**Модуль Б. Разведочный анализ данных**

Время на выполнение модуля – 3 часа

**Задания:**

|  |
| --- |
| 2.1 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. |

|  |
| --- |
| 2.2 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. |

|  |
| --- |
| 2.3 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 |