

NOTE: QUESTIONS ARE HIGHLIGHTED IN GREEN AND QUERIES ARE HIGHLIGHTED IN YELLOW

Q1. Read the data, show it and Count the number of records.

```
hadoop fs -put project/*.csv /project
```

```
>>> from pyspark.sql.functions import *
```

```
>>> df1=spark.read.option("header",True).option("inferSchema",True).csv("/project/case.csv")
```

```
>>> df1.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+
|case_id|province|    city|group|  infection_case|confirmed| latitude| longitude|
+-----+-----+-----+-----+-----+-----+-----+
|1000001| Seoul|  Yongsan-gu| true|   Itaewon Clubs|   139|37.538621|126.992652|
|1000002| Seoul|   Gwanak-gu| true|      Richway|   119| 37.48208|126.901384|
|1000003| Seoul|   Guro-gu| true| Guro-gu Call Center|   95|37.508163|126.884387|
|1000004| Seoul| Yangcheon-gu| true|Yangcheon Table T...|   43|37.546061|126.874209|
|1000005| Seoul|  Dobong-gu| true|   Day Care Center|   43|37.679422|127.044374|
|1000006| Seoul|   Guro-gu| true|Manmin Central Ch...|   41|37.481059|126.894343|
|1000007| Seoul|from other city| true|SMR Newly Planted...|   36|   -|   -|
|1000008| Seoul| Dongdaemun-gu| true|   Dongan Church|   17|37.592888|127.056766|
|1000009| Seoul|from other city| true|Coupang Logistics...|   25|   -|   -|
|1000010| Seoul|   Gwanak-gu| true|  Wangsung Church|   30|37.481735|126.930121|
|1000011| Seoul| Eunpyeong-gu| true|Eunpyeong St. Mar...|   14| 37.63369| 126.9165|
|1000012| Seoul| Seongdong-gu| true|  Seongdong-gu APT|   13| 37.55713| 127.0403|
|1000013| Seoul|  Jongno-gu| true|Jongno Community ...|   10| 37.57681| 127.006|
|1000014| Seoul|  Gangnam-gu| true|Samsung Medical C...|    7| 37.48825| 127.08559|
|1000015| Seoul|   Jung-gu| true|Jung-gu Fashion C...|    7|37.562405|126.984377|
|1000016| Seoul| Seodaemun-gu| true| Yeonana News Class|    5|37.558147|126.943799|
|1000017| Seoul|  Jongno-gu| true|Korea Campus Crus...|    7|37.594782|126.968022|
|1000018| Seoul|  Gangnam-gu| true|Gangnam Yeoksam-d...|    6|   -|   -|
|1000019| Seoul|from other city| true|Daejeon door-to-d...|    1|   -|   -|
|1000020| Seoul|  Geumcheon-gu| true|Geumcheon-gu rice...|    6|   -|   -|
+-----+-----+-----+-----+-----+-----+-----+
```

only showing top 20 rows

```
>>> df2=spark.read.option("header",True).option("inferSchema",True).csv("/project/region.csv")
```

```
>>> df2.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+
| code|province|    city| latitude|
longitude|elementary_school_count|kindergarten_count|university_count|academy_ratio|elderly_population_rati
o|elderly_alone_ratio|nursing_home_count|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+

|10000| Seoul|    Seoul|37.566953|126.977977|          607|          830|          48|          1.44|
15.38|          5.8|          22739|

|10010| Seoul|  Gangnam-gu|37.518421|127.047222|          33|          38|          0|          4.18|
13.17|          4.3|          3088|

|10020| Seoul|  Gangdong-gu|37.530492|127.123837|          27|          32|          0|          1.54|
14.55|          5.4|          1023|

|10030| Seoul|  Gangbuk-gu|37.639938|127.025508|          14|          21|          0|          0.67|
19.49|          8.5|          628|

|10040| Seoul|  Gangseo-gu|37.551166|126.849506|          36|          56|          1|          1.17|
14.39|          5.7|          1080|

|10050| Seoul|  Gwanak-gu| 37.47829|126.951502|          22|          33|          1|          0.89|
15.12|          4.9|          909|

|10060| Seoul|  Gwangjin-gu|37.538712|127.082366|          22|          33|          3|          1.16|
13.75|          4.8|          723|

|10070| Seoul|  Guro-gu|37.495632| 126.88765|          26|          34|          3|          1.0|
16.21|          5.7|          741|

|10080| Seoul|  Geumcheon-gu|37.456852|126.895229|          18|          19|          0|          0.96|
16.15|          6.7|          475|

|10090| Seoul|  Nowon-gu|37.654259|127.056294|          42|          66|          6|          1.39|
15.4|          7.4|          952|

|10100| Seoul|  Dobong-gu|37.668952|127.047082|          23|          26|          1|          0.95|
17.89|          7.2|          485|

|10110| Seoul|Dongdaemun-gu|37.574552|127.039721|          21|          31|          4|          1.06|
17.26|          6.7|          832|

|10120| Seoul|  Dongjak-gu|37.510571|126.963604|          21|          34|          3|          1.17|
15.85|          5.2|          762|

|10130| Seoul|  Mapo-gu|37.566283|126.901644|          22|          24|          2|          1.83|
14.05|          4.9|          929|

|10140| Seoul|  Seodaemun-gu|37.579428|126.936771|          19|          25|          6|          1.12|
16.77|          6.2|          587|

|10150| Seoul|  Seocho-gu|37.483804|127.032693|          24|          27|          1|          2.6|
13.39|          3.8|          1465|

|10160| Seoul|  Seongdong-gu|37.563277|127.036647|          21|          30|          2|          0.97|
14.76|          5.3|          593|
```

10170	Seoul	Seongbuk-gu	37.589562	127.0167	29	49	6	1.02
16.15	6.0	729						
10180	Seoul	Songpa-gu	37.51462	127.106141	40	51	1	1.65
13.1	4.1	1527						
10190	Seoul	Yangcheon-gu	37.517189	126.866618	30	43	0	2.26
13.55	5.5	816						

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

only showing top 20 rows

```
>>> df3=spark.read.option("header",True).option("inferSchema",True).csv("/project/timeprovince.csv")
>>> df3.show()
```

```
+-----+-----+-----+-----+-----+-----+
|  date|time|    province|confirmed|released|deceased|
+-----+-----+-----+-----+-----+-----+
|2020-01-20| 16|    Seoul|    0|    0|    0|
|2020-01-20| 16|    Busan|    0|    0|    0|
|2020-01-20| 16|    Daegu|    0|    0|    0|
|2020-01-20| 16|   Incheon|    1|    0|    0|
|2020-01-20| 16|   Gwangju|    0|    0|    0|
|2020-01-20| 16|   Daejeon|    0|    0|    0|
|2020-01-20| 16|    Ulsan|    0|    0|    0|
|2020-01-20| 16|   Sejong|    0|    0|    0|
|2020-01-20| 16| Gyeonggi-do|    0|    0|    0|
|2020-01-20| 16| Gangwon-do|    0|    0|    0|
|2020-01-20| 16| Chungcheongbuk-do|    0|    0|    0|
|2020-01-20| 16| Chungcheongnam-do|    0|    0|    0|
|2020-01-20| 16| Jeollabuk-do|    0|    0|    0|
|2020-01-20| 16| Jeollanam-do|    0|    0|    0|
|2020-01-20| 16| Gyeongsangbuk-do|    0|    0|    0|
|2020-01-20| 16| Gyeongsangnam-do|    0|    0|    0|
|2020-01-20| 16|    Jeju-do|    0|    0|    0|
|2020-01-21| 16|    Seoul|    0|    0|    0|
|2020-01-21| 16|    Busan|    0|    0|    0|
|2020-01-21| 16|    Daegu|    0|    0|    0|
+-----+-----+-----+-----+-----+-----+
```

only showing top 20 rows

```
>>> df1.count()
```

```
174
```

```
>>> df2.count()
```

```
244
```

```
>>> df3.count()
```

```
2771
```

Q2. Describe the data with a describe function.

```
>>> df1.describe().show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
|summary| case_id|province| city| infection_case| confirmed| latitude| longitude|
+-----+-----+-----+-----+-----+-----+-----+-----+
| count| 174| 174| 174| 174| 174| 174| 174|
| mean|2686215.7586206896| null| null| null| 65.48850574712644|
36.69405111076924|127.58488500461536|
| stddev|1943218.4834908631| null| null| null|355.09765388939746|0.9114662922487264|
0.823086807800544|
| min| 1000001| Busan| -|Anyang Gunpo Past...| 0| -| -|
| max| 7000004| Ulsan|from other city| overseas inflow| 4511| 37.758635| 129.1256|
+-----+-----+-----+-----+-----+-----+-----+-----+
```

```
>>> df2.describe().show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+
|summary| code|province| city| latitude|
longitude|elementary_school_count|kindergarten_count| university_count|
academy_ratio|elderly_population_ratio|elderly_alone_ratio|nursing_home_count|
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+
| count| 244| 244| 244| 244| 244| 244| 244| 244|
244| 244| 244| 244|
```

```
| mean| 32912.09016393442| null| null| 36.39699581147539|127.66140068442627|
74.18032786885246|107.90163934426229| 4.151639344262295|1.2947540983606531| 20.923729508196736|
10.644672131147537|1159.2581967213114|

| stddev|19373.349735535565| null| null|1.0603044400519337|0.9047812669143711|
402.7134817931963| 588.7883199916743|22.513040514027963|0.5928979025284541| 8.087427906866017|
5.604885829305476| 6384.185084757067|

| min|      10000| Busan| Andong-si|      33.488936|      126.263554|          4|          4|          0|
0.19|          7.69|          3.3|          11|

| max|      80000| Ulsan|Yuseong-gu|      38.380571|      130.905883|        6087|        8837|
340|          4.18|          40.26|          24.7|          94865|

+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+
```

```
>>> df3.describe().show()
```

```
+-----+-----+-----+-----+-----+-----+
|summary|  date|      time|province|  confirmed|  released|  deceased|
+-----+-----+-----+-----+-----+-----+
| count|  2771|      2771|  2771|      2771|      2771|      2771|
| mean|   null|4.122699386503068|   null| 444.3081919884518|320.72645254420786| 9.239985564778058|
| stddev|   null|6.998872580910621|   null|1360.8909633062747|1126.0412546146954|32.638615308276584|

| min|2020-01-20|          0| Busan|          0|          0|          0|
| max|2020-06-30|          16| Ulsan|        6906|        6700|        189|

+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+
```

Q3. If there is any duplicate value drop it.

```
>>> df4=df1.dropDuplicates()
```

```
>>> df4.count()
```

```
174
```

```
>>> df5=df2.dropDuplicates()
```

```
>>> df5.count()
```

```
244
```

```
>>> df6=df3.dropDuplicates()
```

```
>>> df6.count()
```

```
2771
```

Q4. Use limit function for showcasing a limited number of records.

```
>>> df4.limit(10).show()
```

```
+-----+-----+-----+-----+-----+-----+-----+
|case_id| province| city|group| infection_case|confirmed| latitude| longitude|
+-----+-----+-----+-----+-----+-----+-----+
|1500003| Daejeon| Seo-gu| true| Orange Town| 7|36.3398739|127.3819744|
|3000003| Gangwon-do| Wonju-si| true| Wonju-si Apartments| 4| 37.342762| 127.983815|
|1700006| Sejong| -|false| etc| 1| -| -|
|4100004| Chungcheongnam-do|from other city| true| Richway| 3| -| -|
|5000004| Jeollabuk-do| -|false| overseas inflow| 12| -| -|
|5000001| Jeollabuk-do|from other city| true| Itaewon Clubs| 2| -| -|
|1000023| Seoul| Jung-gu| true| KB Life Insurance| 13| 37.560899| 126.966998|
|3000001| Gangwon-do|from other city| true| Shincheonji Church| 17| -| -|
|2000021| Gyeonggi-do| -|false|contact with patient| 63| -| -|
|1000022| Seoul|from other city| true|Guri Collective I...| 5| -| -|
+-----+-----+-----+-----+-----+-----+-----+
```

```
>>> df5.limit(10).show()
```

```
+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+
| code| province| city| latitude|
longitude|elementary_school_count|kindergarten_count|university_count|academy_ratio|elderly_population_rati
o|elderly_alone_ratio|nursing_home_count|
+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+
|20190| Gyeonggi-do|Yangpyeong-gun|37.491744| 127.48757| 22| 25| 1| 0.78|
24.66| 11.1| 178|
|11110| Busan| Seo-gu|35.098135|129.024193| 11| 9| 0| 1.02|
24.36| 12.3| 235|
|11150| Busan| Jung-gu|35.106321|129.032256| 4| 4| 0| 1.36|
26.0| 13.0| 182|
|30030| Gangwon-do| Donghae-si|37.524689|129.114239| 14| 17| 0| 1.6|
19.41| 10.1| 140|
|14010| Incheon| Ganghwa-gun|37.747065|126.487777| 20| 19| 1| 0.68|
31.97| 13.9| 107|
|20210| Gyeonggi-do| Yeoncheon-gun|38.096409|127.075067| 13| 13| 0| 0.82|
25.41| 12.7| 69|
```

51110	Jeollanam-do	Suncheon-si	34.950592	127.487396	42	61	3	1.94
15.2	8.3	452						
50140	Jeollabuk-do	Jinan-gun	35.791669	127.424806	13	12	0	0.39
33.8	19.5	54						
60180	Gyeongsangbuk-do	Uljin-gun	36.993078	129.400585	13	16	0	1.01
27.3	17.6	83						
12070	Daegu	Suseong-gu	35.858395	128.630551	34	49	1	2.28
15.0	6.9	948						

```

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

```

```
>>> df6.limit(10).show()
```

```

+-----+-----+-----+-----+-----+
|  date|time|    province|confirmed|released|deceased|
+-----+-----+-----+-----+-----+
|2020-02-04| 16|  Gyeonggi-do|    8|    0|    0|
|2020-02-15| 16| Jeollanam-do|    0|    0|    0|
|2020-02-16| 16|    Seoul|   14|    3|    0|
|2020-02-19| 16| Chungcheongbuk-do|    0|    0|    0|
|2020-03-25|  0| Chungcheongbuk-do|   38|   13|    0|
|2020-03-29|  0|    Seoul|  410|   93|    0|
|2020-04-09|  0|    Sejong|   46|   16|    0|
|2020-04-23|  0|    Ulsan|   43|   36|    1|
|2020-04-26|  0| Gyeongsangbuk-do| 1364| 1109|   52|
|2020-05-08|  0| Gyeongsangnam-do|  117|  105|    0|
+-----+-----+-----+-----+-----+

```

Q5. If you find the column name is not suitable, change the column name.

```
>>> df2.withColumnRenamed("code","Region_ID").show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+
|Region_ID|province|    city|latitude|
longitude|elementary_school_count|kindergarten_count|university_count|academy_ratio|elderly_population_rati
o|elderly_alone_ratio|nursing_home_count|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+
| 10000| Seoul|    Seoul|37.566953|126.977977|          607|          830|          48|          1.44|
15.38|          5.8|          22739|
| 10010| Seoul| Gangnam-gu|37.518421|127.047222|          33|          38|          0|          4.18|
13.17|          4.3|          3088|
| 10020| Seoul| Gangdong-gu|37.530492|127.123837|          27|          32|          0|          1.54|
14.55|          5.4|          1023|
| 10030| Seoul| Gangbuk-gu|37.639938|127.025508|          14|          21|          0|          0.67|
19.49|          8.5|          628|
| 10040| Seoul| Gangseo-gu|37.551166|126.849506|          36|          56|          1|          1.17|
14.39|          5.7|          1080|
| 10050| Seoul| Gwanak-gu| 37.47829|126.951502|          22|          33|          1|          0.89|
15.12|          4.9|          909|
| 10060| Seoul| Gwangjin-gu|37.538712|127.082366|          22|          33|          3|          1.16|
13.75|          4.8|          723|
| 10070| Seoul| Guro-gu|37.495632| 126.88765|          26|          34|          3|          1.0|
16.21|          5.7|          741|
| 10080| Seoul| Geumcheon-gu|37.456852|126.895229|          18|          19|          0|          0.96|
16.15|          6.7|          475|
| 10090| Seoul| Nowon-gu|37.654259|127.056294|          42|          66|          6|          1.39|
15.4|          7.4|          952|
| 10100| Seoul| Dobong-gu|37.668952|127.047082|          23|          26|          1|          0.95|
17.89|          7.2|          485|
| 10110| Seoul| Dongdaemun-gu|37.574552|127.039721|          21|          31|          4|          1.06|
17.26|          6.7|          832|
| 10120| Seoul| Dongjak-gu|37.510571|126.963604|          21|          34|          3|          1.17|
15.85|          5.2|          762|
| 10130| Seoul| Mapo-gu|37.566283|126.901644|          22|          24|          2|          1.83|
14.05|          4.9|          929|
| 10140| Seoul| Seodaemun-gu|37.579428|126.936771|          19|          25|          6|          1.12|
16.77|          6.2|          587|
| 10150| Seoul| Seocho-gu|37.483804|127.032693|          24|          27|          1|          2.6|
13.39|          3.8|          1465|
```


10160	Seoul	Seongdong-gu	37.563277	127.036647	21	30	2	0.97
14.76	5.3	593						
10170	Seoul	Seongbuk-gu	37.589562	127.0167	29	49	6	1.02
16.15	6.0	729						
10180	Seoul	Songpa-gu	37.51462	127.106141	40	51	1	1.65
13.1	4.1	1527						
10190	Seoul	Yangcheon-gu	37.517189	126.866618	30	43	0	2.26
13.55	5.5	816						

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+
```

only showing top 20 rows

Q6. Select the subset of the columns.

```
>>> df1.select("case_id","province","infection_case").show()
```

```
+-----+-----+-----+
|case_id|province|infection_case|
+-----+-----+-----+
|1000001|Seoul|Itaewon Clubs|
|1000002|Seoul|Richway|
|1000003|Seoul|Guro-gu Call Center|
|1000004|Seoul|Yangcheon Table T...|
|1000005|Seoul|Day Care Center|
|1000006|Seoul|Manmin Central Ch...|
|1000007|Seoul|SMR Newly Planted...|
|1000008|Seoul|Dongan Church|
|1000009|Seoul|Coupang Logistics...|
|1000010|Seoul|Wangsung Church|
|1000011|Seoul|Eunpyeong St. Mar...|
|1000012|Seoul|Seongdong-gu APT|
|1000013|Seoul|Jongno Community ...|
|1000014|Seoul|Samsung Medical C...|
|1000015|Seoul|Jung-gu Fashion C...|
|1000016|Seoul|Yeonana News Class|
|1000017|Seoul|Korea Campus Crus...|
|1000018|Seoul|Gangnam Yeoksam-d...|
|1000019|Seoul|Daejeon door-to-d...|
```

```
|1000020| Seoul|Geumcheon-gu rice...|
```

```
+-----+-----+-----+-----+
```

```
>>> df2.select("code","province","city").show()
```

```
+----+-----+-----+-----+
```

```
| code|province|    city|
```

```
+----+-----+-----+-----+
```

```
|10000| Seoul|    Seoul|
```

```
|10010| Seoul| Gangnam-gu|
```

```
|10020| Seoul| Gangdong-gu|
```

```
|10030| Seoul| Gangbuk-gu|
```

```
|10040| Seoul| Gangseo-gu|
```

```
|10050| Seoul| Gwanak-gu|
```

```
|10060| Seoul| Gwangjin-gu|
```

```
|10070| Seoul| Guro-gu|
```

```
|10080| Seoul| Geumcheon-gu|
```

```
|10090| Seoul| Nowon-gu|
```

```
|10100| Seoul| Dobong-gu|
```

```
|10110| Seoul|Dongdaemun-gu|
```

```
|10120| Seoul| Dongjak-gu|
```

```
|10130| Seoul| Mapo-gu|
```

```
|10140| Seoul| Seodaemun-gu|
```

```
|10150| Seoul| Seocho-gu|
```

```
|10160| Seoul| Seongdong-gu|
```

```
|10170| Seoul| Seongbuk-gu|
```

```
|10180| Seoul| Songpa-gu|
```

```
|10190| Seoul| Yangcheon-gu|
```

```
+----+-----+-----+-----+
```

```
>>> df3.select("date","time","province").show()
```

```
+-----+----+-----+-----+
```

```
|   date|time|    province|
```

```
+-----+----+-----+-----+
```

```
|2020-01-20| 16|    Seoul|
```

2020-01-20	16	Busan
2020-01-20	16	Daegu
2020-01-20	16	Incheon
2020-01-20	16	Gwangju
2020-01-20	16	Daejeon
2020-01-20	16	Ulsan
2020-01-20	16	Sejong
2020-01-20	16	Gyeonggi-do
2020-01-20	16	Gangwon-do
2020-01-20	16	Chungcheongbuk-do
2020-01-20	16	Chungcheongnam-do
2020-01-20	16	Jeollabuk-do
2020-01-20	16	Jeollanam-do
2020-01-20	16	Gyeongsangbuk-do
2020-01-20	16	Gyeongsangnam-do
2020-01-20	16	Jeju-do
2020-01-21	16	Seoul
2020-01-21	16	Busan
2020-01-21	16	Daegu

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

only showing top 20 rows

Q7. If there is any null value, fill it with any random value or drop it

```
>>> df1.dropna().count()
```

174

```
>>> df2.dropna().count()
```

244

```
>>> df3.dropna().count()
```

2771

Note : There are no nulls in any of the dataframe as after using dropna() the count of records is same as earlier

Note: Deliberately adding null to the data and then written a query

```
>>> df4.filter(col("group")!= "NULL").select("case_id","province","city","group").show()
```

```
+-----+-----+-----+-----+
|case_id|province|    city|group|
+-----+-----+-----+-----+
|1000002|  Seoul|  Gwanak-gu| TRUE|
|1000003|  Seoul|   Guro-gu| TRUE|
|1000005|  Seoul| Dobong-gu| TRUE|
|1000006|  Seoul|   Guro-gu| TRUE|
|1000007|  Seoul|from other city| TRUE|
|1000008|  Seoul| Dongdaemun-gu| TRUE|
|1000009|  Seoul|from other city| TRUE|
|1000010|  Seoul|  Gwanak-gu| TRUE|
|1000011|  Seoul| Eunpyeong-gu| TRUE|
|1000012|  Seoul| Seongdong-gu| TRUE|
|1000013|  Seoul|  Jongno-gu| TRUE|
|1000014|  Seoul|  Gangnam-gu| TRUE|
|1000015|  Seoul|   Jung-gu| TRUE|
|1000016|  Seoul| Seodaemun-gu| TRUE|
|1000017|  Seoul|  Jongno-gu| TRUE|
|1000018|  Seoul|  Gangnam-gu| TRUE|
|1000019|  Seoul|from other city| TRUE|
|1000020|  Seoul| Geumcheon-gu| TRUE|
|1000021|  Seoul|from other city| TRUE|
|1000022|  Seoul|from other city| TRUE|
+-----+-----+-----+-----+
```

Q8. find out all the different infection case in Daegu Province with more than 10 confirmed cases.

```
>>> df1.filter((col("province")=="Daegu") &
(col("confirmed")>=10)).select("case_id","infection_case","confirmed","province").show()
```

```
+-----+-----+-----+-----+
|case_id|  infection_case|confirmed|province|
+-----+-----+-----+-----+
|1200001| Shincheonji Church|   4511| Daegu|
|1200002| Second Mi-Ju Hosp...|   196| Daegu|
|1200003| Hansarang Convale...|   124| Daegu|
|1200004| Daesil Convalesce...|   101| Daegu|
|1200005| Fatima Hospital|    39| Daegu|
|1200008| overseas inflow|    41| Daegu|
|1200009| contact with patient|   917| Daegu|
|1200010|          etc|   747| Daegu|
+-----+-----+-----+-----+
```

Q9. Sort the number of confirmed cases. Confirmed column is there in the dataset. Check with descending sort also.

```
>>> df1.orderBy(col("confirmed")).select("*").show(174)
```

```
+-----+-----+-----+-----+-----+-----+-----+
|case_id|  province|  city|group|  infection_case|confirmed| latitude| longitude|
+-----+-----+-----+-----+-----+-----+-----+
|7000002| Jeju-do| -|false|contact with patient|    0|    -|    -|
|3000007| Gangwon-do| -|false|contact with patient|    0|    -|    -|
|1000030| Seoul| Gangseo-gu| true|SJ Investment Cal...|    0| 37.559649| 126.835102|
|1700006| Sejong| -|false|          etc|    1|    -|    -|
|1000019| Seoul|from other city| true|Daejeon door-to-d...|    1|    -|    -|
|5000003| Jeollabuk-do|from other city| true| Shincheonji Church|    1|    -|    -|
|5100002| Jeollanam-do|from other city| true| Shincheonji Church|    1|    -|    -|
|7000004| Jeju-do|from other city| true| Itaewon Clubs|    1|    -|    -|
|1000028| Seoul|from other city| true|Anyang Gunpo Past...|    1|    -|    -|
|1000034| Seoul| -| true| Orange Life|    1|    -|    -|
|1000025| Seoul| Gangnam-gu| true|Gangnam Dongin Ch...|    1| 37.522331| 127.057388|
```

1100007	Busan from other city true Cheongdo Daenam H...	1	-	-
1300005	Gwangju - false etc 1 - -			
1700003	Sejong from other city true Shincheonji Church 1 - -			
4100007	Chungcheongnam-do - false contact with patient 2 - -			
5000001	Jeollabuk-do from other city true Itaewon Clubs 2 - -			
5100001	Jeollanam-do Muan-gun true Manmin Central Ch...	2	35.078825	126.316746
6100008	Gyeongsangnam-do from other city true Itaewon Clubs 2 - -			
6100009	Gyeongsangnam-do from other city true Onchun Church 2 - -			
1000033	Seoul from other city true Uiwang Logistics ... 2 - -			
1200007	Daegu from other city true Cheongdo Daenam H...	2	-	-
1200006	Daegu from other city true Itaewon Clubs 2 - -			
1400004	Incheon from other city true Shincheonji Church 2 - -			
1500006	Daejeon from other city true Shincheonji Church 2 - -			
1500007	Daejeon from other city true Seosan-si Laboratory 2 - -			
4000003	Chungcheongbuk-do from other city true Guro-gu Call Center 2 - -			
1000026	Seoul Yangcheon-gu true Biblical Language...	3	37.524623	126.843118
1000032	Seoul Jung-gu true Seoul City Hall S...	3	37.565699	126.977079
1000035	Seoul Guro-gu true Daezayeon Korea 3 37.486837 126.893163			
1500005	Daejeon Seo-gu true Korea Forest Engi...	3	36.358123	127.388856
4100004	Chungcheongnam-do from other city true Richway 3 - -			
1000024	Seoul Yeongdeungpo-gu true Yeongdeungpo Lear...	3	37.520846	126.931278
4100005	Chungcheongnam-do from other city true Eunpyeong-Boksago...	3	-	-
1600003	Ulsan - false contact with patient 3 - -			
6100007	Gyeongsangnam-do Yangsan-si true Soso Seowon 3 35.338811 129.017508			
1700005	Sejong - false contact with patient 3 - -			
5000002	Jeollabuk-do from other city true Door-to-door sale...	3	-	-
3000004	Gangwon-do from other city true Richway 4 - -			
1000029	Seoul Gangnam-gu true Samsung Fire & Ma...	4	37.498279	127.030139
3000005	Gangwon-do from other city true Geumcheon-gu rice...	4	-	-
5100004	Jeollanam-do - false contact with patient 4 - -			
5100005	Jeollanam-do - false etc 4 - -			
7000003	Jeju-do - false etc 4 - -			
1000031	Seoul from other city true Yongin Brothers 4 - -			
1100006	Busan from other city true Itaewon Clubs 4 - -			

1100005	Busan	Jin-gu true	Jin-gu Academy	4	35.17371	129.0633
1500004	Daejeon	Seo-gu true	Dreaming Church	4	36.346869	127.368594
3000003	Gangwon-do	Wonju-si true	Wonju-si Apartments	4	37.342762	127.983815
1100003	Busan	Suyeong-gu true	Suyeong-gu Kinder...	5	35.16708	129.1124
2000018	Gyeonggi-do	from other city true	Seoul City Hall S...	5	-	-
1000027	Seoul	Seocho-gu true	Seocho Family	5	-	-
1300001	Gwangju	Dong-gu true	Gwangneuksa Temple	5	35.136035	126.956405
1300004	Gwangju	- false	contact with patient	5	-	-
1700004	Sejong	- false	overseas inflow	5	-	-
2000019	Gyeonggi-do	Seongnam-si true	Seongnam neighbor...	5	-	-
5000005	Jeollabuk-do	- false	etc	5	-	-
1000022	Seoul	from other city true	Guri Collective I...	5	-	-
1000016	Seoul	Seodaemun-gu true	Yeonana News Class	5	37.558147	126.943799
1000020	Seoul	Geumcheon-gu true	Geumcheon-gu rice...	6	-	-
1400006	Incheon	- false	contact with patient	6	-	-
6100011	Gyeongsangnam-do	- false	contact with patient	6	-	-
1100004	Busan	Haeundae-gu true	Haeundae-gu Catho...	6	35.20599	129.1256
2000016	Gyeonggi-do	from other city true	Geumcheon-gu rice...	6	-	-
4000004	Chungcheongbuk-do	from other city true	Shincheonji Church	6	-	-
1000018	Seoul	Gangnam-gu true	Gangnam Yeoksam-d...	6	-	-
2000017	Gyeonggi-do	from other city true	Wangsung Church	6	-	-
1600004	Ulsan	- false	etc	7	-	-
3000008	Gangwon-do	- false	etc	7	-	-
1000014	Seoul	Gangnam-gu true	Samsung Medical C...	7	37.48825	127.08559
1500003	Daejeon	Seo-gu true	Orange Town	7	36.3398739	127.3819744
6100005	Gyeongsangnam-do	Changwon-si true	Hanmaeum Changwon...	7	35.22115	128.6866
6100006	Gyeongsangnam-do	Changnyeong-gun true	Changnyeong Coin ...	7	35.54127	128.5008
2000015	Gyeonggi-do	from other city true	Korea Campus Crus...	7	-	-
1000015	Seoul	Jung-gu true	Jung-gu Fashion C...	7	37.562405	126.984377
1000017	Seoul	Jongno-gu true	Korea Campus Crus...	7	37.594782	126.968022
4000006	Chungcheongbuk-do	- false	contact with patient	8	-	-
1000021	Seoul	from other city true	Shincheonji Church	8	-	-
6100004	Gyeongsangnam-do	Geochang-gun true	Geochang-gun Woon...	8	35.805681	127.917805
1700002	Sejong	Sejong true	gym facility in S...	8	36.48025	127.289

1300002	Gwangju	from other city	true	Shincheonji Church	9	-	-
6100003	Gyeongsangnam-do	Jinju-si	true	Wings Tower	9	35.164845	128.126969
4000002	Chungcheongbuk-do	from other city	true	Itaewon Clubs	9	-	-
4100003	Chungcheongnam-do	Seosan-si	true	Seosan-si Laboratory	9	37.000354	126.354443
3000002	Gangwon-do	from other city	true	Uijeongbu St. Mar...	10	-	-
6000010	Gyeongsangbuk-do	Gumi-si	true	Gumi Elim Church	10	-	-
2000014	Gyeonggi-do	Suwon-si	true	Suwon Saeng Myeon...	10	37.2376	127.0517
6100002	Gyeongsangnam-do	Geochang-gun	true	Geochang Church	10	35.68556	127.9127
1000013	Seoul	Jongno-gu	true	Jongno Community ...	10	37.57681	127.006
4100002	Chungcheongnam-do	from other city	true	Door-to-door sale...	10	-	-
1400007	Incheon	- false		etc	11	-	-
4000001	Chungcheongbuk-do	Goesan-gun	true	Goesan-gun Jangye...	11	36.82422	127.9552
4000007	Chungcheongbuk-do	- false		etc	11	-	-
1100002	Busan	from other city	true	Shincheonji Church	12	-	-
4100008	Chungcheongnam-do	- false		etc	12	-	-
5000004	Jeollabuk-do	- false		overseas inflow	12	-	-
1000012	Seoul	Seongdong-gu	true	Seongdong-gu APT	13	37.55713	127.0403
4000005	Chungcheongbuk-do	- false		overseas inflow	13	-	-
1000023	Seoul	Jung-gu	true	KB Life Insurance	13	37.560899	126.966998
1500002	Daejeon	Seo-gu	true	Dunsan Electronic...	13	36.3400973	127.3927099
7000001	Jeju-do	- false		overseas inflow	14	-	-
1000011	Seoul	Eunpyeong-gu	true	Eunpyeong St. Mar...	14	37.63369	126.9165
5100003	Jeollanam-do	- false		overseas inflow	14	-	-
2000012	Gyeonggi-do	Suwon-si	true	Lotte Confectione...	15	37.287356	127.013827
1500010	Daejeon	- false		etc	15	-	-
1500008	Daejeon	- false		overseas inflow	15	-	-
1500009	Daejeon	- false		contact with patient	15	-	-
3000006	Gangwon-do	- false		overseas inflow	16	-	-
4100006	Chungcheongnam-do	- false		overseas inflow	16	-	-
6000009	Gyeongsangbuk-do	Gyeongsan-si	true	Gyeongsan Cham Jo...	16	35.82558	128.7373
1600001	Ulsan	from other city	true	Shincheonji Church	16	-	-
6000008	Gyeongsangbuk-do	Gyeongsan-si	true	Gyeongsan Jeil Si...	17	35.84819	128.7621
3000001	Gangwon-do	from other city	true	Shincheonji Church	17	-	-
1000008	Seoul	Dongdaemun-gu	true	Dongan Church	17	37.592888	127.056766

2000013	Gyeonggi-do	Anyang-si	true	Lord Glory Church	17	37.403722	126.954939
1100009	Busan	-	false	contact with patient	19	-	-
6100012	Gyeongsangnam-do	-	false	etc	20	-	-
1400003	Incheon	from other city	true	Guro-gu Call Center	20	-	-
2000011	Gyeonggi-do	Anyang-si	true	Anyang Gunpo Past...	22	37.381784	126.93615
6000011	Gyeongsangbuk-do	-	false	overseas inflow	22	-	-
2000010	Gyeonggi-do	Seongnam-si	true	Bundang Jesaeng H...	22	37.38833	127.1218
1300003	Gwangju	-	false	overseas inflow	23	-	-
1000009	Seoul	from other city	true	Coupang Logistics...	25	-	-
1600002	Ulsan	-	false	overseas inflow	25	-	-
2000009	Gyeonggi-do	-	true	SMR Newly Planted...	25	-	-
6100010	Gyeongsangnam-do	-	false	overseas inflow	26	-	-
2000008	Gyeonggi-do	from other city	true	Yangcheon Table T...	28	-	-
2000007	Gyeonggi-do	from other city	true	Shincheonji Church	29	-	-
1000010	Seoul	Gwanak-gu	true	Wangsung Church	30	37.481735	126.930121
1100010	Busan	-	false	etc	30	-	-
1700001	Sejong	Sejong	true	Ministry of Ocean...	31	36.504713	127.265172
6100001	Gyeongsangnam-do	from other city	true	Shincheonji Church	32	-	-
1000007	Seoul	from other city	true	SMR Newly Planted...	36	-	-
1100008	Busan	-	false	overseas inflow	36	-	-
6000007	Gyeongsangbuk-do	Chilgok-gun	true	Milal Shelter	36	36.0581	128.4941
1200005	Daegu	Dong-gu	true	Fatima Hospital	39	35.88395	128.624059
1100001	Busan	Dongnae-gu	true	Onchun Church	39	35.21628	129.0771
6000006	Gyeongsangbuk-do	Yechun-gun	true	Yechun-gun	40	36.646845	128.437416
1000006	Seoul	Guro-gu	true	Manmin Central Ch...	41	37.481059	126.894343
1200008	Daegu	-	false	overseas inflow	41	-	-
6000005	Gyeongsangbuk-do	from other city	true	Pilgrimage to Israel	41	-	-
1400002	Incheon	from other city	true	Coupang Logistics...	42	-	-
1000005	Seoul	Dobong-gu	true	Day Care Center	43	37.679422	127.044374
1000004	Seoul	Yangcheon-gu	true	Yangcheon Table T...	43	37.546061	126.874209
2000005	Gyeonggi-do	Uijeongbu-si	true	Uijeongbu St. Mar...	50	37.758635	127.077716
2000006	Gyeonggi-do	from other city	true	Guro-gu Call Center	50	-	-
1400001	Incheon	from other city	true	Itaewon Clubs	53	-	-
1500001	Daejeon	-	true	Door-to-door sale...	55	-	-

2000004	Gyeonggi-do	from other city	true	Richway	58	-	-
2000003	Gyeonggi-do	from other city	true	Itaewon Clubs	59	-	-
2000021	Gyeonggi-do	- false	contact with patient	63	-	-	
6000004	Gyeongsangbuk-do	Gyeongsan-si	true	Gyeongsan Seorin ...	66	35.782149	128.801498
2000001	Gyeonggi-do	Seongnam-si	true	River of Grace Co...	67	37.455687	127.161627
2000002	Gyeonggi-do	Bucheon-si	true	Coupang Logistics...	67	37.530579	126.775254
1400005	Incheon	- false	overseas inflow	68	-	-	
6000003	Gyeongsangbuk-do	Bonghwa-gun	true	Bonghwa Pureun Nu...	68	36.92757	128.9099
2000022	Gyeonggi-do	- false	etc	84	-	-	
1000003	Seoul	Guro-gu	true	Guro-gu Call Center	95	37.508163	126.884387
1000038	Seoul	- false	etc	100	-	-	
1200004	Daegu	Dalseong-gun	true	Daesil Convalesce...	101	35.857393	128.466653
4100001	Chungcheongnam-do	Cheonan-si	true	gym facility in C...	103	36.81503	127.1139
1000002	Seoul	Gwanak-gu	true	Richway	119	37.48208	126.901384
6000002	Gyeongsangbuk-do	Cheongdo-gun	true	Cheongdo Daenam H...	119	35.64887	128.7368
1200003	Daegu	Seo-gu	true	Hansarang Convale...	124	35.885592	128.556649
6000013	Gyeongsangbuk-do	- false	etc	133	-	-	
1000001	Seoul	Yongsan-gu	true	Itaewon Clubs	139	37.538621	126.992652
1000037	Seoul	- false	contact with patient	162	-	-	
6000012	Gyeongsangbuk-do	- false	contact with patient	190	-	-	
1200002	Daegu	Dalseong-gun	true	Second Mi-Ju Hosp...	196	35.857375	128.466651
1000036	Seoul	- false	overseas inflow	298	-	-	
2000020	Gyeonggi-do	- false	overseas inflow	305	-	-	
6000001	Gyeongsangbuk-do	from other city	true	Shincheonji Church	566	-	-
1200010	Daegu	- false	etc	747	-	-	
1200009	Daegu	- false	contact with patient	917	-	-	
1200001	Daegu	Nam-gu	true	Shincheonji Church	4511	35.84008	128.5667

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

```
>>> df1.orderBy(col("confirmed").desc()).select("*").show(174)
```

case_id	province	city	group	infection_case	confirmed	latitude	longitude
1200001	Daegu	Nam-gu	true	Shincheonji Church	4511	35.84008	128.5667
1200009	Daegu	-	false	contact with patient	917	-	-
1200010	Daegu	-	false	etc	747	-	-
6000001	Gyeongsangbuk-do	from other city	true	Shincheonji Church	566	-	-
2000020	Gyeonggi-do	-	false	overseas inflow	305	-	-
1000036	Seoul	-	false	overseas inflow	298	-	-
1200002	Daegu	Dalseong-gun	true	Second Mi-Ju Hosp...	196	35.857375	128.466651
6000012	Gyeongsangbuk-do	-	false	contact with patient	190	-	-
1000037	Seoul	-	false	contact with patient	162	-	-
1000001	Seoul	Yongsan-gu	true	Itaewon Clubs	139	37.538621	126.992652
6000013	Gyeongsangbuk-do	-	false	etc	133	-	-
1200003	Daegu	Seo-gu	true	Hansarang Convale...	124	35.885592	128.556649
1000002	Seoul	Gwanak-gu	true	Richway	119	37.48208	126.901384
6000002	Gyeongsangbuk-do	Cheongdo-gun	true	Cheongdo Daenam H...	119	35.64887	128.7368
4100001	Chungcheongnam-do	Cheonan-si	true	gym facility in C...	103	36.81503	127.1139
1200004	Daegu	Dalseong-gun	true	Daesil Convalesce...	101	35.857393	128.466653
1000038	Seoul	-	false	etc	100	-	-
1000003	Seoul	Guro-gu	true	Guro-gu Call Center	95	37.508163	126.884387
2000022	Gyeonggi-do	-	false	etc	84	-	-
6000003	Gyeongsangbuk-do	Bonghwa-gun	true	Bonghwa Pureun Nu...	68	36.92757	128.9099
1400005	Incheon	-	false	overseas inflow	68	-	-
2000001	Gyeonggi-do	Seongnam-si	true	River of Grace Co...	67	37.455687	127.161627
2000002	Gyeonggi-do	Bucheon-si	true	Coupang Logistics...	67	37.530579	126.775254
6000004	Gyeongsangbuk-do	Gyeongsan-si	true	Gyeongsan Seorin ...	66	35.782149	128.801498
2000021	Gyeonggi-do	-	false	contact with patient	63	-	-
2000003	Gyeonggi-do	from other city	true	Itaewon Clubs	59	-	-
2000004	Gyeonggi-do	from other city	true	Richway	58	-	-
1500001	Daejeon	-	true	Door-to-door sale...	55	-	-
1400001	Incheon	from other city	true	Itaewon Clubs	53	-	-

2000006	Gyeonggi-do	from other city	true	Guro-gu Call Center	50	-	-
2000005	Gyeonggi-do	Uijeongbu-si	true	Uijeongbu St. Mar...	50	37.758635	127.077716
1000005	Seoul	Dobong-gu	true	Day Care Center	43	37.679422	127.044374
1000004	Seoul	Yangcheon-gu	true	Yangcheon Table T...	43	37.546061	126.874209
1400002	Incheon	from other city	true	Coupang Logistics...	42	-	-
1000006	Seoul	Guro-gu	true	Manmin Central Ch...	41	37.481059	126.894343
1200008	Daegu	-	false	overseas inflow	41	-	-
6000005	Gyeongsangbuk-do	from other city	true	Pilgrimage to Israel	41	-	-
6000006	Gyeongsangbuk-do	Yechun-gun	true	Yechun-gun	40	36.646845	128.437416
1100001	Busan	Dongnae-gu	true	Onchun Church	39	35.21628	129.0771
1200005	Daegu	Dong-gu	true	Fatima Hospital	39	35.88395	128.624059
6000007	Gyeongsangbuk-do	Chilgok-gun	true	Milal Shelter	36	36.0581	128.4941
1000007	Seoul	from other city	true	SMR Newly Planted...	36	-	-
1100008	Busan	-	false	overseas inflow	36	-	-
6100001	Gyeongsangnam-do	from other city	true	Shincheonji Church	32	-	-
1700001	Sejong	Sejong	true	Ministry of Ocean...	31	36.504713	127.265172
1000010	Seoul	Gwanak-gu	true	Wangsung Church	30	37.481735	126.930121
1100010	Busan	-	false	etc	30	-	-
2000007	Gyeonggi-do	from other city	true	Shincheonji Church	29	-	-
2000008	Gyeonggi-do	from other city	true	Yangcheon Table T...	28	-	-
6100010	Gyeongsangnam-do	-	false	overseas inflow	26	-	-
1600002	Ulsan	-	false	overseas inflow	25	-	-
1000009	Seoul	from other city	true	Coupang Logistics...	25	-	-
2000009	Gyeonggi-do	-	true	SMR Newly Planted...	25	-	-
1300003	Gwangju	-	false	overseas inflow	23	-	-
2000011	Gyeonggi-do	Anyang-si	true	Anyang Gunpo Past...	22	37.381784	126.93615
6000011	Gyeongsangbuk-do	-	false	overseas inflow	22	-	-
2000010	Gyeonggi-do	Seongnam-si	true	Bundang Jesaeng H...	22	37.38833	127.1218
1400003	Incheon	from other city	true	Guro-gu Call Center	20	-	-
6100012	Gyeongsangnam-do	-	false	etc	20	-	-
1100009	Busan	-	false	contact with patient	19	-	-
2000013	Gyeonggi-do	Anyang-si	true	Lord Glory Church	17	37.403722	126.954939
6000008	Gyeongsangbuk-do	Gyeongsan-si	true	Gyeongsan Jeil Si...	17	35.84819	128.7621
3000001	Gangwon-do	from other city	true	Shincheonji Church	17	-	-

1000008	Seoul Dongdaemun-gu true Dongan Church	17 37.592888 127.056766
3000006	Gangwon-do - false overseas inflow 16 - -	
1600001	Ulsan from other city true Shincheonji Church 16 - -	
4100006	Chungcheongnam-do - false overseas inflow 16 - -	
6000009	Gyeongsangbuk-do Gyeongsan-si true Gyeongsan Cham Jo... 16 35.82558 128.7373	
1500009	Daejeon - false contact with patient 15 - -	
1500010	Daejeon - false etc 15 - -	
2000012	Gyeonggi-do Suwon-si true Lotte Confectione... 15 37.287356 127.013827	
1500008	Daejeon - false overseas inflow 15 - -	
1000011	Seoul Eunpyeong-gu true Eunpyeong St. Mar... 14 37.63369 126.9165	
5100003	Jeollanam-do - false overseas inflow 14 - -	
7000001	Jeju-do - false overseas inflow 14 - -	
1500002	Daejeon Seo-gu true Dunsan Electronic... 13 36.3400973 127.3927099	
1000012	Seoul Seongdong-gu true Seongdong-gu APT 13 37.55713 127.0403	
4000005	Chungcheongbuk-do - false overseas inflow 13 - -	
1000023	Seoul Jung-gu true KB Life Insurance 13 37.560899 126.966998	
5000004	Jeollabuk-do - false overseas inflow 12 - -	
1100002	Busan from other city true Shincheonji Church 12 - -	
4100008	Chungcheongnam-do - false etc 12 - -	
4000007	Chungcheongbuk-do - false etc 11 - -	
1400007	Incheon - false etc 11 - -	
4000001	Chungcheongbuk-do Goesan-gun true Goesan-gun Jangye... 11 36.82422 127.9552	
6000010	Gyeongsangbuk-do Gumi-si true Gumi Elim Church 10 - -	
1000013	Seoul Jongno-gu true Jongno Community ... 10 37.57681 127.006	
6100002	Gyeongsangnam-do Geochang-gun true Geochang Church 10 35.68556 127.9127	
4100002	Chungcheongnam-do from other city true Door-to-door sale... 10 - -	
2000014	Gyeonggi-do Suwon-si true Suwon Saeng Myeon... 10 37.2376 127.0517	
3000002	Gangwon-do from other city true Uijeongbu St. Mar... 10 - -	
6100003	Gyeongsangnam-do Jinju-si true Wings Tower 9 35.164845 128.126969	
1300002	Gwangju from other city true Shincheonji Church 9 - -	
4000002	Chungcheongbuk-do from other city true Itaewon Clubs 9 - -	
4100003	Chungcheongnam-do Seosan-si true Seosan-si Laboratory 9 37.000354 126.354443	
1000021	Seoul from other city true Shincheonji Church 8 - -	
6100004	Gyeongsangnam-do Geochang-gun true Geochang-gun Woon... 8 35.805681 127.917805	

1700002	Sejong	Sejong	true gym facility in S...	8	36.48025	127.289
4000006	Chungcheongbuk-do	- false	contact with patient	8	-	-
1500003	Daejeon	Seo-gu	true Orange Town	7	36.3398739	127.3819744
1000014	Seoul	Gangnam-gu	true Samsung Medical C...	7	37.48825	127.08559
2000015	Gyeonggi-do	from other city	true Korea Campus Crus...	7	-	-
1000015	Seoul	Jung-gu	true Jung-gu Fashion C...	7	37.562405	126.984377
3000008	Gangwon-do	- false	etc	7	-	-
6100006	Gyeongsangnam-do	Changnyeong-gun	true Changnyeong Coin ...	7	35.54127	128.5008
1000017	Seoul	Jongno-gu	true Korea Campus Crus...	7	37.594782	126.968022
6100005	Gyeongsangnam-do	Changwon-si	true Hanmaeum Changwon...	7	35.22115	128.6866
1600004	Ulsan	- false	etc	7	-	-
1100004	Busan	Haeundae-gu	true Haeundae-gu Catho...	6	35.20599	129.1256
1000018	Seoul	Gangnam-gu	true Gangnam Yeoksam-d...	6	-	-
2000017	Gyeonggi-do	from other city	true Wangsung Church	6	-	-
1000020	Seoul	Geumcheon-gu	true Geumcheon-gu rice...	6	-	-
4000004	Chungcheongbuk-do	from other city	true Shincheonji Church	6	-	-
6100011	Gyeongsangnam-do	- false	contact with patient	6	-	-
1400006	Incheon	- false	contact with patient	6	-	-
2000016	Gyeonggi-do	from other city	true Geumcheon-gu rice...	6	-	-
1000016	Seoul	Seodaemun-gu	true Yeonana News Class	5	37.558147	126.943799
1000022	Seoul	from other city	true Guri Collective I...	5	-	-
1000027	Seoul	Seocho-gu	true Seocho Family	5	-	-
1300001	Gwangju	Dong-gu	true Gwangneuksa Temple	5	35.136035	126.956405
1100003	Busan	Suyeong-gu	true Suyeong-gu Kinder...	5	35.16708	129.1124
1300004	Gwangju	- false	contact with patient	5	-	-
2000019	Gyeonggi-do	Seongnam-si	true Seongnam neighbor...	5	-	-
2000018	Gyeonggi-do	from other city	true Seoul City Hall S...	5	-	-
1700004	Sejong	- false	overseas inflow	5	-	-
5000005	Jeollabuk-do	- false	etc	5	-	-
1500004	Daejeon	Seo-gu	true Dreaming Church	4	36.346869	127.368594
3000003	Gangwon-do	Wonju-si	true Wonju-si Apartments	4	37.342762	127.983815
1100006	Busan	from other city	true Itaewon Clubs	4	-	-
3000004	Gangwon-do	from other city	true Richway	4	-	-
3000005	Gangwon-do	from other city	true Geumcheon-gu rice...	4	-	-

5100005	Jeollanam-do	- false	etc	4	-	-
1100005	Busan	Jin-gu	true	Jin-gu Academy	4	35.17371 129.0633
1000029	Seoul	Gangnam-gu	true	Samsung Fire & Ma...	4	37.498279 127.030139
5100004	Jeollanam-do	- false	contact with patient	4	-	-
7000003	Jeju-do	- false	etc	4	-	-
1000031	Seoul	from other city	true	Yongin Brothers	4	- -
1000035	Seoul	Guro-gu	true	Daezayeon Korea	3	37.486837 126.893163
1500005	Daejeon	Seo-gu	true	Korea Forest Engi...	3	36.358123 127.388856
1000026	Seoul	Yangcheon-gu	true	Biblical Language...	3	37.524623 126.843118
1600003	Ulsan	- false	contact with patient	3	-	-
4100005	Chungcheongnam-do	from other city	true	Eunpyeong-Boksago...	3	- -
1000032	Seoul	Jung-gu	true	Seoul City Hall S...	3	37.565699 126.977079
5000002	Jeollabuk-do	from other city	true	Door-to-door sale...	3	- -
6100007	Gyeongsangnam-do	Yangsan-si	true	Soso Seowon	3	35.338811 129.017508
1700005	Sejong	- false	contact with patient	3	-	-
4100004	Chungcheongnam-do	from other city	true	Richway	3	- -
1000024	Seoul	Yeongdeungpo-gu	true	Yeongdeungpo Lear...	3	37.520846 126.931278
1200007	Daegu	from other city	true	Cheongdo Daenam H...	2	- -
1400004	Incheon	from other city	true	Shincheonji Church	2	- -
1500006	Daejeon	from other city	true	Shincheonji Church	2	- -
1500007	Daejeon	from other city	true	Seosan-si Laboratory	2	- -
4100007	Chungcheongnam-do	- false	contact with patient	2	-	-
6100009	Gyeongsangnam-do	from other city	true	Onchun Church	2	- -
4000003	Chungcheongbuk-do	from other city	true	Guro-gu Call Center	2	- -
5000001	Jeollabuk-do	from other city	true	Itaewon Clubs	2	- -
5100001	Jeollanam-do	Muan-gun	true	Manmin Central Ch...	2	35.078825 126.316746
1200006	Daegu	from other city	true	Itaewon Clubs	2	- -
6100008	Gyeongsangnam-do	from other city	true	Itaewon Clubs	2	- -
1000033	Seoul	from other city	true	Uiwang Logistics ...	2	- -
1000019	Seoul	from other city	true	Daejeon door-to-d...	1	- -
1100007	Busan	from other city	true	Cheongdo Daenam H...	1	- -
1000025	Seoul	Gangnam-gu	true	Gangnam Dongin Ch...	1	37.522331 127.057388
1300005	Gwangju	- false	etc	1	-	-
5000003	Jeollabuk-do	from other city	true	Shincheonji Church	1	- -

1000028	Seoul	from other city	true	Anyang Gunpo Past...	1	-	-
1700003	Sejong	from other city	true	Shincheonji Church	1	-	-
5100002	Jeollanam-do	from other city	true	Shincheonji Church	1	-	-
1700006	Sejong	- false	etc	1	-	-	
1000034	Seoul	- true	Orange Life	1	-	-	
7000004	Jeju-do	from other city	true	Itaewon Clubs	1	-	-
1000030	Seoul	Gangseo-gu	true	SJ Investment Cal...	0	37.559649	126.835102
3000007	Gangwon-do	- false	contact with patient	0	-	-	
7000002	Jeju-do	- false	contact with patient	0	-	-	

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

Q10. In case of any wrong data type, cast that data type from integer to string or string to integer.

Converting df1.latitude, df1.longitude from string to float

```
>>> df1.printSchema()
```

```
root
```

```
|-- case_id: integer (nullable = true)
|-- province: string (nullable = true)
|-- city: string (nullable = true)
|-- group: boolean (nullable = true)
|-- infection_case: string (nullable = true)
|-- confirmed: integer (nullable = true)
|-- latitude: string (nullable = true)
|-- longitude: string (nullable = true)
```

```
>>> df1_castf=df1. withColumn("latitude", col("latitude").cast("float"))
```

```
>>> df1_castd = df1_castf.withColumn("latitude", col("latitude").cast("float"))
```

```
>>> df1_castd.printSchema()
```

```
root
```

```
|-- case_id: integer (nullable = true)
|-- province: string (nullable = true)
|-- city: string (nullable = true)
|-- group: boolean (nullable = true)
|-- infection_case: string (nullable = true)
```



```
|-- confirmed: integer (nullable = true)
|-- latitude: float (nullable = true)
|-- longitude: float (nullable = true)
```

Q11. Use group by on top of province and city column and agg it with sum of confirmed cases.

```
>>> df.groupby(["province", "city"]).agg(sum("confirmed").alias("total_sum")).show()
```

```
+-----+-----+-----+
| province| city|total_sum|
+-----+-----+-----+
|Gyeongsangnam-do| Jinju-si| 9|
| Seoul| Guro-gu| 139|
| Seoul| Gangnam-gu| 18|
| Daejeon| -| 100|
| Jeollabuk-do|from other city| 6|
|Gyeongsangnam-do|Changnyeong-gun| 7|
| Seoul| -| 561|
| Jeju-do|from other city| 1|
|Gyeongsangbuk-do| -| 345|
|Gyeongsangnam-do| Geochang-gun| 18|
|Gyeongsangbuk-do| Gumi-si| 10|
| Incheon|from other city| 117|
| Busan| -| 85|
| Daegu| Seo-gu| 124|
| Busan| Suyeong-gu| 5|
| Gyeonggi-do| Uijeongbu-si| 50|
| Seoul| Yongsan-gu| 139|
| Daegu| -| 1705|
| Seoul| Seodaemun-gu| 5|
| Gyeonggi-do| Seongnam-si| 94|
+-----+-----+-----+
```

Q11. For joins we will need one more file.you can use region file. User different different join methods.for example cases.join(regions, ['province','city'],how='left') You can do your best analysis.

```
>>> df4=df1.join(df2,df1.city==df2.city,"inner").select("*")
```

```
>>> df4.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|case_id|province|city|group|infection_case|confirmed|latitude|longitude|code|province|city|
latitude|
longitude|elementary_school_count|kindergarten_count|university_count|academy_ratio|elderly_population_rati
o|elderly_alone_ratio|nursing_home_count|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|1000029|Seoul|Gangnam-gu|true|Samsung Fire & Ma...|4|37.498279|127.030139|10010|Seoul|
Gangnam-gu|37.518421|127.047222|33|38|0|4.18|13.17|
4.3|3088|
|1000025|Seoul|Gangnam-gu|true|Gangnam Dongin Ch...|1|37.522331|127.057388|10010|Seoul|
Gangnam-gu|37.518421|127.047222|33|38|0|4.18|13.17|
4.3|3088|
|1000018|Seoul|Gangnam-gu|true|Gangnam Yeoksam-d...|6|-|-|10010|Seoul|Gangnam-
gu|37.518421|127.047222|33|38|0|4.18|13.17|4.3|
3088|
|1000014|Seoul|Gangnam-gu|true|Samsung Medical C...|7|37.48825|127.08559|10010|Seoul|
Gangnam-gu|37.518421|127.047222|33|38|0|4.18|13.17|
4.3|3088|
|1000030|Seoul|Gangseo-gu|true|SJ Investment Cal...|0|37.559649|126.835102|10040|Seoul|
Gangseo-gu|37.551166|126.849506|36|56|1|1.17|14.39|
5.7|1080|
|1000010|Seoul|Gwanak-gu|true|Wangsung Church|30|37.481735|126.930121|10050|Seoul|
Gwanak-gu|37.47829|126.951502|22|33|1|0.89|15.12|4.9|
909|
|1000002|Seoul|Gwanak-gu|true|Richway|119|37.48208|126.901384|10050|Seoul|
Gwanak-gu|37.47829|126.951502|22|33|1|0.89|15.12|4.9|
909|
|1000035|Seoul|Guro-gu|true|Daezayeon Korea|3|37.486837|126.893163|10070|Seoul|
Guro-gu|37.495632|126.88765|26|34|3|1.0|16.21|5.7|
741|
|1000006|Seoul|Guro-gu|true|Manmin Central Ch...|41|37.481059|126.894343|10070|Seoul|
Guro-gu|37.495632|126.88765|26|34|3|1.0|16.21|5.7|
741|
|1000003|Seoul|Guro-gu|true|Guro-gu Call Center|95|37.508163|126.884387|10070|Seoul|
Guro-gu|37.495632|126.88765|26|34|3|1.0|16.21|5.7|
741|
```

1000020	Seoul	Geumcheon-gu	true	Geumcheon-gu rice...	6	-	- 10080	Seoul	Geumcheon-
gu 37.456852	126.895229	18	19	0	0.96	16.15	6.7	475	
1000005	Seoul	Dobong-gu	true	Day Care Center	43	37.679422	127.044374	10100	Seoul
Dobong-gu 37.668952	127.047082	23	26	1	0.95	17.89	7.2	485	
1000008	Seoul	Dongdaemun-gu	true	Dongan Church	17	37.592888	127.056766	10110	Seoul
Dongdaemun-gu 37.574552	127.039721	21	31	4	1.06	17.26	6.7	832	
1000016	Seoul	Seodaemun-gu	true	Yeonana News Class	5	37.558147	126.943799	10140	Seoul
Seodaemun-gu 37.579428	126.936771	19	25	6	1.12	16.77	6.2	587	
1000027	Seoul	Seocho-gu	true	Seocho Family	5	-	- 10150	Seoul	Seocho-
gu 37.483804	127.032693	24	27	1	2.6	13.39	3.8	1465	
1000012	Seoul	Seongdong-gu	true	Seongdong-gu APT	13	37.55713	127.0403	10160	Seoul
Seongdong-gu 37.563277	127.036647	21	30	2	0.97	14.76	5.3	593	
1000026	Seoul	Yangcheon-gu	true	Biblical Language...	3	37.524623	126.843118	10190	Seoul
Yangcheon-gu 37.517189	126.866618	30	43	0	2.26	13.55	5.5	816	
1000004	Seoul	Yangcheon-gu	true	Yangcheon Table T...	43	37.546061	126.874209	10190	Seoul
Yangcheon-gu 37.517189	126.866618	30	43	0	2.26	13.55	5.5	816	
1000024	Seoul	Yeongdeungpo-gu	true	Yeongdeungpo Lear...	3	37.520846	126.931278	10200	
Seoul Yeongdeungpo-gu 37.526505	126.89619	23	39	0	1.21	15.6	5.8	1001	
1000001	Seoul	Yongsan-gu	true	Itaewon Clubs	139	37.538621	126.992652	10210	Seoul
Yongsan-gu 37.532768	126.990021	15	13	1	0.68	16.87	6.5	435	

```

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

```

Q12. If you want, you can also use SQL with data frames. Let us try to run some SQL on the cases tab

```
>>> df1.registerTempTable('cases_table')
```

```
>>> newDF = sqlContext.sql('select * from cases_table where confirmed > 100')
```

```
>>> newDf.show()
```

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|case_id|  province|    city|group|  infection_case|confirmed| latitude| longitude|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|1000001|    Seoul|  Yongsan-gu| true|  Itaewon Clubs|    139|37.538621|126.992652|
|1000002|    Seoul|  Gwanak-gu| true|    Richway|    119| 37.48208|126.901384|

```

1000036	Seoul	- false	overseas inflow	298	-	-
1000037	Seoul	- false	contact with patient	162	-	-
1200001	Daegu	Nam-gu	true Shincheonji Church	4511	35.84008	128.5667
1200002	Daegu	Dalseong-gun	true Second Mi-Ju Hosp...	196	35.857375	128.466651
1200003	Daegu	Seo-gu	true Hansarang Convale...	124	35.885592	128.556649
1200004	Daegu	Dalseong-gun	true Daesil Convalesce...	101	35.857393	128.466653
1200009	Daegu	- false	contact with patient	917	-	-
1200010	Daegu	- false	etc	747	-	-
2000020	Gyeonggi-do	- false	overseas inflow	305	-	-
4100001	Chungcheongnam-do	Cheonan-si	true gym facility in C...	103	36.81503	127.1139
6000001	Gyeongsangbuk-do	from other city	true Shincheonji Church	566	-	-
6000002	Gyeongsangbuk-do	Cheongdo-gun	true Cheongdo Daenam H...	119	35.64887	128.7368
6000012	Gyeongsangbuk-do	- false	contact with patient	190	-	-
6000013	Gyeongsangbuk-do	- false	etc	133	-	-

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

Q13. Create Spark UDFs Create function casehighlow() If case is less than 50 return low else return high convert into a UDF Function and mention the return type of function. Note: You can create as many as udf based on analysis.

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import udf
from pyspark.sql.types import StringType

spark = SparkSession.builder.getOrCreate()

def casehighlow(case_value):
    if case_value < 50:
        return "low"
    else:
        return "high"

casehighlow_udf = udf(casehighlow, StringType())
df1=spark.read.option("header",True).option("inferSchema",True).csv("/project/case.csv")
df = df1.withColumn("highlow", casehighlow_udf("confirmed"))
df.show()
```

case_id	province	city	group	infection_case	confirmed	latitude	longitude	highlow
1000001	Seoul	Yongsan-gu	true	Itaewon Clubs	139	37.538621	126.992652	high
1000002	Seoul	Gwanak-gu	true	Richway	119	37.48208	126.901384	high
1000003	Seoul	Guro-gu	true	Guro-gu Call Center	95	37.508163	126.884387	high
1000004	Seoul	Yangcheon-gu	true	Yangcheon Table T...	43	37.546061	126.874209	low
1000005	Seoul	Dobong-gu	true	Day Care Center	43	37.679422	127.044374	low
1000006	Seoul	Guro-gu	true	Manmin Central Ch...	41	37.481059	126.894343	low
1000007	Seoul	from other city	true	SMR Newly Planted...	36	-	-	low
1000008	Seoul	Dongdaemun-gu	true	Dongan Church	17	37.592888	127.056766	low
1000009	Seoul	from other city	true	Coupang Logistics...	25	-	-	low
1000010	Seoul	Gwanak-gu	true	Wangsung Church	30	37.481735	126.930121	low
1000011	Seoul	Eunpyeong-gu	true	Eunpyeong St. Mar...	14	37.63369	126.9165	low
1000012	Seoul	Seongdong-gu	true	Seongdong-gu APT	13	37.55713	127.0403	low
1000013	Seoul	Jongno-gu	true	Jongno Community ...	10	37.57681	127.006	low
1000014	Seoul	Gangnam-gu	true	Samsung Medical C...	7	37.48825	127.08559	low
1000015	Seoul	Jung-gu	true	Jung-gu Fashion C...	7	37.562405	126.984377	low
1000016	Seoul	Seodaemun-gu	true	Yeonana News Class	5	37.558147	126.943799	low
1000017	Seoul	Jongno-gu	true	Korea Campus Crus...	7	37.594782	126.968022	low
1000018	Seoul	Gangnam-gu	true	Gangnam Yeoksam-d...	6	-	-	low
1000019	Seoul	from other city	true	Daejeon door-to-d...	1	-	-	low
1000020	Seoul	Geumcheon-gu	true	Geumcheon-gu rice...	6	-	-	low