Итоговый проект по программе «Data-аналитик: старт карьеры»

Поток № 4 — сентябрь 2022 г.

Розенков Андрей Сергеевич





академия аналитики данных при Томском государственном университете



Название проекта:

Аналитика вакансий и курсов ІТ для выбора ниши образовательных курсов

Бизнес-цель заказчика:

Найти нишу для бизнеса по продажам образовательных продуктов для

ІТ-специалистов в 2022 году.

Объект исследования:

Рынок труда ІТ-отрасли;

Крупные игроки на образовательном рынке IT и их продукты.

Предмет исследования:

Вакансии на рынке IT;

Информация о самых крупных игроках рынке IT.

Цель исследования:

Оценить востребованность образовательных продуктов и ситуацию на рынке труда на основе:

- спроса по количеству вакансий;
- предложения по количеству опубликованных резюме.

Сравнить содержание предлагаемых курсов с требованиями работодателей, предъявляемыми к соискателям на позиции ІТ-специалистов разных направлений.

Требования к результату анализа:

- Наглядное описание ситуации на рынке труда и ситуации в дополнительном профессиональном образовании (онлайн-курсы);
- Представление выводов и рекомендаций: какие образовательные продукты нужны рынку, какая цена (или диапазон цен) можно поставить, чтобы не продешевить, но и люди могли покупать.

Риски:

- Не уложиться в сроки;
- Финансовые риски;
- Малое количество или плохое качество данных;
- Данные качественные, но закономерности в принципе отсутствуют;
- В данных отсутствует необходимая информация (взаимосвязи);
- Поломка программного обеспечения или компьютера;
- Форс-мажор

Источники данных, типы данных:

hh.ru — крупнейшая российская компания интернет-рекрутинга Tutortop.ru — сайт для сравнения онлайн-курсов по digital и IT.

Качественная и количественная характеристика собранных данных

| Качество данных | Соответствует/Нет |
|----------------------------------|---|
| Доступность | + |
| Точность | + |
| Взаимосвязанность | + |
| Полнота | + |
| Непротиворечивость | + |
| Однозначность | + |
| Релевантность | + |
| Надежность | + |
| Своевременность | +/- |
| Данные, предоставленные заказчин | ком дополнительно обрабатывались и очищались. |

Способ(ы) получения данных:

Данные были предоставлены заказчиком и получины в результате парсинга сайтов hh.ru и Tutortop.ru

Потребности клиентов:

Что влияет на выбор курса у покупателей и как мы можем помочь нашим клиентам планируемыми курсами?

1)Перспектива – 3П после обучения и в дальнейшем;

(посчитать среднюю зп по вакансиям, зависимость зп от региона)

2)Лёгкость получения первой работы (конкуренция на рынке, помощь с трудоустойством);

(Не можем оценить, нет данных)

3)Сроки обучения и необходимое количество времени на обучение; (Можем оценить по датасету из курсов)

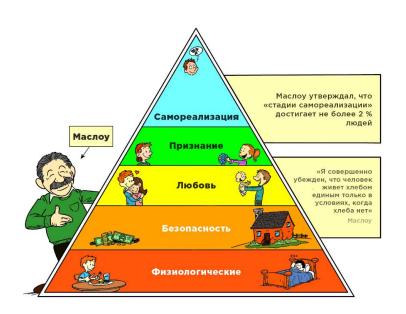
4) Лёгкость обучения;

5)Стоимость курса (+маркетинг, скидки); (Можем оценить)

6)Качество обучения;

7) Можно ли получить обучение по квоте от Государства или нет;

8)Общий тренд (комьюнити, нравится или нет обучение).



Этапы исследования:

На этапе планирования исследования:

Были проработано техническое задание заказчика, заполнена форма с основными параметрами исследования, такими как выбор методологии, постановка бизнес цели, определение предмета и объекта исследования, цели и инструментов, оценка рисков и другими;

На этапе обработки данных:

данные, подвергались очистке и предобработке. В данных были удалены лишние столбцы и строки, оценены качественные и количественные характеристики, проведён разведочный анализ, визуализация с помощью столбчатых диаграмм самых востребованные профессий, навыков, топ-работодателей и приведены заработные платы к одной валюте.

На этапе статистического исследования данных были составлены:

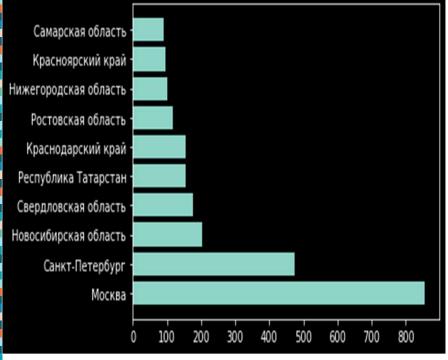
- -зависимости зп от формата работы, опыта работы, оценены различия в данных группах и доказана их статистическая значимость.
- -произведена фильтрация по навыкам и выбраны самые высокооплачиваемые и востребованные навыки. По этим навыкам выбраны курсы конкурентов.
- -доказана зависимости стоимости курсов от их длительности и построена соответствующая модель.
- -проведена оценка стоимости курса от его длительности.

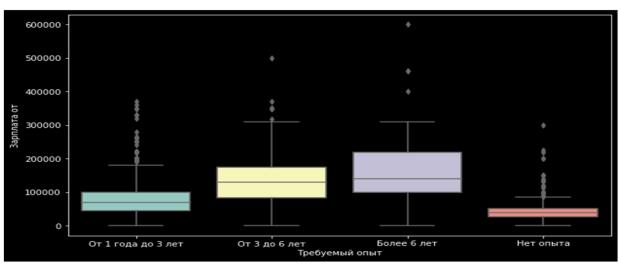
На этапе **визуализации данных** был построен дашборд с основными параметрами и метриками для анализа ранка труда для IT специалистов.

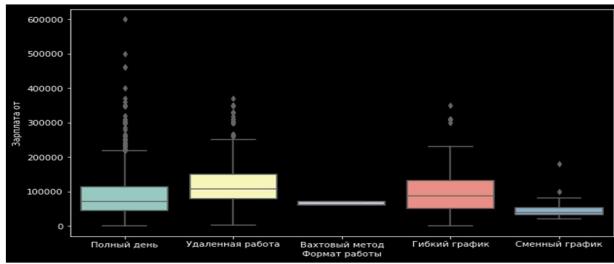
На этапе **статистического исследования** посмотрел распределение вакансий по регионам, как зависит зп от стажа работы и формата. Проверил принадлежность распределение зп в группах по тесту Шапиро-Уилка и по критерию Спирмена доказал значимость различия зп в

группах.

| Формат работы | Единичная доля вакансий |
|---------------------------|-------------------------|
| Нет опыта работы | 0.112723 |
| Опыт работы от 1 до 3 лет | 0.549873 |
| Опыт работы от 3 до 6 лет | 0.314504 |
| Опыт работы более 6 лет | 0.022901 |



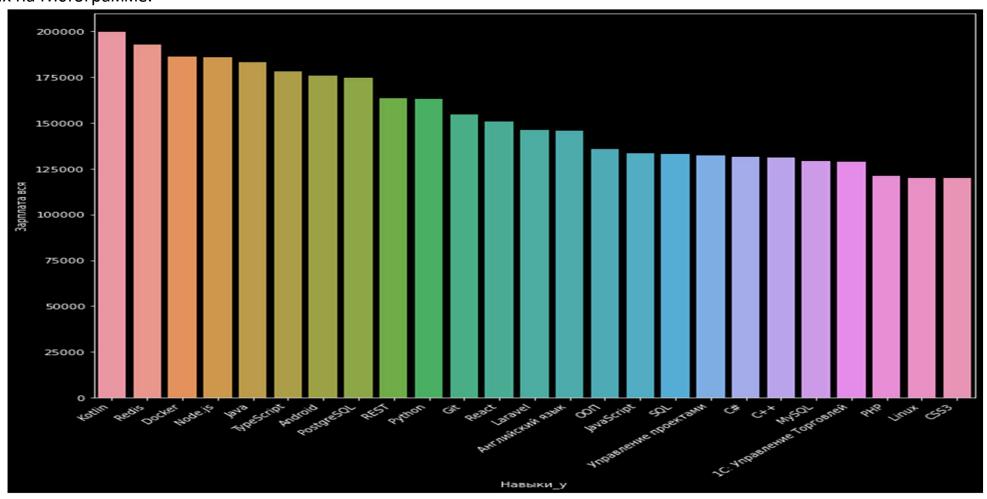




Проверил корреляцию опыта работы, заработной платы и навыков на основании критерия Спирмена для не нормального распределения и проверил статистическую значимость данных корреляций. Корреляция между опытом работы и заработной платой высокая, между зп и навыками – низкая. Данные корреляции визуализировал с помощью тепловой карты.



Далее выбрал топ-50 самых популярных навыков и среди них 25 самых высокооплачиваемых. Визуализировал их на гистограмме.



Далее навыки сравнивались с образовательными курсами топовых компаний по данным направлениям и были выбраны 5 основных ниш для заказчика. Также был добавлен курс по трудоустройству, как очень важный для любого направления в программировании и учтены требования заказчика, что курсы должны быть короткие и быстро собираться.

| Наименование навыка | Средняя <u>зп</u> | Количество вакансий | Срок обучения, месяцев (по конкурентам) |
|---|----------------------|------------------------|---|
| Курс по английскому для IT специалистов | (E) | | 6 |
| Linux | 106922 | 725 | 5 |
| Bash | (2) | 98 | 1 |
| Администрирование сетевого оборудования | 66929 | 110 | 5 |
| Курсы по трудоустройству + пробные собеседования для самых востребованных специальностей. Курс по трудоустройству за рубежом. | .=0 | - | 0,5-1 |

Результаты анализа

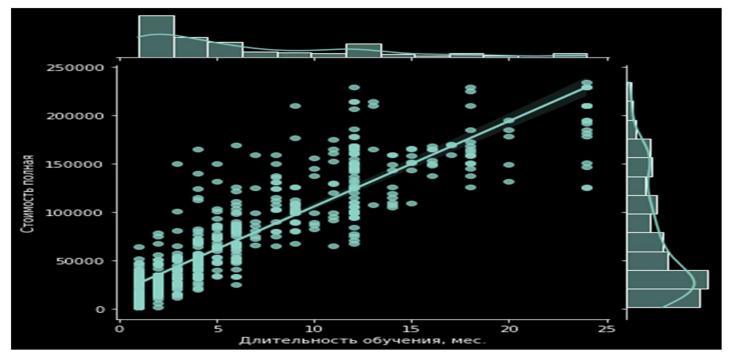
Гипотеза исследования: Чем больше срок обучения на курсе, тем выше у него стоимость.

Метод(ы) проверки гипотезы: Количественная корреляция Пирсона.

Результаты проверки гипотезы (графическое представление):

Гипотеза подтвердилась и чем дольше обучение, тем дороже оно стоит, если сравнивать независимые группы одного сегмента. На основании гипотезы была построена модель, по которой:

Стоимость обучения = 17560 + (8815,7152*длительность обучения мес.)



Полученные результаты и рекомендации для заказчика:

| Название курса | Школа | Длительность, | Стоимость в | Стоимость по |
|----------------------|-------------------------|---------------|---|--------------------------------|
| | | Mec. | школе | модели |
| Administrator Linux. | at charge regulation of | 100,400 | 0.075 077 300 000 AV | 61638,58 |
| Professional | Otus | 5 | 101000 | es. |
| Администрирование | 00000000 | | | 61638,58 |
| OC Linux | Skillbox | 5 | 73255 | 66 |
| Работа в командной | | | | 26375,72 |
| строке Bash | Skillbox | 1 | 11495 | 000 1000 0 000 000 000 000 000 |
| Linux. Рабочая | Contraction along | | | |
| станция Практика | | | | |
| администрирования | | | | |
| OC Linux на | | | | 26375,72 |
| компьютере | GeekBrains | 1 | 15000 | 50 |
| Английский для | Яндекс | 6 | 79000 | 70454,29 |
| работы в | Практикум | 10000 | 95.00.000000000000000000000000000000000 | |
| международной IT | 900 MM | | | |
| команде | | | | TO. |

Полученные результаты и рекомендации для заказчика:

Найдены ниши для бизнеса по продажам образовательных продуктов для ІТ-специалистов в 2022 году.

Наиболее привлекательные по востребованности и оплате труда являются профессии в сфере:

- -Администрирование сетевого оборудования,
- -Курс по английскому для ІТ специалистов,
- -Linux,
- -Bash,
- -Курсы по трудоустройству + пробные собеседования для самых востребованных специальностей. Курс по трудоустройству за рубежом.

На основе проведённого исследования заказчику рекомендую открыть курсы в сфере Администрирование ОС Linux, потому что:

- 1) Очень популярная и оплачиваемая сфера для специалистов;
- 2) Создание курсом можно организовать на разный тип клиентов по платёжеспособности: для начинающих, продолжающих и профессионалов;
- 3) В соответствии с требованиями заказчика курсы можно собирать быстро со сроком обучения от 1 до 5 месяцев;
- 4) Минимальная конкуренция со стороны крупных образовательных школ в этой сфере.

Цена на курсы будет во многом зависеть от длительности и по оценке получается:

| 1 | 26375,72 |
|---|----------|
| 2 | 35191,43 |
| 3 | 44007,15 |
| 4 | 52822,86 |
| 5 | 61638,58 |
| 6 | 70454,29 |
| | |

Также нужно учитывать другие параметры, например, какие специалисты будут вести курс и их заработная плата, траты на маркетинг, цены на курсы у конкурентов и многие другие факторы, которые не учитывались в данном исследовании.

• Перспективные направления для дальнейшего анализа:

- Оценить позитивные и негативные отзывы по курсам.
- Сделать соцопросы в чатах IT на счёт того, насколько сейчас востребованы выбранные курсы.

Дашборд

На дашборде представленны фильтрация по уровню зарплат, срезы по формату работы и требуемому опыту работы, накопительные диаграммы по наиболее востребованным навыкам с разбивкой по опыту работы, диаграммы с количеством вакансий и топ-работодателям по убыванию, карта и кнопка возврата к первоначальным значениям, меры с количеством вакансий и средним значением заработной платы.



Спасибо за внимание!

Розенков Андрей Сергеевич and-rozenkov@yandex.ru