

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

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



Описание проекта

Название проекта:

**“Исследование рынка труда и дополнительных
образовательных услуг РФ”**

Описание проекта

Бизнес-цель заказчика:

- ✓ Определиться со стратегией подготовки образовательных программ.
- ✓ Выявление целевой аудитории Работодатели /Работники

Объекты исследования:

- ✓ Рынок труда РФ
- ✓ Рынок предоставления образовательных услуг

Описание проекта

Предмет исследования:

- ✓ Спрос/предложение на рынке труда
- ✓ Требования к компетенциям и навыкам специалистов

Цель исследования:

- ✓ Выявление потребностей и тенденций рынка труда
- ✓ Выявление возможностей имеющихся образовательных курсов
- ✓ Определение диапазонов цен существующих образовательных курсов



Описание проекта

Условия реализации проекта:

Ресурсы. Данные

- ✓ Базы вакансий
- ✓ Базы предлагаемых курсов

Описание проекта

Риски реализации проекта:

- ✓ Данные не подходят для задачи (плохо понято ТЗ, выбран неверный метод сбора и анализа данных)
- ✓ Недостаточное количество данных / Плохое качество данных (недостаточно квалификации для сбора данных)
- ✓ В данных отсутствует необходимая информация (взаимосвязь)
 - неверно выбраны источники данных
- ✓ Поломка техники/недоступность программного обеспечения
- ✓ Недостаточная квалификация исполнителя



Описание проекта

Требования к результату анализа:

- ✓ Интерактивные дашборды
- ✓ Предоставление выводов и рекомендаций

Описание проекта

Источники данных, типы данных:

✓ Источники:

- Базы вакансий и резюме России
- Агрегаторы курсов дополнительного образования

✓ Типы данных:

- для вакансий собрана полная информация о имеющихся вакансиях с сайта headhunter.ru по состоянию на 26.05.2022. Данные представлены в формате таблицы , содержащей как структурированные JSON описания, так и текстовые данные.
- для курсов собраны данные о направлении обучения (ИТ-тематика), образовательном учреждении, стоимости обучения. Данные несут недостаточно качественной информации для анализа нарабатываемых в процессе обучения навыков.



Описание проекта

Способ(ы) получения данных:

Используем только открытые данные в соответствии с законодательством Российской Федерации.

Данные о вакансиях с сайта headhunter.ru и агрегатора coursotur.online предоставлены заказчиком.

Для сравнения также используются данные с сайта-агрегатора tutortop.ru. Данные собраны с помощью парсинга сайта tutortop.ru сторонним исполнителем (вторичные источники).



Описание проекта

Этапы исследования

Планирование дизайна исследования:

Проведен анализ технического задания

- подготовлен план исследования,
- определены объекты и предмет исследования,
- определены цели анализа данных,
- оценены риски и условия реализации проекта,
- выбраны методы и инструменты анализа,
- сформулированы требования к результату исследования.



Описание проекта

Этапы исследования

Обработка данных:

В среде Google Colab с использованием языка программирования Python проведен разведочный анализ данных, предоставленных заказчиком.

Проведена обработка и очистка датасета - оценены и удалены пустоты, проведена проверка на наличие дубликатов.

Описание проекта

Этапы исследования

Статистическое исследование данных:

В среде Google Colab с помощью языка программирования Python выполнены следующие исследования:

- Расчитаны основные статистические показатели для количественных данных (минимальные и максимальные зарплаты)
- Проведен анализ данных на соответствие нормальному закону распределения с помощью критериев асимметрии и эксцесса, а так же теста д'Агостино
- С помощью критерия сравнения групп Манна-Уитни проведен анализ связи количественного показателя (заработная плата) с категориальным показателем (программисты/иные профессии)



Описание проекта

Этапы исследования

Интерпретация данных:

Проведен дополнительный визуальный анализ наборов данных, размещенных на интерактивных дашбордах. Проведенна аналогия на соответствие полученных результатов поставленным задачам.

Сформулированы выводы и рекомендации.

Описание проекта

Этапы исследования

Оформление результатов анализа:

С помощью среды Power BI подготовлены интерактивные дашборды для анализа информации о вакансиях, полученным с сайта hh.ru, а так же для анализа информации об обучающих курсах от фирм-агрегаторов Coursetor и TutorTop

В программном обеспечении WPS Office подготовлена презентация с результатами исследования.

С помощью программного обеспечения Zoom записана видеопрезентация выполненного проекта.

Результаты анализа

Гипотеза исследования:

“На рынке труда наблюдается нехватка специалистов отрасли Информационные технологии”

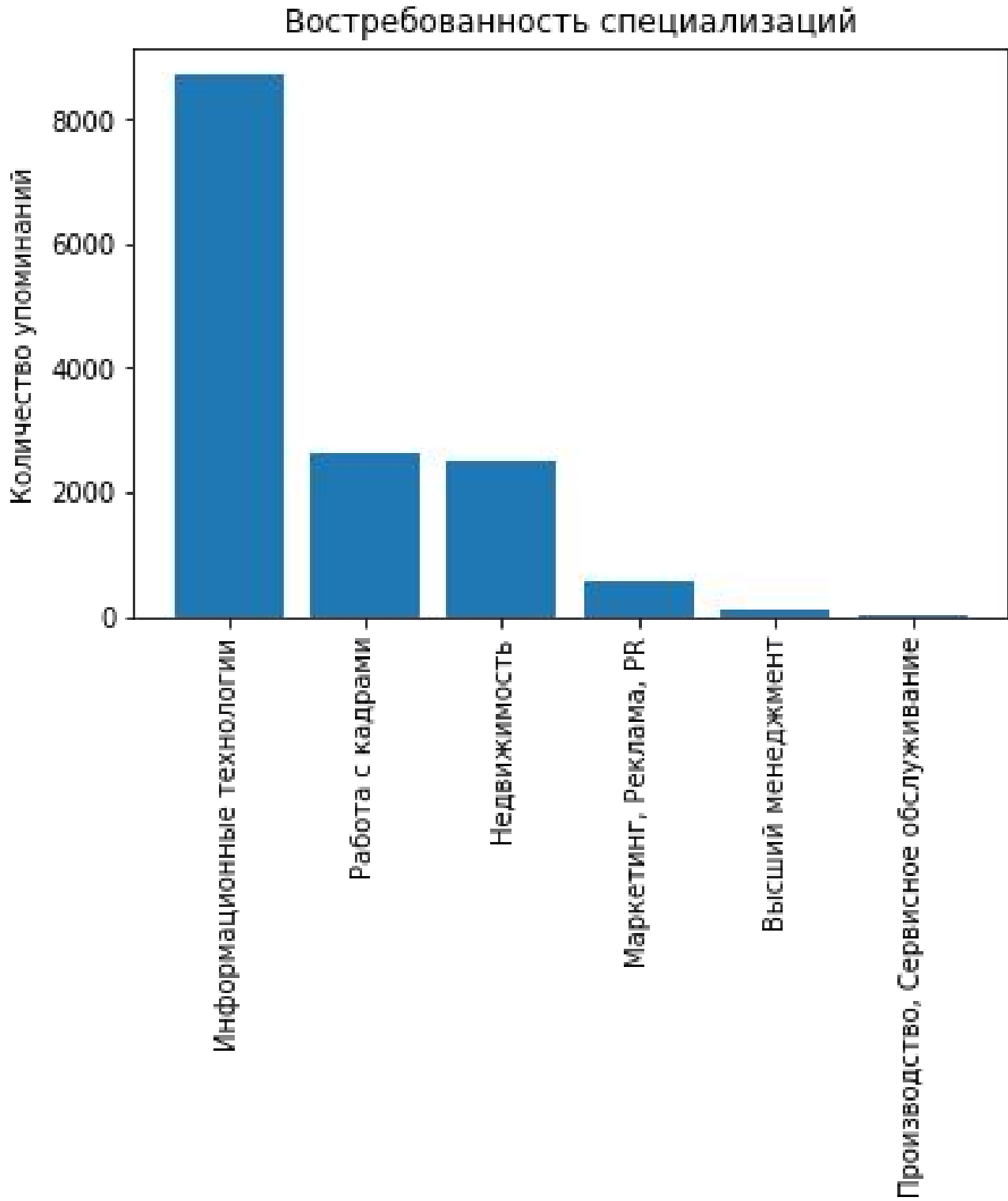
Методы проверки гипотезы:

- Оценка статистических показателей имеющихся в наборе данных вакансий на рынке труда hh.ru для различных отраслей.
- Оценка KPI по интерактивному дашборду

Результаты анализа

Результаты проверки гипотезы

| | count | mean | std |
|--------------------------------------|--------|---------------|--------------|
| unif_name | | | |
| Высший менеджмент | 109.0 | 113742.477064 | 67992.614187 |
| Информационные технологии | 8713.0 | 87883.622059 | 64092.805029 |
| Маркетинг, Реклама, PR | 561.0 | 45481.140820 | 28213.421784 |
| Недвижимость | 2525.0 | 95324.411089 | 61746.606313 |
| Производство, Сервисное обслуживание | 5.0 | 48276.000000 | 14425.214089 |
| Работа с кадрами | 2621.0 | 49832.465090 | 24641.681475 |



Результаты анализа

Результаты проверки гипотезы



Результаты анализа

Результаты проверки гипотезы



Результаты анализа

Описание результатов

- По результатам исследования, полученным с помощью разведочного и статистического анализа данных, для чуть более 15 тысяч вакансий с указанной заработной платой почти половина набора данных приходится на отрасль «Информационные технологии»
- Дашборд подготовлен для всего набора данных, так же проведена очистка от данных, не содержащих сведений о заработной плате. Доля вакансий в ИТ отрасли : 7937 из 13920
- Гипотеза о том, что ИТ-специалистов не хватает - **ПОДТВЕРЖДАЕТСЯ.**

Результаты анализа

Гипотеза исследования:

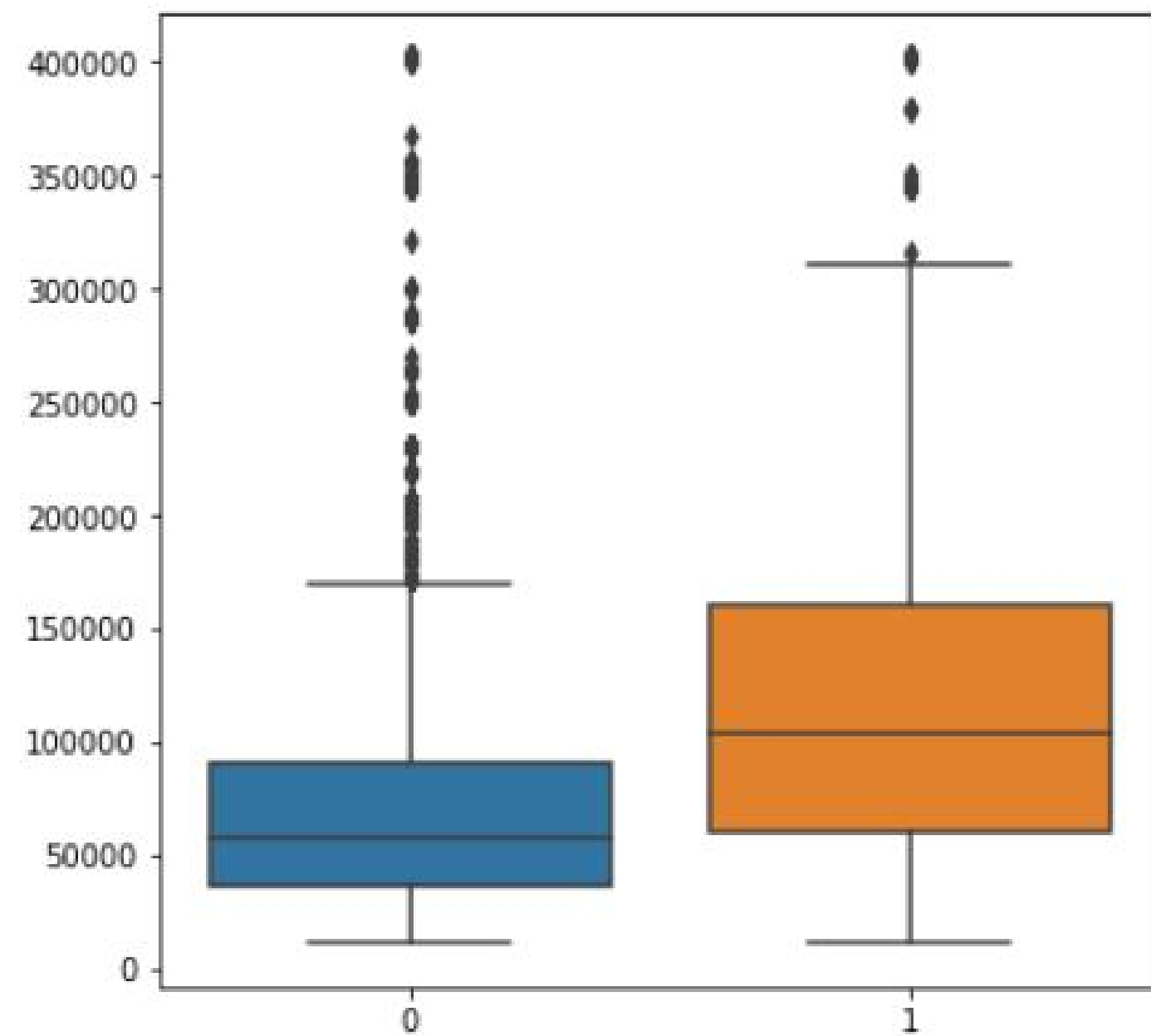
“Вакансии, связанные с программированием, оплачиваются выше остальных в ИТ-отрасли ”

Методы проверки гипотезы:

- Анализ влияния категориальных признаков по критерию Манна-Уитни.
- Оценка статистических показателей имеющихся в наборе данных вакансий на рынке труда hh.ru для различных отраслей.
- Оценка KPI по интерактивному дашборду

Результаты анализа

Результаты проверки гипотезы



```
df_vacancy_IT_0 = df_vacancy_IT[df_vacancy_IT['new_name']==0]
df_vacancy_IT_1 = df_vacancy_IT[df_vacancy_IT['new_name']==1]
scipy.stats.mannwhitneyu(df_vacancy_IT_0['min_salary'],df_vacancy_IT_1['min_salary'])
```

MannwhitneyuResult(statistic=4781885.0, pvalue=6.200976703727796e-250)

Результаты анализа

Результаты проверки гипотезы



Результаты анализа

Описание результатов

Визуально минимальная заработная плата в отрасли Информационные технологии для категории 1 (программисты) выше, чем для остальных должностей.

Для подтверждения проведен статистический анализ влияния категориальных признаков по критерию Манна-Уитни. В результате которого нулевая гипотеза H_0 об отсутствии различий в оплате ОТВЕРГАЕТСЯ. (p-value много меньше 0,05).

А, следовательно, можно утверждать, что должность программиста является более высокооплачиваемой в области ИТ.

Результаты анализа

Гипотеза исследования:

“На рынке ИТ-курсов курсы по программированию занимают равную долю с остальными направлениями”

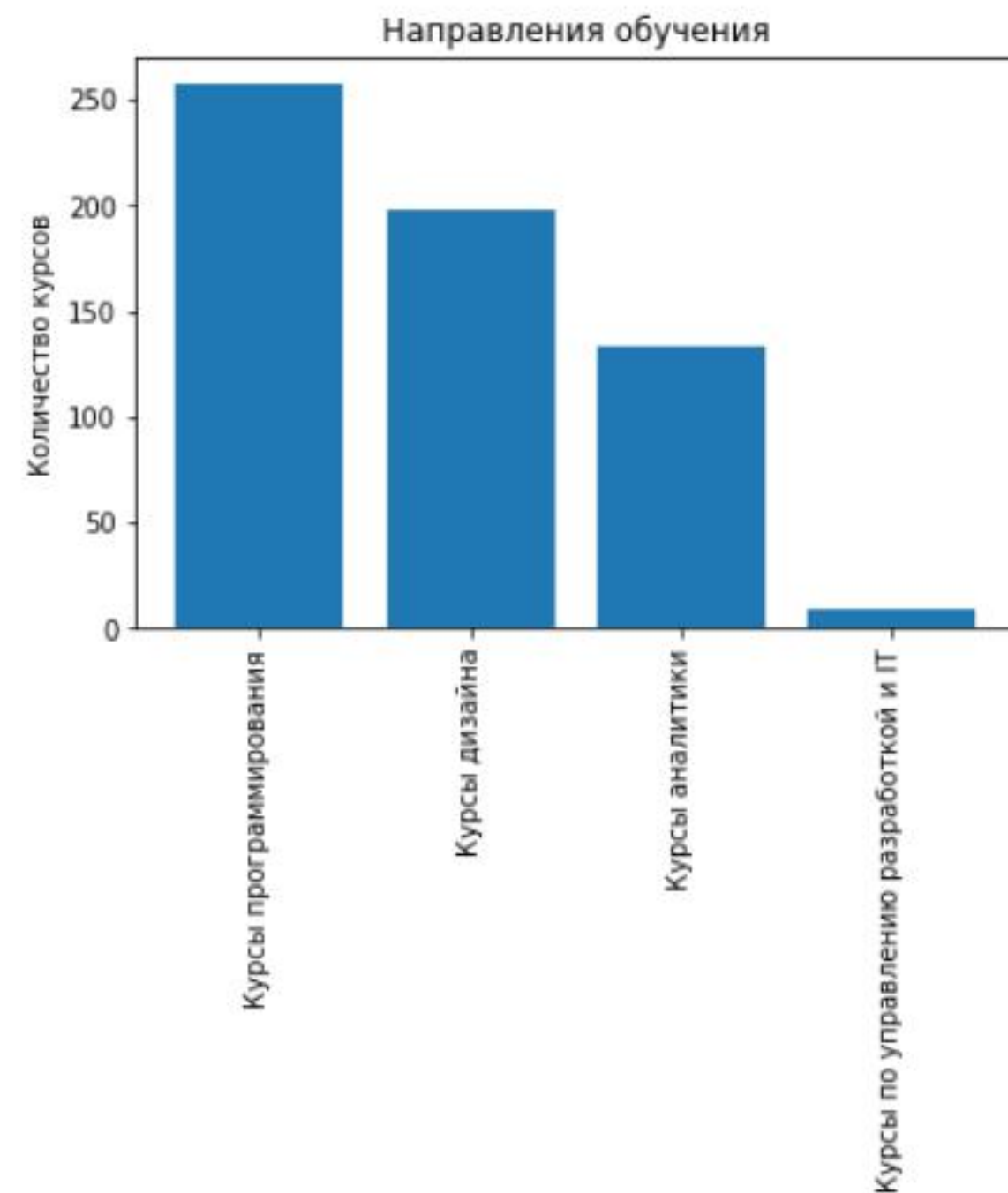
Методы проверки гипотезы:

- Оценка статистических показателей имеющихся в наборе данных агрегаторов курсов Coursotor и TutorTop.
- Оценка KPI по интерактивному дашборду

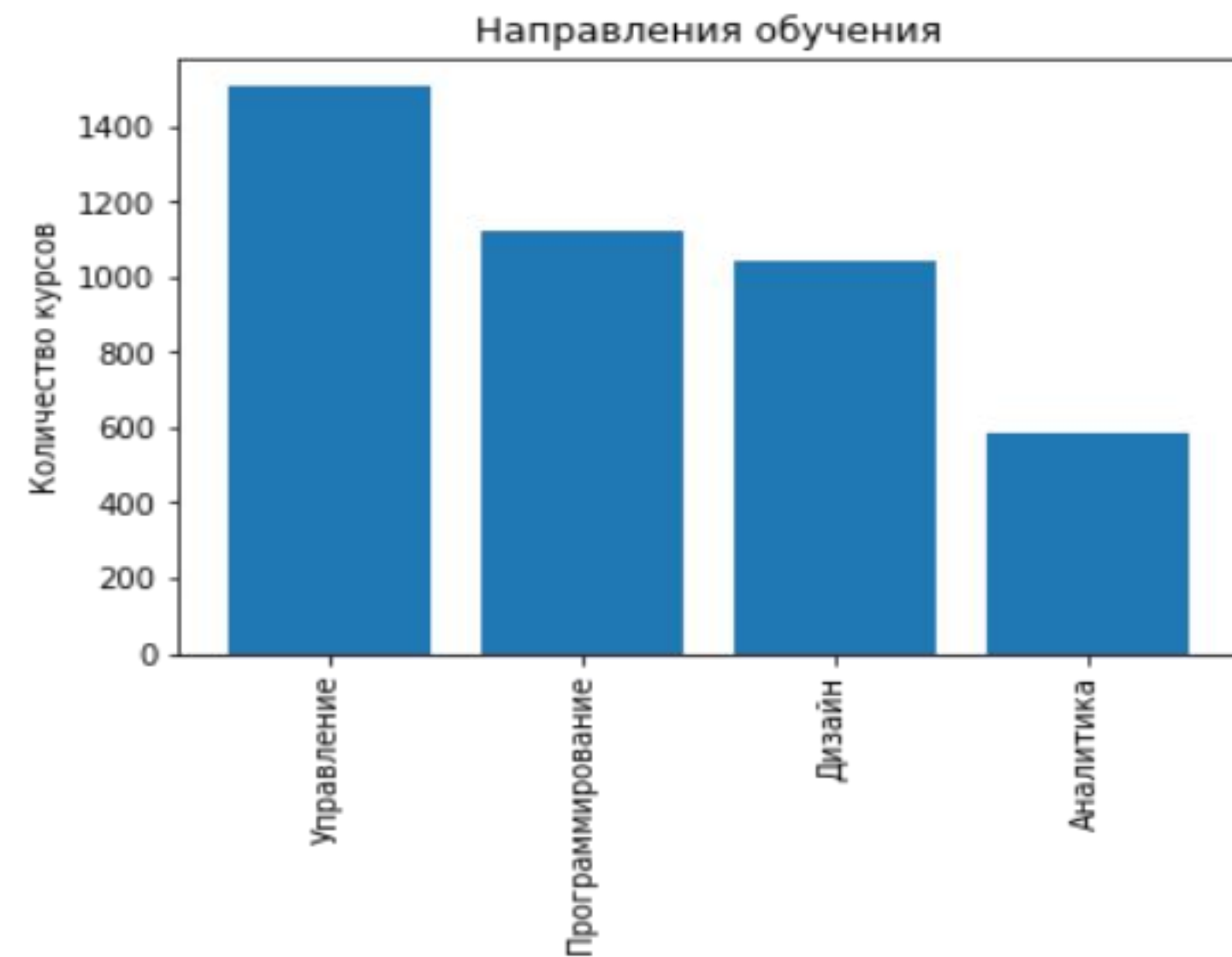
Результаты анализа

Результаты проверки гипотезы

агрегатор Coursotor



агрегатор TutorTop



Результаты анализа

Результаты проверки гипотезы

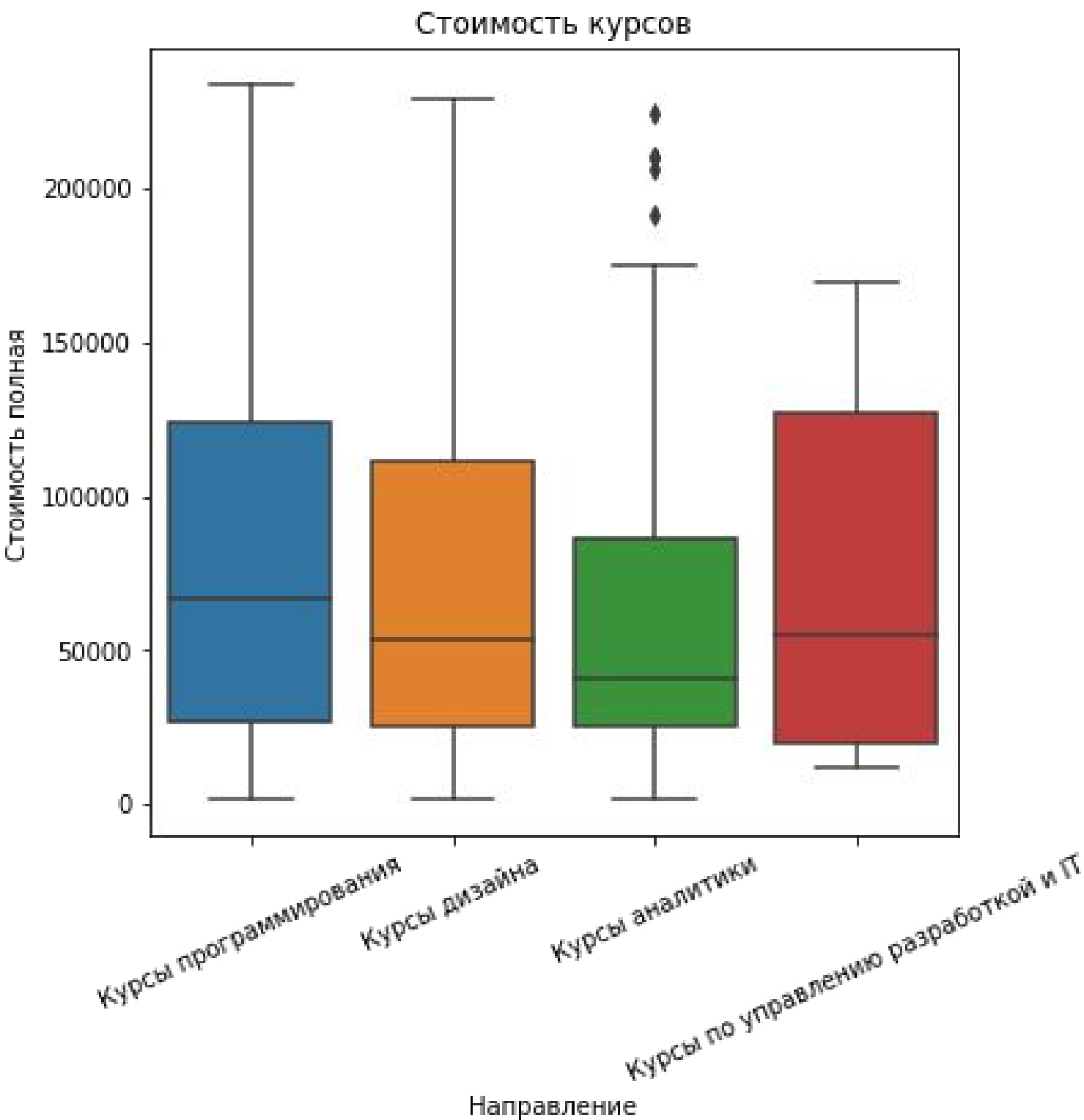
Медианные значения Стоимости курсов

| | |
|--|---------|
| Направление | |
| Курсы аналитики | 41300.0 |
| Курсы дизайна | 53201.0 |
| Курсы по управлению разработкой и IT | 54700.0 |
| Курсы программирования | 66500.0 |
| Name: Стоимость полная, dtype: float64 | |

Средние значения Стоимости курсов

| | |
|--|--------------|
| Направление | |
| Курсы аналитики | 63400.541353 |
| Курсы дизайна | 73890.015152 |
| Курсы по управлению разработкой и IT | 77699.111111 |
| Курсы программирования | 76907.062257 |
| Name: Стоимость полная, dtype: float64 | |

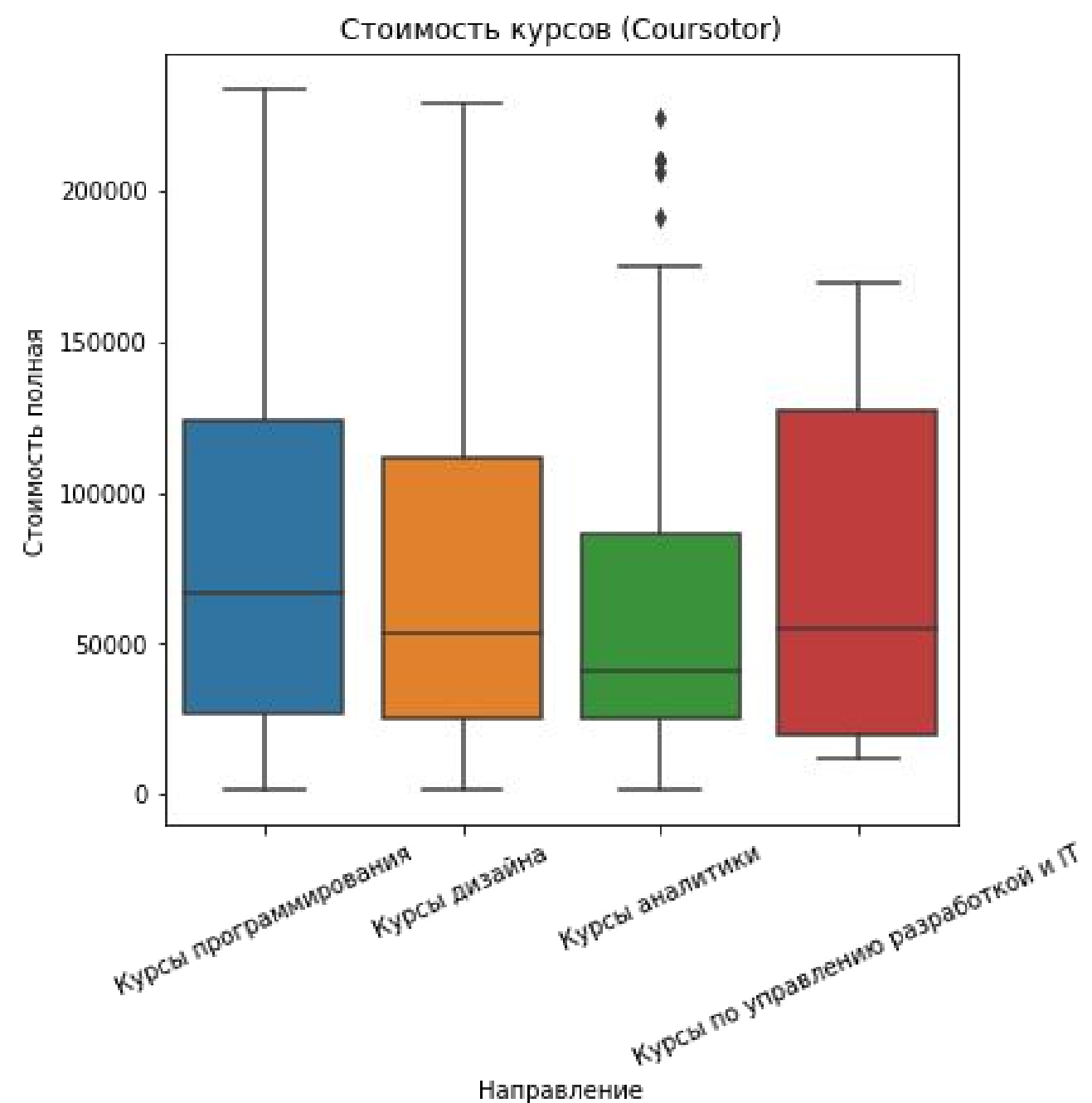
агрегатор Coursotor



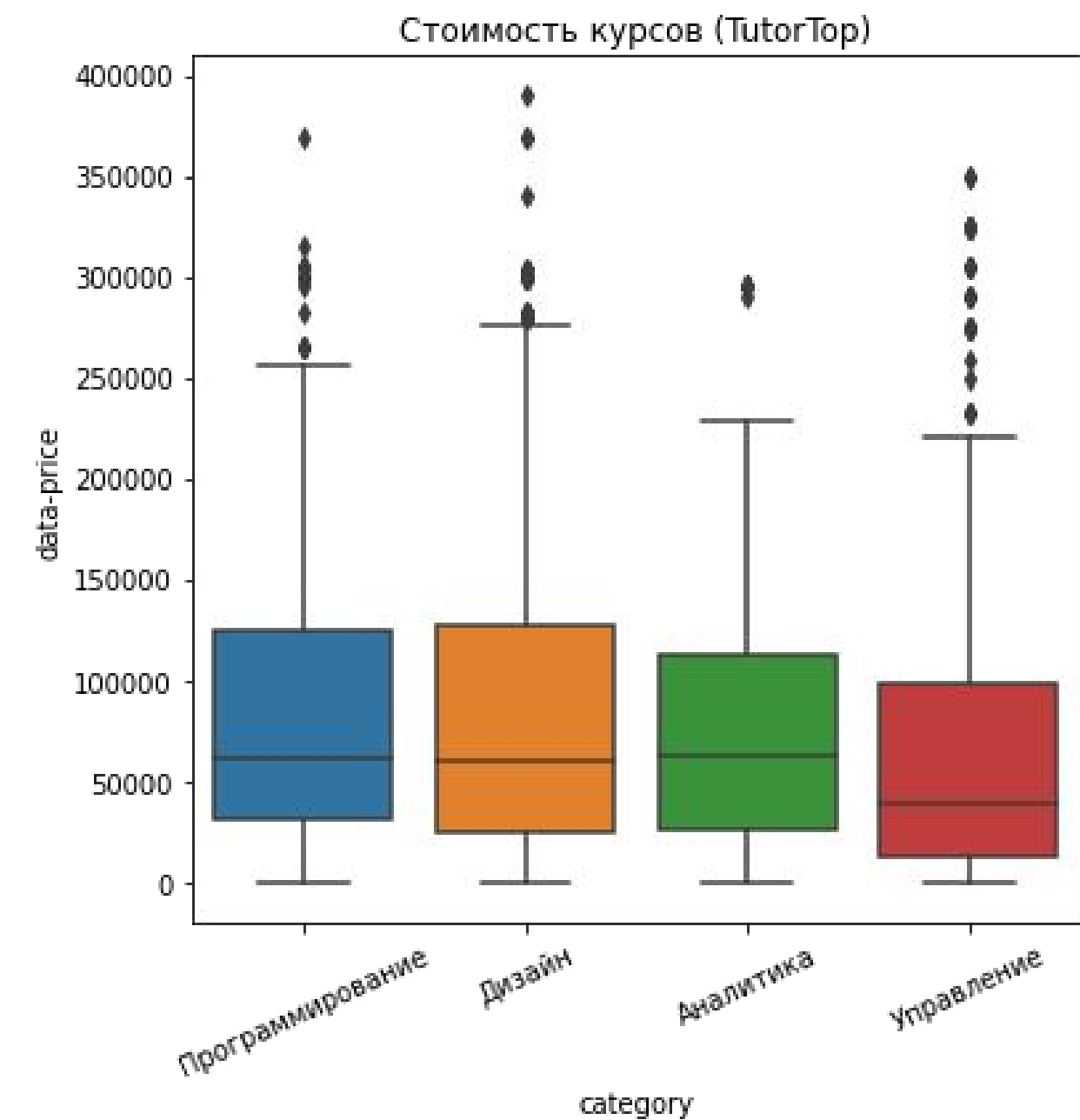
Результаты анализа

Результаты проверки гипотезы

агрегатор Coursotor

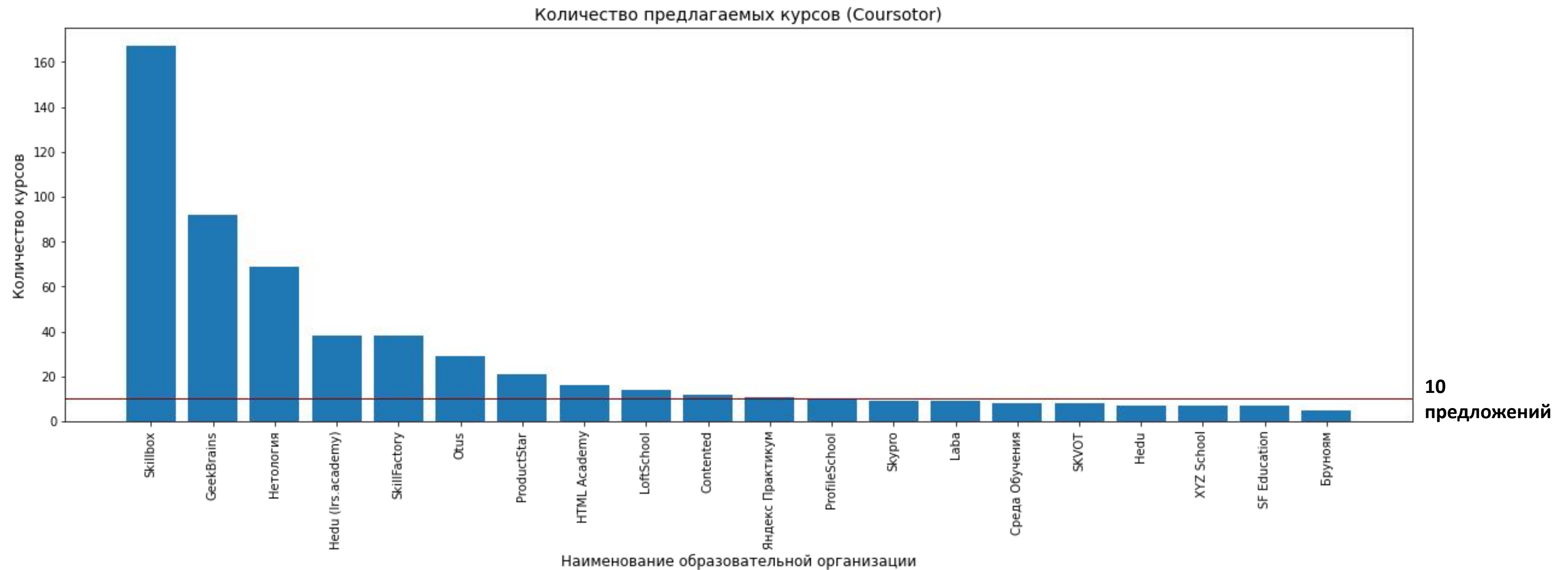


агрегатор TutorTop



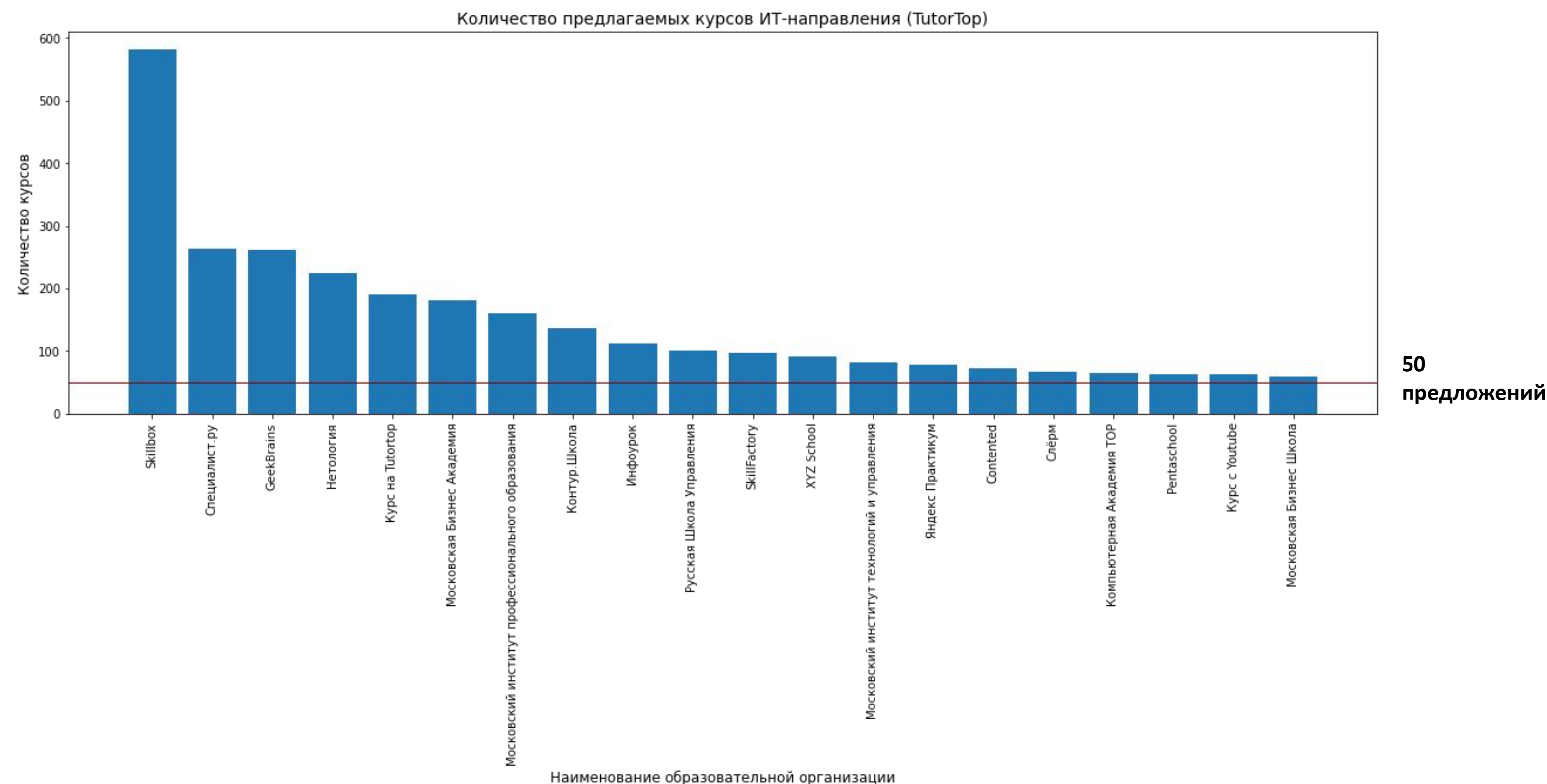
Результаты анализа

Результаты проверки гипотезы



Результаты анализа

Результаты проверки гипотезы



Результаты анализа

Описание результатов

1. Данные , полученные от двух разных агрегаторов курсов разнятся как по составу образовательных заведений, так и по количеству и направлениям, предлагаемых ими курсов.
2. Подтверждения или опровержения гипотезы о том, что рынок ИТ-курсов примерно равномерно разделен не найдено. Необходимо провести дополнительный сбор и анализ данных.
3. Для анализа навыков и знаний, получаемых слушателями на существующих курсах, из имеющегося датасета данных недостаточно, необходим более глубокий анализ программ обучения на сайтах образовательных организаций.

Интерпретация данных

Полученные результаты:

1. Рынок вакансий действительно нуждается в ИТ-специалистах.
2. Большая доля рынка так же принадлежит специалистам по работе с кадрами и недвижимостью.
3. На данный момент наилучшие условия по оплате труда в ИТ-области предоставляются специалистам, так или иначе связанным с программированием.
4. Вакансии в ИТ-области с удаленным форматом работы предлагают лучшие условия оплаты труда, следовательно нужно уметь работать самостоятельно, и знать инструменты и технологии управления проектами, хранения данных, контроля версий разрабатываемого программного обеспечения.
5. Рынок образовательных услуг нуждается в более глубоком изучении. Информации из датасета с сайта Coursotor недостаточно.

Интерпретация данных

Рекомендации для Заказчика:

- Для понимания какие навыки наиболее необходимы работодателям (слушатели курсов, именно в таких навыках и заинтересованы) рекомендуется изучить дашборд «Вакансии». В отдельном поле «Необходимые навыки» выводится градация востребованности знаний и умений.
- Дашборды с Курсами посвящены данным с сайтов-агрегаторов курсов , и из анализа можно получить сведения о стоимости и длительности курсов для различных направлений.

Интерпретация данных

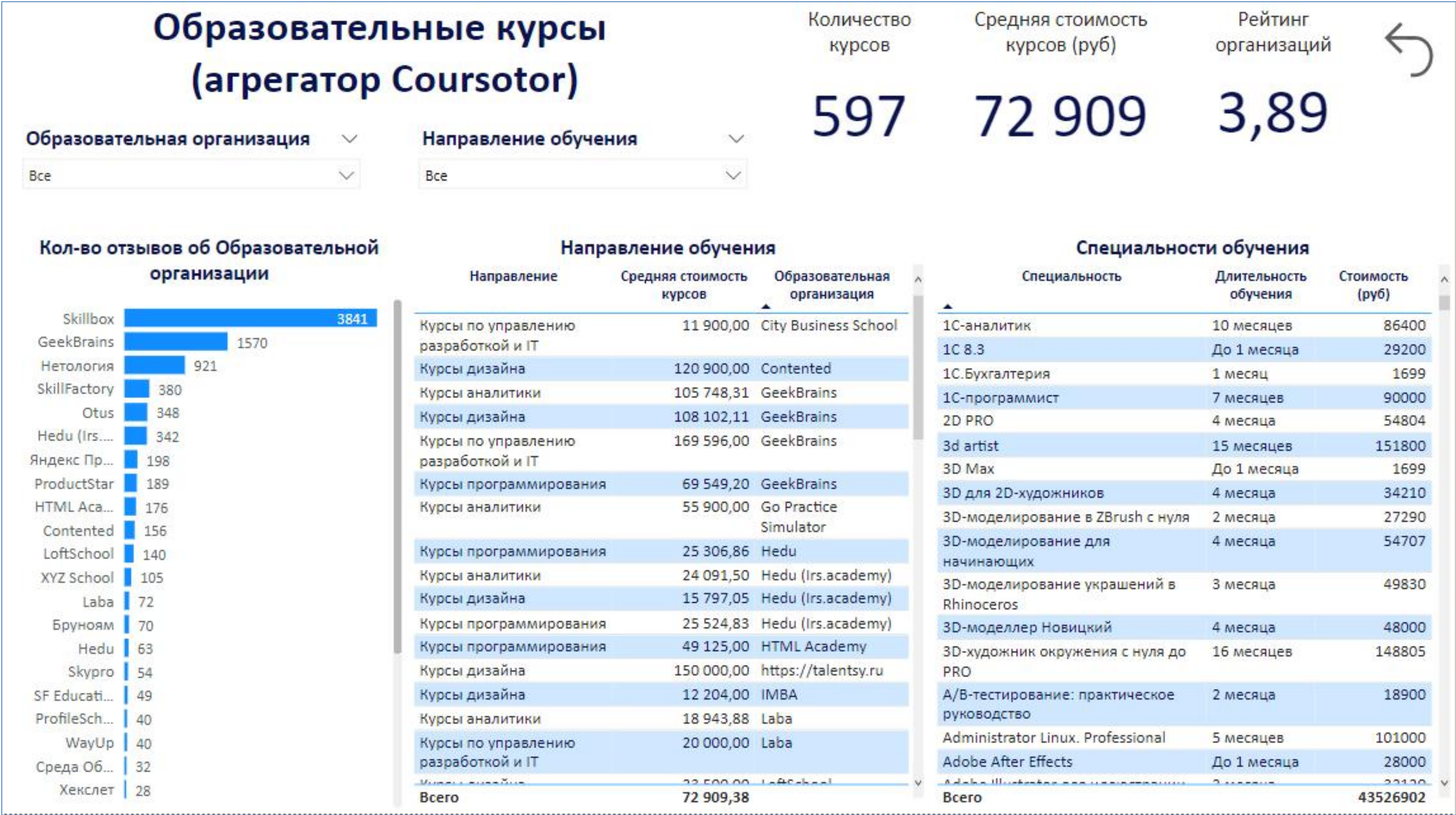
Перспективные направления для дальнейшего анализа :

- Сбор данных и анализ рынка образовательных услуг на более глубоком уровне, необходимо понимать качество предлагаемого обучения.
- Сбор и анализ данных баз резюме соискателей, для понимания имеющихся у специалистов навыков, а так же уровня образования и имеющихся данных о ДПО.
- Анализ рынка труда на других площадках-агрегаторах вакансий.
- Анализ соцсетей для выявления реальных потребителей рынка образовательных услуг.
- По результатам анализа рынка труда было выявлено, что существенную долю рынка занимают вакансии по работе с кадрами и недвижимостью - рекомендуется уделить внимание данным направлениям.

Дашборд



Дашборд



Дашборд

