|  |  |
| --- | --- |
| 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 **09.04.01/12 Интеллектуальный анализ больших**

**данных в системах поддержки принятия решений**

**Отчет**

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

Вариант 6

**Название:** Spark

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6-23М |  |  | В.А. Гордеев |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | П.В. Степанов |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

Москва, 2023

**ВВЕДЕНИЕ**

**Задание**

1. Выбрать любой датасет на kaggle.com
2. Cделать 10 выборок данных по выбранной предметной области с помощью spark

**ОСНОВНАЯ ЧАСТЬ**

**Задание 1.1**

Был выбран датасет с информацией о пиве на kaggle.com. Далее сделано 10 выборок данных по данному датасету с помощью spark.

Код программы приведен в листинге 1. Результат выполнения программы представлен на рисунках 1, 2, 3, 4, 5.

Листинг 1 – Код программы

package org.example;  
  
import org.apache.spark.sql.Dataset;  
import org.apache.spark.sql.Row;  
import org.apache.spark.sql.SparkSession;  
import java.nio.file.Paths;  
  
public class Main {  
 public static void main(String[] args) {  
 String fileName = "beers.csv";  
 String currentDir = System.getProperty("user.dir");  
 String filePath = Paths.get(currentDir).resolve(fileName).toString();  
  
 SparkSession spark = SparkSession  
 .builder()  
 .appName("Lab10.1")  
 .master("local")  
 .getOrCreate();  
  
 Dataset<Row> df = spark.read()  
 .option("header", true)  
 .option("inferSchema", true)  
 .csv(filePath);  
 df.createOrReplaceTempView("beers");  
  
 spark.sql("SELECT \* FROM beers").show();  
 spark.sql("SELECT \* FROM beers WHERE abv IS NOT null ORDER BY abv ASC ").show();  
 spark.sql("SELECT \* FROM beers WHERE abv < 0.03").show();  
 spark.sql("SELECT abv, name FROM beers WHERE abv = 0.09 OR abv = 0.08").show();  
 spark.sql("SELECT \* FROM beers WHERE style = 'Cider'").show();  
 spark.sql("SELECT name, style, ounces FROM beers WHERE style = 'Cider' ORDER BY ounces DESC ").show();  
 spark.sql("SELECT abv, name, style FROM beers WHERE name LIKE '%Cider%' ").show();  
 spark.sql("SELECT abv, name FROM beers WHERE name LIKE '%Cider%' AND abv = 0.06 ").show();  
 spark.sql("SELECT name, brewery\_id FROM beers WHERE brewery\_id = 154 ").show();  
 spark.sql("SELECT name, brewery\_id FROM beers WHERE brewery\_id = 154 AND name NOT LIKE '%Cider%'").show();  
 }  
}

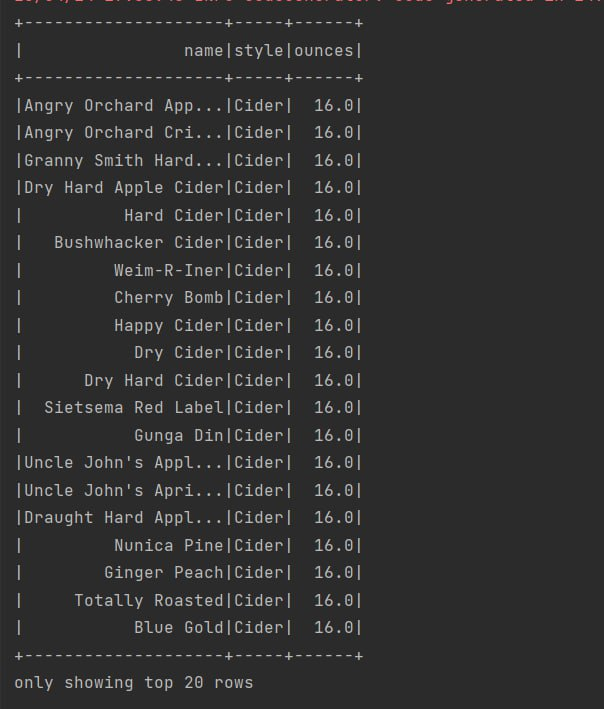


Рисунок 1 – Результат работы программы

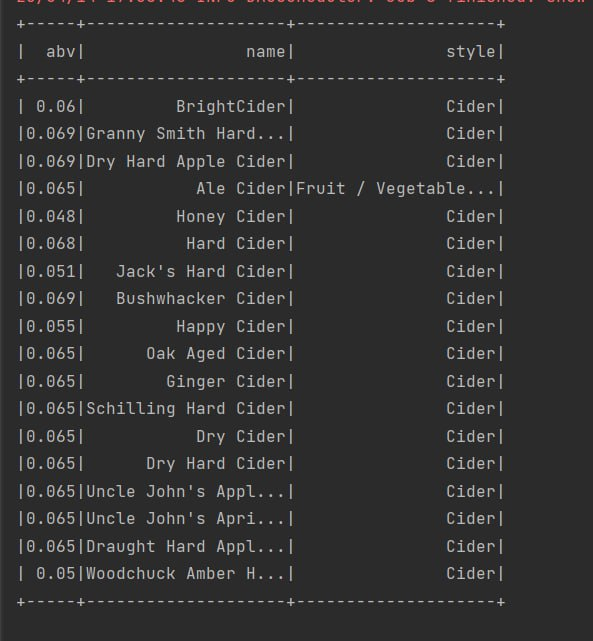


Рисунок 2 – Результат работы программы

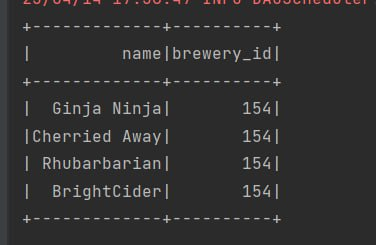


Рисунок 3 – Результат работы программы

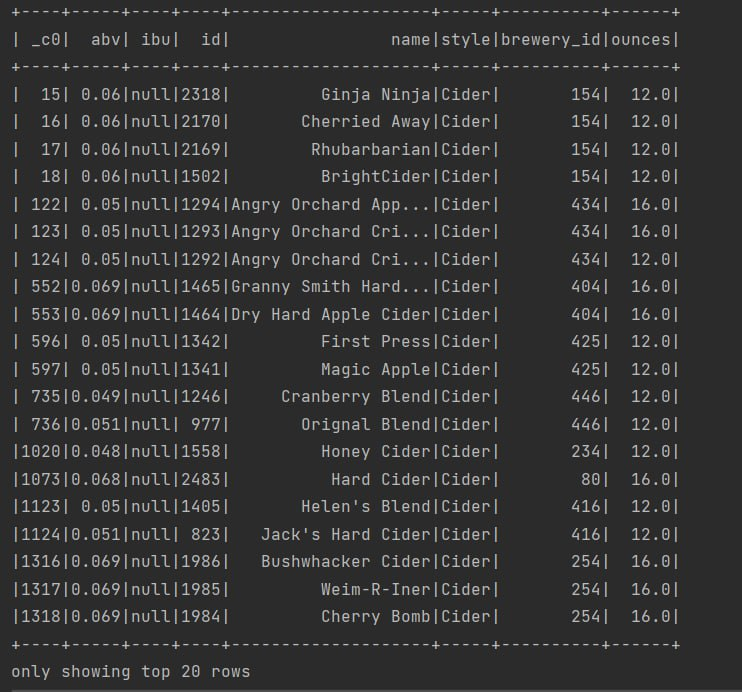


Рисунок 4 – Результат работы программы

**Вывод:** В результате выполнения лабораторной работы были получены практические навыки при работе со Spark. Написано приложение, которое делает несколько выборок из датасета.