

# Nike test

November 2, 2020

```
[132]: import org.apache.spark.ml.stat.Summarizer
import org.apache.spark.sql.functions.udf
```

```
[132]: import org.apache.spark.ml.stat.Summarizer
import org.apache.spark.sql.functions.udf
```

## 1 Constants

```
[11]: val CALENDAR_PATH = "./data/calendar.csv"
val PRODUCT_PATH = "./data/product.csv"
val SALES_PATH = "./data/sales.csv"
val STORE_PATH = "./data/store.csv"
```

```
[11]: CALENDAR_PATH: String = ./data/calendar.csv
CONSUMPTION_PATH: String = ./data/consumption.csv
PRODUCT_PATH: String = ./data/product.csv
SALES_PATH: String = ./data/sales.csv
STORE_PATH: String = ./data/store.csv
```

## 2 Loading data into dataframes

### 2.1 Calendar

```
[259]: val calendarDf = spark.read.options(Map("header" -> "true")).csv(CALENDAR_PATH)
calendarDf.registerTempTable("calendar")
calendarDf.orderBy($"weeknumberofseason", $"datecalendarday").show
```

```
+-----+-----+-----+-----+
|datekey|datecalendarday|datecalendaryear|weeknumberofseason|
+-----+-----+-----+-----+
|  5230|           23|         2018|                1|
|  5321|           23|         2018|                1|
|  5139|           24|         2018|                1|
|  5231|           24|         2018|                1|
|  5322|           24|         2018|                1|
```

	5232	25	2018	1
	5048	25	2018	1
	5323	25	2018	1
	5140	25	2018	1
	5324	26	2018	1
	5141	26	2018	1
	5049	26	2018	1
	5233	26	2018	1
	5142	27	2018	1
	5050	27	2018	1
	5234	27	2018	1
	5325	27	2018	1
	5143	28	2018	1
	5051	28	2018	1
	5235	28	2018	1

+-----+-----+-----+-----+

only showing top 20 rows

[259]: calendarDf: org.apache.spark.sql.DataFrame = [datekey: string, datecalendarday: string ... 2 more fields]

## 2.2 Product

```
[37]: val productDf = spark.read.options(Map("header" -> "true")).csv(PRODUCT_PATH)
      productDf.registerTempTable("product");
      productDf.show
```

+-----+-----+-----+-----+
productid division gender  category
+-----+-----+-----+-----+
00567228914507  APPAREL  KIDS  CRICKET
00565177969035  FOOTWEAR  MENS  COLLECTIONS
00565177969059  APPAREL  KIDS  COLLECTIONS
00567228317407  APPAREL  MENS  HOCKEY
00565177969042  APPAREL  WOMENS  COLLECTIONS
00564802779339  APPAREL  MENS  KABBADI
00564802779193  APPAREL  MENS  BASEBALL
00565177969080  FOOTWEAR  MENS  COLLECTIONS
00564802779230  APPAREL  MENS  GOLF
00567228914452  APPAREL  KIDS  CRICKET
00562801460043  APPAREL  MENS  SPINE
00568407032141  APPAREL  MENS  COLLECTIONS
00562801651212  APPAREL  MENS  COLLECTIONS
00091203971983  APPAREL  WOMENS  TENNIS
00564802779179  APPAREL  UNISEX  SWIMMING
00564802779209  APPAREL  MENS  DART

```
|00567228914476| APPAREL|UNISEX|    CRICKET|
|00565177969076| APPAREL|  MENS|COLLECTIONS|
|00564802779346| APPAREL|  MENS|      GOLF|
|00566915022258|FOOTWEAR|  KIDS| BASKETBALL|
+-----+-----+-----+-----+
```

```
[37]: productDf: org.apache.spark.sql.DataFrame = [productid: string, division: string
... 2 more fields]
```

## 2.3 Sales

```
[36]: val salesDf = spark.read.options(Map("header"->"true")).csv(SALES_PATH)
salesDf.registerTempTable("sales")
salesDf.show
```

```
+-----+-----+-----+-----+-----+-----+
|saleId|netSales|salesUnits|storeId|dateId|    productId|
+-----+-----+-----+-----+-----+-----+
|    1|    300.24|         5|    409|  4965|00567228914507|
|    2|    300.24|         5|    409|  4965|00567228914507|
|    3|    300.24|         5|    409|  4965|00567228914507|
|    4|    300.24|         5|    409|  4965|00567228914507|
|    5|    300.24|         5|    411|  4965|00567228914507|
|    6|    300.24|         5|    411|  4965|00567228914507|
|    7|    300.24|         5|    409|  4965|00567228914507|
|    8|    300.24|         5|    411|  4965|00567228914507|
|    9|    300.24|         5|    409|  4965|00567228914507|
|   10|    300.24|         5|    409|  4965|00567228914507|
|   11|    300.24|         5|    409|  4965|00567228914507|
|   12|    300.24|         5|    411|  4965|00567228914507|
|   13|    300.24|         5|    409|  4965|00567228914507|
|   14|    300.24|         5|    409|  4965|00567228914507|
|   15|    300.24|         5|    409|  4965|00567228914507|
|   16|    300.24|         5|    409|  4965|00567228914507|
|   17|    300.24|         2|    409|  4965|00567228914507|
|   18|    300.24|         5|    409|  4965|00567228914507|
|   19|    300.24|         5|    409|  4965|00567228914507|
|   20|    300.24|         5|    409|  4965|00567228914507|
+-----+-----+-----+-----+-----+-----+
```

only showing top 20 rows

```
[36]: salesDf: org.apache.spark.sql.DataFrame = [saleId: string, netSales: string ...
4 more fields]
```

## 2.4 Store

```
[35]: val storeDf = spark.read.options(Map("header" -> "true")).csv(STORE_PATH)
      storeDf.registerTempTable("store")
      storeDf.show
```

```
+-----+-----+-----+
|storeid|channel|country|
+-----+-----+-----+
| 409|Digital|INDIA|
| 410|Digital|CHINA|
| 411|Digital|AUSTRIA|
| 412|Store|BELGIUM|
| 413|Digital|DENMARK|
| 414|Digital|FINLAND|
| 415|Store|FRANCE|
| 416|Digital|GERMANY|
| 417|Digital|IRELAND|
| 418|Digital|ITALY|
| 419|Digital|LUXEMBOURG|
| 420|Store|NETHERLANDS|
| 421|Digital|SPAIN|
| 422|Digital|SWEDEN|
| 423|Digital|UK|
| 424|Store|PORTUGAL|
| 425|Digital|GREECE|
| 426|Digital|SLOVENIA|
| 427|Store|POLAND|
| 428|Digital|HUNGARY|
+-----+-----+-----+
```

```
[35]: storeDf: org.apache.spark.sql.DataFrame = [storeid: string, channel: string ...
      1 more field]
```

## 3 Statistics

```
[28]: calendarDf.describe().show()
```

```
+-----+-----+-----+-----+-----+
+
|summary|datekey|
datecalendarday|datecalendaryear|weeknumberofseason|
+-----+-----+-----+-----+-----+
+
|count|365|365|365|
```

```

365|
|   mean|          5147.0|15.72054794520548|          2018.0|
6.986301369863014|
|  stddev|105.5106629682517|8.808321472985572|
0.0|3.7507990016375086|
|   min|          4965|          1|          2018|
1|
|   max|          5329|          9|          2018|
9|
+-----+-----+-----+-----+-----+
+

```

```
[32]: productDf.groupBy("division", "gender", "category").count.show
```

```

+-----+-----+-----+-----+
|division|gender|  category|count|
+-----+-----+-----+-----+
| APPAREL|  MENS|  BASEBALL|    1|
|FOOTWEAR|  MENS|COLLECTIONS|    2|
| APPAREL|  MENS|   KABBADI|    1|
| APPAREL|  MENS|    SPINE|    1|
|FOOTWEAR| KIDS| BASKETBALL|    1|
| APPAREL|WOMENS|   TENNIS|    1|
| APPAREL| KIDS|COLLECTIONS|    1|
| APPAREL|  MENS|   HOCKEY|    1|
| APPAREL|WOMENS|COLLECTIONS|    1|
| APPAREL|  MENS|    GOLF|    2|
| APPAREL|UNISEX|  SWIMMING|    1|
| APPAREL| KIDS|   CRICKET|    2|
| APPAREL|UNISEX|   CRICKET|    1|
| APPAREL|  MENS|    DART|    1|
| APPAREL|  MENS|COLLECTIONS|    3|
+-----+-----+-----+-----+

```

```
[33]: salesDf.select("netSales", "salesUnits").describe().show()
```

```

+-----+-----+-----+
|summary|          netSales|          salesUnits|
+-----+-----+-----+
|  count|          70|          70|
|   mean| 342.424285714286| 5.214285714285714|
|  stddev|150.7352238877725|1.1150207269517867|
|   min|          200.24|          2|
|   max|          909.24|          8|
+-----+-----+-----+

```

## 4 Join

```
[232]: val totalSalesDf = spark.sqlContext.sql("""
        SELECT p.division,
               p.gender,
               p.category,

               st.channel,
               st.country,

               c.datecalendaryear,
               c.weeknumberofseason,

               s.netSales,
               s.salesUnits
        FROM sales s

        INNER JOIN product p
        ON s.productId = p.productid

        INNER JOIN store st
        ON s.storeId = st.storeid

        INNER JOIN calendar c
        ON s.dateId = c.datekey
        """)
totalSalesDf.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+
--+-----+-----+
|division|gender|category|channel|
country|datecalendaryear|weeknumberofseason|netSales|salesUnits|
+-----+-----+-----+-----+-----+-----+-----+
--+-----+-----+
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|AUSTRIA    |          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|AUSTRIA    |          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
```

```

| APPAREL| KIDS| CRICKET|Digital|AUSTRIA      |          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|AUSTRIA      |          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          2|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
| APPAREL| KIDS| CRICKET|Digital|          INDIA|          2018|
2| 300.24|          5|
+-----+-----+-----+-----+-----+-----+-----+
--+-----+-----+
only showing top 20 rows

```

[232]: totalSalesDf: org.apache.spark.sql.DataFrame = [division: string, gender: string ... 7 more fields]

## 5 UDFs

```

[233]: val createUniqueIdFn = (year: String, channel: String, division: String, gender:
      ↪ String, category: String) => Seq(year, channel, division, gender, category).
      ↪mkString("_")
val createUniqueIdFnUdf = udf(createUniqueIdFn)

```

[233]: createUniqueIdFn: (String, String, String, String, String) => String =  
 \$Lambda\$5132/0x00000008418bec40@62bf993c  
 createUniqueIdFnUdf: org.apache.spark.sql.expressions.UserDefinedFunction = Spar  
 kUserDefinedFunction(\$Lambda\$5132/0x00000008418bec40@62bf993c,StringType,List(So

```
me(class[value[0]: string]), Some(class[value[0]: string]), Some(class[value[0]:
string]), Some(class[value[0]: string]), Some(class[value[0]:
string])),None,true,true)
```

```
[234]: createUniqueIdFn("2018", "1", "2", "3", "4")
```

```
[234]: res179: String = 2018_1_2_3_4
```

```
[239]: val createYearFn = (year: Int) => "RY" + (year%100).toString
      val createYearFnUdf = udf(createYearFn)
```

```
[239]: createYearFn: Int => String = $Lambda$5134/0x000000084199b840@6b299a56
createYearFnUdf: org.apache.spark.sql.expressions.UserDefinedFunction = SparkUse
rDefinedFunction($Lambda$5134/0x000000084199b840@6b299a56,StringType,List(Some(c
lass[value[0]: int])),None,true,true)
```

```
[240]: createYearFn(2020)
```

```
[240]: res183: String = RY20
```

## 6 TRANSFORMATIONS

```
[242]: val totalSalesWithUniqueKeyDf = totalSalesDf.withColumn("year",  
    ↳ createYearFnUdf($"datecalendaryear"))  
    .withColumn("uniqueId",  
    ↳ createUniqueIdFnUdf(  
    ↳     $"year",  
    ↳     $"channel",  
    ↳     $"division",  
    ↳     $"gender",  
    ↳     $"category"  
    ↳ )  
    )  
totalSalesWithUniqueKeyDf.show(false)
```



+-----+-----+-----+-----+-----+-----+-----+-----+													
division gender		category channel		country									
date calendar year		week number of season		netSales salesUnits		year uniqueId							
+-----+-----+-----+-----+-----+-----+-----+-----+													
+-----+-----+-----+-----+-----+-----+-----+-----+													
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	AUSTRIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	AUSTRIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	AUSTRIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital	INDIA	2018	2							
300.24	5	RY18	RY18_Digital_APPAREL_KIDS_CRICKET										
APPAREL	KIDS	CRICKET	Digital										

only showing top 20 rows

```
[242]: totalSalesWithUniqueKeyDf: org.apache.spark.sql.DataFrame = [division: string,
gender: string ... 9 more fields]
```

```
[243]: val totalNetSalesAggDf = totalSalesWithUniqueKeyDf
      .groupBy("uniqueId", "weeknumberofseason", "year", "channel", "division", "gender", "category")
      .agg(
        sum($"netSales").as("netSales"),
        sum($"salesUnits").as("salesUnits"),
      )
totalNetSalesAggDf.show(false)
```

```
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|uniqueId
|weeknumberofseason|year|channel|division|gender|category    |netSales
|salesUnits|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|2
|RY18|Digital|FOOTWEAR|MENS    |COLLECTIONS|3767.3999999999996|55.0    |
|RY18_Digital_APPAREL_KIDS_COLLECTIONS |2
|KIDS    |COLLECTIONS|8192.6999999999999 |111.0    |
|RY18_Digital_APPAREL_KIDS_CRICKET      |2
|KIDS    |CRICKET      |12009.5999999999993|199.0    |
|RY18|Digital|APPAREL
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
```

```
[243]: totalNetSalesAggDf: org.apache.spark.sql.DataFrame = [uniqueId: string,
weeknumberofseason: string ... 7 more fields]
```

```
[244]: totalNetSalesAggDf.registerTempTable("total_agg_sales")
(1 to 53).toDF("weekofyear").registerTempTable("weeks_of_year")
```

```
[260]: val resultPerIdPerWeek = spark.sqlContext.sql("""
      SELECT
        t.uniqueId,
        CONCAT("W", CAST(w.weekofyear AS STRING)) AS weekofyear,
        channel,
        division,
        gender,
```

```

        category,
        IF(t.weeknumberofseason = w.weekofyear, netSales, 0) AS_
↪netSales,
        IF(t.weeknumberofseason = w.weekofyear, salesUnits, 0) AS_
↪salesUnits

    FROM total_agg_sales t, weeks_of_year w
""")

resultPerIdPerWeek.show(false)

```

```

+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+
|uniqueId
|weekofyear|channel|division|gender|category    |netSales      |salesUnits|
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W1          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W2          |Digital|FOOTWEAR|MENS
|COLLECTIONS|3767.3999999999996|55.0     |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W3          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W4          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W5          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W6          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W7          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W8          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W9          |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W10         |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W11         |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W12         |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W13         |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W14         |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |
|RY18_Digital_FOOTWEAR_MENS_COLLECTIONS|W15         |Digital|FOOTWEAR|MENS
|COLLECTIONS|0.0              |0.0      |

```



[illegible]



```
[250]: salesPerWeek: org.apache.spark.sql.DataFrame = [uniqueId: string, channel:
      string ... 5 more fields]
```

uniqueId	channel	division	gender	category
netSales				
salesUnits				







```
[258]: val rows = salesPerWeek.count()  
       salesPerWeek.repartition(rows.toInt).write.mode("overwrite").json("./json/")
```

```
[258]: rows: Long = 3
```

```
[ ]:
```