PA   PART1   Version control

Current File

Project

PART1 ~/PycharmProjects/PART1
.venv
bin
lib
share
.gitignore
pyvenv.cfg

main.py

```python
df.show(10)


df_with_json = df.withColumn( colName: "value", to_json(struct(*df.columns)))


df to kafka = df with json withColumn( colName: "key", col("athlete id") cast("string")) \
```

Run   main

```
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
+----------------+----------+-----------+----------+----------------+---------+------------+-------+----+-----------+
|         edition|edition_id|country_noc|     sport|           event|result_id|     athlete|athlete_id| pos|medal|isTeamSport|
+----------------+----------+-----------+----------+----------------+---------+------------+-------+----+-----------+
|1908SummerOlympics|         5|        ANZ|Athletics|   100metres,Men|    56265|ErnestHutcheon|  64710| DNS| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|   400metres,Men|    56313|   HenryMurray|  64756| DNS| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|   800metres,Men|    56338|  HarveySutton|  64808|3h8r12| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|   800metres,Men|    56338|    GuyHaskins| 922519| DNS| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|   800metres,Men|    56338|   JosephLynch|  64735| DNS| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|   800metres,Men|    56338|   HenryMurray|  64756| DNS| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|1,500metres,Men|    56349|   JosephLynch|  64735|5h2r12| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|1,500metres,Men|    56349|  CharlesSwain|  79576|ACh3r12| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|1,500metres,Men|    56349|    GuyHaskins| 922519| DNS| nan|      False|
|1908SummerOlympics|         5|        ANZ|Athletics|1,500metres,Men|    56349|   GeorgeBlake|  64619| DNS| nan|      False|
+----------------+----------+-----------+----------+----------------+---------+------------+-------+----+-----------+
only showing top 10 rows


Process finished with exit code 0
```

PART1 > main.py                4:1   LF   UTF-8   4 spaces   Python 3.12 (PART1)

ChatGPT can make mistakes. Check important info.

{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"TeamPursuit8laps,Men","result_id":19020410,"athlete":"JoeyMantia","athlete_id":128784,"pos":"3","medal":"Bronze","isTeamSport":"True"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"500metres,Women","result_id":19020424,"athlete":"ErinJackson","athlete_id":138378,"pos":"1","medal":"Gold","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"500metres,Women","result_id":19020424,"athlete":"BrittanyBowe","athlete_id":128778,"pos":"16","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"500metres,Women","result_id":19020424,"athlete":"KimiGoetz","athlete_id":149162,"pos":"18","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"1,000metres,Women","result_id":19020425,"athlete":"BrittanyBowe","athlete_id":128778,"pos":"3","medal":"Bronze","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"1,000metres,Women","result_id":19020425,"athlete":"KimiGoetz","athlete_id":149162,"pos":"7","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"1,500metres,Women","result_id":19020426,"athlete":"BrittanyBowe","athlete_id":128778,"pos":"10","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"1,500metres,Women","result_id":19020426,"athlete":"MiaKilburg","athlete_id":138379,"pos":"20","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"3,000metres,Women","result_id":19020427,"athlete":"MiaKilburg","athlete_id":138379,"pos":"19","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"MassStart,Women","result_id":19020429,"athlete":"MiaKilburg","athlete_id":138379,"pos":"4","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"SpeedSkating","event":"MassStart,Women","result_id":19020429,"athlete":"GiorgiaBirkeland","athlete_id":149161,"pos":"12","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"KaillieHumphries","athlete_id":118672,"pos":"1","medal":"Gold","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"USA","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"ElanaMeyersTaylor","athlete_id":118644,"pos":"2","medal":"Silver","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"CAN","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"ChristinedeBruin","athlete_id":136825,"pos":"3","medal":"Bronze","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"GER","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"LauraNolte","athlete_id":2301452,"pos":"4","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"AUS","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"BreeanaWalker","athlete_id":147822,"pos":"5","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"CHN","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"HuaiMingming","athlete_id":148023,"pos":"6","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"SUI","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"MelanieHasler","athlete_id":148906,"pos":"7","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"CAN","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"CynthiaAppiah","athlete_id":147913,"pos":"8","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"CHN","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"YingQing","athlete_id":148025,"pos":"9","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"ROC","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"NadezhdaSergeyeva","athlete_id":128310,"pos":"10","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"FRA","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"MargotBoch","athlete_id":2301119,"pos":"11","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"ROU","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"AndreeaGrecu","athlete_id":128289,"pos":"12","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"GER","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"MariamaJamanka","athlete_id":137253,"pos":"13","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"AUT","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"KatiBeierl","athlete_id":136713,"pos":"14","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"ITA","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"GiadaAndreutti","athlete_id":148470,"pos":"15","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"NED","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"KarlienSleper","athlete_id":148662,"pos":"16","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"SVK","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"Viktriaeransk","athlete_id":138683,"pos":"17","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"KOR","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"KimYuRan","athlete_id":137542,"pos":"18","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"JAM","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"JazmineFenlatorVictorian","athlete_id":128682,"pos":"19","medal":"nan","isTeamSport":"False"}
{"edition":"2022WinterOlympics","edition_id":62,"country_noc":"UKR","sport":"Bobsleigh","event":"Monobob,Women","result_id":19019671,"athlete":"LidiyaHunko","athlete_id":149047,"pos":"20","medal":"nan","isTeamSport":"False"}

SQL ▼  Commit  Rollback  Auto  ⟲ ▼  olympic_dataset 2 ▼  olympic_dataset ▼

Database Navigator ✕

Enter a part of object name here

> mothers - localhost:5432
∨ olympic_dataset 2 - 217.61.57.46:3306
  > Databases
  > Users
  > Administer
  > System Info
> postgres - localhost:5432
> test

test | mothers | *<olympic_dataset 2> Script-5 ✕

SELECT * FROM athlete_aggregates;

athlete_aggregates 1 ✕

SELECT * FROM athlete_aggregates    Enter a SQL expression to filter results (use Ctrl+Space)

| | sport | medal | gender | country_noc | avg_height | avg_weight | timestamp |
|---|---|---|---|---|---|---|---|
| 1 | Swimming | nan | Male | ANZ | 170 | 65 | 025-04-11 15:18:02 |
| 2 | Swimming | Silver | Male | ANZ | 170 | 65 | 025-04-11 15:18:02 |
| 3 | Rowing | nan | Male | ANZ | 189 | 85 | 025-04-11 15:18:02 |
| 4 | Swimming | Bronze | Male | ANZ | 170 | 65 | 025-04-11 15:18:02 |
| 5 | Athletics | Bronze | Male | ANZ | 184 | 76 | 025-04-11 15:18:02 |
| 6 | Rugby | Gold | Male | ANZ | 175 | 75 | 025-04-11 15:18:02 |
| 7 | Athletics | nan | Male | ANZ | 170.8333333333 | 66 | 025-04-11 15:18:02 |
| 8 | Tennis | Bronze | Male | ANZ | 188 | 84 | 025-04-11 15:18:02 |
| 9 | Boxing | nan | Male | ANZ | 167 | 62 | 025-04-11 15:18:02 |
| 10 | Swimming | nan | Female | POR | 169.1 | 59.4666666667 | 2025-04-11 15:18:10 |
| 11 | Sailing | nan | Female | BER | 169.1428571429 | 64.1428571429 | 2025-04-11 15:18:10 |
| 12 | CyclingRoad | nan | Male | BER | 179.6666666667 | 72.6666666667 | 2025-04-11 15:18:10 |
| 13 | Shooting | Silver | Male | FRA | 179.6 | 75.8 | 2025-04-11 15:18:10 |
| 14 | Swimming | Silver | Female | UKR | 182 | 70 | 2025-04-11 15:18:10 |
| 15 | CanoeSprint | Gold | Female | FIN | 170 | 5,962 | 2025-04-11 15:18:10 |
| 16 | CyclingRoad | nan | Female | TPE | 161.6666666667 | 55.6666666667 | 2025-04-11 15:18:10 |
| 17 | Sailing | nan | Male | DOM | 189 | 82 | 2025-04-11 15:18:10 |
| 18 | Taekwondo | nan | Female | JPN | 170.6666666667 | 58.8333333333 | 2025-04-11 15:18:10 |
| 19 | Boxing | Gold | Male | FRA | 174.3333333333 | 67.6666666667 | 2025-04-11 15:18:10 |
| 20 | Golf | nan | Female | CAN | 165.5 | 59.5 | 2025-04-11 15:18:10 |
| 21 | CanoeSprint | nan | Female | TUN | 173 | 60 | 2025-04-11 15:18:10 |
| | Football | nan | Male | AUS | 182.0849056604 | 78.0433962264 | 2025-04-11 15:18:10 |

Refresh ▼ | Save ▼ | Cancel | ◀ ◀ ▶ ▶ | Export data ▼ | 200 | 200+

200 row(s) fetched - 0.125s (0.081s fetch), on 2025-04-11 at 15:20:39

Error Log ✕

Workspace Log

type filter text

CET | en_GB | Writable | Smart Insert | 1 : 34 : 33 | Sel: 0 | 0