

# Module 5: Playing with Spark RDD's

---

## Case Study 1

edureka!

**edureka!**

© Brain4ce Education Solutions Pvt. Ltd.

## Case Study: Mobile App Store

### Domain: Analytics

The ever-changing mobile landscape is a challenging space to navigate. Android holds about 53.2% of the smartphone market, while iOS is 43%. To get more people to download your app, you need to make sure they can easily find your app. Mobile app analytics is a great way to understand the existing strategy to drive growth and retention of future user.

The data set contains more than 7000 Apple iOS mobile application details. The data was extracted from the iTunes Search API.

#### Tasks:

With millions of apps around nowadays, the data has become very key to getting top trending apps in iOS app store. As a data scientist, you are required to explore the datasets including cleaning and transforming the dataset.

1. Load csv into spark as a text file
2. Parse the data as csv.
3. Convert bytes to MB and GB in a new column
4. List top 10 trending apps
5. The difference in the average number of screenshots displayed of highest and lowest rating apps.
6. What percentage of high rated apps support multiple languages.
7. How does app details contribute to user ratings?
8. Compare the statistics of different app groups/genres.
9. Does length of app description contribute to the ratings?
10. Create a spark-submit application for the same and print the findings in the log