

Write a spark program to find the null values in the columns and provide the column null reason

```
import org.apache.log4j.Level
import org.apache.log4j.Logger
import org.apache.spark.SparkConf
import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.types.StructField
import org.apache.spark.sql.types.IntegerType
import org.apache.spark.sql.types.StringType
import org.apache.spark.sql.types.StructType
import org.apache.spark.sql.Row
import org.apache.spark.sql.functions.when

object DataframeMissingDatainColumns extends App{

    Logger.getLogger("org").setLevel(Level.ERROR)

    val sparkConf = new SparkConf()
    sparkConf.set("spark.app.name", "Dataframe Missing Data in Columns")
    sparkConf.set("spark.master", "local[2]")

    val spark = SparkSession.builder()
        .config(sparkConf)
        .getOrCreate()

    val data = Seq(Row(1,"mangoes",66,"v432"),
        Row(2,"mangoes",66,"v4632"),
        Row(3,"apples",null,"v4312"),
        Row(4,"apples",66,"v432"),
        Row(5,"strawberries",66,null),
        Row(6,"strawberries",null,null),
        Row(7,"cherries",null,"v256"),
        Row(8,"cherries",null,null),
        Row(9,"Grapes",66,null),
        Row(10,"Grapes",66,"v432"),
        Row(11,"mangoes",66,null),
        Row(12,"mangoes",66,"v4632"),
        Row(13,"apples",null,"v4312"),
        Row(14,"apples",66,"v432")
    )
```

```

val schema = List(StructField("id",IntegerType,false),
  StructField("name",StringType,false),
  StructField("qty",IntegerType,true),
  StructField("vendor_Id",StringType,true))

import spark.implicits._

val df = spark.createDataFrame(spark.sparkContext.parallelize(data),
  StructType(schema))

val validData = df.filter('qty.isNotNull && 'vendor_Id.isNotNull)

val invalidData = df.filter('qty.isNull || 'vendor_Id.isNull)
  .withColumn("Reason", when('vendor_Id.isNull && 'qty.isNull, "qty and
vendor_Id is missing")
    .when('vendor_Id.isNull, "vendor_Id is missing")
    .otherwise(when('qty.isNull,"qty is missing")))

validData.show(false)

invalidData.select("id","name","Reason").show(false)

}

```

Output :

```

+---+-----+---+-----+
|id |name   |qty|vendor_Id|
+---+-----+---+-----+
| 1 |mangoes|66 |v432     |
| 2 |mangoes|66 |v4632    |
| 4 |apples |66 |v432     |
|10 |grapes |66 |v432     |
|12 |mangoes|66 |v4632    |
|14 |apples |66 |v432     |
+---+-----+---+-----+

```

```
+---+-----+-----+
|id |name      |reason      |
+---+-----+-----+
| 3 |apples     |qty is missing|
| 5 |strawberries|vendor_Id is missing|
| 6 |strawberries|qty and vendor_Id is missing|
| 7 |cherries   |qty is missing|
| 8 |cherries   |qty and vendor_Id is missing|
| 9 |grapes     |vendor_Id is missing|
|11 |mangoes    |vendor_Id is missing|
|13 |apples     |qty is missing|
+---+-----+-----+
```