



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

Название: Spark

Дисциплина: Языки программирования для работы с большими  
данными

Студент ИУ6-23М  
(Группа)

Д.В. Пешков  
(Подпись, дата) (И.О. Фамилия)

Преподаватель

П.В. Степанов  
(Подпись, дата) (И.О. Фамилия)

Москва, 2024

## Цель работы

Целью лабораторной работы является изучение работы со Spark в языке Kotlin.

## Задание

Выбрать любой датасет на kaggle.com.

Сделать 10 выборок данных по выбранной предметной области.

## Ход работы

Был выбран датасет <https://www.kaggle.com/dwdkills/russian-demography>. Была составлена программа для выполнения требуемой задачи. Был настроен CI для проверки сборки кода.

Фрагмент программного кода приведен в листинге 1.

Листинг 1 — Программный код

```
package org.lab10

/**
 * 1) Выбрать любой датасет на kaggle.com
 * 2) Сделать 10 выборок данных по выбранной предметной
 * области
 */

import org.apache.spark.sql.Session

fun main() {
    val spark =
    SparkSession.builder().appName("Lab10").create
    val russian_demography =
    spark.read().format("csv").option("header",
    "true").load("src/main/resources/russian_demography.csv")

    russian_demography.createOrReplaceTempView("russian_demograp
    hy")

    spark.sql("SELECT * FROM russian_demography").show()
    spark.sql("SELECT AVG(birth_rate) FROM
    russian_demography WHERE year < 2000").show()
    spark.sql("SELECT AVG(death_rate) FROM
    russian_demography WHERE year < 2000").show()
    spark.sql("SELECT * FROM russian_demography WHERE
```

```

birth_rate > 14").show()
    spark.sql("SELECT * FROM russian_demography WHERE
birth_rate > 10").show()
    spark.sql("SELECT          AVG(urbanization)          FROM
russian_demography WHERE year >= 2000").show()
    spark.sql("SELECT * FROM russian_demography WHERE
birth_rate > 14 AND birth_rate > 10").show()
    spark.sql("SELECT          AVG(urbanization)          FROM
russian_demography WHERE year < 2000").show()
    spark.sql("SELECT          AVG(birth_rate)             FROM
russian_demography WHERE year >= 2000").show()
    spark.sql("SELECT          AVG(death_rate)             FROM
russian_demography WHERE year >= 2000").show()
    spark.sql("SELECT * FROM russian_demography WHERE region
= 'Kaliningrad Oblast']").show()
    spark.stop()
}

```

Полные программные коды программ доступны в репозитории:  
<https://github.com/DPeshkoff/PLfBD>.

### **Вывод**

В ходе выполнения данной лабораторной работы были получены требуемые компетенции. Была изучена работа со Spark в языке Kotlin.