```
# Look at the column names
print (ratings.columns)

# Look at the first few rows of data
print (ratings.show())

# Count the total number of ratings in the dataset
numerator = ratings.select("rating").count()

# Count the number of distinct userIds and distinct movieIds.
num_users = ratings.select("userId").distinct().count()
num_movies = ratings.select("movieId").distinct().count()

# Set the denominator equal to the number of users multiplied by the number of movies
denominator = num_users * num_movies

# Divide the numerator by the denominator
sparsity = (1.0 - (numerator *1.0)/denominator)*100
print ("The ratings dataframe is ", "%.2f" % sparsity + "% empty.")
98.36% empty

# Import the requisite packages
from pyspark.sql.functions import col

# View the ratings dataset
ratings.show()

# Filter out all userIds greater than 100
ratings.filter(col("userId") < 100).show()

# Group data by userId, count ratings
ratings.groupby("userId").count().show()

# Min num ratings for movies
print("Movie with the fewest ratings: ")
ratings.groupBy("movieId").count().select(min("count")).show()

Movie with the fewest ratings:
+----------+
|min(count)|
+----------+
|         1|
+----------+
# Avg num ratings per movie
print("Avg num ratings per movie: ")
ratings.groupBy("movieId").count().select(avg("count")).show()
```

```
Avg num ratings per movie:
+------------------+
|        avg(count)|
+------------------+
|11.030664019413193|
+------------------+
```

```
# Min num ratings for user
print("User with the fewest ratings: ")
ratings.groupBy("userId").count().select(min("count")).show()
```

```
User with the fewest ratings:
+----------+
|min(count)|
+----------+
|        20|
+----------+
```

```
# Avg num ratings per users
print("Avg num ratings per user: ")
ratings.groupBy("userId").count().select(avg("count")).show()
```

```
Avg num ratings per user:
+------------------+
|        avg(count)|
+------------------+
|149.03725782414307|
+------------------+
```

View Schema

```
# Use the .printSchema() method to see the datatypes of the ratings dataset.
ratings.printSchema()
```

```
root
 |-- userId: string (nullable = true)
 |-- movieId: string (nullable = true)
 |-- rating: string (nullable = true)
 |-- timestamp: string (nullable = true)
```

```
# Tell Spark to convert the columns to the proper data types.
ratings = ratings.select(ratings.userId.cast("integer"), ratings.movieId.cast("integer"),
ratings.rating.cast("double"))

# Call .printSchema() again to confirm the columns are now in the correct format
ratings.printSchema()

root
 |-- userId: integer (nullable = true)
 |-- movieId: integer (nullable = true)
 |-- rating: integer (nullable = true)


from pyspark.ml.evaluation import RegressionEvaluator
from pyspark.ml.recommendation import ALS
from pyspark.ml.tuning import ParamGridBuilder, CrossValidator

# Create test and train set
(train, test) = ratings.randomSplit([0.8, 0.2], seed = 1234)

# Create ALS model
als = ALS(userCol="userId", itemCol="movieId", ratingCol="rating", nonnegative = True,
implicitPrefs = False)

# Confirm that a model called "als" was created
type(als)

pyspark.ml.recommendation.ALS

# Import the requisite items
from pyspark.ml.evaluation import RegressionEvaluator
from pyspark.ml.tuning import ParamGridBuilder, CrossValidator

# Add hyperparameters and their respective values to param_grid
param_grid = ParamGridBuilder() \
        .addGrid(als.rank, [10, 50, 100, 150]) \
        .addGrid(als.maxIter, [5, 50, 100, 200]) \
        .addGrid(als.regParam, [0.01, .05, .1, .15]) \
        .build()

# Define evaluator as RMSE and print length of evaluator
evaluator = RegressionEvaluator(metricName="rmse", labelCol="rating",
predictionCol="prediction")
```

```
print ("Num models to be tested: ", len(param_grid))

Num models to be tested:  64

# Build cross validator
cv = CrossValidator(estimator=als, estimatorParamMaps=param_grid, evaluator=evaluator,
numFolds=5)

# Confirm cv was built
print (cv)

# Fit cross validation to be the ‘train’ dataset
Model = cv.fit(train)

#Extract best model from the cv model above
best.model= model.bestModel

# Print best_model
print(type(best_model))

# Complete the code below to extract the ALS model parameters
print("**Best Model**")

# Print "Rank"
print("  Rank:", best_model.getRank())

# Print "MaxIter"
print("  MaxIter:", best_model.getMaxIter())

# Print "RegParam"
print("  RegParam:", best_model.getRegParam())

<class 'pyspark.ml.recommendation.ALS'>
**Best Model**
  Rank: 50
  MaxIter: 100
  RegParam: 0.1

# Since the best model is within our range, if it’s on both edges, then we should try different
numbers

# View the predictions
test_predictions.show()
```

```
# Calculate and print the RMSE of the test_predictions
RMSE = evaluator.evaluate(test_predictions)
print (RMSE)

+------+-------+------+------------------+
|userId|movieId|rating|        prediction|
+------+-------+------+------------------+
|   380|    463|   3.0| 4.093334993256898|
|   460|    471|   5.0| 4.789751482535894|
|   440|    471|   3.0| 2.440344619907418|
|   306|    471|   3.0|3.3247629567900976|
|    19|    471|   3.0| 3.067333162723295|
|   299|    471|   4.5| 5.218491499885204|
|   537|    471|   5.0|  5.69083471617962|
|   241|    471|   4.0| 3.816546176254299|
|    23|    471|   3.5| 2.539020466532909|
|   195|    471|   3.0| 3.355342979133588|
|   487|    471|   4.0| 3.105186392315445|
|   242|    471|   5.0| 5.893115933597325|
|    30|    471|   4.0| 4.017221049024606|
|   516|   1088|   3.0|3.3911144131643005|
|   111|   1088|   3.5| 4.504826156475481|
|    57|   1088|   4.0| 3.024549429857915|
|    54|   1088|   5.0| 5.235519746597422|
|    19|   1088|   3.0|  3.70884171609874|
|   387|   1088|   4.0| 4.070842875474657|
|   514|   1088|   3.0| 2.313176685047038|
+------+-------+------+------------------+
only showing top 20 rows

0.6332304339145925

# Look at user 60's ratings
print ("User 60's Ratings:")
original_ratings.filter(col("userId") == 60).sort("rating", ascending = False).show()

# Look at the movies recommended to user 60
print ("User 60s Recommendations:")
recommendations.filter(col("userId") == 60).show()

# Look at user 63's ratings
print ("User 63's Ratings:")
original_ratings.filter(col("userId") == 63).sort("rating", ascending = False).show()
```

```
# Look at the movies recommended to user 63
print ("User 63's Recommendations:")
recommendations.filter(col("userId") == 63).show()
```

```
User 60's Ratings:
   +------+-------+------+--------------------+--------------------+
   |userId|movieId|rating|               title|              genres|
   +------+-------+------+--------------------+--------------------+
   |    60|    858|     5|  GodfatherThe(1972)|         Crime|Drama|
   |    60|    235|     5|        EdWood(1994)|        Comedy|Drama|
   |    60|   1732|     5|BigLebowskiThe(1998)|        Comedy|Crime|
   |    60|   2324|     5|LifeIsBeautiful(L...|Comedy|Drama|Roma...|
   |    60|   3949|     5|RequiemforaDream(...|               Drama|
   |    60|    541|     5|   BladeRunner(1982)|Action|Sci-Fi|Thr...|
   |    60|   5995|     5|    PianistThe(2002)|           Drama|War|
   |    60|   6350|     5|Laputa:Castleinth...|Action|Adventure|...|
   |    60|   7361|     5|EternalSunshineof...|Drama|Romance|Sci-Fi|
   |    60|   8638|     5|  BeforeSunset(2004)|       Drama|Romance|
   |    60|   8981|     5|        Closer(2004)|       Drama|Romance|
   |    60|  27803|     5|SeaInsideThe(Mara...|               Drama|
   |    60|  30749|     5|   HotelRwanda(2004)|           Drama|War|
   |    60|   5060|     5|M*A*S*H(a.k.a.MAS...|    Comedy|Drama|War|
   |    60|   1221|     5|Godfather:PartIIT...|         Crime|Drama|
   |    60|   5690|     5|GraveoftheFirefli...| Animation|Drama|War|
   |    60|   1653|     5|       Gattaca(1997)|Drama|Sci-Fi|Thri...|
   |    60|     16|   4.5|        Casino(1995)|         Crime|Drama|
   |    60|    111|   4.5|    TaxiDriver(1976)|Crime|Drama|Thriller|
   |    60|   1080|   4.5|MontyPython'sLife...|              Comedy|
   +------+-------+------+--------------------+--------------------+
   only showing top 20 rows

   User 60s Recommendations:
   +------+-------+----------+--------------------+--------------------+
   |userId|movieId|prediction|               title|              genres|
   +------+-------+----------+--------------------+--------------------+
   |    60|  83318|  5.810963|       GoatThe(1921)|              Comedy|
   |    60|  83411|  5.810963|          Cops(1922)|              Comedy|
   |    60|  73344|  5.315315|ProphetA(UnProphÃ...|         Crime|Drama|
   |    60|   3309| 5.2298656|     Dog'sLifeA(1918)|             Comedy|
   |    60|   8609| 5.2298656|OurHospitality(1923)|              Comedy|
   |    60|  72647| 5.2298656|    Zorn'sLemma(1970)|               Drama|
   |    60|   5059| 5.2298656|LittleDieterNeeds...|         Documentary|
   |    60|   8797| 5.2298656|       Salesman(1969)|         Documentary|
   |    60|  25764| 5.2298656|  CameramanThe(1928)|Comedy|Drama|Romance|
   |    60|   7074| 5.2298656|  NavigatorThe(1924)|              Comedy|
   |    60|  31547| 5.2298656|LessonsofDarkness...|     Documentary|War|
   |    60|   4405| 5.2298656|LastLaughThe(Letz...|               Drama|
   |    60|  26400| 5.2298656| GatesofHeaven(1978)|         Documentary|
   |    60|  80599| 5.2298656|BusterKeaton:AHar...|         Documentary|
   |    60|  92494| 5.1418443|DylanMoran:Monste...|  Comedy|Documentary|
   |    60|   3216| 5.1418443|VampyrosLesbos(Va...|Fantasy|Horror|Th...|
   |    60|   6918| 5.1184077|UnvanquishedThe(A...|               Drama|
   |    60|  40412| 5.0673676|DeadMan'sShoes(2004)|      Crime|Thriller|
   |    60|  52767|  5.043912|         21Up(1977)|         Documentary|
   |    60|   8955| 5.0317564|      Undertow(2004)|Crime|Drama|Thriller|
   +------+-------+----------+--------------------+--------------------+

   User 63's Ratings:
   +------+-------+------+--------------------+--------------------+
   |userId|movieId|rating|               title|              genres|
   +------+-------+------+--------------------+--------------------+
   |    63|      1|     5|      ToyStory(1995)|Adventure|Animati...|
   |    63|     16|     5|        Casino(1995)|         Crime|Drama|
   |    63|    260|     5|StarWars:EpisodeI...|Action|Adventure|...|
   |    63|    318|     5|ShawshankRedempti...|         Crime|Drama|
   |    63|    592|     5|        Batman(1989)|Action|Crime|Thri...|
   |    63|   1193|     5|OneFlewOvertheCuc...|               Drama|
   |    63|   1198|     5|RaidersoftheLostA...|    Action|Adventure|
   |    63|   1214|     5|         Alien(1979)|       Horror|Sci-Fi|
   |    63|   1221|     5|Godfather:PartIIT...|         Crime|Drama|
   |    63|   1259|     5|     StandbyMe(1986)|     Adventure|Drama|
   |    63|   1356|     5|StarTrek:FirstCon...|Action|Adventure|...|
   |    63|   1639|     5|    ChasingAmy(1997)|Comedy|Drama|Romance|
   |    63|   2797|     5|           Big(1988)|Comedy|Drama|Fant...|
   |    63|   2858|     5|AmericanBeauty(1999)|       Drama|Romance|
   |    63|   2918|     5|FerrisBueller'sDa...|              Comedy|
   |    63|   3114|     5|     ToyStory2(1999)|Adventure|Animati...|
   |    63|   3176|     5|TalentedMr.Ripley...|Drama|Mystery|Thr...|
   |    63|   3481|     5|  HighFidelity(2000)|Comedy|Drama|Romance|
   |    63|   3578|     5|     Gladiator(2000)|Action|Adventure|...|
   |    63|   4306|     5|         Shrek(2001)|Adventure|Animati...|
   +------+-------+------+--------------------+--------------------+
   only showing top 20 rows

   User 63's Recommendations:
   +------+-------+----------+--------------------+--------------------+
   |userId|movieId|prediction|               title|              genres|
   +------+-------+----------+--------------------+--------------------+
   |    63|  92210| 4.8674645|DisappearanceofHa...|Adventure|Animati...|
   |    63| 110873| 4.8674645|CentenarianWhoCli...|Adventure|Comedy|...|
   |    63|   9010| 4.8588977|LoveMeIfYouDare(J...|       Drama|Romance|
   |    63| 108583|  4.836118|FawltyTowers(1975...|              Comedy|
   |    63|   8530| 4.8189244|    DearFrankie(2004)|       Drama|Romance|
   |    63|  83318|  4.813581|       GoatThe(1921)|              Comedy|
   |    63|  83411|  4.813581|          Cops(1922)|              Comedy|
   |    63|  65037| 4.7906556|          BenX(2007)|               Drama|
   |    63|  54328|  4.688013|MyBestFriend(Monm...|              Comedy|
   |    63|   3437|  4.678849|     CoolasIce(1991)|               Drama|
   |    63|   2924|  4.675808|DrunkenMaster(Jui...|       Action|Comedy|
   |    63|   1196| 4.6633716|StarWars:EpisodeV...|Action|Adventure|...|
   |    63|  27156| 4.6382804|NeonGenesisEvange...|Action|Animation|...|
   |    63|  26865| 4.6308517|FistofLegend(Jing...|        Action|Drama|
   |    63|   5244| 4.6302986|ShogunAssassin(1980)|    Action|Adventure|
   |    63|  93320| 4.6302986|TrailerParkBoys(1...|        Comedy|Crime|
   |    63|  50641| 4.6302986|  House(Hausu)(1977)|Comedy|Fantasy|Ho...|
   |    63|   6598|  4.624031|StepIntoLiquid(2002)|         Documentary|
   |    63|   7502| 4.6232696|BandofBrothers(2001)|    Action|Drama|War|
   |    63|  73344|  4.609774|ProphetA(UnProphÃ...|         Crime|Drama|
   +------+-------+----------+--------------------+--------------------+
```