# Задание по стажировке: Раздел 1

Внимательно прочитайте текст задания по стажировке и составьте план предстоящего исследования. Используйте табличную форму:

**ПЛАН ИССЛЕДОВАНИЯ**

|  |  |
| --- | --- |
| **Заказчик** | — IT-компания, последние 15 лет успешно занимавшаяся разработкой и консалтингом на российском рынке |
| **Методология исследования, этапы исследования** | CRISP-DM  1. Понимание бизнеса (Бизнес-цель заказчика, оценка текущей ситуации, определение целей аналитики)  2. Понимание данных (сбор данных с сайтов вакансий и агрегаторов курсов, описание данных (их количество, тип), их изучение, проверка качества ((наличие не корректных данных, соответствие типов, наличие пропусков))  3. Подготовка данных (выборка данных (загрузка в программу для работы с данными), очистка данных (от выбросов, не нужных столбцов), замена или удаление пропусков, изменение типов при не соответствии, разбиение признаков на колонки или слияние в новую)  4. Моделирование (выдвижение и проверка гипотез зависимости критериев, выбор статистик и алгоритмов в зависимости от типов данных, обучение моделей и оценка их качества)  5. Оценка результатов. |
| **Бизнес-цель заказчика** | найти свою нишу, в которой заказчику не пришлось бы конкурировать с «образовательными гигантами», выпускающими большое количество образовательных курсов, ориентированных скорее на массовость, чем на качественную подготовку специалистов. готовить востребованный продукт, который смогут покупать как отдельные потребители, так и компании, заинтересованные в переподготовке и повышении квалификации своих сотрудников.  (наверное правильная цель - конкретная и измеримая - открыть 2 группы переподготовки и повышения квалификации по 2-м специальностям) |
| **Объект исследования**  **(Что анализируем?)** | 1. рынок труда  2. предложения потенциальных конкурентов  ( сравнение содержание предлагаемых курсов с требованиями работодателей, предъявляемыми к соискателям на позиции IT-специалистов разных направлений на самых популярных базах вакансий России) |
| **Предмет исследования**  **(свойства, характеристики)** | Навыки, требующиеся работодателю.  Топ вакансий с предлагаемой ЗП  Курсы предлагаемые в сети(их содержание - какие навыки и знания формируют, их продолжительность и стоимость) |
| **Цель анализа данных** | Выяснить самые востребованные знания и навыки. Оценить количество, продолжительность и стоимость курсов предлагающие эти знания и навыки. |
| **Требования к результату анализа** | * наглядное описание ситуации на рынке труда и ситуации в дополнительном профессиональном образовании (онлайн-курсы); * результат бенчмаркинга продуктов конкурентов; * представление выводов и рекомендаций: какие образовательные продукты нужны рынку, какая цена (или диапазон цен) можно поставить, чтобы не продешевить, но и люди могли покупать. |
| **Планируемые источники данных** | Популярные базы вакансий России — HH.ru, SuperJob, Работа в России и т.п  Сайты сравнения курсов онлайн обучения: Tutortop.ru, topkursov.ru,  journal.tinkoff.ru/sravnyator и т.д. |
| **Типы данных** | структурированные, полуструктурированные |
| **Способ(-ы) получения данных** | Данные заказчика, парсинг |
| **Методы анализа** | описательные статистики(построение графиков и нахождение средних величин), статистические критерии сравнения групп(сравнение по профессиям), корреляционный анализ(зависимость ЗП от требуемых навыков). |
| **Инструменты анализа** | Python, Pandas, Matplotlib, Seaborn, Jupiter notebook, Colab |
| **Риски и условия реализации проекта** | Политическая ситуация в стране и мире после 24.02.2022. Все выводы могут потерять актуальность при изменении ситуации |

Пояснительная записка.

В телеграмм канале и на консультации говорилось о соответствии плана исследования и реальных действий. То есть, если в плане присутствует пункт о данных которых нет, то их надо получить и проанализировать. В то же время, для получения зачета по стажировке (по Варианту 1 - для начинающих аналитиков — БЕЗ сбора данных), достаточно провести анализ только тех данных которые предоставил заказчик, а по ним можно только сказать, кто востребован на рынке и по какой "цене". Так же, если бы была возможность общаться с заказчиком, то попытался конкретнее выяснить бизнес цель (сколько и каких курсов) планируется. Ниша в которой планируется работа насколько "занята". Конкурентное преимущество компании (на мой взгляд) в большой практической деятельности и начинать деятельность именно в том где они считают профессионалами и можно попытаться найти персонал и информацию по программам внутри компании, а не изучать востребованные специальности, в которых нет такого богатого опыта.