

ОТЧЕТ

о проектной работе

по теме: Динамическая модель компетенции

по дисциплине: Проектный практикум

Команда: Команда №2980

Студенты:

Пятков Всеволод Константинович РИ-200023

Ковешникова Татьяна Анатольевна РИ-200023

Веренич Андрей Александрович РИ-200002

Васкевич Юрий Алексеевич РИ-200004

Сундуй Эвелина Зелимовна НМТ-202701

Куратор: Шадрин Денис Борисович

Екатеринбург

2022

**СОДЕРЖАНИЕ**

[**ВВЕДЕНИЕ**](#_heading=h.b223iyalf6nh) **3**

[**ТРЕБОВАНИЕ К ПРОДУКТУ**](#_heading=h.joyc4p2k2724) **4**

[**МЕТОДОЛОГИЯ**](#_heading=h.q6sfeigmr3hv) **6**

[**Получение и подготовка данных**](#_heading=h.bgkmvn7aos7e) **6**

[**Работа с ключевыми словами**](#_heading=h.icih5nfss71l) **8**

[**Вспомогательные программы для построения графиков**](#_heading=h.jj1jxueupxe3) **10**

[**Результат**](#_heading=h.dawbkfy7mczv) **11**

[**ЗАКЛЮЧЕНИЕ**](#_heading=h.mv0noa3ptqk7) **15**

[**ИСТОЧНИКИ ДАННЫХ**](#_heading=h.6xlv1pjjimph) **16**

# ВВЕДЕНИЕ

Сейчас образовательные программы в российских университетах строятся по принципу "что знаем, то и преподаем". Мы хотим изменить подход к созданию образовательных программ, чтобы студенты изучали цифровые технологии, которые востребованы работодателями. Первый шаг на этом пути: узнать, что именно нужно работодателям. Именно для этого разработан наш инструмент - он позволяет загружать данные о востребованных компетенциях из Интернета на регулярной основе.

Целью нашего проекта является - создание инструментария для аналитики и проведение глубокой аналитики выгрузок с сайта hh.ru.

# ТРЕБОВАНИЕ К ПРОДУКТУ

В результате анализа поставленных целей и задач, были выдвинуты следующие требования к проекту:

|  |  |  |  |
| --- | --- | --- | --- |
| **Код** | **Требование** | **Источник** | **Приоритет**  **<Высокий, средний, низкий>** |
| Пользовательские требования | | | |
| Т1\_1 | Вывод графиков. | Заказчик | Средний |
| Т1\_2 | Наличие большого количества вакансий. | Клиент | Высокий |
| Т1\_3 | Актуальная информация. | Клиент | Высокий |
| Функциональные требования | | | |
| Т2\_1 | Наличие функциональной кнопки, с помощью которой можно выбрать интересующий год. | Заказчик | Средний |
| Т2\_2 | Различные виды графиков (столбчатые и линейные) | Команда проекта | Средний |
| Т2\_3 | График сравнения по годам | Заказчик | Высокий |
| Нефункциональные требования | | | |
| Т3\_1 | Скрипт, позволяющий обработать выгрузку с сайта hh.ru. | Команда проекта | Низкий |
| Т3\_2 | Скрипт получения информации для графиков. | Заказчик | Средний |

# МЕТОДОЛОГИЯ

Технологии - Python, JavaScript.

Фреймворки - Chart.js.

Платформа - Web.

Наша гипотеза заключается в том, что мы пытаемся определить по названию профессии ее содержание и проклассифицировать к какой группе ее отнести по названию.

Мы делали графики с 2017, потому что хотели получать более актуальную информацию по рынку труда.

### Получение и подготовка данных

Мы получаем выгрузку готовую выгрузку с вакансиями в формате CSV (<https://alexbers.com/files/vacancy/>).

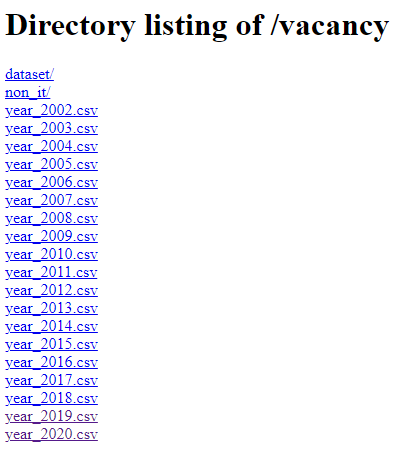


Рисунок 1.1 - Сайт с заготовкой выгрузки с hh.ru

Потом мы использовали скрипт get\_profs\_stat.py, который составлял данные для более удобной работы.

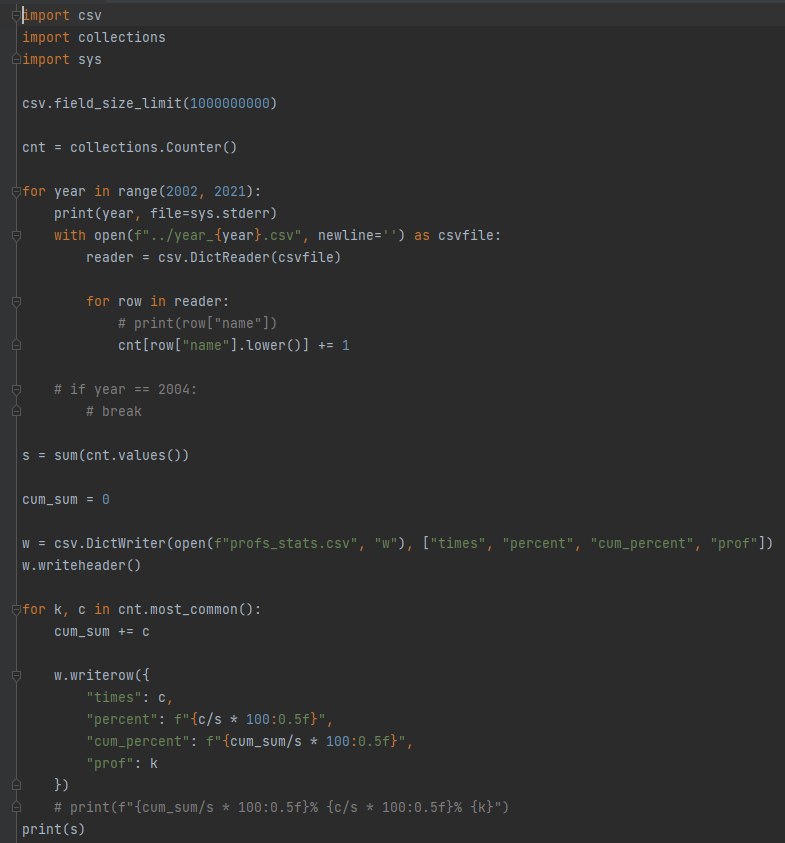


Рисунок 1.2 - Обработчик вакансий

В результате, мы получаем файл со следующими характеристиками:

Частота встречаемости, процент встречаемости, сумма процента встречаемости и название профессии.

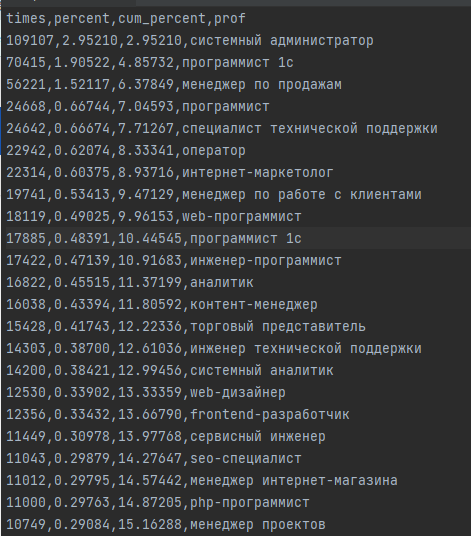


Рисунок 1.3 - Пример составленной базы

### Работа с ключевыми словами

После построения базы, мы начинали сортировку по группам вакансий. Всего мы выделили 20 штук:

* Программист
* Менеджер (IT-проекта)
* ERP
* Системный администратор
* Специалист по интернет-маркетингу
* Аналитик
* Веб
* Специалист техподдержки
* Системный инженер (+QA-инженер)
* Инженер
* Devops
* Mobile - разработчик
* UX/UI дизайнер
* Руководитель
* GameDev
* Специалист по информационной безопасности
* Data Scientist
* IT-специалист
* Администратор баз данных
* Системный программист
* Не относится к IT

Мы смотрели на название вакансии и если оно содержало ключевое слово, например веб, мы вносили ее в одну группу. Если получалось, что в группе много вакансий, мы смотрели на какие подгруппы можно разбить одну большую группу и если в результате остаются большие колонки, выносили их в отдельную категорию. В конечном итоге у нас получилось 20 групп, если не считать неопределенных.

Неопределенные - это те вакансии, которые которые невозможно отнести к какой то конкретной группе, потому что, они являются относятся к сфере IT. Например: Программист инженер.

Не относится к IT - это те вакансии, которые случайно попали в выгрузку. Например: электрик, наладчик, монтажник, техник, оператор и т.д.

В каждой группе профессии находился список подпрофессий:

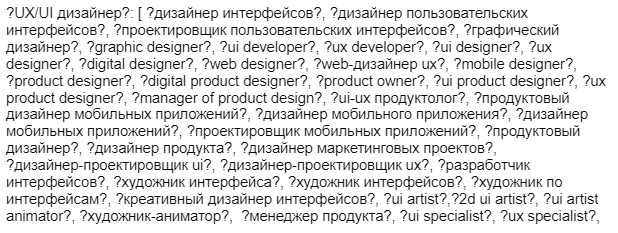


Рисунок 2.1 - Пример вакансий в группе UX/UI дизайнер

После распределения заголовков, мы запускали скрипт allad2.py, который бежал по базе и начинал подсчет количества вакансий для каждой группы.

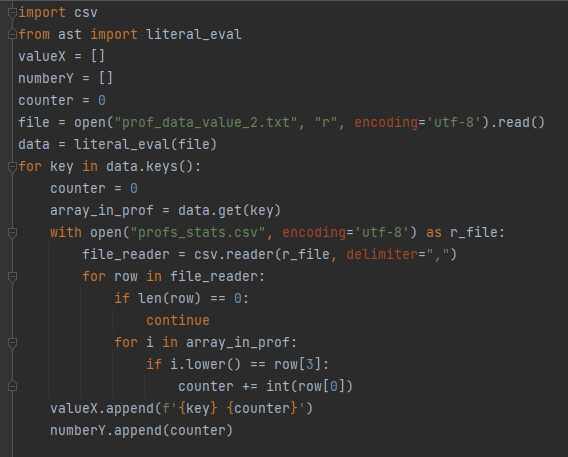


Рисунок 2.2 - Код подсчета вакансий в базе

### Вспомогательные программы для построения графиков

После получения массивов названий и количества упоминаний, мы с помощью Chart.js. В качестве меток мы использовали названия групп, а данные - количество упоминаний в группах. Также мы задавали различные цвета для групп, для более удобного сравнения.

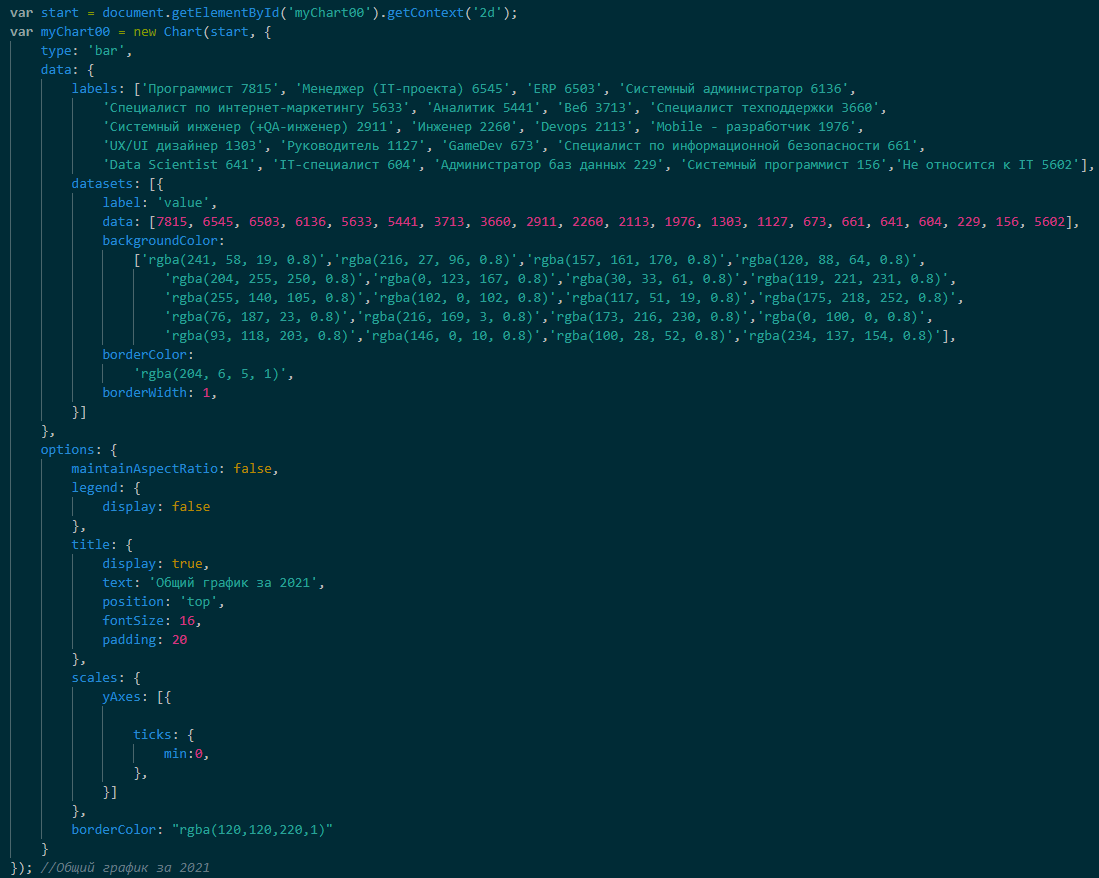


Рисунок 3.1 - Код построения графика “Общий график за 2021”

### Результат

Наша команда реализовала ознакомительную страницу, в которой можно просмотреть результаты аналитики. Сайт прост в использовании, мы выбираем интересующий нас год и сайт выдает нам график.

Первый график - общий. На нем можно увидеть в каком соотношении находится наша база. Можно заметить, что вакансий IT намного больше, чем неопределенных и не относящихся к IT.

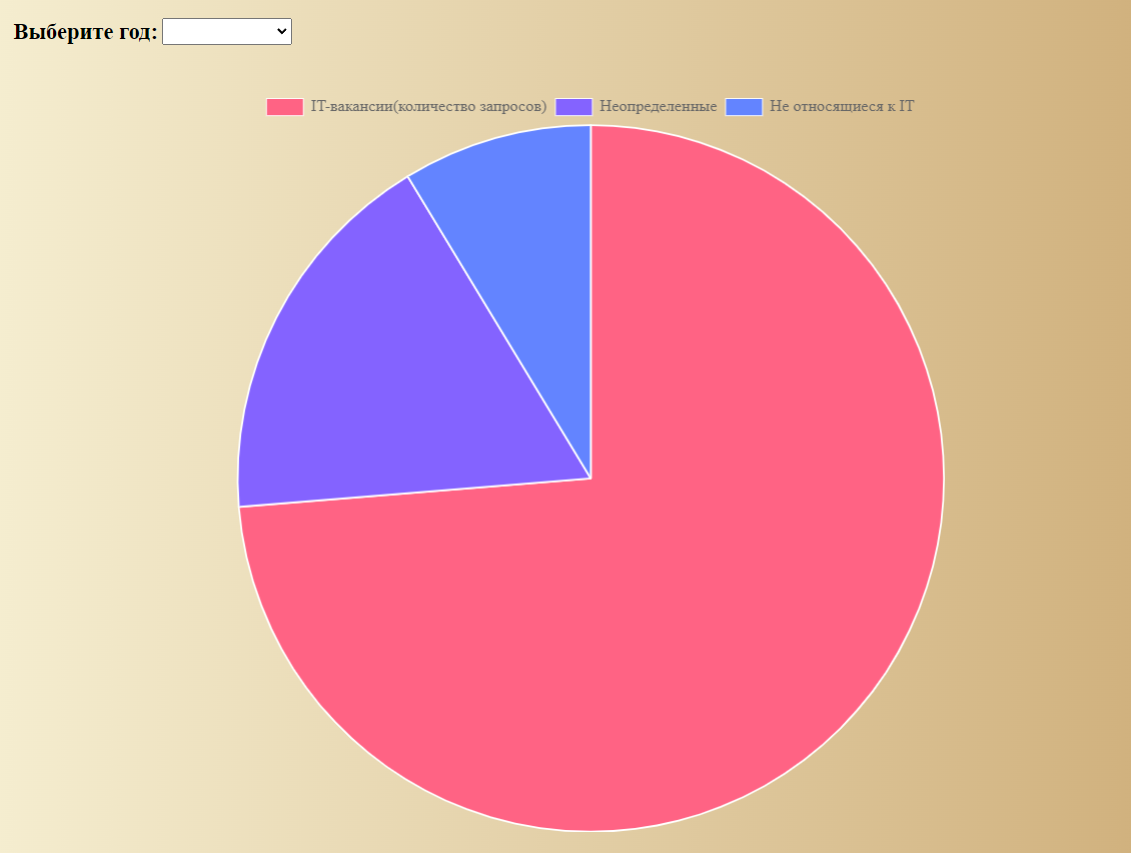


Рисунок 4.1 - График соотношений

Следующая группа графиков - Количество вакансий за определенный год. Предполагается, что пользователь сначала выбирает нужный год, затем выпадает нужный ему график.

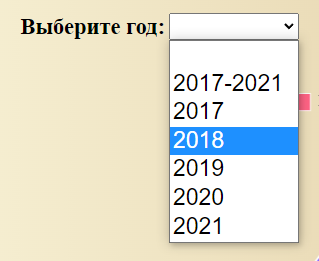


Рисунок 4.2 - Выберите год

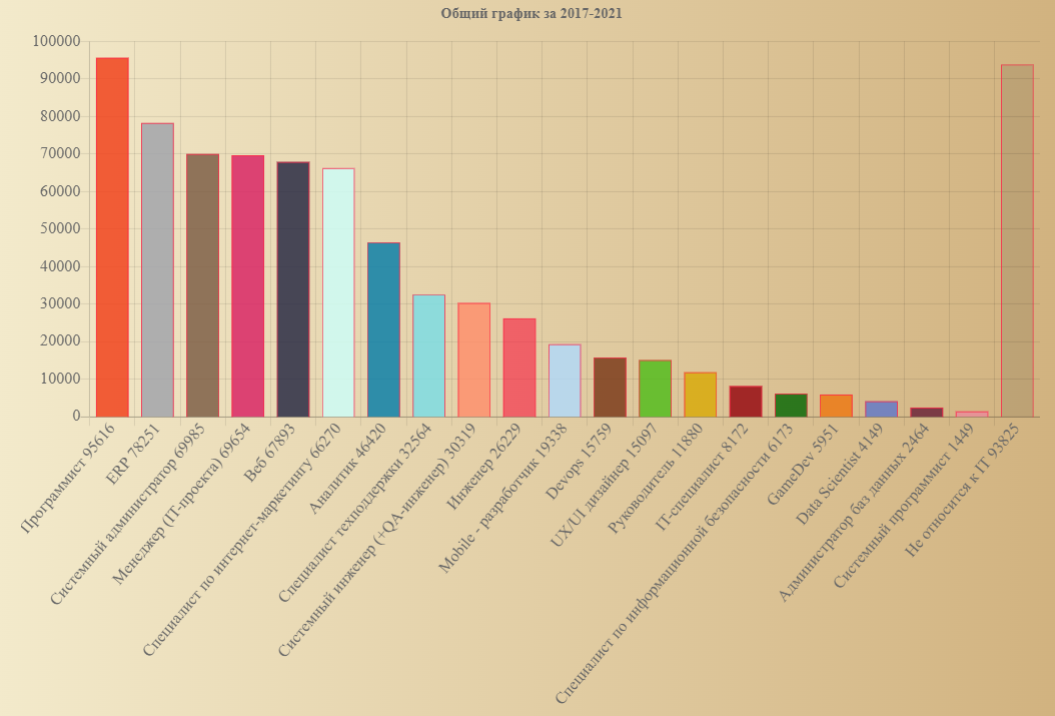


Рисунок 4.3 - График количества профессий в 2017-2021 годах

Следующий график “Динамика”, который показывается, как менялось количество востребованности с 2017 года по 2021. Можно нажимать на названия групп вакансий и они будут исчезать из графика.

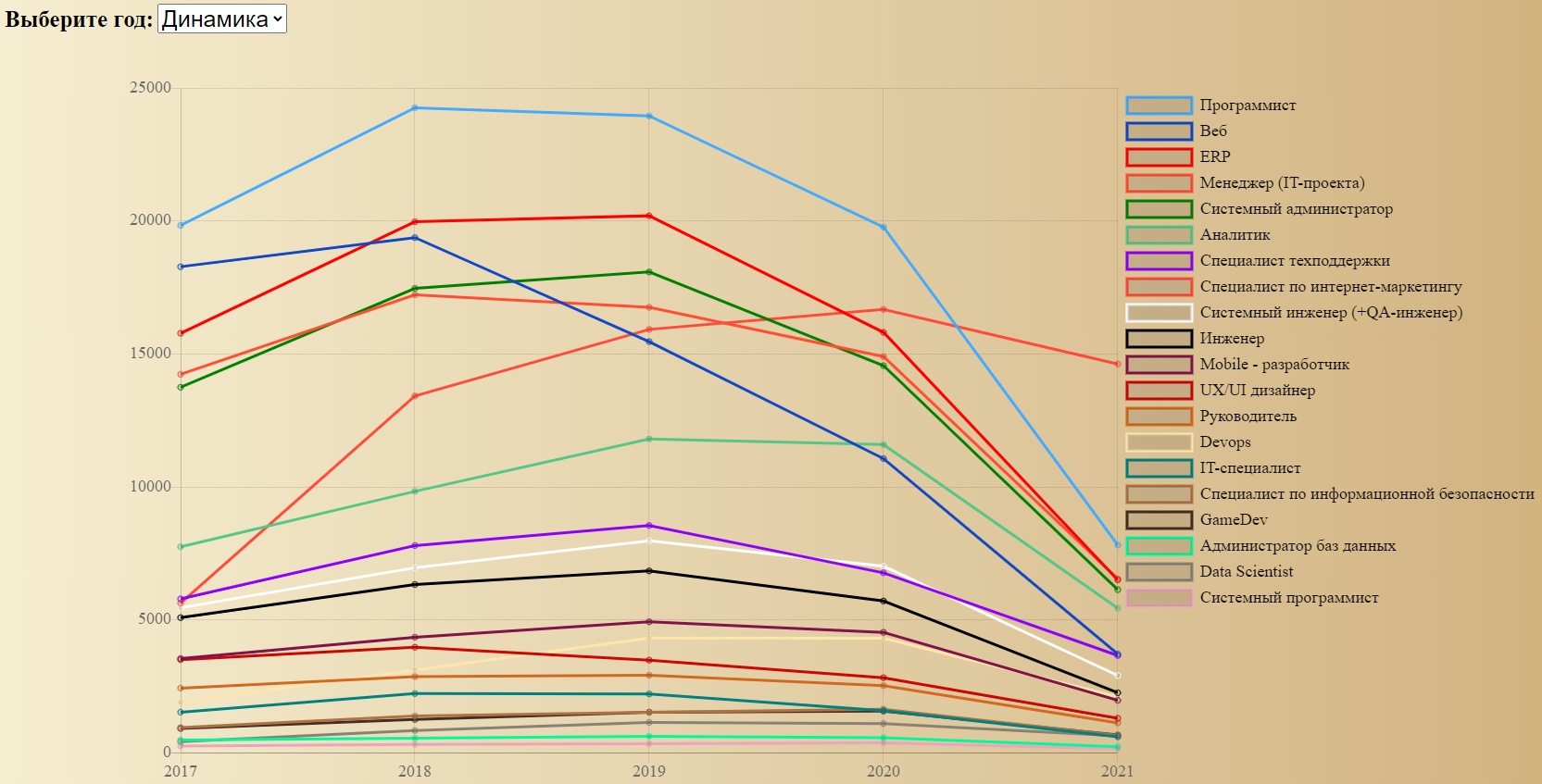


Рисунок 4.4 - Динамика востребованности вакансий

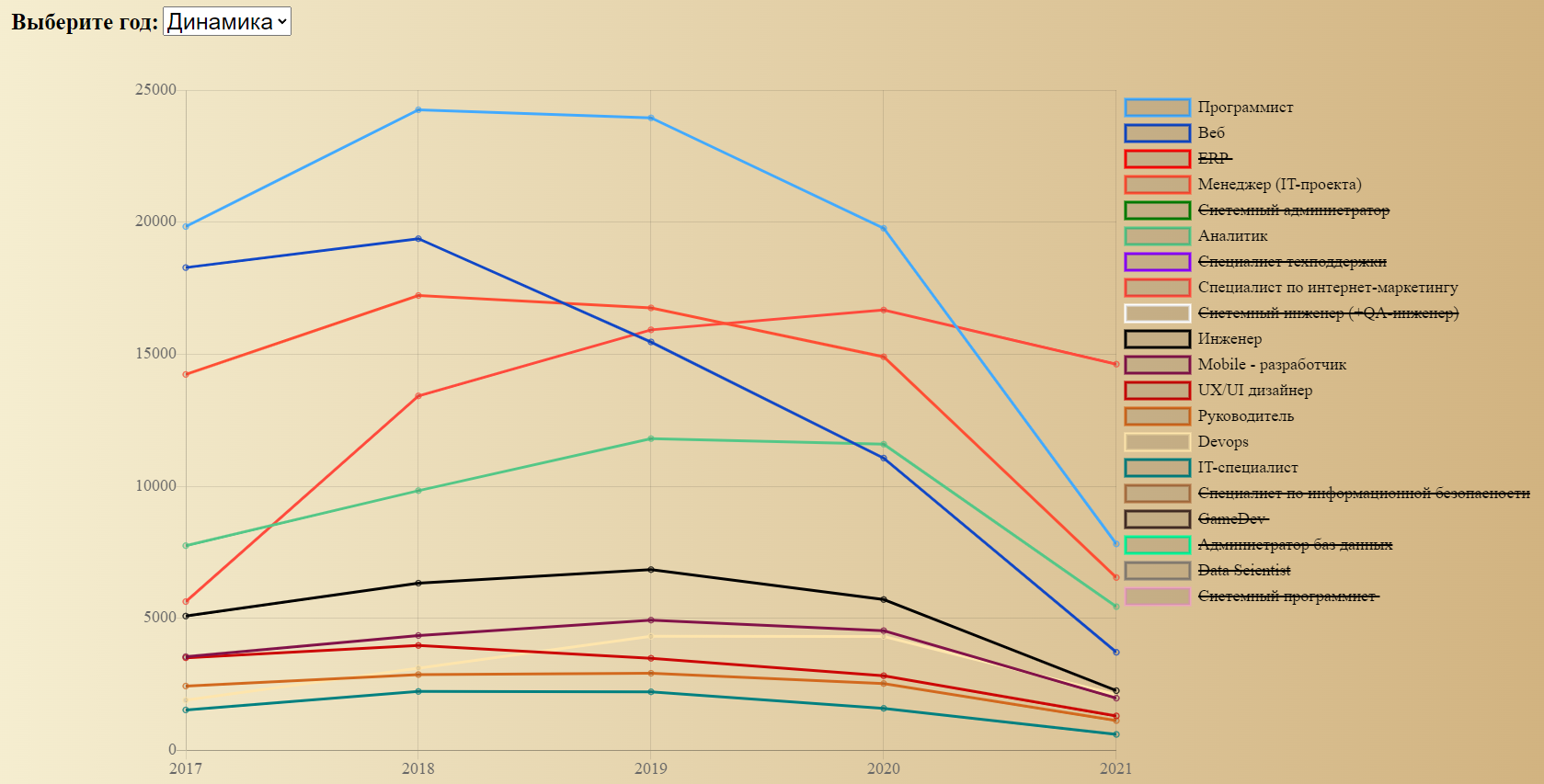


Рисунок 4.5 - Пример исключения групп из динамики

Также можно ознакомиться с соотношениями других годов:

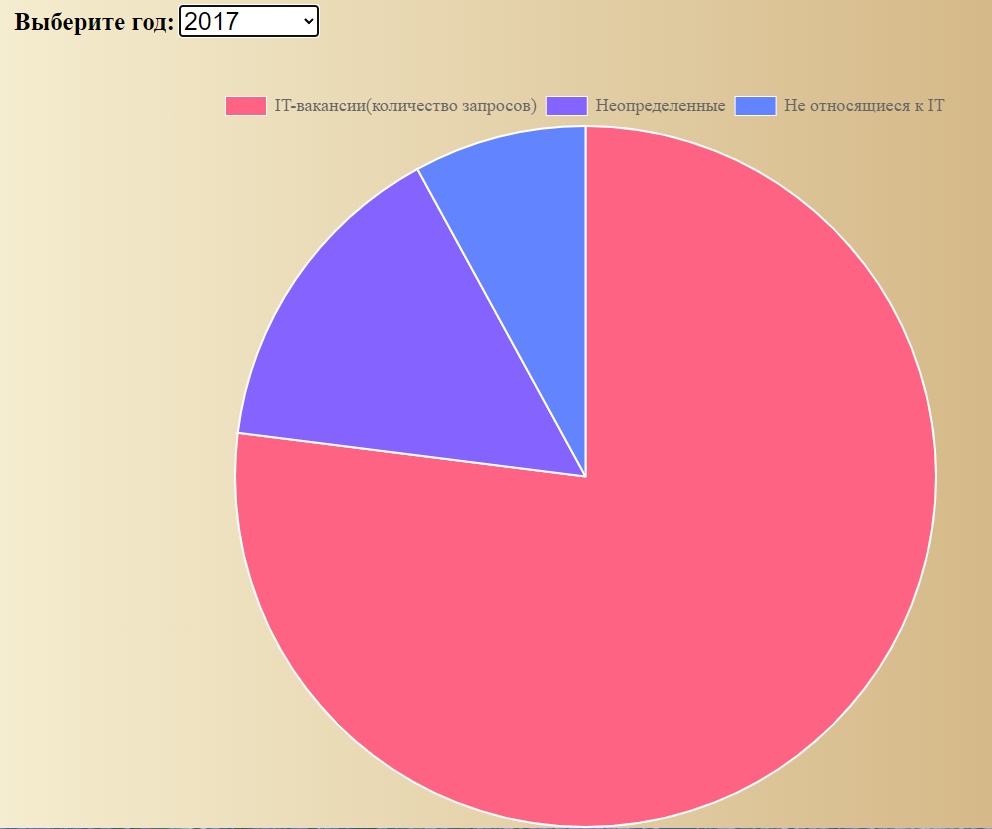


Рисунок 4.6 - Пример соотношений за 2017 год

По результатам аналитики видно, что в базах встречаются вакансии, которые не относятся к сфере IT, также много вакансий, которые носят название “программист”, из-за чего невозможно определить какой специалист требуется данному работодателю. Также встречается много единичных вакансий, потому что работодатель ищет конкретного специалиста и их не так много в базе, например: pay-per-click (ppc) marketing specialist.

# ЗАКЛЮЧЕНИЕ

По итогу всей работы, нам удалось провести глубокую аналитику проекта без реализации сервиса. Сейчас мы имеем более широкое представление о проблеме, знаем пути ее решения. В дальнейшем планируется реализовать сервис, на основе этой аналитики.

Мы смогли подтвердить нашу гипотезу, это подтверждается проделанной аналитической работы.

Каждый из членов команды: тимлид, программисты и аналитики в значительной степени улучшили свои навыки в соответствующих областях.

# ИСТОЧНИКИ ДАННЫХ

1. hh.ru [Электронный ресурс] // hh.ru: <https://ekaterinburg.hh.ru/> - дата обращения: 20.09.2021.
2. Chart.js [Электронный ресурс] // Chart.js: <https://www.chartjs.org/> - дата обращения: 14.10.2021