

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное бюджетное образовательное учреждение высшего образования  
«Тольяттинский государственный университет»

Кафедра \_\_\_\_\_ Прикладная математика и информатика \_\_\_\_\_  
(наименование кафедры)

02.03.03 Математическое обеспечение и администрирование информационных систем  
(код и наименование направления подготовки, специальности)

Мобильные и сетевые технологии  
(наименование профиля, специализации)

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
(БАКАЛАВРСКАЯ РАБОТА)**

на тему: \_\_\_\_\_ «Разработка ПО для отбора потенциальных кандидатов на вакансию по  
\_\_\_\_\_ данным резюме» \_\_\_\_\_

Обучающийся	А.Д. Клишин _____ (И.О. Фамилия)	_____ (личная подпись)
Руководитель	С.В. Митин _____ (ученая степень, звание, И.О. Фамилия)	
Консультант	к.п.н., доцент, А.В. Егорова _____ (ученая степень, звание, И.О. Фамилия)	

Тольятти 2024

## Аннотация

Выпускная квалификационная работа посвящена разработке программного обеспечения для отбора потенциальных кандидатов на вакансию по данным резюме.

Ключевые слова: резюме, отбор кандидатов, служба кадров, анализ данных, кадровая политика.

Объектом исследования данной бакалаврской работы является процесс подбора кандидатов на должность по данным резюме кадровой службой предприятия.

Предметом исследования является программное обеспечение для подбора кандидатов на должность по данным резюме кадровой службой предприятия.

Цель работы - разработка программного обеспечения для отбора потенциальных кандидатов на вакансию по данным резюме для повышения эффективности деятельности кадровой службы предприятия.

В ходе выполнения выпускной работы было проведено исследование предметной области, проектирование программного обеспечения, разработка и тестирование программного обеспечения, расчет эффективности.

Разработанное программное обеспечение обеспечивает эффективное взаимодействие кадровой службы с подразделениями предприятия при формировании условий вакансий и отбору потенциальных кандидатов на вакансию по данным резюме.

Выпускная квалификационная работа представлена на 43 страницах, включает 14 иллюстраций, 35 таблиц, список используемой литературы, состоящий из 20 источников, приложение.

## **Abstract**

The graduate qualification work is devoted to the development of software for selection of potential candidates for a vacancy on the basis of CV data.

Key words: CV, candidate selection, personnel service, data analysis, personnel policy.

The object of research of this bachelor's work is the process of selection of candidates for a position on the basis of CV data by the human resources service of the enterprise.

The subject of the study is the software for selection of candidates for a position on the basis of CV data by the human resources department of the enterprise.

The purpose of the work is to develop software for the selection of potential candidates for a vacancy on the basis of CV data to improve the efficiency of the human resources service of the enterprise.

In the course of graduate work was carried out research of the subject area, software design, development and testing of software, calculation of efficiency.

The developed software provides effective interaction of the personnel service with divisions of the enterprise at formation of conditions of vacancies and selection of potential candidates for a vacancy on the data of a CV.

The graduate qualification work is presented on 43 pages, includes 14 illustrations, 3 tables, list of used literature consisting of 20 sources, appendix.

## Содержание

Введение.....	5
1 Характеристика кадровой службы предприятия .....	8
1.1 Описание деятельности кадровой службы предприятия .....	8
1.2 Обоснование необходимости автоматизации анализа данных резюме в деятельности кадровой службы по работе с вакансиями.....	14
2 Проектирование системы автоматизации анализа данных резюме....	18
2.1 Проектирование системы подготовки данных .....	18
2.2 Проектирование и модульная разработка системы анализа данных .....	26
3 Разработка ПО для анализа данных резюме .....	34
3.1 Разработка приложения.....	34
3.2 Тестирование приложения .....	35
Заключение .....	40
Список используемой литературы и используемых источников .....	41

## Введение

Управление кадровыми ресурсами представляет собой важную составляющую общей эффективности деятельности предприятия.

«Информационные технологии позволяют автоматизировать и оптимизировать процессы управления персоналом, позволяя получать более качественные результаты с минимальными затратами времени. Правильное использование ИТ в HR-менеджменте помогает снизить текучесть кадров путем анализа деятельности компании и оценки эффективности работы персонала. Если говорить об автоматизации данной сферы, важно отметить, что в деятельности, относящейся к типу связи «человек – человек», не все процессы могут быть выстроены по строгому алгоритму, последовательности действий, приводящих к определенному ожидаемому результату. Тем не менее, большинство операций сферы управления персоналом можно автоматизировать.»[5]

Когда-то для поиска нового сотрудника хватало составить объявление о вакантной должности, разместить его в агрегаторе вакансий и прособеседовать пару кандидатов, чтобы одному из них сделать оффер. Однако сейчас на рынке труда всё больше вакансий и соискателей — вручную искать специалистов становится сложнее. Поэтому одним из перспективных направлений в IT-рекрутинге стало использование Big Data: они упрощают поиск и подбор сотрудников, а также процесс управления талантами в компании. [2], [4]

Вместе с тем недостаточное использование современных методов анализа больших данных часто приводит к неэффективному процессу кадрового пополнения, вследствие чего компания либо переплачивает сотруднику, которого могла бы нанять за меньшие деньги, либо не получает приемлемых откликов на вакансию, условия которой ниже рынка труда.

В свете изложенного тема работы - Разработка ПО для отбора потенциальных кандидатов на вакансию по данным резюме – является актуальной.

Часть знаний, подвергающихся исследованию в соответствии с темой работы имеет отношение к деятельности службы кадров в сфере рекрутинга. Таким образом объект исследования – деятельность кадровой службы по подбору новых сотрудников.

Конкретный аспект проблемы, занимаясь рассмотрением которого познается целостный объект, обозначаются и выделяются его характерные свойства в свете темы работы представляет собой программное обеспечение, предназначенное для отбора потенциальных кандидатов на вакансию по данным резюме. Таким образом предмет исследования - программное обеспечение, предназначенное для отбора потенциальных кандидатов на вакансию по данным резюме.

Положительный результат, который может быть получен при устранении недостатка, связанного с публикацией вакансий, не соответствующих требованиям рынка труда (выше рынка), состоит в снижении трудоёмкости службы кадров, связанной с дополнительным отсевом резюме, в случае, когда условия вакансии выше рынка. А с другой стороны, при устранении публикаций вакансий с условиями ниже рынка, организация экономит время, которое тратится на то, чтобы понять что откликов на вакансию не будет и необходимо внести изменения в условия вакансии.

Таким образом, цель работы – повышение эффективности деятельности кадровой службы путем автоматизации обработки данных резюме.

Для достижения цели работы необходимо выполнить следующие задачи:

- Провести анализ деятельности кадровой службы организации,
- Разработать программное обеспечение для анализа данных резюме,

- Провести тестирование разработанного программного обеспечения.

Работа состоит из трех разделов.

В первом разделе дана характеристика кадровой службы предприятия, выполнено описание кадровой службы предприятия, проведено обоснование необходимости автоматизации анализа данных резюме в деятельности кадровой службы по работе с вакансиями.

Во втором разделе выполнено проектирование системы автоматизации анализа данных резюме. Состоящее из проектирования системы подготовки данных и проектирования системы анализа данных.

В третьем разделе выполнена разработка приложения на основе спроектированных систем обработки и анализа данных, разработанное приложение протестировано. Выполнено функциональное тестирование со сравнением результатов автоматической аналитической обработки данных с общепринятыми закономерностями рынка труда.

## **1 Характеристика кадровой службы предприятия**

### **1.1 Описание деятельности кадровой службы предприятия**

Работа кадровых служб предприятий в настоящее время строится на использовании специализированных интернет-сервисов по поиску и предложению работы. [6], [8], [], [] ,

Использование собственного сайта для размещения вакансий нецелесообразно, поскольку посещаемость корпоративных сайтов несопоставима с посещаемостью сайтов по трудоустройству. [3]

Среди сайтов по трудоустройству целесообразно использовать самый крупный и посещаемый.

Использование нескольких сайтов по трудоустройству характерно для кадровых агентств, основная деятельность которых непосредственно связана с подбором резюме по вакансии или наоборот, с подбором вакансий по резюме. При этом полезность кадрового агентства определяется как раз умением сопоставлять информацию из различных источников, в противном случае в их деятельности необходимости не было бы. [7], [9].

Предприятия иной по отношению к кадровым агентствам специализации, как правило, используют единственный сайт по трудоустройству.

При выполнении деятельности в кадровой сфере широко используются информационные технологии. [5], [10], [11], [12], [15], [16], [17], [18], [20].

Далее исследуем предложения в сегменте интернет-услуг по трудоустройству с целью определить наиболее вероятный сервис, который будет использован кадровой службой предприятия в своей деятельности [1].

«По данным рейтинга Similarweb, сервис hh.ru занимает третье место в мире среди сайтов по поиску работы и найму сотрудников. Каждый месяц на hh.ru появляется более полумиллиона вакансий. Найти рабочее место могут как профессионалы в своём деле, так и начинающие специалисты, студенты и



фрилансеры вне зависимости от профессии и места нахождения. На hh.ru можно найти работу по всей России и в любой сфере. Все предложения проходят проверку модераторов.

На сайте могут найти подходящие вакансии специалисты из разных сфер и городов. В основном на hh.ru работодатели ищут кандидатов на полный рабочий день и с опытом работы от года. Однако есть и предложения о стажировках, проектной удалённой работе, также есть предложения для студентов и сотрудников без опыта.»[1]

«На SuperJob много предложений по работе с высоким заработком. В основном работодатели ищут здесь опытных сотрудников. На сайте представлены вакансии не только от крупных частных компаний, но и от государственного сектора, можно найти предложения от российского министерства обороны. У сервиса Superjob.ru есть собственный исследовательский центр, который мониторит российский рынок труда, анализирует заработные платы и востребованность специалистов в разных сегментах.

Сервис подойдет опытным специалистам, которые ищут высокооплачиваемую работу в крупных российских и зарубежных компаниях.»[1]

«На сайте «Работа.ру» собраны предложения для сотрудников «среднего звена». Здесь могут быстро найти работу студенты и люди, которые только начинают свой путь в профессии, а также те, кто ищет подработку или работу вахтовым методом. В партнёрах сервиса несколько крупных компаний, которые готовы предложить студентам стажировку. Сайт полезен тем, кто ищет работу в сфере услуг, торговли, транспорта или строительства.»[1]

«Careerspace создан для тех, кто хочет найти работу с карьерным ростом. На этом сайте от прямых работодателей собраны актуальные вакансии с прямыми контактами нанимателей и чётким описанием необходимых навыков и требований к кандидатам.

Основатели сервиса — авторы популярных телеграм-каналов о карьере: «Без aspera ad astra», «Careerspace» и «Internspace». На Careerspace помимо раздела с вакансиями, есть раздел «карьерные карты». Они помогают определить, на какой карьерной ступени находится специалист и куда ему двигаться, также есть карьерные карты для абитуриентов и тех, кто только выбирает профессию.»[1]

«Больше 12 лет сервис «Зарплата.ру» помогает соискателям найти интересную работу по всей России, а нанимателям – нужных сотрудников. Вакансии постоянно обновляются и проверяются модераторами. На сайте можно найти подборки с предложениями об удалённой или проектной работе.»[1]

«На сайте «Карьерист.ру» можно найти раздел «Лучшие работодатели», где представлены крупные компании, которые чаще всего выбирают соискатели. Также есть список «горячих» вакансий, которые помогают людям находить работу в кратчайшие сроки.

Сайт полезен тем, кто хочет найти рабочее место в кратчайшие сроки. Больше всего вакансий представлено в банковской сфере, логистике, инвестициях и сфере информационных технологий.»[1]

«Trudvsem.ru — это информационная система «Федеральной Службы Российской Федерации по труду и занятости». На сайте представлено более миллиона вакансий от государственных учреждений, центров занятости и крупных коммерческих организаций. Можно найти предложения для студентов и для людей с ограниченными возможностями. Также на Trudvsem.ru можно встать на учёт и в качестве безработного и оформить пособие. Сайт полезен тем, кто ищет работу в крупных компаниях и государственном секторе.»[1]

Достоинства и недостатки рассмотренных интернет-сервисов приведены в таблице 1.

Таблица 1 – Результаты исследования

Сайт	Адрес	Метрика Category Rank / число посещени й за месяц	Достоинства	Недостатки
HeadHunter	<a href="https://hh.ru/">https://hh.ru/</a>	1 / 62,3М	«Есть конструктор резюме Удобная система поиска вакансий — можно сортировать вакансии по уровню дохода, местоположению компании, типу занятости и др.»[1]	«Редкая обратная связь от нанимателей — многие отклики на вакансии остаются не просмотренными и без ответа Отсутствует система оценки работодателей» [1]
SuperJob	<a href="https://www.superjob.ru/">https://www.superjob.ru /</a>	4 / 7.9М	«Большой выбор вакансий для профессионалов Много предложений от зарубежных компаний» [1]	«Мало предложений для тех, у кого нет опыта. Трудно напрямую связаться с работодателем.»[1]
Работа.ру	<a href="https://www.rabota.ru/">https://www.rabota.ru/</a>	9 / 3,1М	«Удобная навигация — грамотно продуман поиск и сортировка подходящих соискателю вакансий Есть предложения о стажировках — люди без опыта могут быстро войти в профессию и прокачать навыки на практике»[1]	«Мало предложений для квалифицированных специалистов — в основном представлены вакансии для специалистов средней квалификации Редко встречаются предложения от топовых нанимателей — большая часть вакансий поступает от маленьких и средних фирм»[1]

Продолжение таблицы 1

Сайт	Адрес	Метрика Category Rank / число посещений за месяц	Достоинства	Недостатки
Careerspace	<a href="https://careerspace.app/">https://careerspace.app/</a>	15 / 50,8К	«Персональная подборка вакансий — сервис учтёт ваш прошлый опыт, желаемую должность и предложит подборку актуальных вакансий»[1]	«Необходимо подтверждение опыта — чтобы получить доступ к вакансиям нужно привязать к аккаунту резюме, составленное на hh.ru, или ссылку на профиль в LinkedIn»[1]
Зарплата.ру	<a href="https://www.zarplata.ru/">https://www.zarplata.ru/</a>	8 / 5М	«Удобный поисковый механизм — легко находить нужную вакансию, есть функция сортировки предложений от работодателей»[1]	«Длительная регистрация — прежде чем откликнуться на вакансию нужно заполнить большую форму персональных данных»[1]
Карьерист.ру	<a href="https://careerist.ru/">https://careerist.ru/</a>	18 / 921К	«Есть вакансии для специалистов с маленьким опытом — много предложений для начинающих специалистов»[1]	«Много предложений от рекрутинговых агентств — часто резюме попадают не напрямую к работодателям, а к посредникам»[1]
Trudvsem.ru	<a href="https://trudvsem.ru/">https://trudvsem.ru/</a>	5 / 3.7М	«Отсутствие рекламы — на сайте нет рекламных баннеров и вакансий, которые продвигают работодатели за	«Регистрация только через портал Госуслуг»[1]

		деньги»[1]	
--	--	------------	--

Для оценки величины ресурса использованы метрики «Место в категории» и «Число посещений за последний месяц», взятые с сервиса Similarweb [19].

Распределение метрики «Число посещений за последний месяц» среди исследованных сайтов показано на рисунке 1.

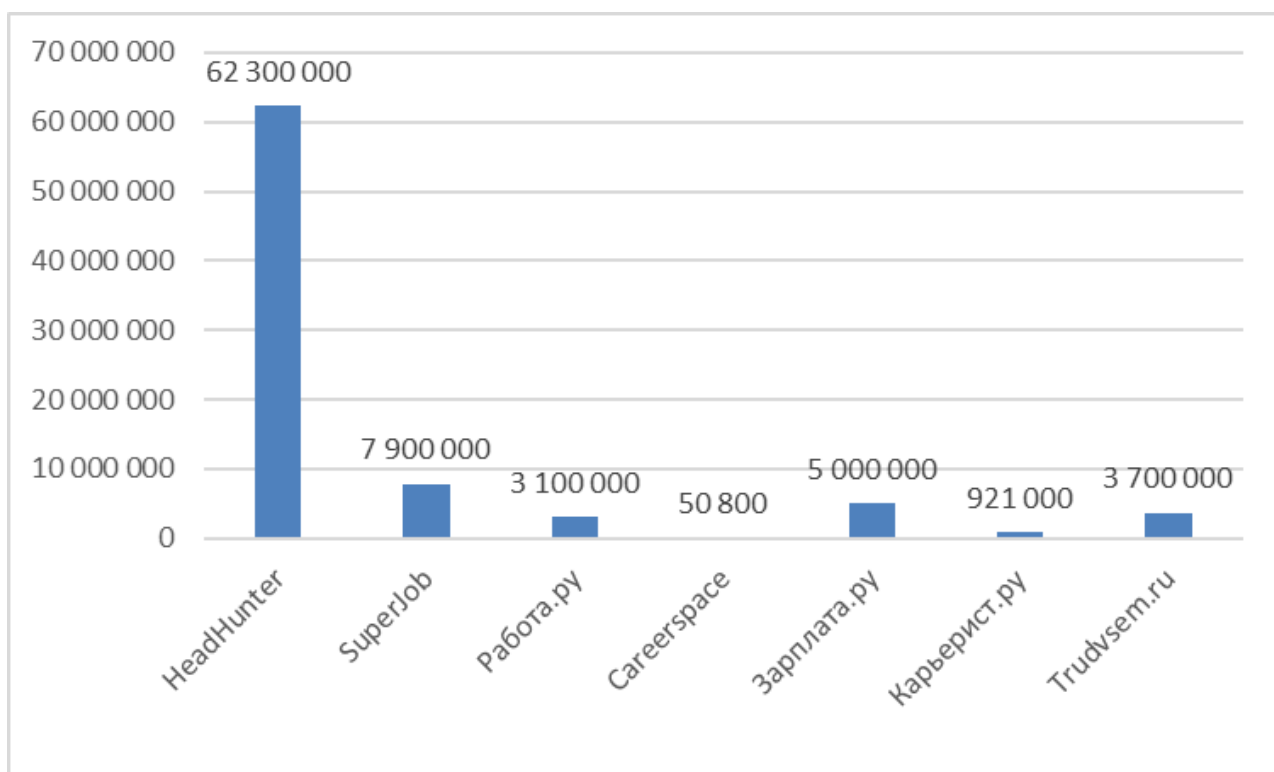


Рисунок 1 – Распределение метрики «Число посещений за последний месяц»

В результате исследования установлено, что все из рассмотренных сайтов имеют свои достоинства и недостатки, но по количеству посещений безусловным лидером является HeadHunter.

Таким образом показано, что использование сайта HeadHunter является наиболее вероятным в деятельности кадровой службы предприятия.

## **1.2 Обоснование необходимости автоматизации анализа данных резюме в деятельности кадровой службы по работе с вакансиями**

Работа с вакансиями в настоящее время выполняется следующим образом: Подразделение, которое имеет потребность в дополнительном сотруднике, направляет заявку с описанием требований к вакансии в службу кадров. Служба кадров оформляет вакансию и размещает на сервисе HeadHunter. По прошествии некоторого времени за которое потенциальные соискатели увидят вакансию и примут решение, наступает отклик. Он может быть выражен в трех вариантах:

- поступило достаточное количество подходящих резюме;
- поступило избыточное количество резюме;
- подходящих резюме не поступило.

В первом случае отдел кадров обрабатывает резюме и направляет соискателей на собеседование в подразделение.

В случае, когда нет подходящих резюме, отдел кадров направляет в подразделение запрос на изменение требований к вакансии с целью обеспечить лучшее соответствие рынку труда. И процесс запускается снова. В случае, если поступило избыточное число подходящих резюме отдел кадров вынужден тратить дополнительное время на предварительный отсев резюме перед обработкой и направлением в подразделение на собеседование.

Модель работы с вакансиями показана на рисунке 2.

В результате анализа модели существующей организации работ с вакансиями выявлено, что при размещении объявлений о вакансиях при назначении условий кадровая служба часто не может учесть текущую ситуацию на рынке труда с учетом всех требований к кандидату, что приводит к следующему:

Если уровень условий ниже, чем обусловлено ситуацией на рынке труда, то на вакансию не откликаются.

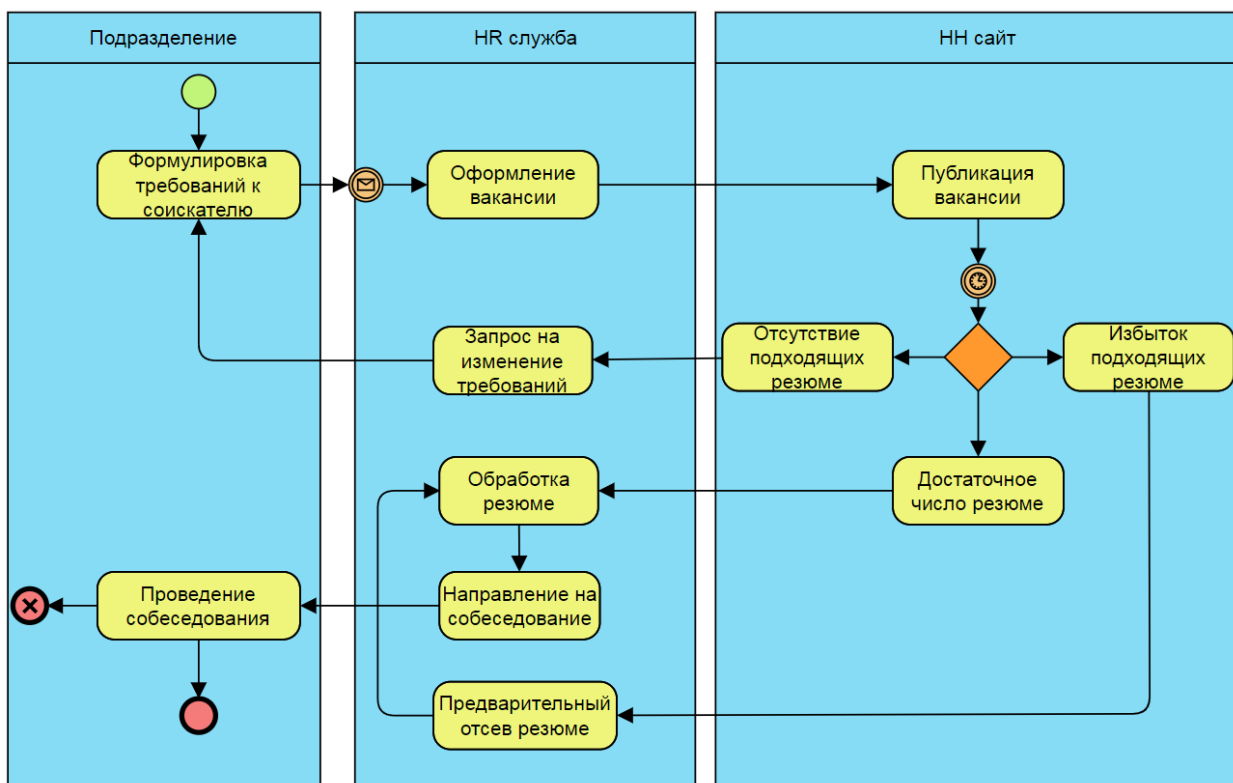


Рисунок 2 – Модель работы с вакансиями

Если уровень условий выше, чем обусловлено ситуацией на рынке труда, то на вакансию откликается слишком много кандидатов, в результате чего:

- кадровая служба испытывает перегрузку при обработке большого количества резюме;
- организация упускает возможность экономии на заработной плате.

Указанные недостатки можно устранить путем применения методов анализа данных при формировании условий вакансии и при обработке резюме. Модель работы кадровой службы с использованием анализа данных резюме показана на рисунке 3. На этом рисунке приведен наиболее

характерный параметр – уровень заработной платы, на условия вакансии этим не ограничиваются и при необходимости наряду или вместо повышения заработной платы могут быть использованы такие параметры как сокращённый рабочий день, снижение требований к опыту работы, снижение требований к образованию.

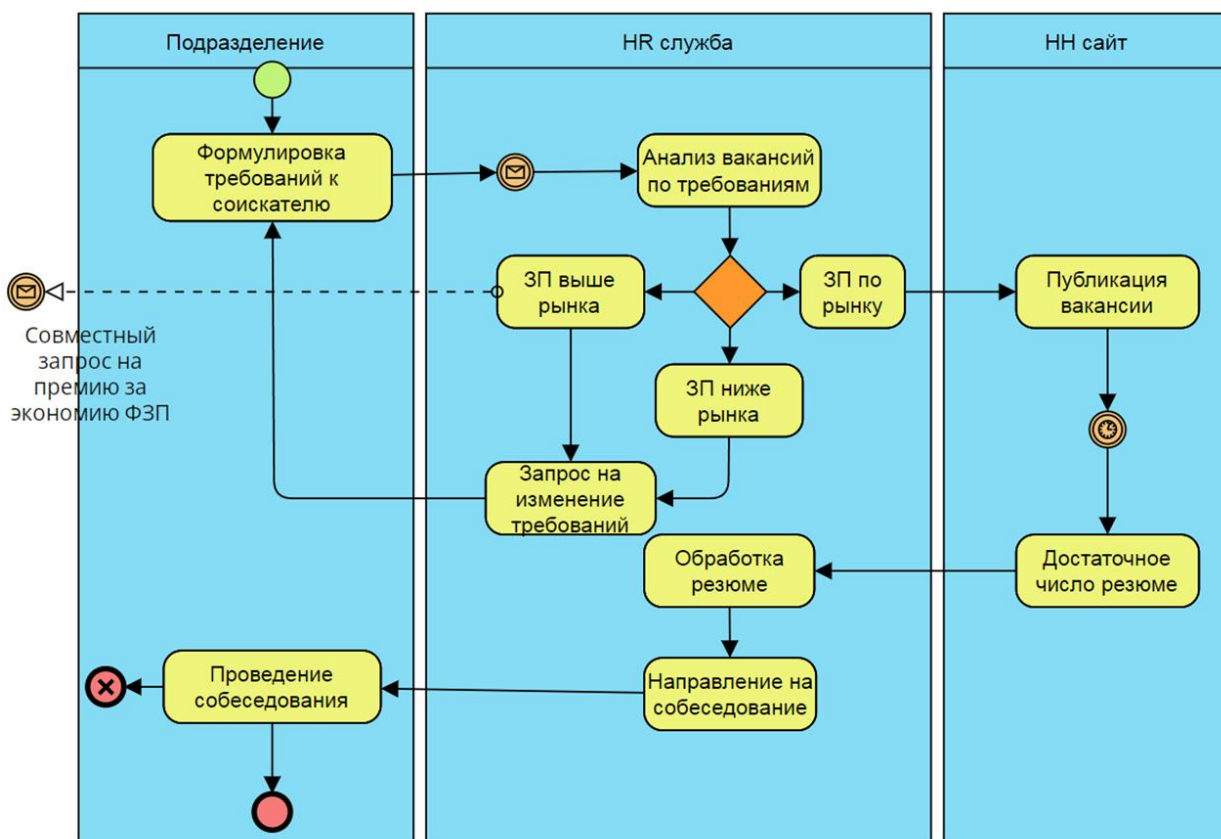


Рисунок 3 - Применение анализа данных резюме при работе с вакансиями

В случае, когда подразделение передаст в кадры описание вакансии с условиями выше рынка – есть возможность корректировки условий в сторону снижения, снизив, например, зарплату. При этом обеспечится экономия фонда заработной платы, поскольку работник будет найден и принят на зарплату меньше, чем изначально закладывалась.

Применение анализа данных резюме при работе с вакансиями обеспечит незамедлительное рассмотрение заявки от подразделения на



соответствие рынку труда. И для корректировки не будет необходимости ждать отклика с сервиса.

Выводы по разделу:

Проведен анализ детальности кадровой службы предприятия, по результатам которого установлено, что в своей работе служба кадров предприятий используют интернет-сервисы для поиска работы.

На основании проведенного интернет исследования установлено, что наиболее вероятным интернет сервисом используемым кадровой службой предприятия в своей деятельности будет HeadHunter.

Разработан алгоритм деятельности кадровой службы в нотации моделирования бизнес-процессов. По результатам анализа полученной модели установлены недостатки, которые могут быть устранены путем внедрения программного обеспечения по анализу резюме.

Разработана модель деятельности кадровой службы с использованием программного обеспечения анализа резюме. В рамках данной модели обеспечивается снижение сроков закрытия вакансии, снижение трудоемкости в кадровой службе за счет соответствия требований резюме текущей ситуации на рынке трудовых ресурсов.

## 2 Проектирование системы автоматизации анализа данных резюме

### 2.1 Проектирование системы подготовки данных

В качестве данных для анализа взят датасет – выгрузка с сервиса HeadHunter [13], [14]. Данные содержат следующие поля:

- пол, возраст;
- ЗП;
- ищет работу на должность;
- город, переезд, командировки;
- занятость;
- график;
- опыт работы;
- последнее/нынешнее место работы;
- последняя/нынешняя должность;
- образование и ВУЗ;
- обновление резюме;
- авто.

Фрагмент датасета, взятого в качестве основы для проектирования аналитического модуля системы показан в таблице 2. Общий алгоритм работы программного обеспечения по анализу данных резюме показан на рисунке 4.

В рамках показанного на рисунке 4 общего алгоритма программное обеспечение для анализа резюме работает на основании датасета, полученного путем парсинга сайта HeadHunter. Парсинг проводится службой ИТ предприятия по заявке от службы кадров или в соответствии с календарным планом. Данный процесс не поддается полной автоматизации, поскольку зависит от изменяющихся во времени настроек hh и проводится ИТ службой вручную.

Таблица 2 – Фрагмент датасета

	Пол, возраст	ЗП	Ищет работу на должность	Город, переезд, командировки	Занятость	График	Опыт работы	Последнее /нынешнее место работы	Последняя /нынешняя должность	Образование и ВУЗ	Обновление резюме	Авто
0	Мужчина, 39 лет, родился 27 ноября 1979	29000 руб.	Системный администратор	Советск (Калининградская область), не готов к...	частичная занятость, проектная работа, полная ...	гибкий график, полный день, сменный график, ва...	Опыт работы 16 лет 10 месяцев Август 2010 — п...	МАОУ "СОШ № 1 г.Немана"	Системный администратор	Неоконченное высшее образование 2000 Балтийск..	16.04.2019 15:59	Имеется собственный автомобиль
44742	Женщина, 48 лет, родилась 26 декабря 1970	40000 руб.	Аналитик данных, Математик	Челябинск, готова к переезду, готова к редки...	полная занятость	полный день, удаленная работа	Опыт работы 21 год 5 месяцев Январь 1998 — по...	ОАО «ЧМК», Исследовательско-Технологический Це...	Начальник группы аналитики	Высшее образование 2000 Южно-Уральский госуда...	09.04.2019 05:07	Не указано
44743	Мужчина, 38 лет, родился 25 апреля 1980	120000 руб.	Руководитель проекта	Москва, не готов к переезду, не готов к кома...	полная занятость	полный день	Опыт работы 15 лет 10 месяцев Руководитель пр...	ПАО ГК ТНС энерго	Руководитель отдела технической поддержк и	Высшее образование 1997 Южно-Российский госуд...	05.07.2018 20:15	Не указано

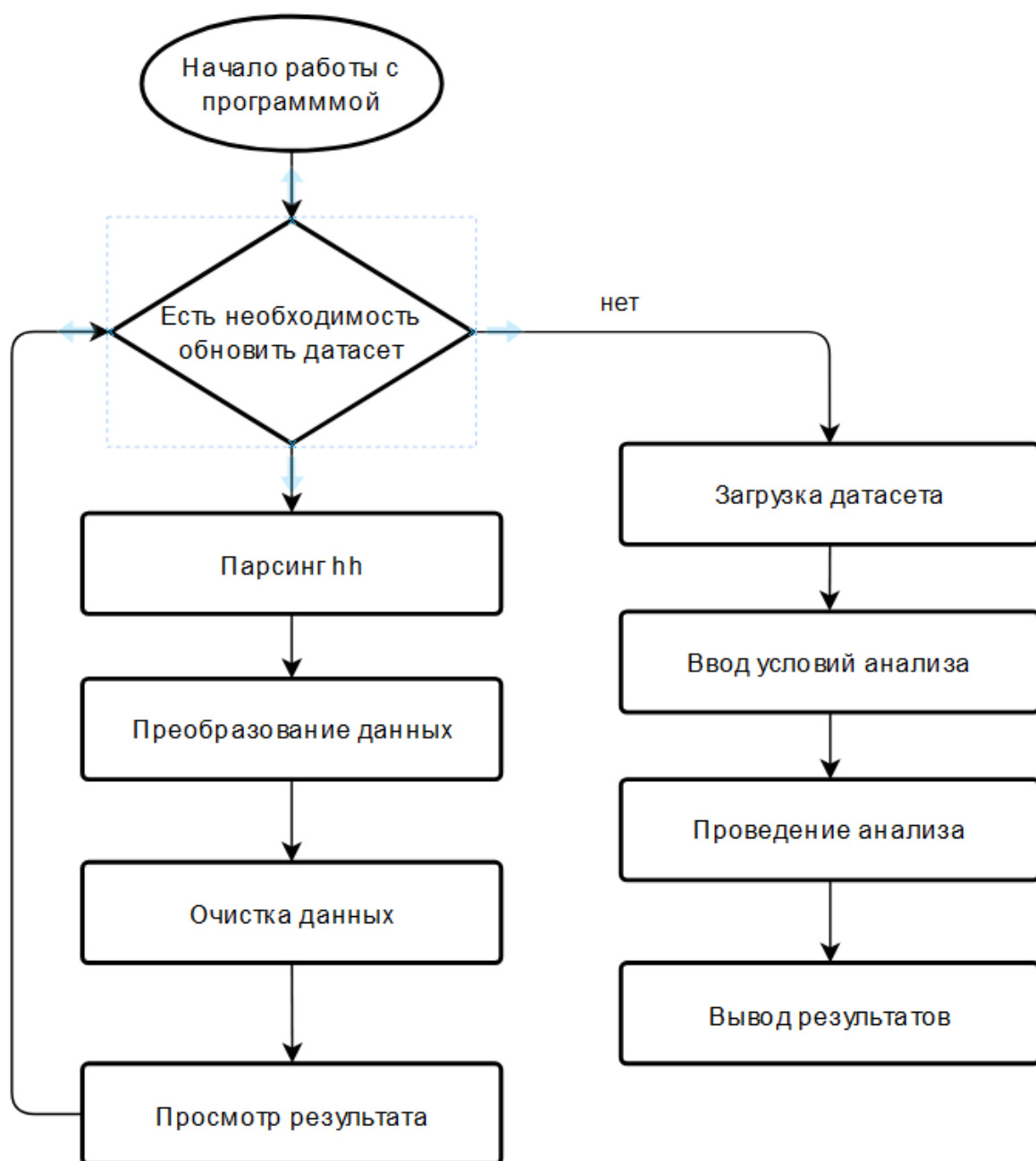


Рисунок 4 - Общий алгоритм работы программного обеспечения по анализу данных резюме

При успешном выполнении на выходе предоставляется объем данных, который в дальнейшем можно автоматически преобразовать к виду удобному для анализа, провести очистку от некорректных данных, просмотреть и подать для проведения анализа.

Далее представлены элементы программного обеспечения, поддерживающие процессы преобразования, очистки и анализа данных.

Выполним предварительный анализ датасета с точки зрения пригодности данных для целей анализа резюме для работы с вакансиями.

Поле «Образование и ВУЗ» имеет формат: «уровень образования, год выпуска, ВУЗ, специальность», например «Высшее образование 2016 Московский авиационный институт (национальный исследовательский университет) инженер-электрик». В контексте решаемой задачи интерес представляет только уровень образования.

Необходимо создать с помощью функции-преобразования новый признак "Образование", который будет иметь 4 категории: "высшее", "неоконченное высшее", "среднее специальное" и "среднее".

Код указанного преобразования (в среде программирования Visual Studio Code) показан на рисунке 5.

```
def get_education(arg):
    if "Высшее" in arg:
        return "высшее"
    elif "Неоконченное" in arg:
        return "неоконченное высшее"
    elif "Среднее специальное" in arg:
        return "среднее специальное"
    elif "Среднее" in arg:
        return "среднее"
    else:
        return pd.NA
hh_df['Образование'] = hh_df['Образование и ВУЗ'].apply(get_education)
print(f"Средний уровень образования имеет {hh_df['Образование'].value_counts}")
hh_df = hh_df.drop('Образование и ВУЗ', axis=1)
```

Рисунок 5 – Код преобразования поля «Образование и ВУЗ»

Столбец «Пол, возраст» представлен в формате «Пол, возраст, дата рождения», например, «Мужчина, 39 лет, родился 27 ноября 1979». В соответствии со спецификой решаемой задачи необходимо выделить каждый

параметр в отдельный столбец. Нужно создать новых признака «Пол» и «Возраст». При этом важно учесть то, что признак пола должен иметь 2 уникальных строковых значения: «М» - мужчина, «Ж» - женщина.

Признак возраста должен быть представлен целыми числами. Код, который выполняет описанное преобразование показан на рисунке 6.

```
def get_sex(arg):
    if "Мужчина" in arg:
        return "М"
    if "Женщина" in arg:
        return "Ж"
def get_age(arg):
    arg_splited = arg.split(" ")
    year_words=["год", "года", "лет"]
    for index, item in enumerate(arg_splited):
        if item in year_words:
            return int(arg_splited[index-1])
hh_df["Пол"] = hh_df["Пол, возраст"].apply(get_sex)
hh_df["Возраст"] = hh_df["Пол, возраст"].apply(get_age)
```

Рисунок 6 – Код преобразования пола и возраста

Признак «Опыт работы» имеет формат «Опыт работы: n лет m месяцев, периоды работы в различных компаниях». В соответствии со спецификой решаемой задачи необходимо выделить общий опыт работы соискателя в месяцах «Опыт работы (месяц)»

Код, дающий столбец, содержащий информацию о том, сколько месяцев проработал соискатель показан на рисунке 7.

```

def get_experience(arg):
    if arg is np.nan or arg == "Не указано":
        return None
    year_words=["год", "года", "лет"]
    month_words=["месяц", "месяцев", "месяца"]
    arg_splited = arg.split(" ")[:7]
    years = 0
    months = 0
    for index, item in enumerate(arg_splited):
        if item in year_words:
            years = int(arg_splited[index-1])
        if item in month_words:
            months = int(arg_splited[index-1])
    return int(years*12 + months)

```

Рисунок 7 – Код преобразования опыта работы

Признак «Город, переезд, командировки» имеет формат «Город, (метро), готовность к переезду (города для переезда), готовность к командировкам», например, «Москва , м. Беломорская , не готов к переезду, не готов к командировкам». В соответствии с решаемой задачей необходимо создать отдельные признаки «Город», «Готовность к переезду», «Готовность к командировкам». При этом признак «Город» должен содержать только 4 категории: «Москва», «Санкт-Петербург», «город-миллионник» и «другие». Признаки «Готовность к переезду» и «Готовность к командировкам» должны иметь два возможных варианта: True или False.

Соответствующий код показан на рисунке 8.

```

def get_city(arg):
    million_cities = ['Новосибирск', 'Екатеринбург', 'Нижний Новгород',
city = arg.split(' ',')[0]
if (city == 'Москва') or (city == 'Санкт-Петербург'):
    return city
if city in million_cities:
    return "город миллионник"
else:
    return "другие"
def get_ready_to_move(arg):
if ("не готов к переезду" in arg) or ("не готова к переезду" in arg):
    return False
elif "хочу" in arg:
    return True
else:
    return True
def get_ready_for_business_trip(arg):
if "командировка" in arg:
    if ("не готова к командировкам" in arg) or ("не готов к командировкам" in arg):
        return False
    else:
        return True
return False
hh_df["Город"] = hh_df["Город, переезд, командировки"].apply(get_city)
hh_df["Готовность к переезду"] = hh_df["Город, переезд, командировки"].apply(get_ready_to_move)
hh_df["Готовность к командировкам"] = hh_df["Город, переезд, командировки"].apply(get_ready_for_business_trip)
hh_df = hh_df.drop("Город, переезд, командировки", axis=1)
mask = hh_df['Готовность к переезду'] & hh_df['Готовность к командировкам']

```

Рисунок 8 – Код преобразования готовности к переезду и командировкам

Признаки «Занятость» и «График» представляют собой набор категорий желаемой занятости (полная занятость, частичная занятость, проектная работа, волонтерство, стажировка) и желаемого графика работы (полный день, сменный график, гибкий график, удаленная работа, вахтовый метод). На сайте hh.ru соискатель может указывать различные комбинации данных категорий. Такой вариант признаков имеет множество различных комбинаций, а значит множество уникальных значений, что мешает анализу. В соответствии со спецификой решаемой задачи необходимо создать признаки-мигалки для каждой категории: если категория присутствует в



списке желаемых соискателем, то в столбце на месте строки рассматриваемого соискателя ставится True, иначе - False. Код, выполняющий данное преобразование, показан на рисунке 9.

```

employments = ["полная занятость", "частичная занятость", "проектная работа", "волонтерство", "стажировка"]
charts = ["полный день", "сменный график", "гибкий график", "удаленная работа", "вахтовый метод"]
for employment, chart in zip(employments, charts):
    hh_df[employment] = hh_df["Занятость"].apply(lambda x: True if employment in x else False)
    hh_df[chart] = hh_df["График"].apply(lambda x: True if chart in x else False)
hh_df = hh_df.drop("Занятость", axis=1)
hh_df = hh_df.drop("График", axis=1)

```

Рисунок 9 – Код преобразования категорий желаемой занятости

Общее преобразование данных показано в таблице 3

Таблица 3 – Преобразование данных датасета

Удаляемое поле	Вновь вводимые поля
Образование и ВУЗ	Образование
Пол, возраст	Пол Возраст
Опыт работы лет, месяцев	Опыт работы месяцев
Город, переезд, командировки	Город Готовность к переезду Готовность к командировкам
Занятость	Полная занятость Частичная занятость Проектная работа Стажировка Волонтерство
График	Полный день Сменный график Гибкий график Удаленная работа Вахтовый метод

Выполнено преобразование данных датасета для облегчения анализа. В частности, разделены поля пол и возраст, опыт работы представлен в единообразном формате – в месяцах, условия качающиеся занятости, графика работы и готовности к командировкам формализованы и из текстовых полей преобразованы в набор логических полей.

## 2.2 Проектирование и модульная разработка системы анализа данных

Система анализа данных имеет целью предоставить возможность специалисту кадровой службы выполнить анализ резюме для оценки соответствия условий размещаемой вакансии текущей ситуации на рынке трудовых ресурсов.

Диаграмма прецедентов (рисунок 10) показывает, что основное взаимодействие с информационной системой, реализующей разрабатываемое ПО будет производить специалист кадровик. ИТ специалист обеспечивает систему «сырыми» данными полученными с сайта hh.

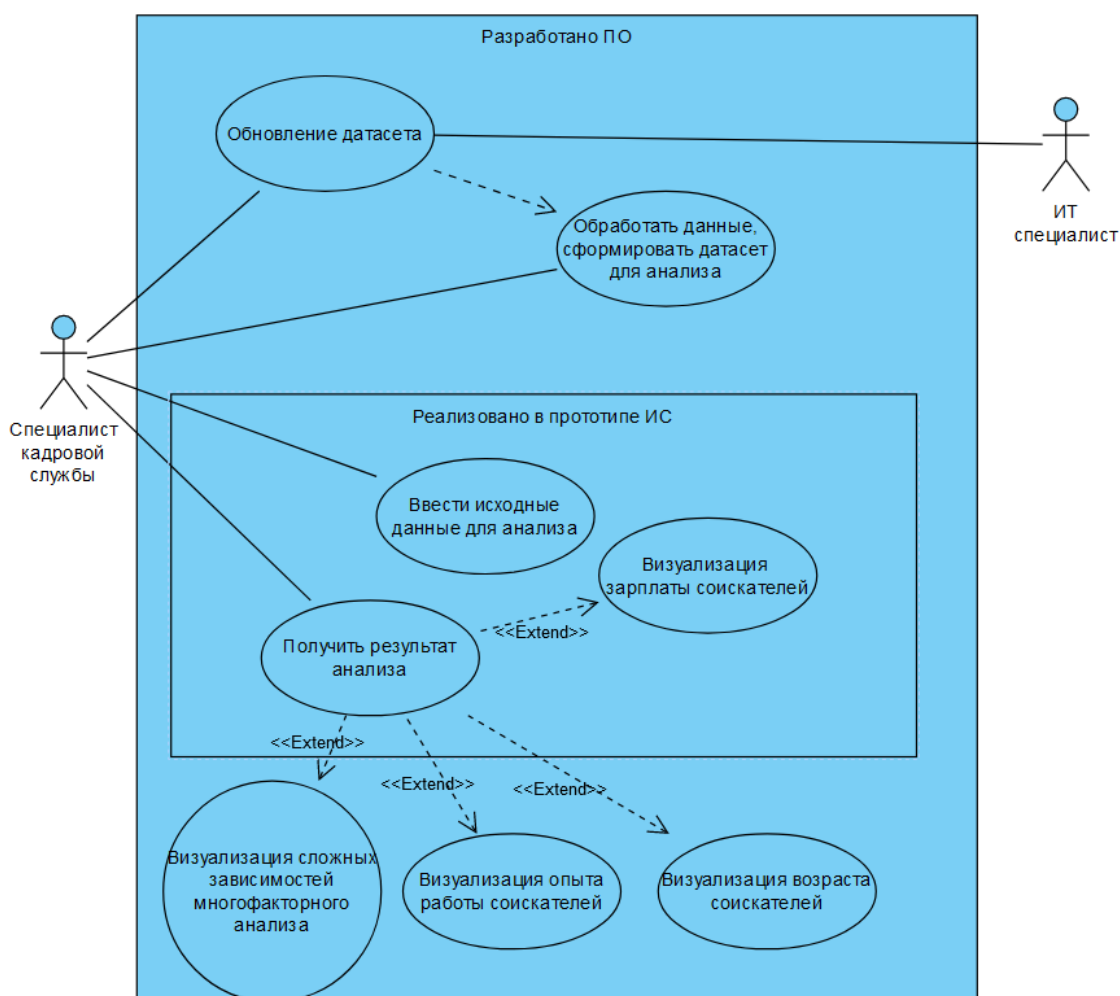


Рисунок 10 – Диаграмма прецедентов

Разрабатываемое ПО реализует прецеденты с рисунка 10. Часть их реализована для демонстрации работоспособности прототипа информационной системы. С дальнейшим развитием систем в нее будут включаться новые дополнительные модули, основанные на разработанном ПО.

Далее разработано программное обеспечение для анализа данных и даны примеры аналитических выводов на основе полученных результатов.

Вывод графика распределения возраста соискателей выполняется следующим кодом (рисунок 11).

```
fig = px.histogram(  
    data_frame=hh_df,  
    x="Возраст",  
    title="Распределение возраста соискателей",  
    histnorm="percent",  
    marginal="box",  
    color_discrete_sequence=['purple']  
)  
fig.show()
```

Рисунок 11 – Код распределения возраста соискателей

Распределение возраста соискателей показано на рисунке 12.

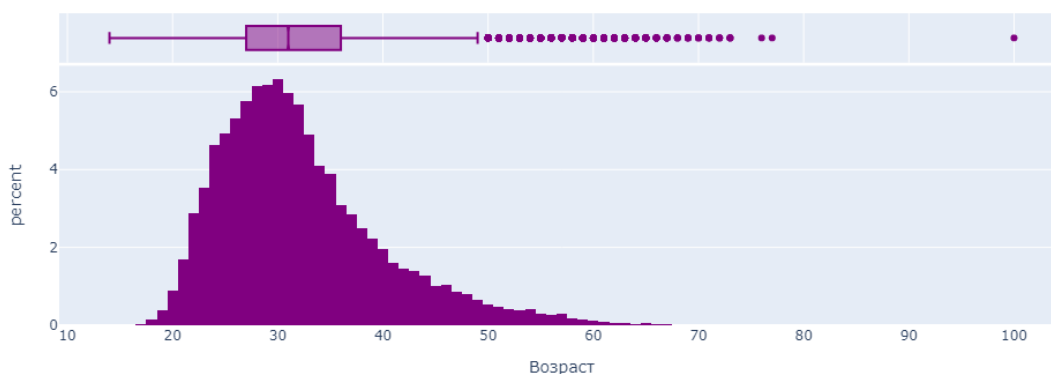


Рисунок 12 - Распределение возраста соискателей

Медианное значение возраста соискателей равно 31 год, мода распределения равна 30 лет. Максимальное значение признака - 100 лет, минимальное - 14 лет. Межквартильный диапазон составляет от 27 до 36 лет.

Гистограмма напоминает логарифмическое нормальное распределение.

Возраст большинства соискателей находится в интервале от 18 до 49 лет. Соискатель с возрастом 100 лет является аномальным значением, поскольку несмотря на опыт работы, в силу физиологических особенностей, человеку сложно поддерживать работоспособность в таком возрасте. Что касается соискателя в возрасте 14 лет, то по трудовому законодательству РФ с этого возраста можно устроиться на работу, для этого понадобится согласие любого из родителей, а также органа опеки и попечительства.

Вывод графика распределения опыта работы соискателей выполняется следующим кодом (рисунок 13).

```
fig = px.histogram(  
    data_frame=hh_df,  
    x="Опыт работы (месяц)",  
    title="Распределение опыта работы соискателей",  
    histnorm="percent",  
    marginal="box",  
    color_discrete_sequence=['purple']  
)  
fig.show()
```

Рисунок 13 – Код распределения опыта работы соискателей

Распределение опыта работы соискателей показано на рисунке 14.

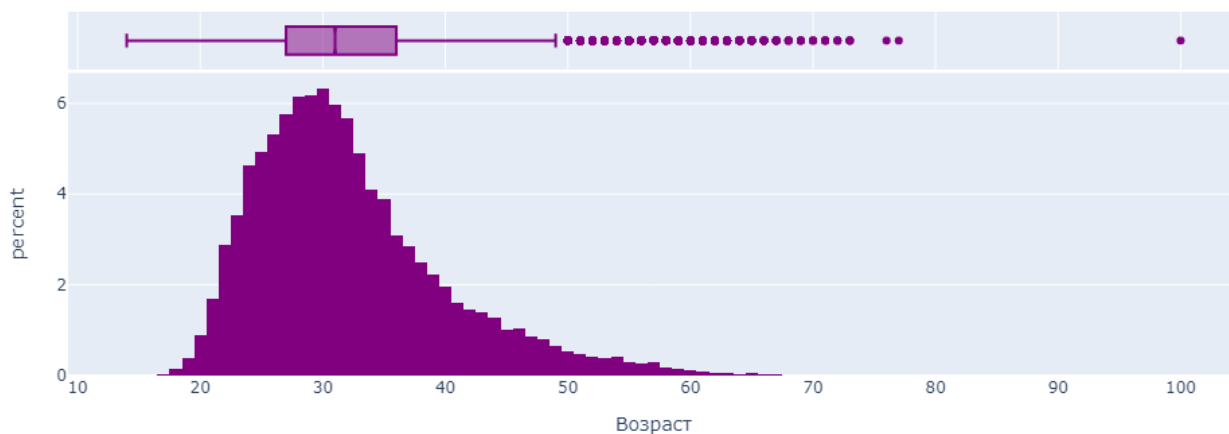


Рисунок 14 - Распределение опыта работы соискателей

Медианное значение опыта работы составляет 100 месяцев, мода распределения находится в интервале от 80 до 84 месяцев. Максимальное значение признака - 1188 месяца, минимальное - 1 месяц. Межквартильный диапазон составляет от 57 до 154 месяца.

Гистограмма напоминает логарифмическое нормальное распределение.

Опыт работы большинства соискателей находится в интервале от 1 до 299 месяцев. Соискатель с опытом работы 1188 месяцев является аномальным значением, поскольку 99 лет рабочего стажа иметь практически никому невозможно. Что касается соискателя с опытом работы - 1 месяц, то это вполне реальная ситуация.

Далее разработаны методы для многофакторного анализа.

Медианное значение желаемой ЗП у соискателей, имеющих высшее образование, составляет - 60 тысяч рублей (а также наибольший разброс), у соискателей с неоконченным высшим, средним и средне специальным - 50 тысяч рублей, 50 тысяч рублей и 40 тысяч рублей соответственно. Зависимость заработной платы от образования показана на рисунке 15.

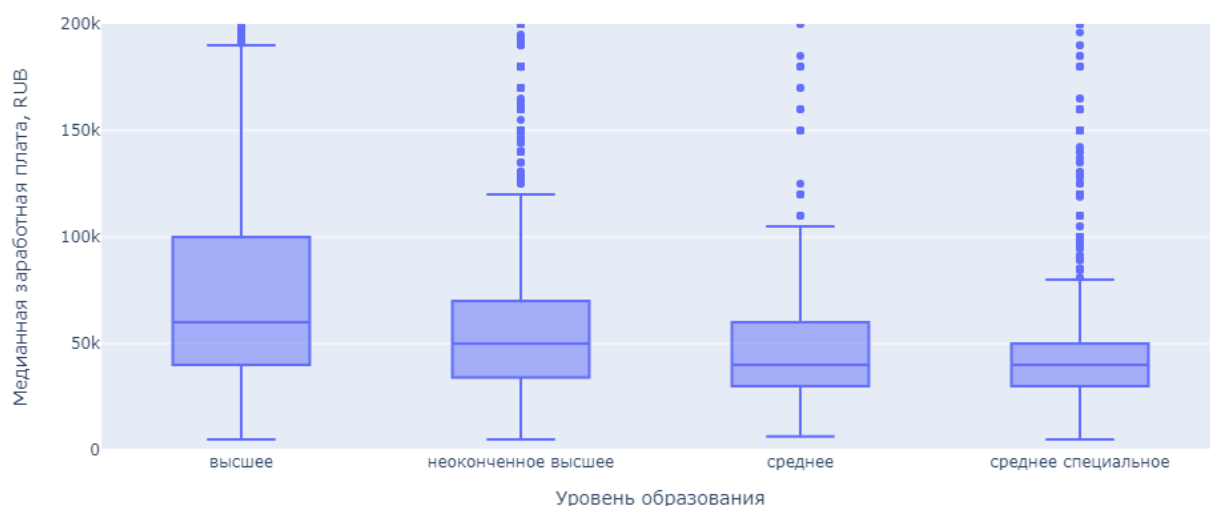


Рисунок 15 - Зависимость заработной платы от образования

В разрезе полового признака значения желаемой ЗП у мужчин на порядок выше, чем у женщин с одинаковым уровнем образования. Так, у мужчин с высшим уровнем образования медианное значение желаемой ЗП в среднем на 20 тысяч рублей выше, чем у женщин и составляет 70 тысяч рублей, в остальных группах по уровню образования - та же тенденция. Зависимость медианной ЗП от образования показана на рисунке 16.

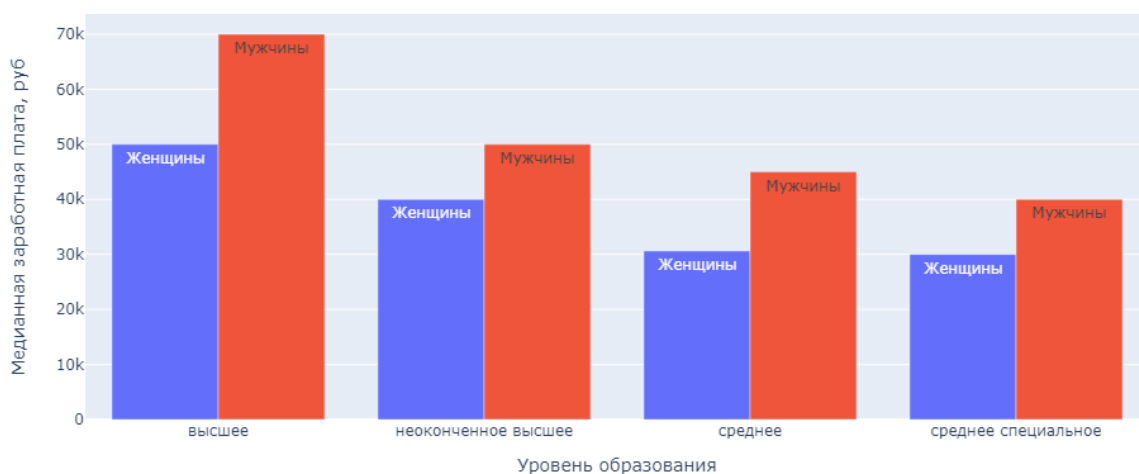


Рисунок 16 - Зависимость медианной ЗП от образования

Медианное значение желаемой ЗП у соискателей, указавших:

- готовность к командировкам и переезду составляет 66 тысяч рублей;
- только готовность к командировкам - 60 тысяч рублей;
- только готовность к переезду - 50 тысяч рублей;
- неготовность к переезду и командировкам - 40 тысяч рублей.

Зависимость медианной ЗП от готовности к командировкам и переезду показана на рисунке 17.

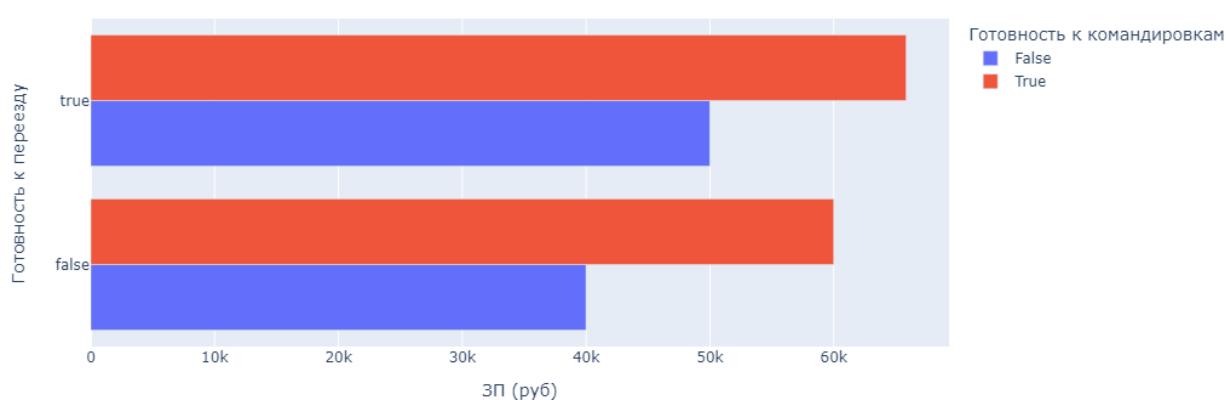


Рисунок 17 - Зависимость медианной ЗП от готовности к командировкам и переезду

Значения медианной ЗП и распределения у категорий полный день и удалённая работа практически идентичны и равны 60 и 100 тысяч рублей соответственно. Затем следует категория гибкий график со значением медианной ЗП - 55 тысяч рублей, вахтовый метод и сменный график по 50 тысяч рублей.

Зависимость ЗП от графика работы показана на рисунке 18.

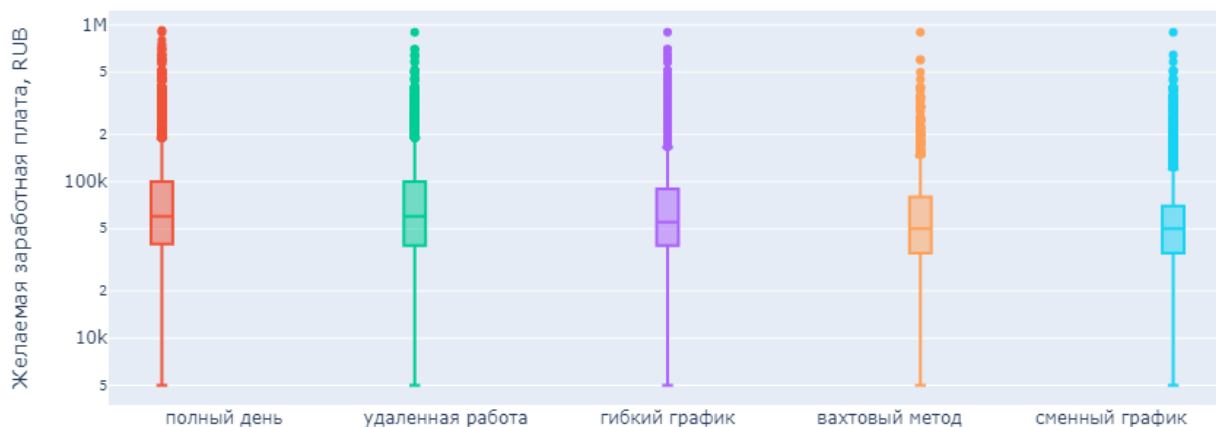


Рисунок 18 - Зависимость ЗП от графика работы

Значения медианной ЗП и распределения у категорий полная занятость и проектная работа практически идентичны и равны 60 и 100 тысяч рублей соответственно. Затем следует категория частичная занятость со значением медианной ЗП - 50 тысяч рублей, волонтерство - 50 тысяч рублей и стажировка - 40 тысяч рублей соответственно.

Зависимость ЗП от занятости показана на рисунке 19.

Распределение заработной платы от занятости соискателей

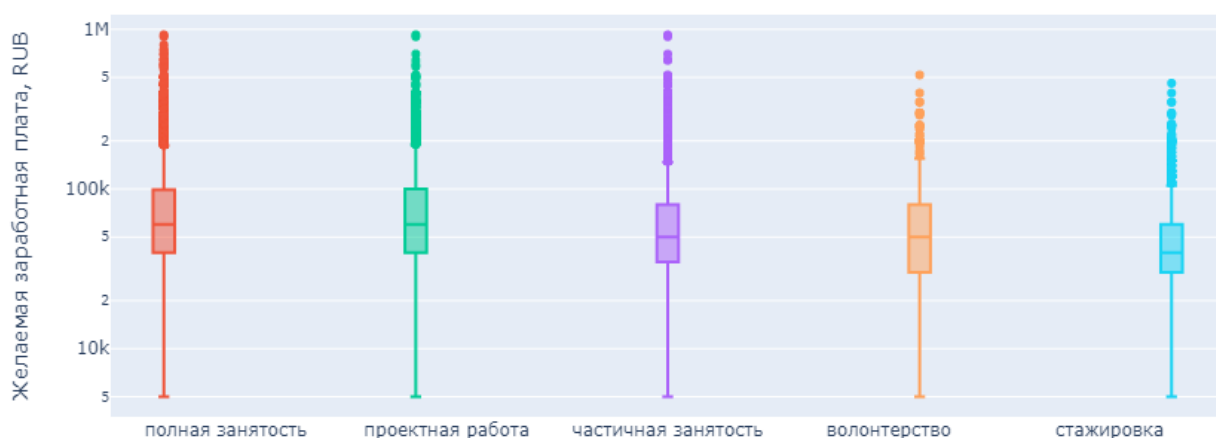


Рисунок 19 - Зависимость ЗП от занятости



На основании полученных исследований формулированы рекомендации по влиянию условий найма на ЗП. Они могут быть эффективны в качестве основы для корректировки требований в случае, когда Подразделение изначально заявило ЗП ниже рынка и не имеет возможности поднять ЗП в вакансии и вынуждено смягчать условия для соответствия своей заявки рынку труда.

Выводы по разделу:

Во втором разделе выполнено проектирование программного обеспечения для отбора потенциальных кандидатов на вакансию по данным резюме.

Выполнено проектирование системы подготовки данных, в рамках которой спроектировано программное обеспечение, позволяющее привести формат выгрузки с сайта Headhunter к виду, удобному для анализа данных резюме с целью подбора кандидатов.

Выполнено проектирование системы анализа данных, которая обеспечивает эффективное представление информации по рынку труда в различных разрезах, обеспечивающую принятие решений при публикации вакансий и своевременное редактирование условий размещаемых и размещенных вакансий для лучшего соответствия текущему состоянию рынка трудовых ресурсов.

## 3 Разработка ПО для анализа данных резюме

### 3.1 Разработка приложения

Разработан вид интерфейса пользователя приложения для анализа данных резюме (рисунок 20).

В левой части отражены характеристики текущего датасета и кнопка позволяющая обновить датасет при необходимости.

Формирование самого датасета выполняется методами парсинга сайта Headhunter.

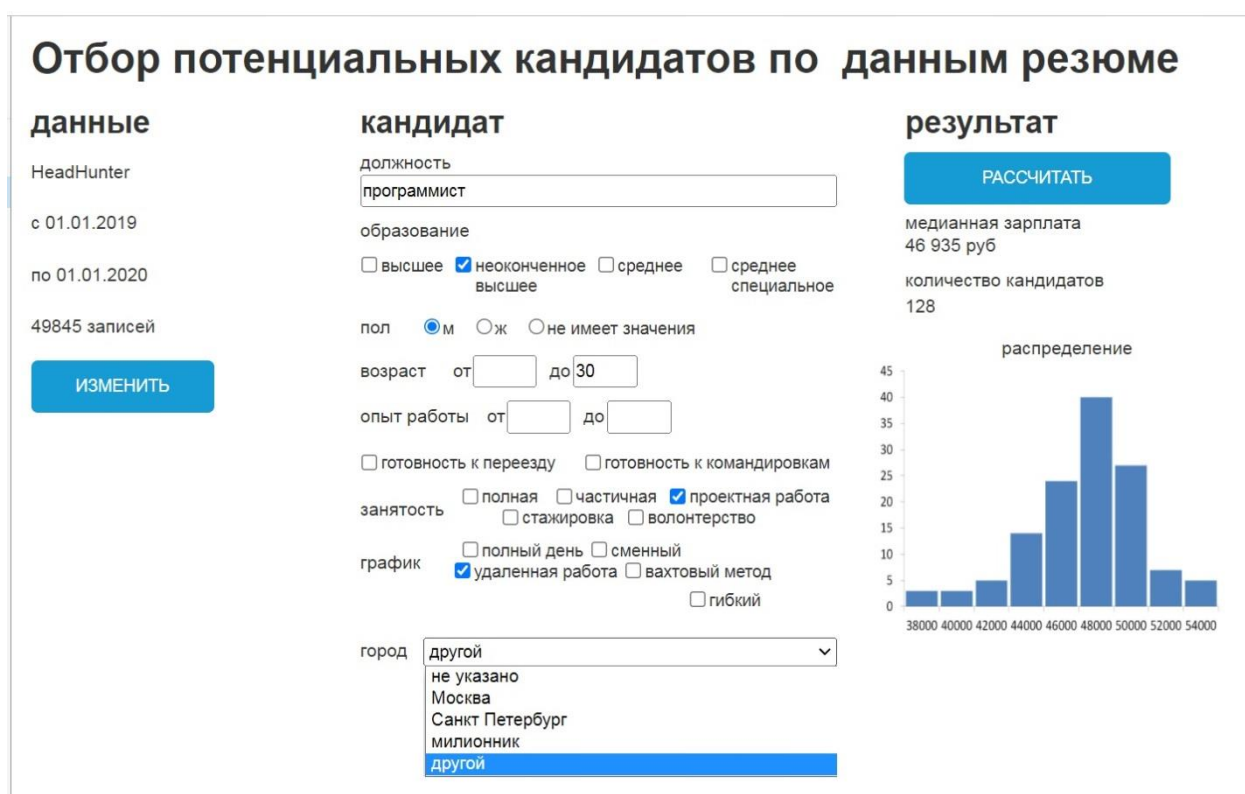


Рисунок 20 - Пользовательский интерфейс приложения

В центральной части задаются требования к соискателю. Незаполненные воспринимаются программой как отсутствие требований по данному критерию.

В правой части представлены результаты анализа датасета, обновляемые по кнопке Рассчитать и содержащие информацию о медианной зарплате, количеству записей, соответствующих требованиям и график распределения зарплаты по отобранным записям.

Разработанный интерфейс является простым интуитивно понятным и обеспечивает требуемый для сотрудника HR службы функционал.

Функционал приложения покрывает наиболее очевидные потребности работника кадровой службы при анализе резюме с целью определения оптимальных условий вакансии по рынку труда – выводит медианную зарплату и гистограмму распределения зарплаты при различных настройках выборки из датасета.

При необходимости функционал может быть расширен с включением более сложных вариантов аналитической обработки данных, программное обеспечение для которых разработано в разделе 2.

### **3.2 Тестирование приложения**

Целью тестирования приложения является проверка вывода аналитики для различных сочетаний исходных данных и определение логичности и непротиворечивости аналитики программы реальным условиям на рынке труда.

Измеряемый параметр – медианная зарплата.

В ходе тестирования необходимо оценить влияние на медианную зарплату варьируемых параметров. Предполагается, что параметры независимые и взаимовлияние параметров отсутствует. Поэтому в каждом эксперименте варьировали один параметр, поддерживая значение остальных на значении «не указано».

Перечень и уровни варьирования параметров представлены в таблице 4.

Таблица 4 – Варьируемые параметры

Параметр	Уровни варьирования
Город	Москва, город-миллионник, другой
Готовность к командировкам	Да, нет
Возраст	до 30, от 30 до 40, от 40 до 50, от 50
Пол	Мужчина, женщина
Образование	Высшее, неоконченное высшее, среднее, среднее специальное

Результаты тестирования показаны в таблице 5.

Таблица 5 – Результаты тестирования

Параметр	Номер эксперимента	Значение параметра	Количество вакансий	Медианная зарплата
Город	1	Москва	3890	68912
	2	город-миллионник	15247	57987
	3	другой	30708	50245
Готовность к командировкам	1	Да	19254	58211
	2	Нет	30591	51463
Возраст	1	до 30	19938	58936
	2	от 30 до 40	13956	65965
	3	от 40 до 50	8473	47040
	4	от 50	7476	37849
Пол	1	Мужчина	27414	56121
	2	Женщина	22430	52019
Образование	1	высшее	28411	58936
	2	неоконченное высшее	1993	26494
	3	среднее	8972	29738
	4	среднее специальное	10467	47040

Пример вывода результатов анализа для четвертого эксперимента серии «Образование» показан на рисунке 21.

## Отбор потенциальных кандидатов по данным резюме

### данные

HeadHunter

с 01.01.2019

по 01.01.2020

49845 записей

ИЗМЕНИТЬ

### кандидат

должность

образование

высшее  неоконченное высшее  среднее  среднее специальное

пол  м  ж  не имеет значения

возраст от  до

опыт работы от  до

готовность к переезду  готовность к командировкам  
занятость  полная  частичная  проектная работа  
 стажировка  волонтерство

график  полный  сменный  вахтовый метод  
 удаленная работа  гибкий

город

### результат

РАССЧИТАТЬ

медианная зарплата  
47 040руб

количество кандидатов  
10467

распределение

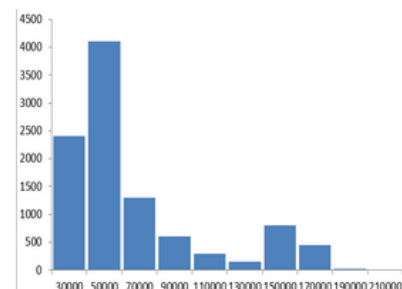


Рисунок 21 – Вывод данных анализа для среднеспециального уровня образования

Анализ результатов по параметру «Город» показывает, что медианные зарплаты в Москве выше, чем в городах-миллионниках, а те в свою очередь выше, чем в остальных городах. Это соответствует действительности.

Анализ результатов по параметру «Готовность к командировкам» показывает, что медианные зарплаты у готовых к командировкам соискателей выше, чем у не готовых. Это соответствует действительности.

Анализ результатов по параметру «Возраст» показан на рисунке 22. Зависимость медианной зарплаты от возраста, полученная в результате анализа данных резюме, соответствует действительности – максимальный уровень соответствует диапазону от 30 до 40 лет, когда оптимальным образом сочетаются опыт и энергичность. Снижение в диапазоне до 30 лет объясняется меньшим опытом работы, а снижение медианной заработной платы в более пожилом возрасте – меньшей энергичностью и большей подверженностью болезням.

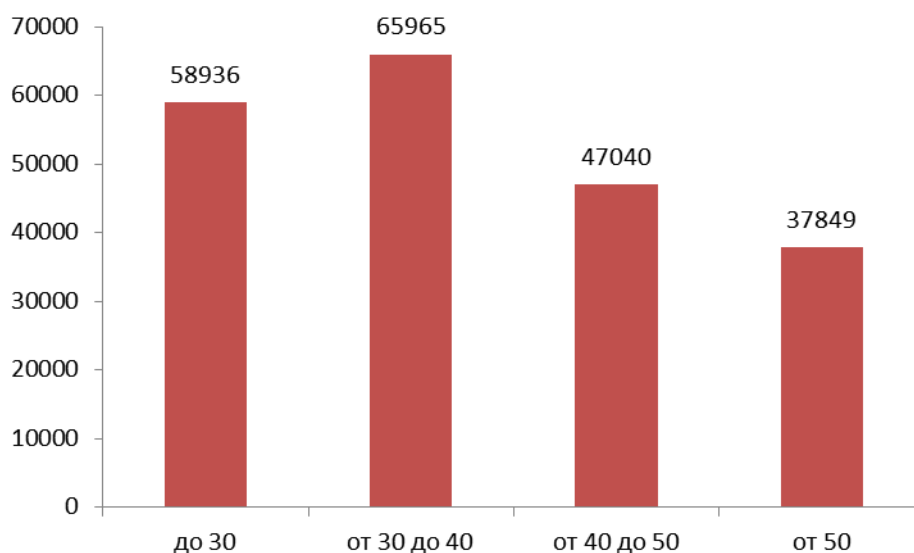


Рисунок 22 – Зависимость медианной зарплаты от возраста

Распределение зарплат по результатам анализа резюме по половому признаку также соответствует фактическому положению дел на рынке труда.

Влияние уровня образования на медианную заработную плату по результатам тестирования программы показано на рисунке 23.



Рисунок 23 – Влияние уровня образования на медианную заработную плату

Самый высокий уровень зарплат имеет место при самом высоком уровне образования – это очевидно. Самая низкая медианная зарплата при неоконченном высшем образовании может быть объяснена тем, что существенную долю соискателей составляют студенты, совмещающие учебу и работу, и на снижение дополнительно влияет фактор неполного рабочего дня и отсутствия опыта. Среди соискателей со средним специальным образованием значительную долю занимают квалифицированные рабочие. Этим объясняется достаточно высокая медианная заработная плата в этой категории.

Выводы по разделу:

В целом результаты тестирования показали, что разработанная программа правильно анализирует данные резюме. Результаты анализа не противоречат общей ситуации на рынке труда. Программа может быть использована в деятельности службы кадров для предварительной оценки условий вакансий на основе анализа данных резюме.

## Заключение

В результате анализа деятельности кадровой службы выявлено, что при работе по поиску не выполняется предварительный анализ резюме, что приводит к повышению трудоемкости/

Проведен анализ детальности кадровой службы предприятия, по результатам которого установлено, что в своей работе служба кадров использует сервис HeadHunter.

По результатам моделирования деятельности по работе с вакансиями и резюме выявлен ряд недостатков, которые можно устранить, внедрив программное обеспечение по анализу данных резюме.

Разработано программное обеспечение для анализа резюме исходя из требований подразделения-заявителя, позволяющее установить уровень зарплаты, соответствующий текущему состоянию рынка труда.

Выполнено проектирование системы подготовки данных, в рамках которой спроектировано программное обеспечение, позволяющее привести формат выгрузки с сайта Headhunter к виду, удобному для анализа данных резюме с целью подбора кандидатов.

Выполнено проектирование системы анализа данных, которая обеспечивает эффективное представление информации по рынку труда в различных разрезах, обеспечивающую принятие решений при публикации вакансий.

В целом результаты тестирования показали, что разработанная программа правильно анализирует данные резюме. Результаты анализа не противоречат общей ситуации на рынке труда. Программа может быть использована в деятельности службы кадров для предварительной оценки условий вакансий на основе анализа данных резюме.

Все задачи работы выполнены, цель работы - повышение эффективности деятельности кадровой службы путем автоматизации обработки данных резюме – достигнута.



## Список используемой литературы и используемых источников

1. Бесплатные сайты для поиска работы [Электронный ресурс] <https://checkroi.ru/blog/sayty-po-poisku-raboty/> (дата обращения 30.05.2024)
2. Глызина М.П., Иванова Е.А. Рекрутинг персонала в условиях цифровой трансформации современных организаций // Вестник Алтайской академии экономики и права. – 2021. – № 6-1. – С. 36-41; URL: <https://vaael.ru/ru/article/view?id=1725> (дата обращения: 18.05.2024).
3. Ермоленко А. А., Косенко И. В. Деструктивные эффекты иллюзорности рынка труда, возникающие при воздействии информационно-коммуникационных технологий на процесс подбора персонала. Часть вторая: описание резюме на онлайн платформе // Российское предпринимательство. 2018. №5. URL: <https://cyberleninka.ru/article/n/destruktivnyye-effekty-illyuzornosti-rynka-truda-voznikayuschie-pri-vozdeystvii-informatsionno-kommunikatsionnyh-tehnologiy-na> (дата обращения: 18.05.2024).
4. Зачем нужны Big Data в IT-рекрутинге [Электронный ресурс]. URL: <https://habr.com/ru/articles/671664/> (дата обращения: 30.04.2024)
5. Идрисова Ж.В., Алихаджиев С.Х., Вагапова М.В. Использование информационных технологий в сфере управления персоналом // Педагогический журнал. 2020. Т. 10. № 1А. с. 398-405. DOI: 10.34670/AR.2020.1.46.147
6. Косенко И. В. Деструктивные эффекты иллюзорности рынка труда, возникающие при воздействии информационно-коммуникационных технологий на процесс подбора персонала. Часть первая: описание вакансий на онлайн-платформе // Российское предпринимательство. 2018. №4. URL: <https://cyberleninka.ru/article/n/destruktivnyye-effekty-illyuzornosti-rynka-truda-voznikayuschie-pri-vozdeystvii-informatsionno-kommunikatsionnyh-tehnologiy-na-1> (дата обращения: 18.05.2024).
7. Никифорова С.В. HR-аналитика для оценки процесса рекрутинга // Форум молодых ученых. 2018. №1 (17). URL:

<https://cyberleninka.ru/article/n/hr-analitika-dlya-otsenki-protssessa-rekrutinga>  
(дата обращения: 18.05.2024).

8. Ребрикова Н.В. Особенности использования рекрутинга в системе управления персоналом организации // Инновации и инвестиции. 2021. №5. URL: <https://cyberleninka.ru/article/n/osobennosti-ispolzovaniya-rekrutinga-v-sisteme-upravleniya-personalom-organizatsii> (дата обращения: 18.05.2024).

9. Стрельникова Л. А., Лембрикова М. М. Актуализация цифровых технологий в управлении процессом подбора персонала // Государственное и муниципальное управление. Ученые записки . 2019. №1. URL: <https://cyberleninka.ru/article/n/aktualizatsiya-tsifrovyyh-tehnologiy-v-upravlenii-protssessom-podbora-personala> (дата обращения: 18.05.2024).

10. Фурсова, А. М. Применение технологий больших данных в сфере подбора и оценки персонала / А. М. Фурсова. — Текст : непосредственный // Молодой ученый. — 2021. — № 51 (393). — С. 118-120. — URL: <https://moluch.ru/archive/393/86860/> (дата обращения: 18.05.2024).

11. Шарапова Н. В., Шарапова В. М., Швецова Д. М. Digital-технологии подбора персонала // КЭ. 2017. №9. URL: <https://cyberleninka.ru/article/n/digital-tehnologii-podbora-personala> (дата обращения: 18.05.2024).

12. Digital Recruitment: Process, Strategies and Best Practices [Электронный ресурс]. URL: <https://content.wisestep.com/digital-recruitment/> (дата обращения: 30.04.2024)

13. HeadHunter [Электронный ресурс]. URL: <https://hh.ru/> (дата обращения: 30.04.2024)

14. HeadHunter vacancies [Электронный ресурс]. URL: <https://www.kaggle.com/datasets/antonbelyaevd/headhunter-vacancies-for-data-search> (дата обращения: 30.04.2024)

15. How to begin with your digital recruitment strategy [Электронный ресурс]. URL: <https://www.hackerearth.com/blog/talent-assessment/begin-digital-recruitment-strategy/> (дата обращения: 30.04.2024)
16. HR Analytics using python. Python. IBM HR Analytics Employee Attrition & Performance [Электронный ресурс]. URL: <https://www.kaggle.com/code/kukreti12/hr-analytics-using-python/> (дата обращения: 30.04.2024)
17. Human Resource Analysis with Python [Электронный ресурс]. URL: <https://thecleverprogrammer.com/2021/01/04/human-resource-analysis-with-python/> (дата обращения: 30.04.2024)
18. Practical Applications of Recruitment Automation in Your Hiring Process [Электронный ресурс]. URL: <https://toggl.com/blog/ai-recruitment-automation/> (дата обращения: 30.04.2024)
19. Similarweb [Электронный ресурс] <https://similarweb.com> (дата обращения 30.05.2024)
20. Waters, S.D. and Streets, V.N. and Johnson-Murray, R. and McFarlane, L. The Practical Guide to HR Analytics: Using Data to Inform, Transform, and Empower HR Decisions. Society for Human Resource Management. 2018 [Электронный ресурс]. URL: <https://books.google.ru/books?id=LOQ-twEACAAJ> (дата обращения: 30.04.2024)