



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.

О Т Ч Е Т

по лабораторной работе № 10

Вариант 6

Название: Spark

Дисциплина: Языки программирования для работы с большими данными

Студент

ИУ6-23М

(Группа)

(Подпись, дата)

Г.Л. Кушнир

(И.О. Фамилия)

Преподаватель

(Подпись, дата)

П.В. Степанов

(И.О. Фамилия)

Москва, 2023

Цель: Ознакомиться с работой Spark в языке программирования Java.

Задание: Сделать 10 выборки данных по выбранной предметной области.

Был выбран Dataset с лапшой рамен, с показателем рейтингов, указанием брендов, типом упаковки лапши, страны создания данного бренда.

Код основного класса Main:

```
package org.example;

import org.apache.spark.sql.Dataset;
import org.apache.spark.sql.Row;
import org.apache.spark.sql.Session;

public class Main {

    public static void main( String[] args )
    {

        SparkSession spark = SparkSession.builder()
            .appName("Ramen Rating")
            .master("local[*]")
            .getOrCreate();

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

        Dataset<Row> dataset = spark.read().format("csv")
            .option("header", "true")
            .load("ramen-ratings.csv");
        dataset.createOrReplaceTempView("ramen_rate");

        System.out.println("1. All dataset:");
        spark.sql("select * from ramen_rate").show();

        System.out.println("2. Count of ramen's brand:");
        spark.sql("select count(distinct Brand) as BrandCount from
ramen_rate").show();

        System.out.println("3. All ramen from Japan with kimchi:");
        spark.sql("select Brand, Variety, Style, Stars from ramen_rate where
Country = 'Japan' and Variety like '%Kimchi%'").show();

        System.out.println("4. Average stars rate of all ramen:");
        spark.sql("select avg(Stars) as StarsAvg from ramen_rate").show();

        System.out.println("5. All Nissin ramen with rate above 4:");
        spark.sql("select Variety, Style, Country, Stars from ramen_rate
where Brand = 'Nissin' and Stars >= 4 order by Stars asc").show();

        System.out.println("6. All not Nissin ramen from USA with rate equal
5:");
        spark.sql("select Brand, Variety, Style from ramen_rate where Brand
<> 'Nissin' and Stars = 5 and Country = 'USA'").show();

        System.out.println("7. Percent of pack, bowl, cup and other types of
ramen's style:");
        spark.sql("select PackAmount*100/TotalAmount as PackPercent,
```

```

BowlAmount*100/TotalAmount as BowlPercent, " +
    "CupAmount*100/TotalAmount as CupPercent, 100-
(PackAmount+BowlAmount+CupAmount)*100/TotalAmount as OtherPercent from " +
"(select sum(case when Style = 'Pack' then 1 else 0 end) as
PackAmount, " +
    "sum(case when Style = 'Bowl' then 1 else 0 end) as
BowlAmount, " +
    "sum(case when Style = 'Cup' then 1 else 0 end) as CupAmount,
" +
    "count(*) as TotalAmount from ramen_rate) as RR").show();

System.out.println("8. Other types of ramen's style except pack, cup
and bowl:");
spark.sql("select distinct Style as OtherStyles from ramen_rate where
Style not in ('Pack', 'Bowl', 'Cup')").show();

System.out.println("9. Average rate of ramen by brands:");
spark.sql("select Brand, avg(Stars) as StarsAvg from ramen_rate group
by Brand order by StarsAvg desc").show();

System.out.println("10. Counts of brands in each country:");
spark.sql("select Country, count(Brand) as BrandCount from ramen_rate
group by Country").show();

spark.stop();

}

}

```

Работа программы представлена на рисунках 1-2.

1. All dataset:

Review #	Brand	Variety	Style	Country	Stars	Top Ten
2580	New Touch	T's Restaurant Ta...	Cup	Japan	3.75	NULL
2579	Just Way	Noodles Spicy Hot...	Pack	Taiwan	1	NULL
2578	Nissin	Cup Noodles Chick...	Cup	USA	2.25	NULL
2577	Wei Lih	66E Ramen Snack T...	Pack	Taiwan	2.75	NULL
2576	Ching's Secret	Singapore Curry	Pack	India	3.75	NULL
2575	Samyang Foods	Kimchi song Song ...	Pack	South Korea	4.75	NULL
2574	Acecook	Spice Deli Tantan...	Cup	Japan	4	NULL
2573	Ikeda Shoku	Nabeyaki Kitsune ...	Tray	Japan	3.75	NULL
2572	Ripe'n'Dry	Hokkaido Soy Sauc...	Pack	Japan	0.25	NULL
2571	KOKA	The Original Spic...	Pack	Singapore	2.5	NULL
2570	Tao Kae Noi	Creamy tom Yum Ku...	Pack	Thailand	5	NULL
2569	Yamachan	Yokohama Tonkotsu...	Pack	USA	5	NULL
2568	Nongshim	Mr. Bibim Stir-Fr...	Pack	South Korea	4.25	NULL
2567	Nissin	DeKa Buto Kimchi ...	Bowl	Japan	4.5	NULL
2566	Nissin	Demae Ramen Bar N...	Pack	Hong Kong	5	NULL
2565	KOKA	Mushroom Flavour ...	Cup	Singapore	3.5	NULL
2564	TRDP	Mario Masala Noodles	Pack	India	3.75	NULL
2563	Yamachan	Tokyo Shoyu Ramen	Pack	USA	5	NULL
2562	Binh Tay	Mi Hai Cua	Pack	Vietnam	4	NULL
2561	Paldo	Premium Gomtang	Pack	South Korea	4	NULL

only showing top 20 rows

2. Count of ramen's brand:

BrandCount
355

Рисунок 1 – Работа программы

```

3. All ramen from Japan with kimchi:
+-----+-----+-----+
|      Brand|      Variety|Style|Stars|
+-----+-----+-----+
|    Nissin|Deka Buto Kimchi ...| Bowl| 4.5|
|    Nissin|Kimchi Chikin Ram...| Bowl| 5.00|
|New Touch|      Kimchi Noodle| Bowl| 3.5|
+-----+-----+-----+

4. Average stars rate of all ramen:
+-----+
|      StarsAvg|
+-----+
|3.654675979821498|
+-----+

```

Рисунок 2 – Работа программы

```

5. All Nissin ramen with rate above 4:
+-----+-----+-----+
|      Variety|Style|  Country|Stars|
+-----+-----+-----+
|Demae Ramen Spicy...| Pack|Hong Kong| 4|
|Instant Noodles M...| Pack| Thailand| 4|
|Cup Noodles Spicy...| Cup|   USA| 4|
|Demae Ramen Tokyo...| Pack| Germany| 4|
|  Cup Noodles Seafood| Cup|  China| 4|
|Cup Noodles BIG X...| Cup|Hong Kong| 4|
|Cup Noodles Tom Y...| Cup|Hong Kong| 4|
|Maxi Sopa Nissin ...| Cup| Mexico| 4|
|Korean Kimchi Fla...| Cup|Hong Kong| 4|
|Cup Noodles Sopa ...| Cup| Mexico| 4|
|Demae Iccho Red H...| Cup|Hong Kong| 4|
|Sopa Nissin Sabor...| Pack| Mexico| 4|
|Demae Iccho Spicy...| Bowl|Hong Kong| 4|
|Maxi Sopa Nissin ...| Cup| Mexico| 4|
|Disney Cuties Ins...| Cup| Thailand| 4|
|Cup Noodles Sopa ...| Cup| Mexico| 4|
|Demae Iccho Spicy...| Pack|Hong Kong| 4|
|Demae Iccho Tokyo...| Cup|Hong Kong| 4|
|Vietnamese Chicke...| Cup|Hong Kong| 4|
|Cup Noodles Black...| Cup|Hong Kong| 4|
+-----+-----+-----+
only showing top 20 rows

```

Рисунок 3 – Работа программы

```

6. All not Nissin ramen from USA with rate equal 5:
+-----+-----+-----+
|      Brand|      Variety|Style|
+-----+-----+-----+
|      Yamachan|Yokohama Tonkotsu...| Pack|
|      Yamachan|  Tokyo Shoyu Ramen| Pack|
|Jackpot Teriyaki|      Beef Ramen| Pack|
|      Daifuku|  Katsuo Bowl Udon| Bowl|
|  Dream Kitchen|Curry Flavour Ins...| Cup|
|Nongshim|  Shin Noodle Soup| Cup|
|Maruchan|Instant Lunch Chi...| Cup|
|Nongshim|Kimchi Flavor Noo...| Cup|
|Maruchan|Bowl Taste Of Asi...| Bowl|
|Nongshim|Arrabiata Rice Bu...| Pack|
|Sapporo Ichiban|Tonkotsu Ramen Ar...| Pack|
|Nongshim|Soon Veggie Noodl...| Cup|
|Nongshim|Shin Black Spicy ...| Pack|
|Nongshim|Spicy Tonkotsu No...| Pack|
|Annie Chun's|Ramen House Thai ...| Pack|
|Nongshim|Shin Black Spicy ...| Cup|
|Komforte ChockoLates|      Savory Ramen | Bar|
|      Myojjo|Japanese Style No...| Pack|
|      Sun Noodle|  Ramen Shoyu Flavor| Tray|
|      Sun Noodle|  Ramen Miso Flavor| Tray|
+-----+-----+-----+
only showing top 20 rows

7. Percent of pack, bowl, cup and other types of ramen's style:
+-----+-----+-----+-----+
|      PackPercent|      BowlPercent|      CupPercent|      OtherPercent|
+-----+-----+-----+-----+
|59.249226006191954|18.61455108359133|17.41486068111455|4.721362229102169|
+-----+-----+-----+-----+

```

Рисунок 4 – Работа программы

```

8. Other types of ramen's style except pack, cup and bowl:
+-----+
|OtherStyles|
+-----+
|      Bar|
|      Box|
|      Tray|
|      Can|
+-----+

9. Average rate of ramen by brands:
+-----+
|      Brand|StarsAvg|
+-----+
|      ORee Garden| 5.0|
|      Song Hak| 5.0|
|      Prima| 5.0|
|      Nyor Nyar| 5.0|
|      Foodmon| 5.0|
|      Daddy| 5.0|
|      MyOri| 5.0|
|      Higashi| 5.0|
|      Kimura| 5.0|
|      Torishi| 5.0|
|      Plats Du Chef| 5.0|
|      Peyang| 5.0|
|      Prima Taste| 5.0|
|      Seven & I| 5.0|
|      Jackpot Teriyaki| 5.0|
|      Takamori| 5.0|
|      The Bridge| 5.0|
|      Patanjali| 5.0|
|The Ramen Rater S...| 5.0|
|      Daifuku| 5.0|
+-----+
only showing top 20 rows

```

Рисунок 5 – Работа программы

```

10. Counts of brands in each country:
+-----+
|      Country|BrandCount|
+-----+
|      Sweden| 3|
| Philippines| 47|
| Singapore| 109|
| Malaysia| 156|
| Fiji| 4|
| Germany| 27|
| Cambodia| 5|
| Taiwan| 224|
| NULL| 0|
| Finland| 3|
| Ghana| 2|
| Myanmar| 14|
| India| 31|
| China| 169|
|United States| 1|
| Nigeria| 1|
| Bangladesh| 7|
| Thailand| 191|
| Hong Kong| 137|
| USA| 323|
+-----+
only showing top 20 rows

```

Рисунок 6 – Работа программы

Вывод: Был изучен Spark в языке программирования Java.