



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 Интеллектуальные системы анализа,  
обработки и интерпретации больших данных

**О Т Ч Е Т**

по лабораторной работе № 1 0

**Вариант 18**

**Название:** Scala Spark

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

Студент

ИУ6-23М

(Группа)

\_\_\_\_\_  
(Подпись, дата)

Д.Н. Хныкин

(И.О. Фамилия)

Преподаватель

\_\_\_\_\_  
(Подпись, дата)

П.В. Степанов

(И.О. Фамилия)

Москва, 2022

## Цель работы

Получение навыков работы со Scala Spark.

## Выполнение

### Задание 1:

1. Задана Выбрать любой датасет на kaggle.com
2. Сделать 10 выборки данных на ваше усмотрение

Листинг выполнения подзадачи 1

#### #SparkSessinTest.scala

```
import org.apache.spark.sql.SparkSession
object SparkSessionTest extends App{
  val spark = SparkSession.builder()
    .master("local[1]")
    .appName("SparkByExample")
    .getOrCreate();
  spark.sparkContext.setLogLevel("ERROR")

  val datafile = spark.read
    .format("com.databricks.spark.csv")
    .option("header", true)
    .load("data/russian_passenger_air_service_2.csv")

  datafile.show()
  datafile.createOrReplaceTempView("passengers")
  spark.sql("select * from passengers").show()
  spark.sql("SELECT COUNT(*) FROM passengers").show()
  spark.sql("SELECT * from passengers WHERE passengers.Year = 2019").show()
  spark.sql("SELECT * from passengers WHERE passengers.January > 10000").show()
  spark.sql("SELECT `Airport name` from passengers").show()
  spark.sql("SELECT * from passengers WHERE February BETWEEN 1000 AND 5000").show()
  spark.sql("SELECT * from passengers WHERE passengers.August >= 500000").show()
  spark.sql("SELECT `Airport coordinates` from passengers").show()
  spark.sql("SELECT `Whole year` from passengers").show()
  spark.sql("SELECT `Year` from passengers").show()
}
```

## Ссылка на программное решение

Программное решение представлено в репозитории распределённой системы управления версиями Git:

[https://github.com/Aranei99/labsJava/tree/main/LB10\\_3](https://github.com/Aranei99/labsJava/tree/main/LB10_3)

## Вывод

При выполнении лабораторной работы были получены навыки работы со Scala Spark.