Spark_Coding_Assignment

```
val Df1=spark.read.format("csv")
                    .option("header","true")
                    .load("file:///home/talentum/Downloads/startup.csv")
```

Df1: org.apache.spark.sql.DataFrame = [Sr_No: string, Date: string ... 8 more fields]

FINISHED

Took 1 sec. Last updated by anonymous at May 20 2021, 5:14:14 PM. (outdated)

```
Df1.show()
```

FINISHED

```
+-----+----------+--------------------+--------------------+--------------------+---------+--------------------+--------------------+-------------+------+
|Sr_No|      Date|        Startup_Name|   Industry_Vertical|          SubVertical|     City|      Investors_Name|      InvestmentnType|Amount_in_USD|Remrks|
+-----+----------+--------------------+--------------------+--------------------+---------+--------------------+--------------------+-------------+------+
|    1|09/01/2020|               BYJU'S|              E-Tech|          E-learning|Bengaluru|Tiger Global Mana...|Private Equity Round| 20,00,00,000|  null|
|    2|13/01/2020|              Shuttl|      Transportation|App based shuttle...|  Gurgaon|Susquehanna Growt...|            Series C|    80,48,394|  null|
|    3|09/01/2020|           Mamaearth|          E-commerce|Retailer of baby ...|Bengaluru|Sequoia Capital I...|            Series B|  1,83,58,860|  null|
|    4|02/01/2020|https://www.wealt...|             FinTech|   Online Investment| New Delhi|       Vinod Khatumal|         Pre-series A|    30,00,000|  null|
|    5|02/01/2020|              Fashor| Fashion and Apparel|Embroiled Clothes...|   Mumbai|Sprout Venture Pa...|          Seed Round|    18,00,000|  null|
|    6|13/01/2020|               Pando|           Logistics|Open-market, frei...|  Chennai|    Chiratae Ventures|            Series A|    90,00,000|  null|
|    7|10/01/2020|              Zomato|         Hospitality|Online Food Deliv...|  Gurgaon|       Ant Financial|Private Equity Round| 15,00,00,000|  null|
|    8|12/12/2019|              Ecozen|          Technology|            Agritech|     Pune|Sathguru Catalyze...|            Series A|    60,00,000|  null|
|    9|06/12/2019|             CarDekho|          E-Commerce|          Automobile|  Gurgaon|Ping An Global Vo...|            Series D|  7,00,00,000|  null|
|   10|03/12/2019|         Dhruva Space|           Aerospace|Satellite Communi...|Bengaluru|Mumbai Angels, Ra...|                Seed|  5,00,00,000|  null|
|   11|13/12/2019|              Rivigo|          Technology|Logistics Service...|  Gurgaon|SAIF Partners, Sp...|            Series F|  2,00,00,000|  null|
|   12|17/12/2019|           Healthians|B2B-focused foodt...|Food Solutions Fo...|Bengaluru|Paytm, NPTK, Sabr...|            Series C|  1,20,00,000|  null|
|   13|16/12/2019|             Licious|          E-Commerce|Online Meat And S...|Bengaluru|   Vertex Growth Fund|            Series E|  3,00,00,000|  null|
|   14|16/12/2019|               InCred|             Finance|Non-Banking Finan...|   Mumbai|                null|        Debt Funding|    59,00,000|  null|
+-----+----------+--------------------+--------------------+--------------------+---------+--------------------+--------------------+-------------+------+
```

Took 1 sec. Last updated by anonymous at May 20 2021, 5:14:19 PM. (outdated)

```
val Df2=spark.read.format("parquet")
                    .option("header","true")
                    .load("file:///home/talentum/Downloads/consumerInternet.parquet")
```

FINISHED

Df2: org.apache.spark.sql.DataFrame = [Sr_No: string, Date: string ... 8 more fields]

Took 2 sec. Last updated by anonymous at May 20 2021, 5:16:28 PM. (outdated)

```
Df2.show()
```

FINISHED

Spark_Coding_Assignment

```
+-----+----------+------------+-----------------+-----------------+---------+------------------+---------------+-------------+-------+
|Sr_No|      Date|Startup_Name|Industry_Vertical|       SubVertical|     City|    Investors_Name|InvestmentnType|Amount_in_USD|Remarks|
+-----+----------+------------+-----------------+-----------------+---------+------------------+---------------+-------------+-------+
|  152|01/09/2018|     Netmeds|Consumer Internet|Online Pharmacy C...|  Chennai|Sistema Asia Fund...|  Private Equity| 3,50,00,000|    nan|
|  154|02/09/2018|   Kunafount| Consumer Internet|...nd ebooks M...| Bengaluru|       Falcon Edge|  Private Equity|   63,90,000|    nan|
|  156|04/09/2018|   HappyGoEasy|Consumer Internet| Online Travel Agecy| Gurugram|Korea Investment ...|  Private Equity|         N/A|    nan|
|  157|05/09/2018|       Nykaa|Consumer Internet|Online Marketplac...|   Mumbai|        Lighthouse|  Private Equity|15,72,00,000|    nan|
|  159|06/09/2018|     Dream11|Consumer Internet|Online Gaming Pla...|   Mumbai|Tencent, Kalaari ...|  Private Equity|10,00,00,000|    nan|
|  162|11/09/2018|   Pharmeasy|Consumer Internet|Online Marketplac...|   Mumbai|    InnoVen Capital|    Debt Funding|   55,00,000|    nan|
|  167|02/08/2018| Samosa Labs|Consumer Internet|Social Media and ...|Hyderabad|Xiaomi and Sequoi...|  Private Equity|   70,00,000|    nan|
|  168|03/08/2018|   ZiffyHomes|Consumer Internet|Online Home Renta...| Gurugram|     Y Combinator|  Private Equity|    1,20,000|    nan|
|  169|07/08/2018|    My OmNamo|Consumer Internet|Holy Platform Off...|   Mumbai|              N/A|Seed/ Angel Funding|   10,00,000|    nan|
|  170|07/08/2018|       ShopX|Consumer Internet|Assisted E-commer...| Bengaluru|Fung Strategic Ho...|  Private Equity| 3,50,00,000|    nan|
|  171|07/08/2018|   MakeMyTrip|Consumer Internet|Online Travel Agg...| Gurugram|  MakeMyTrip Limited|  Private Equity| 1,00,00,000|    nan|
|  172|08/08/2018|   Hansel io|Consumer Internet|Real-time Mobile ...| Bengaluru|    Vertex Ventures|Seed/ Angel Funding|   40,00,000|    nan|
|  173|08/08/2018| Metro Bikes|Consumer Internet|Online Bike Renta...| Bengaluru|Sequoia Capital I...|  Private Equity| 1,22,00,000|    nan|
|  174|08/08/2018|    Phone Pe|Consumer Internet|   UPI Payments App| Bengaluru|         Flipkart|  Private Equity| 6,60,00,000|    nan|
|  177|13/08/2018|      Jethoo|Consumer Internet|Rural E-commerce ...|  Chennai|Ventureast Pring...|Seed/ Angel Funding|   40,00,000|    nan|
```

Took 2 sec. Last updated by anonymous at May 20 2021, 5:17:03 PM. (outdated)

```
val mergeDf = Df1.union(Df2)
mergeDf.show()
```
FINISHED

```
+-----+----------+------------+------------------+------------------+---------+------------------+-----------------+-------------+-------+
|Sr_No|      Date|Startup_Name| Industry_Vertical|       SubVertical|     City|    Investors_Name|  InvestmentnType|Amount_in_USD|Remarks|
+-----+----------+------------+------------------+------------------+---------+------------------+-----------------+-------------+-------+
|    1|09/01/2020|      BYJU'S|            E-Tech|       E-learning| Bengaluru|Tiger Global Mana...|Private Equity Round|20,00,00,000|   null|
|    2|13/01/2020|      Shuttl|    Transportation|App based shuttle...|  Gurgaon|Susquehanna Growt...|        Series C|     80,48,394|   null|
|    3|09/01/2020|   Mamaearth|       E-commerce|Retailer of baby ...| Bengaluru|Sequoia Capital I...|        Series B|   1,83,58,860|   null|
|    4|02/01/2020|https://www.wealt...|          FinTech|  Online Investment| New Delhi|     Vinod Khatumal|    Pre-series A|     30,00,000|   null|
|    5|02/01/2020|      Fashor| Fashion and Apparel|Embroiled Clothes...|   Mumbai|Sprout Venture Pa...|      Seed Round|     18,00,000|   null|
|    6|13/01/2020|       Pando|         Logistics|Open-market, frei...|  Chennai|   Chiratae Ventures|        Series A|     90,00,000|   null|
|    7|10/01/2020|      Zomato|       Hospitality|Online Food Deliv...|  Gurgaon|      Ant Financial|Private Equity Round|15,00,00,000|   null|
|    8|12/12/2019|      Ecozen|        Technology|         Agritech|     Pune|Sathguru Catalyze...|        Series A|     60,00,000|   null|
|    9|06/12/2019|    CarDekho|        E-Commerce|       Automobile|  Gurgaon|Ping An Global Vo...|        Series D|   7,00,00,000|   null|
|   10|03/12/2019| Dhruva Space|         Aerospace|Satellite Communi...| Bengaluru|Mumbai Angels, Ra...|            Seed|   5,00,00,000|   null|
|   11|13/12/2019|      Rivigo|        Technology|Logistics Service...|  Gurgaon|SAIF Partners, Sp...|        Series F|   2,00,00,000|   null|
|   12|17/12/2019|   Healthians|B2B-focused foodt...|Food Solutions Fo...| Bengaluru|Paytm, NPTK, Sabr...|        Series C|   1,20,00,000|   null|
|   13|16/12/2019|     Licious|        E-Commerce|Online Meat And S...| Bengaluru|  Vertex Growth Fund|        Series E|   3,00,00,000|   null|
|   14|16/12/2019|      InCred|           Finance|Non-Banking Finan...|   Mumbai|             null|    Debt Funding|     59,00,000|   null|
|   15|14/12/2019|        Trell|             Video|Experience Disco...| Bengaluru|Rainbow Investment...|       Seed Round|     20,00,000|   null|
```

Took 1 sec. Last updated by anonymous at May 20 2021, 5:19:53 PM. (outdated)

```
mergeDf.count()
```
FINISHED

```
res8: Long = 3044
```

Took 1 sec. Last updated by anonymous at May 20 2021, 5:26:51 PM. (outdated)

Spark_Coding_Assignment

```
mergeDf.createOrReplaceTempView("Data")
```
FINISHED

Took 1 sec. Last updated by anonymous at May 20 2021, 5:28:12 PM. (outdated)

FINISHED

```
spark.sql("""select count(Startup_Name) count_of_satrtups_from_pune from Data where city = 'Pune'""").show()

+---------------------------+
|count_of_satrtups_from_pune|
+---------------------------+
|                        105|
+---------------------------+
```

Took 2 sec. Last updated by anonymous at May 20 2021, 5:33:11 PM. (outdated)

FINISHED

```
spark.sql("""select count(Startup_Name) count_of_satrtups_from_pune_has_Seed_Angel_Funding from Data where InvestmentnType = 'Seed' or InvestmentnType = 'Angel'""").show()

+-------------------------------------------------+
|count_of_satrtups_from_pune_has_Seed_Angel_Funding|
+-------------------------------------------------+
|                                                5|
+-------------------------------------------------+
```

Took 1 sec. Last updated by anonymous at May 20 2021, 5:42:10 PM. (outdated)

FINISHED

```
spark.sql("""select Industry_Vertical top_5_Industry_Vertical from data GROUP BY Industry_Vertical ORDER BY count(Startup_Name)""").show(5)

+--------------------+
|top_5_Industry_Vertical|
+--------------------+
|  IOT Energy Manage...|
|  Smart Safety Wear...|
|  Activity Based So...|
```

```
|    Hyperlocal Deals ...|
|    Online Shopping A...|
+--------------------+
only showing top 5 rows
```

Spark_Coding_Assignment

Took 2 sec. Last updated by anonymous at May 20 2021, 6:31:57 PM. (outdated)

```
spark.sql("""select Investors_Name top_Investors from data GROUP BY Investors_Name ORDER BY sum(Amount_in_USD)""").show()
```

FINISHED

```
+--------------------+
|       top_Investors|
+--------------------+
|       Mumbai Angels|
|Skycatcher Fund, ...|
|     Axilor Ventures|
|Geniee, Inc, Purv...|
|        ah! Ventures|
|        Ravi Agarwal|
|Group of 15 Angel...|
|Anand Rajaraman, ...|
|      Francesco Cara|
|Dutch government ...|
|DG Ventures India...|
|Nandan Nilekani, ...|
|        ResponsAbility|
|       Blume ventures|
```

Took 4 sec. Last updated by anonymous at May 20 2021, 6:21:43 PM. (outdated)

READY