**Паспорт проекта**

**Проект для студентов бакалавриата:**

* «ИТ-сервисы и технологии обработки данных на транспорте»

**Проектный наставник: Кушеверская Мария Александровна   
Архитектор проектов внедрений** ООО «ХФ Лабс»

**Название проекта:** Аналитика данных по метрикам качества (с визуализацией или без)

**Формулировка проблемы**

Любая MDM система (управление мастер-данными,) на вход получает большое количество данных. Полученные данные необходимо правильно обработать, исходя из их специфики. Для этого сначала требуется провести первичный анализ данных, загруженных из источников, выявить ошибки, аномалии, особенности.

Также, после отработки алгоритмов системы, необходимо собирать статистику по обработанным данным с целью улучшения бизнес-процессов

**Предлагается выбрать набор статистических метрик для анализа данных и визуализировать результаты.**

**Это поможет:**

1. **Оценить качество входных данных, выявить аномалии, ошибочные значения в датасете.**
2. **Оценить результат обработки данных системой, скорректировать настройки в зависимости от качества входных данных.**
3. **Дать рекомендации по улучшению качества данных в системе-источнике.**

**Требования к входным компетенциям для участия в проекте:**

1. Уметь работать с большими объемами структурированных данных.
2. Уметь писать sql-запросы к БД.
3. Понимать, как формируется перечень метрик для итоговой визуализации.
4. Уметь визуализировать агрегированные данные.

**Исходные данные:** синтетические данные физических лиц и юридических лиц.

**Описание результата. Результат, который планируется получить в финале проектной работы со студентами:**

**Продуктовый:** продукт, позволяющий увидеть состав и качество данных и по запросу сформировать отчет.

**Образовательный:**

1. опыт анализа предметной области;
2. опыт работы с большими объемами данных;
3. опыт анализа качества данных и формирования наборов метрик для датасета;
4. опыт визуализации результатов на графиках и дашбордах.

**Состав студенческой команды: 2-5 человек**

**Возможные роли в команде:**

1. Аналитик;
2. Разработчик;
3. Тестировщик;
4. UI/UX дизайнер.

**Количество студенческих команд на проектную задачу (если предусмотрено соревнование).**

**График и условия работы над продуктом (этапы и сроки реализации проекта, частные задачи на каждом этапе, задачи для различных ролевых подгрупп).**

Этапы:

1. Описание предметной области задачи и целей исследования;
2. Анализ данных
3. Формирование перечня визуализируемых метрик;
4. Выбор технологического стека;
5. Разработка программного обеспечения;
6. Формирование графических элементов визуализации;
7. Тестирование и отладка работы программного обеспечения;
8. Подготовка итоговой презентации/отчета с результатами, выводами и рекомендациями.

**Заказчик проекта (ИЛИ кому и каким компаниям будут полезны полученные студентами результаты проекта):**

Заказчик: ООО «ХФ Лабс».

**Эксперт от индустрии:**

Кушеверская Мария Александровна – архитектор проектов внедрений ООО «ХФ Лабс»

**Каким образом результаты проекта могут быть востребованы в индустрии (у заказчика) - практическая ценность:**

Результат визуализации выбранных метрик позволит оценить качество загруженных в MDM-систему данных и результат обработки этих данных.

Это позволит корректировать алгоритмы обработки данных системой с учетом особенностей данных, а также формировать гипотезы по улучшению качества данных, приходящих из систем-источников.

**Ресурсное обеспечение проекта. Для получения итогового продукта необходимо следующее ПО:**

1. Язык программирования Python или аналогичный;
2. ПО для визуализации данных, например, Tableau Public, Redash или аналоги.
3. БД на выбор, например, PostgreSQL, ClickHouse, SQLite; клиент БД (Dbeaver и аналоги)