**МОДУЛЬ Б**

**Исходные файлы:**

1. Data.zip (Результаты предыдущего модуля)
2. Модуль Б.pdf (Инструкция к модулю Б)

**Результаты работы:**

1. Data.zip (Результаты модуля Б)
2. Report.html + Report.ipynb (Отчет о проделанной работе)
3. Readme.txt (Дополнительные комментарии)

**ВВЕДЕНИЕ**

В этом модуле вы продолжаете работать с данными, подготовленными в предыдущем модуле. Какая-либо работа, обусловленная задачами предыдущего модуля, выполненная в ходе текущего, оцениваться не будет, поэтому проделывайте её только в случае необходимости.

**ЗАДАНИЕ**

|  |
| --- |
| 2.1 Построение аналитической системы |
| Для своевременного принятия решения на основе данных, минимизации ошибок и быстрого достижения целей, необходимо построить аналитический дашборд. Интерактивная информационная панель должна наглядно представлять, визуализировать и анализировать данные. Для этого необходимо подключить к системе источник данных, т.е. подключиться к базе данных, настроить регламент обновления и уровни доступа.  Аналитическая система должна позволить пользователю:   * Автоматизировать работу с данными; * Анализировать все в реальном времени; * Видеть все главные метрики на одном экране; * Легко считывать информацию и делать выводы. |

|  |
| --- |
| 2.2 Кластеризация набора данных |
| Разработанная информационная панель должна позволить пользователю проводить анализ показателей эффективности и востребованности транспортных средств.  Необходимо вычислить такие характеристики, как среднее количество человек на поездку в день, суммарная стоимость поездок в районах (считается по начальному району) в неделю, среднее суммарное количество чаевых в день за месяц, суммарное количество пассажиров в день, процентное соотношение количества поездок по способам оплаты за неделю.  Визуализируйте динамику полученных метрик во времени. В дашборде должны быть отражены ключевые сводные показатели. |

**После выполнения пунктов 2.1 и 2.2 необходимо продемонстрировать выполненную работу экспертам**

|  |
| --- |
| 2.3 Кластеризация набора данных |
| Выберите модель кластеризации данных. Задача кластеризации – сегментация поездок, определение схожих группы объектов в каждом наборе данных. В результате должно быть выявлено не менее 3 и не более 5 кластеров. При кластеризации могут получиться группы поездок, схожие по различным характеристикам. Необходимо выявить такие характеристики и дать подробное описание каждого кластера. |

|  |
| --- |
| 2.4 Вычисление показателей производительности кластеризации |
| Выберите метрику оценки качества кластеризации. Обоснуйте выбор методов и  приемов. Выполните визуальный анализ кластерных структур и оценки качества кластеризации.  Определите лучший алгоритм кластеризации на основе выбранной метрики. |

|  |
| --- |
| 2.5 Подготовка отчета |
| Подготовьте отчет о проделанной работе по итогам сессии в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива Data.zip |