

Дипломная работа по теме:

“Анализ рынка вакансий по данным сайта «Хабр Карьера» (парсинг данных, поиск инсайтов, составление рекомендаций стейкхолдерам)”

**Профессия «Аналитик данных», DA-70
Летуновский Михаил Валериевич**

г. Москва, 2023 г.



Бизнес-задачи:

1. Изучить и выявить наиболее востребованные профессии на рынке труда в настоящий момент.
2. Выявить компании, которые в настоящий момент наиболее заинтересованы в найме сотрудником
3. Произвести анализ уровня заработной платы на рынке труда, в зависимости от навыков и опыта кандидата.
4. Определить наиболее востребованные навыки и знания, которые востребованы и пользуются большим спросом.
5. Изучить географию предложений, в каких городах требуются специалисты.

Стейкхолдеры:

1. Образовательные онлайн-платформы, школы, университеты, которые специализируются на подготовке специалистов в IT-сфере.
2. Компании, которые используют в своей работе современные технологические и программные решение, нуждающиеся в найме и развитии IT-специалистов.

```
[ ] df_all.head(5)
```

	vacancy_name	company	remote	link	date	skills	salary
0	Руководитель ИТ-проектов (проектный офис, тран...	Hoff Tech	[Москва , Полный рабочий день]	https://career.habr.com/vacancies/1000124365	13 августа	Менеджер проекта, Старший (Senior) , Agile , У...	NaN
1	Разработчик Fullstack (React, Python)	Bell Integrator	[Москва , Санкт-Петербург , Сочи , Полный р...	https://career.habr.com/vacancies/1000127669	13 августа	Бэкенд разработчик, Средний (Middle) , Python ...	NaN
2	Тестировщик/QA Инженер	ГК Астра	[Москва , Полный рабочий день]	https://career.habr.com/vacancies/1000077423	13 августа	Инженер по ручному тестированию, Средний (Midd...	NaN
3	Технический писатель в команду Salute2B Цифров...	Сбер	[Москва , Полный рабочий день , Можно удаленно]	https://career.habr.com/vacancies/1000112549	13 августа	Технический писатель, Средний (Middle) , Техни...	NaN
4	Project Manager в Цифровые поверхности Салют (...)	Сбер	[Москва , Полный рабочий день , Можно удаленно]	https://career.habr.com/vacancies/1000112551	13 августа	Менеджер проекта, Средний (Middle) , Управлени...	NaN

Для написания парсера были использованы библиотеки Requests и BeautifulSoup. Далее данные преобразовывались в табличную структуру при помощи библиотеки Pandas.

Описание столбцов:

1. vacancy_name - название вакансии
2. company - название компании
3. remote - тип занятости/город локации
4. link - ссылка на вакансию
5. date - дата размещения вакансии
6. skills - навыки
7. salary - уровень заработной платы

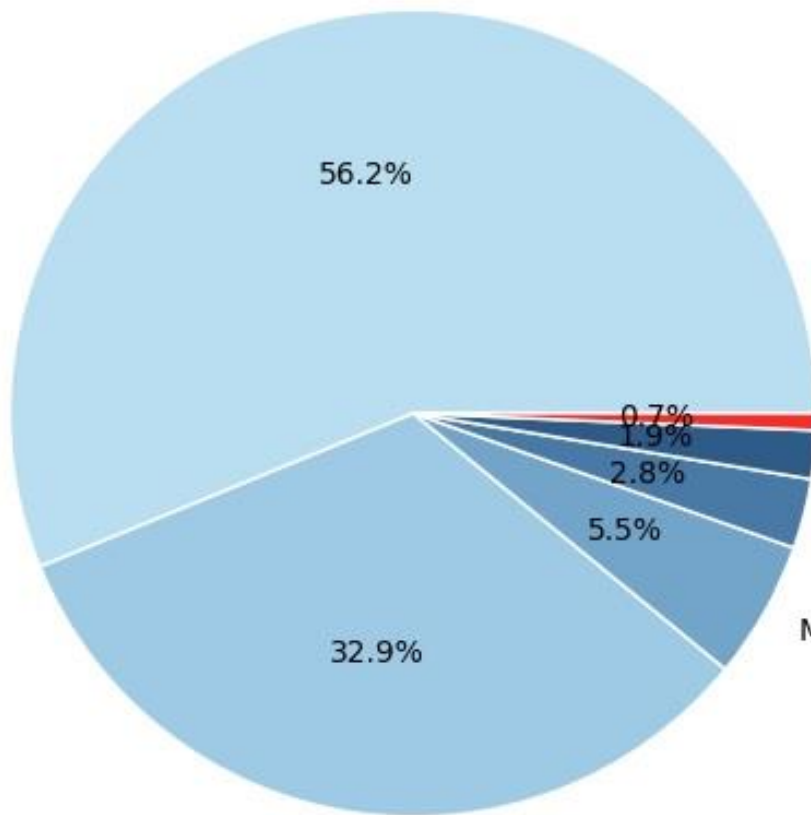
Обработка данных

Данные полученные в результате работы парсера необходимо обработать и привести нужным типам, для удобства их дальнейшего анализа и визуализации.

1. Поиск пропущенных значений.
2. Создание нового столбца `kind_of_work` со значениями типов занятости (Полный рабочий день, Можно удаленно, Неполный рабочий день). Пропущенные значения в данном столбце заполняются значениями «Офис».
3. Удаление типов занятости из столбца `remote`, в нем остаются только названия городов. Пропуски заполняются «Город не указан».
4. Приведения значений в столбце `date` к формату `datetime` (день-месяц-год).
5. Создание нового столбца `salary_from` (зарплата от), куда входят значения нижнего порога зарплат.
6. Создание нового столбца `salary_to` (зарплата до), куда входят значения верхнего порога зарплат.
7. Удаление символов платежных валют (\$, €, £, ₽) из строки, в столбцах `salary_from` и `salary_to`, далее приведение значений к целочисленному виду (`int`).

Классификация типов занятости

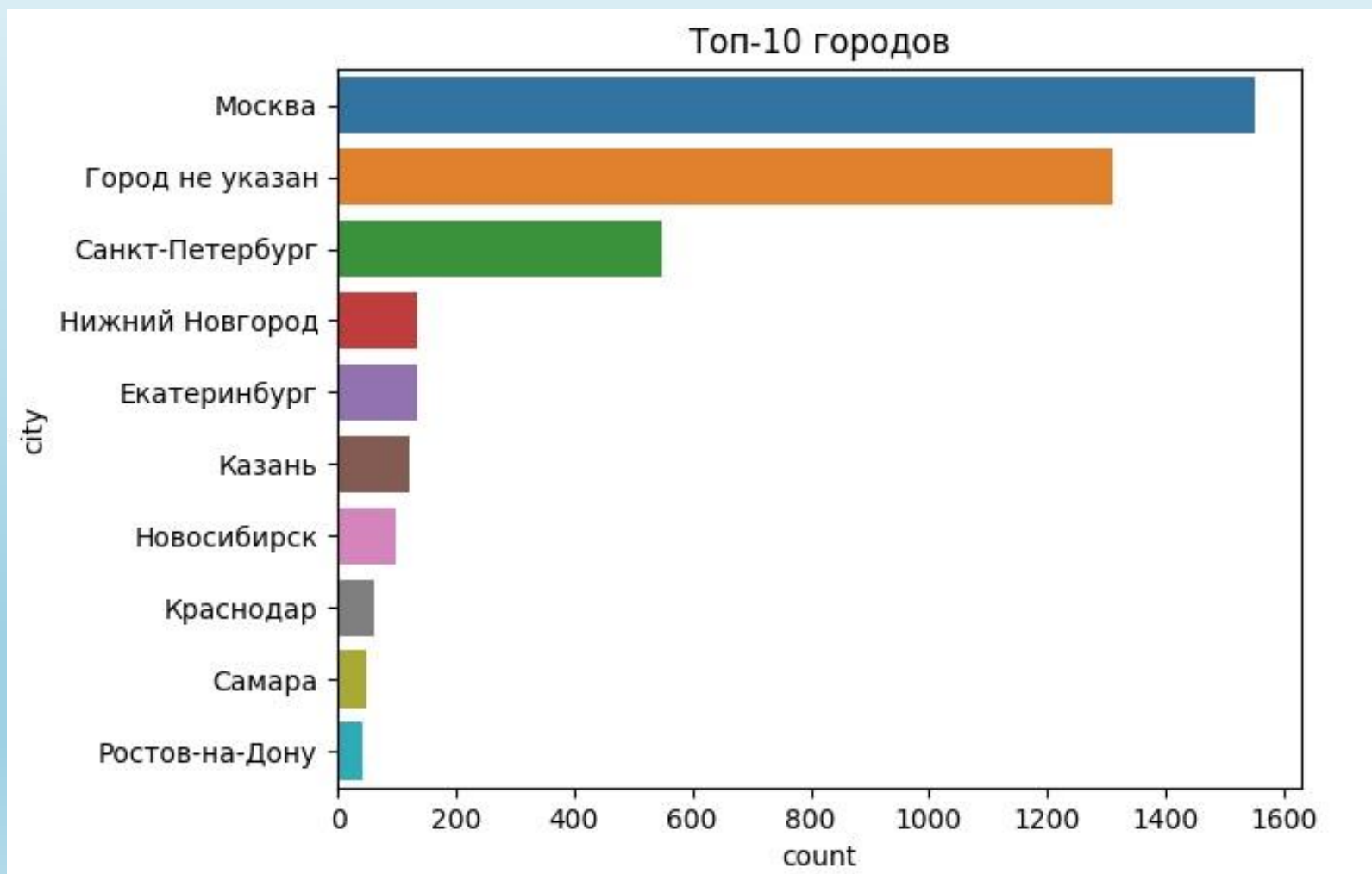
Полный рабочий день, Можно удаленно



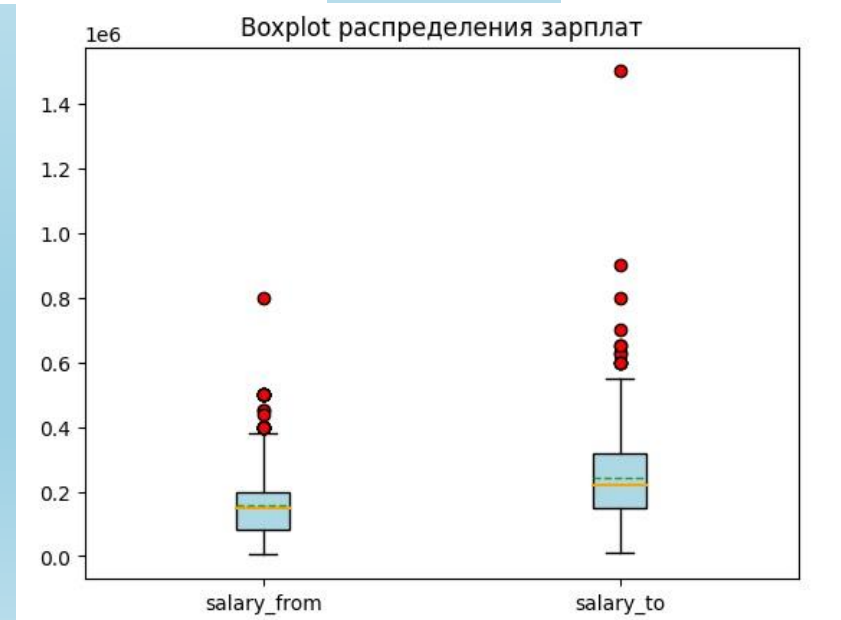
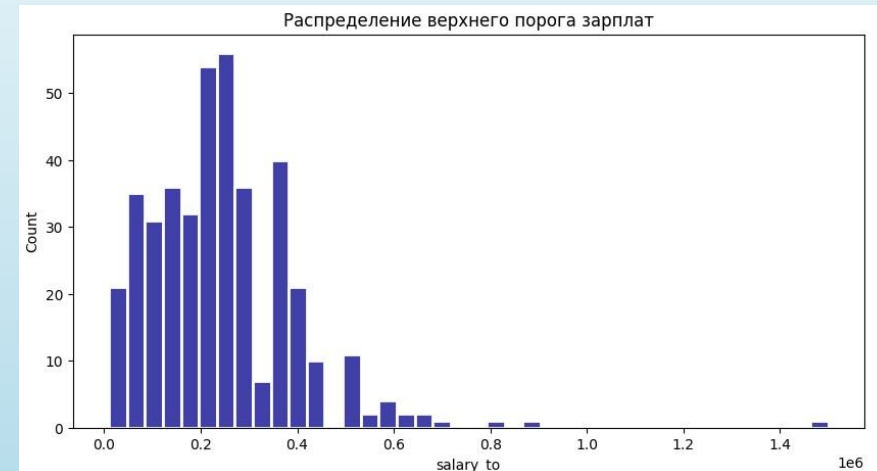
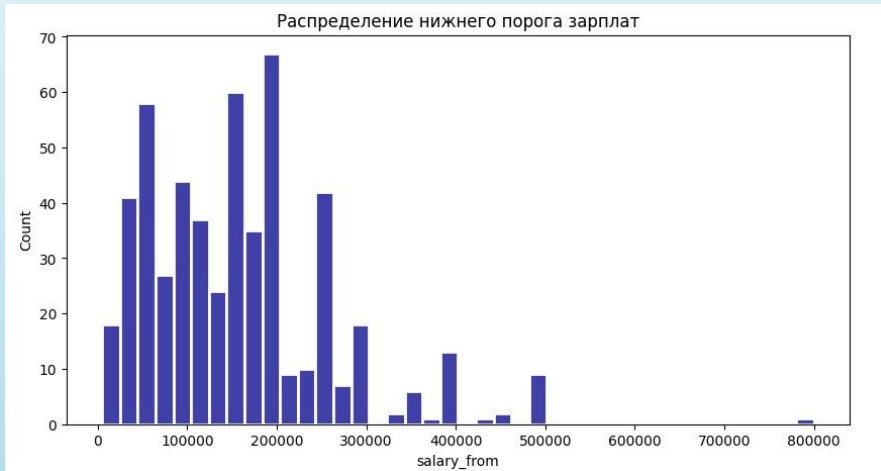
Полный рабочий день

Неполный рабочий день
Неполный рабочий день, Можно удаленно
Офис
Можно удаленно

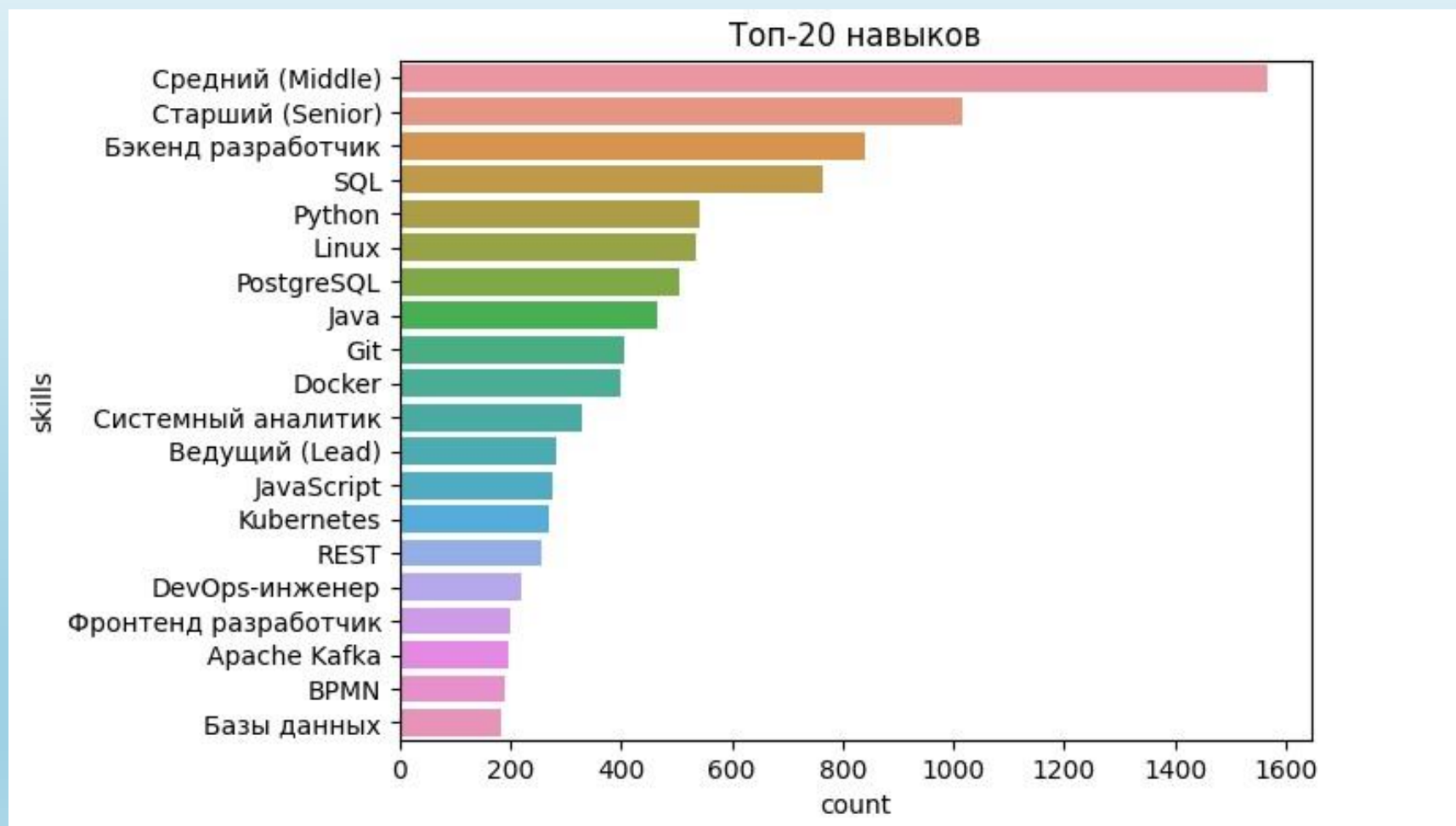
Топ-10 городов по количеству вакансий



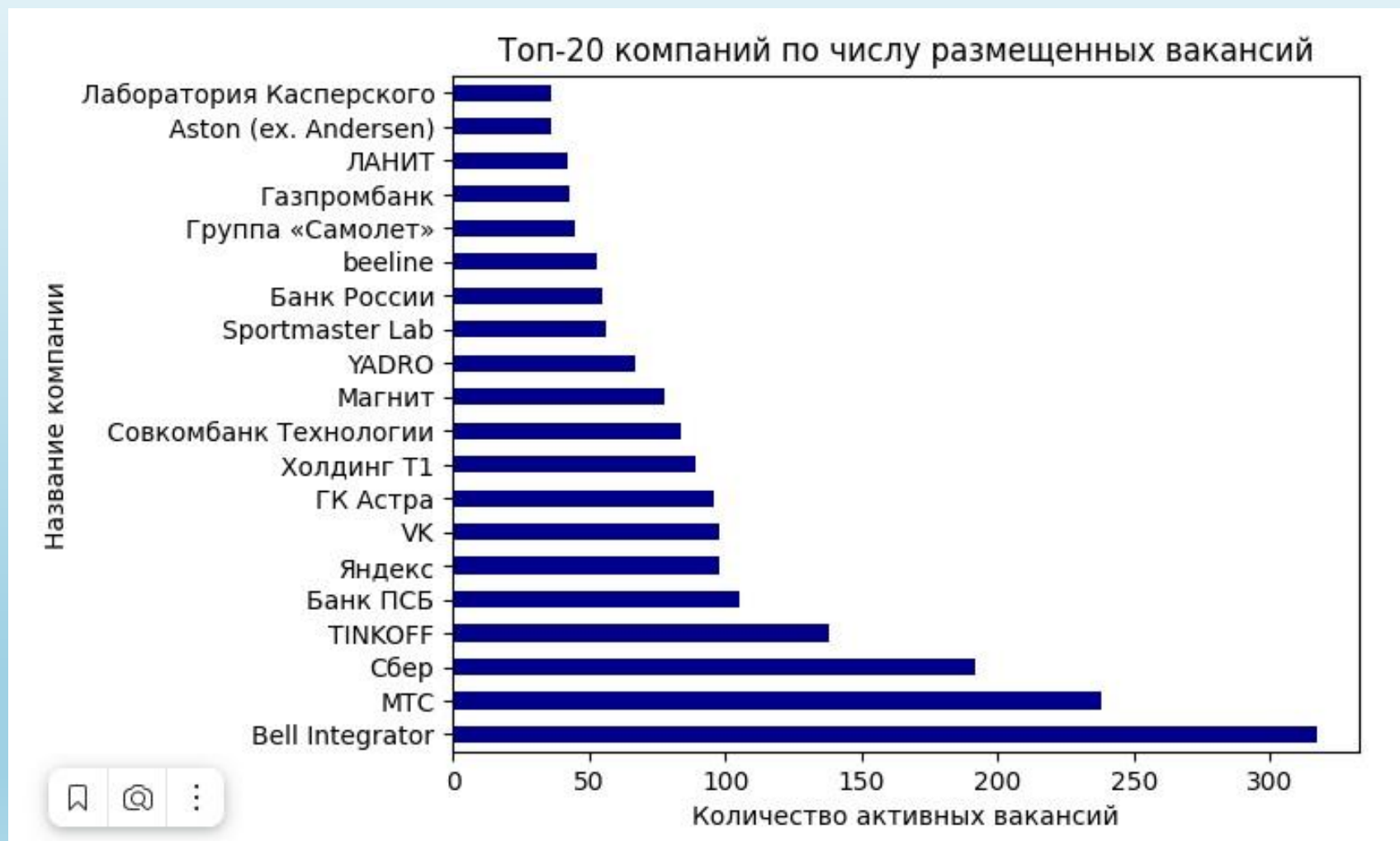
Распределение уровня зарплат



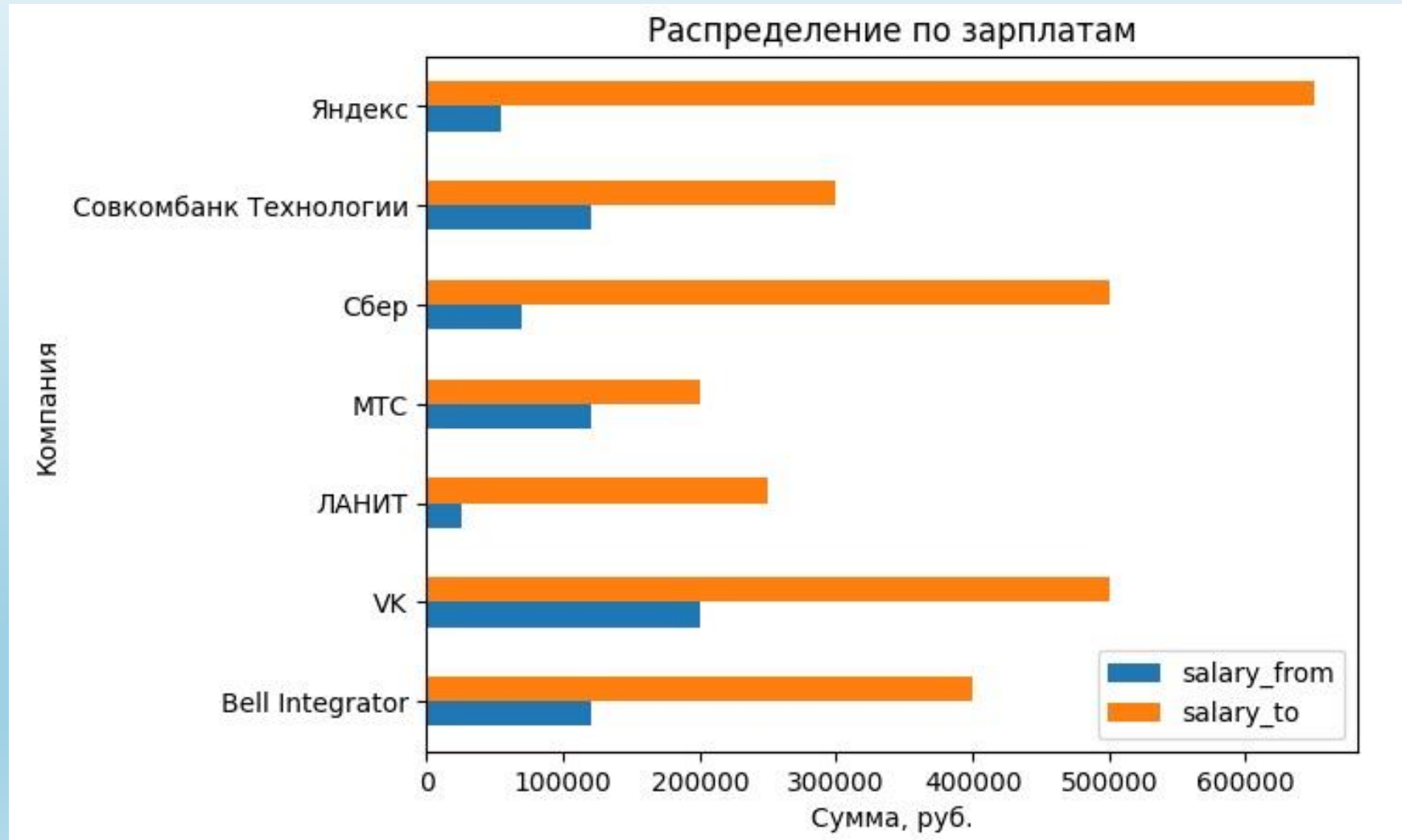
Топ-20 навыков (hard skills)



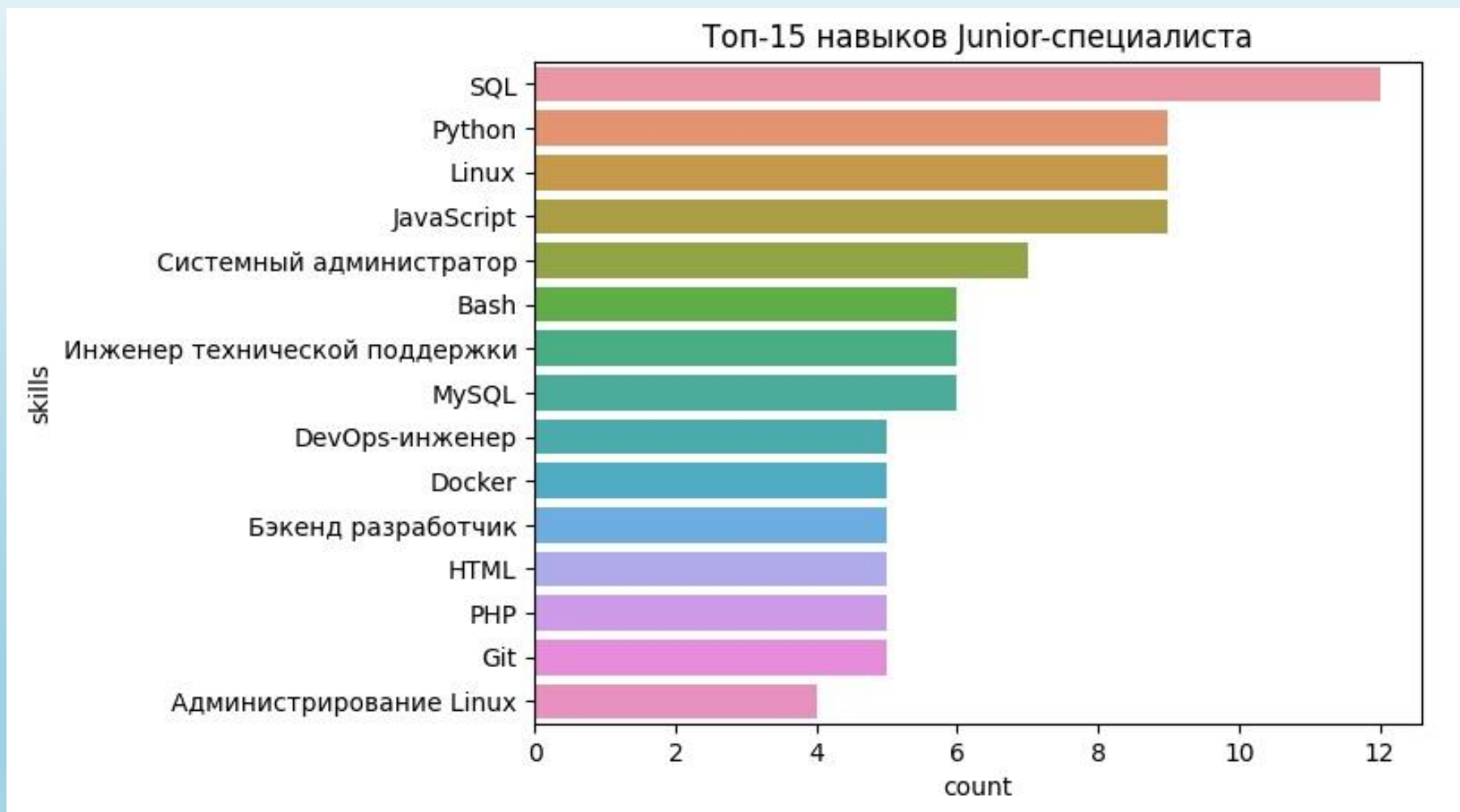
Топ-20 компаний по количеству открытых вакансий



Распределение по размеру зарплаты в компаниях из списка топ-20



Топ-15 навыков Junior-специалиста



Итоги проекта и заключение

- В ходе исследования было выявлено, что в настоящий момент рынок в первую очередь нуждается в специалистах уровня middle и senior. Более чем в 60% случаев работодатель готов предложить сотруднику удаленный формат работы.
- Наиболее востребованной профессией на рынке является backend-разработчик.
- Работодатели в первую очередь хотят, чтобы сотрудники владели следующими навыками: SQL, Python, Linux, Java.
- Самой востребованной профессией является «Системный аналитик».
- В среднем работодатели готовы платить сотруднику от 80 000 до 316 000 рублей, все конечно сильно зависит от опыта и навыков.
- Самым привлекательным городом для поиска работы является город Москва.

Рекомендации стейкхолдерам

- подбирать и корректировать учебные программы согласно требованиям рынка.
- при проведении рекламных компаний опираться на достоверные данные.
- выявлять компании, которым нужны квалифицированные кадры.
- оценивать рынок труда, чтобы избегать утечки кадров к конкурентам (когда дело только в зарплате)
- получать информацию о городах, чтобы находить свободные локации, для расширения географии присутствия.