

# Программа для анализа рынка труда аналитиков данных

Software for Data Analytics Labor Market  
Analysis

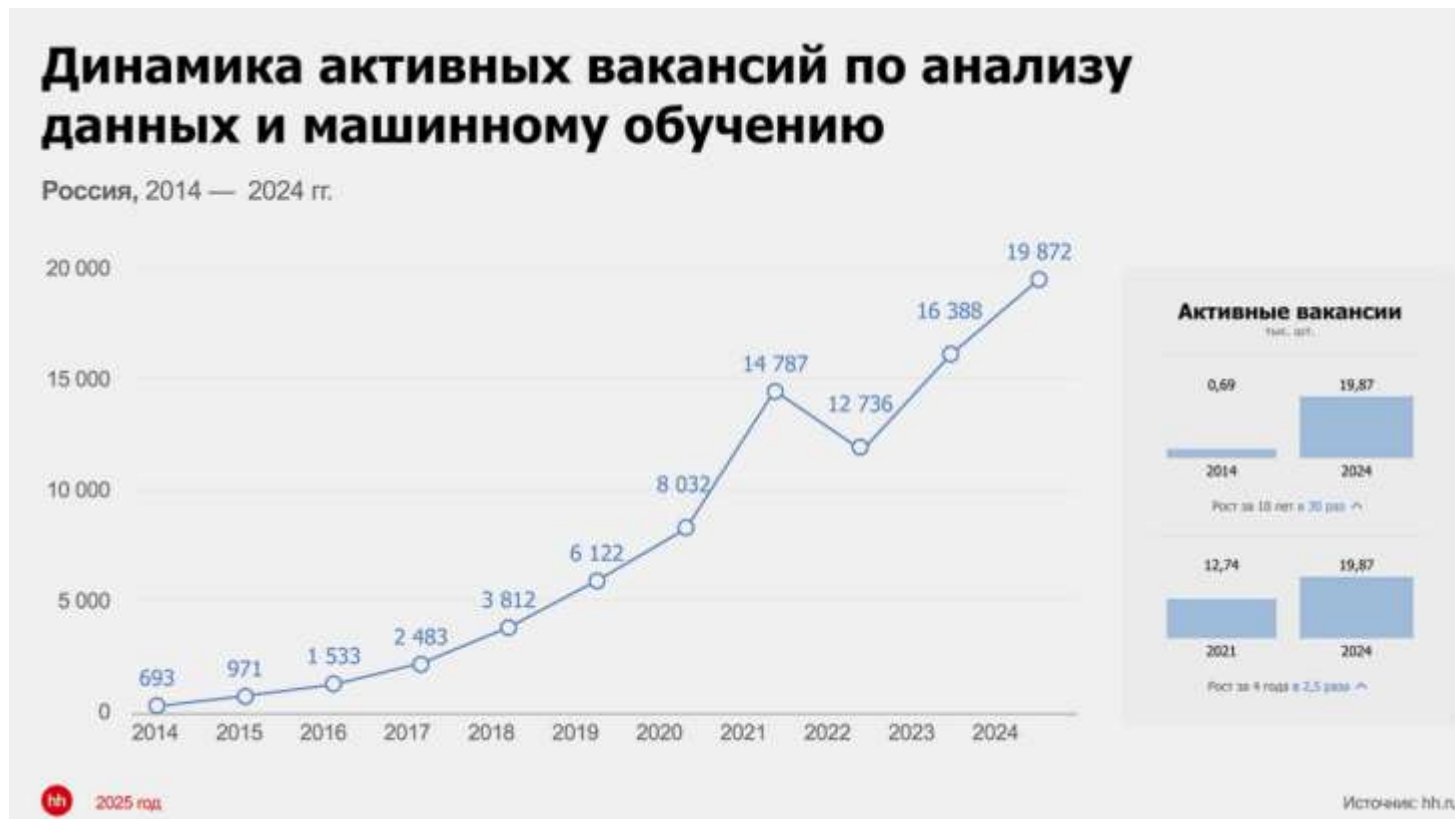
Курсовая работа | Программный проект

**Выполнил студент БПИ2310:**  
Муслин Артемий Михайлович

**Научный руководитель:**  
доцент департамента программной  
инженерии факультета компьютерных наук,  
кандидат наук  
Меликян Алиса Валерьевна

# Описание предметной области

Проект посвящён анализу рынка труда специалистов в области аналитики данных в России. В условиях стремительного роста спроса на специалистов по данным, компании публикуют тысячи вакансий с различными требованиями, а соискатели размещают подробные резюме. Актуальная задача — автоматизированный сбор, структурирование и визуализация информации о вакансиях и резюме



# Объект исследования и его пользователи

**Объект исследования** – рынок труда аналитиков данных в России:

- Зарплатные предложения работодателей
- Требования к навыкам и квалификации
- Основные тенденции рынка
- Различие предложений в разных городах
- Информация о соискателях

## Пользователи и их цели

Категория	Задача
Соискатели	Анализ зарплат, прогноз своей ЗП, сравнение навыков
Работодатели	Оценка рыночных условий, корректировка вакансий
Аналитики	Исследование трендов (например, спрос на Python vs. SQL)

# Актуальность работы

## Проблема рынка труда

- Отсутствие **интерактивных инструментов** для анализа вакансий и резюме аналитиков данных в РФ
- Пользователи вынуждены вручную собирать и обрабатывать данные с hh.ru

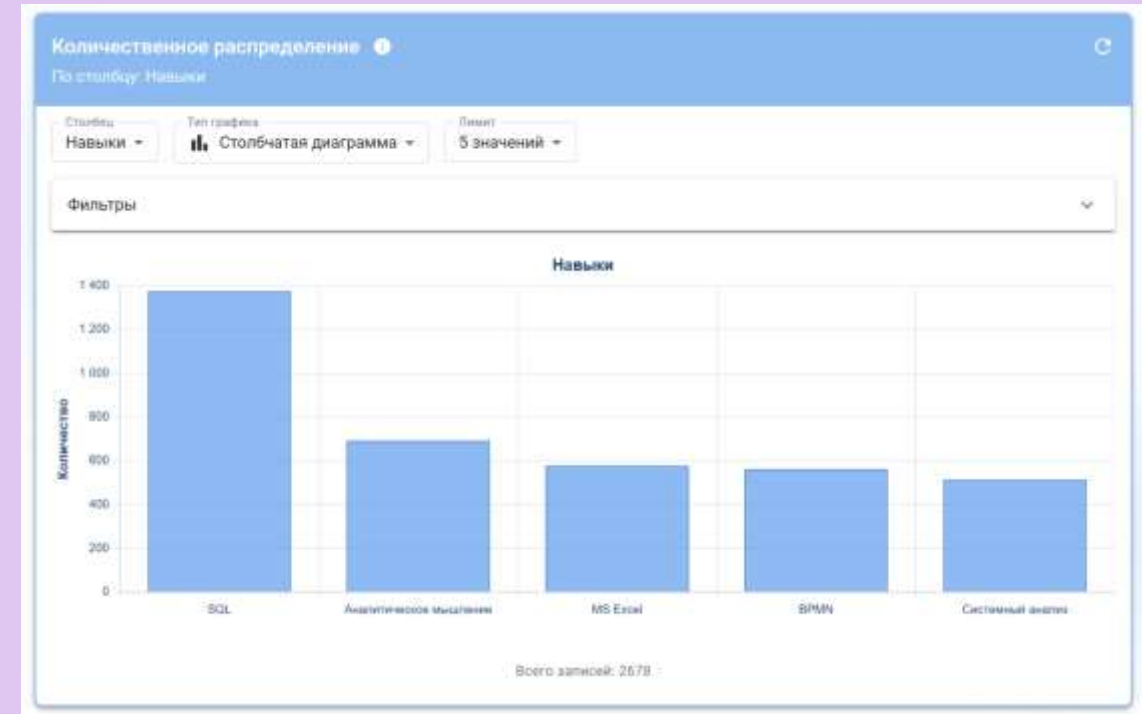
## Почему это важно?

Для соискателей:

- Нет точных данных для переговоров о зарплате
- Сложно оценить востребованность навыков

Для работодателей:

- Нет инструментов для анализа конкурентных предложений



# Анализ существующих решений

Критерий	<u>Habr</u>	<u>NewHR</u>	<u>navigator.hse.ru</u>	<u>Tableau Public</u>	DataCareerInsight
Интерактивные графики	✗	✗	✗	✓	✓
Данные по вакансиям	✓	✗	✓	✓	✓
Данные по резюме	✗	✗	✗	✗	✓
Прогноз зарплат	✗	✗	✗	✗	✓
Дополнение данных	✗	✗	✗	✗	✓
Расширенная фильтрация данных	✗	✗	✗	✗	✓

\*Показаны самые значимые критерии

# Цель и задачи работы

## Цель

**Разработка веб-приложения** для анализа рынка труда аналитиков данных с:

- Автоматизированным сбором данных с hh.ru
- Интерактивной визуализацией
- Прогнозированием зарплат на основе линейной регрессии

## Задачи

- Реализация парсера вакансий и резюме с hh.ru
- Проектирование БД
- Разработка API для обработки запросов:
  - Фильтрация, сортировка, группировка данных
  - Генерация аналитических графиков
- Создание фронтенда с таблицами и визуализациями
- Построение и интеграция ML-модели прогнозирования ЗП

# Архитектура приложения



# Выбор используемых методов и алгоритмов

## Парсинг данных:

Выбор API hh.ru (а не веб-скрапинга) для:

- Легального доступа к данным
- Поддержки всех полей вакансий/резюме
- Обхода блокировок

## Хранение данных:

PostgreSQL за:

- **Поддержку сложных типов данных**
  - Работа с массивами
  - Хранимые компоненты
- **Производительность на аналитических запросах**
- **Надежность и соответствие стандартам**
  - Интеграция с Python (библиотека psycopg2)
  - Поддержка в FastAPI через SQLAlchemy

## Backend (FastAPI):

- Автоматическая генерация OpenAPI-документации
- Асинхронная обработка 1000+ запросов/сек
- Интеграция с SQLAlchemy для сложных SQL-запросов

## Frontend (React + Material-UI):

- Компонентный подход (Позволил создать переиспользуемые компоненты для таблиц, фильтров и т.д.)
- Корректное отображение интерфейса на современных браузерах для React
- Material-UI обеспечивает единый стиль и адаптивность интерфейса
- Готовые компоненты Chart.js для визуализации интерактивных графиков

## Предсказание зарплаты (линейная регрессия)

- Интерпретируемость результатов
- Работа с категориальными признаками
- Быстрое предсказание



# Реализация БД

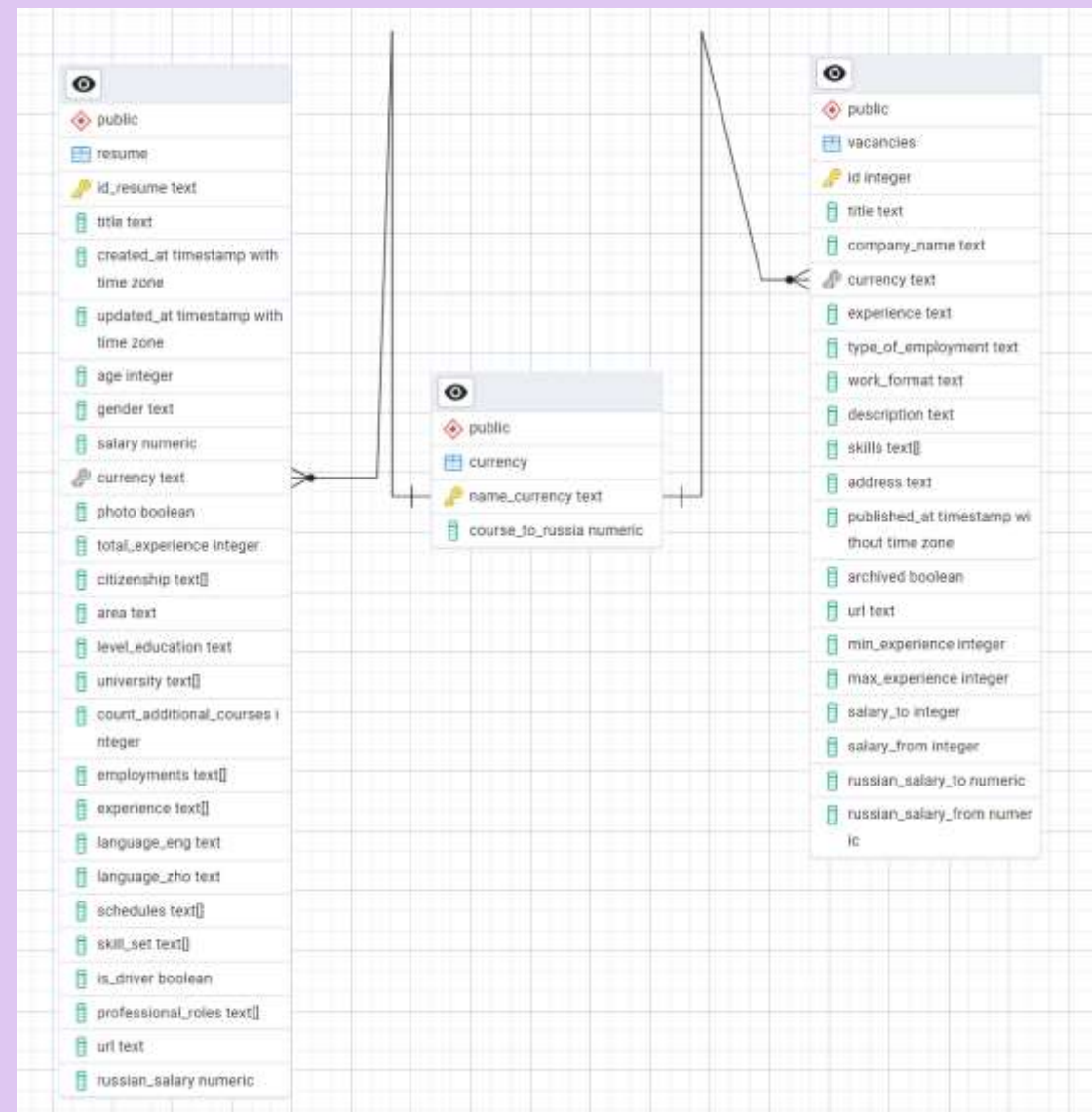
Есть 2 основных таблицы – вакансии и резюме, где собраны все данные с hh.ru

Также есть дополнительная таблица с валютами (данные из ЦБ) для перевода зарплаты в рубли

Для того, чтобы данные пересчитывалась корректно, были созданы триггеры

## Trigger Functions (4)

- `calculate_russian_salaries_resume()`
- `calculate_russian_salaries_vacancy()`
- `update_experience_fields()`
- `update_russian_salaries()`

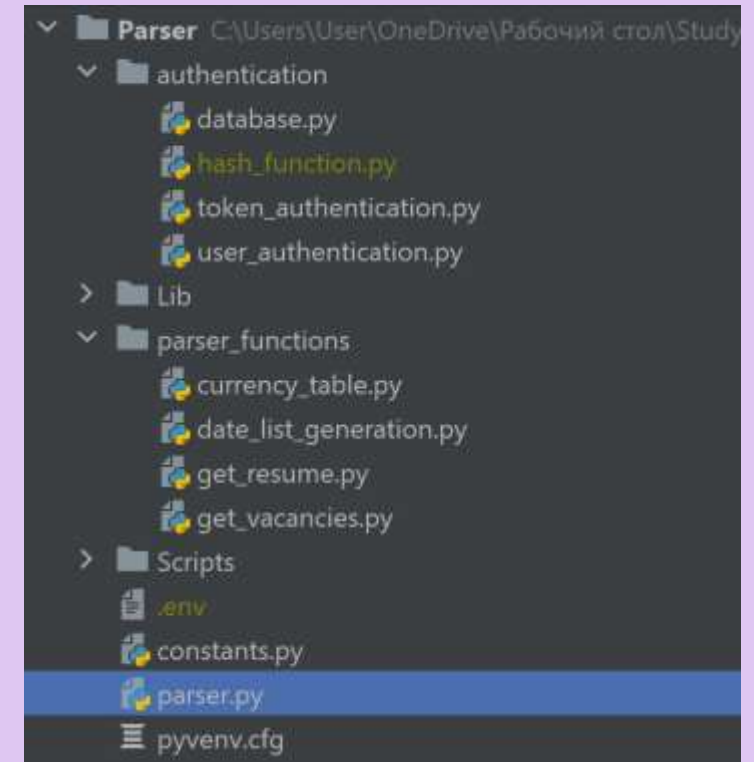


# Реализация Парсера

В .env хранятся захешированные данные о доступе к БД, а также к API.hh.ru

Парсер реализован в виде консольного приложения, где мы можем либо запустить сбор данных о вакансиях/резюме, либо обновить таблицу с валютами через API ЦБ

```
PASSWORD_HASH=6399c11b3e40d2c8545cd76f1b7f68ec9b5dbce347af646a16f2
USER_HASH=a942b37ccfaf5a813b1432caa209a43b9d144e47ad0de1549c289c25
CLIENT_ID=T7SU91LK70ENTG9CH5CL926RD6FEHIMM2NINSFCOEMBL0PFF8PLN5EE5
CLIENT_SECRET=NAAUJUHS3CIEEHNTUFD1DS4U5N66GNN4U1AB85B3RIOBQMFHJOUF
REDIRECT_URI=http://localhost
```



Откройте эту ссылку в браузере и подтвердите доступ:

[https://hh.ru/oauth/authorize?response\\_type=code&client\\_id=T7SU91LK70ENTG9CH5CL926RD6FEHIMM2NINSFCOEMBL0PFF8PLN5EE5](https://hh.ru/oauth/authorize?response_type=code&client_id=T7SU91LK70ENTG9CH5CL926RD6FEHIMM2NINSFCOEMBL0PFF8PLN5EE5)

Введите URL, на который вас перенаправило после авторизации: <http://localhost/?code=Q5DPJRSA05LVEA0902>

Токен успешно получен!

Введите пользователя:

Введите пароль:

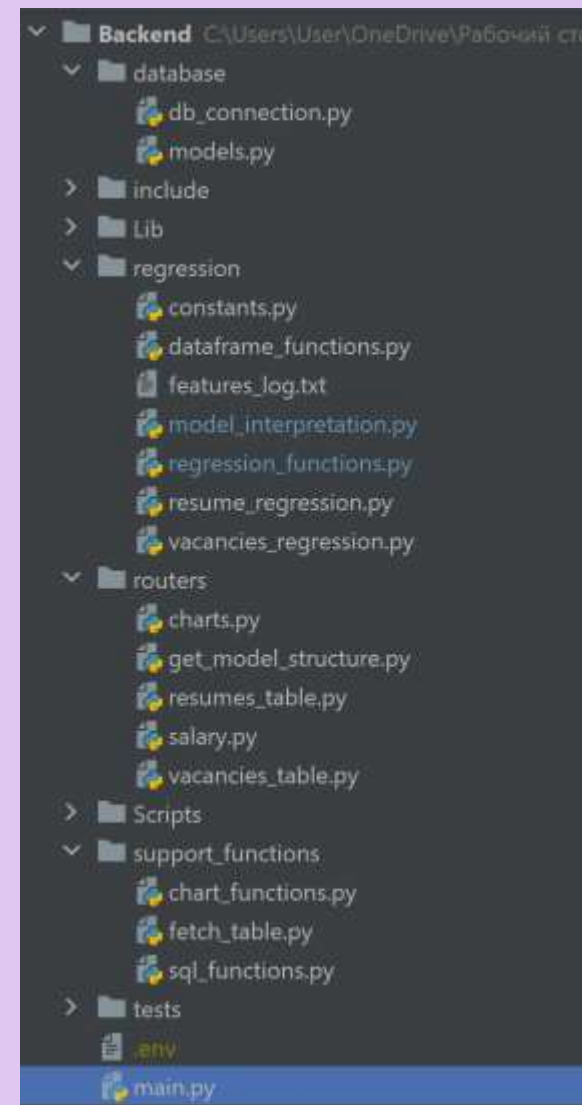
Доступ разрешён

Для какой таблицы идет сбор данных: Вакансии (v), Резюме (r), Валюты (c):

# Реализация Сервера

Серверная часть приложения построена на FastAPI и обеспечивает взаимодействие между клиентом (фронтендом) и базой данных (PostgreSQL). Она отвечает за:

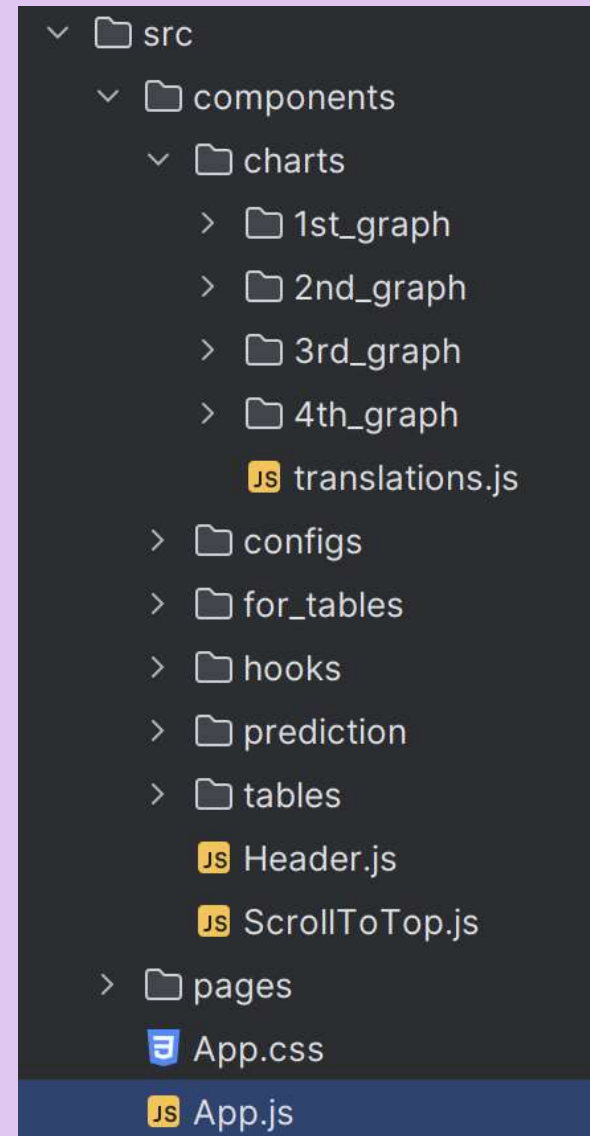
- Обработку API-запросов
- Фильтрацию, агрегацию и сортировку данных
- Генерацию данных для графиков
- Прогнозирование зарплат с помощью моделей линейной регрессии



# Реализация фронтенда

Фронтенд-часть приложения построена на React с использованием Material-UI для стилизации и Chart.js для визуализации данных. Она обеспечивает:

- Интерактивные таблицы с фильтрацией, сортировкой и группировкой
- 4 типа графиков для анализа данных
- Прогнозирование зарплаты через форму с динамическим UI
- Адаптивный дизайн и удобную навигацию между разделами



# Результаты

Был разработан веб-сервис для анализа рынка труда аналитиков данных с:

- Разными видами фильтрации, агрегации и группировки для динамических таблиц
- Разными видами графиков для графического анализа ситуации
- Моделью предсказания зарплаты на основе данных



# Планы на будущее

- Добавление новых платформ для сбора данных
- Внедрение дополнительных модулей для аналитики (кластеризация данных, новые виды графиков, временные ряды)
- Улучшение модели линейной регрессии или переход на другой метод (например, дерево решений)
- Добавление статичных данных для получения первичной информации о рынке труда
- Улучшение UX-дизайна
- Автоматическое обновление данных (валюты, новые вакансии/ резюме)
- Возможность импорта данных (как первичных, так и уже подготовленных)
- Извлечение требуемых навыков из описания вакансии
- Загрузка сайта на хостинг

# Демонстрация

**Спасибо за  
внимание!**