

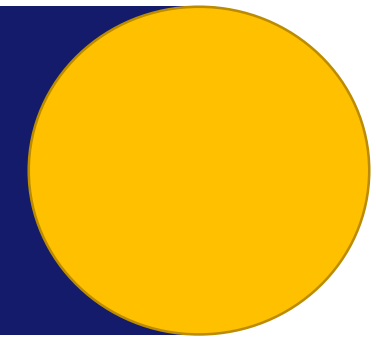


# Оценка рынка труда IT- специалистов в России



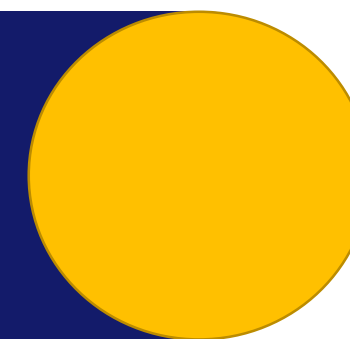
Турчин Максим Владимирович

# Структура



- Постановка задачи
- Гипотезы
- Цели и задачи
- Аналитическая модель проекта
- План реализации исследования
- Используемые методы и подходы
- Источники данных
- Решение
- Результаты

# Постановка задачи. Гипотеза



## ■ Постановка задачи:

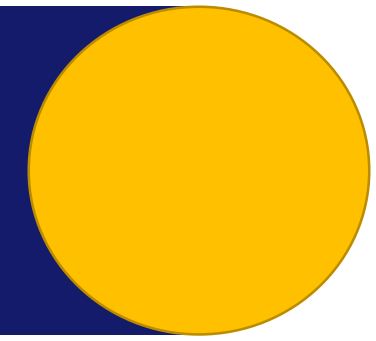
В связи с ростом онлайн-образования остро встает проблема оценки рынка с целью переработки образовательных программ.

Необходимо понять, какие специалисты наиболее востребованы на рынке, какие квалификационные требования к ним предъявляются, какие умения и знания необходимо им передать в процессе обучения.

## ■ Гипотеза:

- Рынок вакансий позволяет получить сведения о необходимых компетенциях, умениях и знаниях, предъявляемых к IT-специалистам

# Цели и задачи



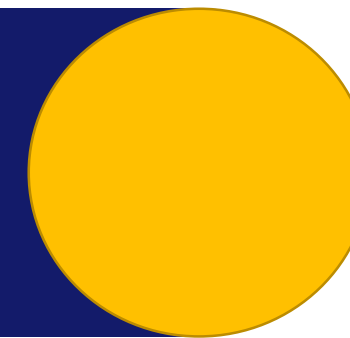
## ■ Цель:

Разработать математическую модель оценки потребности рынка в IT-специалистах.

## Задачи:

- получить результаты анализа о потребности рынка в IT-специалистах;
- Получить данные о требованиях, предъявляемых к соискателям на IT-должности;
- Получить данные о навыках и умениях соискателей на IT-должности;
- Выяснить условия, предоставляемые работодателями соискателям.
- Получить отраслевую и профессиональную структуру IT-рынка в РФ.

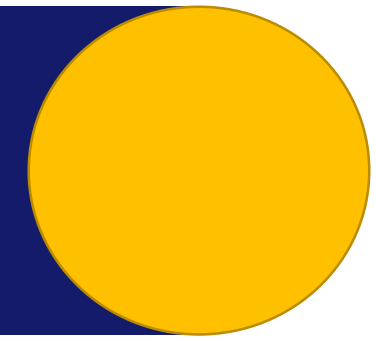
# Аналитическая модель проекта



## ■ Аналитическая модель проекта

Из открытых источников получить данные по IT-вакансиям,  
На основании этих данных  
проанализировать требования к специалистам,  
выяснить наиболее востребованные профессии  
выяснить наиболее нуждающиеся отрасли  
выявить условия, предлагаемые работодателями

# План реализации исследования

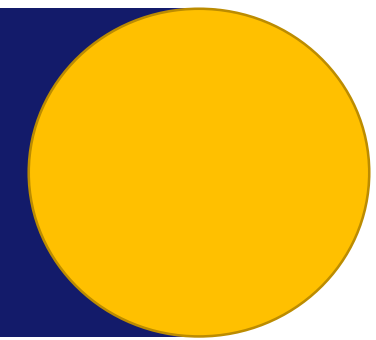


## ■ План реализации исследования:

- Изучить структуру данных вакансий (hh.ru)
- Собрать данные с сайта вакансий (hh.ru)
- Произвести предобработку и нормализацию данных
- Построить математическую модель анализа
- Визуализировать результаты
- Сделать выводы по полученным результатам



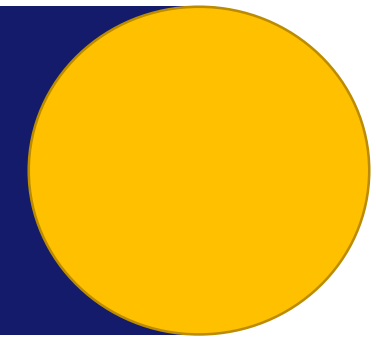
# Используемые методы и подходы



Используемые методы и подходы:

- Для сбора данных был написан код парсера на Python, предусматривающий работу с api сайта hh.ru.
- Для дополнительного сбора данных был построен вспомогательный парсер (на основе сбора HTML кода с помощью Xpath - сбор данных о работодателях)
- Для обхода ограничений api hh.ru по получению данных была применена БД MongoDB (исключение дублирования и дополнительные возможности по сохранению сведений в случае прерывания связи или отказа удаленного сервера).
- Для анализа текстовой информации применено тематическое моделирование

# Источники данных



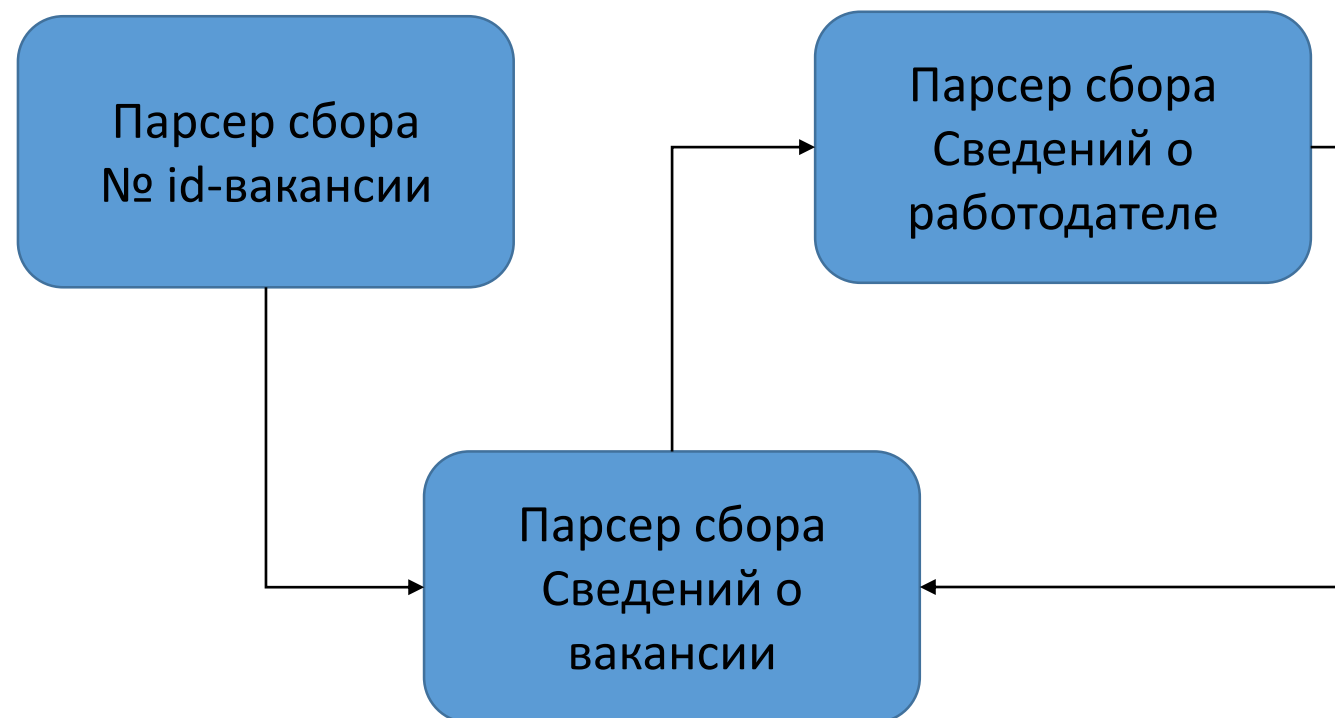
Источники данных:

- Наполнение модели осуществлялось на основе справочников о вакансиях, специализациях, работодателях, резюме и т.д. сайта «hh.ru»



# Решение

## 1. Сбор данных.



## Данные по регионам

	region_id	region_name	area_id
0	1620	Республика Марий Эл	4228
1	1620	Республика Марий Эл	1621
2	1620	Республика Марий Эл	1622
3	1620	Республика Марий Эл	4229
4	1620	Республика Марий Эл	61
...	...	...	...
105	1530	Ростовская область	4562
106	1530	Ростовская область	1552
107	1530	Ростовская область	4563
108	1530	Ростовская область	5299
109	1530	Ростовская область	4564

3920 rows x 3 columns

## Датафрейм с данными

	_id	area	area_id	description	employer	employment	experience	industry	key_skills	name	...	_prof_roles
0	49133800	Рубцовск	1226	<strong>Обязанности: </strong> <ul> <li>Админис...	Глобал Логистик	Полная занятость	Нет опыта	["Перевозки, логистика, склад, ВЭД"]	{["name":"1С: Предприятие 8"], {"name":"1С прог...	Начинающий программист 1С	...	Програ разра
1	44460274	Рубцовск	1226	<strong>Обязанности: </strong> <ul> <li> <p>Обе...	Глобал Логистик	Полная занятость	От 1 года до 3 лет	["Перевозки, логистика, склад, ВЭД"]	{["name":"1С: Зарплата и управление персоналом...	Программист 1С	...	Програ разра
2	49111123	Рубцовск	1226	<strong>Обязанности: </strong> <ul> <li>Обеспеч...	Глобал Логистик	Полная занятость	Нет опыта	["Перевозки, логистика, склад, ВЭД"]	{["name":"Настройка ПК"}, {"name":"Настройка ПО...	Специалист по информационным технологиям	...	Спец техни под
3	49894148	Барнаул	11	<strong>Обязанности: </strong> <ul> <li>Работа ...	Гранат	Полная занятость	От 1 года до 3 лет	["Перевозки, логистика, склад, ВЭД"]	{["name":"1С программирование"], {"name":"1С: П...	Программист 1С8	...	Програ разра
4	49873900	Барнаул	11	<p><em>Сегодня Группа «ПАРФЮМ» — это националь...	Группа компаний Парфюм	Полная занятость	От 1 года до 3 лет	["Перевозки, логистика, склад, ВЭД", "Товары на...	{["name":"1С программирование"], {"name":"1С: У...	Программист 1 С/ Разработчик	...	Програ разра

# Решение

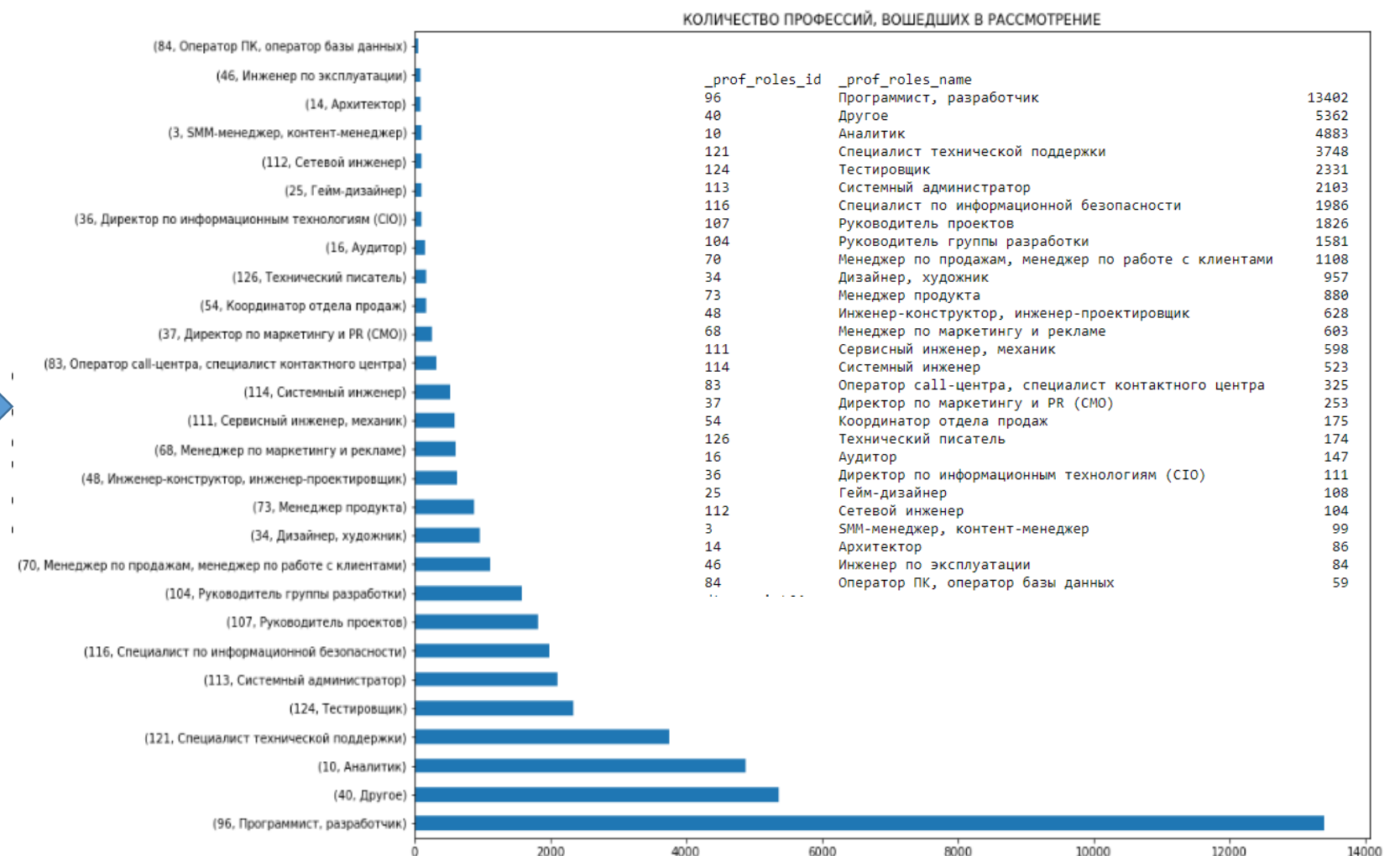
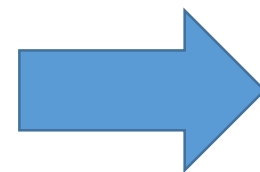
## 1. Предобработка данных.

Несмотря на то, что сбор осуществлялся по направлению «Информационные технологии, интернет, телеком», в данные попали сведения из других направлений. Для их очистки были проанализированы справочники «specializations», «professional\_roles», «employer». Были произведены обработки текстовой информации, содержащейся в описании вакансии («description») для выяснения степени принадлежности вакансии к IT-сфере (при отсутствии прямых указаний на IT-профессию или сферу).

```
# Список профессий и частота их в датафрейме df_it
df_it[['_prof_roles_id', '_prof_roles_name']].value_counts()
```

_prof_roles_id	_prof_roles_name	
96	Программист, разработчик	13402
40	Другое	5362
10	Аналитик	4883
121	Специалист технической поддержки	3748
124	Тестировщик	2331
...	...	...
86	Оператор станков с ЧПУ	1
76	Менеджер/руководитель АХО	1
41	Журналист, корреспондент	1
31	Грузчик	1
146	Юрист	1

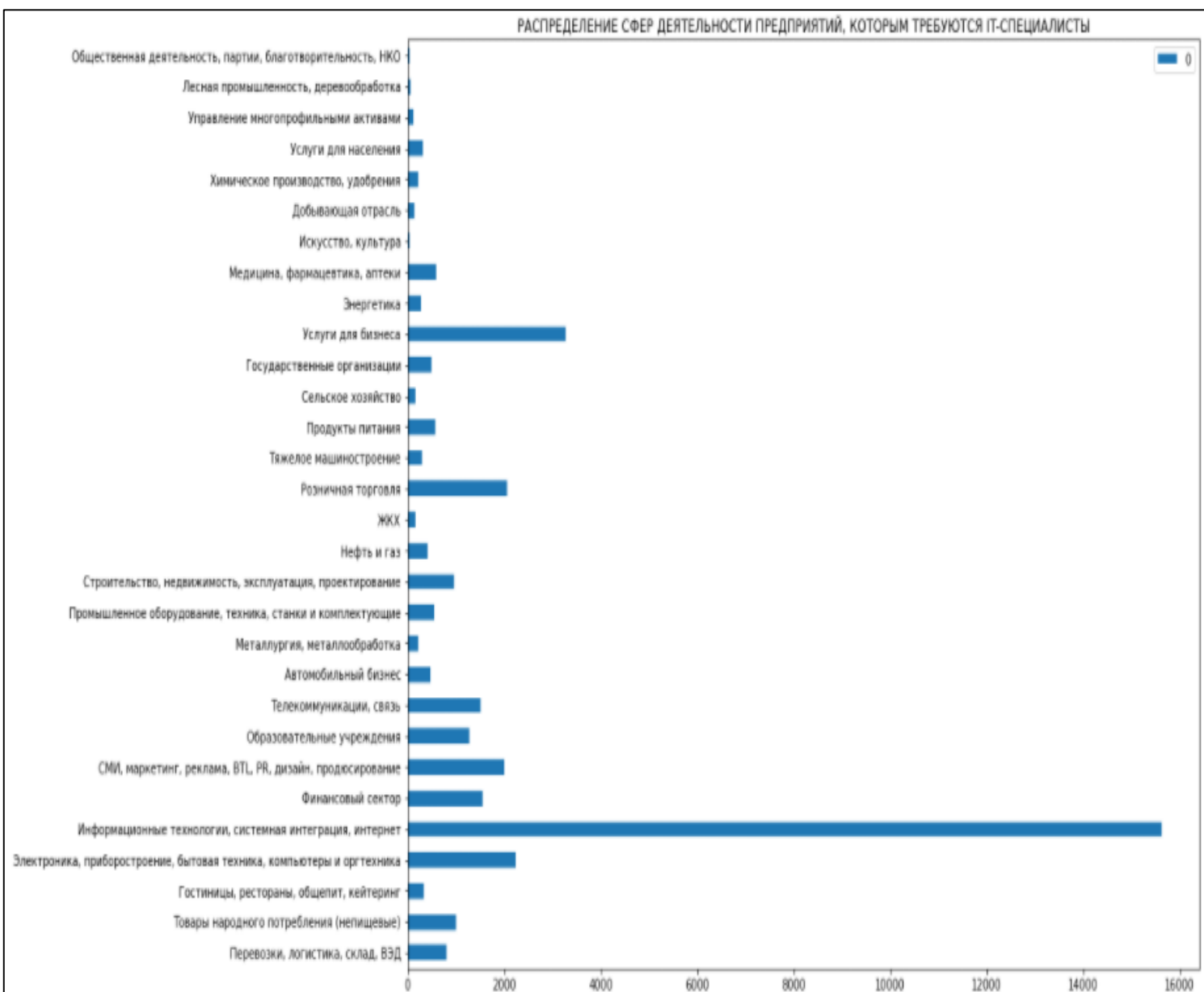
Length: 108, dtype: int64



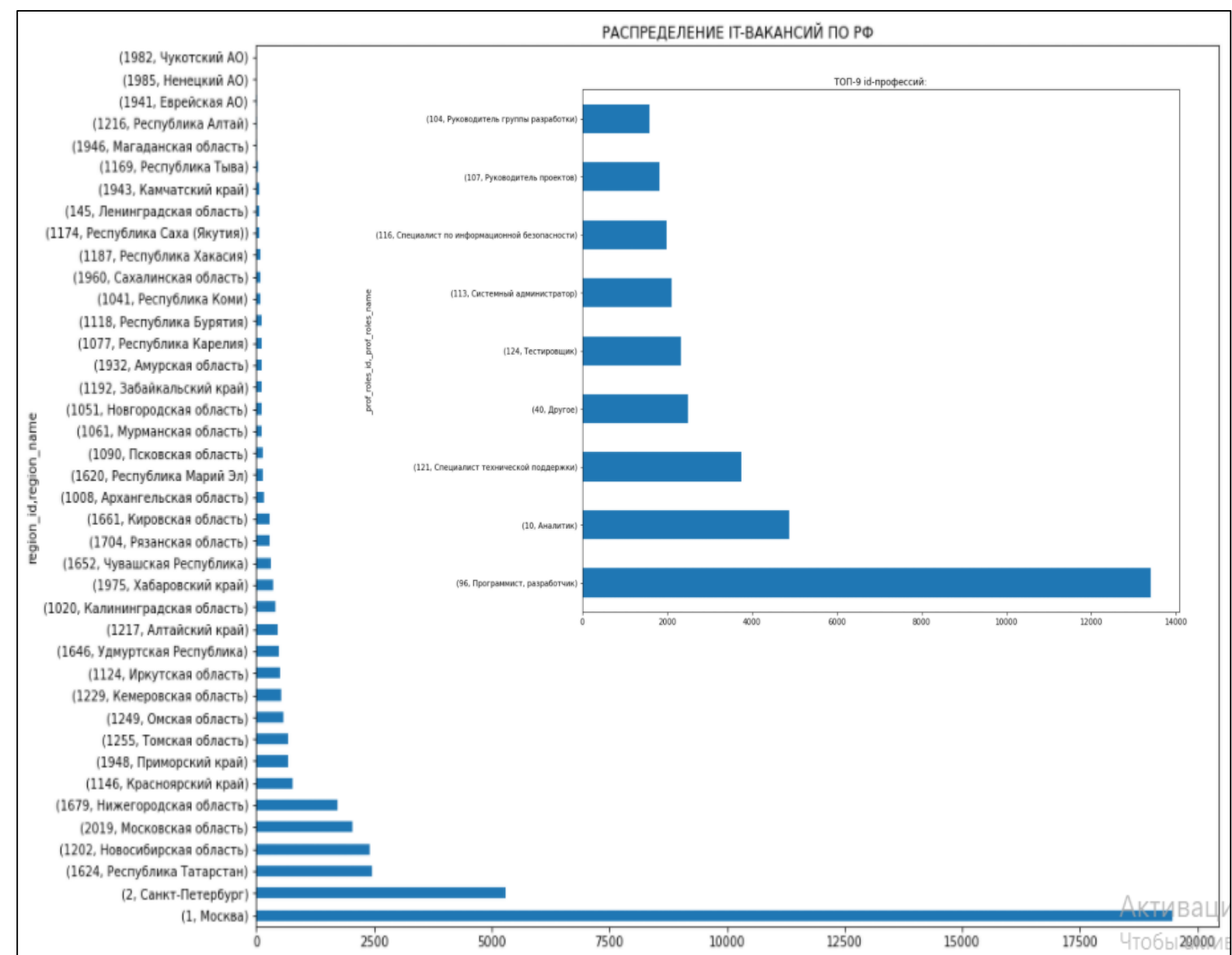
# Результаты

## 1. Построение математической модели и анализ данных.

Построим распределение вакансий по **ОТРАСЛЯМ ПРОМЫШЛЕННОСТИ**



Построим распределение вакансий по **РЕГИОНАМ**

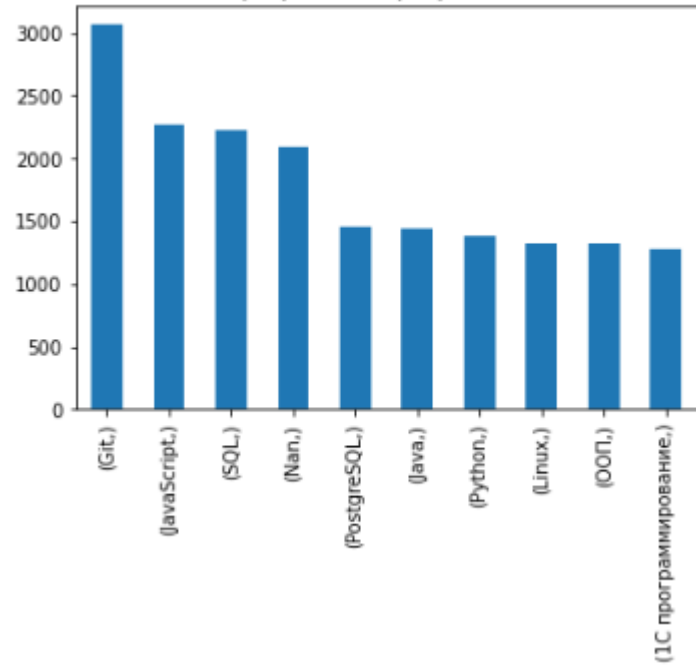


Активаци  
Чтобы 20000

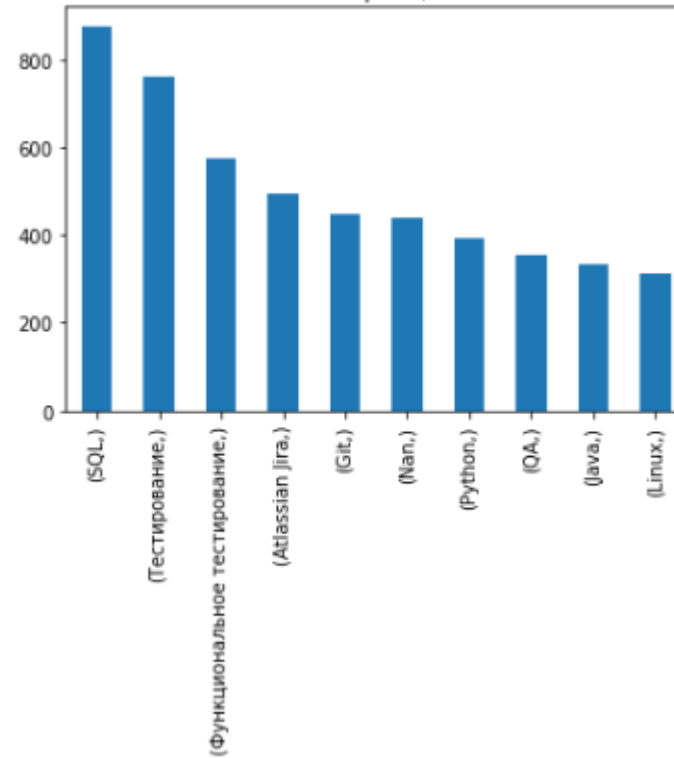
# Результаты

Для ТОП-10 id-профессии найдем самые востребованные компетенции

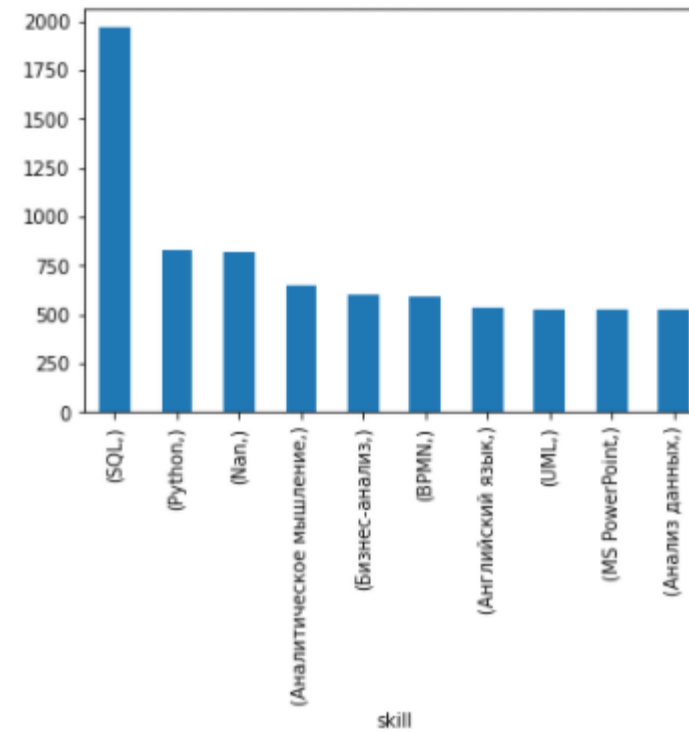
Программист, разработчик



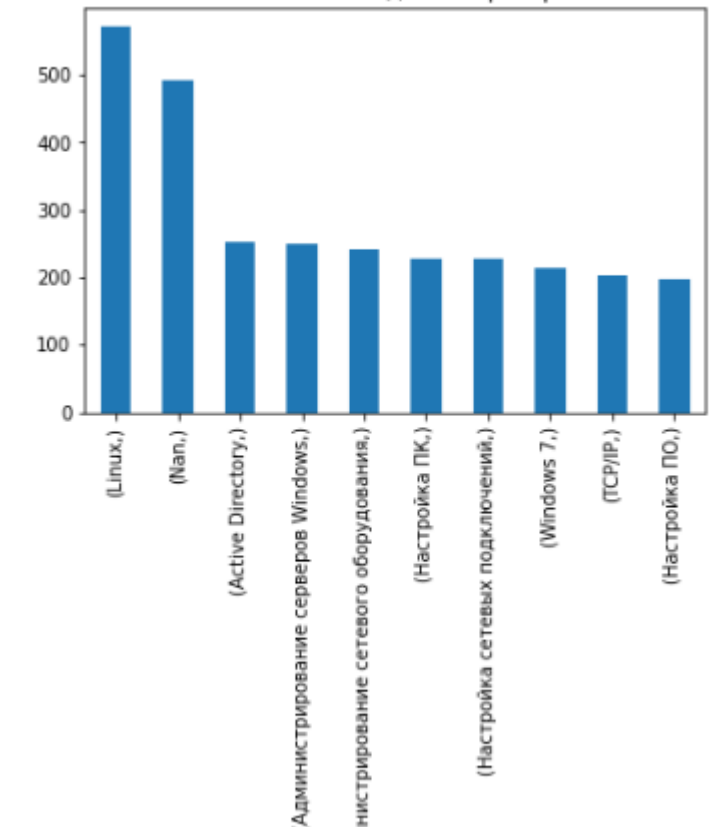
Тестировщик



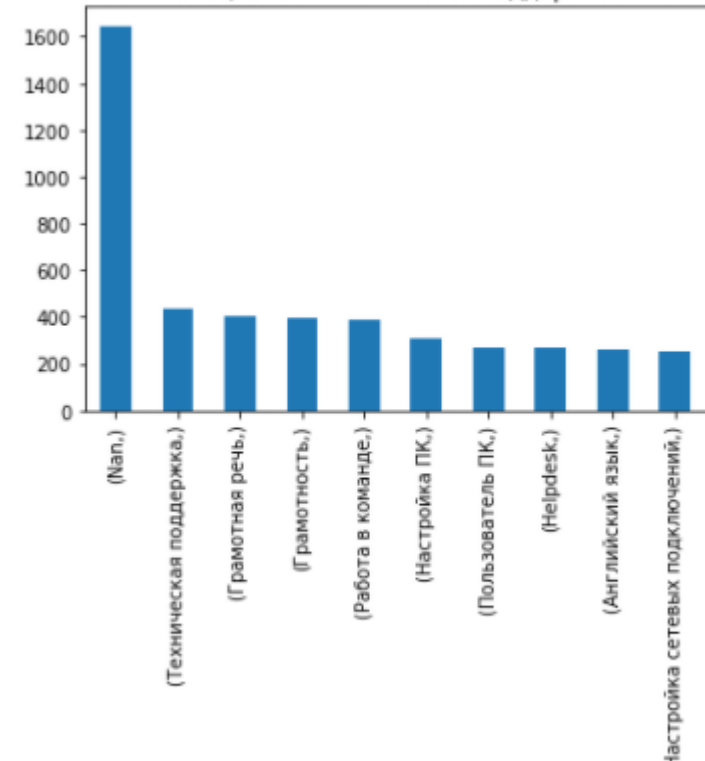
Аналитик



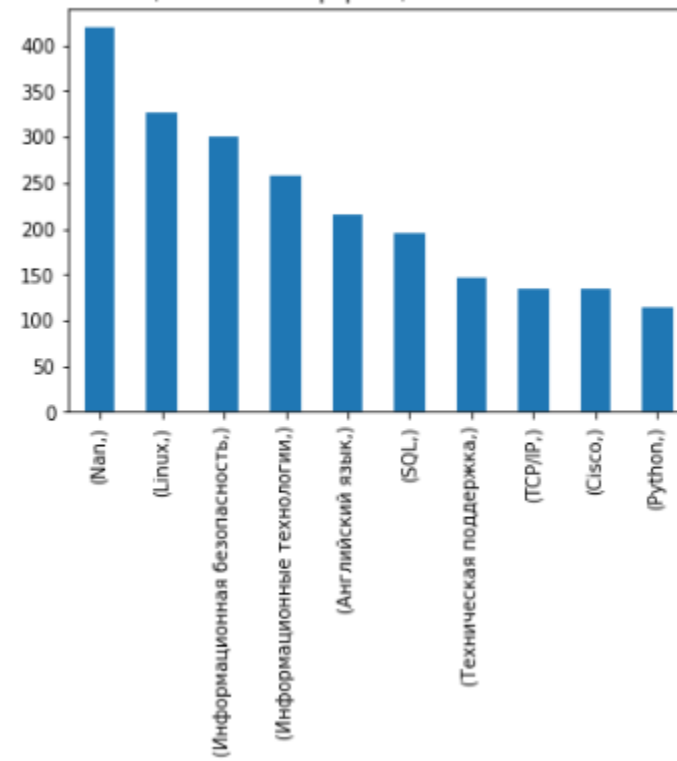
Системный администратор



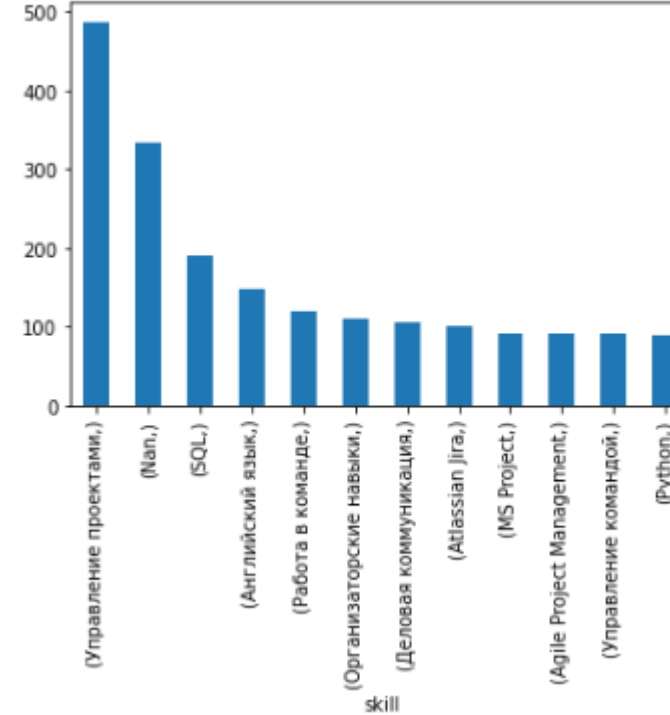
Специалист технической поддержки



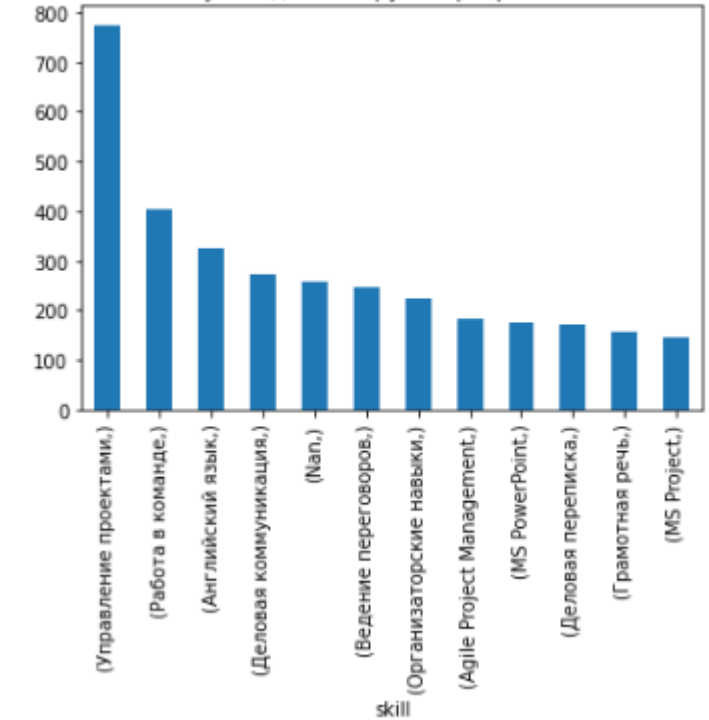
Специалист по информационной безопасности



Руководитель проектов



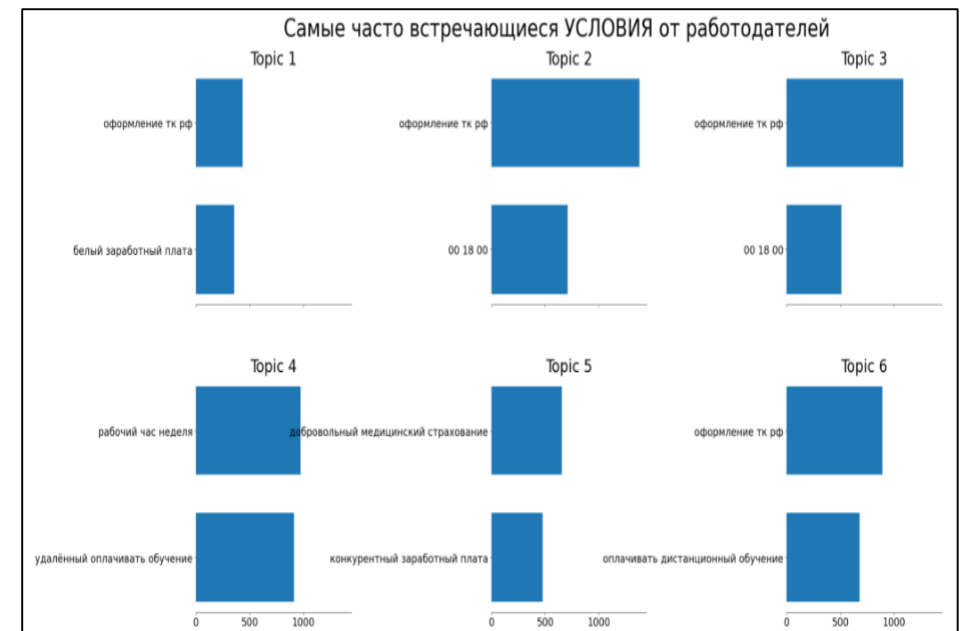
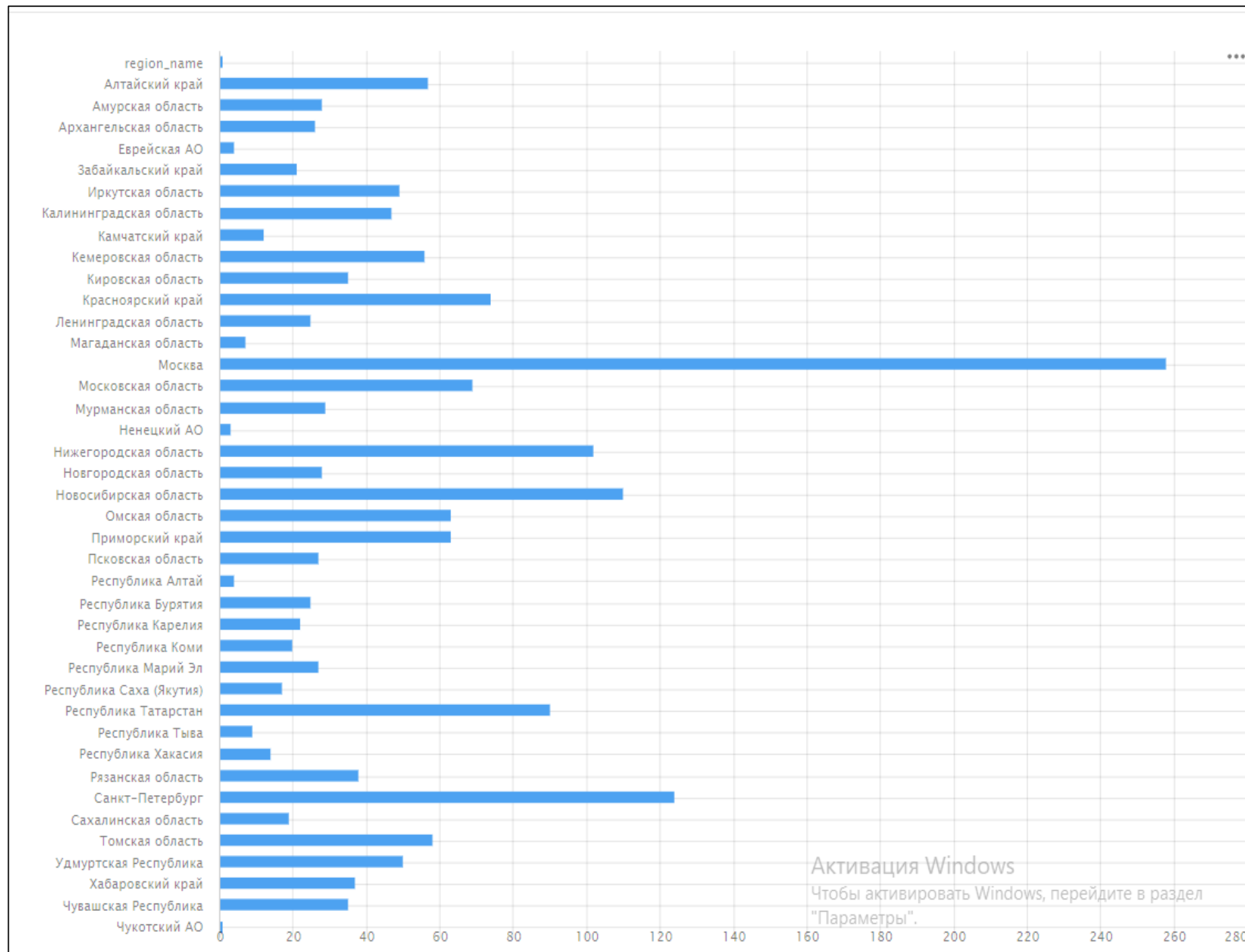
Руководитель группы разработки



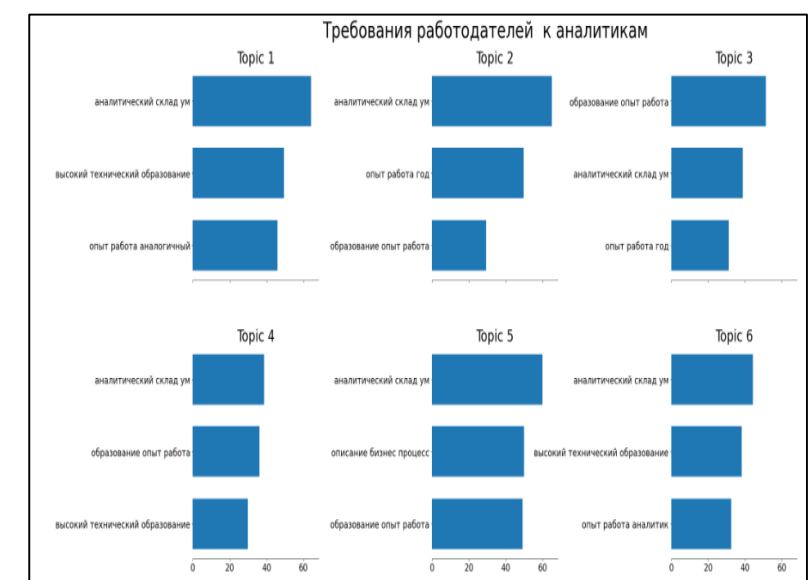


# Результаты

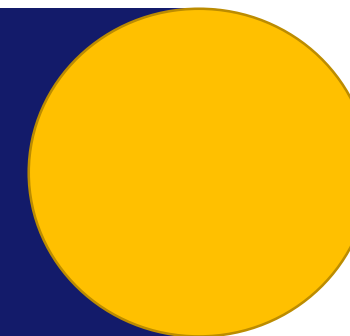
## УСЛОВИЯ, предлагаемые работодателями



## ТРЕБОВАНИЯ к АНАЛИТИКАМ



# ВЫВОДЫ



На основании анализа можно сделать вывод:

К ТОП-3 наиболее востребованным специальностям относятся:

- Программист, разработчик - 33% вакансий
- Аналитик - 12% вакансий
- Специалист технической поддержки - 10% вакансий

Основные навыки, компетенции, которыми должны обладать эти категории:

- Знание SQL – 4% всех требований
- Git – 3% всех требований
- JavaScript / Python / Английский язык – по 2% каждое

Наибольшая потребность в IT- специалистах сосредоточена в:

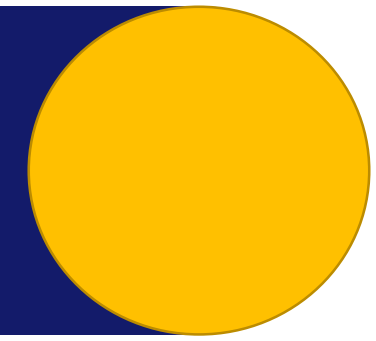
- Москве - 47%
- С-Петербурге – 13%
- Татарстане - 6%

Наибольшую потребность в IT- специалистах испытывают отрасли:

- Информационные технологии, интернет – 38%
- Услуги для бизнеса – 8%
- Розничная торговля – 5%



# Итоговые материалы



Видео-презентация

[https://drive.google.com/file/d/1jS7nfxUKsOyWPMoS6sOT\\_VB8Km\\_ybCQz/view?usp=sharing](https://drive.google.com/file/d/1jS7nfxUKsOyWPMoS6sOT_VB8Km_ybCQz/view?usp=sharing)

Презентация

Код парсера для сбора данных

[Скачать код парсера для сбора данных \(3 файла\)](#)

Ссылка на файл данных

<https://drive.google.com/file/d/1XD-81w9FPRsY8lr8bBtB8FM8EWgxIqKI/view?usp=sharing>

Ссылка на jupyter-notebook

[https://drive.google.com/file/d/1lcAWQfe\\_4azjEJRszwHjlzqVS\\_z0xtl3/view?usp=sharing](https://drive.google.com/file/d/1lcAWQfe_4azjEJRszwHjlzqVS_z0xtl3/view?usp=sharing)

