

## Women chess players Program using PySpark

1. Find the total count of Indian women chess players.
2. Find the names of the chess players belonging to India.
3. Find the names of the Indian chess players who are born in 1995.

**File: topwomen\_chessplayers.csv**

### Program using Pyspark Shell

```
>>> hdfsfile = sc.textFile("hdfs://localhost:9000/user/topwomen_chessplayers.csv")
>>> data = hdfsfile.map(lambda line: line.split(", "))
>>> data.collect()
[['Fide', 'id', 'Name', 'Federation', 'Gender', 'Year_of_birth', 'Title', 'Standard_Rating', 'Rapid_rating',
'Blitz_rating', 'Inactive_flag', "", ""], ['700070', "'Polgar', 'Judit'", 'HUN', 'F', '1976', 'GM', '2675',
'2646', '2736', 'wi', "", "", ""], ['8602980', "'Hou', 'Yifan'", 'CHN', 'F', '1994', 'GM', '2658', '2621', '2601',
'', '', '', ""], ['5008123', "'Koneru', 'Humpy'", 'IND', 'F', '1987', 'GM', '2586', '2483', '2483', "", "", "", ""],
['4147103', "'Goryachkina', 'Aleksandra'", 'RUS', 'F', '1998', 'GM', '2582', '2502', '2441', "", "", "", ""],
['700088', "'Polgar', 'Susan'", 'HUN', 'F', '1969', 'GM', '2577', "", "", 'wi', "", "", ""], ['8600147', "'Xie', '
Jun'", 'CHN', 'F', '1970', 'GM', '2574', "", "", 'wi', "", "", "]]

>>> indiplayers = data.filter(lambda value: 'IND' in value)
>>> indiplayers.collect()
[['5008123', "'Koneru', 'Humpy'", 'IND', 'F', '1987', 'GM', '2586', '2483', '2483', "", "", "", ""], ['5015197',
"'Harika', 'Dronavalli'", 'IND', 'F', '1991', 'GM', '2515', '2450', '2422', "", "", "", ""], ['5091756', 'Vaishali',
'R', 'IND', 'F', '2001', 'WGM', '2393', '2149', '2313', "", "", "", ""], ['5007844', "'Tania', 'Sachdev'", 'IND',
'F', '1986', 'IM', '2392', '2339', '2280', "", "", "", ""], ['5019516', 'Kulkarni', 'Bhakti', 'IND', 'F', '1992', 'IM',
'2391', '2325', '2224', "", "", "", ""], ['5029295', "'Padmini', 'Rout'", 'IND', 'F', '1994', 'IM', '2370', '2355',
'2196', "", "", "", ""], ['5050847', "'Nandhidhaa', 'P V'", 'IND', 'F', '1996', 'WGM', '2365', '2131', '1989', "",
'', '', ""], ['5013623', "'Gomes', 'Mary Ann'", 'IND', 'F', '1989', 'WGM', '2354', '2305', '2106', "", "", "", ""],
['5016193', "'Soumya', 'Swaminathan'", 'IND', 'F', '1989', 'IM', '2351', '2249', '2115', "", "", "", "]]

>>> indiplayers_names = indiplayers.map(lambda value: [value[i] for i in [1,2]])
>>> indiplayers_names.collect()
[[["Koneru", 'Humpy'], ["Harika", 'Dronavalli'], ['Vaishali', 'R'], ["Tania", 'Sachdev'], ['Kulkarni',
'Bhakti'], ["Padmini", 'Rout'], ["Nandhidhaa", 'P V'], ["Gomes", 'Mary Ann'], ["Soumya", '
Swaminathan"]]]

>>> indiplayers_born = indiplayers.filter(lambda value: '1995' in value)
>>> indiplayers_born.collect()
[['5055903', "'Srija', 'Seshadri'", 'IND', 'F', '1997', 'WGM', '2219', '2087', '1995', "", "", "", ""],
['5031931', "'Saranya', 'J'", 'IND', 'F', '1995', 'WFM', '2045', '2052', '2115', "", "", "", ""], ['5029422',
"'Veena', 'Mounica M.'", 'IND', 'F', '1995', "", '2044', "", "", 'wi', "", "", ""], ['5029040', "'Priyanka', '
Kumari'", 'IND', 'F', '1995', "", '1910', '1958', '1944', "", "", "", ""], ['5088313', 'Hinduja', 'Reddy', 'IND',
'F', '1995', "", '1881', "", '1962', 'wi', "", ""], ['5039673', 'Vaishnavi', 'J', 'IND', 'F', '1995', "", '1849', "",
'', 'wi', "", ""], ['5048958', "'Arunima', 'Kalra'", 'IND', 'F', '1995', "", '1846', "", '', 'wi', "", "", ""], ['5067421',
"'Vanessa', 'D'souza'", 'IND', 'F', '1995', "", '1833', "", '', 'wi', "", "", ""], ['5025532', "'Tejaswini', 'Reddy
S.'", 'IND', 'F', '1995', "", '1832', "", '', 'wi', "", "", ""], ['5046424', "'Anjana Krishna', 'S'", 'IND', 'F', '1995',
'', '1827', '1947', '1895', 'wi', "", "", ""], ['5006570', 'Priya', 'V', 'S', 'S', 'IND', 'F', '1995', "", '1803', "",
'', 'wi']]

>>> indiplayers_bornnames = indiplayers_born.map(lambda value: [value[i] for i in [1,2]])
>>> indiplayers_bornnames.collect()
```

```
[["Srija", ' Seshadri'], ["Saranya", ' J'], ["Veena", ' Mounica M.'], ["Priyanka", ' Kumari'],  
['Hinduja', 'Reddy'], ['Vaishnavi', 'J'], ["Arunima", ' Kalra'], ["Vanessa", ' D`souza'], ["Tejaswini", '  
Reddy S.'], ["Anjana Krishna", ' S'], ['Priya', 'V']]
```

```
>>> indiplayers.count()  
284
```

### **Program using Python (chessplayer\_spark.py)**

```
from pyspark import SparkContext
```

```
def ChessPlayer_Analysis():  
    sc = SparkContext(appName='WomanChessPlayerAnalysis')  
    final_result = sc.emptyRDD()  
    result_list = []  
    ind_pl_count = 0  
    hdfsfile = sc.textFile("hdfs://localhost:9000/user/topwomen_chessplayers.csv")  
    data = hdfsfile.map(lambda line: line.split(","))  
  
    #display names of indian chess players  
    indiplayers = data.filter(lambda value: 'IND' in value)  
    indiplayers_names = indiplayers.map(lambda value: [value[i] for i in [1,2]])  
    indiplayers_names.saveAsTextFile("hdfs://localhost:9000/Chess_Analysis1")  
  
    #display indian chess players born in 1995  
    indiplayers_born = indiplayers.filter(lambda value: '1995' in value)  
    indiplayers_bornnames = indiplayers_born.map(lambda value: [value[i] for i in [1,2]])  
    indiplayers_bornnames.saveAsTextFile("hdfs://localhost:9000/Chess_Analysis2")  
  
    #count total indian players  
    ind_pl_count = indiplayers.count()  
    result_list.append(ind_pl_count)  
    final_result = sc.parallelize(result_list)  
    final_result.saveAsTextFile("hdfs://localhost:9000/Chess_Analysis3")  
    sc.stop()  
  
if __name__ == '__main__':  
    ChessPlayer_Analysis()
```

### **Executing file in hadoop**

```
faculty@glrica:~/hadoop-3.1.3$ spark-submit --master local chessplayer_spark.py
```

### **Viewing output on hdfs**

```
faculty@glrica:~/hadoop-3.1.3$ hdfs dfs -cat /Chess_Analysis1/part-00000  
2020-09-01 16:17:58,533 INFO sasl.SaslDataTransferClient: SASL encryption trust check:  
localhostTrusted = false, remoteHostTrusted = false  
["Koneru", ' Humpy']  
["Harika", ' Dronavalli']  
['Vaishali', 'R']  
["Tania", ' Sachdev']  
['Kulkarni', 'Bhakti']  
["Padmini", ' Rout']
```

['Nandhidhaa', ' P V']  
['Gomes', ' Mary Ann']  
['Soumya', ' Swaminathan']  
['Karavade', ' Eesha']  
['Vijayalakshmi', ' Subbaraman']  
['Pratyusha', ' Bodda']  
['Divya', 'Deshmukh']  
['Vantika', 'Agrawal']  
['Sharma', 'Isha']  
['Priyanka', 'Nutakki']  
['Savitha', 'Shri']  
['Mrudul', 'Dehankar']  
['Mahalakshmi', ' M']  
['Pujari', ' Rucha']  
['Parnali', ' S Dharia']  
['Vinuthna', ' N.']  
['Aakanksha', 'Hagawane']  
['Cholleti', 'Sahajasri']  
['Varshini', ' V']  
['Srija', ' Seshadri']  
['Priyanka', ' K']  
['Saritha', ' M. Reddy']  
['Koneru', ' Chandra Hawsa']  
['Khadilkar', ' Rohini']  
['Arpita', ' Mukherjee']  
['Salonika', ' Saina']  
['Monnisha', 'G']  
['Prathiba', ' Yuvarajan']  
['Mohota', ' Nisha']  
['Rakshitta', 'Ravi']  
['Harshini', 'A']  
['Priya', 'P.']  
['Harshita', 'Guddanti']  
['Michelle Catherina', ' P']  
['Chandika', 'Divyasree']  
['Meenakshi', 'Subbaraman']  
['Shah', ' Pallavi G.']  
['Kiran', ' Manisha Mohanty']  
['Chitlange', 'Sakshi']  
['Tejaswini', ' Sagar']  
['Tarini', 'Goyal']  
['Raghavi', ' N.']  
['Vennela', ' T.']  
['Sapale', 'Saloni']  
['Swati', 'Ghate']  
['Nadig', ' Kruttika']  
['Ivana Maria', ' Furtado']  
['Kavitha', 'J']  
['Sowmya', ' H.']  
['Keerthi', ' B.S.']  
['Toshali', ' V']  
['Kunte', ' Mrunalini']

["'Khadilkar', ' Jayshree'"]  
["'Singh', ' Sushma'"]  
["'Unni', ' Vasanti'"]  
["'Samani', ' Shraddha'"]  
["'Nimmy', ' A.G.'"]  
["'Safira', ' Shahnaz'"]  
["'Saheli', ' Nath'"]  
["'Konakanchi', ' Ninad'"]  
["'Savant', ' Riya'"]  
['Bala', 'Kannamma']  
['Jyothsna', 'L']  
["'Sreeparna', ' Chakraborty'"]  
["'Lakshmi', ' Sahiti P.'"]  
["'Kasturi', ' M'"]  
["'Sivasankari', ' P.'"]  
['Jishitha', 'D']  
["'Devangi', ' Patankar'"]  
["'Bharathi', ' R.'"]  
['Lasya.g', 'IND']  
["'Sangeetha', ' M. R.'"]  
["'Gagare', ' Shalmali'"]  
["'Sandhya', ' N.'"]  
["'Bakshi', ' Rutuja'"]  
["'Kanuri', ' Tejaswi'"]  
["'Mokal', ' Amruta Sunil'"]  
["'Iswarya', ' Shobana L.'"]  
["'Ramaswamy', ' Aarthie'"]  
["'Thipsay', ' Bagyashree Sathe'"]  
['Sahithi', 'Varshini']  
["'Sivakami', ' Ramanathan'"]  
["'Priyadarshini', ' M.'"]  
["'Dave', ' Dhyani'"]  
["'Meera', ' Sai'"]  
["'Manasa', ' P.'"]  
["'Dolan Champa', ' Bose'"]  
['Lakshmi', 'C']  
["'Nivedita', ' Sagade'"]  
["'Pon', ' NKrithika'"]  
["'Shah', ' Arpi'"]  
["'Dadarkar', ' Manasi'"]  
["'Delphin', ' C.'"]  
["'Kavitha', ' Ch'"]  
['Sanskriti', 'Goyal']  
["'Rajadharshini', ' M.'"]  
["'Saranya', ' J'"]  
['Potluri', 'Saye']  
["'Veena', ' Mounica M.'"]  
["'Ashwini', ' Pacchapurkar'"]  
['Anoori', 'Mehul']  
["'C.V.', ' Rajalakshmi'"]  
["'Anuradha', ' Beniwal'"]  
["'Lakshmi', ' Praneetha K'"]

['Vishwa', 'Vasnawala']  
['Anusha', 'N']  
['Kotwal', ' Archana']  
['Baisakhi', 'Das']  
['Bommini', 'Mounika']  
['Niji', ' A']  
['Singha', ' Purabi']  
['Swati', ' Mohota']  
['Agrawal', ' Kiran']  
['Pandey', 'Srishti']  
['Supriya', ' Joshi']  
['Harini', ' S.']  
['Priyamvada', 'Karamcheti']  
['Prabhudesai', ' Ninad']  
['Ratnam', ' Priya']  
['Kulkarni', ' Ashwini A.']  
['Akshaya', 'A']  
['Nilavoli', ' Hariharan']  
['Anuprita', ' Patil']  
['Sujatha', ' S.']  
['Ankitha', 'Goud']  
['Aswathy', ' Balagopalan']  
['Uthra', ' P.']  
['Kulkarni', ' Ketaki']  
['Lakshmi', ' Priya T.T.']  
['Chandreyee', ' Hajra']  
['Jennitha', ' Anto K.']  
['Ramyasri', ' J.']  
['Pragna', ' N.']  
['Chandratreya', 'Prachiti']  
['Jain', 'Nityata']  
['Madanasri', ' Gundala']  
['Mohana', ' Priya J.']  
['Dhar-Barua', ' Saheli']  
['Meghna', ' C H']  
['Pooja', ' Rathod']  
['S.', 'Nabeela']  
['Saba', ' Aziz Mhate']  
['Deodhar', 'Vrushali Umesh']  
['Gayathri', ' S']  
['Ananya', 'Rishi']  
['Bipul', ' Nandy']  
['Velpula', 'Sarayu']  
['Priya', 'IND']  
['Sindhu', ' V.K.']  
['Anjanaa', ' N. Sowjanya']  
['Savetha', ' C H']  
['Bhagyashree', 'Patil']  
['Makhija', ' Aashna']  
['Preethi', ' R.']  
['Almeida', ' Leeann']  
['Bidhar', ' Rutumbara']

['Patil', 'Mitali']  
['Madhurima', 'Shekhar']  
['Thakur', 'Aabha V.']  
['Srimathi', 'R']  
['Dhyana', 'Patel']  
['Gupta', 'Anu']  
['Saranya', 'Y']  
['Jayashree', 'P.K.']  
['Bavana', 'V.']  
['Nayana', 'K.']  
['Sharanya', 'R.']  
['Niharika', 'Ch']  
['Shruke', 'R.k.']  
['Radha', 'S. R.']  
['Sithalatchumi', 'A']  
['Jyothirmmai', 'Chodavarapu']  
['Ayushirai', 'Choudhury']  
['Gokhale', 'Anupama']  
['Bristy', 'Mukherjee']  
['Eesha', 'Ajay']  
['Meenal', 'Gupta']  
['Sandhya', 'G']  
['Ananya', 'Suresh']  
['Shivika', 'Rohilla']  
['Kotepalli', 'Sai Nirupama']  
['Ranjitha', 'K.E.R.']  
['Aatirah', 'C P']  
['Nandhini', 'Saripalli']  
['Shah', 'Vishwa']  
['Ghosh', 'Samriddhaa']  
['Priyanka', 'Kumari']  
['Poorna', 'Sri']  
['Varsha', 'Ramesh']  
['Kiruba', 'Helen Arasu']  
['Bhanot', 'Stuti']  
['Shikha', 'Shah']  
['Sirisha', 'Gutta']  
['Thomas', 'Teenu']  
['Athirai', 'S']  
['Shanya', 'Mishra']  
['G', 'Nagachandrika']  
['Gada', 'Sneha']  
['Deepa', 'A']  
['Kulkarni', 'Juily']  
['Madhuri', 'M.']  
['Jagadhambal', 'M']  
['Divya', 'Lakshmi']  
['Mishra', 'Anwasha']  
['Krithigga', 'K']  
['Roshni', 'S.']  
['Mehendi', 'Sil']  
['Hinduja', 'Reddy']

['Ria', 'Dawar']  
['Vincent', 'Alaina']  
['"Aneeqha", ' Tahseen S."']  
['Sradhanjali', 'Jena']  
['"Niyanta", ' Dave"]  
['Sweety', 'Patel']  
['Potluri', 'Supreetha']  
['Meenu', 'Priya.']  
['"Anuja", ' Bombatkar"]  
['Sai', 'Divya']  
['"Chandana", ' Priya Reddy"]  
['Sunyuktha', 'C']  
['"Rajasurya", ' J."']  
['"Shweta", ' Gole"]  
['"Velu", ' Gayatri"]  
['"Kaur", ' Palkin"]  
['Anigani', 'Kavya']  
['"Rupumoni", ' Duwara"]  
['"Sneha", ' Ghatpande"]  
['"Shradha", ' Mohanan"]  
['Rindhiya', 'V']  
['Arushi', 'Kotwal']  
['"Shristi", ' J. Shetty"]  
['Vaishnavi', 'K']  
['"Malti", ' Vhetwar"]  
['"Shruthi", ' K."']  
['Tanishka', 'Kotia']  
['Vaishnavi', 'J']  
['"Ramya", ' S."']  
['"Dhara", ' Gupta"]  
['"Preetika", ' Tayal"]  
['"Arunima", ' Kalra"]  
['"Nupur", ' Suri"]  
['Harshini', 'Ramakrishna']  
['Piyali', 'Ghosh']  
['"Fathima", ' Abdeen"]  
['"Harivardhini", ' I"]  
['Ishvi', 'Aggarwal']  
['"Meenu", ' Rajendran"]  
['Kumari', 'Rama']  
['Anjali', 'R.']  
['"Kajal", ' Basak"]  
['"Singh", ' Neha"]  
['"Vanessa", ' D`souza"]  
['"Tejaswini", ' Reddy S."']  
['Sagar', 'Siya']  
['"Shoumi", ' Mukerjee"]  
['"Sathya", ' Priya S.V."']  
['"Rathore", ' Sonakshi"]  
['Salini', 'R']  
['"Anjana Krishna", ' S"]  
['"Jimli", ' Barua"]

```
["Manasa", 'K.']  
["Bhanu", 'Priya K.']  
["Bhat", 'S. Rekha']  
["Kanitkar", 'Meghana']  
["Mahasweta", 'Kumar']  
["Pratheeka", 'G.']  
["Thelma", 'Gopal']  
["Rachana", 'Mishra']  
["Munot", 'Surbhi']  
['Sangeetha', 'P']  
["Adane", 'Narayani']  
["Sunyasakta", 'Satpathy']  
["Akshaya", 'Nandakumar']  
["Hilmi", 'Parveen']  
["Deshmukh", 'Sonia']  
['Priya', 'V']  
["Khushboo", 'M']  
["Shetye", 'Siddhali']
```

**faculty@glrica:~/hadoop-3.1.3\$ hdfs dfs -cat /Chess\_Analysis2/part-00000**

2020-09-01 16:18:04,230 INFO sasl.SaslDataTransferClient: SASL encryption trust check:  
localhostTrusted = false, remoteHostTrusted = false

```
["Srija", 'Seshadri']  
["Saranya", 'J']  
["Veena", 'Mounica M.']  
["Priyanka", 'Kumari']  
['Hinduja', 'Reddy']  
['Vaishnavi', 'J']  
["Arunima", 'Kalra']  
["Vanessa", 'D`souza']  
["Tejaswini", 'Reddy S.']  
["Anjana Krishna", 'S']  
['Priya', 'V']
```

**faculty@glrica:~/hadoop-3.1.3\$ hdfs dfs -cat /Chess\_Analysis3/part-00000**

2020-09-01 16:18:16,940 INFO sasl.SaslDataTransferClient: SASL encryption trust check:  
localhostTrusted = false, remoteHostTrusted = false