|  |  |
| --- | --- |
| 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**

МАГИСТЕРСКАЯ ПРОГРАММА **09.04.01/07 Интеллектуальный анализ больших**

**данных в системах поддержки принятия решений.**

**Отчет**

**по лабораторной работе № 10**

**Вариант № 12**

**Название:** Spark

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

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

Москва, 2024

**Цель:** изучить и освоить принципы работы со Spark в Java.

**Задание**: сделать 10 выборок данных по выбранной предметной области.

В качестве предметной области был выбран датасет “USA Cars Dataset”.

Этот набор данных содержит информацию о машинах, продающихся в США. Описание полей датасета представлено на рисунке 1. Часть датасета представлена на рисунке 2.

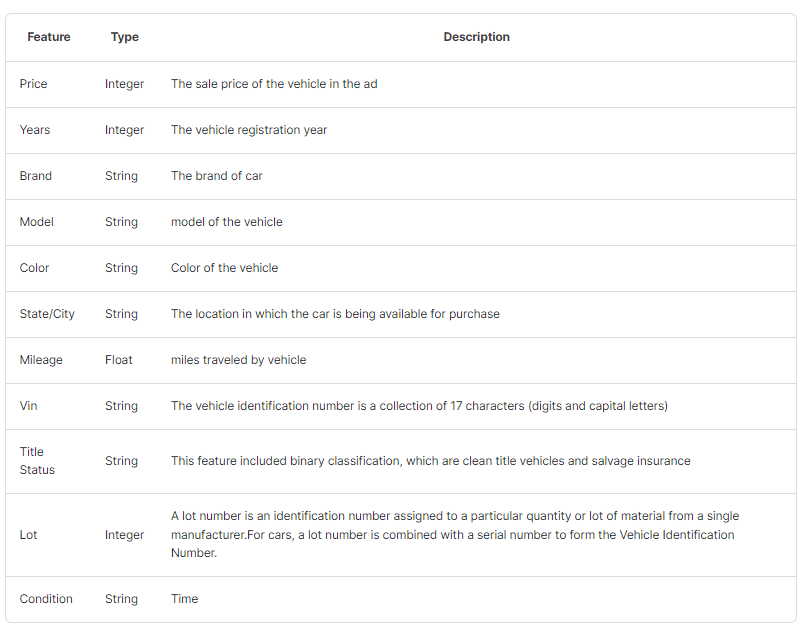


Рисунок 1 – Описание полей датасета

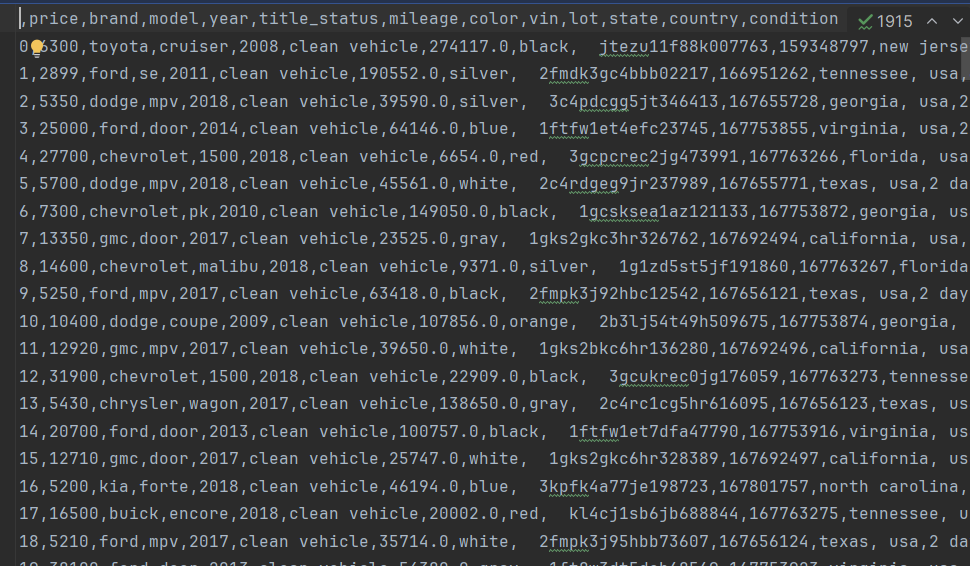


Рисунок 2 – Часть датасета

Код программы:

package org.example

import org.apache.spark.sql.SparkSession

fun main() {

val spark = SparkSession.builder()

.appName("USA Cars")

.master("local[\*]")

.getOrCreate()

spark.sparkContext().setLogLevel("ERROR")

val dataset = spark.read().format("csv")

.option("header", "true")

.option("inferSchema", "true")

.load("USA\_cars\_datasets.csv")

dataset.createOrReplaceTempView("cars")

println("Весь датасет:")

spark.sql("SELECT \* FROM cars").show()

println("1. Сортировка по цене по убыванию")

spark.sql("SELECT \* FROM cars ORDER BY price DESC").show()

println("2. Автомобили после 2015 года")

spark.sql("SELECT \* FROM cars WHERE year > 2015").show()

println("3. Количество автомобилей по бренду")

spark.sql("SELECT brand, COUNT(\*) as count FROM cars GROUP BY brand").show()

println("4. Автомобили с пробегом меньше 50,000")

spark.sql("SELECT \* FROM cars WHERE mileage < 50000").show()

println("5. Средняя цена автомобилей")

spark.sql("SELECT AVG(price) as avg\_price FROM cars").show()

println("6. Автомобили из штата Georgia")

spark.sql("SELECT \* FROM cars WHERE state = 'georgia'").show()

println("7. Автомобили черного цвета")

spark.sql("SELECT \* FROM cars WHERE color = 'black'").show()

println("8. Сортировка по году по возрастанию")

spark.sql("SELECT \* FROM cars ORDER BY year ASC").show()

println("9. Автомобили с указанием оставшихся дней")

spark.sql("SELECT \* FROM cars WHERE condition LIKE '%days left%'").show()

println("10. Максимальная цена по бренду")

spark.sql("SELECT brand, MAX(price) as max\_price FROM cars GROUP BY brand").show()

}

Результат работы программы представлен на рисунке 3.

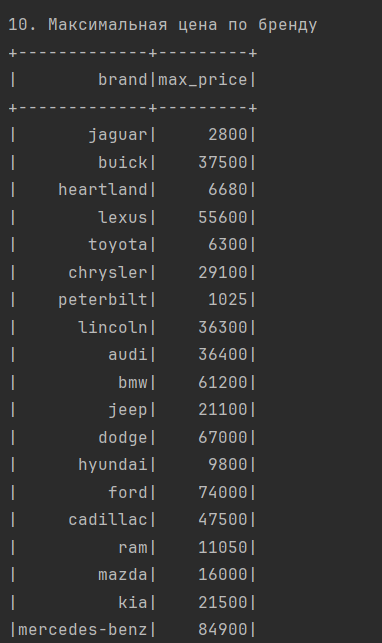


Рисунок 1 – Результат работы программы

**Вывод:** были освоены принципы работы со Spark в Java.