TitanicV3

Titanic visualization assignment.

READY

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

FINISHED

```
%spark

import org.apache.spark.sql.Dataset;
import org.apache.spark.sql.functions.expr

//import org.apache.spark.sql.types.{StructType, StructField, StringType, IntegerType};

// import org.apache.commons.io.IOUtils
// import java.net.URL
// import java.nio.charset.Charset
// import org.apache.spark.ml.classification.DecisionTreeClassificationModel;
// import org.apache.spark.ml.classification.DecisionTreeClassifier;
// import org.apache.spark.ml.feature.StringIndexer;
// import org.apache.spark.ml.feature.StringIndexerModel;
// import org.apache.spark.ml.feature.VectorAssembler;
// import org.apache.spark.mllib.evaluation.MulticlassMetrics;
// import org.apache.spark.sql.SparkSession;
// 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;

// création du dataframe Spark
val df = spark.read.options(Map("inferSchema"->"true","delimiter"->",","header"->"true"))
  .csv("/Users/FXSUIRE/Google Drive/_00_Dev/Titanic/input.csv")

// application du schéma de données
// colonnes d'origine
case class passenger_cls(Survived:Int,Pclass:Int,Name:String,Sex:String,Age:Float,Siblings:Int,ParentAboard:Int,Fare:Float)

case class passenger_addedcol(Survived:Int,Pclass:Int,Name:String,Sex:String,Age:Float,Siblings:Int,ParentAboard:Int,Fare:Float,Fa

//Ratio d'actualisation du prix du ticket en liver sterling 2020
```

TitanicV3

```
// source: https://www.thisismoney.co.uk/money/bills/article-1633409/Historic-inflation-calculator-value-money-changed-1900.html
val UpdatingRatio = 115.8

// FareActualised = "Fare*UpdatingRatio"
val df2 = df.withColumn("FareUpdated", df("Fare") * UpdatingRatio)

// x0: Surived:Int
// x1: Pclass:Int
// x2: Name:String
// x3: Sex:String
// x4: Age:Float
// x5: Siblings:Int
// x6: ParentAboard:Int
// x7: Fare:Float

// x8: Actualised (2020) Fare: float

// mapping des colonnes originales
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

//enregistrement dans une table temporaire des données des passagers
passenger.registerTempTable("PASSENGER")

//sélection des données que l'on souhaite utiliser dans nos requêtes de visualisations

// Décès-survie du passager, classe, sex, prix du ticket
val SurvivalAgeClassSexFare = sqlContext.sql("select Survived, Pclass, Sex, Fare from PASSENGER")
SurvivalAgeClassSexFare.registerTempTable("SMALLPASSTABLE")

// Statistiques basiques sur le prix des tickets
val BasicStats = sqlContext.sql("select min(fare), max(fare), avg(fare), stddev(fare) from PASSENGER where fare>0").toDF

val BasicStatsRenamed = BasicStats.withColumnRenamed("min(fare)","Minimum fare price")
            .withColumnRenamed("max(fare)","Maximum fare price")
            .withColumnRenamed("avg(fare)","Average fare price")
            .withColumnRenamed("stddev(fare)","Standard deviation")

BasicStatsRenamed.registerTempTable("BASICFARETABLE")

// Décès-survie du passager, classe, sex, prix du ticket 2020

//enregistrement dans une table temporaire des données des passagers avec ajout des colonnes Fareupdated
df2.registerTempTable("PASSPROCDATA")

val SurvivalAgeClassSexFare2020 = sqlContext.sql("select Survived, Pclass, Sex, Fare, FareUpdated from PASSPROCDATA")
```

```
79 SurvivalAgeClassSexFare2020.registerTempTable("SMALLPASSTABLE2020")
80
81 // statistiques sur le prix du ticket actualisé, c'est à dire en GBP de 2020
82
83 val BasicUpdFareStats = sqlContext.sql("select min(FareUpdated), max(FareUpdated), avg(FareUpdated), stddev(FareUpdated) from SMAL
84
85 val BasicUpdStatsRenamed = BasicUpdFareStats.withColumnRenamed("min(FareUpdated)","Minimum fare price")
86             .withColumnRenamed("max(FareUpdated)","Maximum fare price")
87             .withColumnRenamed("avg(FareUpdated)","Average fare price")
88             .withColumnRenamed("stddev(FareUpdated)","Standard deviation")
89
90 BasicUpdStatsRenamed.registerTempTable("BASICUPDFARETABLE")
91
```

TitanicV3

```
warning: there were 6 deprecation warnings; re-run with -deprecation for details
import sqlContext.implicits._
import org.apache.spark.sql.Dataset
import org.apache.spark.sql.functions.expr
df: org.apache.spark.sql.DataFrame = [Survived: int, Pclass: int ... 6 more fields]
defined class passenger_cls
defined class passenger_addedcol
UpdatingRatio: Double = 115.8
df2: org.apache.spark.sql.DataFrame = [Survived: int, Pclass: int ... 7 more fields]
```

Took 6 sec. Last updated by anonymous at February 21 2020, 6:03:56 PM. (outdated)

Extract of Titanic passenger list (source data)

READY

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

```
|0      |1     |Mr. Timothy J McCarthy                              |male  |54.0|0                    |0                    |51.8
625|
|0      |3     |Master. Gosta Leonard Palsson                       |male  |2.0 |3                    |1                    |21.0
75 |
|1      |3     |Mrs. Oscar W (Elisabeth Vilhelmina Berg) Johnson     |female|27.0|0                    |2                    |11.1
333|
|1      |2     |Mrs. Nicholas (Adele Achem) Nasser                  |female|14.0|1                    |0                    |30.0
708|
```

```
|1       |3     |Miss. Marguerite Rut Sandstrom                |female|4.0 |1 |1 |16.7
|
|1       |1     |Miss. Elizabeth Bonnell                       |female|58.0|0 |0 |26.5
5  |
|0       |3     |Mr. William Henry Saundercock                 |male  |20.0|0 |0 |8.05
|
|0       |3     |Mr. Anders Johan Andersson                    |male  |39.0|1 |5 |31.2
75 |
|0       |3     |Miss. Hulda Amanda Adolfina Vestrom           |female|14.0|0 |0 |7.85
```

TitanicV3

Extract of Titanic passenger list (source data + added columns)

FINISHED

```
1 %spark
2
3 // we display the raw data
4     df2.show(false)
5     //df2.printSchema()
```

```
|1       |1     |Mrs. John Bradley (Florence Briggs Thayer) Cumings    |female|38.0|1 |0 |71.2
833|8254.60614        |
|1       |3     |Miss. Laina Heikkinen                         |female|26.0|0 |0 |7.92
5  |917.7149999999999 |
|1       |1     |Mrs. Jacques Heath (Lily May Peel) Futrelle   |female|35.0|1 |0 |53.1
|6148.98           |
|0       |3     |Mr. William Henry Allen                       |male  |35.0|0 |0 |8.05
|932.19            |
|0       |3     |Mr. James Moran                               |male  |27.0|0 |0 |8.45
83 |979.4711399999999 |
|0       |1     |Mr. Timothy J McCarthy                        |male  |54.0|0 |0 |51.8
625|6005.6775         |
|0       |3     |Master. Gosta Leonard Palsson                 |male  |2.0 |3 |1 |21.0
75 |2440.4849999999997|
|1       |3     |Mrs. Oscar W (Elisabeth Vilhelmina Berg) Johnson|female|27.0|0 |2 |11.1
333|1289.23614        |
|1       |2     |Mrs. Nicholas (Adele Achem) Nasser            |female|14.0|1 |0 |30.0
708|3482.1986399999996|
```

Took 2 sec. Last updated by anonymous at February 21 2020, 6:04:10 PM.

Number of deceased or living passenger by sex after the Shipwreck

FINISHED

```
%spark.sql
select Survived,Sex, count(Survived) from SMALLPASSTABLE group by Survived, Sex
```

settings

TitanicV3

Survived	Sex
0	female
1	male
1	female
0	male

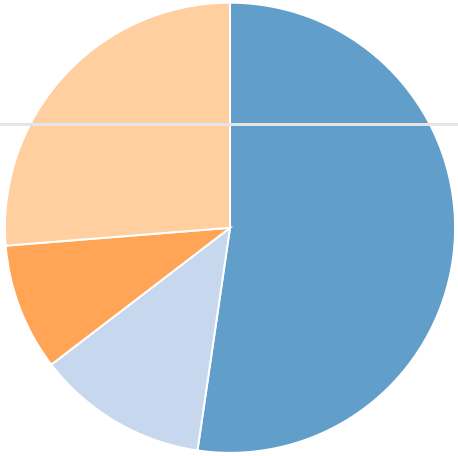
Took 1 sec. Last updated by anonymous at February 21 2020, 5:50:50 PM.

Share of Passenger who survived or not by sex

FINISHED

settings

TitanicV3

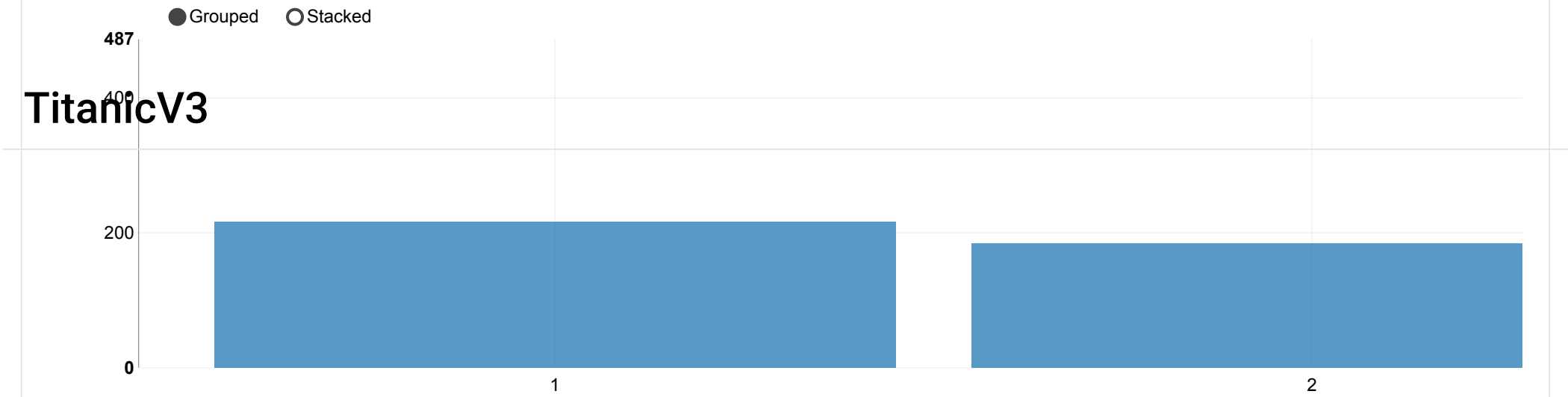


Took 1 sec. Last updated by anonymous at February 21 2020, 6:06:22 PM.

Titanic passenger class distribution

READY

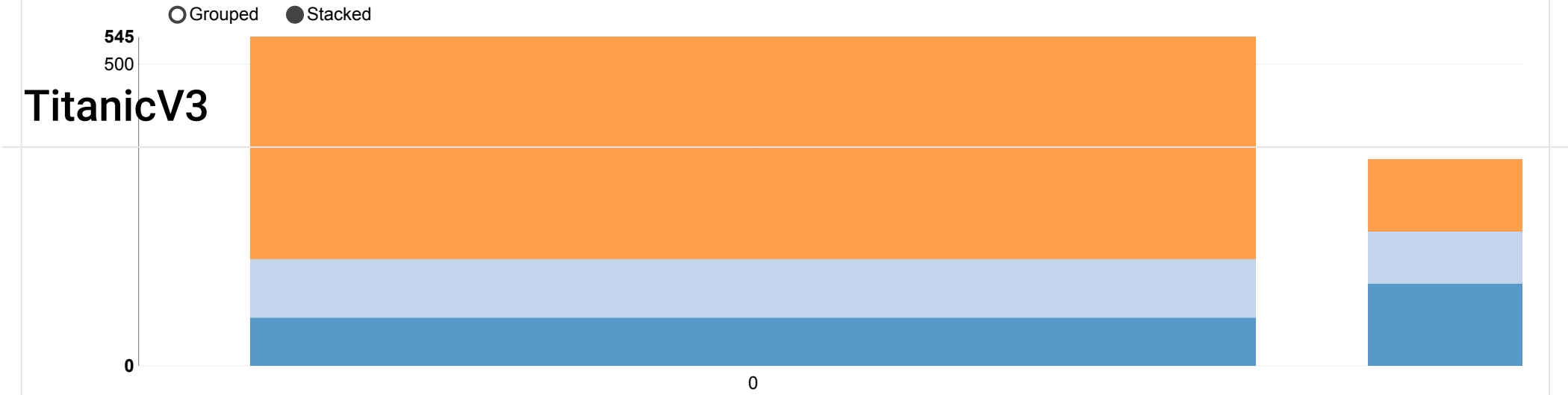
settings




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

READY

settings

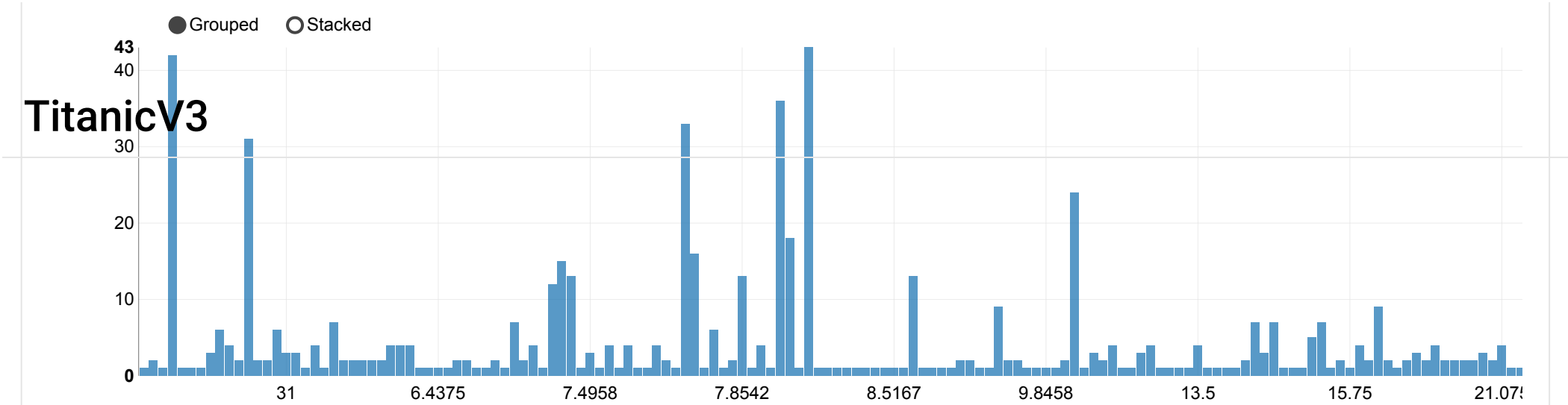


Titanic fare price frequency distribution

READY

```
%spark.sql
select fare, count(1) nInBin
from SMALLPASSTABLE
where fare>0
group by fare
order by fare
```

settings



Basic statistics about Titanic fare price (in year 1912 GBP)

READY

```
%spark.sql
select * from BASICFARETABLE
```

settings

Minimum fare price	Maximum fare price	Average fare price
4.0125	512.3292	32.86113275737937

TitanicV3

Basic statistics about titanic fare price (in GBP actualised as of 2020)

FINISHED

```
%spark.sql
select * from BASICUPDFARETABLE
```

settings

Minimum fare price	Maximum fare price	Average fare price
464.64750000000004	59327.72136	3805.319164747697

Took 0 sec. Last updated by anonymous at February 21 2020, 6:05:00 PM. (outdated)

```
%spark.sql
```

READY