In [0]:
```
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.types import StructType, StructField, StringType, IntegerType
from pyspark.sql.functions import *
```

In [0]:
```
df = spark.read.load('/FileStore/tables/googleplaystore-1.csv', format="csv", sep=","
```

In [0]:
```
df.count()
```

Out[3]: 10841

In [0]:
```
df.show(1)
```

```
+--------------------+--------------+------+-------+----+--------+----+-----+---------
-----+------------+---------------+-----------+------------+
|                 App|      Category|Rating|Reviews|Size|Installs|Type|Price|Content R
ating|      Genres|   Last Updated|Current Ver| Android Ver|
+--------------------+--------------+------+-------+----+--------+----+-----+---------
-----+------------+---------------+-----------+------------+
|Photo Editor & Ca...|ART_AND_DESIGN|   4.1|    159| 19M| 10,000+|Free|    0|      Eve
ryone|Art & Design|January 7, 2018|      1.0.0|4.0.3 and up|
+--------------------+--------------+------+-------+----+--------+----+-----+---------
-----+------------+---------------+-----------+------------+
only showing top 1 row
```

In [0]:
```
df.printSchema()
```

```
root
 |-- App: string (nullable = true)
 |-- Category: string (nullable = true)
 |-- Rating: double (nullable = true)
 |-- Reviews: string (nullable = true)
 |-- Size: string (nullable = true)
 |-- Installs: string (nullable = true)
 |-- Type: string (nullable = true)
 |-- Price: string (nullable = true)
 |-- Content Rating: string (nullable = true)
 |-- Genres: string (nullable = true)
 |-- Last Updated: string (nullable = true)
 |-- Current Ver: string (nullable = true)
 |-- Android Ver: string (nullable = true)
```

Data Cleaning

In [0]:
```
df = df.drop("Size","Content Rating", "Last Updated","Android Ver")
```

In [0]:
```
df.show()
```

```
+--------------------+--------------+------+-------+-----------+----+-----+-----------
--------+-----------------+
|                 App|      Category|Rating|Reviews|   Installs|Type|Price|
Genres|       Current Ver|
+--------------------+--------------+------+-------+-----------+----+-----+-----------
--------+-----------------+
|Photo Editor & Ca...|ART_AND_DESIGN|   4.1|    159|    10,000+|Free|    0|        Art
& Design|             1.0.0|
| Coloring book moana|ART_AND_DESIGN|   3.9|    967|   500,000+|Free|    0|Art & Desig
n;Pret...|              2.0.0|
|U Launcher Lite –...|ART_AND_DESIGN|   4.7|  87510| 5,000,000+|Free|    0|        Art
& Design|             1.2.4|
|Sketch - Draw & P...|ART_AND_DESIGN|   4.5| 215644|50,000,000+|Free|    0|        Art
& Design|Varies with device|
|Pixel Draw - Numb...|ART_AND_DESIGN|   4.3|    967|   100,000+|Free|    0|Art & Desig
n;Crea...|               1.1|
|Paper flowers ins...|ART_AND_DESIGN|   4.4|    167|    50,000+|Free|    0|        Art
& Design|               1.0|
|Smoke Effect Phot...|ART_AND_DESIGN|   3.8|    178|    50,000+|Free|    0|        Art
& Design|               1.1|
|    Infinite Painter|ART_AND_DESIGN|   4.1|  36815| 1,000,000+|Free|    0|        Art
& Design|          6.1.61.1|
|Garden Coloring Book|ART_AND_DESIGN|   4.4|  13791| 1,000,000+|Free|    0|        Art
& Design|             2.9.2|
|Kids Paint Free –...|ART_AND_DESIGN|   4.7|    121|    10,000+|Free|    0|Art & Desig
n;Crea...|               2.8|
|Text on Photo - F...|ART_AND_DESIGN|   4.4|  13880| 1,000,000+|Free|    0|        Art
& Design|             1.0.4|
|Name Art Photo Ed...|ART_AND_DESIGN|   4.4|   8788| 1,000,000+|Free|    0|        Art
& Design|            1.0.15|
|Tattoo Name On My...|ART_AND_DESIGN|   4.2|  44829|10,000,000+|Free|    0|        Art
& Design|               3.8|
|Mandala Coloring ...|ART_AND_DESIGN|   4.6|   4326|   100,000+|Free|    0|        Art
& Design|             1.0.4|
|3D Color Pixel by...|ART_AND_DESIGN|   4.4|   1518|   100,000+|Free|    0|        Art
& Design|             1.2.3|
|Learn To Draw Kaw...|ART_AND_DESIGN|   3.2|     55|     5,000+|Free|    0|        Art
& Design|               NaN|
|Photo Designer - ...|ART_AND_DESIGN|   4.7|   3632|   500,000+|Free|    0|        Art
& Design|               3.1|
|350 Diy Room Deco...|ART_AND_DESIGN|   4.5|     27|    10,000+|Free|    0|        Art
& Design|               1.0|
|FlipaClip - Carto...|ART_AND_DESIGN|   4.3| 194216| 5,000,000+|Free|    0|        Art
& Design|             2.2.5|
|        ibis Paint X|ART_AND_DESIGN|   4.6| 224399|10,000,000+|Free|    0|        Art
& Design|             5.5.4|
+--------------------+--------------+------+-------+-----------+----+-----+-----------
--------+-----------------+
only showing top 20 rows
```

In [0]:
```
df=df.drop("Current Ver")
```

In [0]:
```
df.show()
```

```
+--------------------+--------------+------+-------+-----------+----+-----+-----------
---------+
|                 App|      Category|Rating|Reviews|   Installs|Type|Price|
Genres|
+--------------------+--------------+------+-------+-----------+----+-----+-----------
---------+
|Photo Editor & Ca...|ART_AND_DESIGN|   4.1|    159|    10,000+|Free|    0|        Art
& Design|
| Coloring book moana|ART_AND_DESIGN|   3.9|    967|   500,000+|Free|    0|Art & Desig
n;Pret...|
|U Launcher Lite –...|ART_AND_DESIGN|   4.7|  87510| 5,000,000+|Free|    0|        Art
& Design|
|Sketch - Draw & P...|ART_AND_DESIGN|   4.5| 215644|50,000,000+|Free|    0|        Art
& Design|
|Pixel Draw - Numb...|ART_AND_DESIGN|   4.3|    967|   100,000+|Free|    0|Art & Desig
n;Crea...|
|Paper flowers ins...|ART_AND_DESIGN|   4.4|    167|    50,000+|Free|    0|        Art
& Design|
|Smoke Effect Phot...|ART_AND_DESIGN|   3.8|    178|    50,000+|Free|    0|        Art
& Design|
|    Infinite Painter|ART_AND_DESIGN|   4.1|  36815| 1,000,000+|Free|    0|        Art
& Design|
|Garden Coloring Book|ART_AND_DESIGN|   4.4|  13791| 1,000,000+|Free|    0|        Art
& Design|
|Kids Paint Free –...|ART_AND_DESIGN|   4.7|    121|    10,000+|Free|    0|Art & Desig
n;Crea...|
|Text on Photo - F...|ART_AND_DESIGN|   4.4|  13880| 1,000,000+|Free|    0|        Art
& Design|
|Name Art Photo Ed...|ART_AND_DESIGN|   4.4|   8788| 1,000,000+|Free|    0|        Art
& Design|
|Tattoo Name On My...|ART_AND_DESIGN|   4.2|  44829|10,000,000+|Free|    0|        Art
& Design|
|Mandala Coloring ...|ART_AND_DESIGN|   4.6|   4326|   100,000+|Free|    0|        Art
& Design|
|3D Color Pixel by...|ART_AND_DESIGN|   4.4|   1518|   100,000+|Free|    0|        Art
& Design|
|Learn To Draw Kaw...|ART_AND_DESIGN|   3.2|     55|     5,000+|Free|    0|        Art
& Design|
|Photo Designer - ...|ART_AND_DESIGN|   4.7|   3632|   500,000+|Free|    0|        Art
& Design|
|350 Diy Room Deco...|ART_AND_DESIGN|   4.5|     27|    10,000+|Free|    0|        Art
& Design|
|FlipaClip - Carto...|ART_AND_DESIGN|   4.3| 194216| 5,000,000+|Free|    0|        Art
& Design|
|        ibis Paint X|ART_AND_DESIGN|   4.6| 224399|10,000,000+|Free|    0|        Art
& Design|
+--------------------+--------------+------+-------+-----------+----+-----+-----------
---------+
only showing top 20 rows
```

In [0]:
```
df.printSchema()
```

```
root
 |-- App: string (nullable = true)
 |-- Category: string (nullable = true)
 |-- Rating: double (nullable = true)
 |-- Reviews: string (nullable = true)
 |-- Installs: string (nullable = true)
 |-- Type: string (nullable = true)
 |-- Price: string (nullable = true)
 |-- Genres: string (nullable = true)
```

In [0]:
```
from pyspark.sql.functions import regexp_replace, col
df = df.withColumn("Reviews", col('Reviews').cast(IntegerType()))\
```

```
    .withColumn("Installs",regexp_replace(col('Installs'),"[^0-9]",""))\
        .withColumn("Installs",col("Installs").cast(IntegerType()))\
            .withColumn("Price",regexp_replace(col('Price'),"[$]",""))\
                .withColumn('Price',col('Price').cast(IntegerType()))
```

In [0]:
```
df.printSchema()
```

```
root
 |-- App: string (nullable = true)
 |-- Category: string (nullable = true)
 |-- Rating: double (nullable = true)
 |-- Reviews: integer (nullable = true)
 |-- Installs: integer (nullable = true)
 |-- Type: string (nullable = true)
 |-- Price: integer (nullable = true)
 |-- Genres: string (nullable = true)
```

In [0]:
```
df.show(10)
```

```
+--------------------+--------------+------+-------+--------+----+-----+--------------
------+
|                 App|      Category|Rating|Reviews|Installs|Type|Price|
Genres|
+--------------------+--------------+------+-------+--------+----+-----+--------------
------+
|Photo Editor & Ca...|ART_AND_DESIGN|   4.1|    159|   10000|Free|    0|        Art &
Design|
| Coloring book moana|ART_AND_DESIGN|   3.9|    967|  500000|Free|    0|Art & Design;P
ret...|
|U Launcher Lite –...|ART_AND_DESIGN|   4.7|  87510| 5000000|Free|    0|        Art &
Design|
|Sketch - Draw & P...|ART_AND_DESIGN|   4.5| 215644|50000000|Free|    0|        Art &
Design|
|Pixel Draw - Numb...|ART_AND_DESIGN|   4.3|    967|  100000|Free|    0|Art & Design;C
rea...|
|Paper flowers ins...|ART_AND_DESIGN|   4.4|    167|   50000|Free|    0|        Art &
Design|
|Smoke Effect Phot...|ART_AND_DESIGN|   3.8|    178|   50000|Free|    0|        Art &
Design|
|    Infinite Painter|ART_AND_DESIGN|   4.1|  36815| 1000000|Free|    0|        Art &
Design|
|Garden Coloring Book|ART_AND_DESIGN|   4.4|  13791| 1000000|Free|    0|        Art &
Design|
|Kids Paint Free –...|ART_AND_DESIGN|   4.7|    121|   10000|Free|    0|Art & Design;C
rea...|
+--------------------+--------------+------+-------+--------+----+-----+--------------
------+
only showing top 10 rows
```

In [0]:
```
df.createOrReplaceTempView('apps')
```

In [0]:
```
%sql Select * from apps
```

App	Category	Rating	Reviews	Installs	Typ
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	10000	Fre
Coloring book moana	ART_AND_DESIGN	3.9	967	500000	Fre
U Launcher Lite – FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	5000000	Fre
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	50000000	Fre
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	100000	Fre
Paper flowers instructions	ART_AND_DESIGN	4.4	167	50000	Fre
Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	50000	Fre

In [0]:
```
%sql Select APP, sum(Reviews) FROM apps
group by 1
ORDER BY 2 desc
```

APP	sum(Reviews)
Instagram	266241989
WhatsApp Messenger	207348304
Clash of Clans	179558781
Messenger – Text and Video Chat for Free	169932272
Subway Surfers	166331958
Candy Crush Saga	156993136
Facebook	156286514
8 Ball Pool	99386198
Clash Royale	92530298
Snapchat	68045010

In [0]:
```
%sql SELECT APP, sum(Installs) FROM Apps
group by 1
ORDER BY 2 DESC
```

APP	sum(Installs)
Subway Surfers	6000000000
Instagram	4000000000
Hangouts	4000000000
Google Drive	4000000000
Google News	4000000000
Google Photos	4000000000
Candy Crush Saga	3500000000
Google Chrome: Fast & Secure	3000000000
WhatsApp Messenger	3000000000
Messenger – Text and Video Chat for Free	3000000000

In [0]:
```
%sql SELECT Category, sum(Installs) FROM Apps
```

```
group by 1
ORDER BY 2 DESC
```

Category	sum(Installs)
GAME	35086024415
COMMUNICATION	32647276251
PRODUCTIVITY	14176091369
SOCIAL	14069867902
TOOLS	11452771915
FAMILY	10258263505
PHOTOGRAPHY	10088247655
NEWS_AND_MAGAZINES	7496317760
TRAVEL_AND_LOCAL	6868887146
VIDEO_PLAYERS	6222002720

In [0]:
```
%sql SELECT APP,sum(Price) FROM Apps
WHERE Type="Paid"
GROUP BY 1
ORDER BY 2 DESC
```

APP	sum(Price)
I'm Rich - Trump Edition	400
I am Rich Plus	399
I AM RICH PRO PLUS	399
I'm Rich/Eu sou Rico/أنا غني/我很有錢	399
I Am Rich Premium	399
most expensive app (H)	399
I Am Rich Pro	399
I am rich(premium)	399
I am Rich	399
I am Rich!	399

In [0]:
```
```