Московский государственный технический университет им. Н. Э. Баумана

Курс «Технологии машинного обучения»
Отчёт по лабораторной работе №1

Выполнил:	Проверил:
Горенков А.А.	Гапанюк Ю.Е
группа ИУ5-63Б	

Подпись:

Дата: 14.02.25

Дата:

Цель лабораторной работы: изучение различных методов визуализация данных.

Краткое описание. Построение основных графиков, входящих в этап разведочного анализа данных.

Рекомендуемые инструментальные средства можно посмотреть здесь.

Задание:

- Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов <u>здесь.</u>
- Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных, например из <u>Scikit-learn.</u>
- Пример преобразования датасетов Scikit-learn в Pandas Dataframe можно посмотреть здесь.

Для лабораторных работ не рекомендуется выбирать датасеты большого размера.

- Создать ноутбук, который содержит следующие разделы:
- 1. Текстовое описание выбранного Вами набора данных.
- 2. Основные характеристики датасета.
- 3. Визуальное исследование датасета.
- 4. Информация о корреляции признаков.
- Сформировать отчет и разместить его в своем репозитории на github.

Ход выполнения:

Загрузка и первичный анализ

Анализ наличия ЗП в вакансии

0 49313809 Space307

Golang

Developer

```
[ ] value_counts = df['Salary'].value_counts()

plt.figure(figsize=(7, 7))
value_counts.plot(kind='pie', autopct='%1.1f%', startangle=90)

plt.title('Наличие ЗП в описании вакансии')

plt.ylabel('')
plt.show()
```

Полный

От 3 до 6

Мы в Ѕрасе307

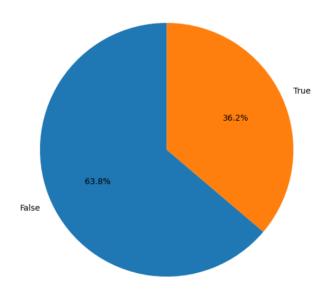
марабатываем Санкт- ['Программист, ['Программиров международную Петербург разработчик'] Разраб

['Docker', 'Golang', 'Redis', 'Английский язык...

____*

Наличие ЗП в описании вакансии

True 251322.0



Определение корреляции между максимальной заработной платой и требуемым опытом.

```
[ ] experience_order = ['Her onыта', 'Ot 1 roда до 3 лет', 'Ot 3 до 6 лет', 'Более 6
df['Experience'] = pd.Categorical(df['Experience'], categories=experience_order,

plt.figure(figsize=(8, 6))
df.boxplot(column='To', by='Experience', grid=False)

plt.title('3П для различного уровня опыта')
plt.suptitle('')
plt.xlabel('Уровень опыта')
plt.ylabel('Максимальная 3П (руб)')

plt.tight_layout()
plt.show()
```

→ <Figure size 800x600 with 0 Axes>

