|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ **Информатика и системы управления**

КАФЕДРА **Компьютерные системы и сети (ИУ6)**

НАПРАВЛЕНИЕ ПОДГОТОВКИ **09.04.01 Информатика и вычислительная техника**

**Отчет**

|  |  |
| --- | --- |
| **по лабораторной работе №** | 10 |

**Название:** Обработка больших данных

**Дисциплина:** Языки программирования для работы с большими данными

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6-21М |  |  | Ю.А. Вишневская |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | П. В. Степанов |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

Москва, 2020

**Вариант 1**

Скачать датасет с Kernel и выполнить несколько запросов.

|  |
| --- |
| package LR10  import org.apache.spark.sql.SparkSession  import org.apache.spark.{SparkConf, SparkContext}  object LR10\_V1 {  def main(args: Array[String]): Unit = {  val conf = new SparkConf().setAppName("LR10\_V1").setMaster("local[\*]")  val sc = new SparkContext(conf);  val spark = SparkSession.builder.appName("Test app").getOrCreate()  val datafile = spark.read  .format("com.databricks.spark.csv")  .option("header",true)  .load("C:/Users/User/IdeaProjects/LR10/HRR\_Scorecard.csv")  //datafile.show()  datafile.createOrReplaceTempView("HRR\_Scorecard")  // Максимальное количество прогнозируемых зараженных COVID-19 в США старше 18 лет  spark.sql("SELECT MAX(`Projected Infected Individuals`) as max FROM HRR\_Scorecard group by `HRR`").show()  // Среднее количество пенсионеров в больницах США  spark.sql("SELECT `HRR`, AVG(`Population 65+`) as avg FROM HRR\_Scorecard group by `HRR`").show()  // Количество мест в больницах Hospital и ICU  spark.sql("select `Total ICU Beds`, `Total Hospital Beds`, `HRR` FROM HRR\_Scorecard WHERE `Total Hospital Beds` > 100").show()  spark.stop()  }  } |