```
import os
# Find the latest version of spark 3.0  from http://www-us.apache.org/dist/spark/ and
# For example:
# spark_version = 'spark-3.0.0'
spark_version = 'spark-3.1.2'
os.environ['SPARK_VERSION']=spark_version

# Install Spark and Java
!apt-get update
!apt-get install openjdk-11-jdk-headless -qq > /dev/null
!wget -q http://www-us.apache.org/dist/spark/$SPARK_VERSION/$SPARK_VERSION-bin-hadoop
!tar xf $SPARK_VERSION-bin-hadoop2.7.tgz
!pip install -q findspark

# Set Environment Variables
import os
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-11-openjdk-amd64"
os.environ["SPARK_HOME"] = f"/content/{spark_version}-bin-hadoop2.7"

# Start a SparkSession
import findspark
findspark.init()

# Other imports
from pyspark.sql.functions import count
```

```
    Ign:1 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
    Get:2 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.7 kB]
    Ign:3 https://developer.download.nvidia.com/compute/machine-learning/repos/ubunt
    Get:4 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
    Hit:5 https://developer.download.nvidia.com/compute/machine-learning/repos/ubunt
    Get:6 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
    Hit:7 http://archive.ubuntu.com/ubuntu bionic InRelease
    Hit:8 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic InRelease
    Get:9 http://archive.ubuntu.com/ubuntu bionic-updates InRelease [88.7 kB]
    Hit:11 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ InRelease
    Hit:12 http://ppa.launchpad.net/cran/libgit2/ubuntu bionic InRelease
    Ign:13 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_6
    Get:13 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_6
    Get:14 http://ppa.launchpad.net/deadsnakes/ppa/ubuntu bionic InRelease [15.9 kB]
    Get:15 http://archive.ubuntu.com/ubuntu bionic-backports InRelease [74.6 kB]
    Get:16 http://security.ubuntu.com/ubuntu bionic-security/main amd64 Packages [2,
    Hit:17 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic InRelease
    Get:18 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 Packages
    Get:19 http://archive.ubuntu.com/ubuntu bionic-updates/universe amd64 Packages [
    Get:20 http://ppa.launchpad.net/deadsnakes/ppa/ubuntu bionic/main amd64 Packages
    Get:21 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 Packages [2,61
    Fetched 9,312 kB in 4s (2,613 kB/s)
    Reading package lists... Done
```

```
# Download the Postgres driver that will allow Spark to interact with Postgres.
```

```
!wget https://jdbc.postgresql.org/download/postgresql-42.2.16.jar
```

```
--2021-06-20 21:25:23--  https://jdbc.postgresql.org/download/postgresql-42.2.16
Resolving jdbc.postgresql.org (jdbc.postgresql.org)... 72.32.157.228, 2001:4800:
Connecting to jdbc.postgresql.org (jdbc.postgresql.org)|72.32.157.228|:443... co
HTTP request sent, awaiting response... 200 OK
Length: 1002883 (979K) [application/java-archive]
Saving to: ‘postgresql-42.2.16.jar’

postgresql-42.2.16. 100%[===================>] 979.38K  4.75MB/s    in 0.2s

2021-06-20 21:25:23 (4.75 MB/s) - ‘postgresql-42.2.16.jar’ saved [1002883/100288
```

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName("BigData-Challenge").config("spark.driver.extraCl
```

```
from pyspark import SparkFiles
url = "https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Wireless_v1_
spark.sparkContext.addFile(url)
df = spark.read.option("encoding", "UTF-8").csv(SparkFiles.get(""), sep="\t", header=
df.show()
```

```
+-----------+-----------+--------------+----------+--------------+---------------
|marketplace|customer_id|     review_id|product_id|product_parent|       product_
+-----------+-----------+--------------+----------+--------------+---------------
|         US|   16414143|R3W4P9UBGNGH1U|B00YL0EKWE|     852431543|LG G4 Case Har
|         US|   50800750|R15V54KBMTQWAY|B00XK95RPQ|     516894650|Selfie Stick F
|         US|   15184378| RY8I449HNXSVF|B00SXRXUKO|     984297154|Tribe AB40 Wat
|         US|   10203548|R18TLJYCKJFLSR|B009V5X1CE|     279912704|RAVPower® Elem
|         US|     488280|R1NK26SWS53B8Q|B00D93OVF0|     662791300|Fosmon Micro U
|         US|   13334021|R11LOHEDYJALTN|B00XVGJMDQ|     421688488|iPhone 6 Case,
|         US|   27520697|R3ALQVQB2P9LA7|B00KQW1X1C|     554285554|Nokia Lumia 63
|         US|   48086021|R3MWLXLNO21PDQ|B00IP1MQNK|     488006702|Lumsing 10400m
|         US|   12738196|R2L15IS24CX0LI|B00HVORET8|     389677711|iPhone 5S Batt
|         US|   15867807|R1DJ8976WPWVZU|B00HX3G6J6|     299654876|HTC One M8 Scr
|         US|    1972249|R3MRWNNR8CBTB7|B00U4NATNQ|     577878727|S6 Case - Bear
|         US|   10956619|R1DS6DKTUXAQK3|B00SZEFDH8|     654620704|BLU Studio X,
|         US|   14805911| RWJM5E0TWUJD2|B00JRJUL9U|     391166958|EZOPower 5-Por
|         US|   15611116|R1XTJKDYNCRGAC|B00KQ4T0HE|     481551630|iPhone 6S Case
|         US|   39298603|R2UZL3DPWEU1XW|B00M0YWKPM|     685107474| iPhone 6s Plu
|         US|   17552454|R2EZXET9KBFFU3|B00KDZEE68|     148320945|zBoost ZB575-A
|         US|   12218556|R26VY1L1FD3LPU|B00BJN45GM|      47788188|OtterBox Defen
|         US|   21872923|R2SSA4NSFCV18T|B00SA86SXW|     748759272|Aduro PowerUP
|         US|   16264332|R1G6333JHJNEUQ|B00Q3I68TU|     974085141|LilGadgets Con
|         US|    6042304|R2DRG0UZXJQ0PE|B00TN4J1TA|     716174627|Anker Aluminum
+-----------+-----------+--------------+----------+--------------+---------------
only showing top 20 rows
```

```
# Filter df on total_vote greater than or equal to 20
totalvotes_df = df.filter("total_votes>=20")
```

```
totalvotes_df.show()
```

```
+-----------+-----------+--------------+----------+--------------+----------------
|marketplace|customer_id|     review_id|product_id|product_parent|        product_
+-----------+-----------+--------------+----------+--------------+----------------
|         US|   44689470|R2WOW0TURNXB26|B00YY3UBV2|     310491927|              Ga
|         US|     112342|R13VL62Y2HBQ0B|B010VFZJD6|     129632031|iTaste MVP3 PR
|         US|   13557708|R22G55KAPZKJQV|B00C8S8S4W|     998105706|SPOT 3 Satelli
|         US|    3346419|R1610PGTJS7G3N|B011I4XMXS|     431799284|   Ausdom Dash
|         US|   46029442| RLQL04BL0QXOJ|B00OSTKZWM|     736471392|RCA 5.5-Inch Q
|         US|   19380011|R2AYJHH8WJNGAU|B013D32WVA|     138975975|Samsung Galaxy
|         US|   22953177|R111DJA10Y6CMU|B013BHLU66|     396362963|Waterproof Cas
|         US|   16980808|R2EE2TR4MRDV0U|B00Y1Z87UU|     956867660|IBESTWIN Li-io
|         US|   37339397| RD4A80I5JDHED|B00UY29N8Y|     384094161|IncrediSonic M
|         US|    7830540|R1GU6IYZQWQE8X|B00NPZG6DW|     267673298|      Parrot Z
|         US|     113760| RZOPM62JMW97V|B00X0X3EQ6|     657509542|OtterBox SYMME
|         US|   52589608|R21GLR3TD27ISV|B013IVO8FK|      84781086|NomadPlus Gene
|         US|   35621482|R1D3NR5GREEXXJ|B013F8C9X4|     561786475|Galaxy Note 5 
|         US|   45872422|R24BMEHX5EWGEY|B00Z9P06DW|     459992506|ATian 9 inch T
|         US|   44137838|R2WYKBQS8OR08O|B010GYYSU2|     518226514|Z-Edge 2.4- in
|         US|   11179629|R3DL0Y1KWYLD5X|B010LXUQNG|     179041214|iPhone 6 Case,
|         US|   21851130|R3EGDTFDMCOOG4|B010MP3K0O|     868762206|Jackery Titan 
|         US|     459473|R1MC93W6WG9R3O|B00TRNCN5Q|     736335718|Galaxy Grand P
|         US|    5374752|R35B0B69DYR54L|B00TRC3YF0|     396308462|Rexing F9 2.7”
|         US|   29179205|R1593EM56412NH|B011VRO5M4|      28128248|iPhone 6S Plus
+-----------+-----------+--------------+----------+--------------+----------------
only showing top 20 rows
```

```
helpful_greater_50_df = totalvotes_df.filter("helpful_votes/total_votes>=.50")
helpful_greater_50_df.show()
```

```
+-----------+-----------+--------------+----------+--------------+----------------
|marketplace|customer_id|     review_id|product_id|product_parent|        product_
+-----------+-----------+--------------+----------+--------------+----------------
|         US|   44689470|R2WOW0TURNXB26|B00YY3UBV2|     310491927|              Ga
|         US|     112342|R13VL62Y2HBQ0B|B010VFZJD6|     129632031|iTaste MVP3 PR
|         US|   13557708|R22G55KAPZKJQV|B00C8S8S4W|     998105706|SPOT 3 Satelli
|         US|    3346419|R1610PGTJS7G3N|B011I4XMXS|     431799284|   Ausdom Dash
|         US|   46029442| RLQL04BL0QXOJ|B00OSTKZWM|     736471392|RCA 5.5-Inch Q
|         US|   19380011|R2AYJHH8WJNGAU|B013D32WVA|     138975975|Samsung Galaxy
|         US|   22953177|R111DJA10Y6CMU|B013BHLU66|     396362963|Waterproof Cas
|         US|   16980808|R2EE2TR4MRDV0U|B00Y1Z87UU|     956867660|IBESTWIN Li-io
|         US|   37339397| RD4A80I5JDHED|B00UY29N8Y|     384094161|IncrediSonic M
|         US|    7830540|R1GU6IYZQWQE8X|B00NPZG6DW|     267673298|      Parrot Z
|         US|     113760| RZOPM62JMW97V|B00X0X3EQ6|     657509542|OtterBox SYMME
|         US|   35621482|R1D3NR5GREEXXJ|B013F8C9X4|     561786475|Galaxy Note 5 
|         US|   45872422|R24BMEHX5EWGEY|B00Z9P06DW|     459992506|ATian 9 inch T
|         US|   44137838|R2WYKBQS8OR08O|B010GYYSU2|     518226514|Z-Edge 2.4- in
|         US|   11179629|R3DL0Y1KWYLD5X|B010LXUQNG|     179041214|iPhone 6 Case,
|         US|   21851130|R3EGDTFDMCOOG4|B010MP3K0O|     868762206|Jackery Titan 
|         US|     459473|R1MC93W6WG9R3O|B00TRNCN5Q|     736335718|Galaxy Grand P
|         US|    5374752|R35B0B69DYR54L|B00TRC3YF0|     396308462|Rexing F9 2.7”
|         US|   29179205|R1593EM56412NH|B011VRO5M4|      28128248|iPhone 6S Plus
|         US|   14971124|R1KEP1DUJK2LD5|B00UH3L82Y|     792159590|Armorsuit - Ap
+-----------+-----------+--------------+----------+--------------+---------------
only showing top 20 rows
```

```
vine_review_df = helpful_greater_50_df.filter(helpful_greater_50_df["vine"] == "Y")
vine_review_df.show()
```

```
+-----------+-----------+--------------+----------+--------------+---------------
|marketplace|customer_id|     review_id|product_id|product_parent|       product_
+-----------+-----------+--------------+----------+--------------+---------------
|         US|   48852155|R1MAOLI5FJHAFM|B013X0V11K|     610966690|BLU Studio 7.0
|         US|   11556116| R9PYAUDIBJVEC|B013X0V4VM|     672788343|BLU Studio C S
|         US|   46671309| R6V9SHMMG5M8F|B013X0V11K|     610966690|BLU Studio 7.0
|         US|   49598970|R37PVLT6ELL5J4|B011HT9AL2|     557568833|Tile (Gen 2) -
|         US|   52057325| R2FSFGWZF24V9|B0129T0XXS|     592405023|BLU Studio C 5-
|         US|   53019847|R3SRW1E8J56IGV|B0129TQLPW|     226174255|BLU Energy X P
|         US|   31302915| R86Z11D4CWOFM|B0129T0XXS|     592405023|BLU Studio C 5-
|         US|   50885906| RNP01HW9YISJO|B00W7S34HY|     920588860|Optrix  Waterp
|         US|   49110251|R3KLACA6LCDZ0S|B00W75BKQ4|     566439653|Motorola T460
|         US|   50125011| RZEQYOT2RE0N7|B0129T0XXS|     592405023|BLU Studio C 5-
|         US|   39749647|R2WBPX441TH495|B0129TQLPW|     226174255|BLU Energy X P
|         US|   53058973|R2BYBSYHS66ZN8|B0129T0XXS|     592405023|BLU Studio C 5-
|         US|   12537483|R3IF59PJGCNU3Q|B011YNPPME|     242359747|BLU Vivo Selfi
|         US|   35304626|R2IXC6U7W4OCQ9|B0129TQLPW|     226174255|BLU Energy X P
|         US|   30057302|R1JEI3H9QRP6PH|B011YNPPME|     242359747|BLU Vivo Selfi
|         US|   52109863|R37L3KGRRR6JTL|B00MCJ4CKG|      91423181|Recon Jet Head
|         US|   43791073|R1YJ7OKAEML92P|B00TYTBHKU|     399571814|OtterBox Symme
|         US|   50441881|R3JZJOD2512UVY|B0102OM1IC|     845336843|Sony Xperia M4
|         US|   52591230|R38MY3TK17MXDH|B00N9E6DUK|     272244000|JAWBONE UP3 Ac
|         US|   52594065|R2LWISZ4DSM0I4|B0102OM1IC|     845336843|Sony Xperia M4
+-----------+-----------+--------------+----------+--------------+---------------
only showing top 20 rows
```

```
no_vine_review_df = helpful_greater_50_df.filter(helpful_greater_50_df["vine"] == "N"
no_vine_review_df.show()
```

```
+-----------+-----------+--------------+----------+--------------+---------------
|marketplace|customer_id|     review_id|product_id|product_parent|       product_
+-----------+-----------+--------------+----------+--------------+---------------
|         US|   44689470|R2WOW0TURNXB26|B00YY3UBV2|     310491927|             Ga
|         US|     112342|R13VL62Y2HBQ0B|B010VFZJD6|     129632031|iTaste MVP3 PR
|         US|   13557708|R22G55KAPZKJQV|B00C8S8S4W|     998105706|SPOT 3 Satelli
|         US|    3346419|R1610PGTJS7G3N|B011I4XMXS|     431799284|   Ausdom Dash
|         US|   46029442| RLQL04BL0QXOJ|B00OSTKZWM|     736471392|RCA 5.5-Inch Q
|         US|   19380011|R2AYJHH8WJNGAU|B013D32WVA|     138975975|Samsung Galaxy
|         US|   22953177|R111DJA10Y6CMU|B013BHLU66|     396362963|Waterproof Cas
|         US|   16980808|R2EE2TR4MRDV0U|B00Y1Z87UU|     956867660|IBESTWIN Li-io
|         US|   37339397| RD4A80I5JDHED|B00UY29N8Y|     384094161|IncrediSonic M
|         US|    7830540|R1GU6IYZQWQE8X|B00NPZG6DW|     267673298|       Parrot Z
|         US|     113760| RZOPM62JMW97V|B00X0X3EQ6|     657509542|OtterBox SYMME
|         US|   35621482|R1D3NR5GREEXXJ|B013F8C9X4|     561786475|Galaxy Note 5
|         US|   45872422|R24BMEHX5EWGEY|B00Z9P06DW|     459992506|ATian 9 inch T
|         US|   44137838|R2WYKBQS8OR08O|B010GYYSU2|     518226514|Z-Edge 2.4- in
|         US|   11179629|R3DL0Y1KWYLD5X|B010LXUQNG|     179041214|iPhone 6 Case,
|         US|   21851130|R3EGDTFDMCOOG4|B010MP3K0O|     868762206|Jackery Titan 
|         US|     459473|R1MC93W6WG9R3O|B00TRNCN5Q|     736335718|Galaxy Grand P
|         US|    5374752|R35B0B69DYR54L|B00TRC3YF0|     396308462|Rexing F9 2.7”
|         US|   29179205|R1593EM56412NH|B011VRO5M4|      28128248|iPhone 6S Plus
|         US|   14971124|R1KEP1DUJK2LD5|B00UH3L82Y|     792159590|Armorsuit - Ap
+-----------+-----------+--------------+----------+--------------+--------------
only showing top 20 rows
```

PAID (VINE) & UN-PAID (NO-VINE) COUNT

```
vine_review_count = vine_review_df.count()
vine_review_count
```

```
613
```

```
no_vine_review_count = no_vine_review_df.count()
no_vine_review_count
```

```
64968
```

5 STAR REVIEWS VINE & NO-VINE

```
vine_review_5_star = vine_review_df.filter(vine_review_df["star_rating"]=="5")
vine_review_5_star_count = vine_review_5_star.count()
vine_review_5_star_count
```

```
222
```

```
no_vine_review_5_star = no_vine_review_df.filter(no_vine_review_df["star_rating"]=="5'
no_vine_review_5_star_count = no_vine_review_5_star.count()
no_vine_review_5_star_count
```

```
30543
```

% OF 5-STAR REVIEWS VINE & NO-VINE

```
vine_5star_pct = (vine_review_5_star_count/vine_review_count)*100
vine_5star_pct
```

```
36.215334420880914
```

```
no_vine_5star_pct = (no_vine_review_5_star_count/no_vine_review_count)*100
```

```
no_vine_5star_pct
```

```
47.01237532323606
```

TOTAL NUMBER OF REVIEWS

```
total_reviews_count = df.count()
total_reviews_count
```

```
9002021
```

✓ 33s completed at 6:04 PM ● ✕