

## ПЛАН ИССЛЕДОВАНИЯ

<b>Заказчик</b>	IT-компания
<b>Методология исследования, этапы исследования</b>	<p>CRISP-DM</p> <ol style="list-style-type: none"> <li>1. Понимание бизнеса (business understanding)</li> <li>2. Понимание данных (data understanding)</li> <li>3. Подготовка данных (data preparation)</li> <li>4. Моделирование (modeling)</li> <li>5. Оценка результатов (evaluation)</li> <li>6. Внедрение (deployment)</li> <li>7. Контроль</li> </ol>
<b>Бизнес-цель заказчика</b>	<p>Найти нишу для бизнеса по продажам образовательных продуктов для IT-специалистов в 2022 году.</p> <p>Востребованность на рынке (заказчики могут быть компании).</p>
<b>Объект исследования (Что анализируем?)</b>	<p>Рынок труда IT-отрасли;</p> <p>Крупные игроки на образовательном рынке IT и их продукты.</p>
<b>Предмет исследования (свойства, характеристики)</b>	<p>Какие есть вакансии на рынке IT, Сколько их?</p> <p>Требования к сотрудникам?</p> <p>Меняются ли требования к сотрудникам?</p> <p>какие специалисты ищут работу?</p> <p>Какие навыки и компетенции наиболее распространены среди соискателей?</p> <p>Можно ли найти несоответствия между спросом на отдельные компетенции и предложением со стороны кандидатов?</p> <p>Информация о самых крупных игроках рынке IT:</p> <ul style="list-style-type: none"> <li>-количество программ в сфере,</li> <li>-их тематическая направленность,</li> <li>-объем (продолжительность обучения) и содержание курсов,</li> <li>их стоимость,</li> <li>-среднюю оценку потребителей,</li> <li>-возможно, их отзывы.</li> </ul> <p>Информация о средней заработной плате для специалистов в данной области, количество вакансий.</p>

<b>Цель анализа данных</b>	<p>Оценить востребованность образовательных продуктов на основании и ситуацию на рынке труда на основе:</p> <ul style="list-style-type: none"> <li>- спроса по количеству вакансий;</li> <li>- предложения по количеству опубликованных резюме.</li> </ul> <p>Сравнить содержание предлагаемых курсов с требованиями работодателей, предъявляемыми к соискателям на позиции IT-специалистов разных направлений</p> <p>Изучить и выбрать стратегию производства продукта на основании вакансий по IT-специальностям, их количеству и требованиям в них, предложениям конкурентов для оценки цены.</p> <p>Декомпозиция:          Что подразумевает заказчик под образовательным продуктом?          Кто является целевой аудиторией?          Какие услуги оказывает заказчик сейчас?</p>
<b>Требования к результату анализа</b>	<ul style="list-style-type: none"> <li>• Наглядное описание ситуации на рынке труда и ситуации в дополнительном профессиональном образовании (онлайн-курсы);</li> <li>• Результат бенчмаркинга продуктов конкурентов;</li> <li>• Представление выводов и рекомендаций: какие образовательные продукты нужны рынку, какая цена (или диапазон цен) можно поставить, чтобы не продешевить, но и люди могли покупать.</li> </ul>
<b>Планируемые источники данных</b>	<p>Tutortop.ru (банк онлайн-курсов) или аналогичное          HH.ru, SuperJob, Работа в России и т.п.</p>
<b>Типы данных</b>	<p>Структурированные данные, Неструктурированные данные, Полуструктурированные данные, Открытые.          Форматы: XLSX-файлы, TXT-файлы, docx, XML-файлы.</p>
<b>Способ(-ы) получения данных</b>	<ul style="list-style-type: none"> <li>• Сбор/получение первичной исходной информации</li> <li>• Получение данных из вторичных (сторонних) источников</li> <li>• Парсинг — автоматизированный сбор информации с сайта, ее анализ, преобразование и выдача в структурированном виде</li> </ul>
<b>Методы анализа</b>	<p>Описательные статистики (описание продуктов конкурентов)          Корреляционный анализ (взаимосвязь между ценой курса и содержанием)          Регрессионный анализ (предсказать стоимость курсов)          Классификация (Будут востребованы курсы или нет?)          Кластеризация (Сегментация курсов, python, java и востребованность их)          Социально-сетевой анализ (описать кто лучше и почему, оценка отзывов)</p>
<b>Инструменты анализа</b>	<ul style="list-style-type: none"> <li>• MS Excel ,</li> <li>• Python (Pandas, matplotlib и др.)</li> </ul>
<b>Риски и условия реализации проекта</b>	<ul style="list-style-type: none"> <li>• Не уложиться в сроки</li> <li>• Финансовые риски</li> <li>• Малое количество или плохое качество данных</li> <li>• Данные качественные, но закономерности в принципе отсутствуют</li> <li>• В данных отсутствует необходимая информация (взаимосвязи)</li> <li>• Поломка программного обеспечения или компьютера</li> </ul>