Файл Правка Поиск Вид Кодировки Синтаксисы Опции Инструменты Макросы Запуск Плагины Вкладки ?

d2.scala

```scala
/*
chcp 65001 && spark-shell -i \spark\d2.scala --conf "spark.driver.extraJavaOptions=-Dfile.encoding=utf-8"
*/
import org.apache.spark.internal.Logging
import org.apache.spark.sql.functions.{col, collect_list, concat_ws}
import org.apache.spark.sql.{DataFrame, SparkSession}
import org.apache.spark.sql.functions._
import org.apache.spark.sql.expressions.Window
val t1 = System.currentTimeMillis()
if(1==1){
var df1 = spark.read.option("delimiter",",")
        .option("header", "true")
        //.option("encoding", "windows-1251")
        .csv("/spark/fifa_s2.csv")
        df1=df1
        .withColumn("ID",col("ID").cast("int"))
        .withColumn("Age",col("Age").cast("int"))
        .withColumn("Overall",col("Overall").cast("int"))
        .withColumn("Value",col("Value").cast("int"))
        .withColumn("Wage",col("Wage").cast("int"))
        .withColumn("International Reputation",col("International Reputation").cast("int"))
        .withColumn("Skill Moves",col("Skill Moves").cast("int"))
        .withColumn("Joined",col("Joined").cast("int"))
        .withColumn("Weight",col("Weight").cast("int"))
        .withColumn("Height",col("Height").cast("float"))
        .withColumn("Release Clause",col("Release Clause").cast("float"))
        .withColumn("Name", lower(col("Name")))                  // Перевод в нижний регистр
        .withColumn("Nationality", lower(col("Nationality")))    // Перевод в нижний регистр
        .withColumn("Club", lower(col("Club")))                  // Перевод в нижний регистр
        .withColumn("Preferred Foot", lower(col("Preferred Foot"))) // Перевод в нижний регистр
        .withColumn("Position", lower(col("Position")))          // Перевод в нижний регистр

        df1 = df1.drop("Potential")                // Удаляем дублирующуся колонку
        df1 = df1.dropDuplicates()                 // Убираются дубли строк
        df1 = df1.withColumn("Category", lit(""))   // Добавляем новую колонку "Категория футболиста"

        df1.write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
        .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketl2a")
        .mode("overwrite").save()

        // Разбитие колонки Категория по группам до 20, от 20 до 30, от 30 до 36 и старше 36
        import java.sql._;
        def sqlexecute(sql: String) = {
            var conn: Connection = null;
            var stmt: Statement = null;
            try {
                Class.forName("com.mysql.cj.jdbc.Driver");
                conn = DriverManager.getConnection("jdbc:mysql://localhost:3306/spark?user=root&password=root");
                stmt = conn.createStatement();
                stmt.executeUpdate(sql);
                println(sql+" complete");
            } catch {
                case e: Exception => println ("exception caught" + e);
            }
        }
        sqlexecute("""UPDATE spark.tasketl2a SET Category = (CASE
                            WHEN Age<='19' THEN 'ДО 20 лет'
                            WHEN Age BETWEEN '20' AND '29' THEN 'С 20 ДО 30 лет'
                            WHEN Age BETWEEN '30' AND '35' THEN 'С 30 ДО 36 лет'
                            ELSE 'СТАРШЕ 36 лет'
                        END)""")
```

```scala
                                        END)""")

        df1.show()

        // Количество футболистов в каждой категории
        val df_load = spark.read.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
        .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketl2a").load()
        val df_distinct = df_load.select("Name","Category").distinct()
        val df_Category = df_distinct.groupBy("Category").count()
        df_Category.show
    }
    val s0 = (System.currentTimeMillis() - t1)/1000
    val s = s0 % 60
    val m = (s0/60) % 60
    val h = (s0/60/60) % 24
    println("%02d:%02d:%02d".format(h, m, s))
    System.exit(0)
```

Java source file

```
Microsoft Windows [Version 10.0.19045.4412]
(c) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.

C:\Users\Dom>chcp 65001 && spark-shell -i \spark\d2.scala --conf "spark.driver.extraJavaOptions=-Dfile.encoding=utf-8"
Active code page: 65001
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://DESKTOP-TIN8UGG:4040
Spark context available as 'sc' (master = local[*], app id = local-1716048089903).
Spark session available as 'spark'.
UPDATE spark.tasketl2a SET Category = (CASE
                                        WHEN Age<='19' THEN 'ДО 20 лет'
                                        WHEN Age BETWEEN '20' AND '29' THEN 'С 20 ДО 30 лет'
                                        WHEN Age BETWEEN '30' AND '35' THEN 'С 30 ДО 36 лет'
                                        ELSE 'СТАРШРЕ 36 лет'
                                      END) complete
```

ID	Name	Age	Nationality	Overall	Club	Value	Wage	Preferred Foot	International Reputation	Skill Moves	Position	Joined	Contract Valid Until	Height	Weight	Release Clause	Category
183907	j. boateng	29	germany	85	fc bayern münchen	30000	115	right	4	2	rcb	2011	01/01/2021	6.3333335	198	49500.0	
188567	p. aubameyang	29	gabon	88	arsenal	59000	265	right	3	4	lm	2018	01/01/2021	6.1666665	176	112100.0	
189242	coutinho	26	brazil	88	fc barcelona	69500	340	right	3	5	lw	2018	01/01/2023	5.6666665	150	147700.0	
199347	t. beermann	27	germany	69	1. fc heidenheim ...	1000	6	left	1	2	lcb	2013	01/01/2020	6.1666665	187	1600.0	
203795	b. romo	29	chile	67	cd antofagasta	600	2	right	1	2	lcb	2017	01/01/2018	6.0833335	190	810.0	
204213	i. latovlevici	32	romania	71	bursaspor	1300	11	left	1	3	lm	2018	01/01/2019	5.75	163	2600.0	
208156	g. piscitella	25	italy	62	carpi	350	1	right	1	3	lw	2018	01/01/2021	5.9166665	152	569.0	
215785	k. baldé	23	senegal	80	inter	20000	65	right	2	4	rm	2016	30/06/2019	5.9166665	176	4585.061	
216742	m. mevlja	28	slovenia	75	null	null	0	right	1	2	lcb	2016	null	6.25	183	4585.061	
223283	f. diedhiou	25	senegal	73	bristol city	5000	35	right	1	3	ls	2017	01/01/2021	6.1666665	176	9900.0	
169721	m. richards	30	england	68	aston villa	675	8	right	2	2	cb	2015	01/01/2019	5.9166665	183	1300.0	
176944	m. fellaini	30	belgium	79	manchester united	11500	120	right	3	3	cm	2013	01/01/2020	6.3333335	187	21300.0	
190941	l. hrádecký	28	finland	84	bayer 04 leverkusen	null	74	right	2	1	gk	2018	01/01/2023	6.3333335	187	37400.0	
193133	j. stryger larsen	27	denmark	73	udinese	3900	19	right	1	3	lb	2017	01/01/2022	6.0	181	7300.0	
193348	x. shaqiri	26	switzerland	81	liverpool	19000	130	left	3	5	cam	2018	01/01/2023	5.5833335	159	36100.0	
210489	l. schaub	23	austria	75	1. fc köln	9000	18	left	1	3	rm	2018	01/01/2022	5.8333335	154	14200.0	
211177	a. traustason	25	iceland	69	malmö ff	1300	4	right	1	3	lm	2018	01/01/2021	5.9166665	170	1800.0	
212783	o. gaber	26	egypt	68	null	null	0	right	1	2	rb	2016	null	5.75	157	4585.061	
219091	j. fuentes	23	chile	69	cd o'higgins	1400	3	right	1	2	lcb	2013	01/01/2023	5.75	159	2200.0	
220051	m. miloš	25	croatia	69	cd aves	1100	5	right	1	3	rb	2018	01/01/2022	5.75	154	2400.0	

```
only showing top 20 rows
```

Category	count
ДО 20 лет	269
СТАРШРЕ 36 лет	23
С 30 ДО 36 лет	457
С 20 ДО 30 лет	1633

```
00:00:06

C:\Users\Dom>
```

Файл Редактировать Поиск Запрос Инструменты Переход Помощь

Unnamed | База данных: spark | Таблица: tasketl2a | Данные | Запрос* |

spark.tasketl2a: 2 397 строк (приблизительно), показано 1 000   Далее   Показать все   Сортировка

#	ID	Name	Age	Nationality	Overall	Club	Value	Wage	Preferred Foot	International Reputation	Skill Moves	Position	Joined	Contract Valid Until	Height	Weight	Release Clause	Category
1	183 907	j. boateng	29	germany	85	fc bayern münchen	30 000	115	right	4	2	rcb	2 011	01/01/2021	6,33333	198	49 500	С 20 до 30 лет
2	188 567	p. aubameyang	29	gabon	88	arsenal	59 000	265	right	3	4	lm	2 018	01/01/2021	6,16667	176	112 100	С 20 до 30 лет
3	189 242	coutinho	26	brazil	88	fc barcelona	69 500	340	right	3	5	lw	2 018	01/01/2023	5,66667	150	147 700	С 20 до 30 лет
4	199 347	t. beermann	27	germany	69	1. fc heidenheim 1846	1 000	6	left	1	2	lcb	2 013	01/01/2020	6,16667	187	1 600	С 20 до 30 лет
5	203 795	b. romo	29	chile	67	cd antofagasta	600	2	right	1	2	lcb	2 017	01/01/2018	6,08333	190	810	С 20 до 30 лет
6	204 213	i. latovlevici	32	romania	71	bursaspor	1 300	11	left	1	3	lm	2 018	01/01/2019	5,75	163	2 600	С 30 ДО 36 лет
7	208 156	g. piscitella	25	italy	62	carpi	350	1	right	1	3	lw	2 018	01/01/2021	5,91667	152	569	С 20 до 30 лет
8	215 785	k. baldé	23	senegal	80	inter	20 000	65	right	2	4	rm	2 016	30/06/2019	5,91667	176	4 585,06	С 20 до 30 лет
9	216 742	m. mevlja	28	slovenia	75	(NULL)	(NULL)	0	right	1	2	lcb	2 016	(NULL)	6,25	183	4 585,06	С 20 до 30 лет
10	223 283	f. diedhiou	25	senegal	73	bristol city	5 000	35	right	1	3	ls	2 017	01/01/2020	6,16667	176	9 900	С 20 до 30 лет
11	169 721	m. richards	30	england	68	aston villa	675	8	right	2	2	cb	2 015	01/01/2019	5,91667	183	1 300	С 30 ДО 36 лет
12	176 944	m. fellaini	30	belgium	79	manchester united	11 500	120	right	3	3	cm	2 013	01/01/2020	6,33333	187	21 300	С 30 ДО 36 лет
13	190 941	l. hrádecký	28	finland	84	bayer 04 leverkusen	(NULL)	74	right	3	1	gk	2 018	01/01/2023	6,33333	187	37 400	С 20 до 30 лет
14	193 133	j. stryger larsen	27	denmark	73	udinese	3 900	19	right	1	3	lb	2 017	01/01/2022	6	181	7 300	С 20 до 30 лет
15	193 348	x. shaqiri	26	switzerland	81	liverpool	19 000	130	left	3	5	cam	2 018	01/01/2023	5,58333	159	36 100	С 20 до 30 лет
16	210 489	l. schaub	23	austria	75	1. fc köln	9 000	18	left	1	3	rm	2 018	01/01/2022	5,83333	154	14 200	С 20 до 30 лет
17	211 177	a. traustason	25	iceland	69	malmö ff	1 300	4	right	1	3	lm	2 018	01/01/2021	5,91667	170	1 800	С 20 до 30 лет
18	212 783	o. gaber	26	egypt	68	(NULL)	(NULL)	0	right	1	2	rb	2 016	(NULL)	5,75	157	4 585,06	С 20 до 30 лет
19	219 091	j. fuentes	23	chile	69	cd o'higgins	1 400	3	right	1	2	lcb	2 013	01/01/2023	5,75	159	2 200	С 20 до 30 лет
20	220 051	m. miloš	25	croatia	69	cd aves	1 100	5	right	1	3	rb	2 018	01/01/2022	5,75	154	2 400	С 20 до 30 лет
21	224 165	j. huentelaf	29	chile	68	cd universidad de concepción	875	3	right	1	3	rw	2 017	01/01/2018	5,83333	161	1 200	С 20 до 30 лет
22	237 832	i. meite	21	england	58	cardiff city	210	5	left	1	2	st	2 017	01/01/2019	6,16667	159	462	С 20 до 30 лет
23	245 978	b. tsonev	23	bulgaria	65	(NULL)	(NULL)	0	right	1	3	cam	2 016	(NULL)	5,58333	143	4 585,06	С 20 до 30 лет
24	158 685	defendi	32	brazil	71	cd aves	1 400	8	left	1	2	lb	2 017	01/01/2021	6,25	183	2 900	С 30 ДО 36 лет
25	169 345	m. suttner	31	austria	73	brighton & hove albion	2 600	33	left	1	2	lb	2 017	01/01/2019	5,83333	161	5 000	С 30 ДО 36 лет
26	181 795	f. kudryashov	31	russia	73	(NULL)	(NULL)	0	left	1	2	lcb	2 016	(NULL)	5,91667	179	4 585,06	С 30 ДО 36 лет
27	190 430	j. bennett	28	england	72	cardiff city	2 500	30	left	1	3	lb	2 016	01/01/2021	5,83333	163	4 900	С 20 до 30 лет
28	190 982	alberto perea	27	spain	68	cádiz cf	950	4	right	1	3	cam	2 017	01/01/2020	5,75	152	1 500	С 20 до 30 лет
29	210 617	samu castillejo	23	spain	82	milan	29 000	94	left	2	4	rm	2 018	01/01/2023	6	148	51 500	С 20 до 30 лет
30	215 314	carlos gutiérrez	26	spain	70	cd numancia	1 700	6	right	1	2	lcb	2 016	01/01/2020	6,33333	181	2 900	С 20 до 30 лет
31	215 330	j. correa	23	argentina	78	lazio	(NULL)	42	right	1	4	lm	2 018	01/01/2023	6,16667	170	24 900	С 20 до 30 лет
32	215 965	c. strandberg	22	sweden	72	malmö ff	4 300	6	left	1	3	st	2 017	01/01/2021	6,08333	187	6 500	С 20 до 30 лет
33	221 491	n. elvedi	21	switzerland	77	borussia mönchengladbach	11 000	19	right	2	3	rcb	2 015	01/01/2021	6,16667	172	20 900	С 20 до 30 лет
34	223 608	eraso	28	spain	76	cd leganés	8 000	29	right	1	3	cam	2 017	01/01/2020	5,91667	157	16 800	С 20 до 30 лет
35	228 019	cao ziheng	23	china pr	58	changchun yatai fc	180	2	right	1	2	st	2 014	01/01/2020	5,75	148	423	С 20 до 30 лет
36	179 846	s. khedira	31	germany	85	juventus	33 000	160	right	3	2	rcm	2 015	01/01/2020	6,16667	198	54 500	С 30 ДО 36 лет
37	192 073	a. bertolacci	27	italy	77	milan	9 000	75	left	2	4	cm	2 015	01/01/2019	5,83333	165	15 300	С 20 до 30 лет
38	197 242	g. krychowiak	28	poland	78	lokomotiv moscow	9 500	1	right	3	3	lcm	2 016	01/01/2021	6,08333	187	20 700	С 20 до 30 лет
39	197 781	isco	26	spain	88	real madrid	73 500	315	right	3	4	lw	2 013	01/01/2022	5,75	174	156 200	С 20 до 30 лет
40	197 837	d. boyata	27	belgium	75	celtic	6 500	42	right	1	2	rcb	2 015	01/01/2019	6,16667	190	12 200	С 20 до 30 лет
41	199 577	sergi gómez	26	spain	78	sevilla fc	11 500	17	right	1	2	lcb	2 018	01/01/2023	6,08333	170	25 000	С 20 до 30 лет
42	208 589	k. pusch	25	germany	67	1. fc heidenheim 1846	950	5	left	1	2	cam	2 017	01/01/2020	5,91667	168	1 500	С 20 до 30 лет
43	208 777	óliver torres	23	spain	78	fc porto	13 500	13	right	3	3	cm	2 017	01/01/2023	5,75	139	28 000	С 20 до 30 лет
44	224 470	he chao	23	china pr	66	changchun yatai fc	725	3	right	1	2	lcm	2 016	01/01/2021	5,83333	170	1 600	С 20 до 30 лет
45	236 588	q. bena	20	france	62	chamois niortais football club	425	1	right	1	2	rb	2 017	01/01/2020	6,08333	174	829	С 20 до 30 лет
46	241 852	m. diaby	18	france	70	paris saint-germain	3 700	19	left	1	3	rw	2 018	01/01/2022	5,58333	143	9 100	ДО 20 лет

Фильтр Регулярное выражение

```
2305  SELECT tc.CONSTRAINT_NAME, cc.CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` AS cc, `information_schema`.`TABLE_CONSTRAINTS` AS tc WHERE tc.CONSTRAINT_SCHEMA='spark' AND tc.TABLE_NAME='tasketl2a' AND tc.CONSTRAINT_TYPE='CHECK' AND tc.CONSTRAINT_SCHEMA=cc.CONSTRAIN
2306  /* Открытие сеанса «Unnamed» */
2307  SELECT * FROM `spark`.`tasketl2a` LIMIT 1000;
2308  SHOW TABLE STATUS LIKE 'tasketl2a';
```

r1 000 : c19   Подключено: 10:29 h   MySQL 8.4.0   Время работы: 10:29 h   Серверное время: