

##SALES DATA ANALYSIS

###Setup for the pySpark

```
!apt-get install openjdk-8-jdk-headless -qq > /dev/null
!wget -q http://archive.apache.org/dist/spark/spark-3.1.1/spark-3.1.1-
bin-hadoop3.2.tgz
!tar xf spark-3.1.1-bin-hadoop3.2.tgz
!pip install -q findspark

import os
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-8-openjdk-amd64"
os.environ["SPARK_HOME"] = "/content/spark-3.1.1-bin-hadoop3.2"

import findspark
findspark.init()
from pyspark.sql import SparkSession
spark = SparkSession.builder.master("local[*]").getOrCreate()
#for MaxRows
#spark.conf.set('spark.sql.repl.eagerEval.maxNumRows', True)
#To Enable the Spark Rows Show
spark.conf.set('spark.sql.repl.eagerEval.enabled', True)
#spark.conf.set("spark.sql.repl.eagerEval.enabled", True) # Property
used to format output tables better
```

###Importing Libraries

```
from pyspark.sql import functions as F
from functools import reduce
from operator import add
from pyspark.sql.functions import col
from pyspark.sql import DataFrame
import plotly.express as px
import pandas as pd
import numpy as np

spark = SparkSession.builder \
    .appName("SalesDataAnalysis") \
    .getOrCreate()
```

###Importing Sales Dataest

```
sales_data = spark.read.csv("sales_data.csv", header=True,
inferSchema=True)
```

The given dataset contains total 2824 rows and 25 columns.

###Analysing Dataset

```
sales_data.printSchema()
```

```
#Before Data Cleaning
```

```
sales_data.show()
```

```
root
```

```
|-- ORDERNUMBER: integer (nullable = true)
|-- QUANTITYORDERED: integer (nullable = true)
|-- PRICEEACH: double (nullable = true)
|-- ORDERLINENUMBER: integer (nullable = true)
|-- SALES: double (nullable = true)
|-- ORDERDATE: string (nullable = true)
|-- STATUS: string (nullable = true)
|-- QTR_ID: integer (nullable = true)
|-- MONTH_ID: integer (nullable = true)
|-- YEAR_ID: integer (nullable = true)
|-- PRODUCTLINE: string (nullable = true)
|-- MSRP: integer (nullable = true)
|-- PRODUCTCODE: string (nullable = true)
|-- CUSTOMERNAME: string (nullable = true)
|-- PHONE: string (nullable = true)
|-- ADDRESSLINE1: string (nullable = true)
|-- ADDRESSLINE2: string (nullable = true)
|-- CITY: string (nullable = true)
|-- STATE: string (nullable = true)
|-- POSTALCODE: string (nullable = true)
|-- COUNTRY: string (nullable = true)
|-- TERRITORY: string (nullable = true)
|-- CONTACTLASTNAME: string (nullable = true)
|-- CONTACTFIRSTNAME: string (nullable = true)
|-- DEALSIZE: string (nullable = true)
```

```
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|ORDERNUMBER|QUANTITYORDERED|PRICEEACH|ORDERLINENUMBER|  SALES|
ORDERDATE|  STATUS|QTR_ID|MONTH_ID|YEAR_ID|PRODUCTLINE|MSRP|
PRODUCTCODE|      CUSTOMERNAME|      PHONE|
ADDRESSLINE1|ADDRESSLINE2|      CITY|  STATE|POSTALCODE|  COUNTRY|
TERRITORY|CONTACTLASTNAME|CONTACTFIRSTNAME|DEALSIZE|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|      10107|      30|      95.7|      2| 2871.0|
2/24/2003 0:00|Shipped|      1|      2|    2003|Motorcycles| 95|
S10_1678|    Land of Toys Inc.|      2125557818|897 Long Airport ...|
```

Yu	NYC	NY	10022	USA	NA
	Kwai	Small			
	10121	34	81.35	5	2765.9
5/7/2003 0:00	Shipped	2	5	2003	Motorcycles 95
S10_1678	Reims Collectables		26.47.1555	59 rue de l'Abbaye	
null	Reims	null	51100	France	EMEA
Henriot	Paul	Small			
	10134	41	94.74	2	3884.34
7/1/2003 0:00	Shipped	3	7	2003	Motorcycles 95
S10_1678	Lyon Souvenirs	+33 1 46 62 7555	27 rue du Colonel...		
null	Paris	null	75508	France	EMEA Da
Cunha	Daniel	Medium			
	10145	45	83.26	6	3746.7
8/25/2003 0:00	Shipped	3	8	2003	Motorcycles 95
S10_1678	Toys4GrownUps.com		6265557265	78934 Hillside Dr.	
null	Pasadena	CA	90003	USA	NA
Young	Julie	Medium			
	10159	49	100.0	14	5205.27
10/10/2003 0:00	Shipped	4	10	2003	Motorcycles 95
S10_1678	Corporate Gift Id...		6505551386	7734 Strong St.	
null	San Francisco	CA	null	USA	NA
Brown	Julie	Medium			
	10168	36	96.66	1	3479.76
10/28/2003 0:00	Shipped	4	10	2003	Motorcycles 95
S10_1678	Technics Stores Inc.		6505556809	9408 Furth Circle	
null	Burlingame	CA	94217	USA	NA
Hirano	Juri	Medium			
	10180	29	86.13	9	2497.77
11/11/2003 0:00	Shipped	4	11	2003	Motorcycles 95
S10_1678	Daedalus Designs ...		20.16.1555	184, chausse de T...	
null	Lille	null	59000	France	EMEA
Rance	Martine	Small			
	10188	48	100.0	1	5512.32
11/18/2003 0:00	Shipped	4	11	2003	Motorcycles 95
S10_1678	Herkku Gifts		+47 2267 3215	Drammen 121, PR 7...	
null	Bergen	null	N 5804	Norway	EMEA
Oeztan	Veysel	Medium			
	10201	22	98.57	2	2168.54
12/1/2003 0:00	Shipped	4	12	2003	Motorcycles 95
S10_1678	Mini Wheels Co.		6505555787	5557 North Penda...	
null	San Francisco	CA	null	USA	NA
Murphy	Julie	Small			
	10211	41	100.0	14	4708.44
1/15/2004 0:00	Shipped	1	1	2004	Motorcycles 95
S10_1678	Auto Canal Petit	(1) 47.55.6555	25, rue Lauriston		
null	Paris	null	75016	France	EMEA
Perrier	Dominique	Medium			
	10223	37	100.0	1	3965.66
2/20/2004 0:00	Shipped	1	2	2004	Motorcycles 95

S10_1678	Australian Collec...	03 9520 4555	636 St Kilda Road
Level 3	Melbourne Victoria	3004 Australia	APAC
Ferguson	Peter	Medium	
10237	23	100.0	7 2333.12
4/5/2004 0:00	Shipped	2	4 2004 Motorcycles 95
S10_1678	Vitachrome Inc.	2125551500	2678 Kingston Rd.
Suite 101	NYC	NY	10022 USA NA
Frick	Michael	Small	
10251	28	100.0	2 3188.64
5/18/2004 0:00	Shipped	2	5 2004 Motorcycles 95
S10_1678	Tekni Collectable...	2015559350	7476 Moss Rd.
null	Newark	NJ	94019 USA NA
Brown	William	Medium	
10263	34	100.0	2 3676.76
6/28/2004 0:00	Shipped	2	6 2004 Motorcycles 95
S10_1678	Gift Depot Inc.	2035552570	25593 South Bay Ln.
null	Bridgewater	CT	97562 USA NA
King	Julie	Medium	
10275	45	92.83	1 4177.35
7/23/2004 0:00	Shipped	3	7 2004 Motorcycles 95
S10_1678	La Rochelle Gifts	40.67.8555	67, rue des Cinqu...
null	Nantes	null	44000 France EMEA
Labrune	Janine	Medium	
10285	36	100.0	6 4099.68
8/27/2004 0:00	Shipped	3	8 2004 Motorcycles 95
S10_1678	Marta's Replicas Co.	6175558555	39323 Spinnaker Dr.
null	Cambridge	MA	51247 USA NA
Hernandez	Marta	Medium	
10299	23	100.0	9 2597.39
9/30/2004 0:00	Shipped	3	9 2004 Motorcycles 95
S10_1678	Toys of Finland, Co.	90-224 8555	Keskuskatu 45
null	Helsinki	null	21240 Finland EMEA
Karttunen	Matti	Small	
10309	41	100.0	5 4394.38
10/15/2004 0:00	Shipped	4	10 2004 Motorcycles 95
S10_1678	Baane Mini Imports	07-98 9555	Erling Skakkes ga...
null	Stavern	null	4110 Norway EMEA
Bergulfsen	Jonas	Medium	
10318	46	94.74	1 4358.04
11/2/2004 0:00	Shipped	4	11 2004 Motorcycles 95
S10_1678	Diecast Classics ...	2155551555	7586 Pompton St.
null	Allentown	PA	70267 USA NA
Yu	Kyung	Medium	
10329	42	100.0	1 4396.14
11/15/2004 0:00	Shipped	4	11 2004 Motorcycles 95
S10_1678	Land of Toys Inc.	2125557818	897 Long Airport ...
null	NYC	NY	10022 USA NA
Yu	Kwai	Medium	

+-----+-----+-----+-----+-----+

```

+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

```

###Dropping NULL Values

```
cleaned_sales_data=sales_data.na.drop()
```

#After Data cleaning

```
cleaned_sales_data.show()
```

```

+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
|ORDERNUMBER|QUANTITYORDERED|PRICEEACH|ORDERLINENUMBER|  SALES|
ORDERDATE| STATUS|QTR_ID|MONTH_ID|YEAR_ID| PRODUCTLINE|MSRP|
PRODUCTCODE|      CUSTOMERNAME|      PHONE|      ADDRESSLINE1|
ADDRESSLINE2|      CITY|      STATE|POSTALCODE|  COUNTRY|TERRITORY|
CONTACTLASTNAME|CONTACTFIRSTNAME|DEALSIZE|
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
|      10223|      37|      100.0|      1|3965.66|
2/20/2004 0:00|Shipped|      1|      2|      2004| Motorcycles| 95|
S10_1678|Australian Collec...|      03 9520 4555|      636 St Kilda Road|
Level 3|      Melbourne|Victoria|      3004|Australia|      APAC|
Ferguson|      Peter|      Medium|
|      10237|      23|      100.0|      7|2333.12|
4/5/2004 0:00|Shipped|      2|      4|      2004| Motorcycles| 95|
S10_1678|      Vitachrome Inc.|      2125551500|      2678 Kingston Rd.|
Suite 101|      NYC|      NY|      10022|      USA|      NA|
Frick|      Michael|      Small|
|      10361|      20|      72.55|      13| 1451.0|
12/17/2004 0:00|Shipped|      4|      12|      2004| Motorcycles| 95|
S10_1678|Souvenirs And Th...|+61 2 9495 8555|Monitor Money Bui...|
Level 6|      Chatswood|      NSW|      2067|Australia|      APAC|
Huxley|      Adrian|      Small|
|      10163|      21|      100.0|      1|4860.24|
10/20/2003 0:00|Shipped|      4|      10|      2003|Classic Cars| 214|
S10_1949|Classic Legends Inc.|      2125558493|      5905 Pompton St.|
Suite 750|      NYC|      NY|      10022|      USA|      NA|
Hernandez|      Maria|      Medium|

```

	10270	21	100.0	9	4905.39
7/19/2004 0:00	Shipped	3	7	2004	Classic Cars 214
S10_1949	Souvenirs And Th...	+61 2 9495 8555	Monitor Money Bui...		
Level 6	Chatswood	NSW	2067	Australia	APAC
Huxley	Adrian	Medium			
	10347	30	100.0	1	3944.7
11/29/2004 0:00	Shipped	4	11	2004	Classic Cars 214
S10_1949	Australian Collec...	03 9520 4555	636 St Kilda Road		
Level 3	Melbourne	Victoria	3004	Australia	APAC
Ferguson	Peter	Medium			
	10391	24	100.0	4	2416.56
3/9/2005 0:00	Shipped	1	3	2005	Classic Cars 214
S10_1949	Anna's Decoration...	02 9936 8555	201 Miller Street		
Level 15	North Sydney	NSW	2060	Australia	APAC
O'Hara	Anna	Small			
	10120	29	96.34	3	2793.86
4/29/2003 0:00	Shipped	2	4	2003	Motorcycles 118
S10_2016	Australian Collec...	03 9520 4555	636 St Kilda Road		
Level 3	Melbourne	Victoria	3004	Australia	APAC
Ferguson	Peter	Small			
	10223	47	100.0	4	5422.39
2/20/2004 0:00	Shipped	1	2	2004	Motorcycles 118
S10_2016	Australian Collec...	03 9520 4555	636 St Kilda Road		
Level 3	Melbourne	Victoria	3004	Australia	APAC
Ferguson	Peter	Medium			
	10361	26	51.15	8	1329.9
12/17/2004 0:00	Shipped	4	12	2004	Motorcycles 118
S10_2016	Souvenirs And Th...	+61 2 9495 8555	Monitor Money Bui...		
Level 6	Chatswood	NSW	2067	Australia	APAC
Huxley	Adrian	Small			
	10120	46	100.0	2	9264.86
4/29/2003 0:00	Shipped	2	4	2003	Motorcycles 193
S10_4698	Australian Collec...	03 9520 4555	636 St Kilda Road		
Level 3	Melbourne	Victoria	3004	Australia	APAC
Ferguson	Peter	Large			
	10223	49	100.0	3	9774.03
2/20/2004 0:00	Shipped	1	2	2004	Motorcycles 193
S10_4698	Australian Collec...	03 9520 4555	636 St Kilda Road		
Level 3	Melbourne	Victoria	3004	Australia	APAC
Ferguson	Peter	Large			
	10237	39	100.0	9	7023.9
4/5/2004 0:00	Shipped	2	4	2004	Motorcycles 193
S10_4698	Vitachrome Inc.	2125551500	2678 Kingston Rd.		
Suite 101	NYC	NY	10022	USA	NA
Frick	Michael	Large			
	10337	25	48.05	8	1201.25
11/21/2004 0:00	Shipped	4	11	2004	Classic Cars 136
S10_4757	Classic Legends Inc.	2125558493	5905 Pompton St.		
Suite 750	NYC	NY	10022	USA	NA

```

Hernandez|          Maria|  Small|
|      10270|          32|    100.0|          2|4302.08|
7/19/2004 0:00|Shipped|      3|      7|    2004|Classic Cars| 147|
S10_4962|Souvenirs And Th...|+61 2 9495 8555|Monitor Money Bui...|
Level 6|    Chatswood|    NSW|    2067|Australia|    APAC|
Huxley|          Adrian|  Medium|
|      10347|          27|    100.0|          2| 4428.0|
11/29/2004 0:00|Shipped|      4|     11|    2004|Classic Cars| 147|
S10_4962|Australian Collec...|    03 9520 4555|    636 St Kilda Road|
Level 3|    Melbourne|Victoria|    3004|Australia|    APAC|
Ferguson|          Peter|  Medium|
|      10370|          35|    65.63|          4|2297.05|
1/20/2005 0:00|Shipped|      1|      1|    2005|Classic Cars| 147|
S10_4962|Anna's Decoration...|    02 9936 8555|    201 Miller Street|
Level 15|North Sydney|    NSW|    2060|Australia|    APAC|
O'Hara|          Anna|  Small|
|      10391|          37|    46.9|          7| 1735.3|
3/9/2005 0:00|Shipped|      1|      3|    2005|Classic Cars| 147|
S10_4962|Anna's Decoration...|    02 9936 8555|    201 Miller Street|
Level 15|North Sydney|    NSW|    2060|Australia|    APAC|
O'Hara|          Anna|  Small|
|      10169|          30|    100.0|          2| 5019.9|
11/4/2003 0:00|Shipped|      4|     11|    2003|Classic Cars| 194|
S12_1099|Anna's Decoration...|    02 9936 8555|    201 Miller Street|
Level 15|North Sydney|    NSW|    2060|Australia|    APAC|
O'Hara|          Anna|  Medium|
|      10127|          46|    100.0|          2|11279.2|
6/3/2003 0:00|Shipped|      2|      6|    2003|Classic Cars| 207|
S12_1108|    Muscle Machine Inc|    2125557413|    4092 Furth Circle|
Suite 400|          NYC|    NY|    10022|    USA|    NA|
Young|          Jeff|  Large|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+
only showing top 20 rows

```

###Dropping Duplicate Values

```

dropped_duplicates_data = cleaned_sales_data.dropna().dropDuplicates()
dropped_duplicates_data.show()

+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+---+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+

```

ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES
ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID
PRODUCTCODE	CUSTOMERNAME	PHONE	ADDRESSLINE1	ADDRESSLINE2
CITY	STATE	POSTALCODE	COUNTRY	TERRITORY
CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE		
10163	31	100.0	2	3329.09
10/20/2003 0:00	Shipped	4	10	2003
S18_2949	Classic Legends Inc.	2125558493	5905 Pompton	
St. Suite 750	NYC	NY	10022	USA
Hernandez	Maria	Medium		NA
10361	34	100.0	6	3871.92
12/17/2004 0:00	Shipped	4	12	2004
S24_3420	Souvenirs And Th...	+61 2 9495 8555	Monitor Money	
Bui... Level 6	Chatswood	NSW	2067	Australia
APAC	Huxley	Adrian	Medium	
10337	29	100.0	2	4498.19
11/21/2004 0:00	Shipped	4	11	2004
S18_4522	Classic Legends Inc.	2125558493	5905 Pompton	
St. Suite 750	NYC	NY	10022	USA
Hernandez	Maria	Medium		NA
10139	31	100.0	7	3184.94
7/16/2003 0:00	Shipped	3	7	2003
S18_1342	Souvenirs And Th...	+61 2 9495 8555	Monitor Money	
Bui... Level 6	Chatswood	NSW	2067	Australia
APAC	Huxley	Adrian	Medium	
10361	25	62.46	1	1561.5
12/17/2004 0:00	Shipped	4	12	2004
S24_4278	Souvenirs And Th...	+61 2 9495 8555	Monitor Money	
Bui... Level 6	Chatswood	NSW	2067	Australia
APAC	Huxley	Adrian	Small	
10267	44	96.74	4	4256.56
7/7/2004 0:00	Shipped	3	7	2004
S24_3191	Muscle Machine Inc	2125557413	4092 Furth	
Circle Suite 400	NYC	NY	10022	USA
NA	Young	Jeff	Medium	
10169	30	100.0	2	5019.9
11/4/2003 0:00	Shipped	4	11	2003
S12_1099	Anna's Decoration...	02 9936 8555	201 Miller	
Street Level 15	North Sydney	NSW	2060	Australia
APAC	O'Hara	Anna	Medium	
10324	27	54.33	1	1466.91
11/5/2004 0:00	Shipped	4	11	2004
S12_3148	Vitachrome Inc.	2125551500	2678 Kingston	
Rd. Suite 101	NYC	NY	10022	USA
				NA

[illegible]

```

St.| Suite 750| NYC| NY| 10022| USA| NA|
Hernandez| Maria| Medium|
| 10120| 29| 96.34| 3|2793.86|
4/29/2003 0:00| Shipped| 2| 4| 2003| Motorcycles|
118| S10_2016|Australian Collec...| 03 9520 4555| 636 St Kilda
Road| Level 3| Melbourne|Victoria| 3004|Australia|
APAC| Ferguson| Peter| Small|
| 10127| 24| 100.0| 11| 2559.6|
6/3/2003 0:00| Shipped| 2| 6| 2003|Trucks and Buses|
118| S12_4473| Muscle Machine Inc| 2125557413| 4092 Furth
Circle| Suite 400| NYC| NY| 10022| USA|
NA| Young| Jeff| Small|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+---+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
only showing top 20 rows

```

###Calculating Totals Sales By Each Product

```

calculate_total_sales_data =
dropped_duplicates_data.groupBy("PRODUCTLINE").sum("SALES").withColumn
Renamed("sum(SALES)", "TotalSales")
calculate_total_sales_data.show()

```

```

+-----+-----+
| PRODUCTLINE| TotalSales|
+-----+-----+
| Motorcycles| 116738.31|
| Vintage Cars| 199436.24|
| Ships| 15146.83|
| Trucks and Buses| 120698.5|
| Classic Cars| 379958.2299999999|
| Trains| 7530.65|
| Planes| 63772.090000000001|
+-----+-----+

```

###Creating New CSV File

```

calculate_total_sales_data.coalesce(1).write.csv("product_sales_summar
y.csv", header=True)
spark.stop()

```