



Оригинальная статья

УДК 331.522

JEL J22, J23

https://doi.org/10.52180/1999-9836_2025_21_2_5_233_244

EDN GLNHGV

Использование открытых данных онлайн-вакансий в сравнении с данными официальной статистики для мониторинга и прогнозирования динамики рынка труда

Виталий Владиславович Алтухов¹, Алексей Дмитриевич Кудрявцев²

^{1,2} Московский государственный университет имени М.В. Ломоносова, Москва, Россия

^{1,2} Профилум, Москва, Россия

¹ (vitaly@profilum.ru), (<https://orcid.org/0009-0000-9307-4276>)

² (kudryavtsev@profilum.ru)

Аннотация

Цифровизация трудовых процессов и рост популярности онлайн-платформ открывают новые возможности для мониторинга и прогнозирования динамики рынка труда. Однако остаются нерешёнными вопросы, связанные с репрезентативностью данных онлайн-вакансий, их оперативностью и полнотой. Научный интерес исследования заключается в развитии подходов к интеграции данных онлайн-источников с официальной статистикой, что позволит повысить точность прогнозирования и оперативность оценки состояния рынка труда. При традиционном анализе ситуации на рынке труда вакансии используются для измерения напряжённости на рынке труда и могут сигнализировать о наличии дисбалансов на рынке труда, когда спрос и предложение не соответствуют друг другу (по качественным характеристикам, географически и т. д.). Целью статьи является сопоставление данных онлайн-вакансий и официальной статистики для разработки подходов к мониторингу и прогнозированию динамики рынка труда. В статье приводится пример реализации мониторинга рынка труда на основании больших данных и сопоставление данных онлайн-вакансий с источниками официальной статистики. Основными источниками данных для сопоставления стали данные Росстата и hh.ru (открытые данные вакансий). При сопоставлении использовалась авторская методология агрегации данных вакансий в группы профессиональных сфер и профессий с опорой на официальные классификаторы, а также методы вычисления и оценки уровня заработных плат. В результате исследования было выявлено, что полученные и агрегированные данные портала поиска работы онлайн hh.ru достоверно соотносятся с официальной поквартальной и помесечной статистикой относительно динамики количества открытых вакансий и заработной платы. В завершении обсуждаются методы прогнозирования динамики рынка труда с использованием методов машинного обучения с опорой на открытые большие данные. Возможность сопоставления динамики показателей онлайн-порталов с официальной статистикой предприятий, по мнению авторов, могла бы дополнить методологию мониторинга рынка труда и повысить достоверность прогнозов.

Ключевые слова: региональный рынок труда, онлайн-вакансии, динамика вакансий, заработная плата, оценка спроса на рынке труда, отрасли экономики, большие данные

Для цитирования: Алтухов В.В., Кудрявцев А.Д. Использование открытых данных онлайн-вакансий в сравнении с данными официальной статистики для мониторинга и прогнозирования динамики рынка труда // Уровень жизни населения регионов России. 2025. Том 21. № 2. С. 233–244. https://doi.org/10.52180/1999-9836_2025_21_2_5_233_244 EDN GLNHGV



RAR (Research Article Report)

JEL J22, J23

https://doi.org/10.52180/1999-9836_2025_21_2_5_233_244

Using Open Data Online Vacancies in Comparison with Official Statistics to Monitor and Forecast Labor Market Dynamics

Vitalii V. Altukhov¹, Aleksei D. Kudryavtsev²

^{1,2} Lomonosov Moscow State University, Moscow, Russia

^{1,2} Profilum, Moscow, Russia

¹ (vitaly@profilum.ru), (<https://orcid.org/0009-0000-9307-4276>)

² (kudryavtsev@profilum.ru)

Abstract

Digitalization of labor processes and the growing popularity of online platforms open up new opportunities for monitoring and forecasting labor market dynamics. However, the issues related to the representativeness of online vacancies data, their timeliness and completeness remain unresolved. The scientific interest of the study lies in the development of approaches to the integration of data from online sources with official statistics, which will improve the accuracy of forecasting and promptness of labor market assessment. In traditional labor market analysis, vacancies are used to measure labor market tensions and can signal the presence of imbalances in the labor market, when supply and demand do not match each other (in terms of qualitative characteristics, geographically, etc.). The purpose of the article is to compare the data of online vacancies and official statistics to develop approaches to monitoring and forecasting labor market dynamics. The article gives an example of implementation of labor market monitoring based on big data and comparison of online vacancies data with the sources of official statistics. The main sources of data for comparison were Rosstat and hh.ru (open vacancy data). The author's methodology of aggregation of vacancy data into groups of professional spheres and professions based on official classifiers, as well as methods of calcula-

tion and estimation of salary levels were used in the comparison. As a result of the study, it was revealed that the obtained and aggregated data of the online job search portal hh.ru reliably correlates with the official quarterly and monthly statistics on the dynamics of the number of open vacancies and salaries. Finally, we discuss methods of forecasting labor market dynamics using machine learning methods based on open big data. According to the authors, the possibility of correlating the dynamics of the indicators of online portals with official statistics of enterprises could complement the methodology of labor market monitoring and increase the reliability of forecasts.

Keywords: regional labor market, online vacancies, vacancy dynamics, wage, estimation of demand in the labor market, economic sectors, big data

For citation: Altukhov V.V., Kudryavtsev A.D. Using Open Data Online Vacancies in Comparison with Official Statistics to Monitor and Forecast Labor Market Dynamics. *Uroven' zhizni naseleniya regionov Rossii=Living Standards of the Population in the Regions of Russia*. 2025;21(2):233–244. https://doi.org/10.52180/1999-9836_2025_21_2_5_233_244 (In Russ.)

Введение

Цифровизация трудовых процессов и рост популярности онлайн-платформ открывают новые возможности для мониторинга и прогнозирования динамики рынка труда. Однако остаются нерешёнными вопросы, связанные с репрезентативностью данных онлайн-вакансий, их оперативностью и полнотой. Научный интерес исследования заключается в развитии подходов к интеграции данных онлайн-источников с официальной статистикой, что позволит повысить точность прогнозирования и оперативность оценки состояния рынка труда.

Цель исследования – сопоставление данных онлайн-вакансий и официальной статистики для разработки подходов к мониторингу и прогнозированию динамики рынка труда.

Задачи исследования:

1. Анализ существующих методологий мониторинга.
2. Разработка алгоритмов обработки данных, включающих в себя классификацию вакансий по профессиональным отраслям и обработку заработных плат.
3. Сравнение данных hh.ru с официальной статистикой Росстата.
4. Оценка результатов, выработка подходов к мониторингу и прогнозированию показателей рынка труда на основе открытых данных.

Объект исследования – показатели структуры и динамики рынка труда (вакансии и заработные платы).

Предмет исследования – данные по показателям (вакансиям и заработным платам из официальных и открытых источников), а также методы и инструменты мониторинга и прогнозирования структуры и динамики рынка труда.

Исследование вакансий представляет собой полезный срез информации и является перспективным источником данных. При традиционном анализе ситуации на рынке труда вакансии используются для измерения напряжённости на рынке труда, то есть для анализа соотношения неудовлетворённого спроса на труд и свободной рабочей

силы (по данным Росстата). Таким образом, данные по вакансиям могут сигнализировать о наличии дисбалансов на рынке труда, когда спрос и предложение не соответствуют друг другу (по качественным характеристикам, географически и т. д.).

Также, согласно исследованиям [1], к 2025 году человечество начнет генерировать 175 зеттабайт данных. Такой большой объём данных мотивирует исследователей активно развивать методы анализа, основанные на анализе больших данных, не ограничиваясь только их частью или статистикой [2].

Гипотезой исследования является применимость данных онлайн-вакансий для мониторинга рынка труда. Различные данные Росстата собираются и подсчитываются ежемесячно или поквартально, однако с учётом развития онлайн-сервисов поиска работы становится актуальным принимать во внимание и их, а также рассматривать возможность соотнесения динамики показателей онлайн-порталов с официальной статистикой предприятий. Такой подход мог бы дополнить методологию мониторинга рынка труда и, возможно, помочь делать достоверные прогнозы на будущее.

Изучая научное поле вопроса достоверности данных порталов по поиску работы онлайн, стоит обратить внимание на ряд исследований, посвящённых сравнению данных официальной статистики (опросов работодателей) с массивом больших данных онлайн-вакансий. На выборке сразу нескольких стран Европейского союза [3; 4] проведены сравнительные исследования распределения вакансий онлайн-платформ и вакансий в отраслевом распределении (согласно европейскому статистическому отраслевому классификатору *NACE*), а также в разрезе профессиональных групп (согласно международной классификации профессий *ISCO-08*) на базе опросов официальной статистики в различных секторах. Авторы отмечают разрывы в данных, исходя из локации (географии стран) и отраслей, – количество вакансий на онлайн-платформах может быть примерно сравнимым с данными официальных опросов, но существуют и случаи несоответствия, когда вакансий может быть больше или значительно меньше в

тех или иных видах деятельности, например, как в случае с государственным управлением, где традиционно меньше онлайн-вакансий. Однако в то же время исследователи отмечают преимущество анализа онлайн-вакансий в силу более оперативного получения информации и возможности охватить различные тренды рынка труда. Исследования, проведённые на выборке данных по отдельным странам – например, по Нидерландам [5], Италии [6] или Австралии [7], – указывают на статистическую значимость сходства между полученными данными онлайн-вакансий и официальной статистикой и являются, согласно выводам, надёжными предикторами трендов на рынке труда.

Теоретические и методологические положения

Исследования репрезентативности данных онлайн-вакансий имеют определенные огра-

ничения, прежде всего в силу того, что не все представители рынка труда имеют равные шансы попасть в выборку в силу разных причин. Вакансии, открытые публично, не представляют весь неудовлетворённый спрос на рабочую силу (рисунок 1). Вакансии, опубликованные в открытом доступе, могут преследовать несколько целей: сделать поиск более эффективным, если другие инструменты этого не позволяют, либо увеличить охват потенциальных кандидатов, повысив таким образом конкуренцию; в некоторых случаях государственные компании обязаны публиковать вакансии на определённых платформах. Некоторые вакансии никогда не становятся публичными, а заполняются иными способами (через внутренние каналы поиска компании, рекрутинговые агентства, неформальные связи и т. д.).

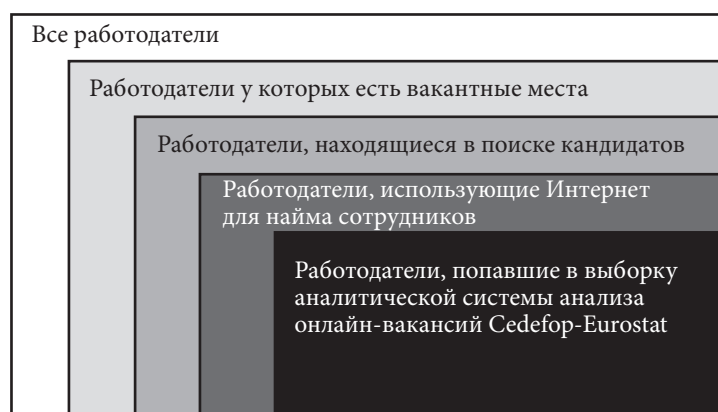


Рисунок 1. Схема механизма отбора онлайн-вакансий из общего количества вакансий на рынке труда в информационно-аналитической системе Cedefop

Figure 1. Scheme of the Mechanism for Selecting Online Vacancies from the Total Number of Vacancies on the Labor Market in the Cedefop Information-Analytical System

Источник: [3].

Онлайн-вакансии являются источником множества данных и содержат: название компании, описание требований, необходимых навыков и образования, заработной платы. В то же время существуют некоторые ограничения или искажения при сборе и интерпретации данных [8].

Согласно ранее изученным данным на европейском рынке труда [9], вакансии, опубликованные на порталах поиска работы онлайн, составляют порядка 80–85% от всех опубликованных вакансий. Больше всего данный способ подходит при поиске менеджеров среднего звена, специалистов разного уровня, технического и административного персонала. Однако могут отсутствовать вакансии для работников низкоквалифицированного труда или, наоборот, для специалистов

высшего звена, которые чаще всего нанимаются через рекрутинговые агентства или нетворкинг.

В исследовании на основе эмпирических данных опроса 1799 российских компаний [10], проведённого в 2015 г., было выявлено, что наиболее популярным способом найма новых сотрудников является поиск через интернет и через социальные сети (знакомых, коллег, родственников). Было выявлено, что порядка 76% компаний из представительной выборки российских предприятий (на момент проведения исследования) используют интернет для поиска сотрудников, и чаще всего в финансовой сфере и трейдинге, реже – в строительстве, промышленности, добыче. Стремительный переход компаний к использованию различных интернет-сервисов и платформ подчёркивает **актуальность** данного исследования.

Есть подтверждения, что на сайтах вакансий недопредставлены низкоквалифицированные и низкооплачиваемые профессии в такой профессиональной группе, как, например, специалисты в сельском хозяйстве, лесной промышленности и др., и излишне представлены профессии, требующие высокой квалификации [11].

Вакансии также могут быть смещены в сторону поиска высококвалифицированных кадров, а также в сторону предложений от компаний, уже успешно зарекомендовавших себя на рынке, или быстрорастущих компаний [12]. Исследования по рынку труда США [13] также говорят о смещённости данных в сторону вакансий, требующих высшего образования (диплом бакалавра), а также о других искажениях при сравнении онлайн-данных и данных официальной статистики. Вакансии в отрасли зависят также от скорости оборота в ней: чем выше оборот, тем больше увеличивается доля представленности – например, в торговле, ресторанном или гостиничном бизнесе оборот выше.

Среди других ограничений – данные открытые вакансий могут зависеть от региональной специфики, популярных ресурсов и платформ, а также от стратегий найма. Например, от возможности открывать вакансию сразу для нескольких кандидатов или для массового найма или, открывая вакансию, задаваться целью мониторинга потенциальных кандидатов, без планов нанимать работников.

На рынок труда может влиять и особенность эволюции институтов труда. Так, стоит упомянуть уже устоявшееся понятие «российская модель рынка труда»: когда в кризисные моменты экономики адаптация происходит в том числе за счёт снижения оплаты труда и сокращения рабочего времени работников [14].

Данные онлайн-вакансий активно применяются в аналитике [15; 16], однако, российский контекст остаётся недостаточно изученным. Основные пробелы включают отсутствие детальных сопоставлений данных порталов онлайн-вакансий с региональной статистикой Росстата. Кроме того, необходимо изучить преимущества и ограничения использования онлайн-источников для мониторинга трудового рынка.

Таким образом, принимая во внимание различные аспекты сопоставления «больших данных» онлайн-вакансий и источников официальной статистики, мы ставим перед собой проблему и задачу выполнить сравнительный анализ этих статистик на российском рынке труда.

Данные и методы работы с ними

На данный момент технологии сбора и обработки данных онлайн-вакансий, то есть «больших

данных», уже достаточно изучены и опробованы. На зарубежном рынке примечателен проект Cedefop, который посвящён сразу 27 странам, входящим в Европейский союз (ЕС), а также Великобритании: распределение вакансий и навыков в отраслевом (по видам экономической деятельности) и профессиональном разрезах¹. Схожий сервис, собирающий данные по вакансиям практически в режиме онлайн, – Lightcast (*бывш. Burning Glass Technologies*)² – уже зарекомендовал себя как надёжный аналитический сервис, использующийся как инструмент для проведения различных совместных исследований рынков труда США и ЕС. Также примечателен российский сервис, созданный «ВНИИ труда» Минтруда России – аналитическая система мониторинга вакансий и резюме³. В целом базовая схема работы заключается в сборе сырых данных, их обработке, преобразовании в структурированную базу данных и визуализацию полученных результатов (например, Cedefop⁴).

Целью нашей работы является соотнесение данных онлайн-вакансий на российском рынке труда с данными официальной статистики Росстата. Важность этой работы определяет дальнейшие шаги для проведения более фундаментальных исследовательских работ: если картина онлайн-вакансий схожа с официальной статистикой, это может в дальнейшем стать валидным инструментом при мониторинге и прогнозировании ситуации на рынке труда.

В своей работе мы будем опираться на несколько **источников**. Из официальной статистики Росстата будут использоваться следующие **показатели**:

1. Заявленная работодателями потребность в работниках в течение отчётного периода (количество свободных вакансий, заявленных работодателями в органы службы занятости населения в течение отчётного периода, по данным служб занятости, доступно по РФ в целом, ежемесячно)⁵.

¹ Cedefop: Skills-OVATE: Online Vacancy Analysis Tool for Europe // Cedefop: [сайт]. URL: <https://www.cedefop.europa.eu/en/tools/skills-online-vacancies> (дата обращения: 04.11.2024).

² Lightcast Data: Basic Overview // Lightcast: [сайт]. URL: https://kb.lightcast.io/en/articles/6957498-lightcast-data-basic-overview#h_31fa7af939 (дата обращения: 04.11.2024).

³ ВНИИ труда представил аналитическую систему мониторинга вакансий и резюме // ВНИИ труда: [сайт]. URL: <https://vcot.info/news/vnii-truda-predstavil-analiticeskuyu-sistemu-monitoringa-vakansij-i-rezume> (дата обращения: 04.11.2024).

⁴ Cedefop. Online job vacancies and skills analysis: a Cedefop pan-European approach. Luxembourg: Publications Office, 2019 // Cedefop: [сайт]. URL: <http://data.europa.eu/doi/10.2801/097022> (дата обращения: 04.11.2024).

⁵ Росстат. Заявленная работодателями потребность в работниках в течение отчетного периода // Росстат: [сайт]. URL: <https://www.fedstat.ru/indicator/43141> (дата обращения: 04.11.2024).

2. Численность требуемых работников списочного состава на вакантные рабочие места на конец отчётного квартала (численность работников списочного состава, которых предполагается принять на вакантные рабочие места, по состоянию на последнее число отчётного квартала, по данным отчётности сектора крупных и средних предприятий, доступна по регионам и отраслям)⁶.

3. Среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по субъектам Российской Федерации⁷.

Эти показатели будут сопоставлены с собранными и проанализированными «Профилум» открытыми данными платформы по найму сотрудников hh.ru, крупнейшей платформы поиска работы и найма сотрудников в России⁸, которая предоставляет открытый и публичный НТТР API. Период для анализа данных – с января по декабрь 2023 года.

Можно выделить следующие **этапы сбора вакансий и выявления показателей заработной платы**:

1. Обработка названий вакансий и их общего количества. Было собрано более 16 млн. вакансий по всем регионам страны. В текстах каждой вакансии проводился процесс удаления служебных частей речи и выделения лемм (начальных словарных форм слов). Затем вакансии были сгруппированы по названиям, тем самым был получен первоначальный список профессий. Также был экспертно сформирован список из двух тысяч профессий, основанный на анализе вакансий (наиболее часто встречаемые) и общероссийском

классификаторе специальностей. Далее специалистами по рынку труда и профориентации были сопоставлены первоначальный и экспертный список профессий. Таким образом были классифицированы вакансии для дальнейшего анализа.

2. Обработка заработных плат. Анализ данных по средней заработной плате указывает на проблему некорректно низких или высоких значений (выбросов) и недостаточной заполненности данных: доля пропусков, где отсутствует хотя бы один из пределов (минимум или максимум), составляет порядка 45%. Для решения первой проблемы для каждой профессии/вакансии рассчитывается 90%-й интервал, а выбросы заменяются крайними значениями (выбросы в меньшую сторону – 5% перцентилем, в большую – 95% перцентилем). Затем для заполнения пропусков модель обучается линейной регрессии и предсказывает отсутствующие значения. Таким образом были получены значения средней заработной платы всех доступных значений в вакансиях, а также средние заработные платы, когда значения были досчитаны в случае, когда работодатель их не указывал.

Для корректного отраслевого соотнесения статистических показателей (т. е. согласно ОКВЭД) и отраслевых категорий⁹ портала hh.ru была осуществлена перекодировка значений портала в соответствии с разделами ОКВЭД¹⁰.

Стоит также принять во внимание методологические ограничения при сравнении показателей Росстата и портала поиска работы hh.ru, которые могут вносить коррективы (таблица 1).

Таблица 1

Сравнительная характеристика определения вакансии в Росстате и hh.ru

Table 1

Comparative Characterization of Vacancy Definition in Rosstat and hh.ru

Росстат	hh.ru
Данные «запаса» – сколько вакансий актуально на определенную дату	Данные «потока» – сколько вакансий открыто за определенный период
Данные по крупным и средним предприятиям	Все предприятия, включая малые
Вакансии для работников списочного состава	Включает вакансии в т. ч. по ГПХ
Одна вакансия – одно рабочее место	За одной вакансией может стоять несколько рабочих мест

Источник: разработано авторами.

⁶ Росстат. Численность требуемых работников списочного состава на вакантные рабочие места на конец отчетного квартала с 2017 г. // Росстат: [сайт]. URL: <https://www.fedstat.ru/indicator/57768> (дата обращения: 04.11.2024).

⁷ Росстат. Среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по субъектам Российской Федерации // Росстат: [сайт]. URL: https://rosstat.gov.ru/labor_market_employment_salaries (дата обращения: 04.11.2024).

⁸ HeadHunter API // HeadHunter: [сайт]. URL: <https://api.hh.ru/openapi/redoc#section/Obshaya-informaciya> (дата обращения: 04.11.2024).

⁹ Отраслевую принадлежность работодатель заполняет самостоятельно, исходя из специфики продукции или сервисов.

¹⁰ Общероссийский классификатор видов экономической деятельности (утв. Приказом Росстандарта от 31.01.2014 N 14-ст) (ред. от 16.08.2024) // КонсультантПлюс: [сайт]. URL: https://www.consultant.ru/document/cons_doc_LAW_163320/ (дата обращения: 04.11.2024).

Результаты и обсуждение

Показатели динамики вакансий

Полученные и обработанные данные портала поиска работы hh.ru были сопоставлены с поквартальной и ежемесячной динамикой количества вакансий Росстата. В целом данные Росстата менее волатильны, особенно при сравнении кварталов в абсолютных и процентных значениях (например, в III квартале 2023 года, рисунок 2–3), однако при этом

число вакансий, размещённых за квартал, одного порядка с числом доступных вакансий от Росстата с точки зрения восходящего/нисходящего трендов.

Проведённый корреляционный анализ поквартального распределения количества вакансий по регионам указывает на высокую положительную связь ($r = 0,98$, $p < .001$; Приложение, рисунок А), что указывает на пропорциональные изменения во времени двух источников вакансий.

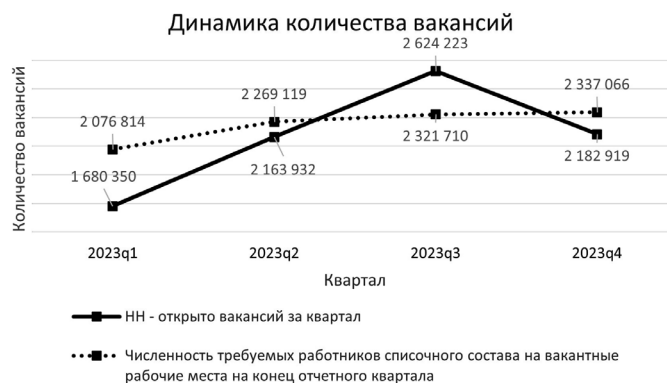


Рисунок 2. Сравнение поквартальной динамики количества вакансий портала hh.ru и Росстата (обследование предприятий), шт.

Figure 2. Comparison of Quarterly Dynamics of the Number of Vacancies on hh.ru Portal and Rosstat (Survey of Enterprises), pcs.

Источник: разработано авторами.

Помесячная динамика более наглядна. Мы можем наблюдать схожие тренды от месяца к месяцу. Пик спроса вакансий на портале hh.ru приходится на август 2023 года, данные Росстата

сопоставимы – пик приходится на май – август. Возможно, это является следствием сезонности. В целом начиная с августа данные уже более синхронны.

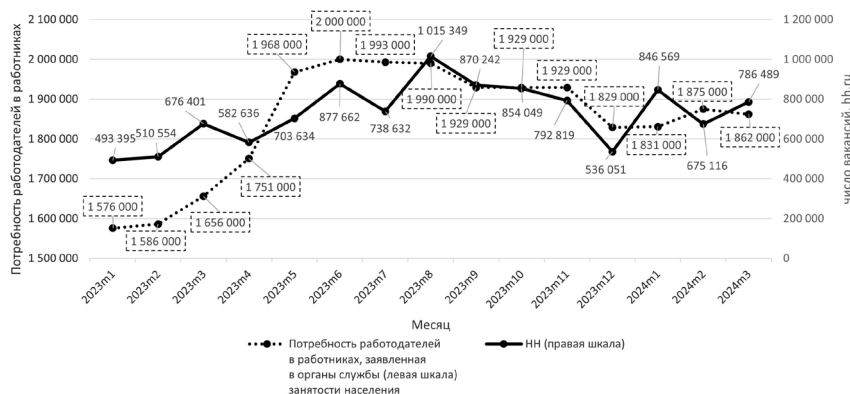


Рисунок 3. Сравнение динамики количества вакансий портала hh.ru («поток») и службы занятости населения («запас») по всем субъектам Российской Федерации, шт.

Figure 3. Comparison of Dynamics of the Number of Vacancies on the hh.ru Portal (“Flow”) and the Employment Service (“Stock”) for All Constituent Entities of the Russian Federation, pcs.

Источник: разработано авторами.

Сравнение отраслевой структуры вакансий

При сравнении отраслевой структуры вакансий портала hh.ru и Росстата (численность требуемых работников списочного состава на вакант-

ные рабочие места на конец отчетного квартала) можно отметить некоторые различия в отраслевой структуре вакансий: онлайн-вакансии хуже отражают (недопредставлены) деятельность в сельском

хозяйстве, государственные сервисы (здравоохранение, образование, государственное управление, культура). Сверхпредставлены рыночные услуги: административные услуги, профессиональная деятельность, финансы, информация и связь, гости-

ницы и общественное питание, и особенно торговля (рисунок 4). При этом сопоставимы данные в транспорте и логистике, строительстве, обрабатывающей промышленности, добыче полезных ископаемых, административной деятельности.

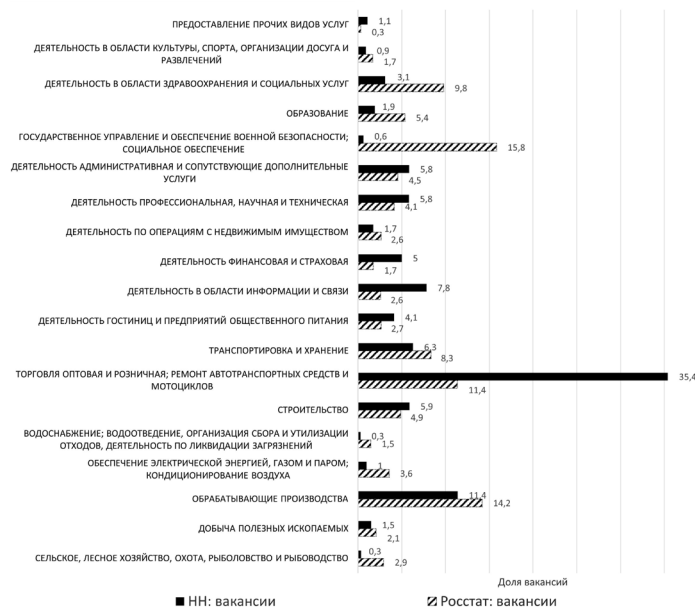


Рисунок 4. Сравнение отраслевой структуры вакансий (по общероссийскому классификатору видов экономической деятельности): доля вакансий hh.ru / Росстата от всех онлайн-вакансий и заявленных вакансий по Росстату соответственно, в среднем за 2023 г., %

Figure 4. Comparison of Sectoral Structure of Vacancies (by All-Russian Classifier of Types of Economic Activities): Share of hh.ru / Rosstat Vacancies from All Online Vacancies and Announced Vacancies According to Rosstat, Respectively, on Average for 2023, %

Источник: разработано авторами.

Сравнение источников данных о заработных платах

При сравнении источников заработной платы в целом по регионам страны стоит уточнить, что зар-

ботная плата от Росстата – это плата за стандартную продолжительность рабочего времени, фактически выплаченная, в то время как у hh.ru – предлагаемая работодателем, то есть некоторое усреднение (рисунок 5).

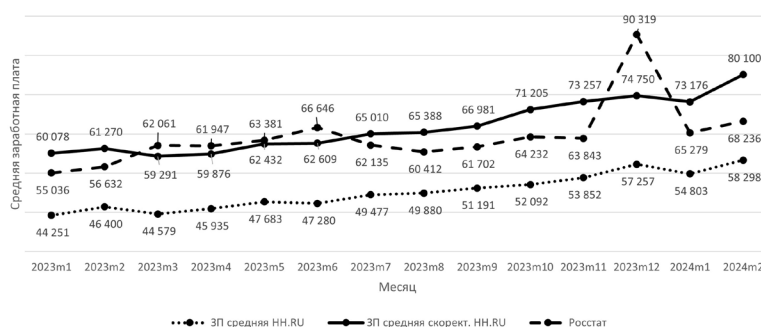


Рисунок 5. Сравнение источников данных о заработных платах: средняя и средняя скорректированная заработная плата портала hh.ru и среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по всем субъектам Российской Федерации, руб., 2023 г.

Figure 5. Comparison of Sources of Data on Salaries: Average and Average Adjusted Salary of the hh.ru Portal and Average Monthly Nominal Accrued Salary of Employees in the Full Range of Organizations by All Constituent Entities of the Russian Federation, Rubles, 2023

Источник: разработано авторами.

Скорректированная средняя заработная плата по данным портала hh.ru, то есть с учётом смоделированных значений заработной платы в вакансиях (с помощью линейной регрессии), где эта информация была пропущена, в целом достаточно близки к значениям Росстата. А корреляционный анализ квартальных и помесечных заработных плат Росстата с данными, агрегированными с hh.ru, показывает также высокую положительную связь между данными ($r = 0,85$, $p < .001$ и $r = 0,82$, $p < .001$ соответственно; Приложение, рисунок Б-В).

С июля заработная плата, предлагаемая работодателями на hh.ru, выше фактически выплаченной Росстатом, за исключением декабря 2023 года – месяца премий, что может свидетельствовать о признаках нарастающего дефицита рабочей силы.

Отраслевое сравнение источников данных о заработных платах

Значения заработных плат по отраслям в целом близки, но внимание привлекает существенная недооценка зарплат в высокооплачиваемых отраслях в онлайн-вакансиях: финансовая деятельность, деятельность в области информации и связи – вероятно, это связано с тем, что часть предложений по заработной плате обсуждается на финальных стадиях найма, а также большим оборотом организаций. Видны и области, где предложение выше, чем в данных Росстата, – операции с недвижимым имуществом, транспорт и логистика, строительство. Также отмечают отрасли с наиболее близкими значениями – торговля и добыча сырья (рисунок 6).

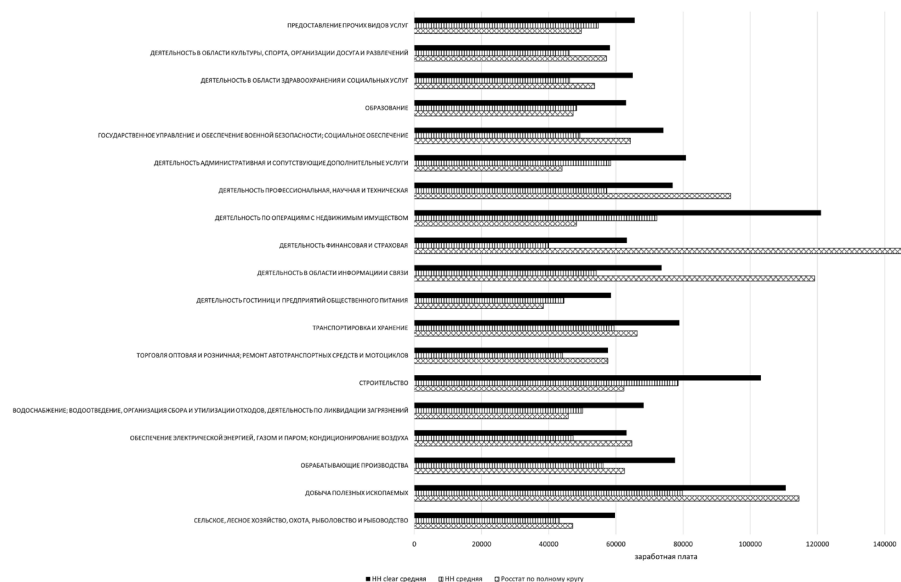


Рисунок 6. Сравнение источников данных о заработных платах по видам экономической деятельности: средняя и средняя скорректированная заработная плата портала hh.ru и среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по всем субъектам Российской Федерации, руб., 2023 г.

Figure 6. Comparison of Sources of Data on Wages by Type of Economic Activity: Average and Average Adjusted Wages of the hh.ru Portal and Average Monthly Nominal Accrued Wages of Employees for a Full Range of Organizations for All Constituent Entities of the Russian Federation, Rubles, 2023

Источник: разработано авторами.

Заключение

1. Полученные данные подтверждают, что онлайн-вакансии могут служить полезным дополнением к официальной статистике для мониторинга рынка труда. В будущем рекомендуется провести дополнительные исследования для учёта сезонных факторов и расширить использование методов машинного обучения, чтобы включить другие источники данных и платформы.

2. Полученное отраслевое распределение данных hh.ru позволило получить осмысленное со-

поставление с данными официальной статистики о вакантных рабочих местах и номинальной начисленной заработной платой.

3. При сопоставлении вакансий в отраслевом разрезе можно отметить лишь некоторые виды экономической деятельности, когда данные онлайн-вакансий соотносятся с данными Росстата (например, транспорт и логистика, строительство, обрабатывающая промышленность, добыча полезных ископаемых, административная деятельность), в других случаях мы можем отметить несовпадение.

4. При сопоставлении отраслевых показателей заработных плат также отмечаются зоны совпадения скорректированной заработной платы и номинальной начисленной зарплаты по Росстату: это деятельность в сфере торговли, добычи полезных ископаемых и др. В случаях, когда показатели онлайн-вакансий незначительно превышают данные Росстата, вероятно, можно предположить зоны дефицита рабочей силы, например в обрабатывающей промышленности, строительстве, транспортировке.

5. В отраслевом разрезе также существуют некоторые аномалии, которые требуют дополнительного исследования: высокая доля вакансий в сфере торговли (с резким скачком в мае 2024 года); заниженные заработные платы в высокодоходных отраслях (финансовая деятельность, информация и связь).

6. В дальнейшем необходимо провести анализ данных на более длительном временном отрезке, чтобы охватить как можно больше трендов, например явление сезонности за 2023–2024 гг.

Работа по анализу данных рынка труда с точки зрения публикуемых вакансий на порталах поиска работы в формате онлайн представляет огромный интерес, это можно отметить по возрастающему количеству проводимых исследований, а также возросшему числу таких платформ в России. Достоверность и валидность результатов сопоставления официальных статистик и данных онлайн-вакансий зависит от многих факторов и государственных политик стран: цифровизации сервисов, открытости экономики, используемых инструментов сбора и обработки данных.

Основным ограничивающим фактором, влияющим на восприятие полноты данных онлайн-вакансий, является их строгость (представлено ранее в таблице 1) – на одну вакансию может приходиться несколько открытых позиций, в то же время мы не можем достоверно знать, когда именно открываются и закрываются вакансии.

В результате в исследовательском плане не вполне ясно, какую именно часть рынка труда представляют онлайн-вакансии: более оперативно отражают рынок труда или дополняют данные официальной статистики.

Между тем динамика данных, полученная после обработки почти 16 млн онлайн-вакансий, и её сопоставление с данными Росстата говорят о схожести трендов, на это также указывает положительная корреляция при сравнении как количества вакансий, так и заработной платы. Одним из полезных результатов является перекодировка экономических отраслей и ниш портала hh.ru в систему классификатора ОКВЭД, что позволит в дальнейшем использовать данный «словарь» для последующих сопоставлений.

В качестве дальнейших исследовательских шагов становится актуальным вопрос о прогнозировании потребности в кадрах на рынке труда на основании данных онлайн-вакансий и резюме при помощи статистических методов или машинного обучения. Для решения этой задачи применимы модели с авторегрессией и интегрированным скользящим средним (ARIMA). Прогнозная модель ARIMA для стационарного временного ряда представляет собой уравнение регрессионного типа, в котором предикторы состоят из лагов зависимой переменной и/или лагов ошибок прогноза [17].

Также перспективным методологическим инструментом для прогнозирования в данном вопросе может стать использование алгоритма градиентного бустинга XGBoost, где в качестве признаков используются данные спроса и предложения на рынке труда, а также исторические данные о нехватке рабочей силы [18]. Применённый алгоритм показывает высокую эффективность (F1 до 86%) и прогностическую ценность. Данный подход, в отличие от источников официальной статистики, позволит намного быстрее выполнять прогноз, опираясь на данные в открытом доступе.

Список литературы

1. Макшанов А.В., Журавлев А.Е., Тындыкарь Л.Н. Большие данные. Big Data: учебник для вузов. 2022. 188 с. ISBN 978-5-8114-9690-7
2. Микрюков А.А., Гранатов М.Г., Абдрахманова З.А. Методы анализа больших данных в экономике // Общество, экономика, управление. 2023. Том 8. № 4. С. 70–74. <https://doi.org/10.47475/2618-9852-2023-8-4-70-74> EDN IXWNXY
3. Napierala J., Kvetan V., Branka J. Assessing the representativeness of online job advertisements. Cedefop working paper. No. 17. Luxembourg: Publications Office of the European Union, 2022. 26 p. <http://data.europa.eu/doi/10.2801/807500>
4. Vermeulen W., Gutierrez Amaros F. How Well do Online Job Postings Match National Sources in European Countries?: Benchmarking Lightcast data against statistical and labour agency sources across regions, sectors and occupation. OECD Local Economic and Employment Development (LEED) Papers. No. 2024/02. Paris: OECD Publishing, 2024. <https://doi.org/10.1787/e1026d81-en>

5. De Pedraza P., Visintin S., Tijdens K., Kismihók G. Survey vs scraped data // *IZA Journal of Labor Economics*. 2019. Vol. 8. No. 1. P. 1-23. <https://doi.org/10.2478/izajole-2019-0004>
6. Lovaglio P.G., Mezzanzanica M., Colombo E. Comparing time series characteristics of official and web job vacancy data // *Quality and Quantity*. 2020. Vol. 54. No. 1. P. 85-98. <https://doi.org/10.1007/s11135-019-00940-3>
7. Evans D., Mason C., Chen H., Andrew R. An algorithm for predicting job vacancies using online job postings in Australia // *Humanities and Social Sciences Communications*. 2023. Vol. 10. No. 102. P. 1-9. <https://doi.org/10.1057/s41599-023-01562-9>
8. Kureková L.M., Beblavý M., Thum-Thysen A. Using online vacancies and web surveys to analyse the labour market // *IZA Journal of Labor Economics*. 2015. Vol. 4. No. 18. P. 1-20. <https://doi.org/10.1186/s40172-015-0034-4>
9. Romanko O., O'Mahony M. The use of online job sites for measuring skills and labour market trends: a review. ESCoE Technical Report. No. 2022-19. London, 2022. 48 p. ISSN 2515-4664
10. Roshchin S., Solntsev S., Vasilyev D. Recruiting and Job Search Technologies in the Age of Internet // *Foresight and STI Governance*. 2017. Vol. 11. No. 4. P. 33-43. <https://doi.org/10.17323/2500-2597.2017.4.33.43>
11. Cammeraat E., Squicciarini M. Burning Glass Technologies' data use in policy-relevant analysis. OECD Science, Technology and Industry Working Papers. No. 2021/05. Paris: OECD Publishing, 2021. 69 p. <https://doi.org/10.1787/cd75c3e7-en>
12. Davis S., Faberman R., Haltiwanger J. The establishment-level behavior of vacancies and hiring // *Quarterly Journal of Economics*. 2013. Vol. 128. No. 2. P. 581-622. <https://doi.org/10.1093/qje/qjt002>
13. Carnevale A.P., Jayasundera T., Repnikov D. Understanding online job ads data: Technical Report. Georgetown: Center on Education and the Workforce, 2014. 28 p. <http://dx.doi.org/10.13140/RG.2.2.13283.57124>
14. Капелюшников Р.И. Российский рынок труда: статистический портрет на фоне кризисов // *Вопросы экономики*. 2023. № 8. С. 5-37. <https://doi.org/10.32609/0042-8736-2023-8-5-37> EDN USUZBY
15. Горина Т., Натсак О. Анализ спроса и предложения на современном рынке труда Республики Тыва // *Бюллетень Калмыцкого научного центра РАН*. 2022. № 1. С. 204-235. <https://doi.org/10.22162/2587-6503-2022-1-21-204-235> EDN CPSPPI
16. Терников А.А. Профессиональные стандарты и спрос на рынке труда в сфере высшего менеджмента в России // *Вестник СПбГУ. Менеджмент*. 2022. Том 21. № 1. С. 131-148. <http://doi.org/10.21638/11701/spbu08.2022.106> EDN JEXYZX
17. Aletdinova A., Bakaev M. Intelligent Data Analysis and Predictive Models for Regional Labor Markets // *Digital Transformation and Global Society: 6th International Conference Proceedings*, 23-25 June 2021, St. Petersburg, Russia. Springer, 2022. P. 351-363. https://doi.org/10.1007/978-3-030-37858-5_29
18. Predicting Skill Shortages in Labor Markets: A Machine Learning / N. Dawson, M.A. Rizioiu, B. Johnston, M.A. Williams // *IEEE International Conference on Big Data Conference Proceedings*, 10-13 December 2020, Atlanta, USA. Atlanta, 2020. P. 3052-3061. <https://doi.org/10.1109/BigData50022.2020.9377773>

Информация об авторах:

Виталий Владиславович Алтухов – младший научный сотрудник, лаборатория социальных и экономических исследований «Технологии развития человеческого капитала и построения институциональных и компетентностных моделей развития человека» при кафедре экономики труда и персонала, экономический факультет, Московский государственный университет имени М.В. Ломоносова, директор по разработке и исследованиям, «Профилум»

Алексей Дмитриевич Кудрявцев – младший научный сотрудник, лаборатория социальных и экономических исследований «Технологии развития человеческого капитала и построения институциональных и компетентностных моделей развития человека» при кафедре экономики труда и персонала, экономический факультет, Московский государственный университет имени М.В. Ломоносова, аналитик данных, «Профилум».

Заявленный вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Авторы заявляют об отсутствии конфликта интересов.

Автор, ответственный за переписку – Виталий Владиславович Алтухов.

Статья поступила в редакцию 25.02.2025; одобрена после рецензирования 19.03.2025; принята к публикации 05.06.2025.

References

1. Makshanov AV, Zhuravlev AE, Tyndykar' LN. Bol'shie dannye. Big Data. Textbook for Universities. St. Petersburg; 2022. 188 p. ISBN 978-5-8114-9690-7 (In Russ.).
2. Mikryukov AA, Granatov MG, Abdrakhmanova ZA. Big Data Analysis Methods. *Obshchestvo, ekonomika, upravlenie=Society, economy, management*. 2023;8(4):70-74. <https://doi.org/10.47475/2618-9852-2023-8-4-70-74> (In Russ.)
3. Napierala J., Kvetan V., Branka J. Assessing the Representativeness of Online Job Advertisements. Cedefop working paper. No. 17. Luxembourg: Publications Office of the European Union; 2022. 26 p. <http://data.europa.eu/doi/10.2801/807500>
4. Vermeulen W., Gutierrez Amaro F. How Well do Online Job Postings Match National Sources in European Countries?: Benchmarking Lightcast Data against Statistical and Labour Agency Sources across Regions, Sectors and Occupation. OECD Local Economic and Employment Development (LEED) Papers. No. 2024/02. Paris: OECD Publishing; 2024. <https://doi.org/10.1787/e1026d81-en>
5. De Pedraza P., Visintin S., Tijdens K., et al. Survey vs Scraped Data: Comparing Time Series Properties of Web and Survey Vacancy Data. *IZA Journal of Labor Economics*. 2019;8(1):1-23. <https://doi.org/10.2478/izajole-2019-0004>
6. Lovaglio P.G., Mezzanzanica M., Colombo E. Comparing Time Series Characteristics of Official and Web Job Vacancy Data. *Quality and Quantity*. 2020;54(1):85-98. <https://doi.org/10.1007/s11135-019-00940-3>
7. Evans D., Mason C., Chen H., et al. An algorithm for Predicting Job Vacancies Using Online Job Postings in Australia. *Humanities and Social Sciences Communications*. 2023;10(102):1-9. <https://doi.org/10.1057/s41599-023-01562-9>
8. Kureková L.M., Beblavý M., Thum-Thysen A. Using Online Vacancies and Web Surveys to Analyse the Labour Market: a Methodological Inquiry. *IZA Journal of Labor Economics*. 2015;4(18):1-20. <https://doi.org/10.1186/s40172-015-0034-4>

9. Romanko O., O'Mahony M. The Use of Online Job Sites for Measuring Skills and Labour Market Trends: a Review. ESCoE Technical Report No. 2022-19. London; 2022. 48 p. ISSN 2515-4664
10. Roshchin S., Solntsev S., Vasilyev D. Recruiting and Job Search Technologies in the Age of Internet. *Foresight and STI Governance*. 2017;11(4):33-43. <https://doi.org/10.17323/2500-2597.2017.4.33.43>.
11. Cammeraat E., Squicciarini M. Burning Glass Technologies' Data Use in Policy-relevant Analysis: an Occupation-level Assessment. OECD Science, Technology and Industry Working Papers. No. 2021/05. Paris: OECD Publishing; 2021. 69 p. <https://doi.org/10.1787/cd75c3e7-en>
12. Davis S., Faberman R., Haltiwanger J. The Establishment-level behavior of Vacancies and Hiring. *Quarterly Journal of Economics*. 2013;128(2):581-622. <https://doi.org/10.1093/qje/qjt002>
13. Carnevale A.P., Jayasundera T., Repnikov D. Understanding Online Job Ads Data: Technical Report. Georgetown: Center on Education and the Workforce; 2014. 28 p. <http://dx.doi.org/10.13140/RG.2.2.13283.57124>
14. Kapeliushnikov R.I. The Russian Labor Market: a Statistical Portrait on the Crises Background. *Voprosy Ekonomiki*. 2023;(8):5-37. <https://doi.org/10.32609/0042-8736-2023-8-5-37> (In Russ.)
15. Gorina T., Natsak. O. Analyzing Supply and Demand in Tuva's Labor Market: Professional and Gender Aspects Revisited. *Byulleten' Kalmytskogo nauchnogo tsentra RAN=Bulletin of the Kalmyk Scientific Center of the RAS*. 2022;(1):204-235. <https://doi.org/10.22162/2587-6503-2022-1-21-204-235> (In Russ.)
16. Ternikov A. A. Professional Standards and Demand in the Labor Market in the Sphere of Higher Management in Russia. *Vestnik Sankt-Peterburgskogo universiteta. Menedzhment=Vestnik of Saint Petersburg University. Management*. 2022;21(1):131-148. <http://doi.org/10.21638/11701/spbu08.2022.106> (In Russ.)
17. Aletdinova A., Bakaev M. Intelligent Data Analysis and Predictive Models for Regional Labor Markets. In: Digital Transformation and Global Society. 6th International Conference Proceedings; June 23-25, 2021; St. Petersburg, Russia. 2022:351-363. https://doi.org/10.1007/978-3-030-37858-5_29
18. Dawson N., Rizoio M. A., Johnston B., et al. Predicting Skill Shortages in Labor Markets: A Machine Learning Approach. IEEE International Conference on Big Data. Conference Proceedings; December 10-13, 2020; Atlanta, USA. 2020:3052-3061. <https://doi.org/10.1109/BigData50022.2020.9377773>

Information about the authors:

Vitalii V. Altukhov – Junior Research Fellow, Laboratory of Social and Economic Research «Technologies for the Development of Human Capital and the Construction of Institutional and Competence-Based Models of Human Development» at the Department of Labor and Personnel Economics, Faculty of Economics, Lomonosov Moscow State University; Director of Development and Research, Profilum

Aleksei D. Kudryavtsev – Junior Research Fellow, Laboratory of Social and Economic Research «Technologies for the Development of Human Capital and the Construction of Institutional and Competence-Based Models of Human Development» at the Department of Labor and Personnel Economics, Faculty of Economics, Lomonosov Moscow State University; Data Scientist, Profilum

Contribution of the authors: the authors contributed equally to this article.

The authors declare no conflicts of interests.

The author responsible for the correspondence is Vitalii V. Altukhov.

The article was submitted 25.02.2025; approved after reviewing 19.03.2025; accepted for publication 05.06.2025.

Приложение Корреляционный анализ

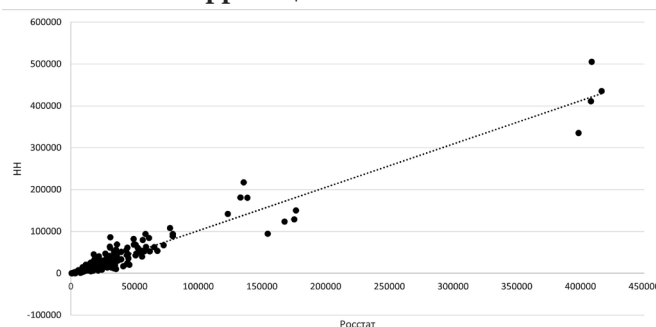


Рисунок А. Корреляционный анализ количества вакансий по всем субъектам Российской Федерации, данные Росстата (численность требуемых работников списочного состава на вакантные рабочие места на конец отчетного квартала) и портала hh.ru, 2023 г.; коэффициент корреляции Пирсона: $r=0,98$, где уровень значимости $p<.001$
Figure A. Correlation Analysis of the Number of Vacancies by All Constituent Entities of the Russian Federation, Data from Rosstat (the Number of Required List Workers for Vacant Jobs at the End of the Reporting Quarter) and hh.ru Portal, 2023; Pearson Correlation Coefficient: $r=0,98$, where the significance level $p<.001$

Источник: разработано авторами.

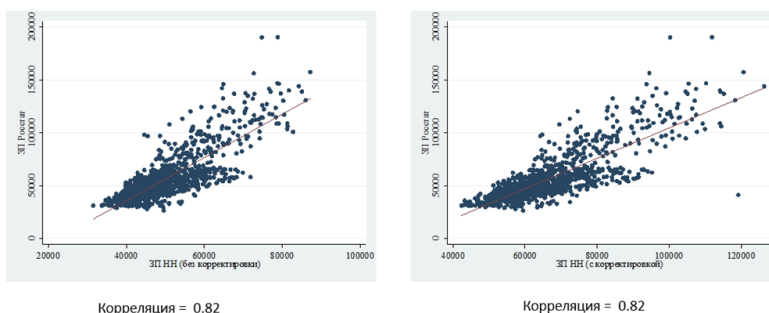


Рисунок Б. Корреляционный анализ среднемесячной заработной платы: данные Росстата (среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по всем субъектам Российской Федерации) и портала hh.ru, 2023 г.
Figure B. Correlation Analysis of Average Monthly Wages: Data from Rosstat (Average Monthly Nominal Accrued Wages of Employees for a Full Range of Organizations by RF Subjects) And hh.ru Portal, 2023.
 Источник: разработано авторами.

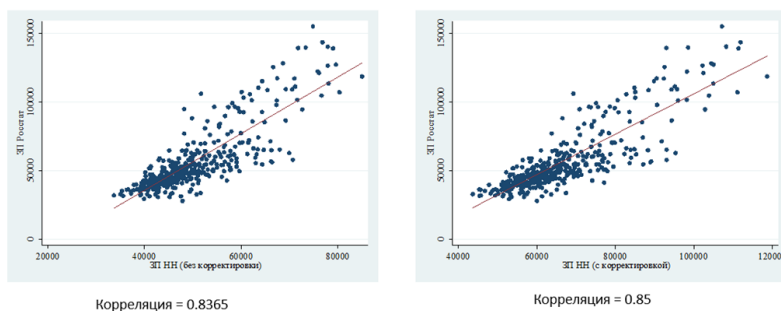


Рисунок В. Корреляционный анализ среднеквартальной заработной платы: данные Росстата (среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по всем субъектам Российской Федерации) и портала hh.ru, 2023 г.
Figure B. Correlation Analysis of Average Quarterly Wages: Data From Rosstat (Average Monthly Nominal Accrued Wages of Employees for a Full Range of Organizations by Constituent Entities of All Russian Federation) and hh.ru Portal, 2023.
 Источник: разработано авторами.