УДК 004.04

Применение SQL-запросов для анализа открытых данных рынка труда Еврейской автономной области

Афанасьева Екатерина Анатольевна
Приамурский государственный университет имени Шолом-Алейхема
Студент

Аннотация
Цель исследования — разработать и продемонстрировать методологию применения языка структурированных запросов SQL для анализа открытых данных рынка труда Еврейской автономной области (ЕАО). Методология включает проектирование реляционной базы данных и разработку набора SQL-скриптов для ETL-предобработки (извлечение, преобразование, загрузка), агрегации и анализа данных. На примере официальной статистики Росстата показано, что применение SQL-конвейера обеспечивает высокую скорость, прозрачность и воспроизводимость обработки неструктурированных данных по сравнению с ручной обработкой в табличных процессорах. Проведенный анализ выявил высокую концентрацию занятости в бюджетном секторе ЕАО (42,58%), низкую долю производственных отраслей (15,67%), а также выраженную асинхронную сезонность в строительном и сельскохозяйственном секторах. Вывод: SQL является высокоэффективным инструментом предварительной обработки и первичного анализа региональных данных, оптимально дополняющим последующую визуализацию в BI-платформах.
Ключевые слова: SQL, рынок труда, Еврейская автономная область, открытые данные, ETL-процесс, оконные функции, бизнес-аналитика.

Application of SQL queries for analyzing open data of the labor market of the Jewish Autonomous Region

Afanasyeva Ekaterina Anatolyevna
Sholom-Aleichem Priamursky State University
Student

Abstract
The aim of the study is to develop and demonstrate a methodology for applying the Structured Query Language (SQL) to analyze open data of the labor market in the Jewish Autonomous Region (JAR). The methodology includes designing a relational database and developing a set of SQL scripts for ETL preprocessing (extract, transform, load), aggregation, and data analysis. Using official Rosstat statistics as an example, it is shown that the use of an SQL pipeline ensures high speed, transparency, and reproducibility of unstructured data processing compared to manual processing in spreadsheets. The analysis revealed a high concentration of

employment in the public sector of the JAR (42.58%), a low share of industrial sectors (15.67%), and a pronounced asynchronous seasonality in the construction and agricultural sectors. Conclusion: SQL is a highly efficient tool for the preliminary processing and primary analysis of regional data, optimally complementing subsequent visualization in BI platforms.
Keywords: SQL, labor market, Jewish Autonomous Region, open data, ETL process, window functions, business intelligence.

Введение

Для субъектов Российской Федерации с небольшими объёмами экономики и ограниченными трудовыми ресурсами, к которым относится Еврейская автономная область, вопрос выбора доступных и воспроизводимых инструментов анализа рынка труда приобретает особую значимость. В условиях межрегиональной конкуренции за трудовые ресурсы, баланс которых в последние годы сокращается, оперативный и детализированный мониторинг занятости становится критически важным элементом кадровой политики региона.

Современные подходы к анализу рынка труда всё чаще опираются на большие массивы открытых данных: официальную статистику Росстата, агрегаторы вакансий, портал «Работа в России». Однако объём и неоднородность этих массивов смещают фокус аналитической работы на этап предобработки, где язык структурированных запросов SQL занимает центральное место как стандартный инструмент управления реляционными базами данных. Специфика работы с открытыми данными малых регионов заключается в том, что официальная статистика ЕАО зачастую представлена в форматах, требующих значительной предварительной обработки: файлы содержат многоуровневые шапки, объединённые ячейки, служебные строки и числовые значения с разделителями разрядов. Поэтому ключевым этапом анализа становится предобработка данных (ETL-процесс: Extract, Transform, Load), включающая приведение «широких» выгрузок к реляционной схеме, очистку и унификацию значений.

Цель исследования — разработать и продемонстрировать методологию применения SQL-запросов для анализа открытых данных рынка труда Еврейской автономной области.

Изучению проблем регионального рынка труда и методов анализа больших данных посвящён широкий круг публикаций. В работе И. О. Галюкова [3] представлен сравнительный анализ состояния рынка труда и баланса трудовых ресурсов по региональному признаку в России; автор обосновывает, что в условиях сокращения баланса человеческого капитала между регионами ведётся активная конкурентная борьба за трудовые ресурсы. Процессы цифровой трансформации рынка труда, включая формирование «дистанционных» отношений между субъектами рынка и появление цифровой рабочей силы, проанализированы И. Г. Носыревой и Н. А. Белобородовой [4], показавшими, что адаптационные механизмы занятости в условиях

цифровизации требуют повышения прозрачности рынка труда и развития компетентностной мобильности.

Ряд исследований посвящён инструментальным средствам обработки данных о занятости. Д. Д. Вавилова, К. В. Кетова и К. М. Давлетшина [1] систематизировали информационные технологии обработки статистических данных о рынке труда и предложили методологию кластеризации субъектов РФ по показателям занятости и безработицы. Е. В. Ванкевич и И. Н. Калиновская [2] рассмотрели применение технологий больших данных и искусственного интеллекта для онлайн-анализа конъюнктуры рынка труда на основе вакансий и резюме, подтвердив перспективность интеграции разнородных источников в единое информационное пространство. Особый интерес представляет работа Р. Гупты, М. Редди и Ю. Ерастова [5], в которой представлена агентная система Ishara для ответа на вопросы о рабочей силе: система интерпретирует вопросы на естественном языке, генерирует SQL-запросы (Text2SQL) и применяет структурные ограничения через конвейер преобразования SQL в AST, что демонстрирует сохранение роли SQL как фундаментального слоя аналитики рынка труда даже в современных AI-агентных архитектурах.

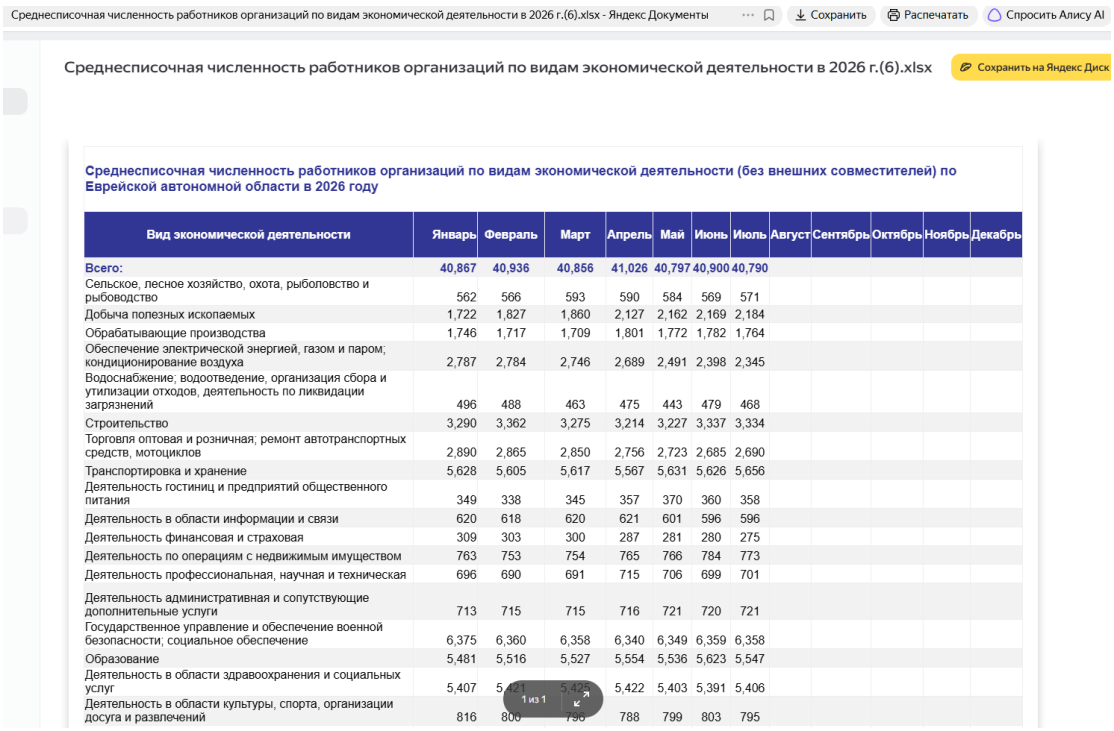
Однако детальная методология применения SQL-запросов для комплексной предобработки, агрегации и анализа открытых данных рынка труда малого региона (ЕАО) в литературе не представлена: работы либо охватывают все субъекты РФ без углубления в регион [1, 3], либо посвящены большим данным и ИИ без раскрытия инструментов предобработки [2, 4, 5].

Практическая значимость: предложенные SQL-скрипты и методология применимы органами власти ЕАО, исследовательскими и образовательными организациями для регулярного мониторинга рынка труда и поддержки решений на основе открытых данных.

Экспериментальная часть основана на файле Росстата «Среднесписочная численность работников организаций по ЕАО в 2026 году» [6]. Исходный XLSX имеет сложную нелинейную структуру (многоуровневые шапки, объединённые ячейки, пробелы в числах), что требует ETL-предобработки перед загрузкой в СУБД. Использовалась SQLite Online IDE [7] (SQL:2016, оконные функции) из-за доступности и воспроизводимости. Применены агрегирующие методы (GROUP BY, SUM, AVG), оконные функции (LAG) для сезонных трендов и ранжирование (RANK, NTILE) для выделения ядра и периферии рынка труда.

Результаты исследования

Исходные файлы Росстата имеют «широкую» структуру (wide format), при которой временные периоды (месяцы) представлены в виде отдельных столбцов, а виды экономической деятельности — в виде строк (рисунок 1). Такая структура удобна для визуального восприятия, но неэффективна для SQL-анализа: она препятствует использованию агрегатных и оконных функций, а также фильтрации по временным периодам.


Среднесписочная численность работников организаций по видам экономической деятельности в 2026 г.(6).xlsx - Яндекс Документы

Сохранить | Распечатать | Спросить Алису AI

Среднесписочная численность работников организаций по видам экономической деятельности в 2026 г.(6).xlsx

Сохранить на Яндекс Диск

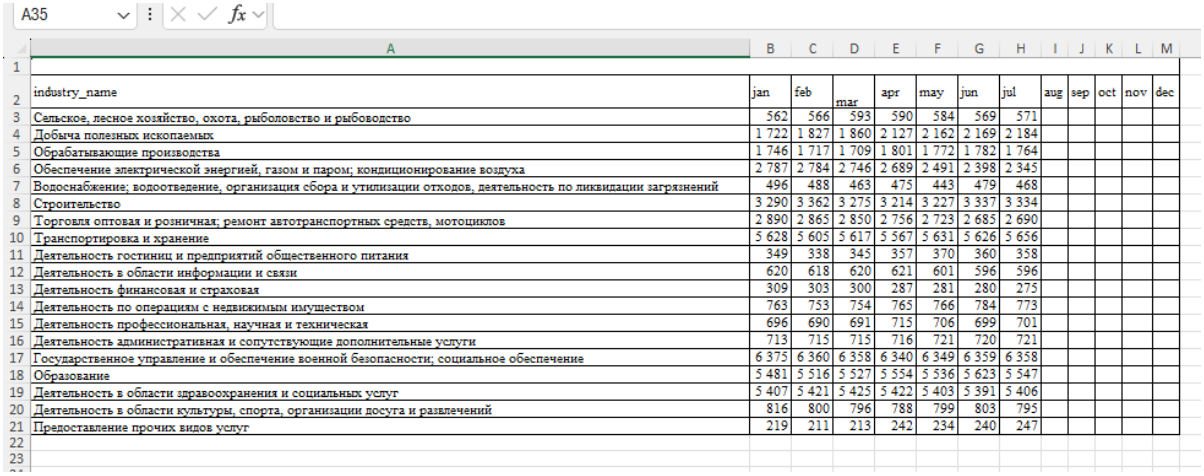
Среднесписочная численность работников организаций по видам экономической деятельности (без внешних совместителей) по Еврейской автономной области в 2026 году

Вид экономической деятельности	Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
Всего:	**40,867**	**40,936**	**40,856**	**41,026**	**40,797**	**40,900**	**40,790**					
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	562	566	593	590	584	569	571					
Добыча полезных ископаемых	1,722	1,827	1,860	2,127	2,162	2,169	2,184					
Обрабатывающие производства	1,746	1,717	1,709	1,801	1,772	1,782	1,764					
Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	2,787	2,784	2,746	2,689	2,491	2,398	2,345					
Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	496	488	463	475	443	479	468					
Строительство	3,290	3,362	3,275	3,214	3,227	3,337	3,334					
Торговля оптовая и розничная; ремонт автотранспортных средств, мотоциклов	2,890	2,865	2,850	2,756	2,723	2,685	2,690					
Транспортировка и хранение	5,628	5,605	5,617	5,567	5,631	5,626	5,656					
Деятельность гостиниц и предприятий общественного питания	349	338	345	357	370	360	358					
Деятельность в области информации и связи	620	618	620	621	601	596	596					
Деятельность финансовая и страховая	309	303	300	287	281	280	275					
Деятельность по операциям с недвижимым имуществом	763	753	754	765	766	784	773					
Деятельность профессиональная, научная и техническая	696	690	691	715	706	699	701					
Деятельность административная и сопутствующие дополнительные услуги	713	715	715	716	721	720	721					
Государственное управление и обеспечение военной безопасности; социальное обеспечение	6,375	6,360	6,358	6,340	6,349	6,359	6,358					
Образование	5,481	5,516	5,527	5,554	5,536	5,623	5,547					
Деятельность в области здравоохранения и социальных услуг	5,407	5,421	5,425	5,422	5,403	5,391	5,406					
Деятельность в области культуры, спорта, организации досуга и развлечений	816	800	796	788	799	803	795					

1 из 1

Рисунок 1 – Исходная структура файла Росстата (фрагмент)

Перед загрузкой в СУБД файл предварительно очищен в табличном процессоре (рисунок 2): удалены многоуровневые шапки и служебная строка «Всего:», наименования столбцов приведены к формату snake_case, числовые значения освобождены от пробелов-разделителей разрядов.


A35

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2	industry_name	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
3	Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	562	566	593	590	584	569	571					
4	Добыча полезных ископаемых	1 722	1 827	1 860	2 127	2 162	2 169	2 184					
5	Обрабатывающие производства	1 746	1 717	1 709	1 801	1 772	1 782	1 764					
6	Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	2 787	2 784	2 746	2 689	2 491	2 398	2 345					
7	Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	496	488	463	475	443	479	468					
8	Строительство	3 290	3 362	3 275	3 214	3 227	3 337	3 334					
9	Торговля оптовая и розничная; ремонт автотранспортных средств, мотоциклов	2 890	2 865	2 850	2 756	2 723	2 685	2 690					
10	Транспортировка и хранение	5 628	5 605	5 617	5 567	5 631	5 626	5 656					
11	Деятельность гостиниц и предприятий общественного питания	349	338	345	357	370	360	358					
12	Деятельность в области информации и связи	620	618	620	621	601	596	596					
13	Деятельность финансовая и страховая	309	303	300	287	281	280	275					
14	Деятельность по операциям с недвижимым имуществом	763	753	754	765	766	784	773					
15	Деятельность профессиональная, научная и техническая	696	690	691	715	706	699	701					
16	Деятельность административная и сопутствующие дополнительные услуги	713	715	715	716	721	720	721					
17	Государственное управление и обеспечение военной безопасности; социальное обеспечение	6 375	6 360	6 358	6 340	6 349	6 359	6 358					
18	Образование	5 481	5 516	5 527	5 554	5 536	5 623	5 547					
19	Деятельность в области здравоохранения и социальных услуг	5 407	5 421	5 425	5 422	5 403	5 391	5 406					
20	Деятельность в области культуры, спорта, организации досуга и развлечений	816	800	796	788	799	803	795					
21	Предоставление прочих видов услуг	219	211	213	242	234	240	247					
22													
23													

Рисунок 2 – Подготовленный CSV-файл со стандартизированной структурой данных

После предварительной очистки данные загружены в среду SQLite (онлайн-платформа sqliteonline.com). SQL-скрипт создания целевой таблицы и загрузки данных представлен на рисунке 3. Для обеспечения идемпотентности скрипта (возможности повторного запуска без ошибок) в рабочую версию конвейера включена команда DROP TABLE IF EXISTS raw_rosstat_2026, удаляющая прежнюю версию таблицы при её наличии.

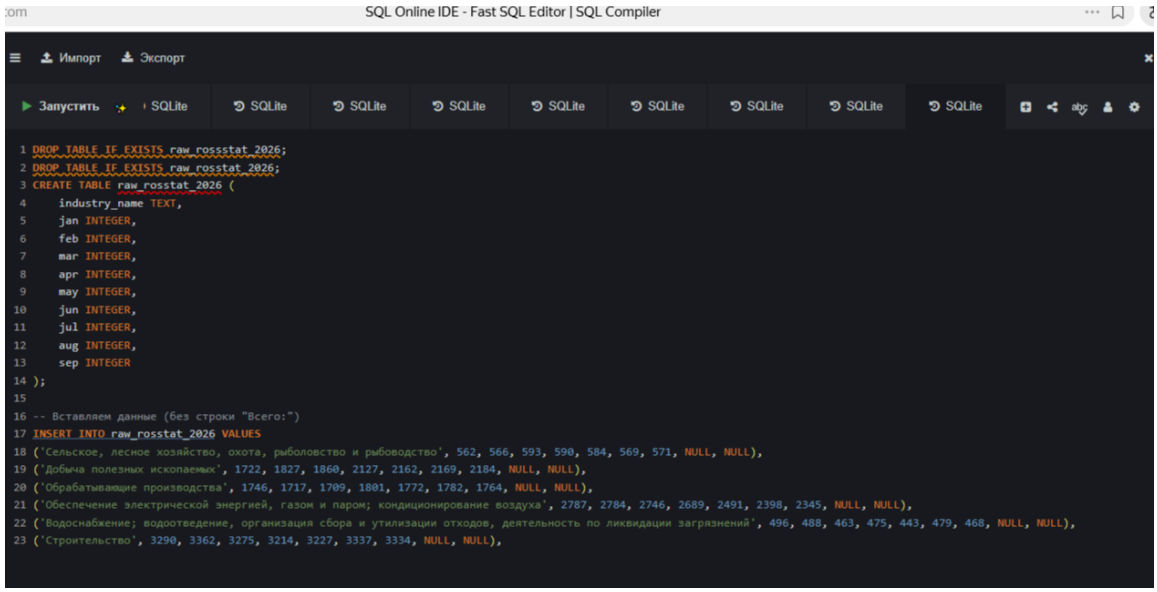


Рисунок 3 – SQL-скрипт создания таблицы raw_rosstat_2026 и загрузки исходных данных

Результат выполнения скрипта представлен на рисунке 4. Контрольный запрос SELECT COUNT(*) вернул значение 18, что соответствует числу видов экономической деятельности в исходной выгрузке [6]. Таким образом, загрузка выполнена без потерь, и данные готовы к последующей трансформации.

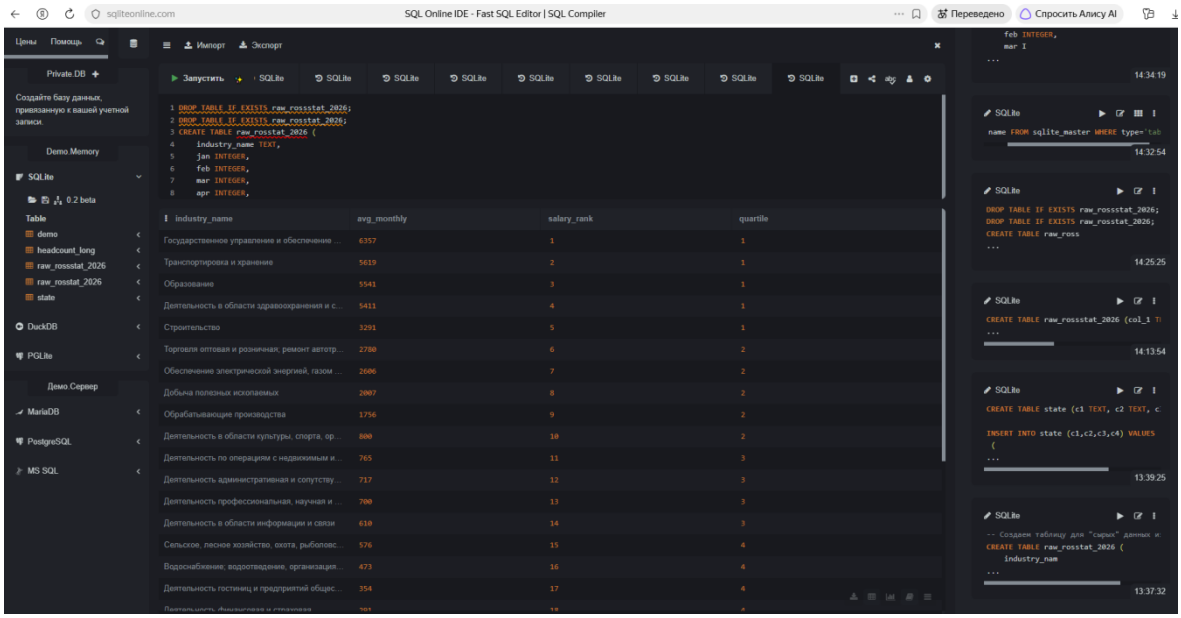


Рисунок 4 – Таблица raw_rosstat_2026 с загруженными исходными данными в среде SQLite

Для приведения данных к форме, оптимальной для реляционной базы данных, была выполнена unpivot-трансформация — преобразование «широкой» таблицы в «длинную» (long format), при котором каждая пара «отрасль–месяц» становится отдельной записью. SQL-запрос с оператором UNION ALL, реализующий эту трансформацию, представлен на рисунке 5.

Каждый блок SELECT извлекает значения одного месяца и снабжает их меткой периода report_month, а условие IS NOT NULL отсекает незаполненные ячейки.

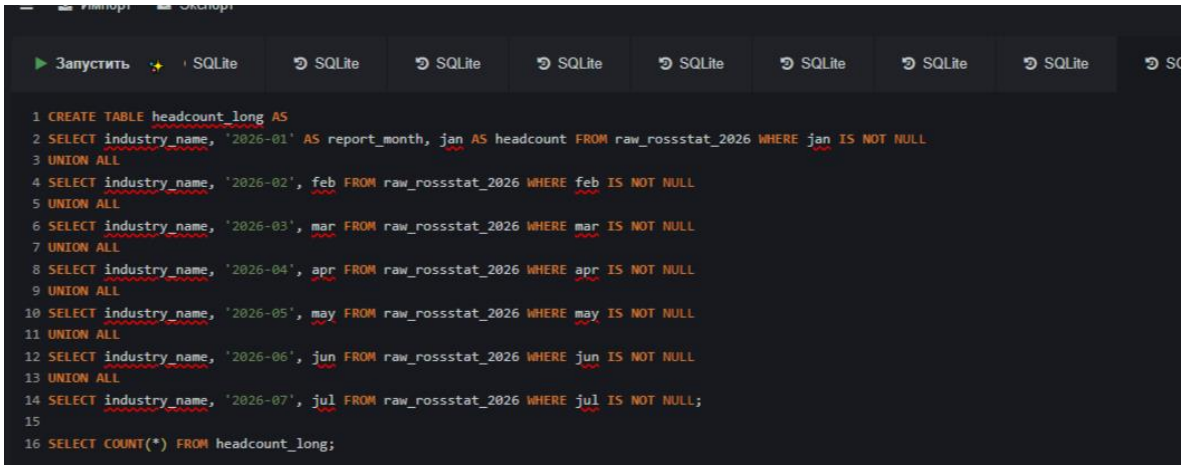


Рисунок 5 – SQL-запрос unpivot-трансформации таблицы raw_rosstat_2026

Как показано на рисунке 6, контрольный запрос вернул значение 126, что в точности равно произведению 18 видов экономической деятельности на 7 заполненных месяцев. Это подтверждает отсутствие потерь и дублирования записей при преобразовании.

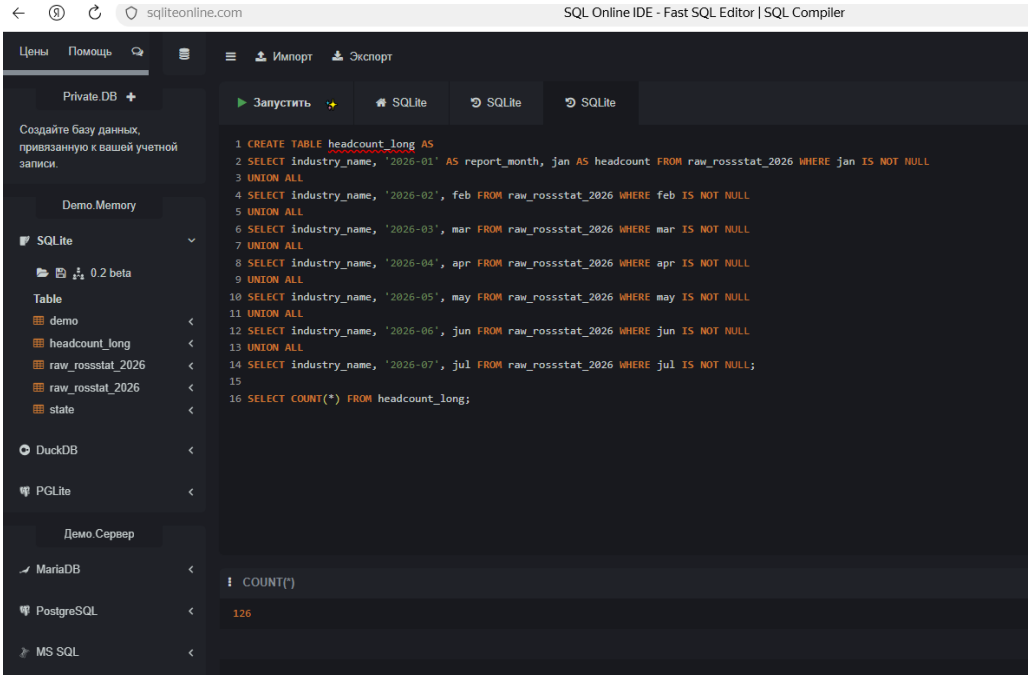


Рисунок 6 – Результат выполнения unpivot-запроса: контрольный подсчет записей таблицы headcount_long (126)

Для наглядной верификации структуры полученной таблицы выполнен отдельный запрос (рисунок 7), выбирающий контрольный фрагмент: две отрасли за первые три месяца.

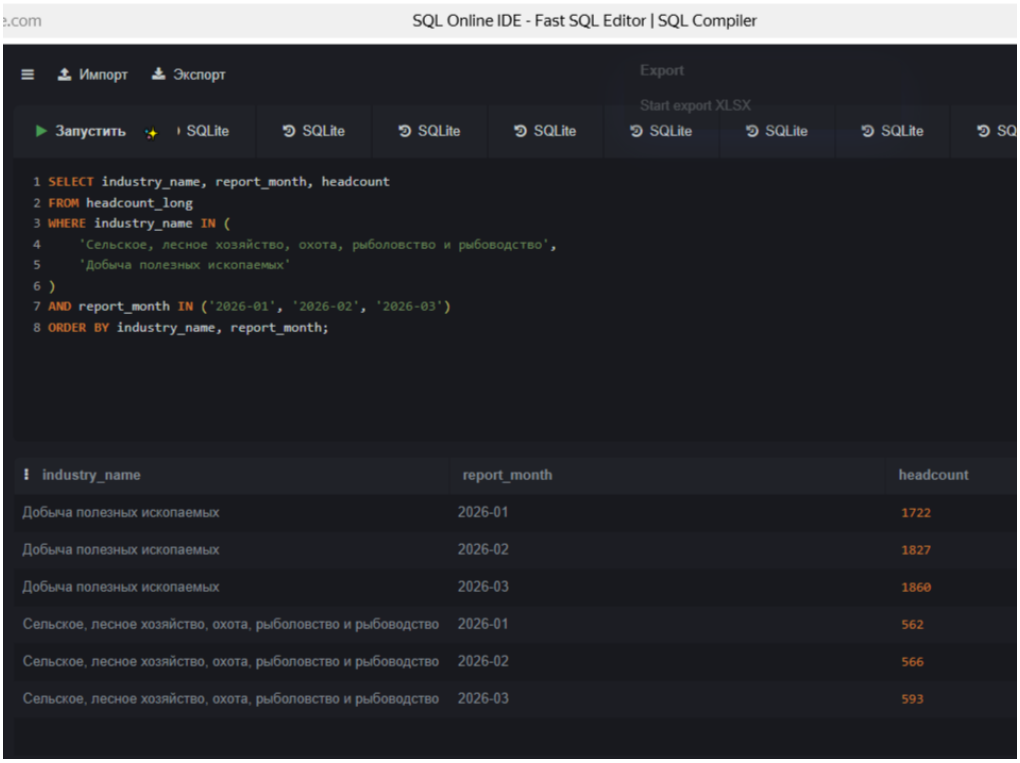


Рисунок 7 – SQL-запрос выборки контрольного фрагмента таблицы headcount_long

Результат выполнения запроса представлен в таблице 1.

Таблица 1 – Результат выборки контрольного фрагмента

industry_name	report_month	headcount
Добыча полезных ископаемых	2026-01	1722
Добыча полезных ископаемых	2026-02	1827
Добыча полезных ископаемых	2026-03	1860
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-01	562
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-02	566
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-03	593

Источник: сформировано автором в СУБД SQLite по данным [6].

Как видно из таблицы 1, каждая запись преобразованной таблицы содержит ровно одну пару «отрасль–период» и одно значение показателя. Именно такая структура делает возможными операции, недоступные для «широкого» формата: фильтрацию по периоду (WHERE report_month = …), агрегацию по отраслям (GROUP BY industry_name) и расчет помесячной динамики через оконные функции, что и используется на следующих этапах анализа.

Первой аналитической задачей является определение структуры занятости региона. SQL-запрос, использующий агрегацию GROUP BY по

наименованию отрасли и рассчитывающий суммарную и среднемесячную численность (SUM, AVG), а также долю каждой отрасли через подзапрос, представлен на рисунке 8.

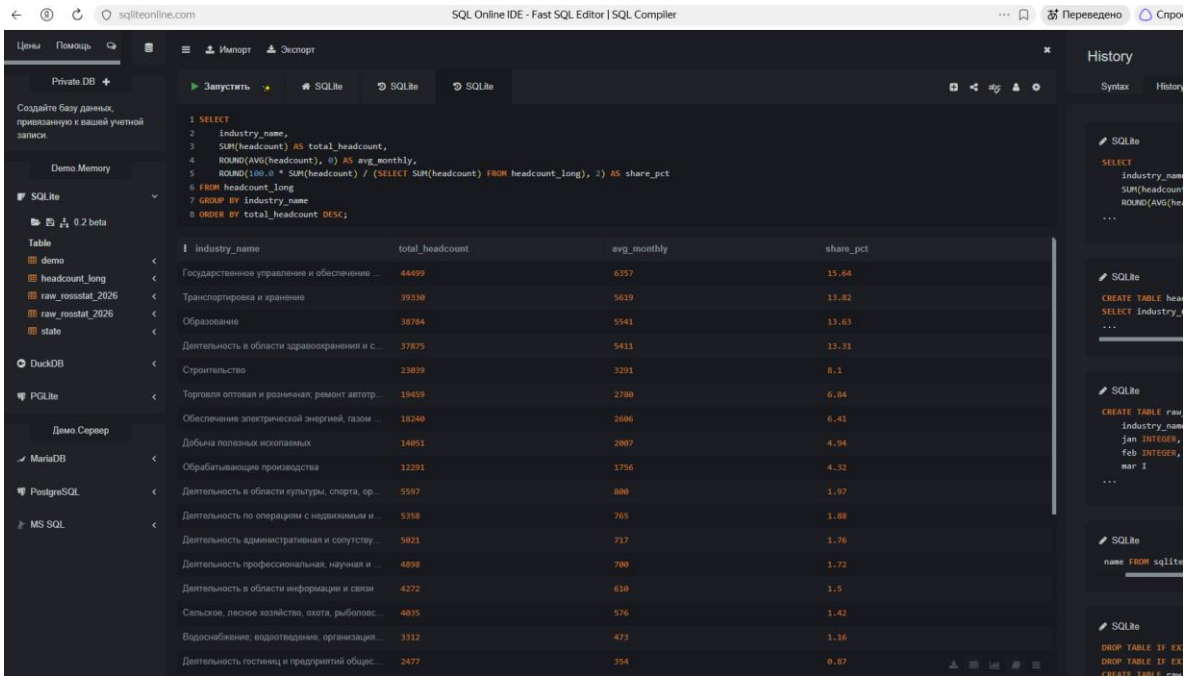


Рисунок 8 – SQL-запрос агрегации и расчета отраслевых долей занятости

Результат выполнения запроса представлен в таблице 2.

Таблица 2 – Структура занятости в ЕАО по видам экономической деятельности (январь–июль 2026 г.)

industry_name	total_headcount	avg_monthly	share_pct
Государственное управление и обеспечение военной безопасности; социальное обеспечение	44499	6357	15,64
Транспортировка и хранение	39330	5619	13,82
Образование	38784	5541	13,63
Деятельность в области здравоохранения и социальных услуг	37875	5411	13,31
Строительство	23039	3291	8,1
Торговля оптовая и розничная; ремонт автотранспортных средств, мотоциклов	19459	2780	6,84
Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	18240	2606	6,41
Добыча полезных ископаемых	14051	2007	4,94
Обрабатывающие производства	12291	1756	4,32
Деятельность в области культуры, спорта, организации досуга и развлечений	5597	800	1,97

Деятельность по операциям с недвижимым имуществом	5358	765	1,88
Деятельность административная и сопутствующие дополнительные услуги	5021	717	1,76
Деятельность профессиональная, научная и техническая	4898	700	1,72
Деятельность в области информации и связи	4272	610	1,5
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	4035	576	1,42
Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	3312	473	1,16
Деятельность гостиниц и предприятий общественного питания	2477	354	0,87
Деятельность финансовая и страховая	2035	291	0,72

Источник: рассчитано автором в СУБД SQLite по данным [6].

Полученные значения подтверждают высокую концентрацию занятости в государственном и социальном секторах: три отрасли — «Государственное управление», «Образование» и «Здравоохранение» — суммарно формируют 42,58% среднесписочной численности работников организаций ЕАО, тогда как доля производственных отраслей составляет лишь 15,67%. Это указывает на сохранение бюджетно-сырьевой ориентации регионального рынка труда, характерной для малых субъектов ДФО [3].

Для выявления сезонных трендов необходима помесячная динамика. SQL-запрос, использующий оконную функцию LAG(headcount, 1) с партиционированием по отрасли и сортировкой по периоду, представлен на рисунке 9. Запрос сопоставляет каждый месяц с предыдущим и рассчитывает темп прироста в процентах.

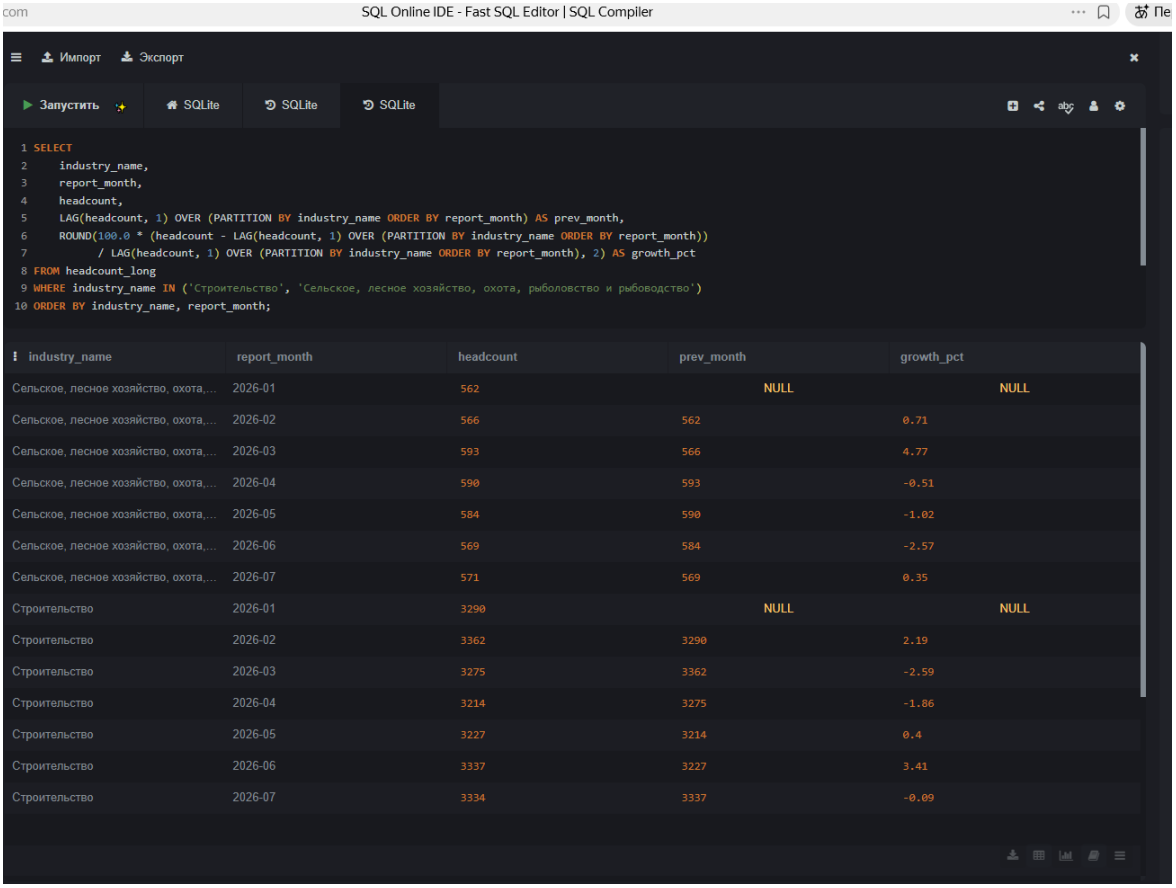


Рисунок 9 – SQL-запрос расчета помесячной динамики с оконной функцией LAG()

Результат выполнения запроса представлен в таблице 3.

Таблица 3 – Помесячная динамика среднесписочной численности работников (темпы прироста к предыдущему месяцу, %)

industry_name	report_month	headcount	prev_month	growth_pct
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-01	562		
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-02	566	562	0,71
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-03	593	566	4,77
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-04	590	593	-0,51
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-05	584	590	-1,02
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-06	569	584	-2,57
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	2026-07	571	569	0,35
Строительство	2026-01	3290		

Строительство	2026-02	3362	3290	2,19
Строительство	2026-03	3275	3362	-2,59
Строительство	2026-04	3214	3275	-1,86
Строительство	2026-05	3227	3214	0,4
Строительство	2026-06	3337	3227	3,41
Строительство	2026-07	3334	3337	-0,09

Источник: рассчитано автором в СУБД SQLite по данным [6].

Выявлена выраженная и асинхронная сезонность: пик найма в сельском хозяйстве приходится на март (+4,77%), а в строительстве — на июнь (+3,41%). Несовпадение сезонных пиков может использоваться службами занятости для временного перераспределения рабочей силы между секторами [4].

Завершающий этап — стратификация отраслей на группы по уровню занятости. SQL-запрос, использующий ранжирующие оконные функции RANK() и NTILE(4), представлен на рисунке 10.

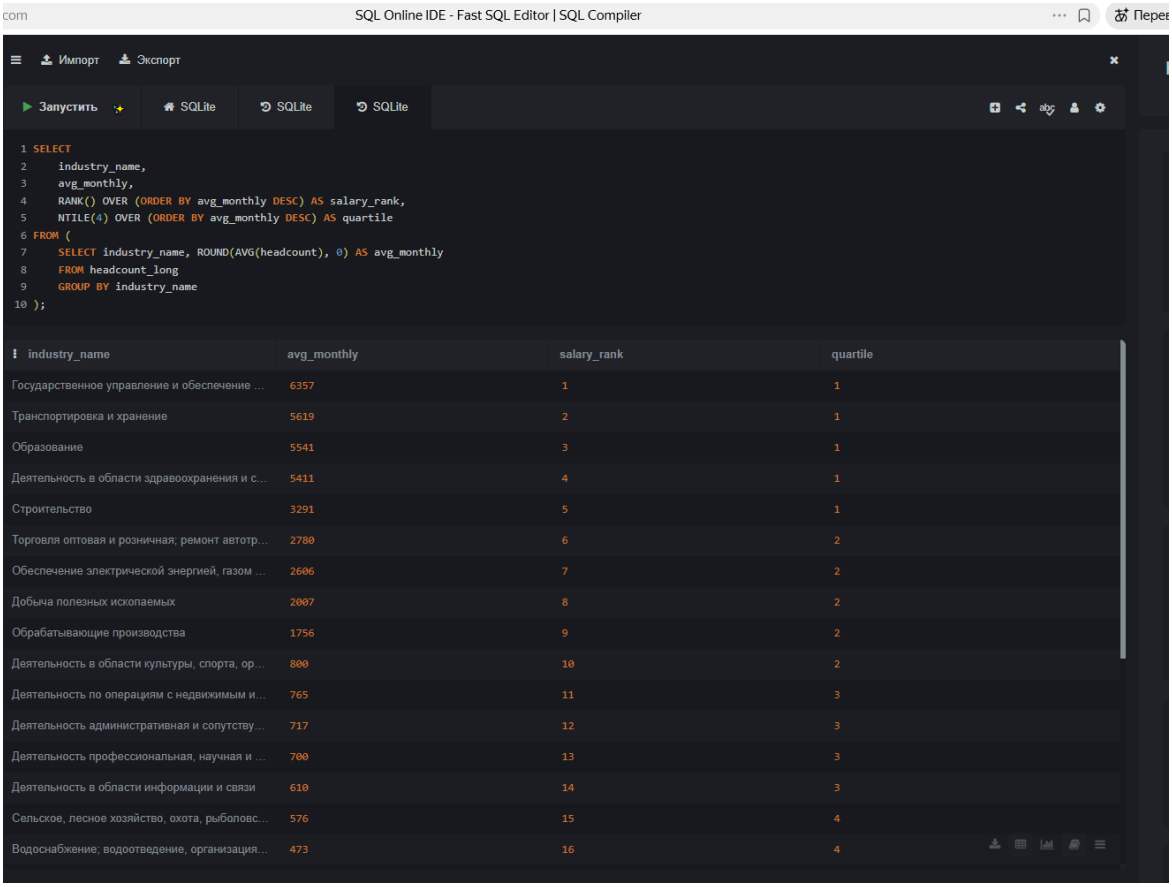


Рисунок 10 – SQL-запрос ранжирования отраслей с функциями RANK() и NTILE(4)

Результат выполнения запроса представлен в таблице 4.

Таблица 4 – Результат ранжирования отраслей ЕАО: значения RANK() и квартильные группы NTILE(4)

industry_name	avg_monthly	industry_rank	quartile
Государственное управление и обеспечение военной безопасности; социальное обеспечение	6357	1	1
Транспортировка и хранение	5619	2	1
Образование	5541	3	1
Деятельность в области здравоохранения и социальных услуг	5411	4	1
Строительство	3291	5	1
Торговля оптовая и розничная; ремонт автотранспортных средств, мотоциклов	2780	6	2
Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	2606	7	2
Добыча полезных ископаемых	2007	8	2
Обрабатывающие производства	1756	9	2
Деятельность в области культуры, спорта, организации досуга и развлечений	800	10	2
Деятельность по операциям с недвижимым имуществом	765	11	3
Деятельность административная и сопутствующие дополнительные услуги	717	12	3
Деятельность профессиональная, научная и техническая	700	13	3
Деятельность в области информации и связи	610	14	3
Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	576	15	4
Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	473	16	4
Деятельность гостиниц и предприятий общественного питания	354	17	4
Деятельность финансовая и страховая	291	18	4

Источник: рассчитано автором в СУБД SQLite по данным [6].

Анализ показал, что ядро рынка труда ЕАО (1-й квартиль) образуют пять отраслей — государственное управление (6 357 чел. среднемесячно), транспортировка и хранение (5 619), образование (5 541), здравоохранение (5 411) и строительство (3 291), — на которые суммарно приходится 64,50% занятости региона. Четвертый квартиль формируют периферийные сектора (сельское хозяйство, водоснабжение, гостиницы, финансы). Разрыв между

квартилями превышает пятнадцатикратное значение, что свидетельствует о низкой диверсификации регионального рынка труда [1, 3].

Практическим результатом эксперимента стал полностью воспроизводимый конвейер, включающий шесть SQL-запросов (рисунки 3, 5, 7, 8, 9, 10), последовательно выполняющих загрузку, нормализацию, агрегацию и ранжирование данных.

Обсуждение

Проведённое исследование позволило количественно охарактеризовать структуру и внутригодовую динамику занятости в ЕАО на основе данных Росстата [6] с применением инструментария SQL.

Анализируя структурные результаты, можно отметить, что высокая концентрация занятости в трёх бюджетных отраслях (42,58%) при доле производственного сектора всего 15,67% подтверждает бюджетно-сырьевую модель экономики малых регионов ДФО, описанную И. О. Галюковым [3]. Квартильный анализ (таблица 4) формализовал это наблюдение: разрыв между ядром рынка труда (5 отраслей) и периферией превышает 15-кратное значение, что указывает на крайне низкую диверсификацию занятости и согласуется с выводами о кластеризации регионов с низкими показателями [1].

Что касается динамических показателей, применение оконной функции LAG() (таблица 3) выявило асинхронную сезонность: пик найма в сельском хозяйстве приходится на март (+4,77%), а в строительстве — на июнь (+3,41%). Это создаёт объективные предпосылки для временного межотраслевого перераспределения рабочей силы, что является важным адаптационным механизмом в условиях цифровой трансформации и необходимости развития компетентностной мобильности [4].

С методологической точки зрения, разработанный конвейер из шести SQL-скриптов обеспечивает полную воспроизводимость и академическую прозрачность ETL-процесса, исключая скрытые ошибки ручной обработки в табличных процессорах. Сформированная «длинная» таблица headcount_long готова к бесшовной интеграции с BI-платформами (Power BI, Loginom), что соответствует современным подходам к созданию единых информационных пространств [2] и подтверждает роль SQL как фундаментального слоя аналитики даже в эпоху AI-агентов [5].

Вместе с тем, результаты следует интерпретировать с учётом ряда ограничений: во-первых, анализ не охватывает неформальную занятость и самозанятых; во-вторых, детализация по ОКВЭД не позволяет оценить дефицит по конкретным профессиям (для этого требуется интеграция с данными агрегаторов вакансий [2, 5]); в-третьих, рассмотрен период лишь в 7 месяцев 2026 г.; в-четвёртых, публикация данных Росстата имеет задержку в 1–2 месяца.

Заключение

В статье разработана и экспериментально апробирована методология применения SQL-запросов для анализа открытых данных рынка труда

Еврейской автономной области на примере официальной статистики Росстата о среднесписочной численности работников организаций за январь–июль 2026 г. Предложенный аналитический конвейер из шести SQL-скриптов покрывает полный цикл ETL-обработки: от создания реляционной схемы и загрузки исходных «широких» выгрузок до unpivot-трансформации, агрегации, анализа внутригодовой динамики через оконные функции и квартильного ранжирования отраслей.

Эмпирические результаты подтвердили наличие выраженных структурных диспропорций на рынке труда ЕАО: три бюджетные отрасли (государственное управление, образование, здравоохранение) суммарно формируют 42,58% занятости, тогда как доля производственного сектора составляет лишь 15,67%, что свидетельствует о низкой диверсификации региональной экономики. Квартильный анализ формализовал это наблюдение, выделив ядро рынка труда (5 отраслей, 64,50% занятости) и периферийные сектора с пятнадцатикратным разрывом по среднемесячной численности. Анализ внутригодовой динамики выявил выраженную асинхронную сезонность в сельскохозяйственном и строительном секторах, что создаёт объективные предпосылки для межотраслевого перераспределения рабочей силы.

Эксперимент наглядно продемонстрировал, что первоначальные трудозатраты на проектирование реляционной схемы и написание SQL-скриптов многократно окупаются за счёт автоматизации рутинных операций, обеспечения воспроизводимости и прозрачности вычислений, а также подготовки данных в форме, пригодной для последующей визуализации в BI-системах. Предложенная методология может быть непосредственно использована органами исполнительной власти ЕАО для регулярного мониторинга, а также легко адаптирована для других субъектов Российской Федерации путём замены исходного файла Росстата.

Дальнейшие исследования могут быть направлены на:

- расширение источников данных за счёт интеграции выгрузок Росстата с API агрегаторов вакансий для оценки реального кадрового дефицита по конкретным профессиям и специальностям;
- построение прогнозных моделей сезонной и межгодовой динамики занятости на основе методов анализа временны́х рядов;
- разработку гибридного аналитического конвейера, в котором SQL выступает слоем предобработки и витрины данных, а отечественные BI-платформы (Loginom, Modus BI, Yandex DataLens) — слоем интерактивной визуализации для лиц, принимающих решения.

Библиографический список

1. Вавилова Д. Д., Кетова К. В., Давлетшина К. М. Информационные технологии обработки данных для анализа показателей рынка труда в разрезе по субъектам Российской Федерации // Интеллектуальные системы в производстве. 2025. Т. 23. № 2. С. 80-92.

URL:https://elibrary.ru/item.asp?id=82505965&ysclid=mvotcs7y18656177089
2. Ванкевич Е. В., Калиновская И. Н. Онлайн анализ рынка труда с применением технологий больших данных и искусственного интеллекта // НАУЧНЫЙ ИМПУЛЬС. 2023. С. 75-84. URL:https://elibrary.ru/item.asp?id=50098560&ysclid=mvotfu801b70183899
3. Галюков, И. О. Сравнительный анализ основных характеристик региональных рынков труда в России // Экономика труда. 2024. Т. 11, № 5. С. 615-630. URL: https://elibrary.ru/item.asp?id=67897663&ysclid=mvotgy5j1h135211213
4. Носырева И. Г., Белобородова Н. А. Трансформация рынка труда в цифровую эпоху // Экономика труда. 2025.Т. 12. № 12. С. 1821-1842. URL:https://elibrary.ru/item.asp?id=88906008&ysclid=mvothqko1k992002880
5. Gupta R., Reddy M., Yerastov Y. A Workforce Agent for Labor Market Question and Answering //2025 IEEE International Conference on Big Data (BigData). IEEE, 2025. С. 6335-6342. URL: https://ieeexplore.ieee.org/abstract/document/11400733
6. Среднесписочная численность работников организаций по видам экономической деятельности в 2026 г. URL: https://docs.yandex.ru/docs/view?url=ya-browser%3A%2F%2F4DT1uXEPRrJRXlUFoewruIklm1dX-UgfNXQ1xphXyPP0iTg7IVwIybvNe_fdb7IyXbuKZhQ_jV4E28sYcSoyFmJ8b14_yN2zJLRlJuupzj1BLK8XXLiKn7seeWQ4-cA2UzvCJIB5KTdK3g4uj4v51A%3D%3D%3Fsign%3D6eGUDv7rzjosqfSdennMq98eNNh9vqEVDkjjxTTb0zs%3D&name=Среднесписочная%20численность%20работников%20организаций%20по%20видам%20экономической%20деятельности%20в%202026%20г.(2).xlsx&nosw=1
7. SQL Online IDE - Fast SQL Editor. URL:https://sqliteonline.com/