

(https://databricks.com)
Import Library

```
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.types import StructType,StructField, StringType, IntegerType
from pyspark.sql.functions import *
```

Init Dataframe

```
df = spark.read.load('/FileStore/tables/googlestore.csv',format='csv',header='true',sep=',',escape="'" )
```

```
df.count()
```

Out[3]: 10841

```
df.show(10)
```

App		Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated
Current Ver	Android Ver										
Photo Editor & Ca...		ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018
1.0.0	4.0.3 and up										
Coloring book moana		ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pret...	January 15, 2018
2.0.0	4.0.3 and up										
U Launcher Lite â		ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018
1.2.4	4.0.3 and up										
Sketch – Draw & P...		ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018
with device											Varies
4.2 and up											
Pixel Draw – Numb...		ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Crea...	June 20, 2018
1.1	4.4 and up										
Paper flowers ins...		ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017
1	2.3 and up										
Smoke Effect Phot...		ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	April 26, 2018
1.1	4.0.3 and up										
Infinite Painter		ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	June 14, 2018

Schema

```
df.printSchema()
```

```
root
|-- App: string (nullable = true)
|-- Category: string (nullable = true)
|-- Rating: string (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

```
# Since Size of the app is a non significant attribute. we can drop that
```

```
df= df.drop("Size","Content Rating");
```

```
df.printSchema()
```

```
df.show(1)
```

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

```
|-- Current Ver: string (nullable = true)
|-- Android Ver: string (nullable = true)

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| App | Category | Rating | Reviews | Installs | Type | Price | Genres | Last Updated | Current Ver | Android Ver |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Photo Editor & Ca... | ART_AND_DESIGN | 4.1 | 159 | 10,000+ | Free | 0 | Art & Design | January 7, 2018 | 1.0.0 | 4.0.3 and up |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

only showing top 1 row
```

```
df=df.withColumn("Reviews",col("Reviews").cast(IntegerType()))

df.printSchema()

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

```
df.show(5)

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| App | Category | Rating | Reviews | Installs | Type | Price | Genres | Last Updated | Current Ver | Android Ver |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Photo Editor & Ca... | ART_AND_DESIGN | 4.1 | 159 | 10,000+ | Free | 0 | Art & Design | January 7, 2018 | 1.0.0 | 4.0.3 and up |
| Coloring book moana | ART_AND_DESIGN | 3.9 | 967 | 500,000+ | Free | 0 | Art & Design; Pret... | January 15, 2018 | 2.0.0 | 4.0.3 and up |
| U Launcher Lite â€” FREE Live Cool Themes, Hide Apps | ART_AND_DESIGN | 4.7 | 87510 | 5,000,000+ | Free | 0 | Art & Design | August 1, 2018 | 1.2.4 | 4.0.3 and up |
| Sketch - Draw & Paint | ART_AND_DESIGN | 4.5 | 215644 | 50,000,000+ | Free | 0 | Art & Design | June 8, 2018 | Varies with device | 4.2 and up |
| Pixel Draw - Number Art Coloring Book | ART_AND_DESIGN | 4.3 | 967 | 100,000+ | Free | 0 | Art & Design; Crea... | June 20, 2018 | 1.1 | 4.4 and up |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

only showing top 5 rows
```

```
df.createOrReplaceTempView("apps")
```

```
%sql select * from apps
```

Table					
	App	Category	Rating	Review	
1	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	
2	Coloring book moana	ART_AND_DESIGN	3.9	967	
3	U Launcher Lite – FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	
4	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	21564	
5	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	
6	Paper flowers instructions	ART_AND_DESIGN	4.4	167	
7	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	
10,000 rows Truncated data					

Top Reviews Given

```
%sql select App,sum(Reviews) from apps
group by App
order by sum(Reviews) desc
```

Table		
	App	sum(Reviews)
1	Instagram	266241989
2	WhatsApp Messenger	207348304
3	Clash of Clans	179558781
4	Messenger â€” Text and Video Chat for Free	169932272
5	Subway Surfers	166331958
6	Candy Crush Saga	156993136
7	Facebook	156286514
9,660 rows		

df=df.withColumn("Installs",col("Installs").cast(IntegerType()))

Top 10 Installed Apps

```
%sql
select App,installs from apps
order by installs desc
limit 10
```

Table		
	App ▲	installs ▲
1	Life Made WI-Fi Touchscreen Photo Frame	Free
2	Viber Messenger	500,000,000+
3	imo free video calls and chat	500,000,000+
4	imo free video calls and chat	500,000,000+
5	Google Duo - High Quality Video Calls	500,000,000+
6	UC Browser - Fast Download Private & Secure	500,000,000+
7	imo free video calls and chat	500,000,000+
10 rows		

Top Paid Apps

Table		
	App ▲	Price ▲
1	ðŸŽ“ I'm rich	\$399.99
2	Ð'Ð,Ð»ÐµÑ,Ñ« ÐŸÐ"Ð" CD 2019 PRO	\$1.49
3	Ã‰galitÃ© et RÃ©conciliation	\$2.99
4	Â¡Ay Caramba!	\$1.99
5	weather HD	\$1.99
6	tTorrent - ad free	\$1.99
7	sugar_sugar	\$1.20
800 rows		