Это позволяет просматривать выборочные данные в разных частях набора — проверка корректности зарплат, возраста компании, навыков.

2.2.2 Логическая индексация

```
# Логическая индексация

dataframe[dataframe['avg_salary'] > 100]

dataframe[(dataframe['python_yn'] == 1) & (dataframe['aws'] == 1)]

dataframe[dataframe['rating'] > 4]

dataframe[dataframe['age'] > 50]

dataframe[(dataframe['job_state'] == 'CA') & (dataframe['avg_salary'] > 120)]
```

Комментарий. Фильтрации показывают:

- высокие зарплаты,
- вакансии, где требуются Python+AWS,
- компании с высоким рейтингом,
- возрастные компании,
- вакансии в конкретных штатах.

2.2.3 Сортировка и анализ

```
# Сортировка данных

dataframe.sort_values(by='job_title').head()

dataframe.sort_values(by='avg_salary', ascending=False).head()

dataframe.sort_values(by='rating').head()  # самые низкие рейтинги

dataframe.sort_values(by='rating', ascending=False).head()  # самые высокие рейтинги
```

Комментарий. Сортировка выявляет:

- наиболее высокооплачиваемые должности,
- рейтинги компаний,
- различия между государствами и отраслями.

2.2.4 Фильтрация через query

```
# Метод query

dataframe.query("avg_salary > 120")

dataframe.query("rating > 4.0")

dataframe.query("python_yn == 1 and excel == 1")

dataframe.query("age < 10")

dataframe.query("avg_salary > 90 and python_yn == 1 and job_state == 'NY'")
```

Комментарий. Метод query() делает код фильтраций более читаемым. Используется для выбора вакансий по зарплате, возрасту компании, навыкам, региону.

2.2.5 Фильтрация через where

```
# Оператор where
show = dataframe.where(dataframe['rating'] > 4).dropna(how='all')
show

show = dataframe.where(dataframe['aws'] == 1).dropna(how='all')
show

show = dataframe.where((dataframe['avg_salary'] > 110) &
                       (dataframe['python_yn'] == 1)).dropna(how='all')
show

show = dataframe.where((dataframe['age'] > 30) &
                       (dataframe['excel'] == 1)).dropna(how='all')
show
```

Комментарий. where сохраняет исходную форму таблицы, что важно при создании масок и последующих объединений.

2.2.6 Сводные таблицы и агрегаты (pivot_table, groupby + agg)

```
# Сводные таблицы
dataframe.pivot_table(values='avg_salary',
                      index='job_state',
                      aggfunc='mean')
dataframe.pivot_table(values=['min_salary', 'max_salary'],
                      index='job_title',
                      aggfunc='mean')
dataframe.pivot_table(values='avg_salary',
                      index='python_yn',
                      aggfunc=['max', 'min', 'mean'])
dataframe.groupby('job_title').agg({
    'avg_salary': ['mean', 'min', 'max'],
    'rating': ['mean'],
    'age': ['mean']
})
```

Комментарий. Полученные сводные таблицы позволяют:

1. анализировать зарплаты по штатам;
2. сравнивать должности;
3. оценивать влияние навыков на зарплату;
4. изучать связь между рейтингом и уровнем компенсации.

2.2.7 Промежуточные выводы

1. Использование логических условий позволяет выделять важные сегменты: вакансии с Python, высокие зарплаты, нужные штаты, компании-с долгой историей.
2. Сортировки помогают выявлять крайние значения: топовые зарплаты, самые рейтинговые работодатели, редкие должности.
3. pivot_table и groupby формируют агрегаты по ключевым срезам рынка труда: отрасли, навыки, география, должности.
4. Эти результаты используются для построения дашбордов и рекомендаций: определение востребованных навыков, регионов с высоким спросом, оптимальных карьерных направлений.

ГЛАВА 3 РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

3.1 ГРАФИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Графическая визуализация данных позволяет выявить закономерности, тенденции и аномалии в распределении зарплат, характеристиках компаний и особенностях вакансий. Ниже приведены ключевые результаты анализа.

3.1.1 Гистограмма распределения ключевых числовых признаков

Для первичного анализа были построены гистограммы распределений следующих признаков: min_salary, max_salary, avg_salary, rating, founded, age.

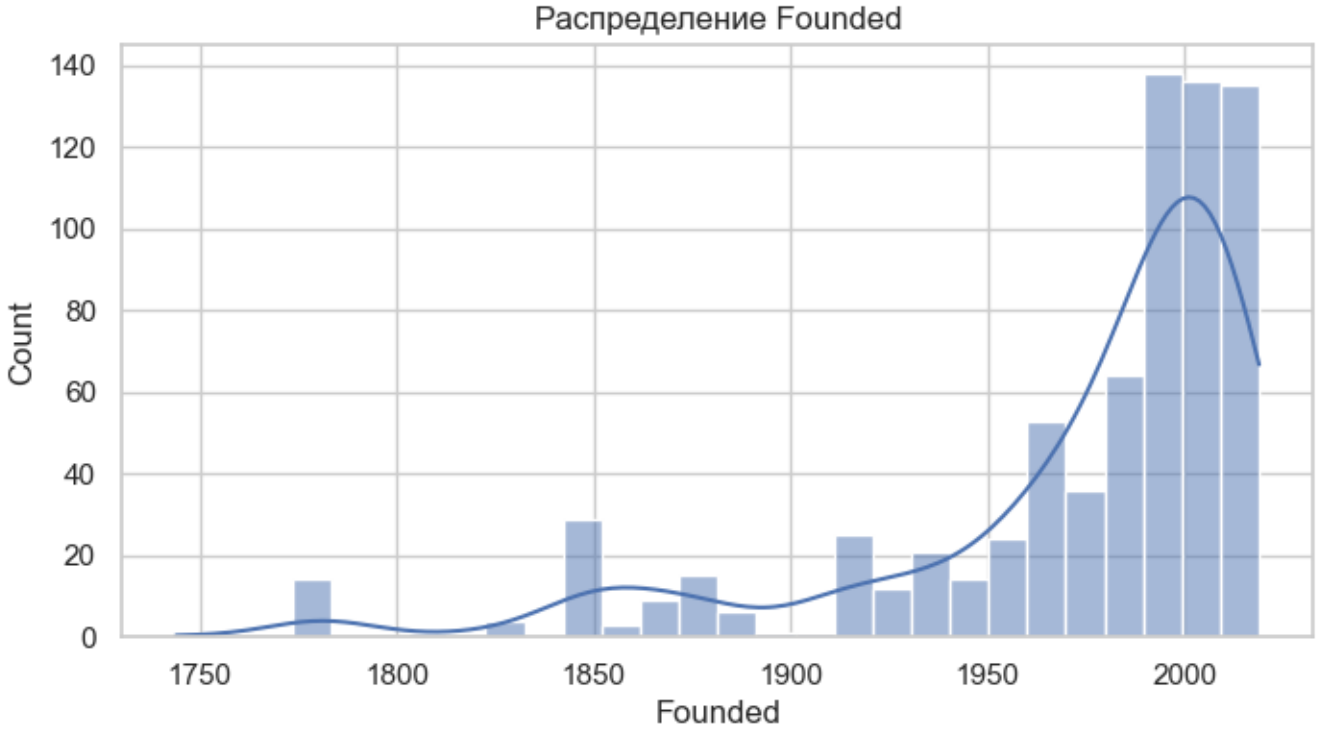

Рисунок 1 – Гистограмма распределения даты основания компании

Комментарий. Мини-вывод:

1. min_salary / max_salary / avg_salary. Распределения имеют выраженную правостороннюю асимметрию — большинство вакансий относится к диапазону зарплат ниже среднего по выборке. Это подтверждает доминирование предложений в нижнем и среднем сегменте рынка труда.
2. rating. Большая концентрация компаний имеет рейтинг от 3 до 4, что указывает на среднее качество условий труда по мнению сотрудников.
3. founded / age. Много компаний возрастом 20–40 лет. Старые компании (старше 80–100 лет) встречаются редко, но присутствуют как крупные корпорации.

3.1.2 Топ-10 должностей по количеству вакансий

```
top10_titles = df['job_title'].value_counts().head(10)
plt.figure(figsize=(8,4))
top10_titles.plot(kind='bar')
plt.title('Топ-10 должностей по числу вакансий')
plt.xlabel('job_title'); plt.ylabel('count')
plt.show()
```

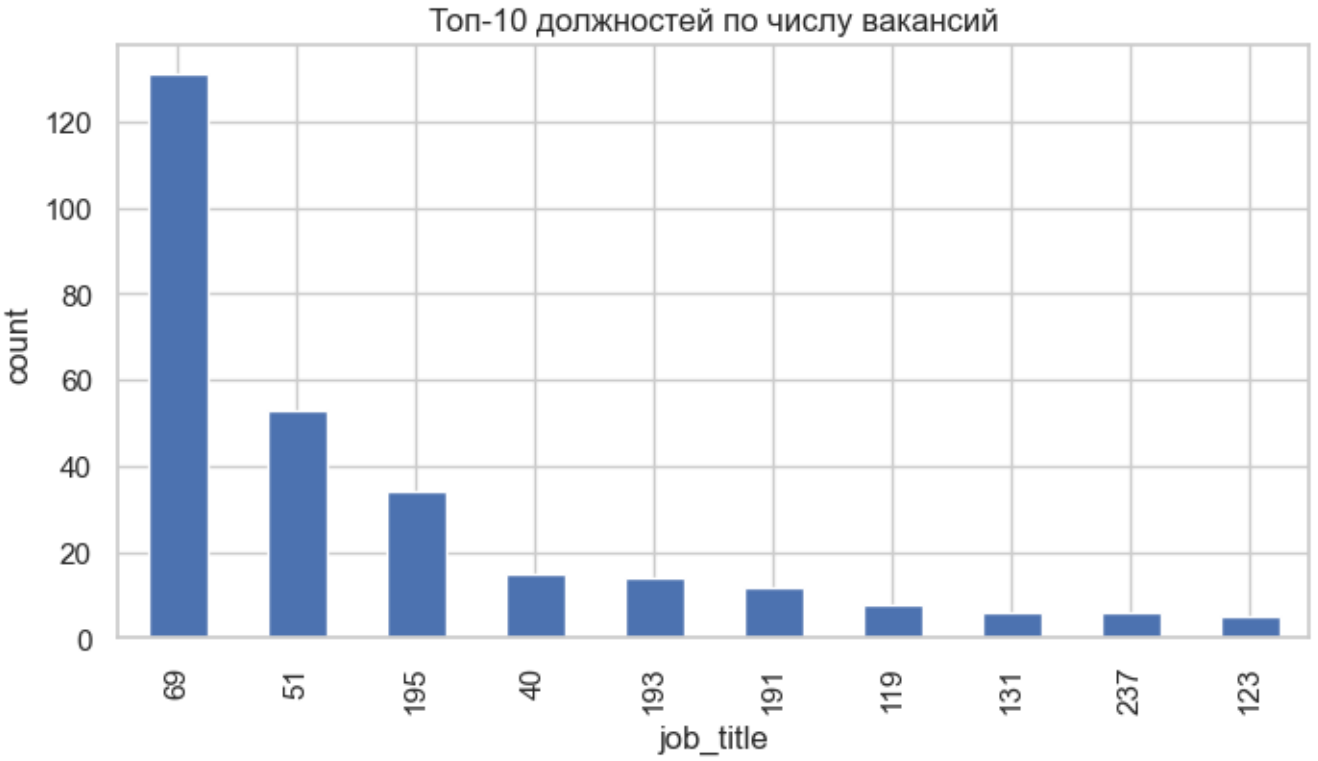

Рисунок 2 – Топ 10 должностей по количеству вакансий

Комментарий. На рынке труда в представленном наборе данных доминируют должности, связанные с анализом данных, машинным обучением и инженерией данных. Наиболее массовыми являются позиции уровня «data scientist», «data engineer», «data analyst», что подтверждает высокий спрос на специалистов в сфере обработки данных.

3.1.3 Корреляционный анализ

```
corr_cols =
['avg_salary','min_salary','max_salary','rating','age','python_yn','aws','spark','excel','r_yn','hourly
']

corr = df[corr_cols].corr(numeric_only=True)

plt.figure(figsize=(8,6))

sns.heatmap(corr, annot=True, fmt='.2f')

plt.title('Корреляции признаков')

plt.show()
```

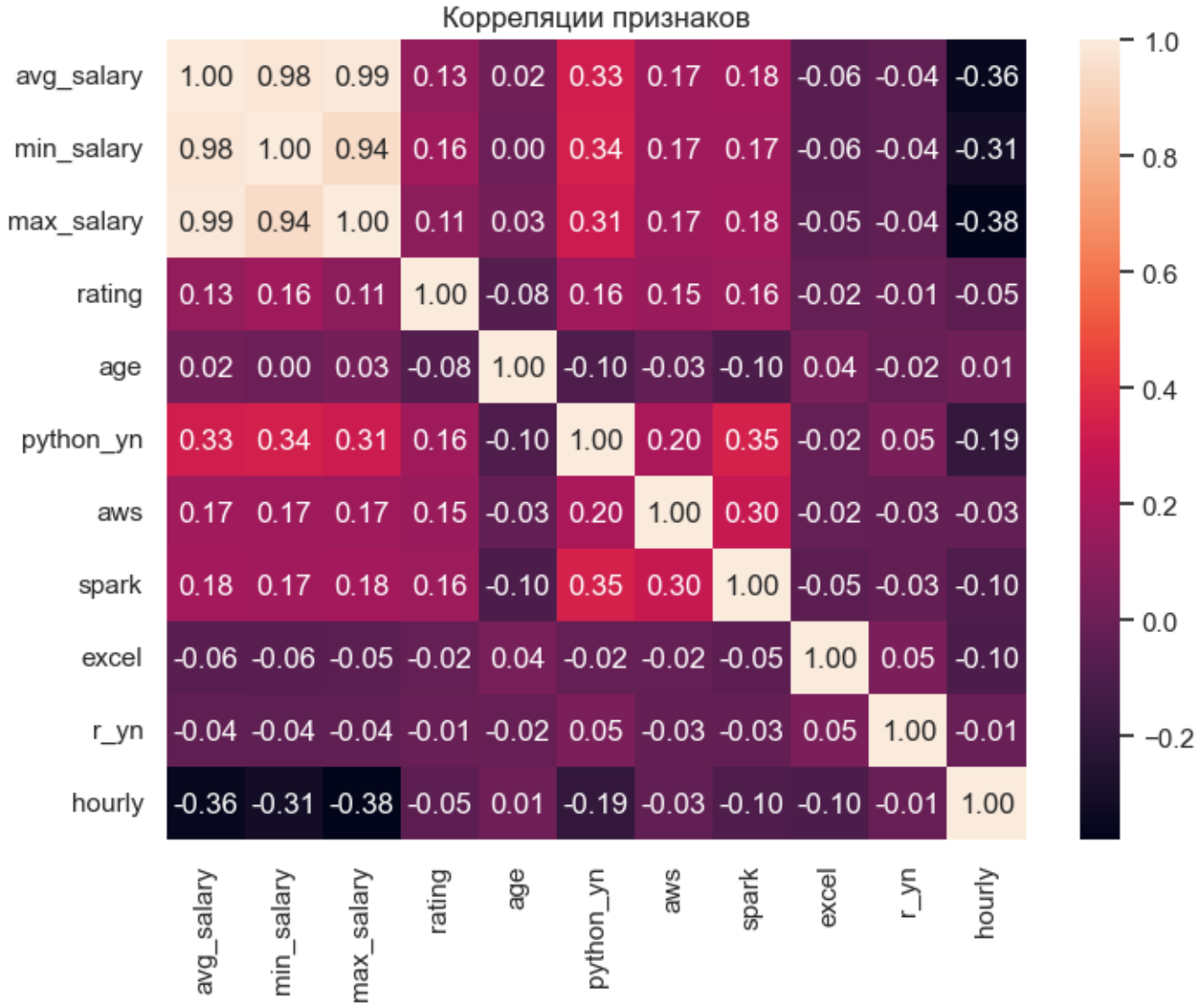

Рисунок 3 – Корреляционный анализ

Комментарий. Мини-вывод:

1. avg_salary ожидаемо сильно коррелирует с min_salary и max_salary, что подтверждает корректность расчётов.
2. Навыки (python_yn, aws, spark) имеют слабую, но положительную корреляцию с уровнем зарплаты.
3. Признак hourly (почасовая оплата) показывает слабую отрицательную корреляцию с годовой зарплатой.
4. Возраст компании (age) слабо влияет на уровень оплаты — это подтверждает, что компенсации определяются скорее ролью и навыками, а не историей компании.

3.1.4 Boxplot зарплат по штатам (топ-8)

```
top_states = df['job_state'].value_counts().head(8).index
plt.figure(figsize=(10,5))
sns.boxplot(data=df[df['job_state'].isin(top_states)], x='job_state', y='avg_salary')
plt.title('Распределение avg_salary по штатам (топ-8)')
plt.show()
```

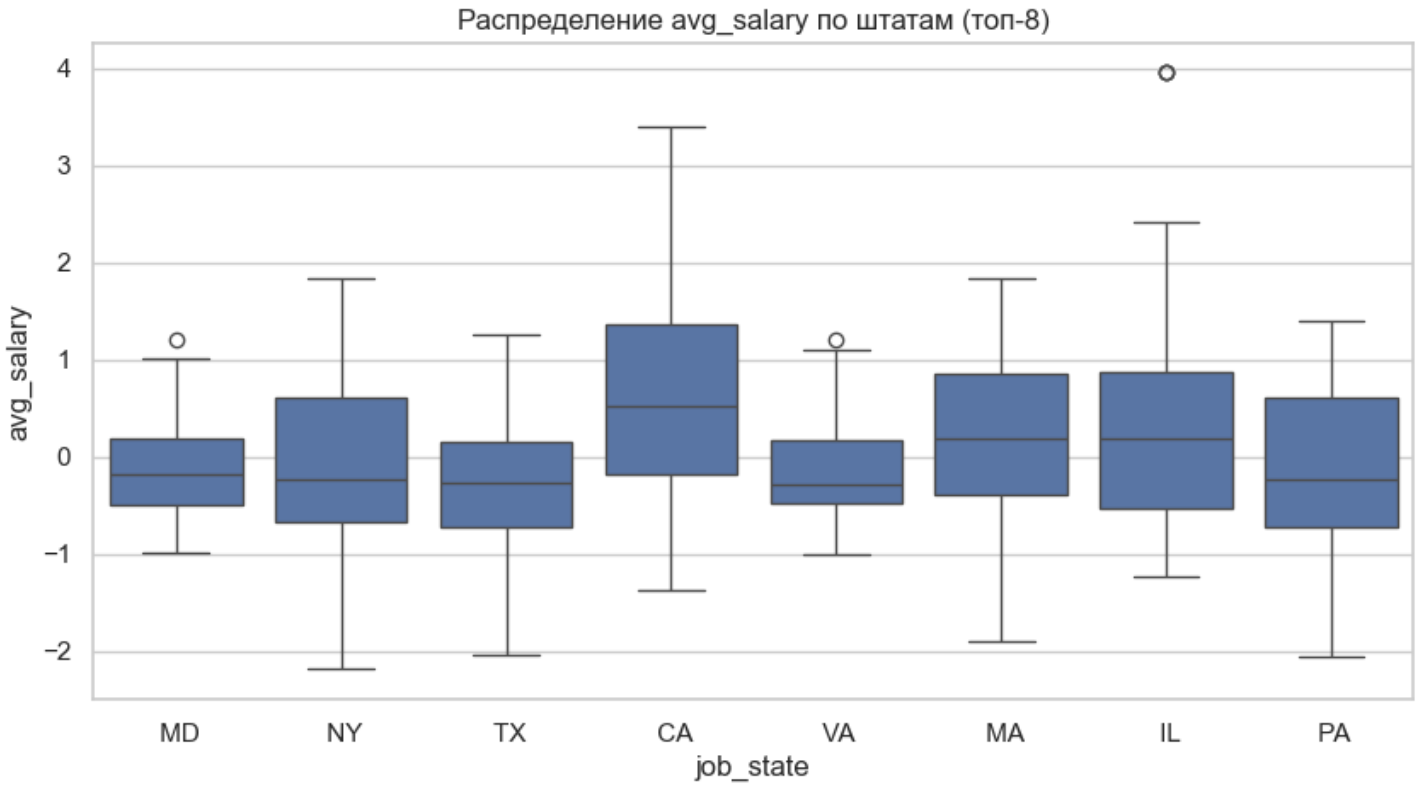

Рисунок 4 – Boxplot зарплат по штатам (топ-8)

Комментарий. Мини-вывод:

1. Наиболее высокая медианная зарплата наблюдается в таких штатах, как CA (Калифорния) и NY (Нью-Йорк) — это объясняется присутствием крупных технологических компаний.
2. Штаты средней категории показывают более низкий и равномерный уровень зарплат.
3. В некоторых штатах присутствуют выбросы, что отражает вакансии для высококвалифицированных специалистов или уникальных позиций с повышенной оплатой.

3.1.5 Выводы по графическому анализу

- Большая часть вакансий сосредоточена в зарплатном диапазоне ниже среднего, что указывает на доминирование массовых позиций аналитиков начального и среднего уровня.

- Python, AWS и Spark являются ключевыми навыками, существенно повышающими уровень дохода.

- Штаты с крупными IT-центрами (CA, NY) имеют заметно более высокие медианные зарплаты.

- Рейтинг компаний сосредоточен в диапазоне 3–4, что говорит об умеренном уровне удовлетворённости сотрудников.

- Корреляционный анализ подтверждает логическую связь зарплатных метрик и слабое влияние возрастных характеристик компаний.

ЗАКЛЮЧЕНИЕ

В ходе выполнения учебной практики была реализована полная последовательность работ по анализу данных и разработке вспомогательного программного инструмента на основе датасета вакансий, содержащего сведения о заработных платах, характеристиках компаний и требованиях к навыкам соискателей.

На первом этапе была выполнена предобработка данных, включающая приведение столбцов к стандартизированным форматам, устранение дубликатов, заполнение и обработку пропусков, извлечение новых признаков (признаки навыков, возраст компании, диапазоны заработной платы).

В рамках исследовательского анализа данных (EDA) были выполнены сортировка, фильтрация, группировка данных, построение сводных таблиц, изучение распределений показателей, анализ влияния отдельных факторов на уровень дохода, а также анализ корреляций между ключевыми характеристиками вакансий. Графический анализ, выполненный с использованием библиотек Matplotlib и Seaborn, позволил визуализировать: распределение зарплатных показателей; географические различия в оплате труда; зависимость уровня зарплаты от ключевых профессиональных навыков; наиболее востребованные должности; взаимосвязи между числовыми и категориальными признаками.

Отдельным этапом работы стало создание простого аналитического приложения на Python с графическим интерфейсом (Tkinter), позволяющего пользователю загружать очищенный CSV-файл, изучать основные сведения о датасете, выполнять фильтрацию данных по заданным параметрам и получать текстовые бизнес-рекомендации, основанные на результатах анализа.

Таким образом, поставленные цели и задачи практики были выполнены в полном объёме. В ходе работы были получены навыки предварительной обработки данных, визуализации, анализа взаимосвязей, построения аналитических выводов и разработки прикладного программного обеспечения на Python. Полученные результаты демонстрируют способность применять методы анализа данных в реальной аналитической задаче и подтверждают значимость освоенных компетенций для дальнейшей профессиональной деятельности в области Data Analysis и Data Science.

СПИСОК ЛИТЕРАТУРЫ

1. Pandas Documentation [Электронный ресурс]. URL: https://pandas.pydata.org/
2. Seaborn: Statistical Data Visualization [Электронный ресурс]. URL: https://seaborn.pydata.org/
3. Matplotlib: Visualization with Python [Электронный ресурс]. URL: https://matplotlib.org/
4. Tkinter — Python Interface to Tcl/Tk [Электронный ресурс] // Python Docs. URL: https://docs.python.org/3/library/tkinter.html
5. Kaggle Dataset: Job Salaries / Data Science Salaries [Электронный ресурс]. URL: https://www.kaggle.com/
6. Работа с пропущенными значениями в pandas [Электронный ресурс] // DevPractice. URL: https://devpractice.ru/pandas-work-with-nan-part4/
7. Сводные таблицы и группировка данных в pandas [Электронный ресурс] // Habr. URL: https://habr.com/ru/articles/713506/
8. NumPy Documentation [Электронный ресурс]. URL: https://numpy.org/doc/stable/
9. Методы визуализации данных в Python [Электронный ресурс] // AndreyEX. URL: https://andreyex.ru/
10. Основы анализа данных в Python [Электронный ресурс] // Яндекс.Практикум. URL: https://practicum.yandex.ru/
11. Python. Типы данных и методы работы со строками [Электронный ресурс] // Docs-Python.ru. URL: https://docs-python.ru/
12. Работа с DataFrame.query в pandas [Электронный ресурс] // Dev-Gang. URL: https://dev-gang.ru/
13. Pandas — базовый курс по обработке данных [Электронный ресурс] // Habr. URL: https://habr.com/ru/companies/skillfactory/articles/683738/
14. Timeweb Cloud — руководство по NumPy и Pandas [Электронный ресурс]. URL: https://timeweb.cloud/tutorials/python/