|  |  |
| --- | --- |
| 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, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ и системы  
 управления**

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

НАПРАВЛЕНИЕ ПОДГОТОВКИ **09.04.01 Информатика и вычислительная техника**

МАГИСТЕРСКАЯ ПРОГРАММА **09.04.01/07 Интеллектуальные системы анализа,**

**обработки и интерпретации больших данных**

**Отчет**

**по лабораторной работе №10**

**Название:** Spark

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

Вариант: 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6-22М |  |  | Н.А. Аскерова |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | П.В. Степанов |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

**Задание**

1) Выбрать любой датасет на kaggle.com

2) Cделать 10 выборок данных по выбранной предметной области Листинг 1 – Код программы

package org.example;

import org.apache.spark.SparkConf;

import org.apache.spark.SparkContext;

import org.apache.spark.api.java.JavaSparkContext;

import org.apache.spark.sql.Dataset;

import org.apache.spark.sql.Row;

import org.apache.spark.sql.SparkSession;

public class Main {

public static void main(String[] args) {

SparkSession spark = SparkSession.builder()

.appName("RandomSplitExample")

.master("local[\*]")

.getOrCreate();

Dataset<Row> data = spark.read().format("csv")

.option("header", "true")

.load("/Users/Admin/Downloads/archivehonda/honda\_sell\_data.csv");

data.createOrReplaceTempView("honda");

spark.sql("select \* from honda ").show();

spark.sql("select \* from honda where Consumer\_Rating='4.4'").show();

spark.sql("select honda.Model, AVG(honda.Consumer\_Rating) from honda group by honda.Model ").show();

spark.sql("select \* from honda where honda.Interior\_Color='Gray' order by honda.Consumer\_Rating").show();

spark.sql("select \* from honda order by 5").show();

spark.sql("select honda.Consumer\_Rating, honda.Model, honda.Price from honda where honda.Model='Crosstour EX' and honda.Consumer\_Rating > 4.4 group by 1,2,3 order by 1").show();

spark.sql("select honda.Model, honda.Year, sum(honda.Comfort\_Rating) from honda where honda.Model='Crosstour EX' group by 1,2").show();

spark.sql("select \* from honda where honda.Drivetrain='All-wheel Drive' order by honda.Price").show();

spark.sql("select honda.Model, honda.Transmission from honda where honda.Mileage=5 group by 1,2").show();

spark.sql("select honda.VIN, honda.Model from honda order by 2").show();

// Останавливаем SparkSession

spark.stop();

}

}

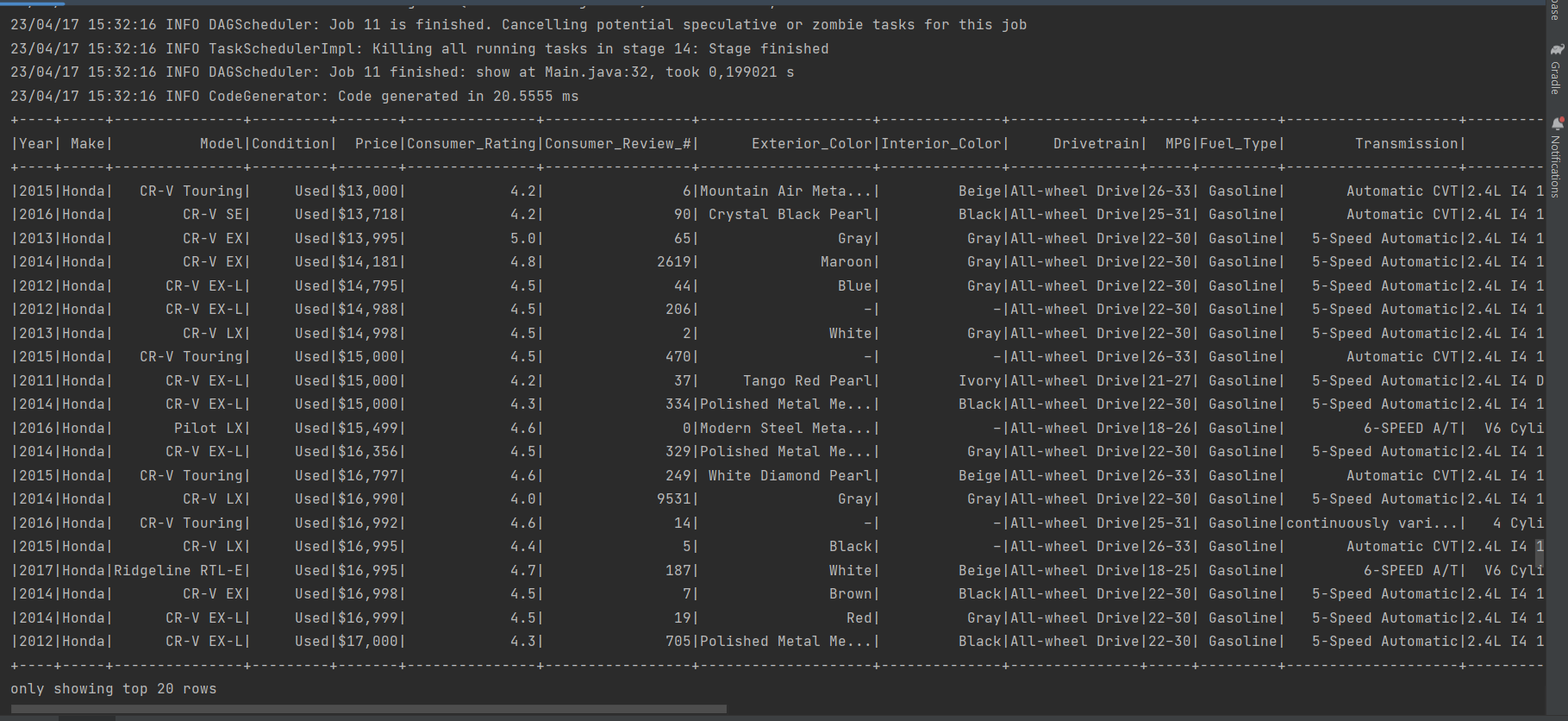


Рисунок 1 – Результат работы программы

**Вывод:** приобретен навык работы с Spark.

**Ссылка на репозиторий с программами:** [**https://github.com/nargi3/BigData**](https://github.com/nargi3/BigData)