```scala
val df = spark.read.option("header", "true")
                   .option("inferSchema","true")
                   .csv("/FileStore/tables/movies.csv")

// Ensure that the data has been uploaded successfully
df.show(false)
df.count

// Register the DataFrame as an SQL table
df.createOrReplaceTempView("movies_table")
```

```
+-----------------+--------------------------+----+
|actor            |title                     |year|
+-----------------+--------------------------+----+
|McClure, Marc (I)|Freaky Friday             |2003|
|McClure, Marc (I)|Coach Carter              |2005|
|McClure, Marc (I)|Superman II               |1980|
|McClure, Marc (I)|Apollo 13                 |1995|
|McClure, Marc (I)|Superman                  |1978|
|McClure, Marc (I)|Back to the Future        |1985|
|McClure, Marc (I)|Back to the Future Part III|1990|
|Cooper, Chris (I)|Me, Myself & Irene        |2000|
|Cooper, Chris (I)|October Sky               |1999|
|Cooper, Chris (I)|Capote                    |2005|
|Cooper, Chris (I)|The Bourne Supremacy      |2004|
|Cooper, Chris (I)|The Patriot               |2000|
|Cooper, Chris (I)|The Town                  |2010|
|Cooper, Chris (I)|Seabiscuit                |2003|
|Cooper, Chris (I)|A Time to Kill            |1996|
|Cooper, Chris (I)|Where the Wild Things Are |2009|
|Cooper, Chris (I)|The Muppets               |2011|
|Cooper, Chris (I)|American Beauty           |1999|
```

```
var moviesCountPerYearSqlWay = spark.sql(
  """
  select year, count(title) as count
  from movies_table
  group by year
  having year is not null
  order by year asc
  """
)

moviesCountPerYearSqlWay.show(false)
```

```
+----+-----+
|year|count|
+----+-----+
|1961|2    |
|1967|2    |
|1972|12   |
|1973|5    |
|1975|5    |
|1977|40   |
|1978|30   |
|1979|37   |
|1980|47   |
|1981|53   |
|1982|103  |
|1983|119  |
|1984|149  |
|1985|133  |
|1986|174  |
|1987|126  |
|1988|111  |
```

```
var moviesCountPerYearDataFrameWay = df.groupBy("year")
                                       .agg(count("*").as("count"))
                                       .where("year is not null")
                                       .orderBy('year.asc)

moviesCountPerYearDataFrameWay.show
```

```
+----+-----+
|year|count|
+----+-----+
|1961|    2|
|1967|    2|
|1972|   12|
|1973|    5|
|1975|    5|
|1977|   40|
|1978|   30|
|1979|   37|
|1980|   47|
|1981|   53|
|1982|  103|
|1983|  119|
|1984|  149|
|1985|  133|
|1986|  174|
|1987|  126|
|1988|  111|
|1989|  152|
```

```
var top5ActorsInNumberOfMoviesSqlWay = spark.sql(
  """
  select actor, count(title) as number_of_movies
  from movies_table
  group by actor
  order by number_of_movies desc
  limit 5
  """
)
```

```
top5ActorsInNumberOfMoviesSqlWay.show
```

```
+------------------+----------------+
|             actor|number_of_movies|
+------------------+----------------+
|  Tatasciore, Fred|              38|
|     Welker, Frank|              38|
|Jackson, Samuel L.|              32|
|     Harnell, Jess|              31|
|       Damon, Matt|              27|
+------------------+----------------+
```

```
top5ActorsInNumberOfMoviesSqlWay: org.apache.spark.sql.DataFrame = [actor: string, number_of_movies: bigint]
```

```
var top5ActorsInNumberOfMoviesDataFrameWay = df.groupBy("actor")
                                               .agg(count("title").as("number_of_movies"))
                                               .orderBy('number_of_movies.desc)
                                               .limit(5)

top5ActorsInNumberOfMoviesDataFrameWay.show

+------------------+----------------+
|             actor|number_of_movies|
+------------------+----------------+
|  Tatasciore, Fred|              38|
|     Welker, Frank|              38|
|Jackson, Samuel L.|              32|
|     Harnell, Jess|              31|
|       Damon, Matt|              27|
+------------------+----------------+

top5ActorsInNumberOfMoviesDataFrameWay: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [actor: string, number_of_movies: bigint]
```

```
var tomHanksMoviesDataFrameWay = df.select("title", "year").where("actor == 'Hanks, Tom'").orderBy('year.asc)

tomHanksMoviesDataFrameWay.show
```

```
+--------------------+----+
|               title|year|
+--------------------+----+
|      Bachelor Party|1984|
|Sleepless in Seattle|1993|
|        Philadelphia|1993|
|        Forrest Gump|1994|
|           Apollo 13|1995|
|           Toy Story|1995|
| Saving Private Ryan|1998|
|     You've Got Mail|1998|
|      The Green Mile|1999|
|         Toy Story 2|1999|
|           Cast Away|2000|
| Catch Me If You Can|2002|
|   Road to Perdition|2002|
|   The Polar Express|2004|
|        The Terminal|2004|
|     The Ladykillers|2004|
|Magnificent Desol...|2005|
|           The Queen|2006|
```