

От чего зависит доход в it?

// и как увеличить свой

просто покажите prediction HR или начальнику
и потребуйте незамедлительного повышения :)

Кто я такая, чтобы знать
способы увеличения
дохода?



Хисаметдинаева Динара Кз241
поток МЕВД 2.1



GOALS AND TASKS

Какие цели преследовала и какие задачи
ставила?

Data Analysis

- Исследовать рынок и узнать, от каких параметров и насколько зависит заработка платя IT-специалистов
- Сделать визуализацию по EDA и выводы

Machine Learning

- Обучить несколько моделей на подготовленных данных и выбрать наиболее точную
- Сделать предсказание зарплаты на тестовых данных

Web Development

- Написать сайт-предсказатель зарплаты по заданным пользователем параметрам
- Добавить страницу с дашбордами, графиками и выводами

Для реализации этих целей было необходимо:

```
def cleand(src_df):
    Loading...
    if 'Yearly stocks' in src_df.columns:
        cleand = src_df.assign(net_salary=lambda x: x['Yearly brutto salary (without bonus and sto
    elif 'Yearly bonus + stocks in EUR' in src_df.columns:
        src_df['Yearly bonus + stocks in EUR'] = pd.to_numeric(src_df['Yearly bonus + stocks in EU
        cleand = src_df.assign(net_salary=lambda x: x['Yearly brutto salary (without bonus and sto
    else:
        cleand = src_df.drop('Are you getting any Stock Options?', axis=1)

    return cleand

def cleand_last_yr(src_df):
    if 'Yearly brutto salary (without bonus and stocks) one year ago. Only answer if staying in sa
        cleand_last_yr = src_df.assign(salary_last_year=lambda x: x['Yearly brutto salary (without
    elif 'Annual brutto salary (without bonus and stocks) one year ago. Only answer if staying in
        src_df['Annual brutto salary (without bonus and stocks) one year ago. Only answer if stayi
        src_df['Annual bonus+stocks one year ago. Only answer if staying in same country'] = pd.to
        cleand_last_yr = src_df.assign(salary_last_year=lambda x: x['Annual brutto salary (without
    else:
        cleand_last_yr = src_df
```

31.0	Male	Munich	0	DevOps	8	AWS	
37.0	Male	Utrecht	Middle	Manager	17	Java	
26.0	Female	Amsterdam	Middle	Backend Developer	4	Java	
33.0	Male	Amsterdam	Senior	Frontend Developer	10	Javascript / Typescript	
28.0	Male	Moscow	Lead	Frontend Developer	5	Javascript / Typescript	
27.0	Male	Berlin	Senior	Data Scientist	5	SAS	
30.0	Male	Amsterdam	Senior	Data Engineer	14	Not Relevant	
46.0	Male	Munich	Lead	SRE	20	Openstack	
41.0	Male	Munich	Senior	Backend Developer	19	.NET	

Tasks for preprocessing

ЭТО БЫЛО МАКСИМАЛЬНО ДОЛГО И СКУЧНОВАТО

- ★ Почистить датасеты

В НИХ СОВЕРШЕННО РАЗНЫЕ КОЛОНКИ, БЫЛО МУТОРНО ПОДГОНЯТЬ
ИХ К ОДНОМУ

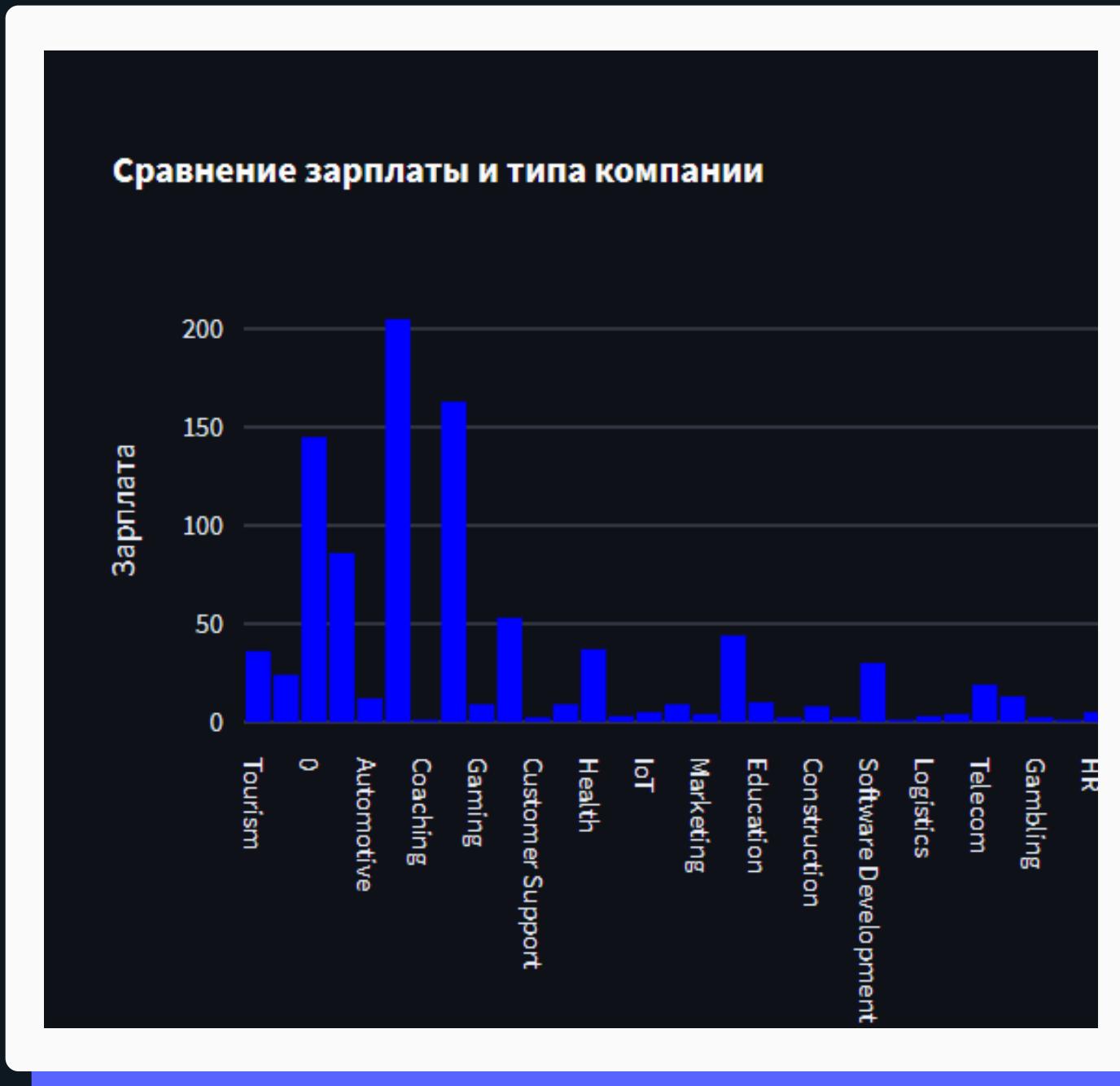
- ★ Объединить датасеты с зарплатами по
разным годам

ПРИВОДИТЬ ДАННЫЕ К ОДНОМУ ЕДИНОМУ ТИПУ В ОДНОЙ КОЛОНКЕ,

ПИСАТЬ ЭНКОДИНГИ, ВЫБИРАТЬ ТОЛЬКО ВЛИЯЮЩИЕ НА
ПРЕДСКАЗАНИЕ ПЕРЕМЕННЫЕ - УТОМИТЕЛЬНОЕ ЗАНЯТИЕ

- ★ Подготовить данные к обучению

Для реализации этих целей было необходимо:



Tasks for visualisation

При помощи дашбордов, установить:

- ★ Какие отрасли наиболее высоко награждают сотрудников?
- ★ Важен ли гендер?
- ★ Выгоднее работать в стартапе или в корпорации?
- ★ Насколько сильно Covid-19 усложнил ситуацию на рынке труда?
- ★ Какой язык программирования лучше изучать?

Для реализации этих целей было необходимо:

```
5: learn: 33284.8394819 total: 19.9ms remaining: 3.3s
6: learn: 33082.7408282 total: 23.6ms remaining: 3.35s
7: learn: 32875.4348552 total: 26.7ms remaining: 3.31s
8: learn: 32705.3973431 total: 28.6ms remaining: 3.15s
9: learn: 32537.1216025 total: 29.8ms remaining: 2.95s
10: learn: 32388.8961164 total: 31.2ms remaining: 2.8s
11: learn: 32226.2088147 total: 32.6ms remaining: 2.69s
12: learn: 32095.8346117 total: 34ms remaining: 2.58s
13: learn: 31964.3774843 total: 35ms remaining: 2.47s
14: learn: 31830.4817748 total: 36.1ms remaining: 2.37s
15: learn: 31720.4682447 total: 37.1ms remaining: 2.28s
16: learn: 31651.2434719 total: 38.1ms remaining: 2.2s
17: learn: 31538.7590845 total: 39.2ms remaining: 2.14s
18: learn: 31429.3169569 total: 40.2ms remaining: 2.08s
19: learn: 31335.0649964 total: 41.4ms remaining: 2.03s
20: learn: 31247.6432285 total: 42.4ms remaining: 1.98s
21: learn: 31165.1731731 total: 43.5ms remaining: 1.93s
22: learn: 31104.5751969 total: 44.5ms remaining: 1.89s
23: learn: 31033.6837077 total: 45.6ms remaining: 1.85s
...
998: learn: 26771.1949520 total: 1.18s remaining: 1.18ms
999: learn: 26768.9314706 total: 1.18s remaining: 0ms
```

Tasks for Machine Learning

для этого пришлось долго тестирувать разные и на разных параметрах, изучать принципы работы либ

- ★ Выбрать оптимальный алгоритм регрессии и библиотеку
- ★ Пришлось углубиться в математические подробности Подбор таких параметров регрессора, чтобы модель получилась наиболее точной
- ★ Звучит просто, однако нужно было исправить тонны ошибок и поменять много фич в датасете, прежде чем заработало
- ★ Сделать тест сплит, энкодинг категориальных признаков, зафитить

Для реализации этих целей было необходимо:

```
genders = (  
    'Male',  
    'Female',  
    'Diverse'  
)  
  
positions = (  
    'QA',  
    'Software Engineer'  
)  
  
experience = st.slider('опыт', 1.0, 40.0, 0.2)  
age = st.slider('возраст', 16.0, 60.0, 0.2)  
seniority = st.selectbox('уровень', seniority_levels)  
main_lang = st.selectbox('основной язык общения с коллегами', main_langs)  
gender = st.selectbox('гендер', genders)  
position = st.selectbox('позиция', positions)  
  
flag = st.button('Рассчитать')  
if flag:  
    X = pd.DataFrame(data=[experience, age, seniority, main_lang,
```

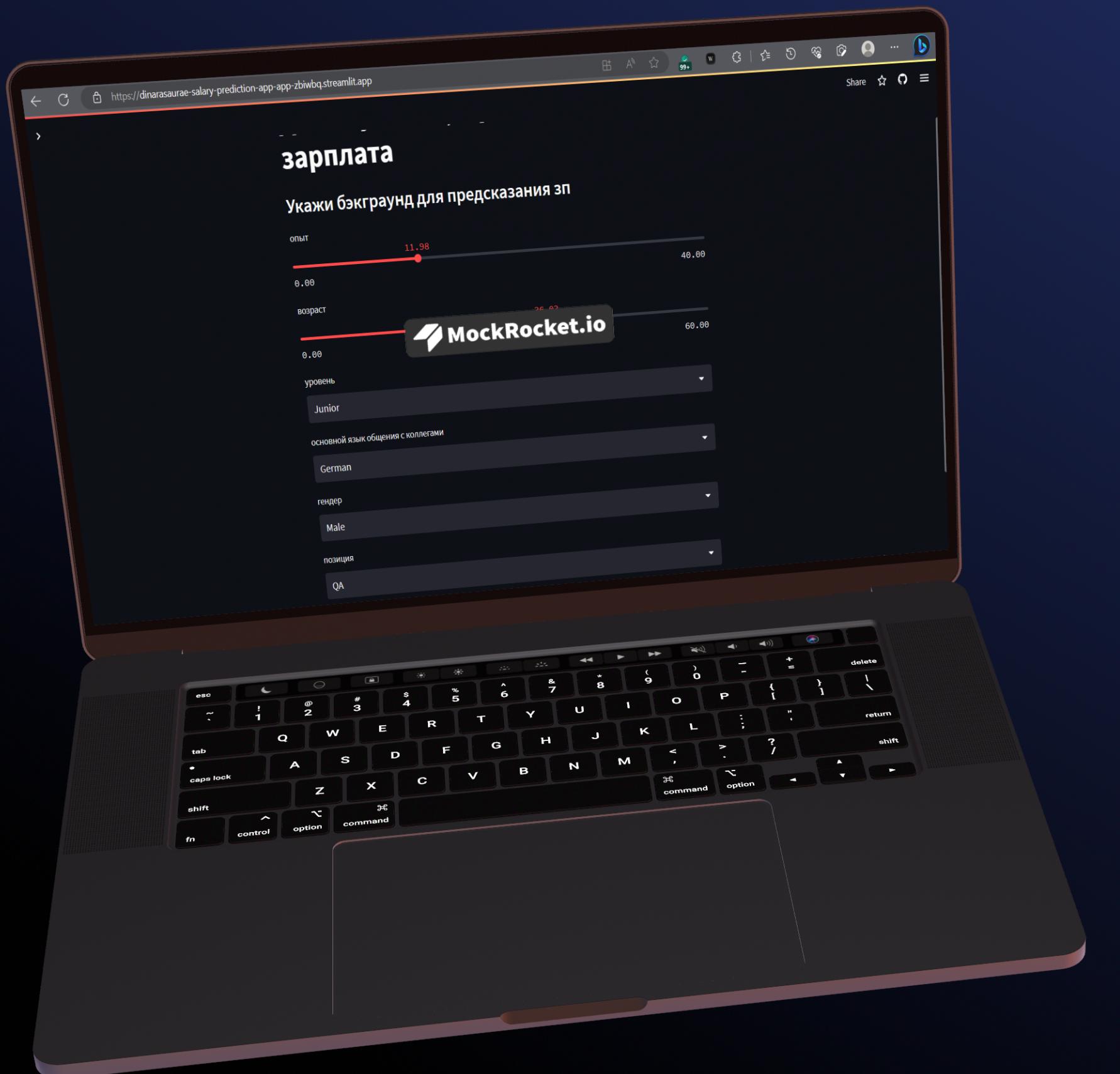
Tasks for Web Development

- ★ Разметить архитектуру многостраничника
написать слайдеры, содержимое
выпадающих списков
- ★ ПРИШЛОСЬ УГЛУБИТЬСЯ В МАТЕМАТИЧЕСКИЕ ПОДРОБНОСТИ
Импортировать модель из бинарника
(pickle) и реализовать предсказание
зарплаты по пользовательским данным
- ★ ПРИШЛОСЬ УГЛУБИТЬСЯ В МАТЕМАТИЧЕСКИЕ ПОДРОБНОСТИ
На другой странице реализовать
дашборды с помощью plotly express,
расписать выводы
- ★ Build and deploy !

Workflows

PRODUCT CONCEPTS

Наконец, то, что можно потыкать руками



SAL
PRE



APP

Здесь можно протестировать приложение

thx for attention :)

THE END