OLYMPIC DATA ANALYSIS PROJECT

REPORT

In this project , I first loaded Olympic dataset on cloudera machine by creating shared folder both on windows and linux OS.

Then I create directory to load the csv files into it using hadoop and log in to hive shell. In hive I created table of data from csv file ,run queries and do analysis of data.

1) Creating Directory and Loading csv files into it

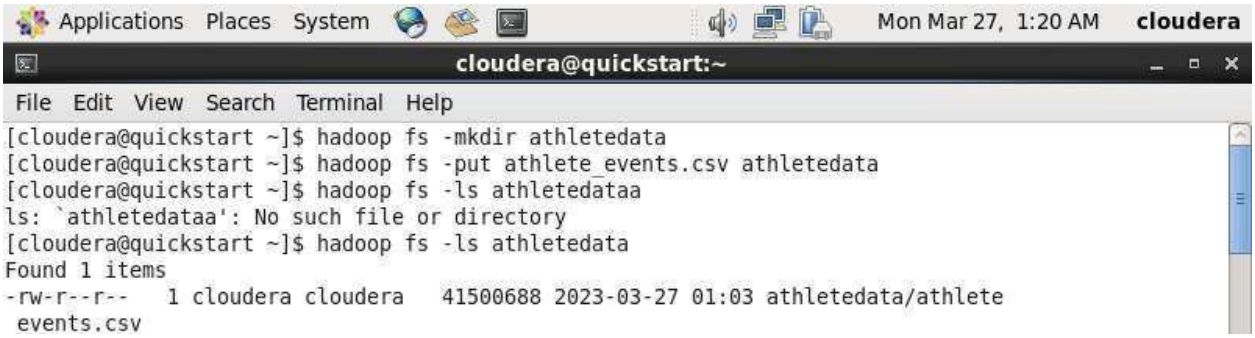
Command :

$ hadoop fs -mkdir athletedata

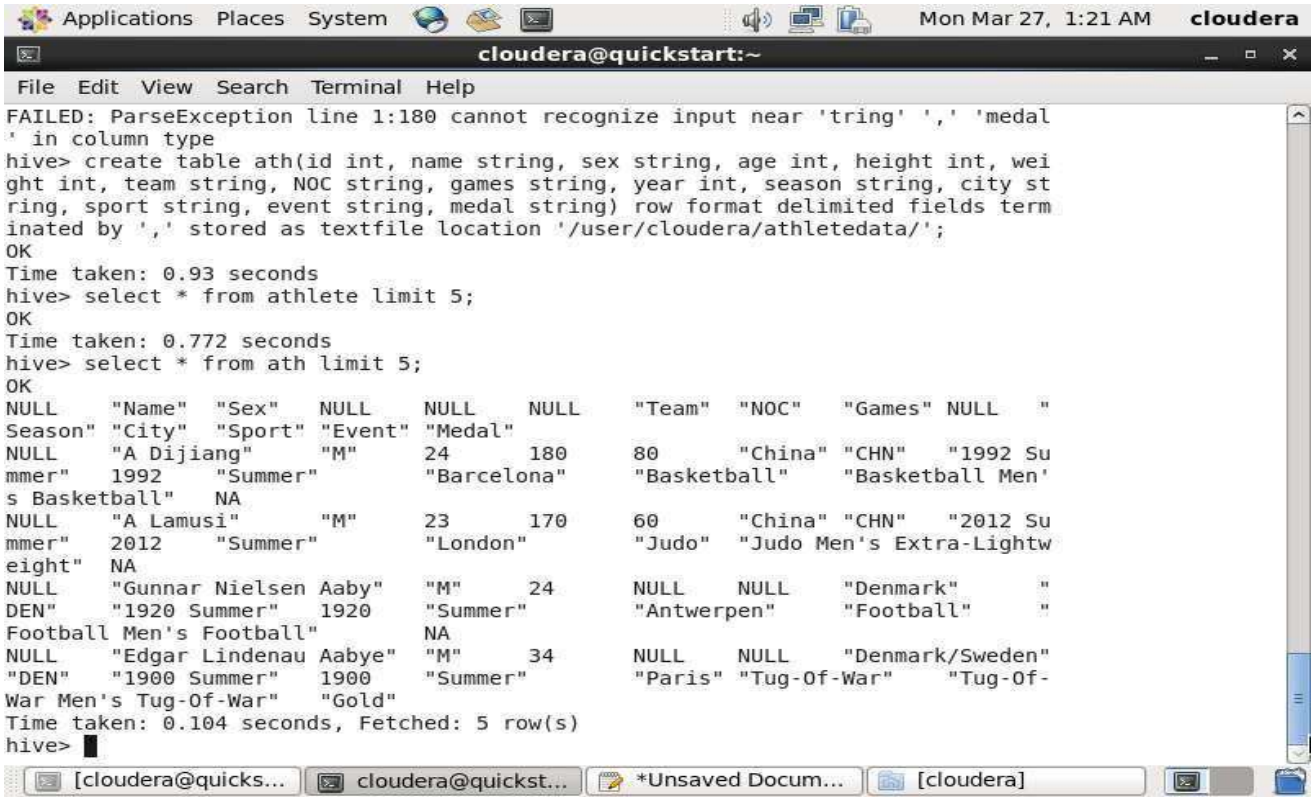
$ hadoop fs -put noc_regions.csv athletedata

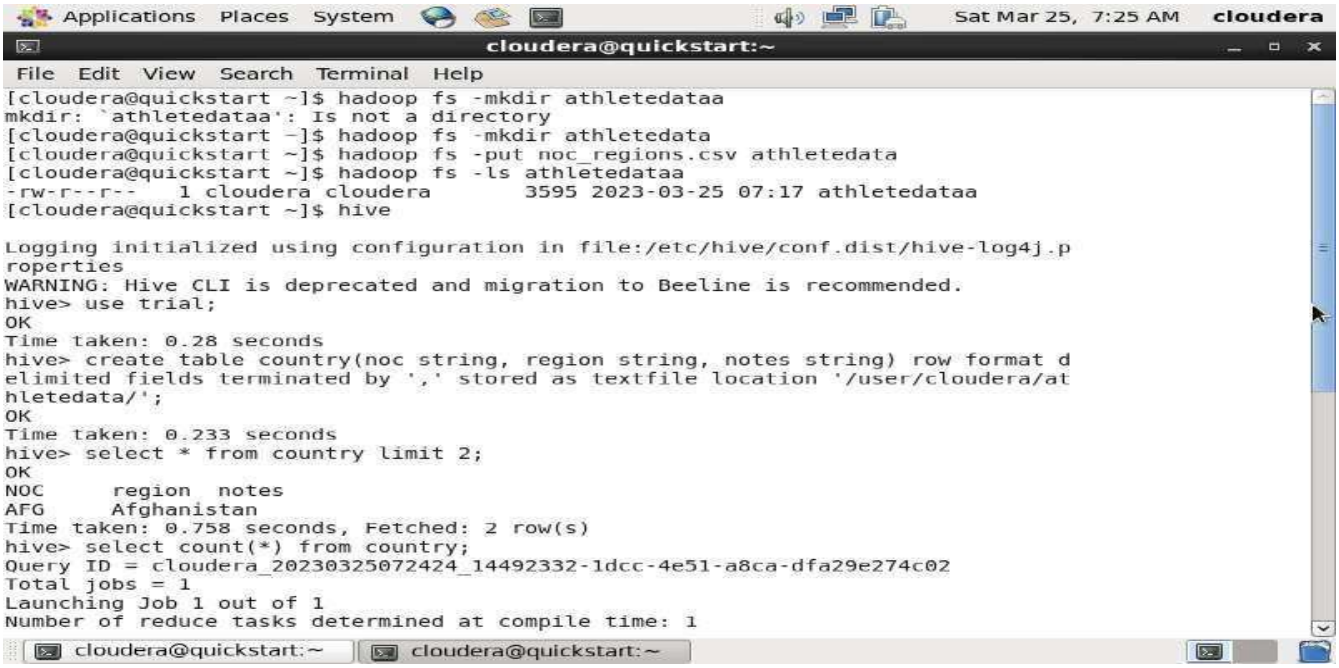
$ hadoop fs -put athlete_events.csv athletedata

$ hadoop fs -ls athletedata



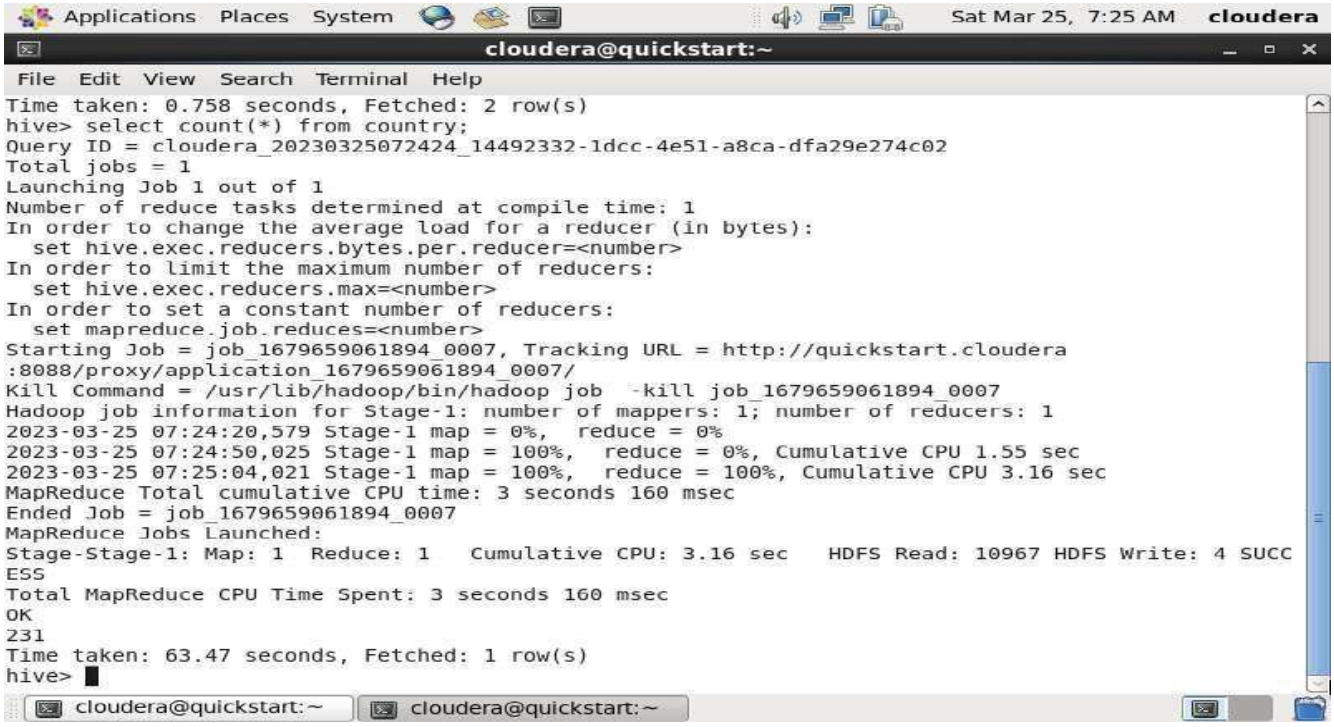
2) Creating Tables



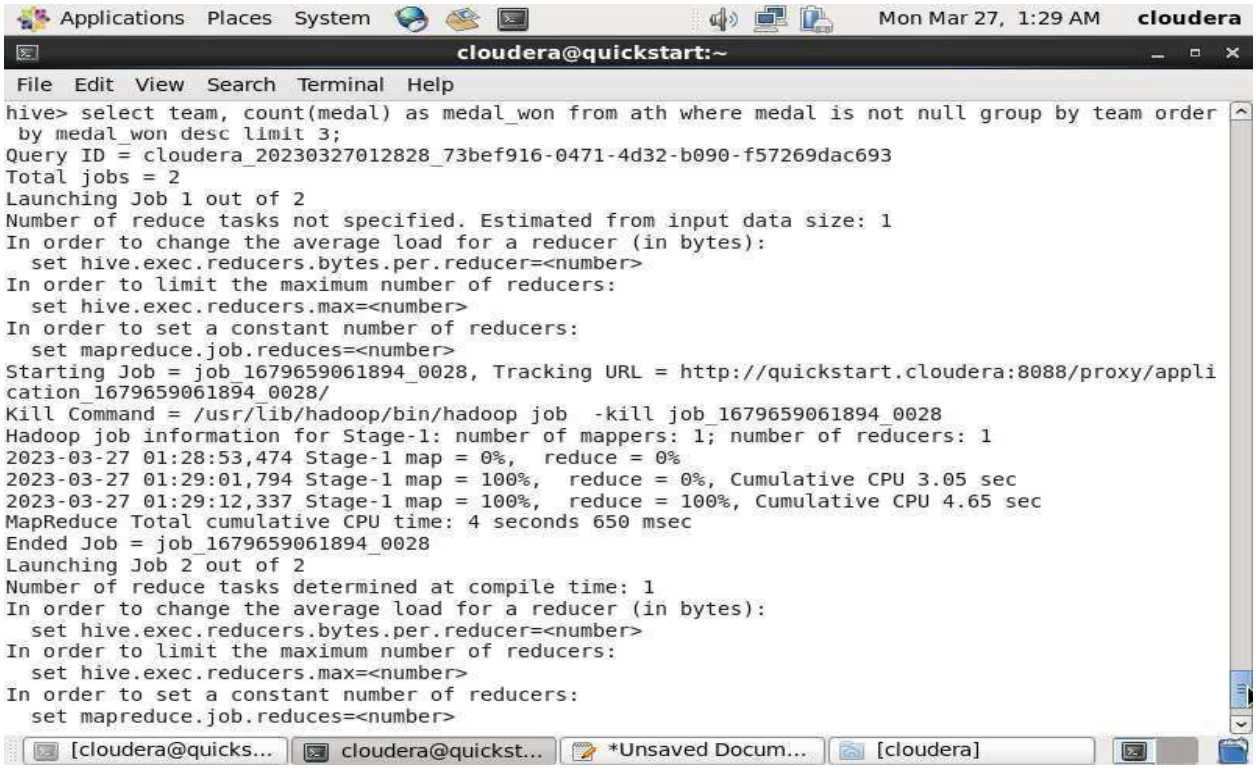


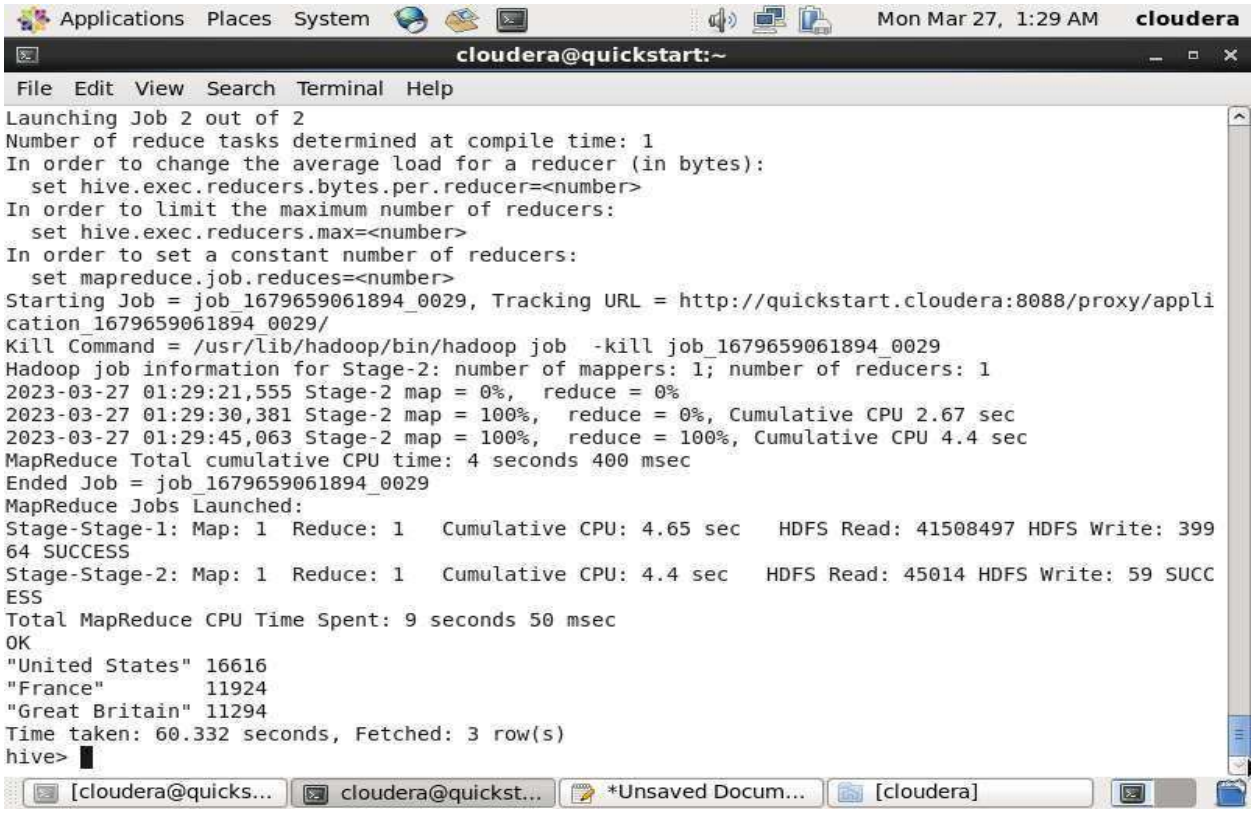
3) Queries

Counting number of countries participated in Olympics

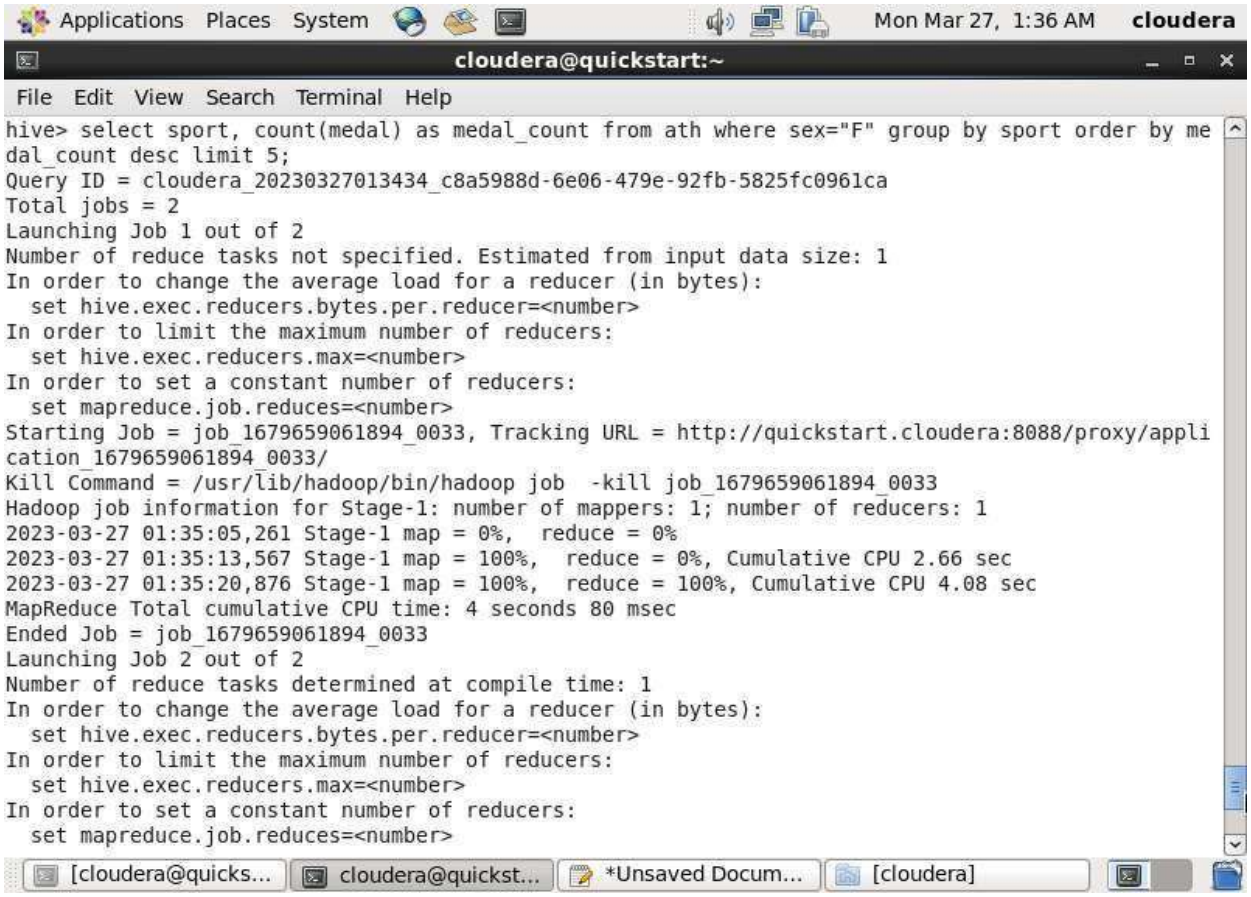


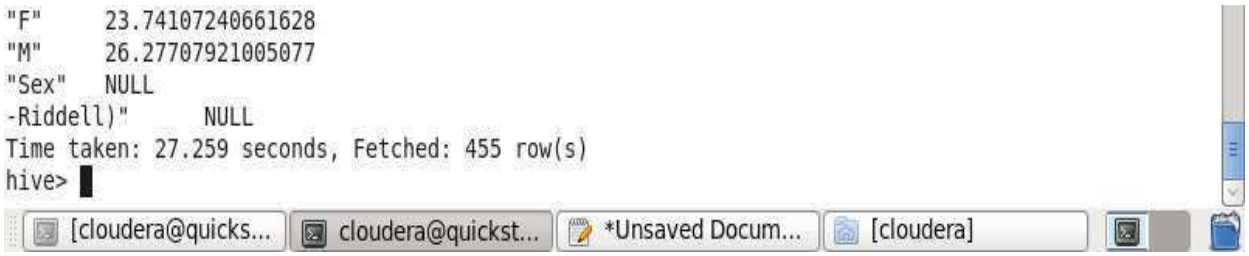
Finding top 3 countries in terms of total number of medals



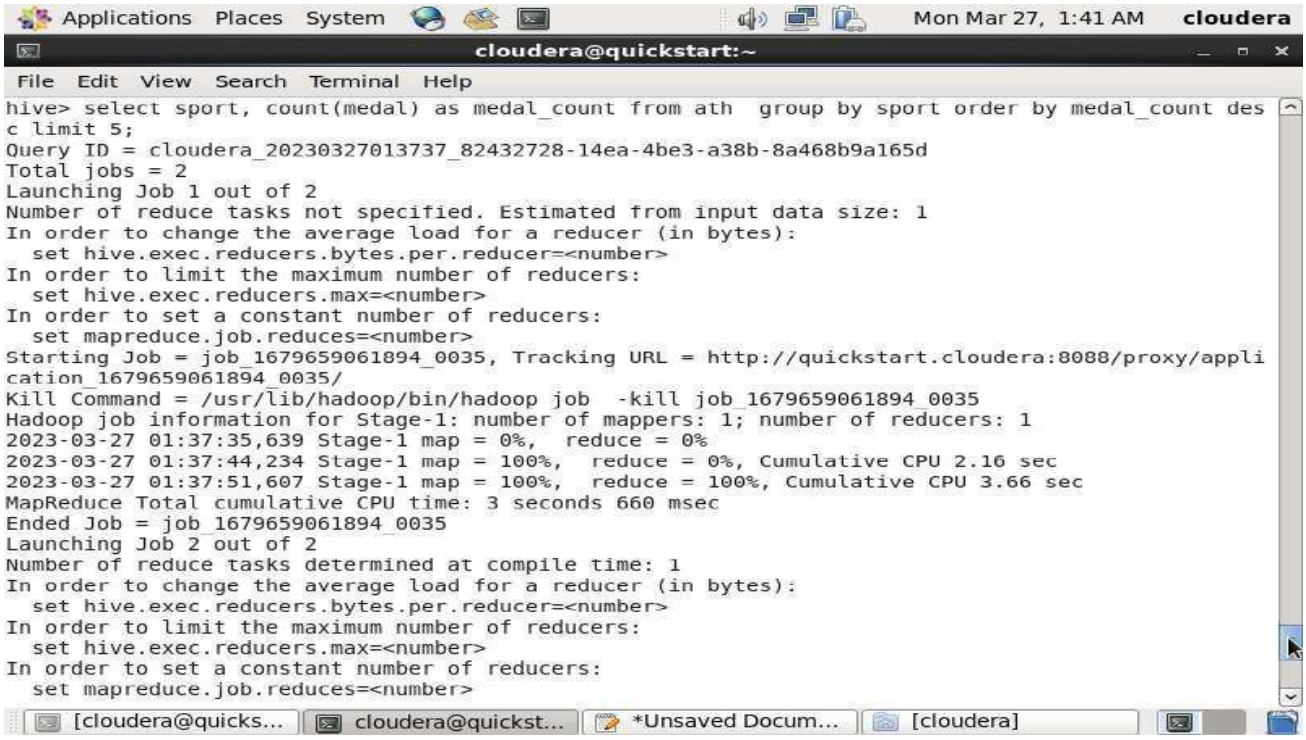


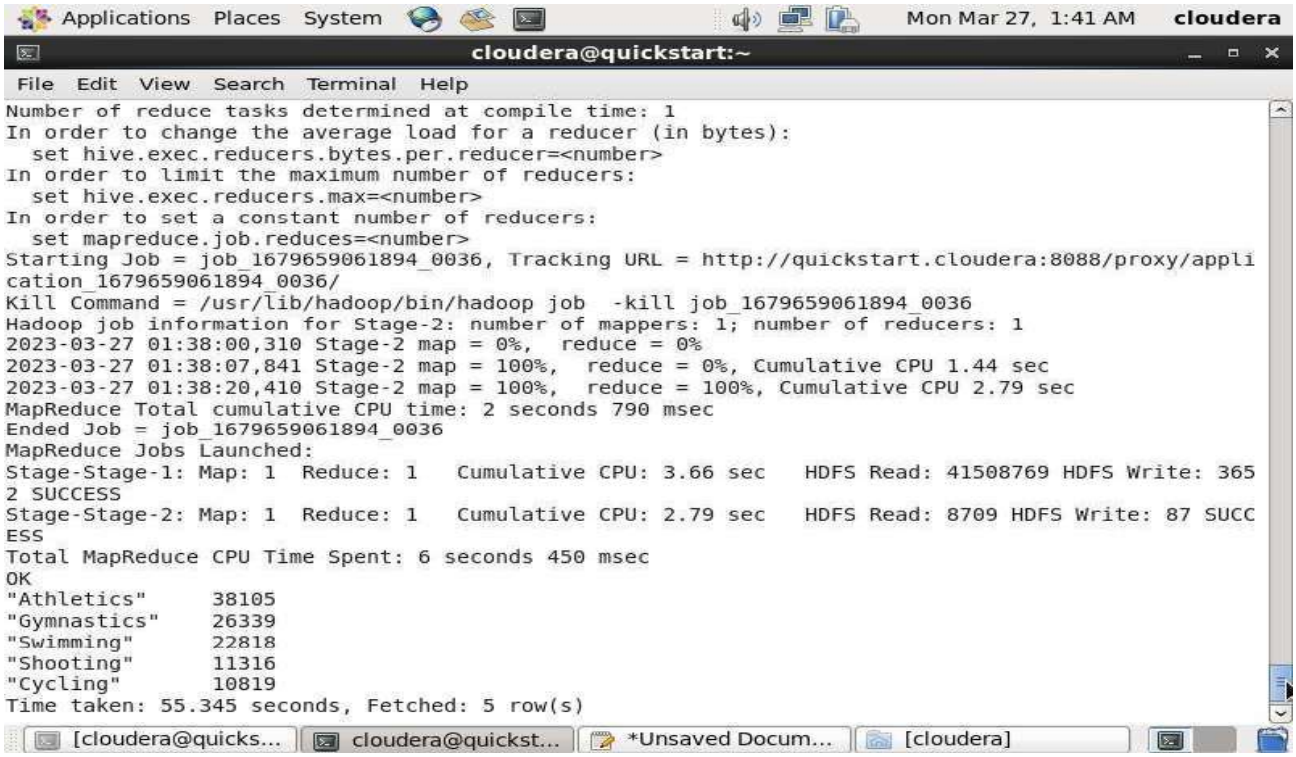
Sports in which female athlete have won medals



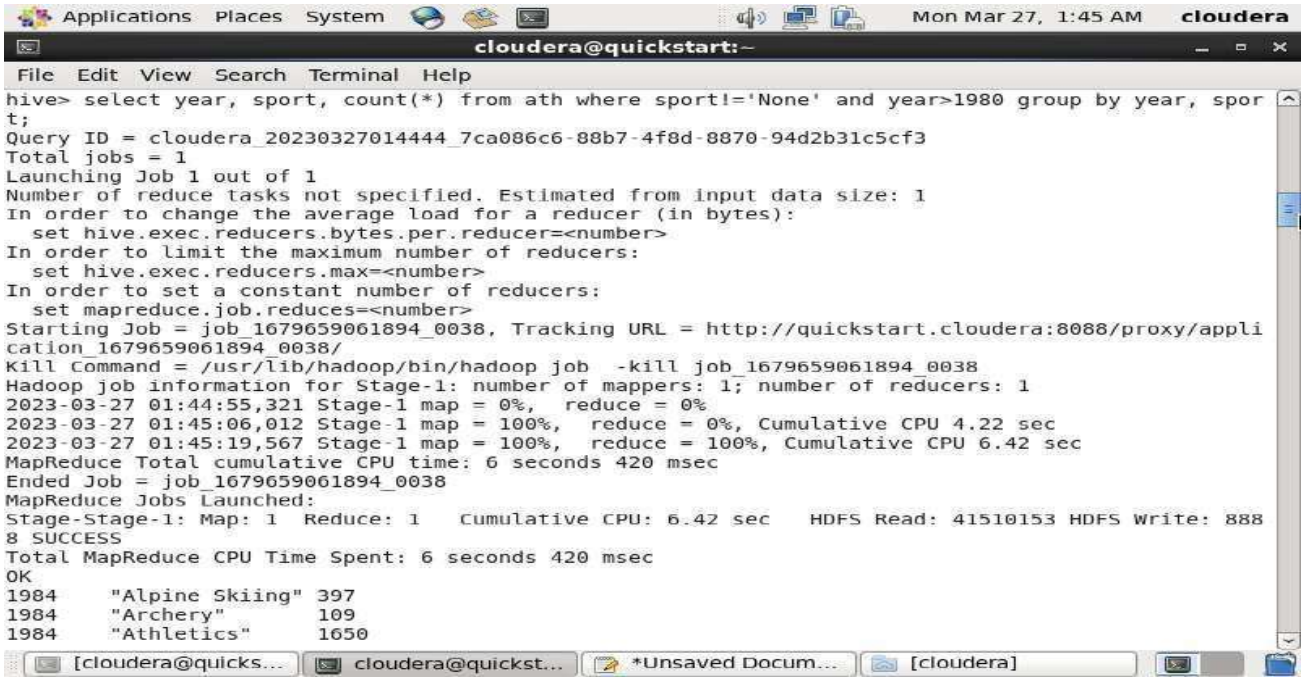


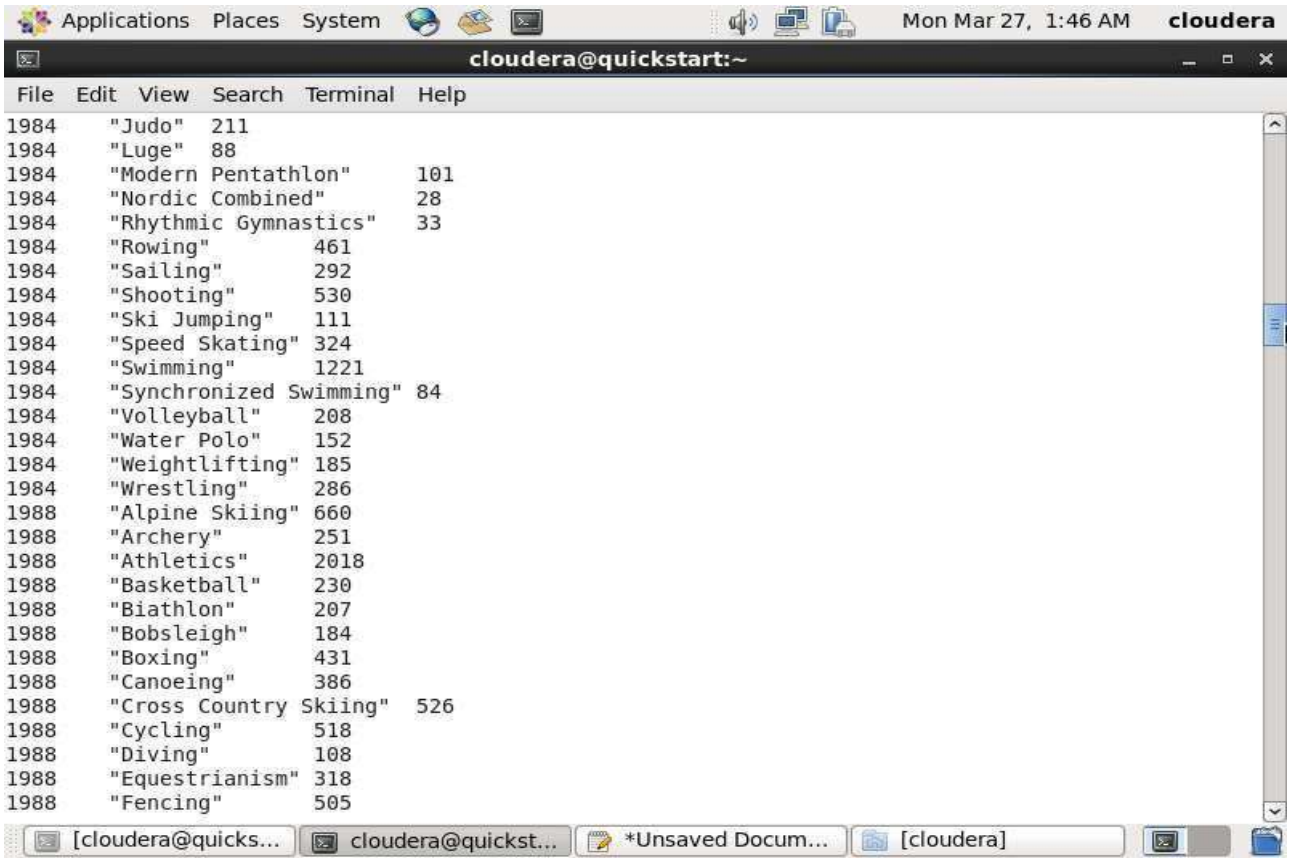
Top 5 sports in which athlete have won medals

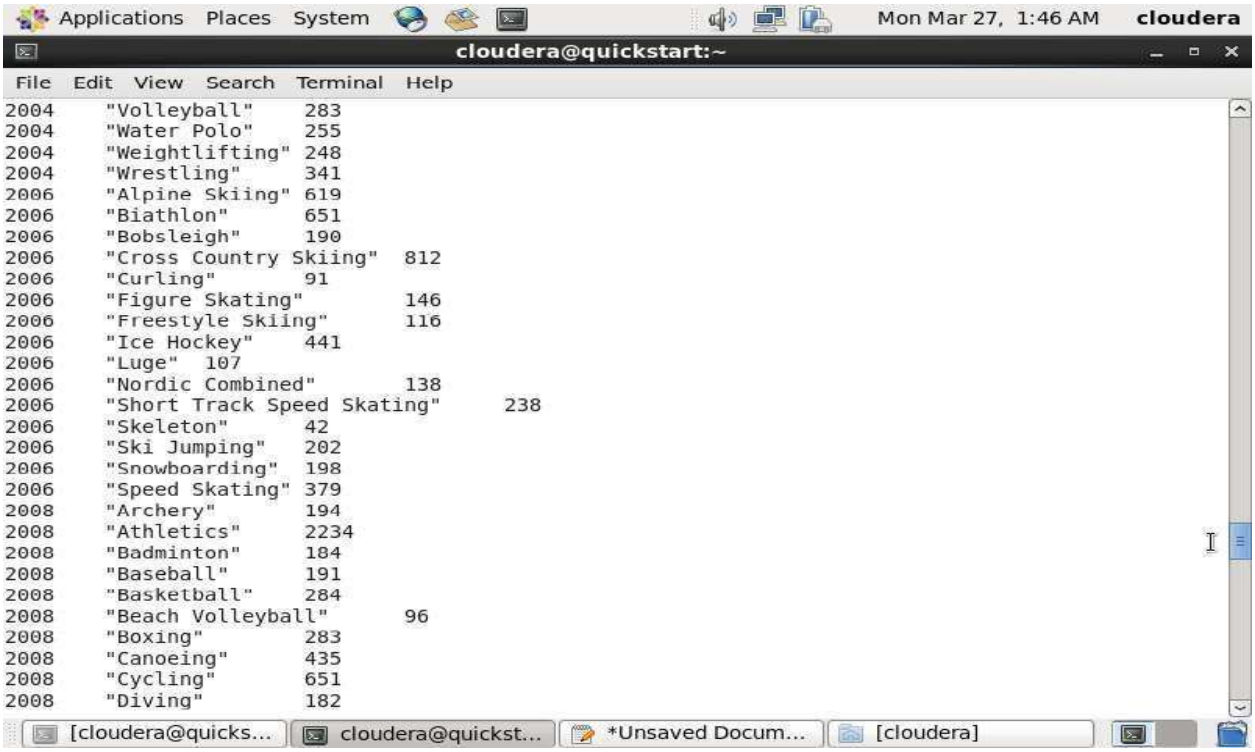


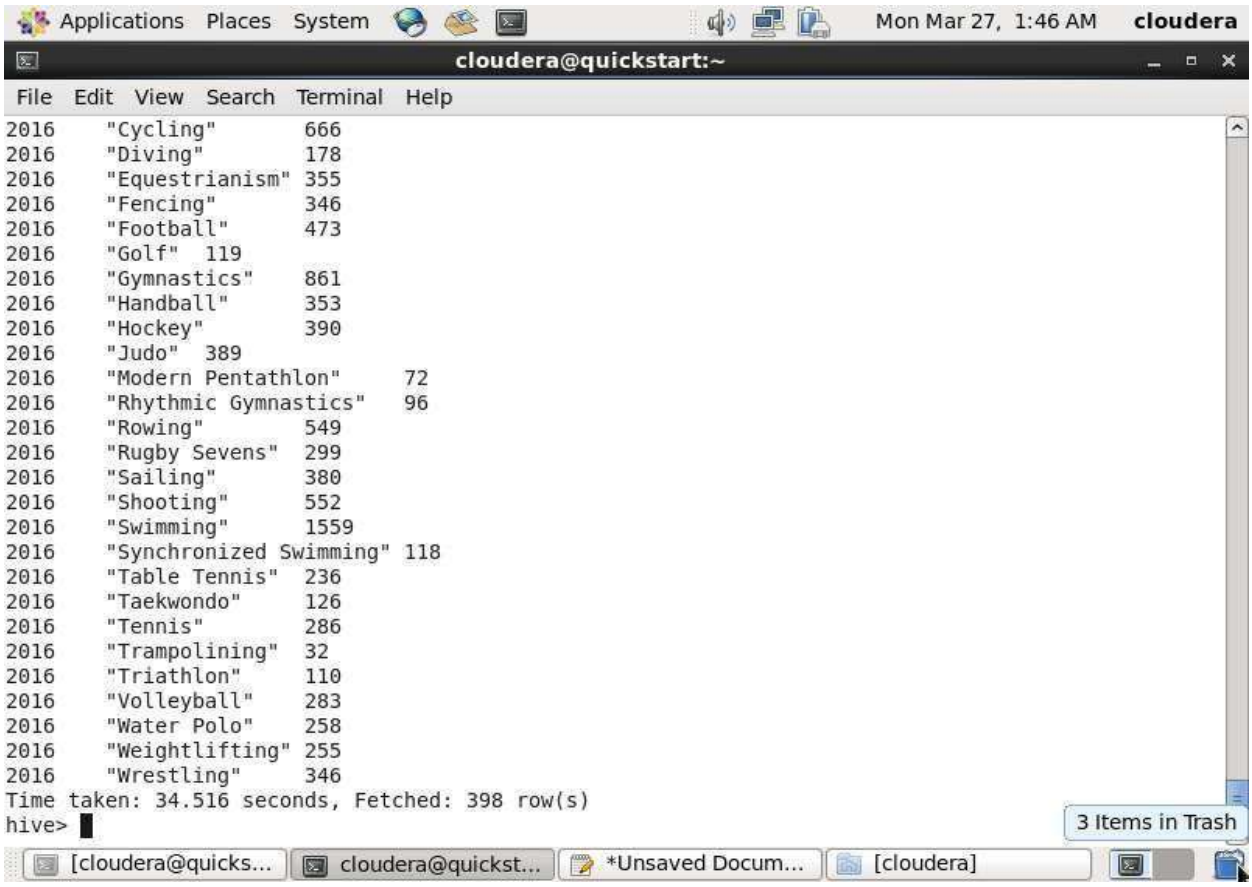


No of events held for each sport in year after 1980

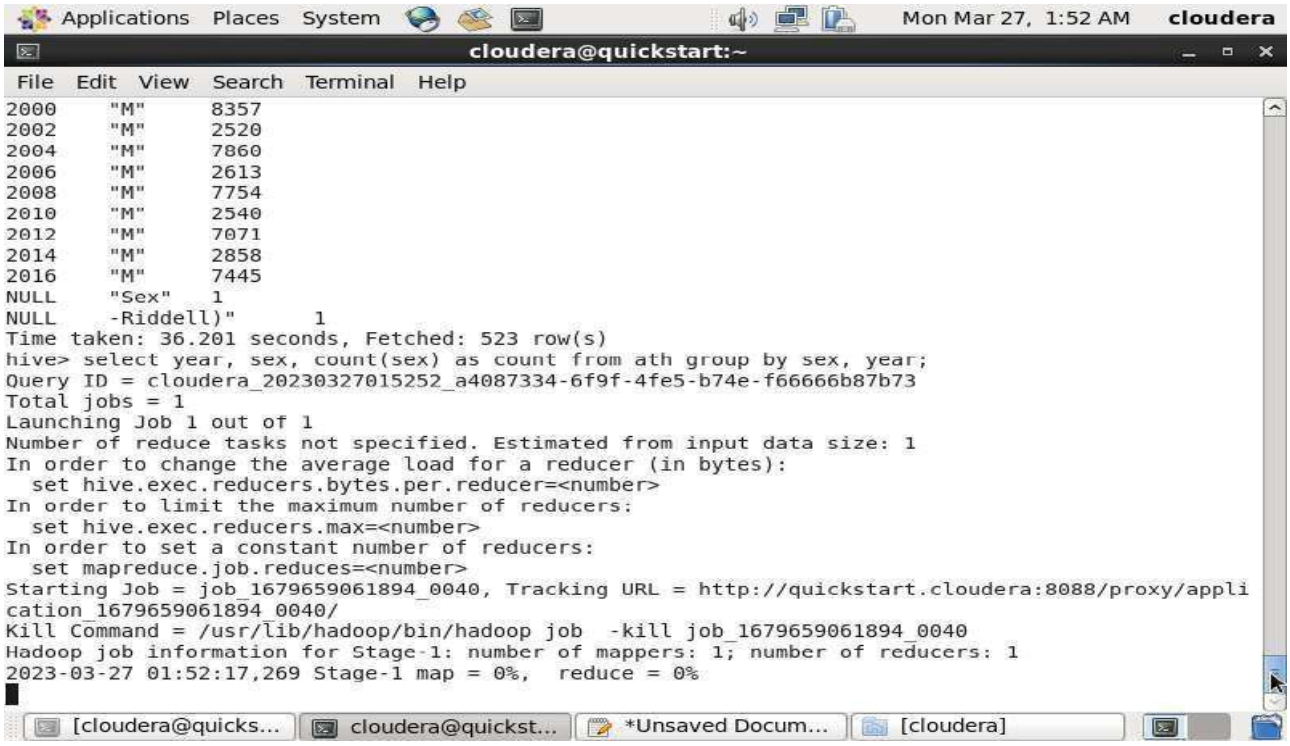


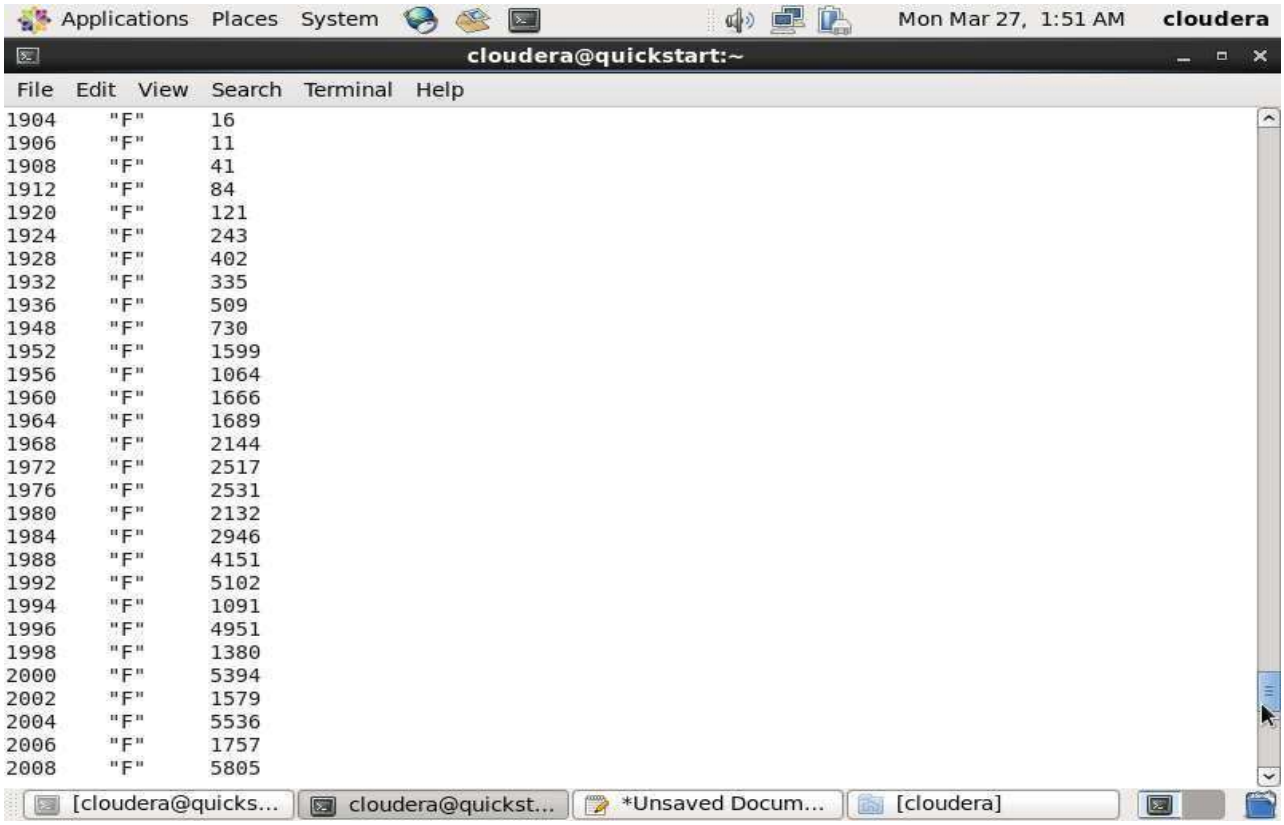


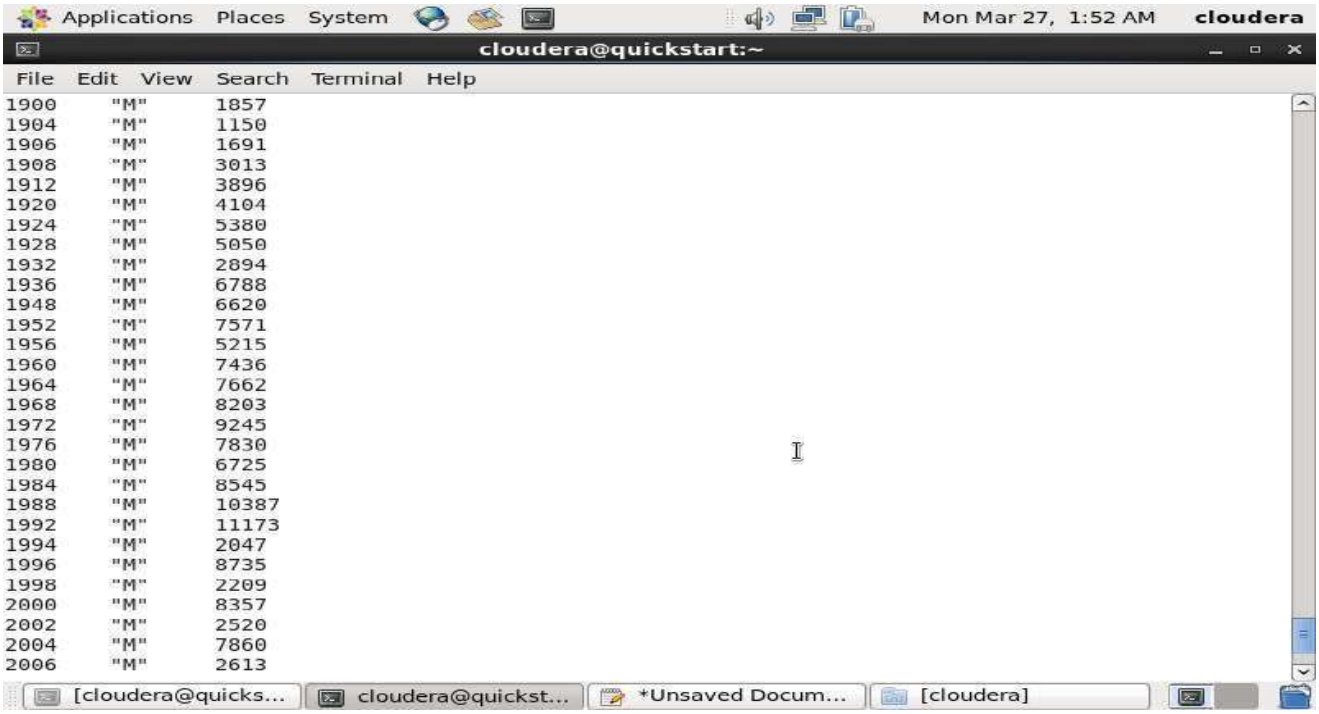

Applications Places System
Mon Mar 27, 1:46 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
2004 "Volleyball" 283
2004 "Water Polo" 255
2004 "Weightlifting" 248
2004 "Wrestling" 341
2006 "Alpine Skiing" 619
2006 "Biathlon" 651
2006 "Bobsleigh" 190
2006 "Cross Country Skiing" 812
2006 "Curling" 91
2006 "Figure Skating" 146
2006 "Freestyle Skiing" 116
2006 "Ice Hockey" 441
2006 "Luge" 107
2006 "Nordic Combined" 138
2006 "Short Track Speed Skating" 238
2006 "Skeleton" 42
2006 "Ski Jumping" 202
2006 "Snowboarding" 198
2006 "Speed Skating" 379
2008 "Archery" 194
2008 "Athletics" 2234
2008 "Badminton" 184
2008 "Baseball" 191
2008 "Basketball" 284
2008 "Beach Volleyball" 96
2008 "Boxing" 283
2008 "Canoeing" 435
2008 "Cycling" 651
2008 "Diving" 182
[cloudera@quicks... cloudera@quickst... *Unsaved Docum... [cloudera]


Applications Places System
Mon Mar 27, 1:46 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
2016 "Cycling" 666
2016 "Diving" 178
2016 "Equestrianism" 355
2016 "Fencing" 346
2016 "Football" 473
2016 "Golf" 119
2016 "Gymnastics" 861
2016 "Handball" 353
2016 "Hockey" 390
2016 "Judo" 389
2016 "Modern Pentathlon" 72
2016 "Rhythmic Gymnastics" 96
2016 "Rowing" 549
2016 "Rugby Sevens" 299
2016 "Sailing" 380
2016 "Shooting" 552
2016 "Swimming" 1559
2016 "Synchronized Swimming" 118
2016 "Table Tennis" 236
2016 "Taekwondo" 126
2016 "Tennis" 286
2016 "Trampolining" 32
2016 "Triathlon" 110
2016 "Volleyball" 283
2016 "Water Polo" 258
2016 "Weightlifting" 255
2016 "Wrestling" 346
Time taken: 34.516 seconds, Fetched: 398 row(s)
hive>
3 Items in Trash
[cloudera@quicks... cloudera@quickst... *Unsaved Docum... [cloudera]

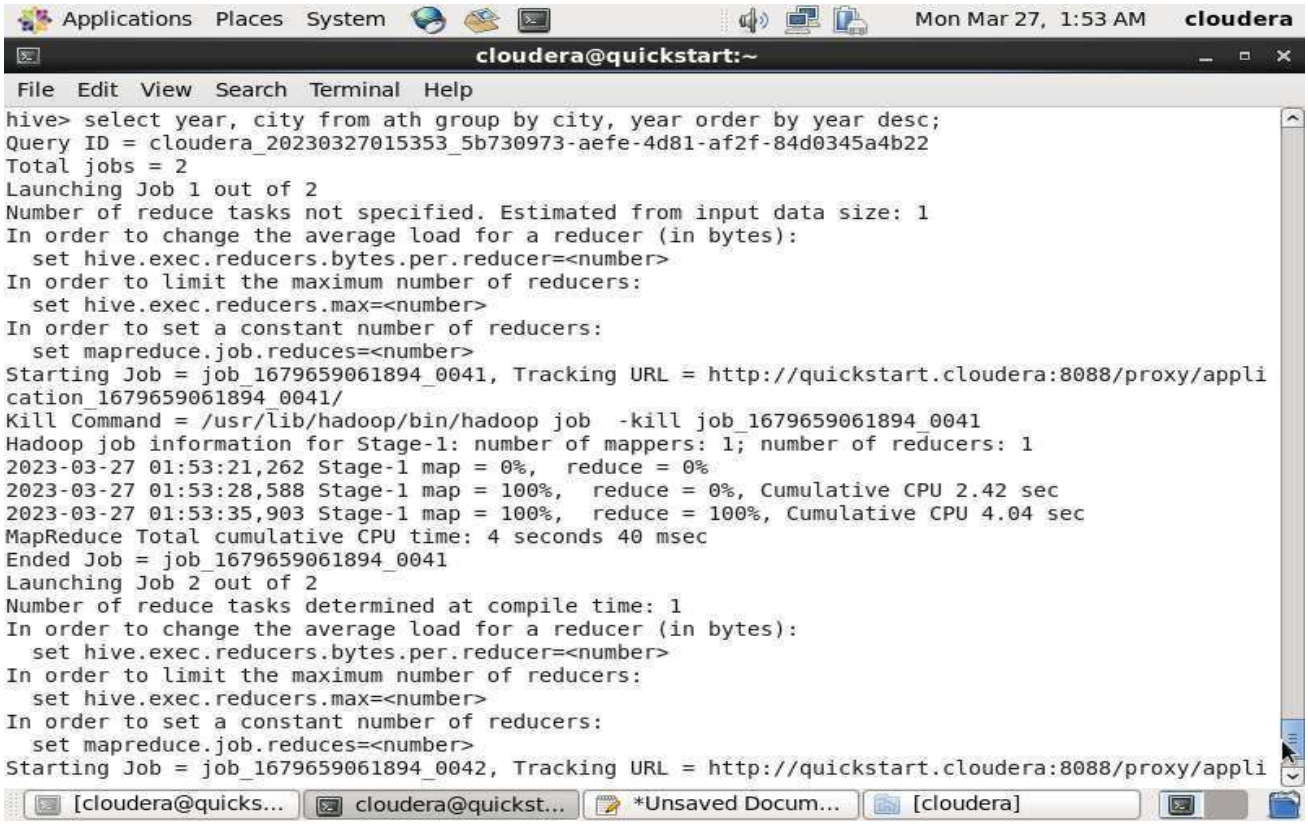
Count of athletes in Olympics in year

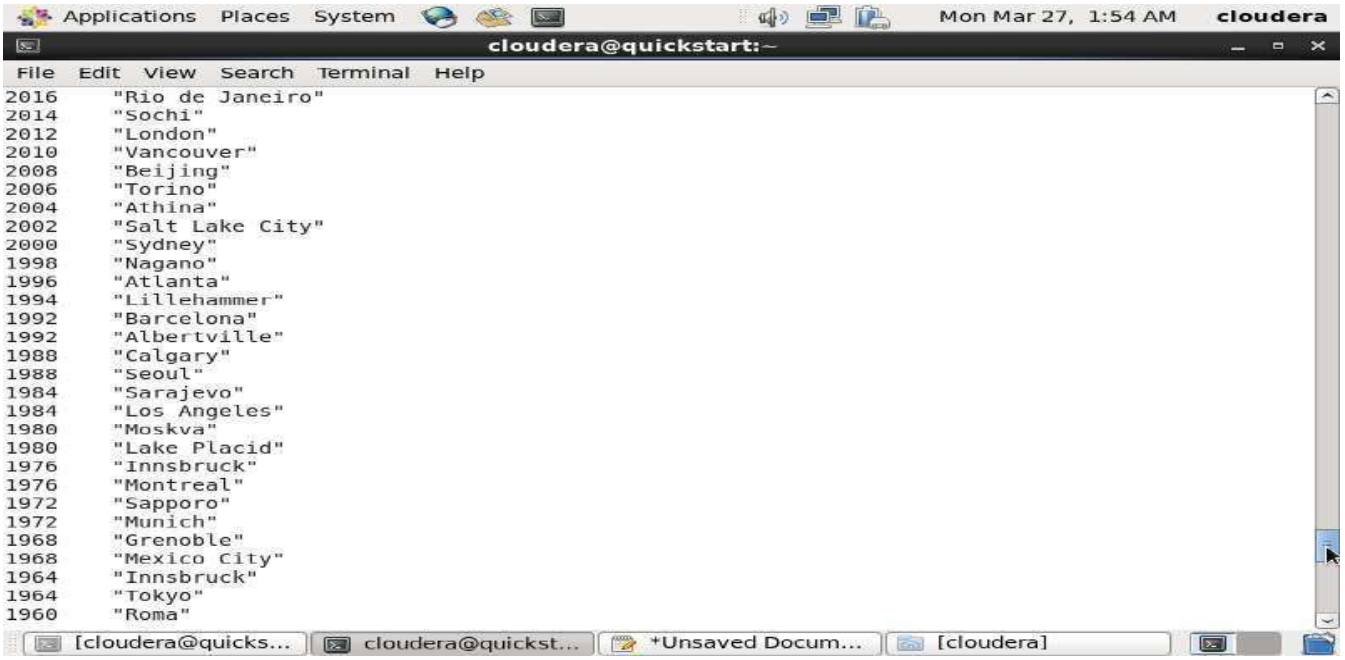




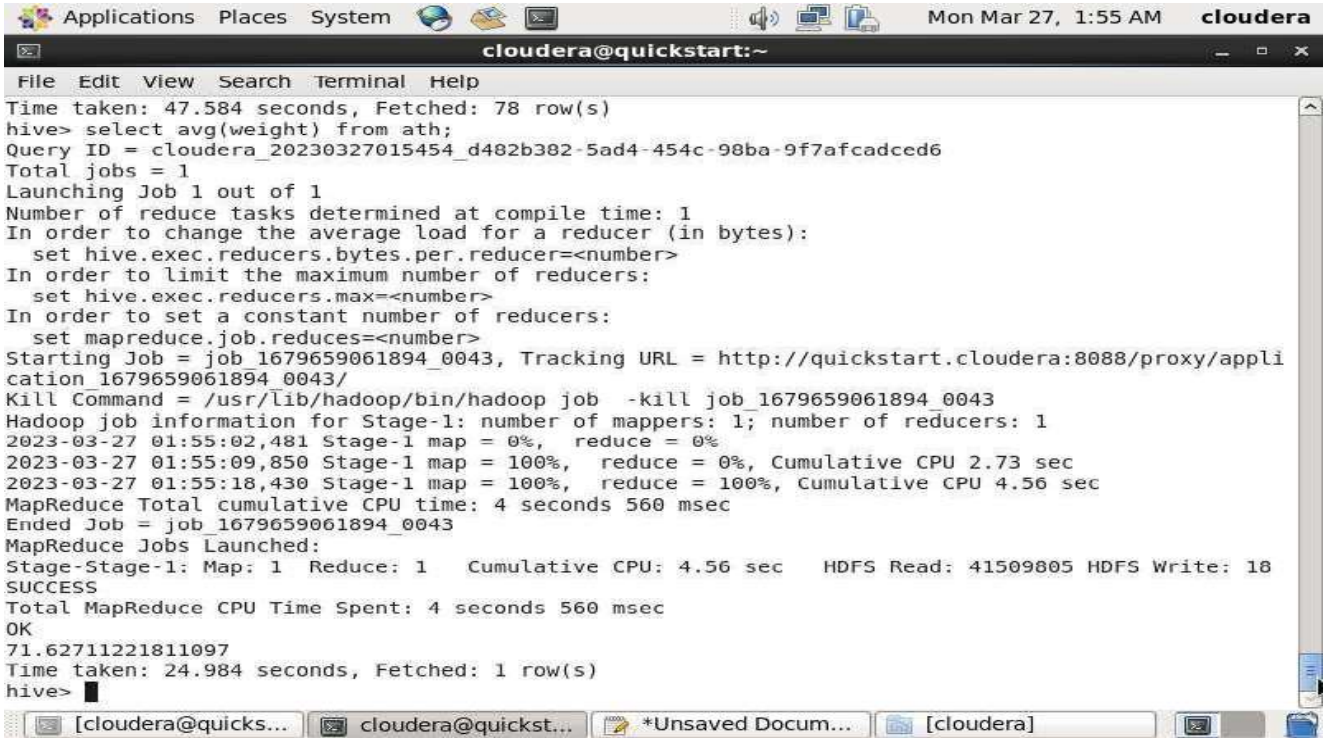


Different cities in which Olympics has held

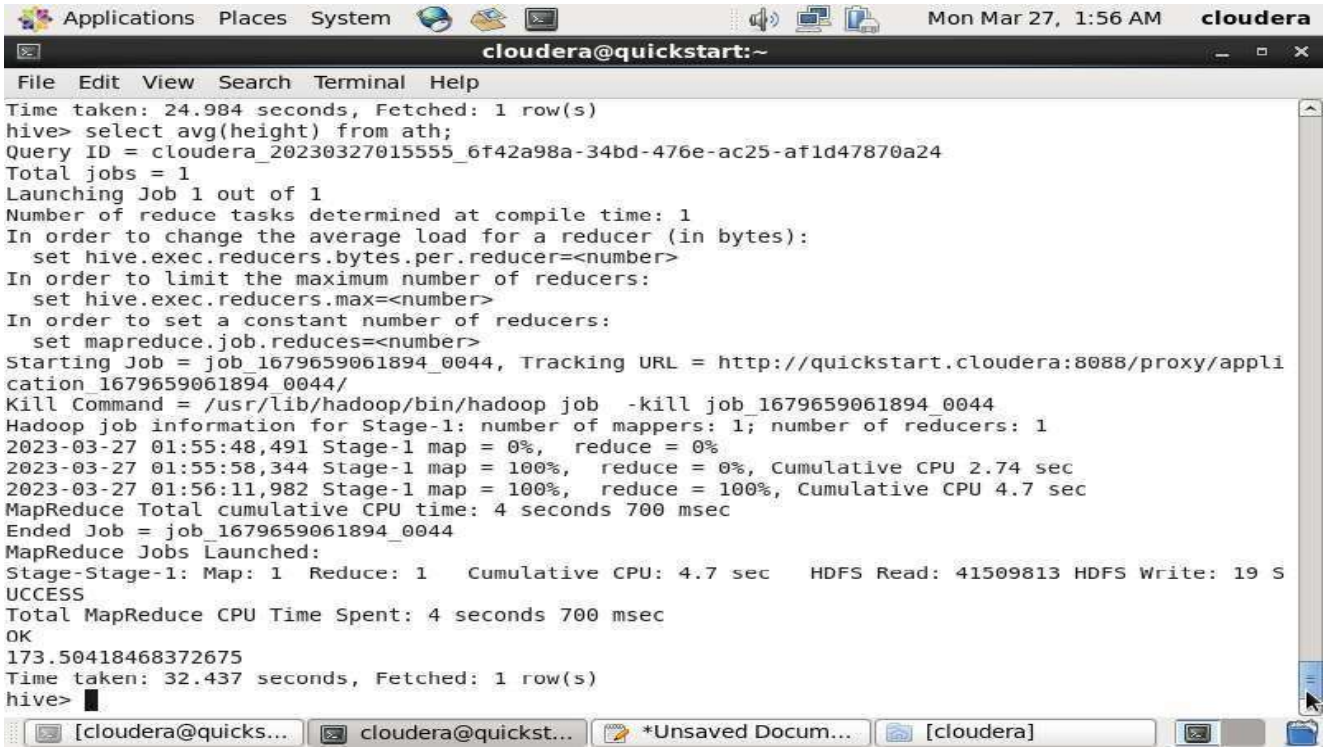


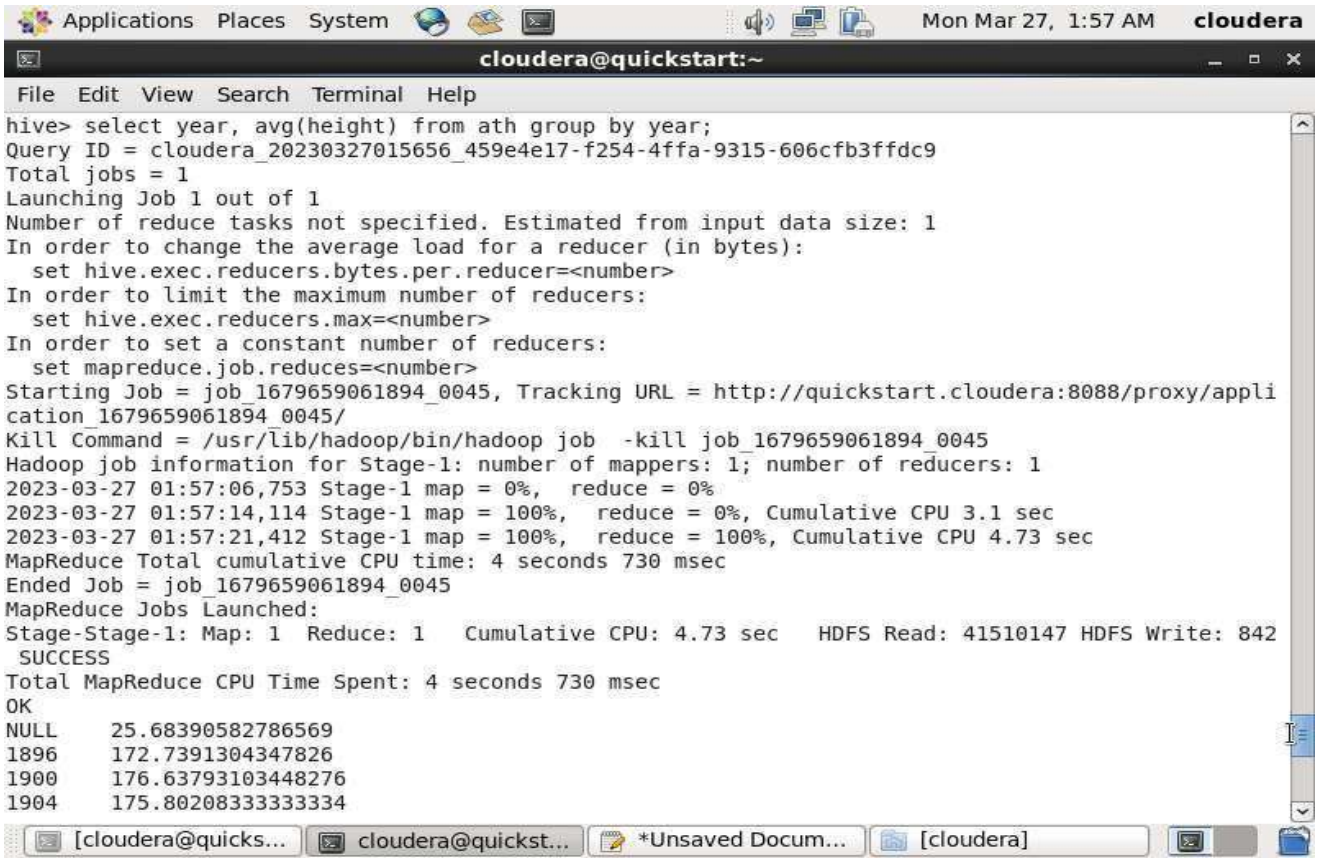


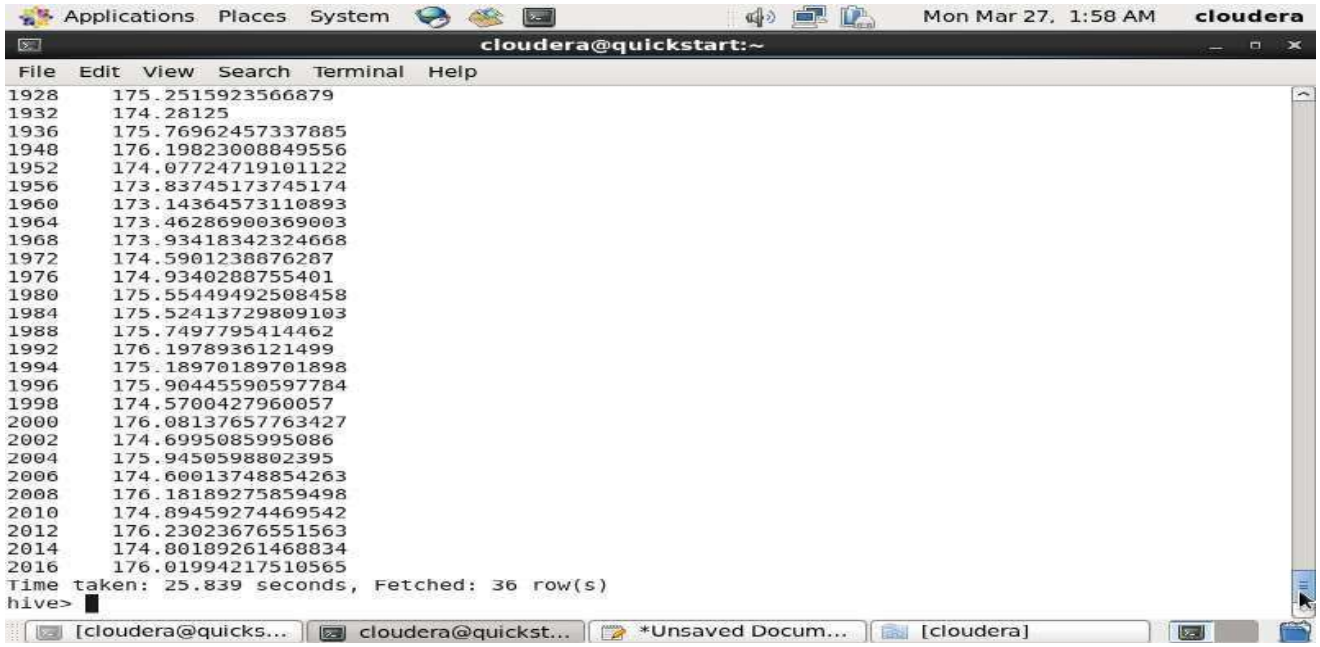
Average weights of all athletes



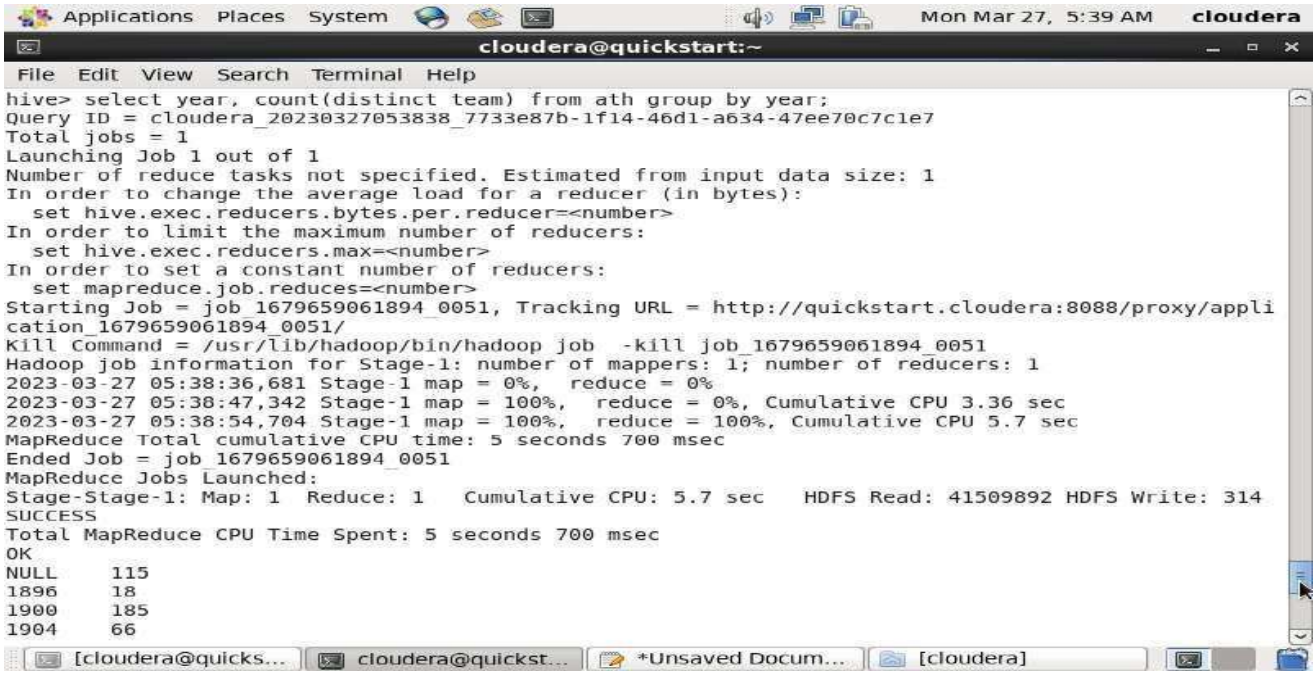
Average heights of all athletes

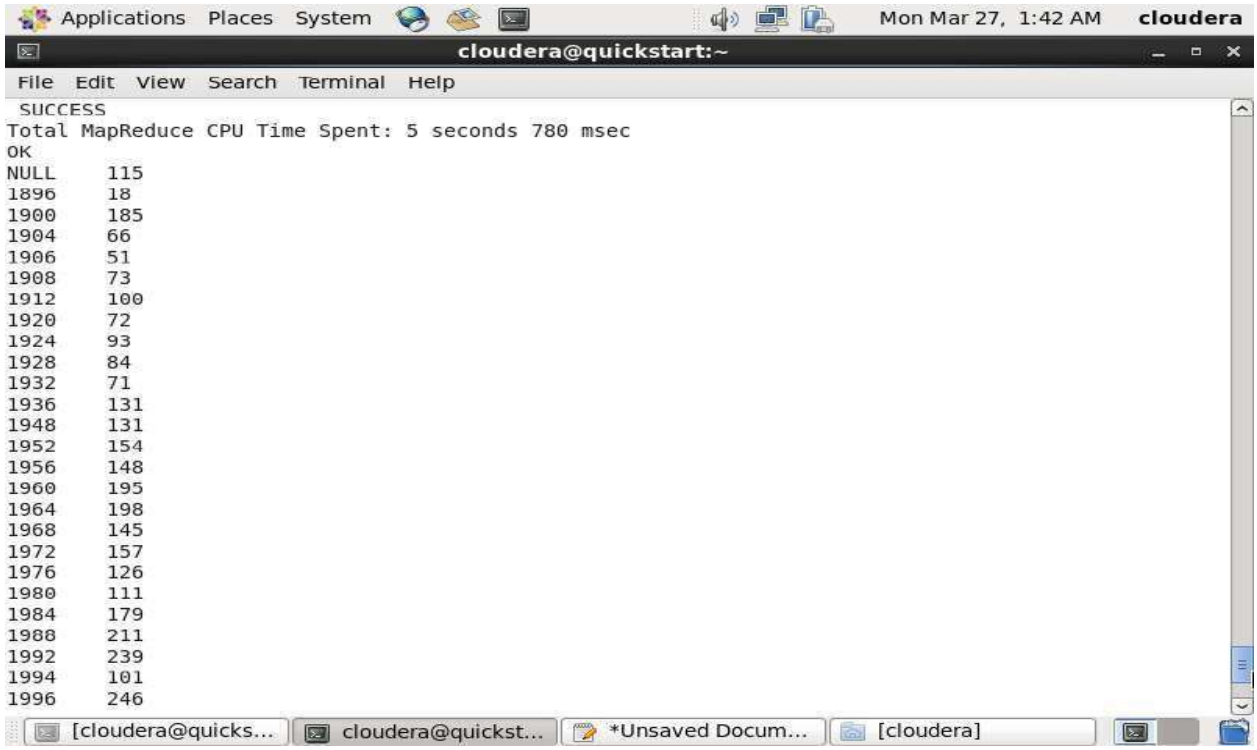


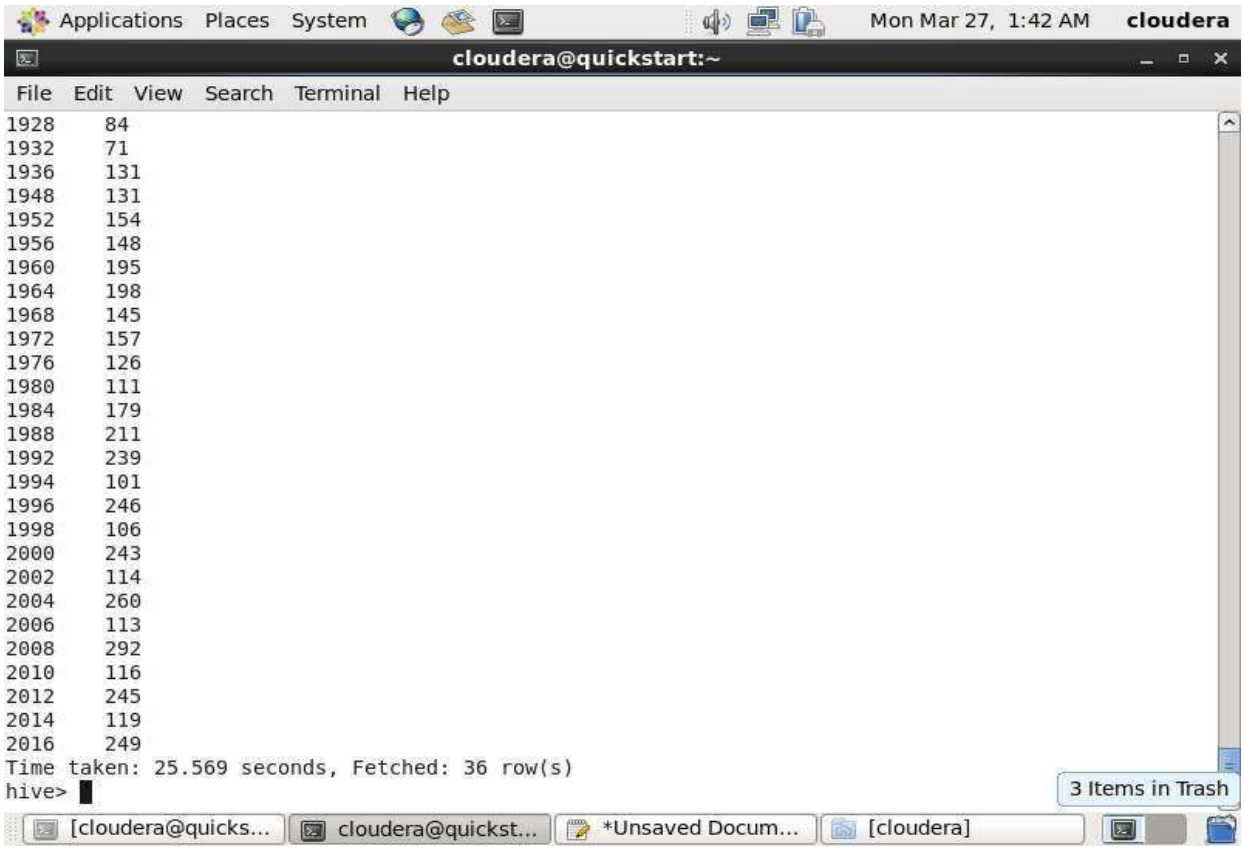


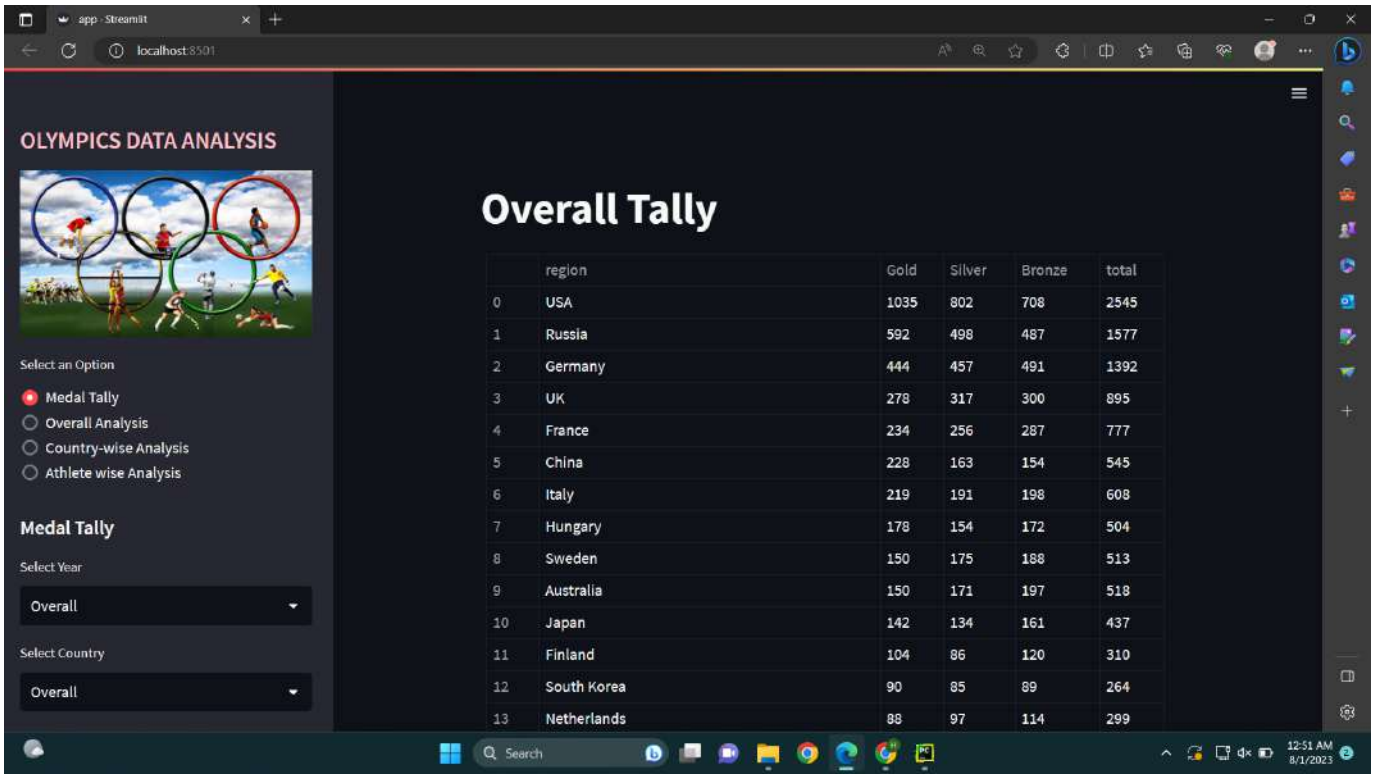


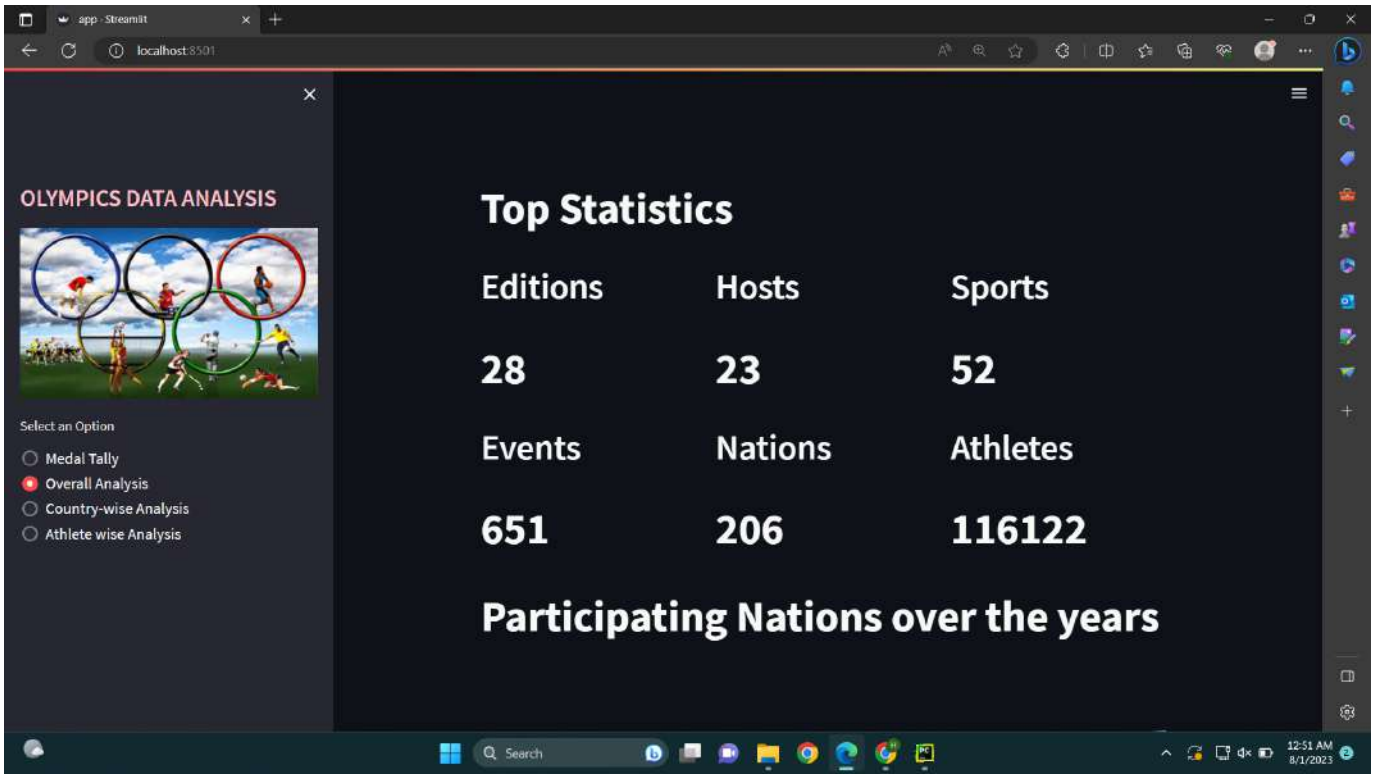
Number of distinct teams participated in every olympics




Applications Places System
Mon Mar 27, 1:42 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 780 msec
OK
NULL 115
1896 18
1900 185
1904 66
1906 51
1908 73
1912 100
1920 72
1924 93
1928 84
1932 71
1936 131
1948 131
1952 154
1956 148
1960 195
1964 198
1968 145
1972 157
1976 126
1980 111
1984 179
1988 211
1992 239
1994 101
1996 246
[cloudera@quicks...
cloudera@quickst...
*Unsaved Docum...
[cloudera]


Applications Places System
Mon Mar 27, 1:42 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
1928 84
1932 71
1936 131
1948 131
1952 154
1956 148
1960 195
1964 198
1968 145
1972 157
1976 126
1980 111
1984 179
1988 211
1992 239
1994 101
1996 246
1998 106
2000 243
2002 114
2004 260
2006 113
2008 292
2010 116
2012 245
2014 119
2016 249
Time taken: 25.569 seconds, Fetched: 36 row(s)
hive>
3 Items in Trash
[cloudera@quicks...
cloudera@quickst...
*Unsaved Docum...
[cloudera]

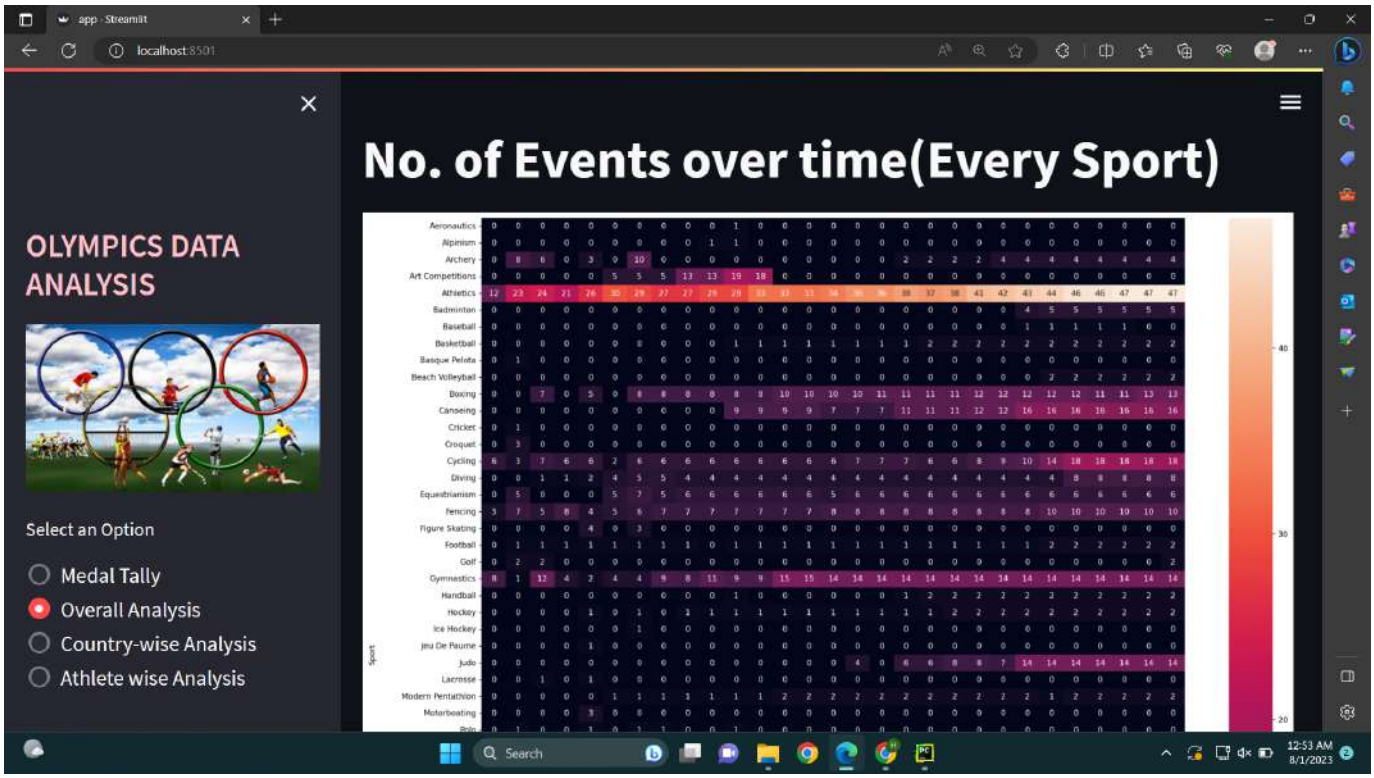

app · Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Medal Tally
Select Year
Overall
Select Country
Overall
Overall Tally
region Gold Silver Bronze total
0 USA 1035 802 708 2545
1 Russia 592 498 487 1577
2 Germany 444 457 491 1392
3 UK 278 317 300 895
4 France 234 256 287 777
5 China 228 163 154 545
6 Italy 219 191 198 608
7 Hungary 178 154 172 504
8 Sweden 150 175 188 513
9 Australia 150 171 197 518
10 Japan 142 134 161 437
11 Finland 104 86 120 310
12 South Korea 90 85 89 264
13 Netherlands 88 97 114 299
Search
12:51 AM
8/1/2023

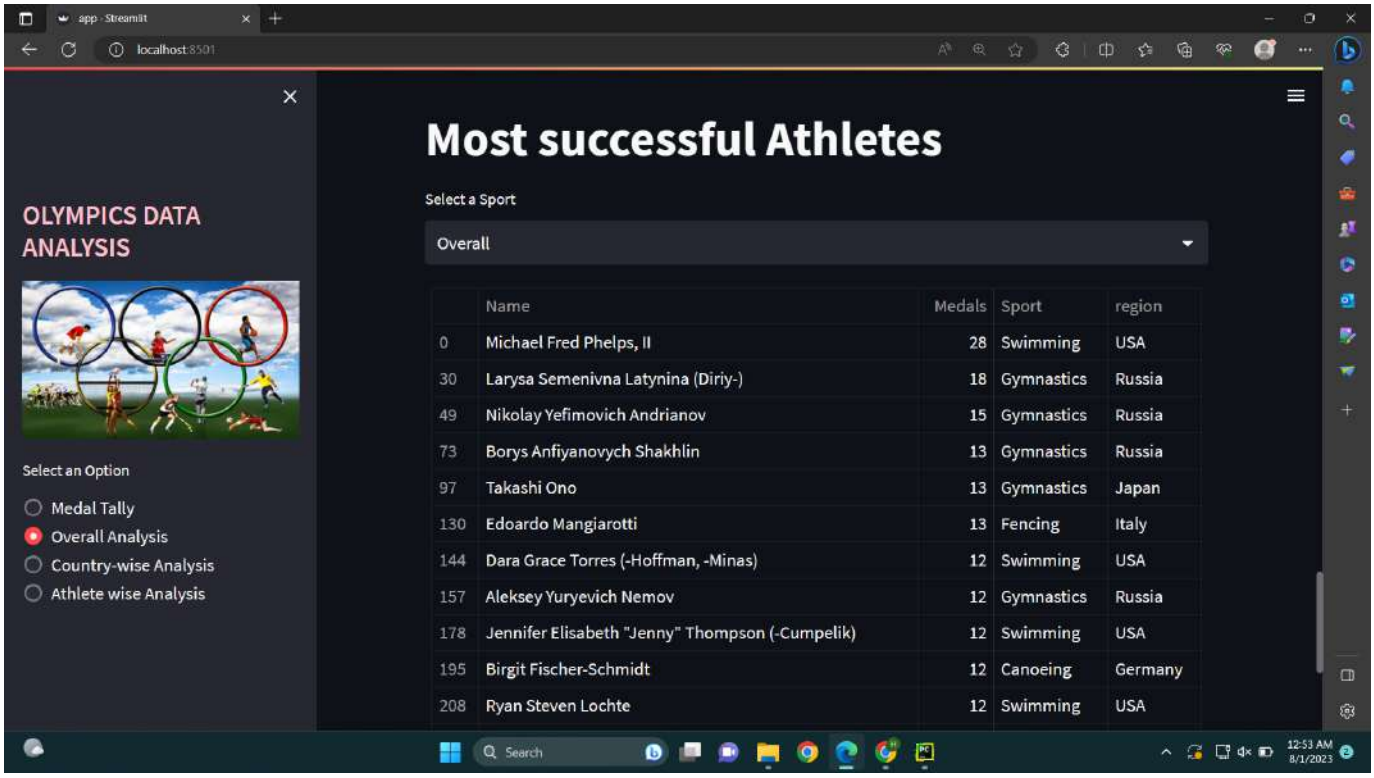

app · Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Top Statistics
Editions
28
Hosts
23
Sports
52
Events
651
Nations
206
Athletes
116122
Participating Nations over the years
Search
12:51 AM
8/1/2023

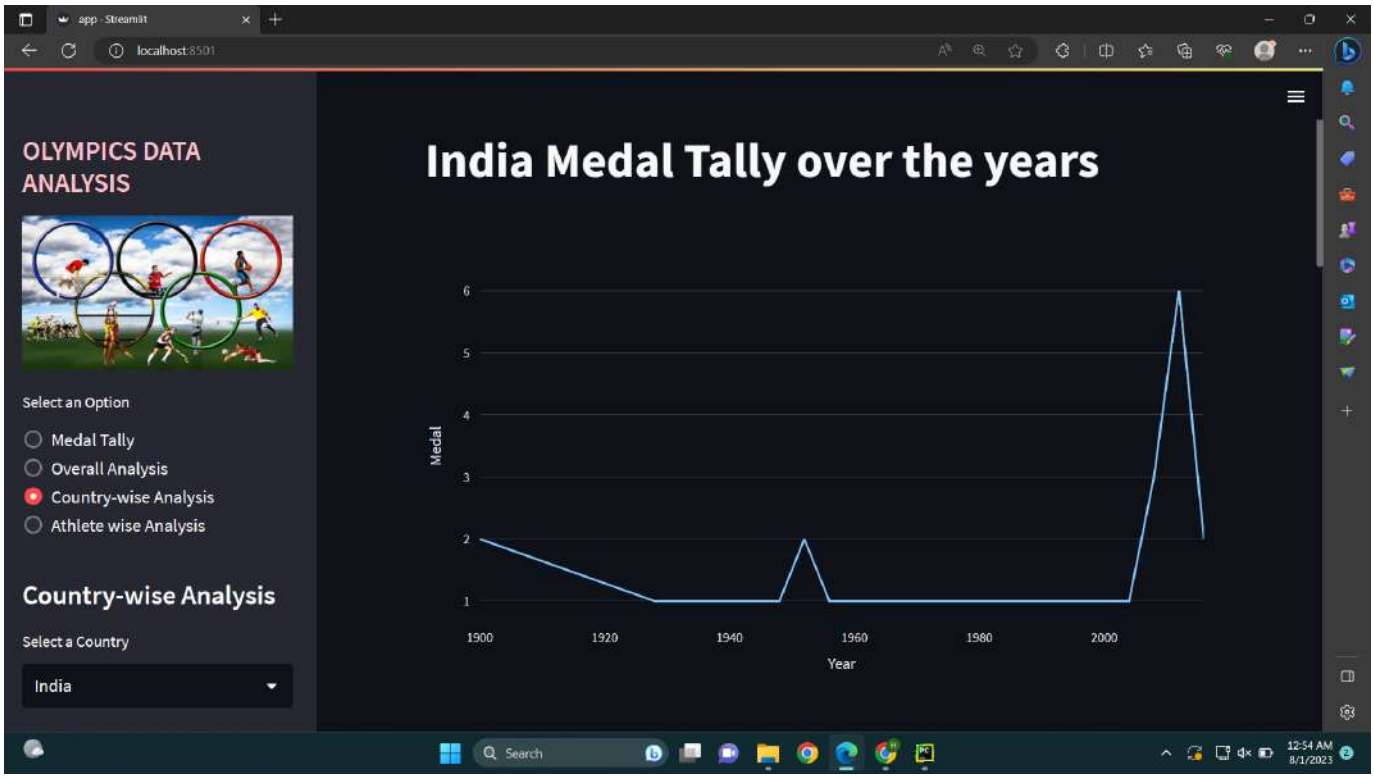

app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Participating Nations over the years
region
200
150
100
50
1900
1920
1940
1960
1980
2000
Edition
Events over the years
Search
12:51 AM
8/1/2023

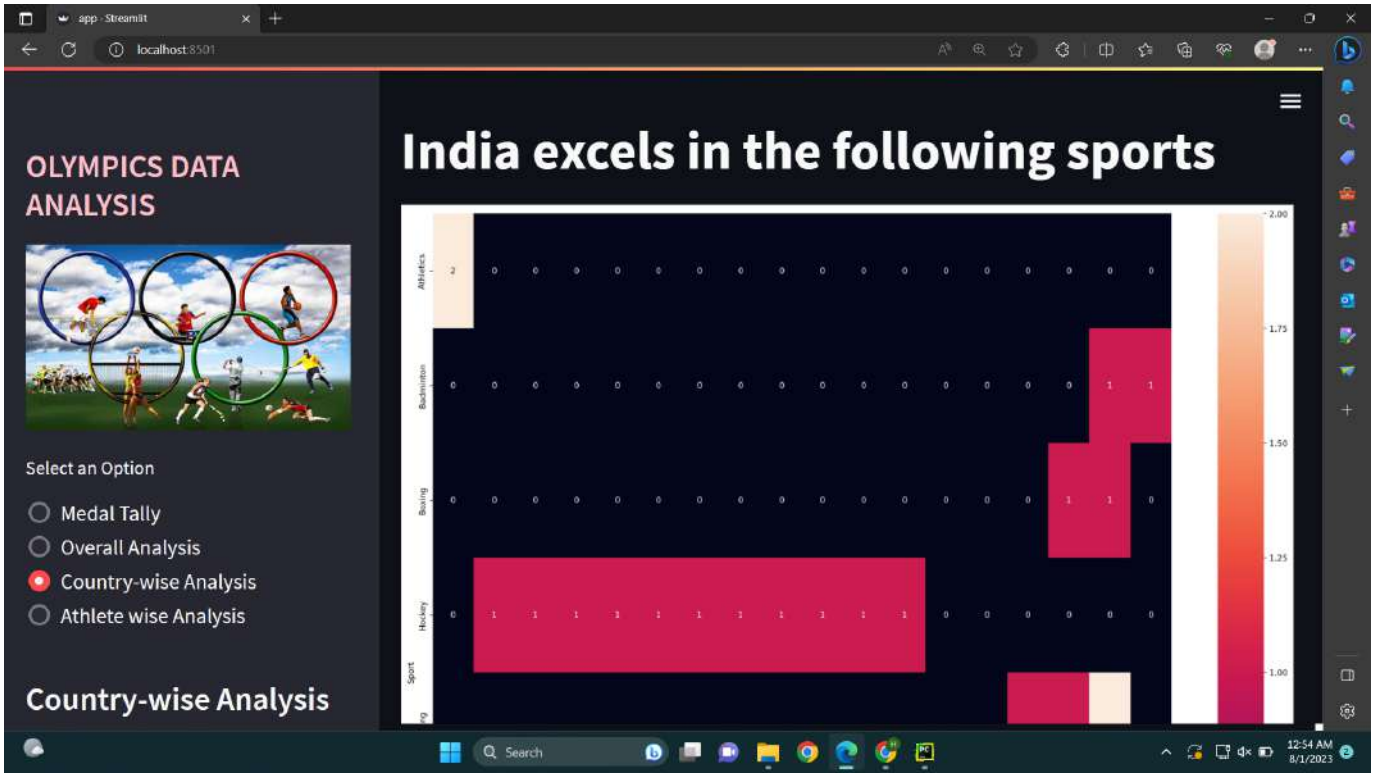

app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Events over the years
Event
300
250
200
150
100
50
1900
1920
1940
1960
1980
2000
Edition
Athletes over the years
Search
12:52 AM
8/1/2023


app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Athletes over the years
Name
Edition
10k
8k
6k
4k
2k
0
1900
1920
1940
1960
1980
2000
Search
12:52 AM
8/1/2023


app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
No. of Events over time(Every Sport)
Sport
Search
12:53 AM
8/1/2023

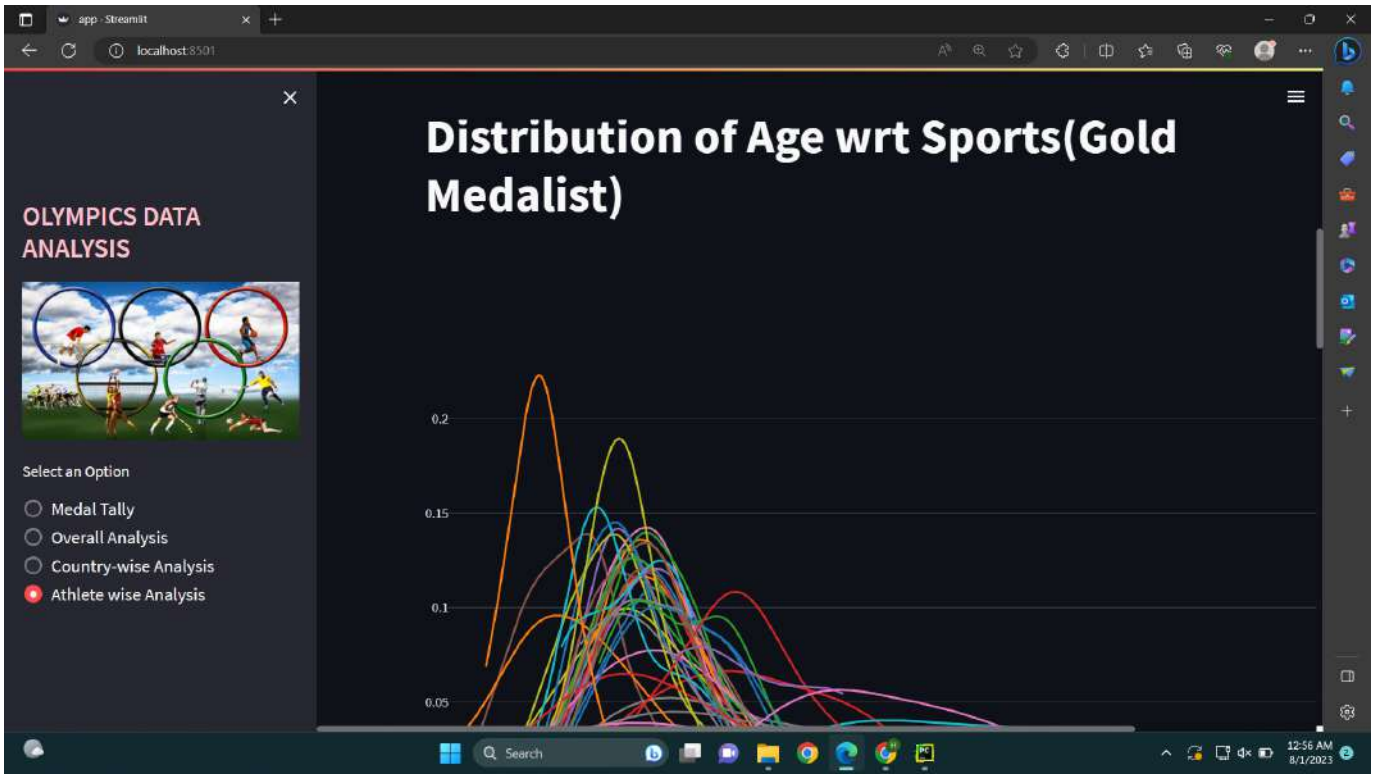

app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Most successful Athletes
Select a Sport
Overall
Name Medals Sport region
0 Michael Fred Phelps, II 28 Swimming USA
30 Larysa Semenivna Latynina (Diriy-) 18 Gymnastics Russia
49 Nikolay Yefimovich Andrianov 15 Gymnastics Russia
73 Borys Anfiyanovych Shakhlin 13 Gymnastics Russia
97 Takashi Ono 13 Gymnastics Japan
130 Edoardo Mangiarotti 13 Fencing Italy
144 Dara Grace Torres (-Hoffman, -Minas) 12 Swimming USA
157 Aleksey Yuryevich Nemov 12 Gymnastics Russia
178 Jennifer Elisabeth "Jenny" Thompson (-Cumpelik) 12 Swimming USA
195 Birgit Fischer-Schmidt 12 Canoeing Germany
208 Ryan Steven Lochte 12 Swimming USA
Search
12:53 AM
8/1/2023


app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
India Medal Tally over the years
Medal
Year
1900
1920
1940
1960
1980
2000
Country-wise Analysis
Select a Country
India
Search
12:54 AM
8/1/2023

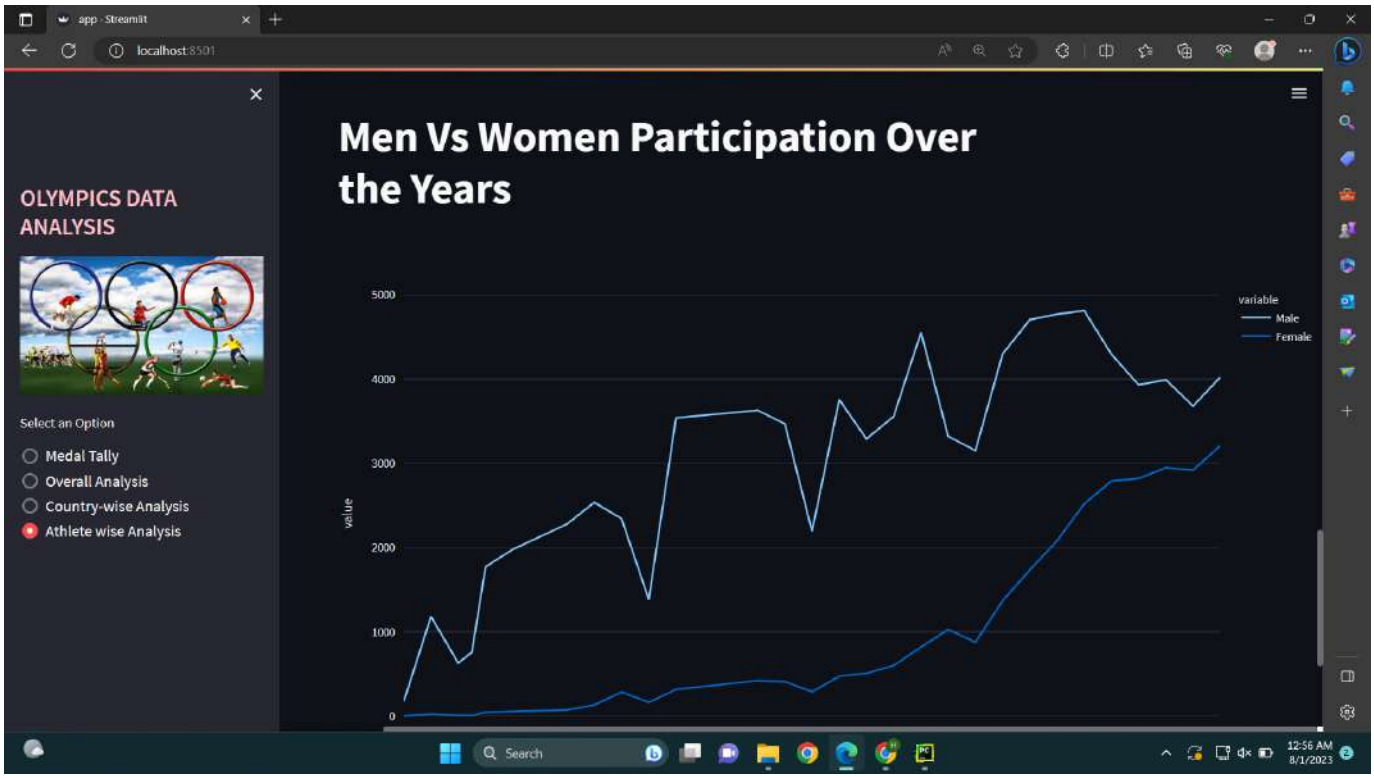


	Name	Medals	Sport
0	Udham Singh Kular	4	Hockey
4	Leslie Walter Claudius	4	Hockey
8	Shankar Pillay Laxman	3	Hockey
11	Ranganathan Francis	3	Hockey
14	Richard James Allen	3	Hockey
17	Balbir Singh	3	Hockey
20	Harbinder Singh Chimni	3	Hockey
23	Balbir Singh Dosanjh, Sr.	3	Hockey
26	Victor John "V. J." Peter	3	Hockey
29	Dhyan Chand Bais	3	Hockey


app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Distribution of Age
0.1
0.08
0.06
0.04
0.02
Search
12:55 AM
8/1/2023


app - Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Distribution of Age wrt Sports(Gold Medalist)
0.2
0.15
0.1
0.05
Search
12:56 AM
8/1/2023


app · Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Height Vs Weight
Select a Sport
Overall
Height
220
200
180
160
Medal
No Medal
Gold
Bronze
Silver
Sex
Search
12:56 AM
8/1/2023


app · Streamlit
localhost:8501
OLYMPICS DATA ANALYSIS
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Men Vs Women Participation Over the Years
value
5000
4000
3000
2000
1000
0
variable
Male
Female
Search
12:56 AM
8/1/2023