

Titanic

Titanic visualization assignment.

FINISHED

This assignment shows in a nice way the data set from Titanic passenger list

Took 0 sec. Last updated by anonymous at February 20 2020, 12:38:53 PM.

```
1 %spark
2
3 // import org.apache.commons.io.IOUtils
4 // import java.net.URL
5 // import java.nio.charset.Charset
6 // import org.apache.spark.ml.classification.DecisionTreeClassificationModel;
7 // import org.apache.spark.ml.classification.DecisionTreeClassifier;
8 // import org.apache.spark.ml.feature.StringIndexer;
9 // import org.apache.spark.ml.feature.StringIndexerModel;
10 // import org.apache.spark.ml.feature.VectorAssembler;
11 // import org.apache.spark.mllib.evaluation.MulticlassMetrics;
12 // import org.apache.spark.sql.Session;
13
14 import org.apache.spark.sql.Dataset;
15 import org.apache.spark.sql.Row;
16 import org.apache.spark.sql.functions;
17 import org.apache.spark.sql.types.DataTypes;
18 import org.apache.spark.sql.types.StructType;
19
20 // création du dataframe Spark
21 val df = spark.read.options(Map("inferSchema"->"true","delimiter"->"", "header"->"true"))
22   .csv("/Users/FXSUIRE/Google Drive/_00_Dev/Titanic/input.csv")
23
24 //case class passenger_cls(Survived:Int,Pclass:Int,Name:String,Sex:String,Age:Int,Siblings:Int,ParentAboard:Int,Fare:Float)
25 case class passenger_cls(Survived:Int,Pclass:Int,Name:String,Sex:String,Age:Float,Siblings:Int,ParentAboard:Int)
26
27 // x0: Survived:Int
28 // x1: Pclass:Int
29 // x2: Name:String
30 // x3: Sex:String
```

FINISHED

Titanic

```

31 // x4: Age:Int
32 // x5: Siblings:Int
33 // x6: ParentAboard:Int
34 // x7: Fare:Float
35
36 // val passenger = data1.map(x=>x.split(",")).map(x => passenger_cls(x(0).toInt,x(1).toInt,x(2),x(3),x(4).toInt,x(5).toInt,x(6).toI
37
38 val passenger = data1.map(x=>x.split(",")).map(x => passenger_cls(x(0).toInt,x(1).toInt,x(2),x(3),x(4).toFloat,x(5).toInt,x(6).toI
39
40 passenger.registerTempTable("PASSENGER")
41
42 // val SurvivalAgeClassSexFare = sqlContext.sql("select Survived, Pclass, Sex, Fare from PASSENGER")
43 val SurvivalAgeClassSex = sqlContext.sql("select Survived, Pclass, Sex from PASSENGER")
44
45 SurvivalAgeClassSex.registerTempTable("SMALLPASSTABLE")
46

```

warning: there were two deprecation warnings; re-run with -deprecation for details

```

import sqlContext.implicits._
import org.apache.spark.sql.Dataset
import org.apache.spark.sql.Row
import org.apache.spark.sql.functions
import org.apache.spark.sql.types.DataTypes
import org.apache.spark.sql.types.StructType
df: org.apache.spark.sql.DataFrame = [Survived: int, Pclass: int ... 6 more fields]
defined class passenger_cls
passenger: org.apache.spark.sql.DataFrame = [Survived: int, Pclass: int ... 5 more fields]
SurvivalAgeClassSex: org.apache.spark.sql.DataFrame = [Survived: int, Pclass: int ... 1 more field]

```

Took 5 sec. Last updated by anonymous at February 20 2020, 3:36:29 PM.

Extract of Titanic passenger list

FINISHED

```

1 %spark
2
3 // we display the raw data
4 df.show(false)
5 df.printSchema()

```

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+---+
|Survived|Pclass|Name                                     |Sex   |Age |Siblings/Spouses Aboard|Parents/Children Aboard|Fare
|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

Titanic

10	13	IMr. Owen Harris Braund	Imale	122.0	11	10	17.25
11	11	IMrs. John Bradley (Florence Briggs Thayer) Cumings	Ifemale	138.0	11	10	171.2
11	13	IMiss. Laina Heikkinen	Ifemale	126.0	10	10	17.92
11	11	IMrs. Jacques Heath (Lily May Peel) Futrelle	Ifemale	135.0	11	10	153.1
10	13	IMr. William Henry Allen	Imale	135.0	10	10	18.05
10	13	IMr. James Moran	Imale	127.0	10	10	18.45

Took 1 sec. Last updated by anonymous at February 20 2020, 12:46:59 PM. (outdated)

Number of passenger who died or survived by sex

FINISHED

settings ▼

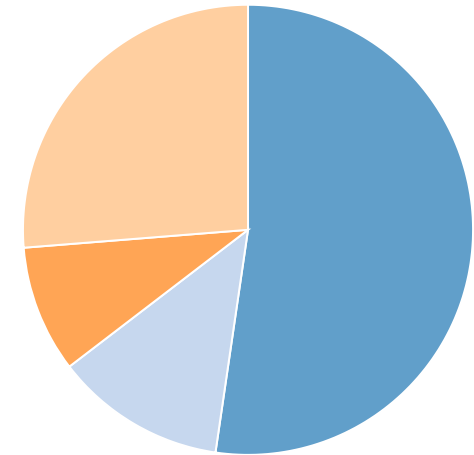
Survived	Sex
0	female
1	male
1	female
0	male

Took 1 sec. Last updated by anonymous at February 20 2020, 2:45:33 PM. (outdated)

Share of Passenger who survived or not by sex

FINISHED

Titanic



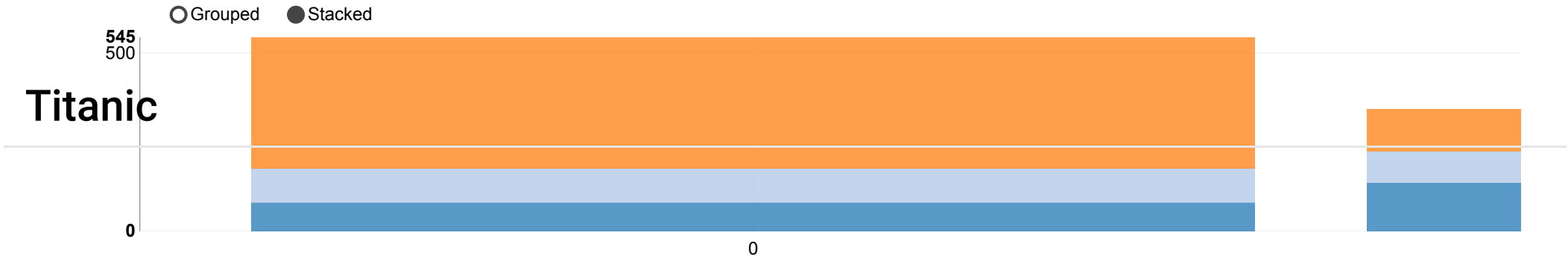
Took 0 sec. Last updated by anonymous at February 20 2020, 3:40:42 PM. (outdated)

Passenger decease or survival by passenger class

FINISHED

```
%spark.sql  
select Survived,Pclass from SMALLPASSTABLE order by Pclass asc
```





Took 0 sec. Last updated by anonymous at February 20 2020, 3:16:17 PM. (outdated)