

Оценка рынка труда IT-специалистов в России



Исполнитель: Воронин Артём

Содержание



- 1 Описание проекта
- 2 Product Brief
- 3 План выполнения проекта
- 4 Парсинг данных
- 5 Результаты

1 Описание проекта



В связи с пандемией коронавируса и общей экономической ситуацией, рынок труда потерпел заметные изменения, которые не могли не сказаться на условиях труда и требованиях, предъявляемых к соискателям. Образовательные программы ВУЗов не всегда готовят специалистов, соответствующих новым требованиям рынка, в частности в сфере IT, что приводит к невостребованности выпускников.

Цель проекта:

Определить какие специалисты наиболее востребованы на рынке IT и какие квалификационные требования к ним предъявляются.

2 Product Brief



Проблема проекта	Снижение востребованности высшего образования в сфере IT в связи с несоответствием программ обучения в ВУЗах реальным требованиям работодателей.
Возможные стейкхолдеры	Томский государственный университет, ВУЗы, площадки интернет-рекрутмента и другие организации, оказывающие образовательные услуги, министерство труда, абитуриенты
Цель анализа	1. Выявление условий работы и квалификационных требований к IT-специалистам на рынке труда. 2. Оценка рынка образовательных программ. 3. Анализ соответствия образовательных программ требованиям рынка к компетенциям специалистов.
Начальная гипотеза/исследовательский вопрос	В связи с пандемией и переходом бизнеса в онлайн-формат резко возрос спрос на специалистов в сфере IT. Однако выпускники ВУЗов не имеют требуемых квалификационных знаний и навыков из-за отсутствия ориентированности образовательных программ на нужды работодателей.
Задачи проекта	1. Предварительный анализ рынка труда в сфере IT. 2. Сбор данных о вакансиях с сайтов интернет-рекрутмента (hh.ru) и предобработка. 3. Анализ данных для выявления условий работы на рынке и квалификационных требований к IT-специалистам. 4. Анализ рынка образовательных программ. 5. Анализ соответствия рынка образовательных программ рынку труда в IT сфере. 6. Подведение итогов проделанной работы.
Используемые данные	Head hunter, Работа в России, Superjob, Rabota.ru Профстандарты, утвержденные министерством труда Сайты ведущих вузов (Образовательные программы, перечень специальностей в сфере IT, количество мест) Хабр
Возможные сложности	Ограничения по количеству данных, которое можно извлечь Разный формат данных
Ожидаемые результаты	Презентация с выводами, отражающими отраслевую и профессиональную структуру рынка труда, требования и условия работы специалистов.
Метрика успеха проекта	Полнота собранных данных; полнота аналитики, логичность, содержательная интерпретация, информативность и визуализация.

3 План выполнения проекта



1. Сбор данных о вакансиях с сайта HeadHunter
2. Предобработка данных
3. Анализ данных для выявления квалификационных требований к IT-специалистами и условий работы
5. Описание полученных результатов и визуализация данных
6. Подведение итогов проделанной работы

4 Решение



Для получения данных о вакансиях с сайта интернет-рекрутмента Headhunter, была разработана программа на языке Python.

```
def getPage(page=0):  
    """  
    Получение страницы со списком вакансий.  
    Аргументы:  
        page - Индекс страницы, начинается с 0. Значение по умолчанию 0, т.е. первая страница  
    """  
  
    # Справочник для параметров GET-запроса  
    params = {  
        'text': 'NAME:', # Текст фильтра. В имени должно быть слово "Аналитик"  
        'area': 113, # Поиск осуществляется по России  
        'page': page, # Индекс страницы поиска на HH  
        'per_page': 100, # Кол-во вакансий на 1 странице (для теста 20, для рабочей программы 100)  
        'specialization' : 1, # Информационные технологии  
        'profarea_id' : 1,  
        'date_from' : '2021-11-15',  
        'date_to' : '2021-11-16'  
    }  
  
    # Пошлaем запрос к API  
    req = requests.get('https://api.hh.ru/vacancies', params)  
    # Декодируем его ответ, чтобы Кириллица отображалась корректно  
    data = req.content.decode()  
    req.close()  
  
    return data
```

4 Решение



```
jsObjs = []

# Считываем первые 2000 вакансий
for page in range(0, 20):

    # Преобразуем текст ответа запроса в справочник Python
    jsonObj = json.loads(getPage(page))

    # добавление текущего ответа запроса в список
    jsObjs.extend(jsonObj["items"])

    # Сохраняем файлы в папку {путь до текущего документа со скриптом}\docs\pagination
    # Определяем количество файлов в папке для сохранения документа с ответом запроса
    # Полученное значение используем для формирования имени документа
    nextFileName = './docs/pagination/{}.json'.format(len(os.listdir('./docs/pagination')))

    # Создаем новый документ, записываем в него ответ запроса, после закрываем
    f = open(nextFileName, mode='w', encoding='utf8')
    f.write(json.dumps(jsonObj, ensure_ascii=False))
    f.close()

    # Проверка на последнюю страницу, если вакансий меньше 2000
    if (jsonObj['pages'] - page) <= 1:
        break

    # Необязательная задержка, но чтобы не нагружать сервисы hh, оставим.
    time.sleep(0.25)

print('Страницы поиска собраны')
```

```
# Получаем карточки вакансий
v = 0

for v in jsObjs: #цикл по справочнику

    # Обращаемся к API и получаем детальную информацию по конкретной вакансии
    req = requests.get(v['url'])
    data = req.content.decode()
    req.close()

    # Создаем файл json с идентификатором вакансии в качестве названия
    fileName = './docs/vacancies/{}.json'.format(v['id'])
    # Записываем в него ответ запроса и закрываем
    f = open(fileName, mode='w', encoding='utf8')
    f.write(data)
    f.close()

    time.sleep(0.25)

print('Вакансии собраны')
```

4 Решение



```
# Модуль для работы с отображением вывода Jupyter
from IPython import display

skills_vac = [] # Список идентификаторов вакансий
skills_name = [] # Список названий навыков

# В выводе будем отображать прогресс
# Для этого узнаем общее количество файлов, которые надо обработать
cnt_docs = len(os.listdir('./docs/vacancies 211121'))
i = 0 # Счетчик обработанных файлов установим в ноль

df_vac = pd.DataFrame([])

# Проходимся по всем файлам в папке vacancies
for fl in os.listdir('./docs/vacancies 211121'):

    # Открываем, читаем и закрываем файл
    f = open('./docs/vacancies 211121/{}'.format(fl), encoding='utf8')
    jsonText = f.read()
    f.close()

    # Текст файла переводим в справочник
    jsonObj = json.loads(jsonText)
    # Добавляем строки в датафрейм
    df_vac = df_vac.append(pd.DataFrame.from_dict(pd.json_normalize(jsonObj), orient='columns'))

    # Увеличиваем счетчик обработанных файлов на 1, очищаем вывод ячейки и выводим прогресс
    i += 1
    display.clear_output(wait=True)
    display.display('Готово {} из {}'.format(i, cnt_docs))
```

'Готово 56577 из 56577'

4 Решение



В результате была получена база вакансий следующей структуры:

id	name	description	key_skills	specializations	professional_roles	published_at	area.id	area.name	salary.from	salary.to	salary.currency	address.lat	address.lng	experience.name	schedule.name	employment.name
34526262	Младший разработчик / Junior Developer (C#)	<p>Привет. Меня зовут Рудниченко Константин, я...		[{'id': '1.221', 'name': 'Программирование, Pa...'}]	[{'id': '96', 'name': 'Программист, разработчи...'}]	2021-11-14T14:21:39+0300	4	Новосибирск	30000.0	NaN	RUR	54.863316	83.090858	Нет опыта	Гибкий график	Полная занятость
35345819	Геймдизайнер	<p> «em>Наши проекты растут, и с...	[{'name': 'рынок мобильных приложений'}, {'nam...'}]	[{'id': '1.475', 'name': 'Игровое ПО', 'profar...'}]	[{'id': '25', 'name': 'Гейм-дизайнер'}]	2021-11-14T15:19:39+0300	24	Волгоград	NaN	NaN	NaN	48.707073	44.516930	От 1 года до 3 лет	Полный день	Полная занятость

При предварительном анализе было выявлено, что требуется очистка данных:

```
[6]: data['professional_roles'].value_counts()

[6]: [{"id": "96", "name": "Программист, разработчик"} 12681
      [{"id": "40", "name": "Другое"}] 7760
      [{"id": "97", "name": "Продавец-консультант, продавец-кассир"}] 3889
      [{"id": "121", "name": "Специалист технической поддержки"}] 3651
      [{"id": "70", "name": "Менеджер по продажам, менеджер по работе с клиентами"}] 3393
      ...
      [{"id": "95", "name": "Полицейский"}] 1
      [{"id": "56", "name": "Косметолог"}] 1
      [{"id": "5", "name": "Автослесарь, автомеханик"}] 1
      [{"id": "24", "name": "Врач"}] 1
      [{"id": "61", "name": "Мастер ногтевого сервиса"}] 1
      Name: professional_roles, Length: 130, dtype: int64
```

4 Решение

Для более детального анализа потребовалось формирование реестра ключевых навыков специалистов и базы регионов:

Регион	id_city	City
Республика Марий Эл	4228	Виловатово
Республика Марий Эл	1621	Волжск
Республика Марий Эл	1622	Звенигово
Республика Марий Эл	4229	Знаменский
Республика Марий Эл	61	Йошкар-Ола
Республика Марий Эл	4230	Кельмаксола
Республика Марий Эл	4231	Килемары
Республика Марий Эл	1623	Козьмодемьянск

id	vacancy	skill	professional_roles
14820007	Программист 1С	1С программирование	Программист, разра...
14820007	Программист 1С	1С: Бухгалтерия	Программист, разра...
14820007	Программист 1С	1С: Предприятие 8	Программист, разра...
14820007	Программист 1С	1С: Зарплата и упр...	Программист, разра...
14820007	Программист 1С	1С: Управление Тор...	Программист, разра...
14820007	Программист 1С	1С: Управление Про...	Программист, разра...
14820007	Программист 1С	Управление проектами	Программист, разра...
14820007	Программист 1С	Обновление конфи...	Программист, разра...
14820007	Программист 1С	1С: Бухгалтерия	Программист, разра...
14820007	Программист 1С	1С: Документооборот	Программист, разра...
14820007	Программист 1С	1С: Комплексная авт...	Программист, разра...
14820007	Программист 1С	Бухгалтерский учет	Программист, разра...
14820007	Программист 1С	1С: Зарплата и кадры	Программист, разра...
14820007	Программист 1С	1С: Торговля	Программист, разра...
14820007	Программист 1С	Кадровое делопроиз...	Программист, разра...
14820007	Программист 1С	1С: Торговля и склад	Программист, разра...
14820007	Программист 1С	Бюджетирование	Программист, разра...
26212752	Тестировщик ПО/ Q...	Тестирование	Тестировщик
26212752	Тестировщик ПО/ Q...	Проведение тестиро...	Тестировщик
26212752	Тестировщик ПО/ Q...	Linux	Тестировщик
26212752	Тестировщик ПО/ Q...	Аналитические спос...	Тестировщик
26212752	Тестировщик ПО/ Q...	Аналитическое мыш...	Тестировщик

4 Решение



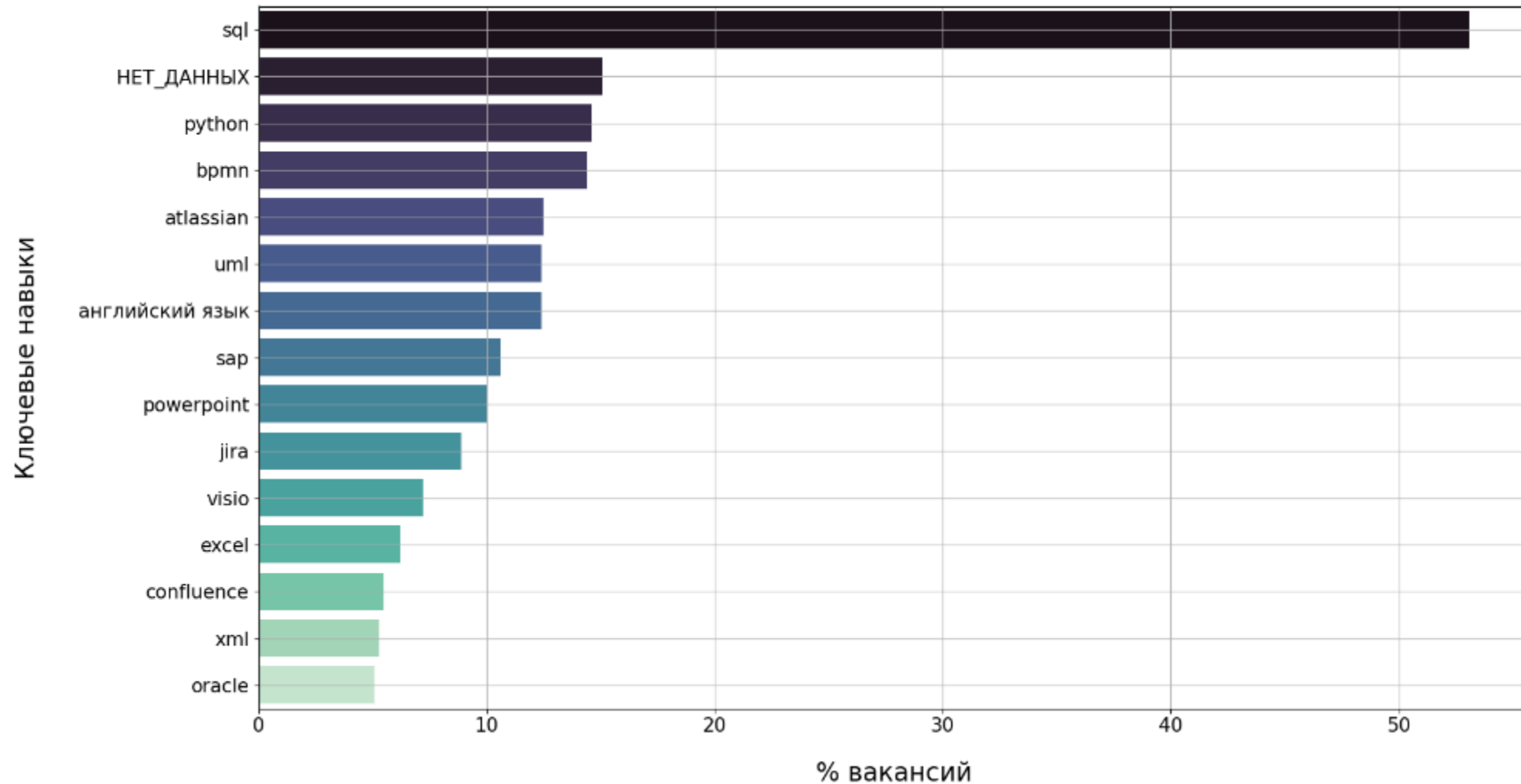
По окончании обработки таблица выглядела следующим образом:

В таком виде таблица была сохранена на жесткий диск и вся дальнейшая работа проходила уже с этой готовой таблицей.

5 Результаты. Аналитик.



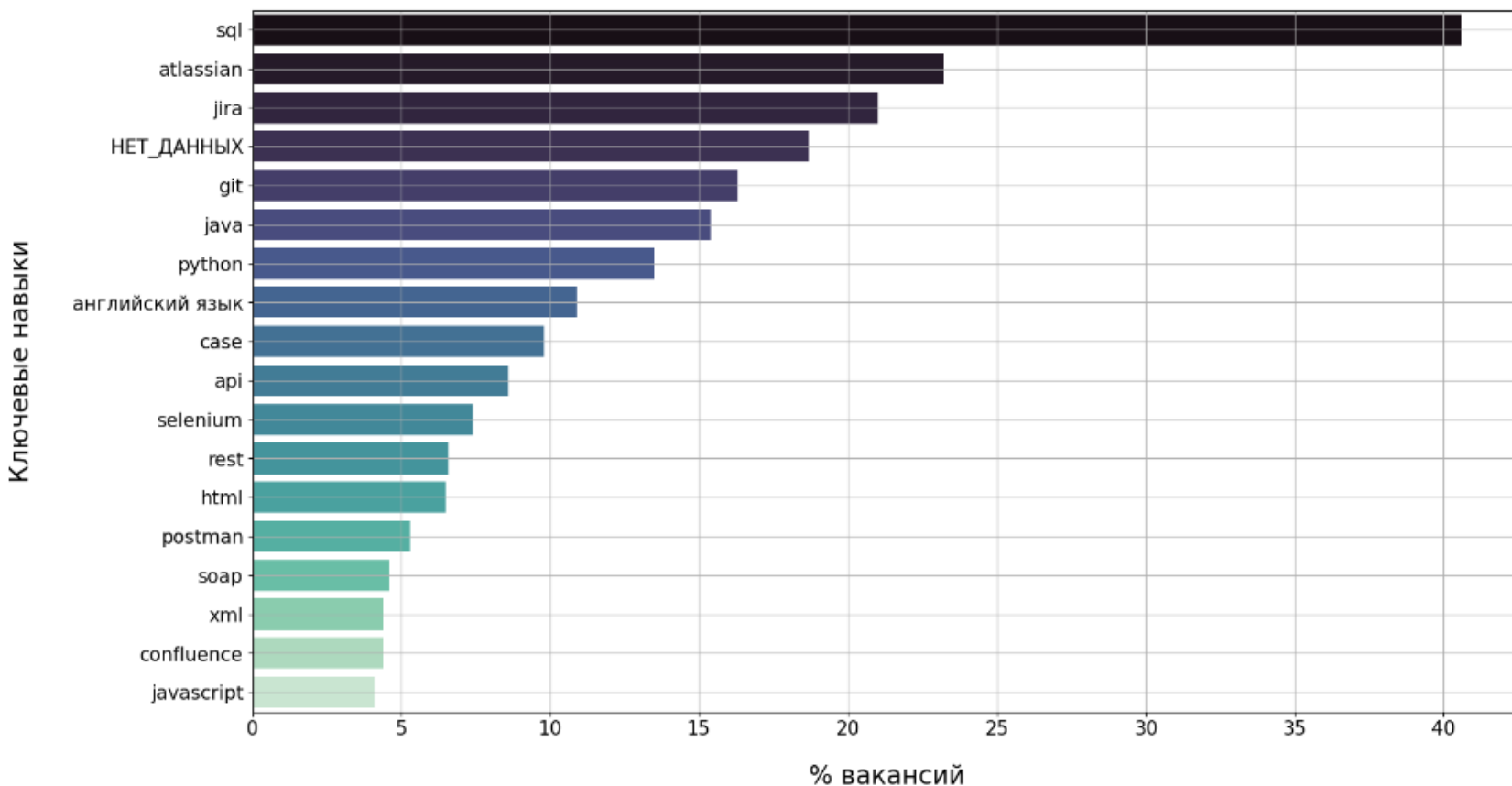
Распределение ключевых навыков в разделе "Аналитик"



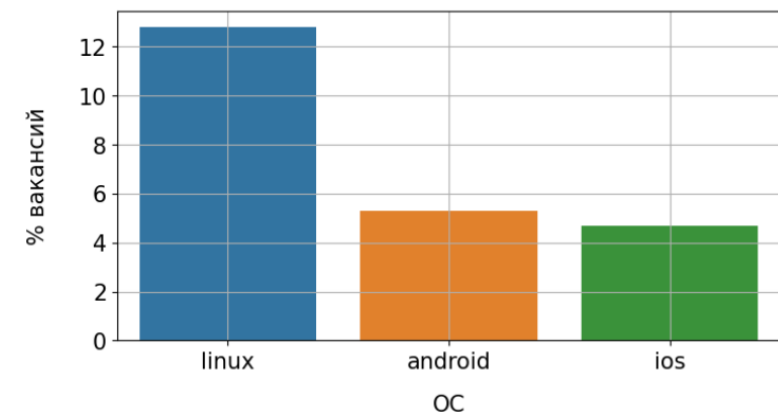
5 Результаты. Тестировщик.



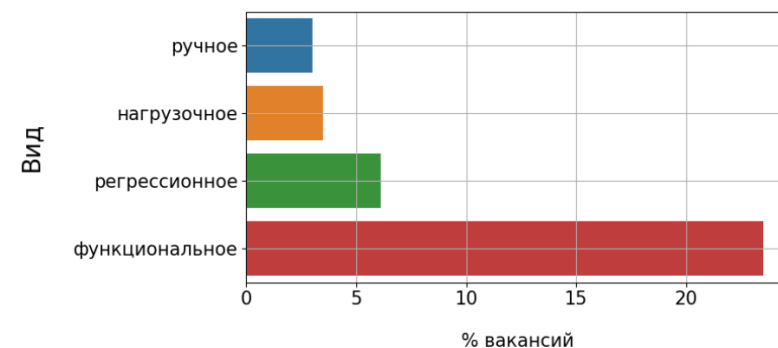
Распределение ключевых навыков в разделе "Тестировщик"



Распределение навыков по операционным системам



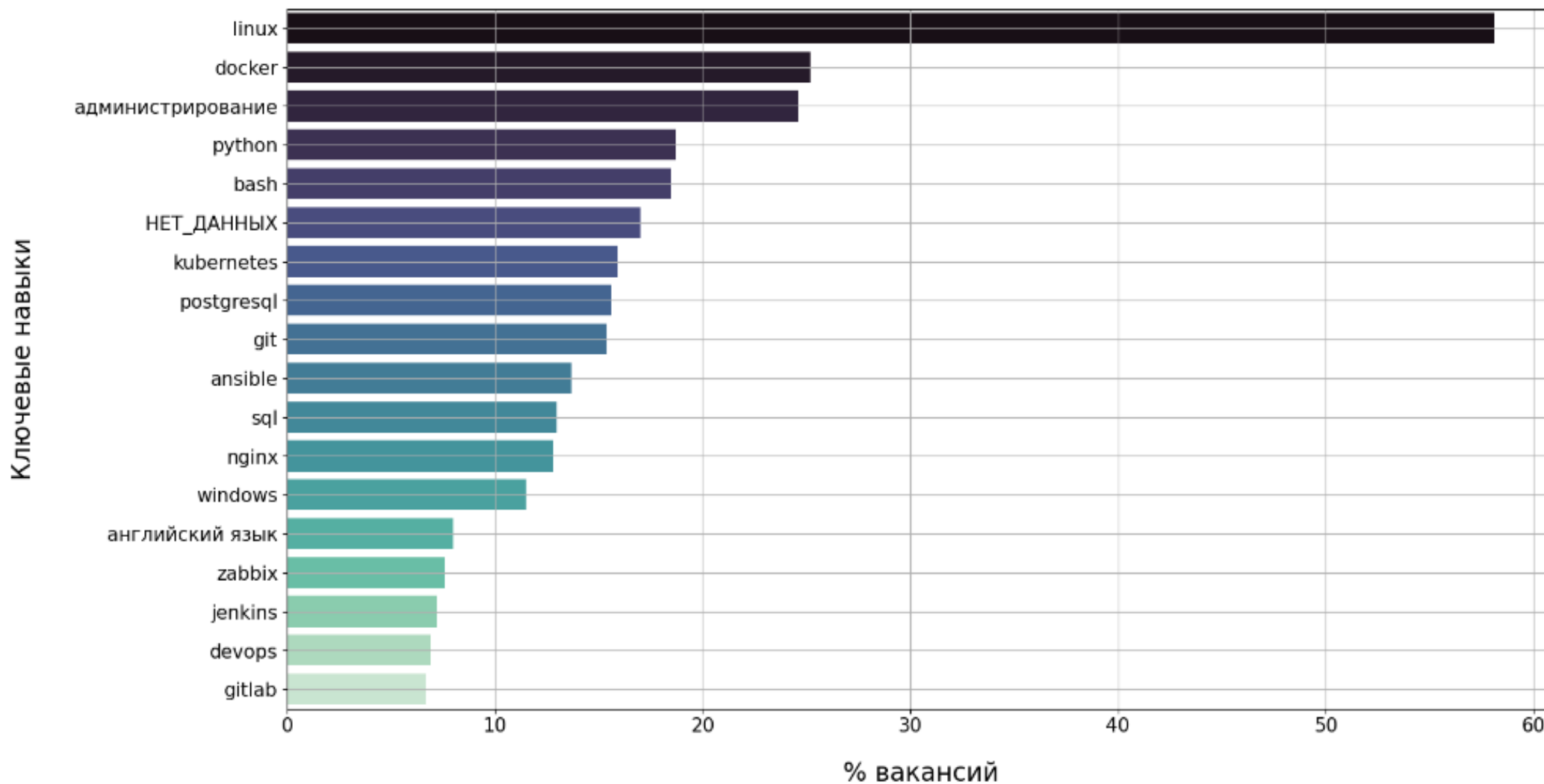
Распределение навыков по виду тестирования



5 Результаты. Системный инженер.



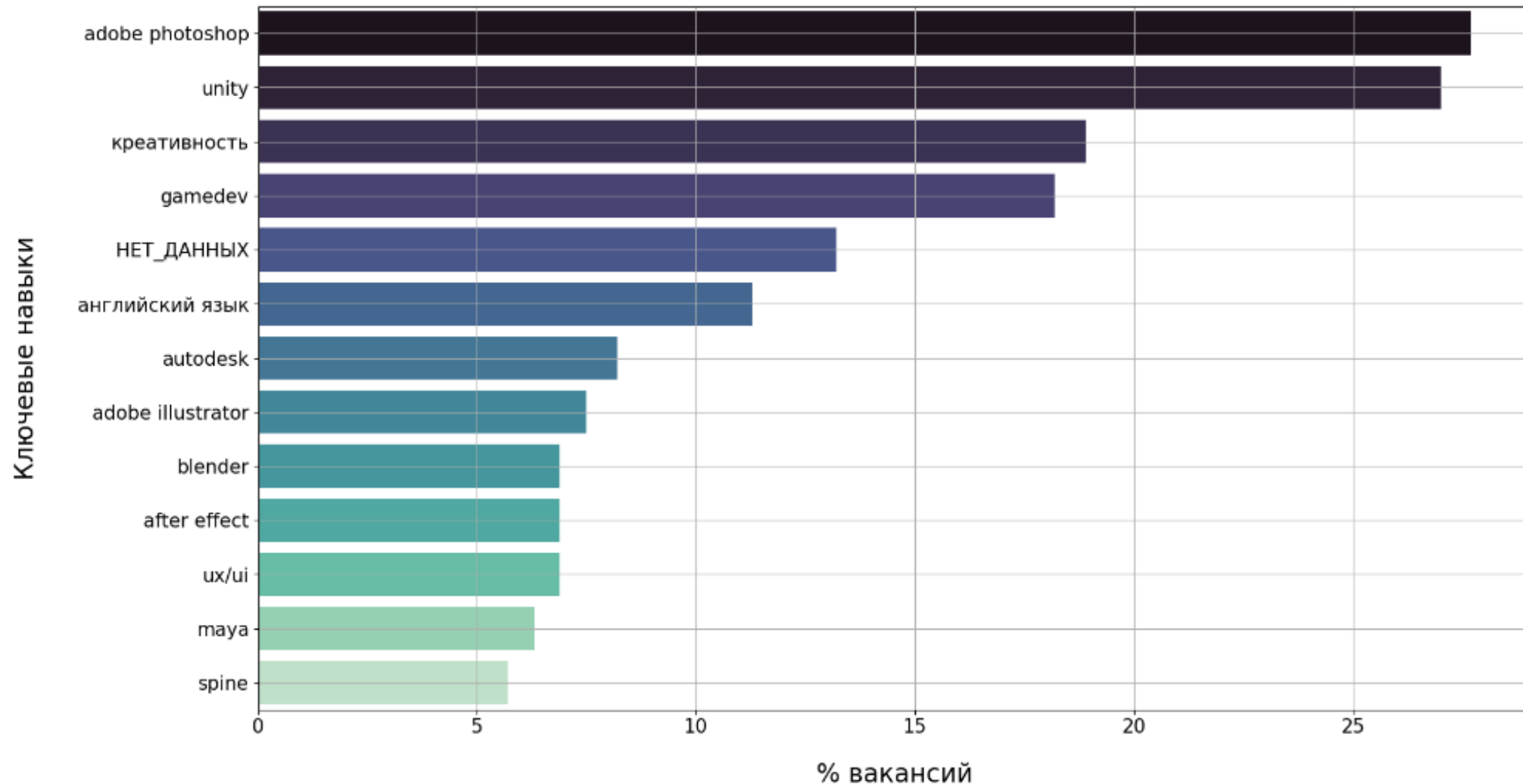
Распределение ключевых навыков в разделе "Системный инженер"



5 Результаты. Гейм дизайнер.



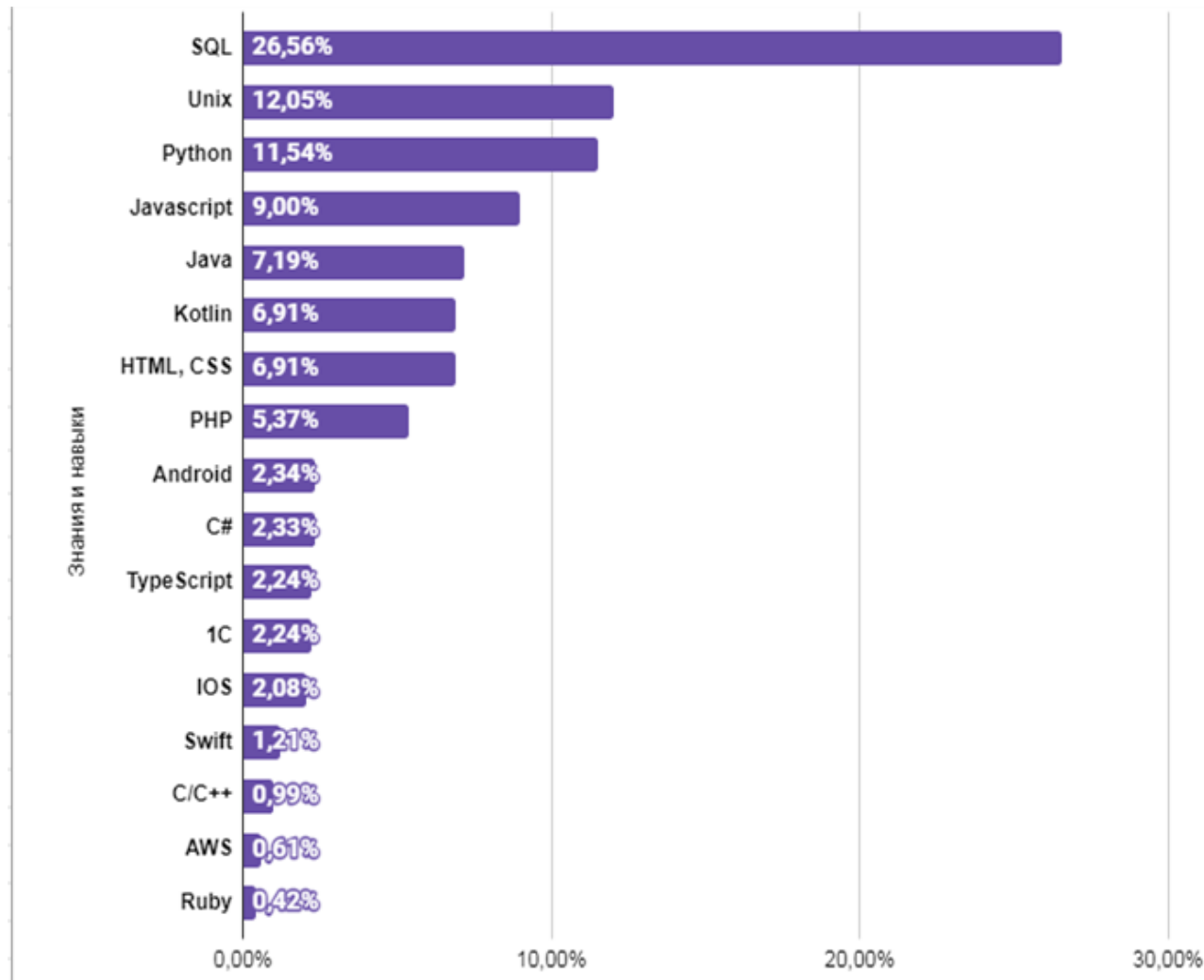
Распределение ключевых навыков в разделе "Гейм дизайнер"



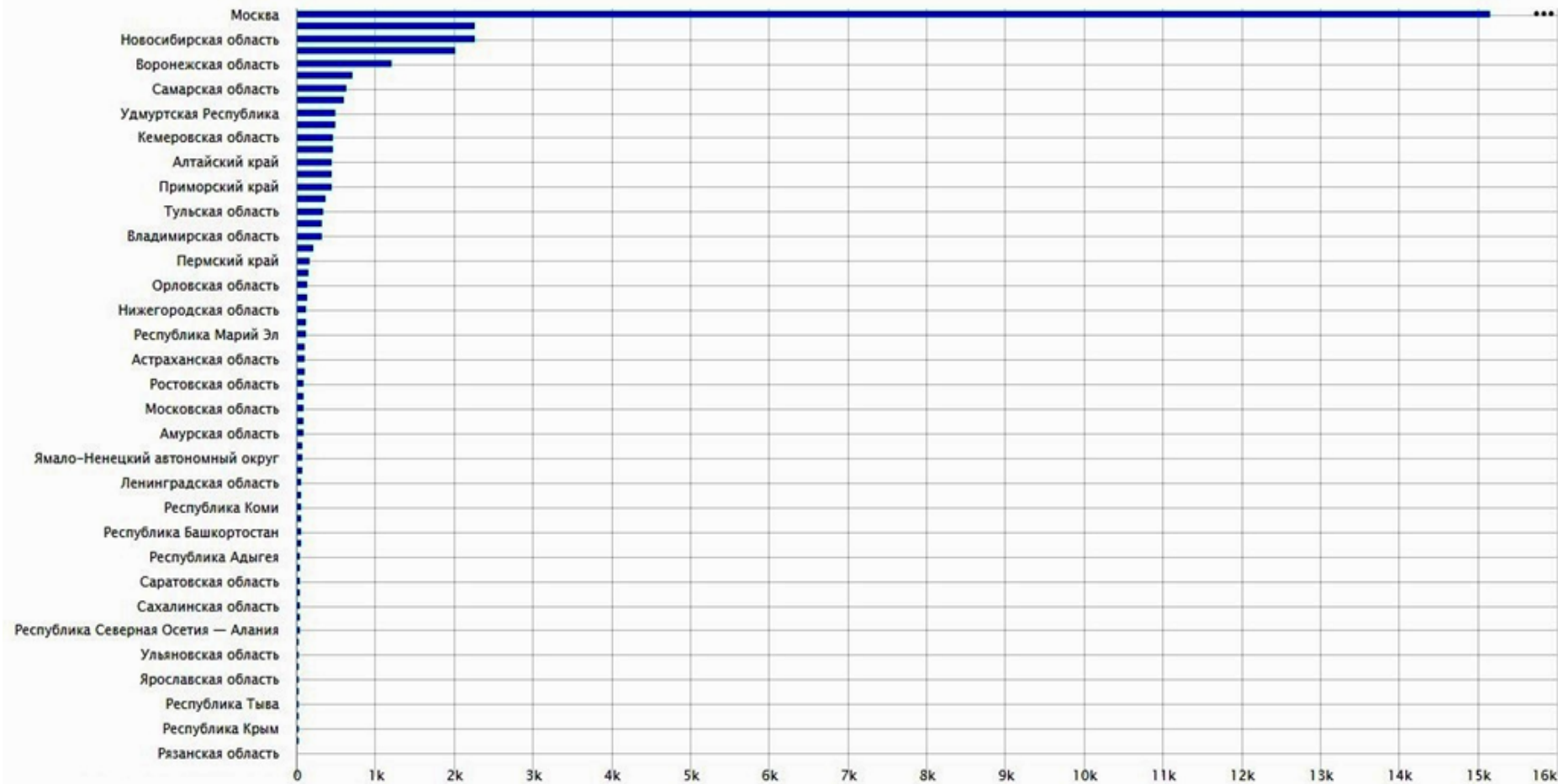
5 Результаты. Разработчики



Для разработчиков востребованы знания и навыки в таких языках программирования как:



5 Результаты. Распределение по городам.

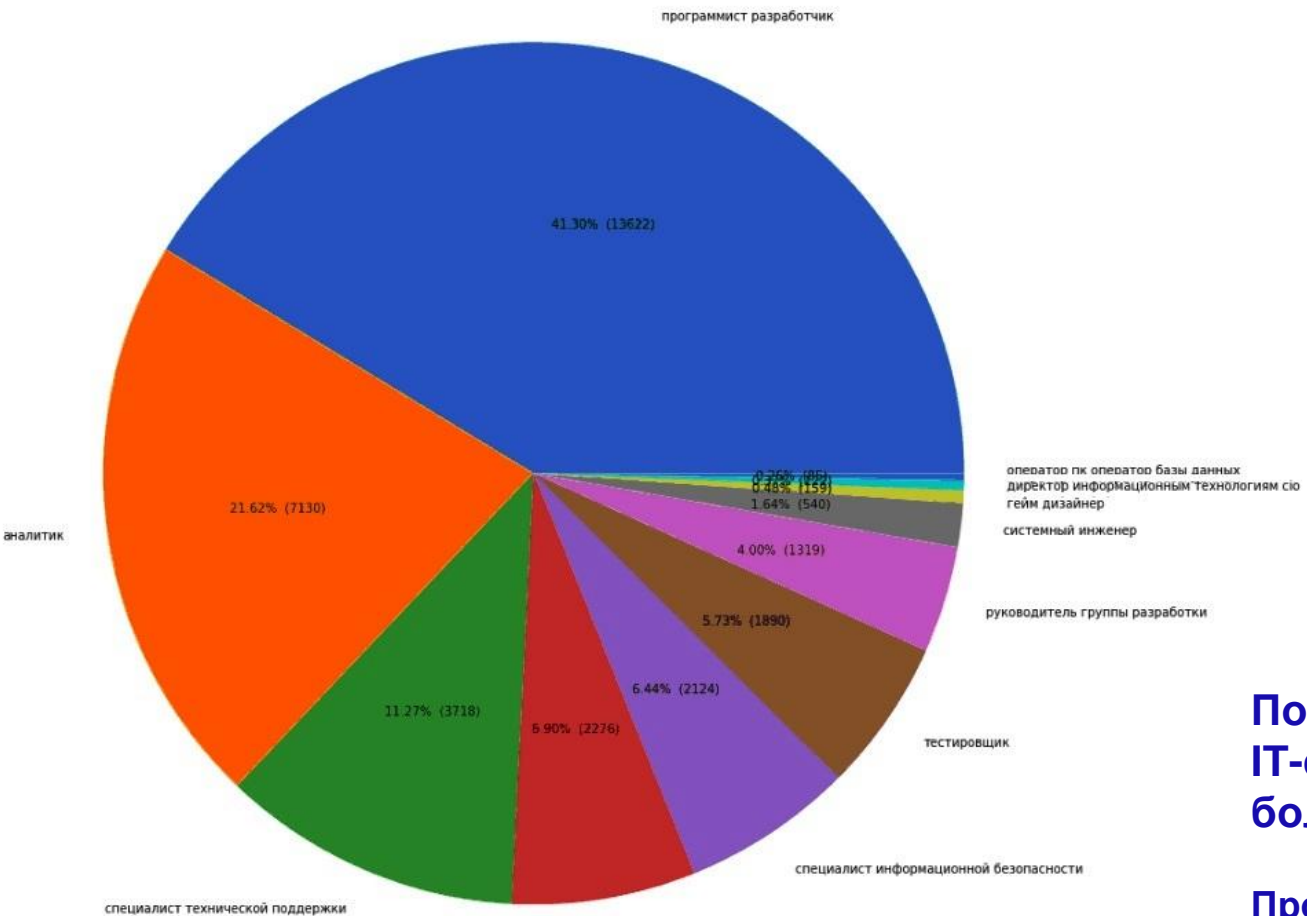


	Количество вакансий
Москва	15 168
Санкт-Петербург	2 266
Новосибирская область	2 264
Республика Татарстан	2 018
Воронежская область	1 213
Томская область	719
Самарская область	644
Омская область	613
Удмуртская Республика	500
Свердловская область	493

Географическим лидером рейтинга по рынку труда специалистов IT является Москва, как мегаполис. Тройка лидеров по регионам:

- Санкт-Петербург, Новосибирская обл., республика Татарстан.

5 Результаты. Разделение по сферам.



	Количество вакансий
Программист, разработчик	13 621
Аналитик	7 129
Специалист технической поддержки	3 718
Системный администратор	2 276
Специалист по информационной безопасности	2 124
Тестировщик	1 890
Руководитель группы разработки	1 319
Системный инженер	540
Гейм-дизайнер	159
Директор по информационным технологиям (CIO)	122
Оператор ПК, оператор базы данных	85

По профессиональному разделению
IT-специалистов работодателей интересуют в
большей степени:

Программист разработчик - 41.3% ,
Аналитик -21,62%,
Специалист технической поддержки - 11.27% ,
остальную долю составляют другие специалисты - 25,81 %

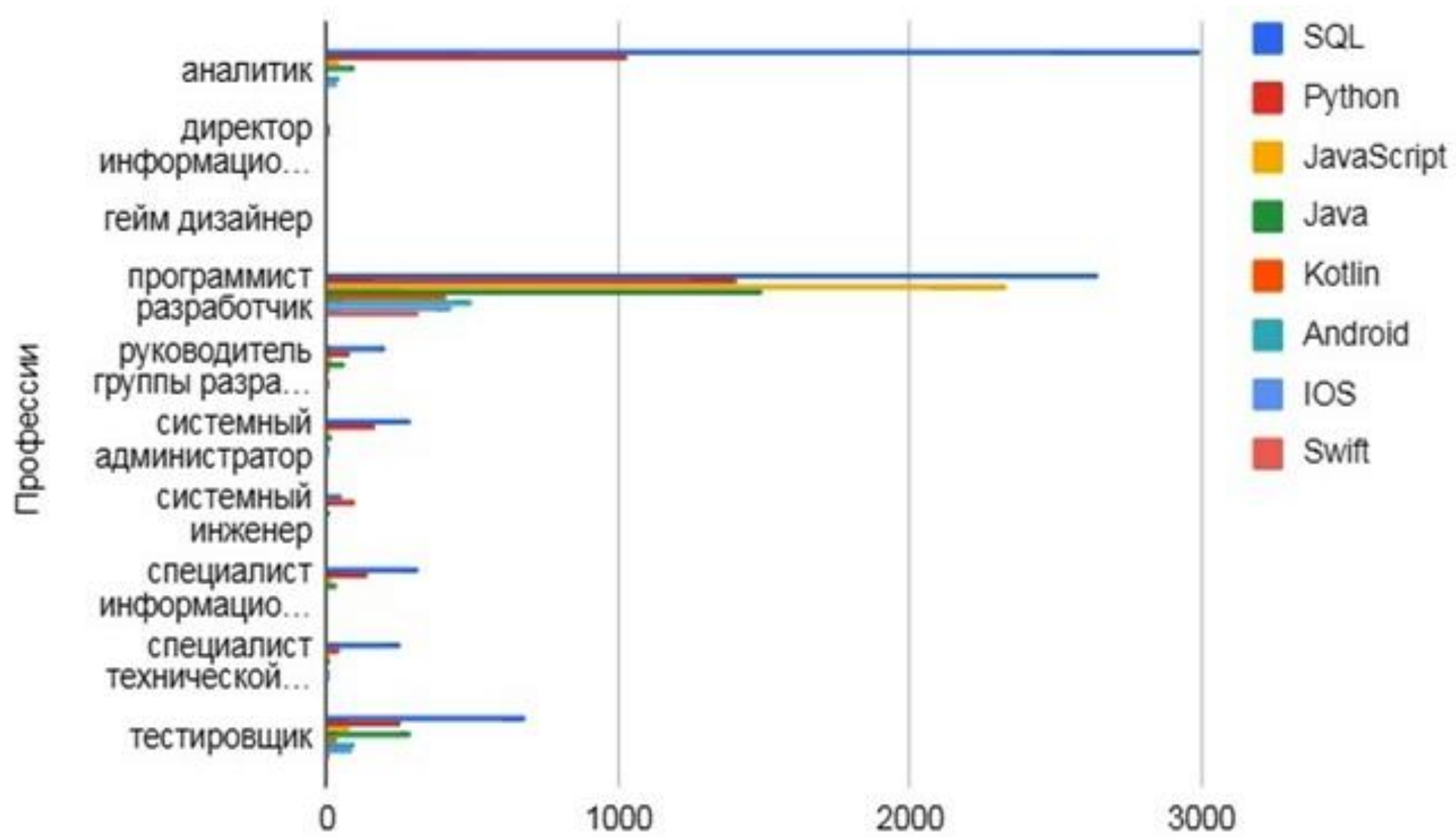
5 Результаты. Навык SQL «по городам».



Города	SQL
Москва	4013
Санкт-Петербург	556
Новосибирск	478
Казань	402
Воронеж	236
Екатеринбург	140
Томск	129
Самара	123
Омск	99
Ижевск	98
Краснодар	93
Владивосток	83
Красноярск	76
Барнаул	75
Иркутск	57
Пермь	51
другие	746
Итого	7455



5 Результаты. Сводная таблица.



5 Результаты. Сводная таблица.



Профессии	SQL	Python	Java Script	Java	Kotlin	Android	IOS	Swift
аналитик	2994	1035	47	95	1	42	39	1
директор информационным технологиям сio	9	9	1	1	0	0	0	0
гейм дизайнер	3	4	0	0	0	3	2	0
программист разработчик	2646	1406	2333	1493	413	495	430	317
руководитель группы разработки	206	80	22	59	9	5	8	9
системный администратор	287	166	9	16	0	6	6	0
системный инженер	57	101	4	13	0	1	0	0
специалист информационной безопасности	314	138	22	38	0	1	4	1
специалист технической поддержки	254	43	10	13	0	10	9	0
тестировщик	685	256	78	289	39	95	88	11
Итого	7455	3238	2526	2017	462	658	586	339

5 Результаты. График работы.



Предлагаемый график работы :

Программисты разработчики:

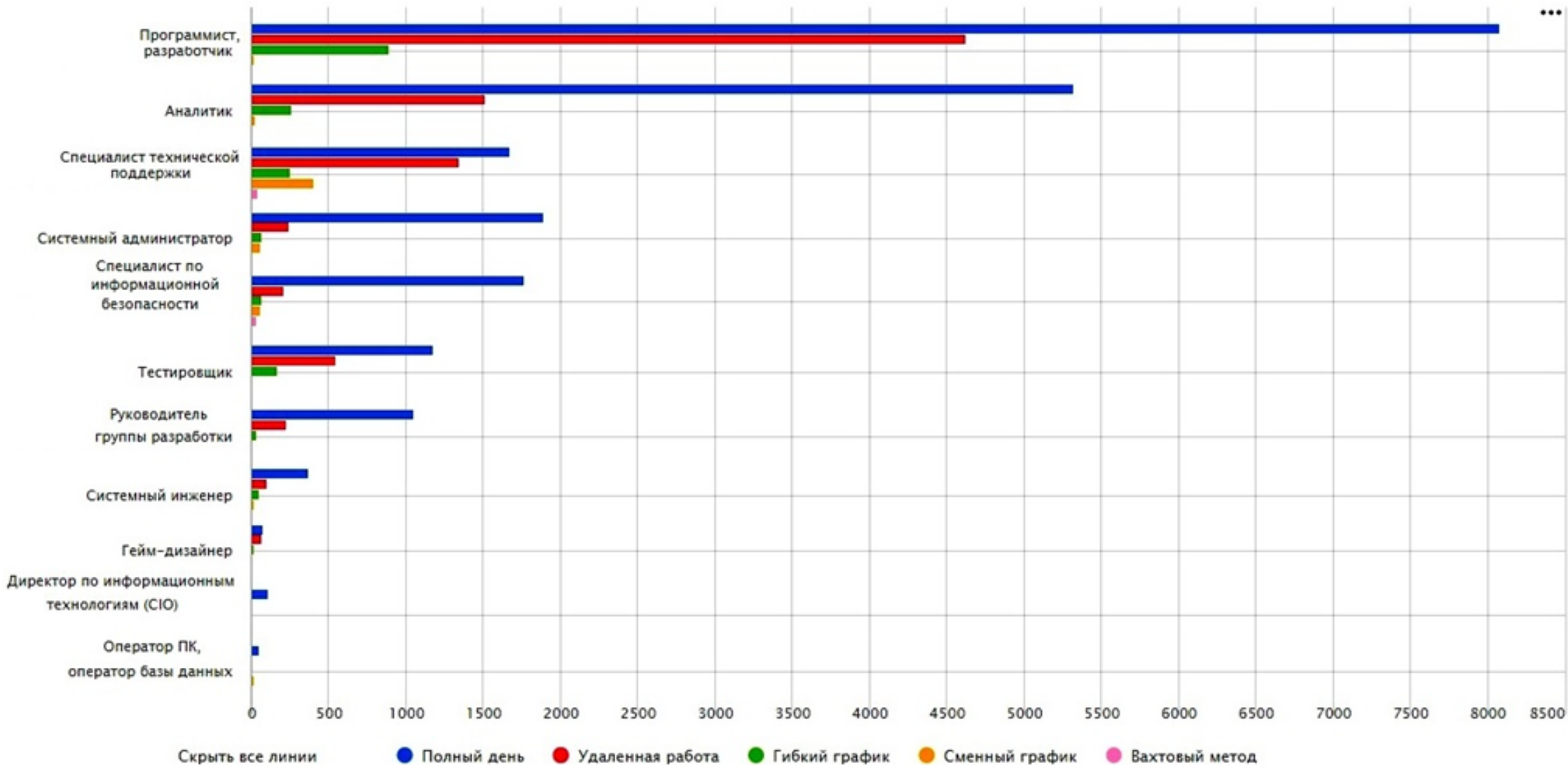
**60% на полный день,
34% на удаленной работе и
6% гибкий график**

В то же время, для специалиста технической поддержки:

- полный день,**
- сменный,**
- гибкий график,**
- удаленная работа**
- вахтовый.**

	Количество вакансий
Полный день	21 570
Удаленная работа	8 900
Гибкий график	1 815
Сменный график	598
Вахтовый метод	100

5 Результаты. График работы.



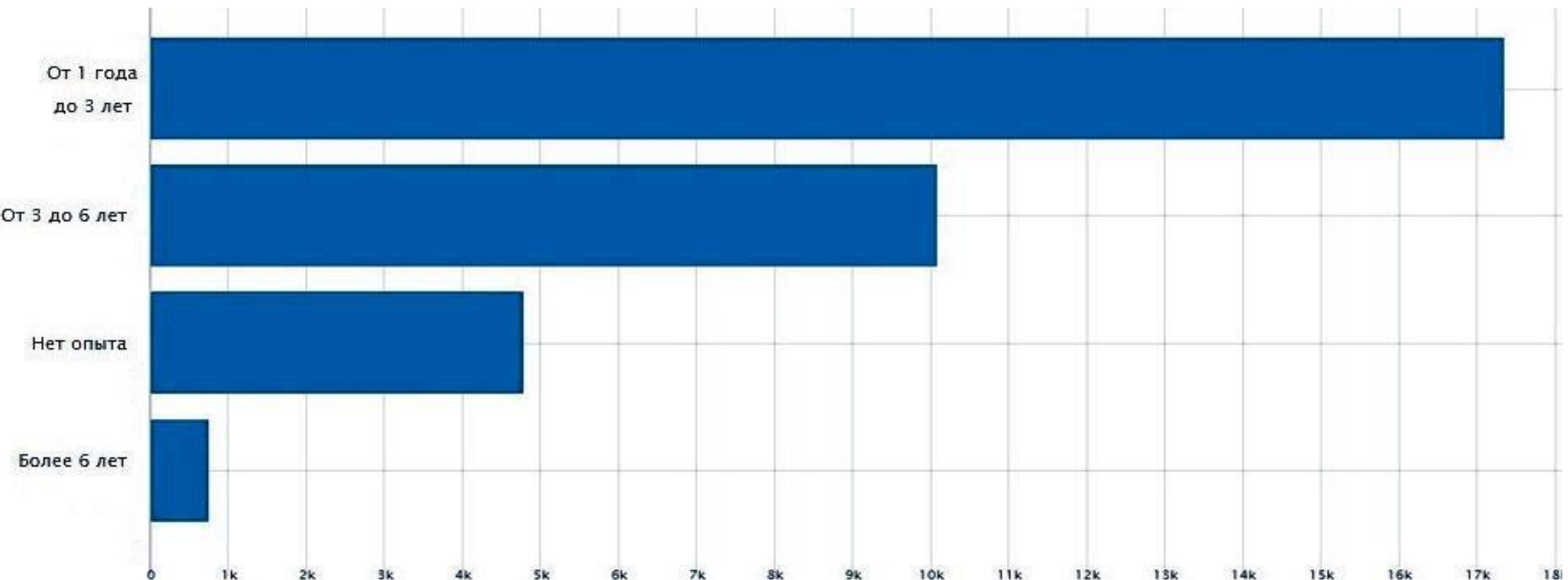
5 Результаты. Стаж.



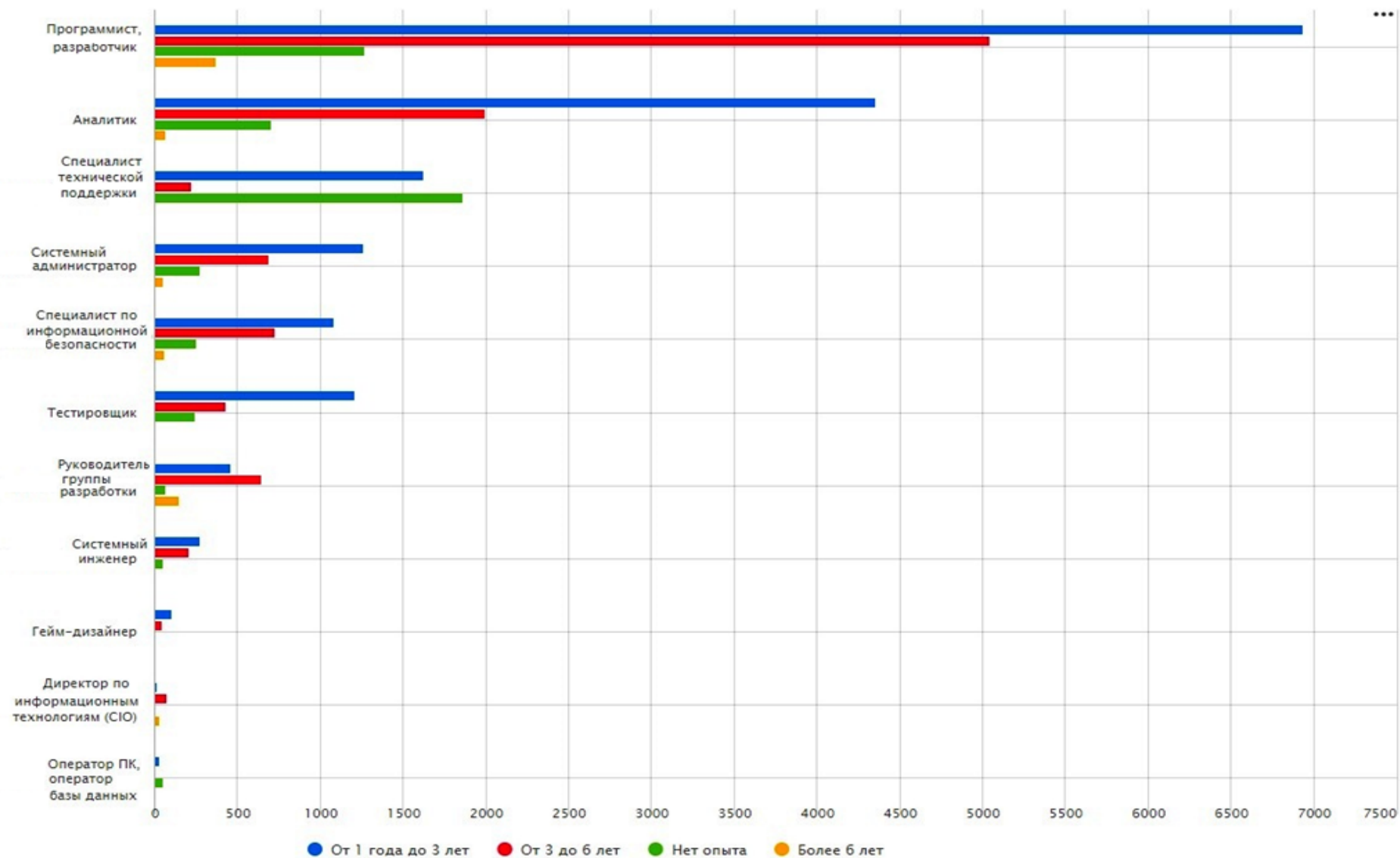
Распределение вакансий по опыту работы

...

	Количество вакансий
От 1 года до 3 лет	17 351
От 3 до 6 лет	10 082
Нет опыта	4 792
Более 6 лет	758



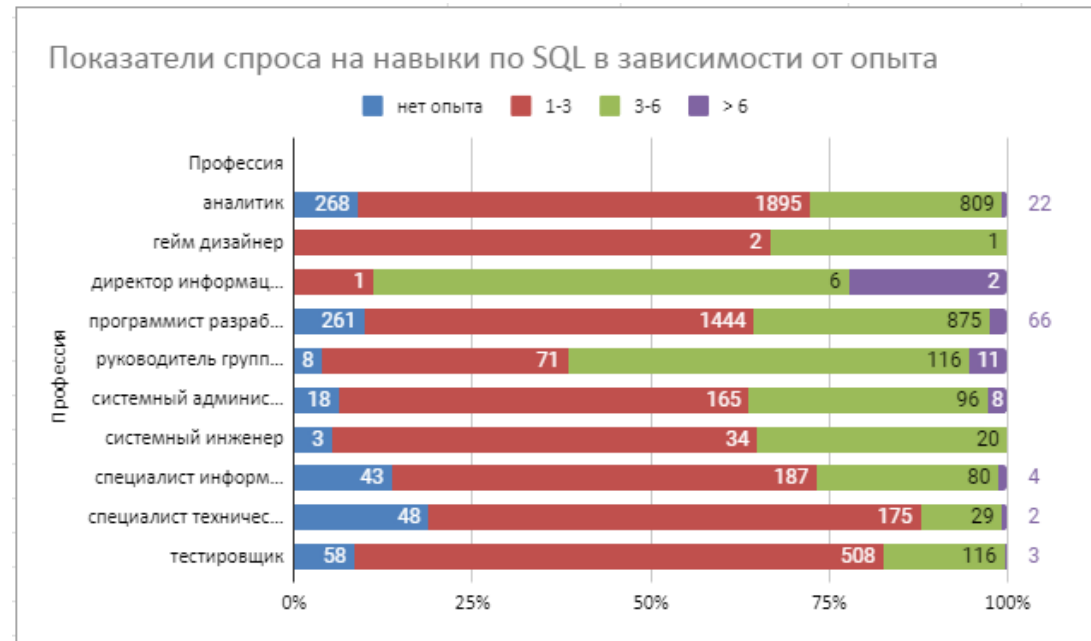
5 Результаты. Стаж.



5 Результаты. Стаж.



Навык SQL	Опыт работы				
Профессия	Нет опыта	От 1 года до 3 лет	От 3 до 6 лет	Более 6 лет	Итого
аналитик	268	1895	809	22	2994
гейм дизайнер	-	2	1	-	3
директор информационным технологиям cjo	-	1	6	2	9
программист разработчик	261	1444	875	66	2646
руководитель группы разработки	8	71	116	11	206
системный администратор	18	165	96	8	287
системный инженер	3	34	20	-	57
специалист информационной безопасности	43	187	80	4	314
специалист технической поддержки	48	175	29	2	254
тестировщик	58	508	116	3	685
Итого	707	4482	2148	118	7455



По навыкам мы видим, что большая часть стажа запрашивается работодателем от 1 до 3 лет

5 Результаты. Зарплата.

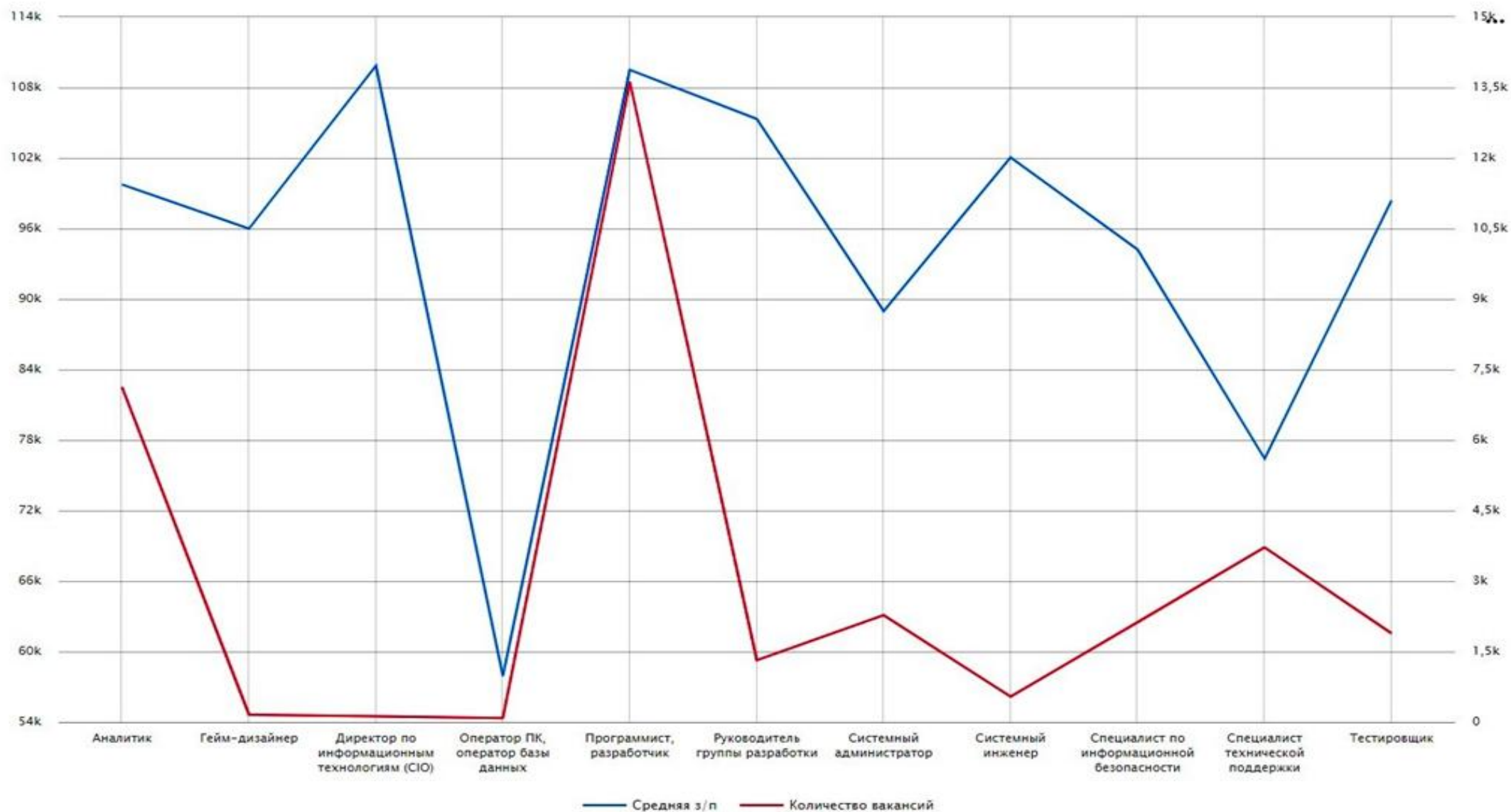


Корреляция спроса на специалистов и средней зарплаты

...

	Количество вакансий	Средняя з/п
Программист, разработчик	13 621	109 519,29
Аналитик	7 129	99 753,46
Специалист технической поддержки	3 718	76 419,73
Системный администратор	2 276	88 965,79
Специалист по информационной безопасности	2 124	94 206,02
Тестировщик	1 890	98 379,35
Руководитель группы разработки	1 319	105 336,18
Системный инженер	540	102 044,35
Гейм-дизайнер	159	96 001,75
Директор по информационным технологиям (CIO)	122	109 867,45
Оператор ПК, оператор базы данных	85	57 926,88

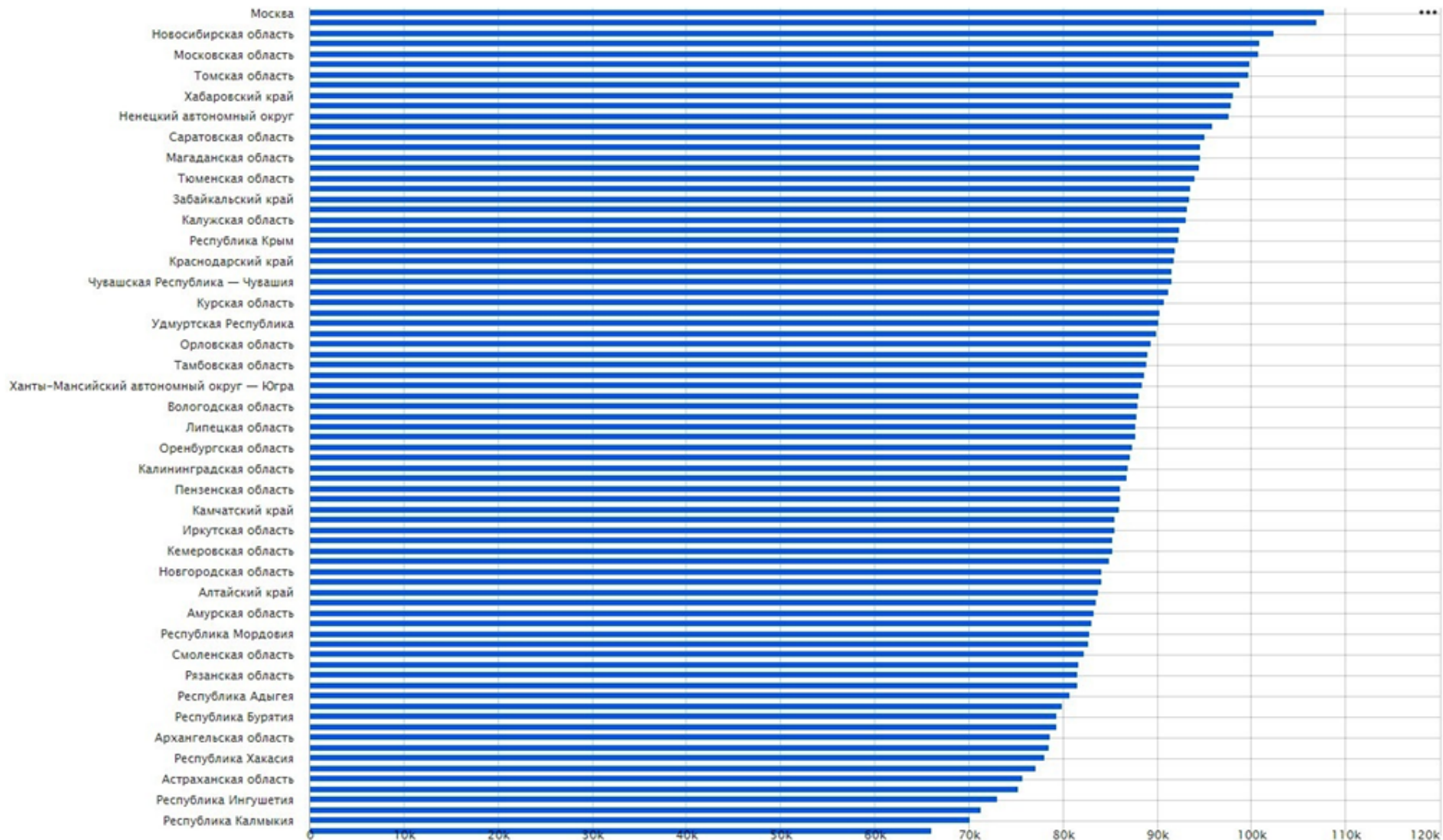
5 Результаты. Зарплата.



5 Результаты. Зарплата.

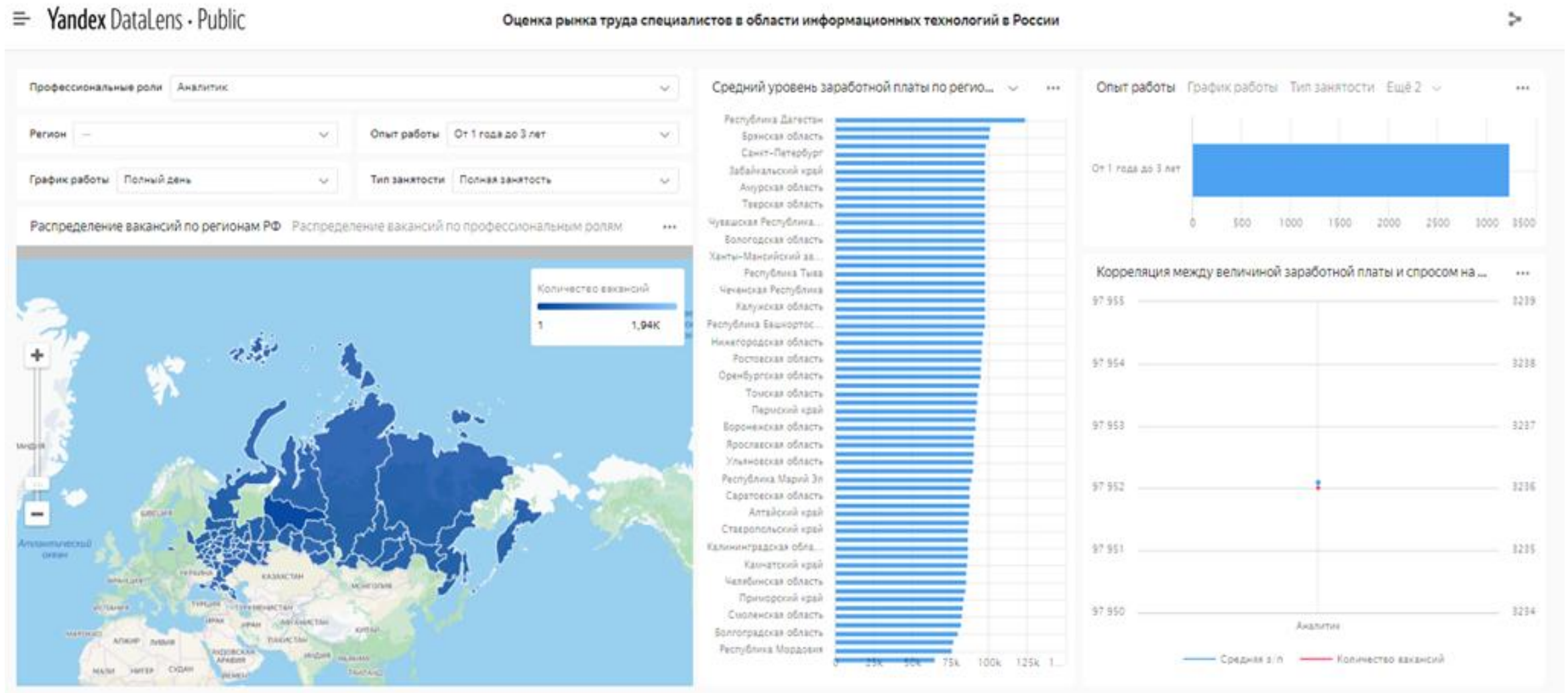


Показатели зарплаты в зависимости от области



5 Результаты. Дашборд.

По результатам исследования был разработан дашборд на платформе Yandex DataLens



<https://datalens.yandex/e1r1pnmzbv3a7?state=287df231396>