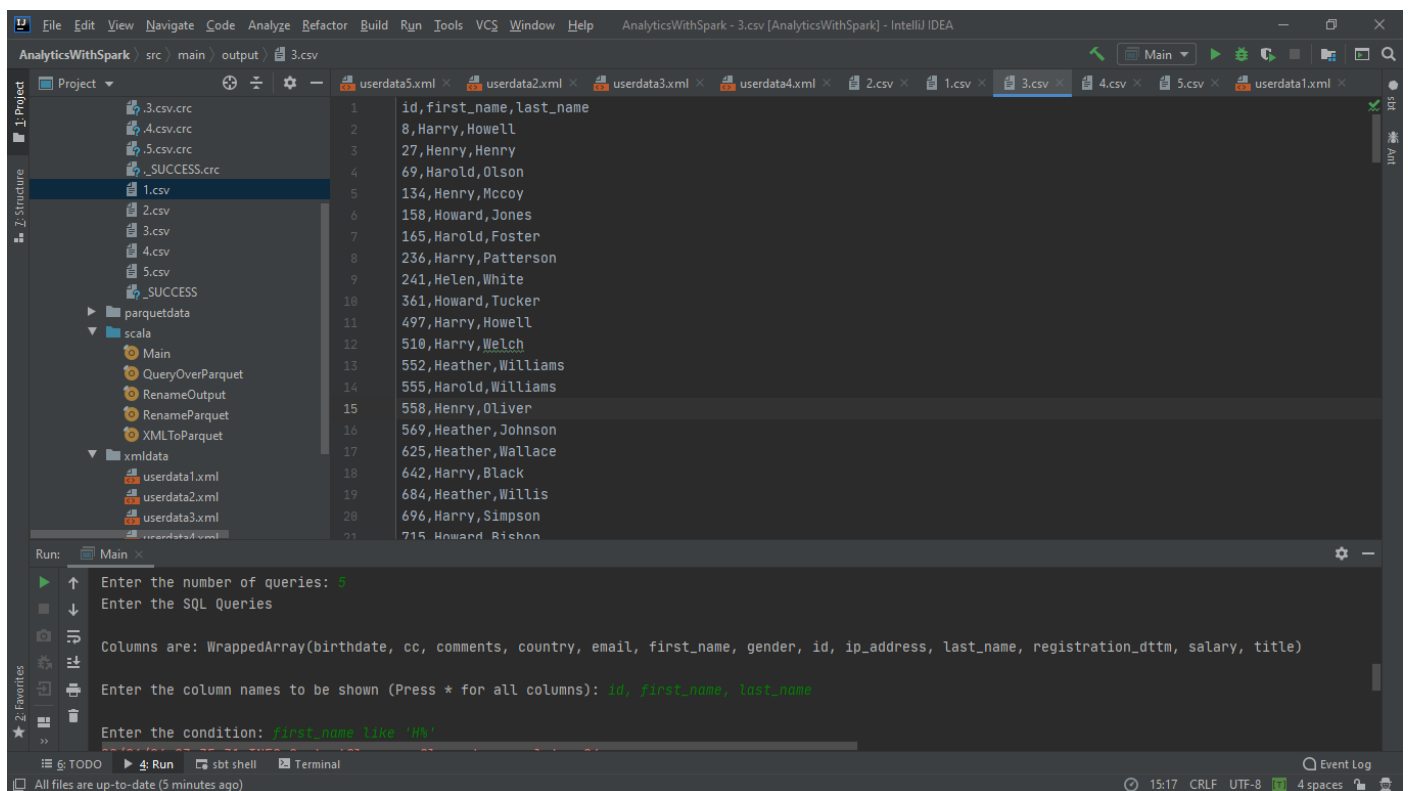


OUTPUT SCREENSHOTS

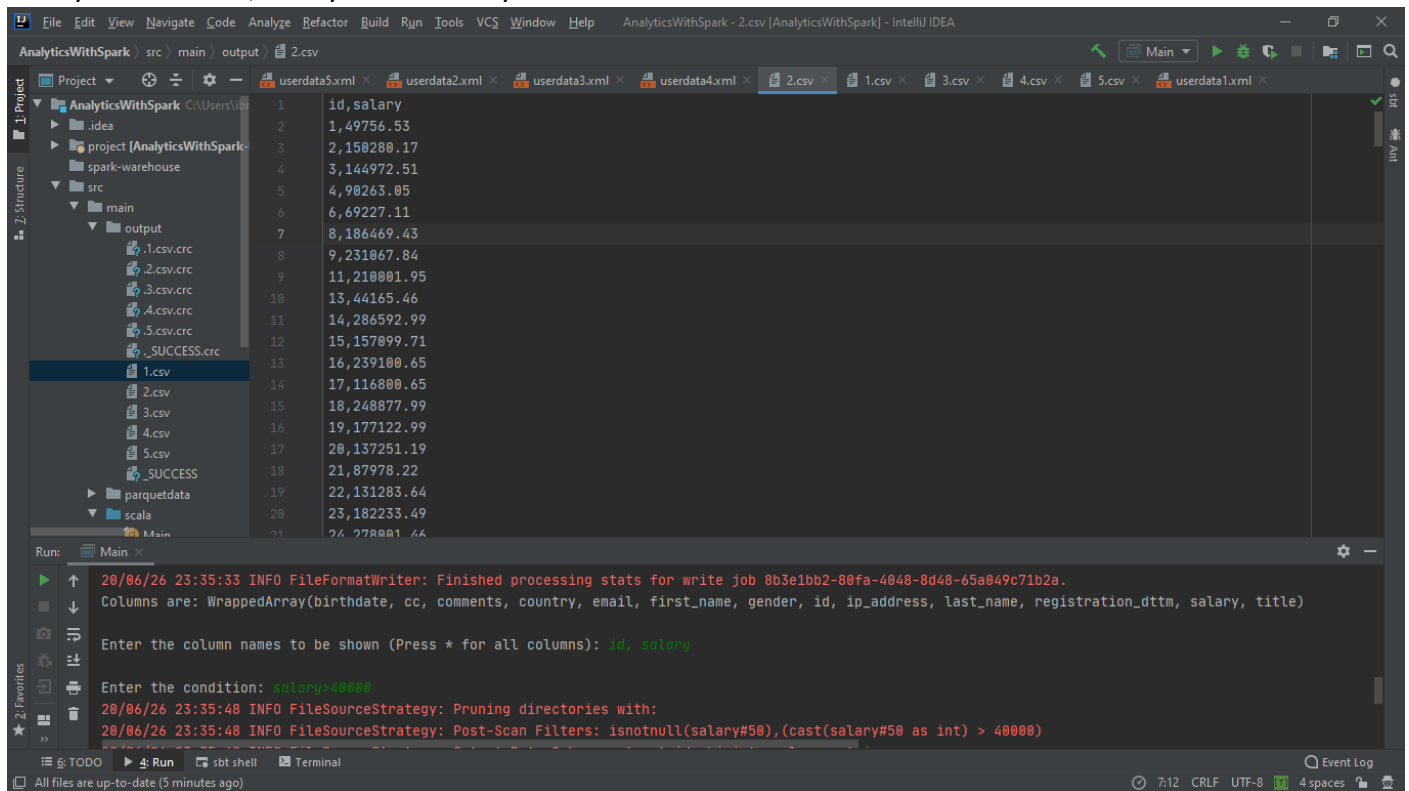
```
20/06/26 23:34:26 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, DESKTOP-7S1QIVB, 65100, None)
20/06/26 23:34:26 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, DESKTOP-7S1QIVB, 65100, None)
Enter the name of the file: *
Enter the number of files: 3
Enter the root tag: user
20/06/26 23:34:34 INFO SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/C:/
20/06/26 23:34:34 INFO SharedState: Warehouse path is 'file:/C:/Users/ibran/IdeaProjects/AnalyticsWithSpark/spark-warehouse/'.
20/06/26 23:34:35 INFO StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint
```

The column names of the transformed parquet dataframe are shown, and the user has to enter the desired columns to be displayed and the condition. The loop runs for as many times the user says. Finally, the output is saved in a folder in csv format.

Query #1: select id, first_name, last_name where first_name like 'H%'



Query #2: select id, salary where salary>40000



```
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help AnalyticsWithSpark - 2.csv [AnalyticsWithSpark] - IntelliJ IDEA

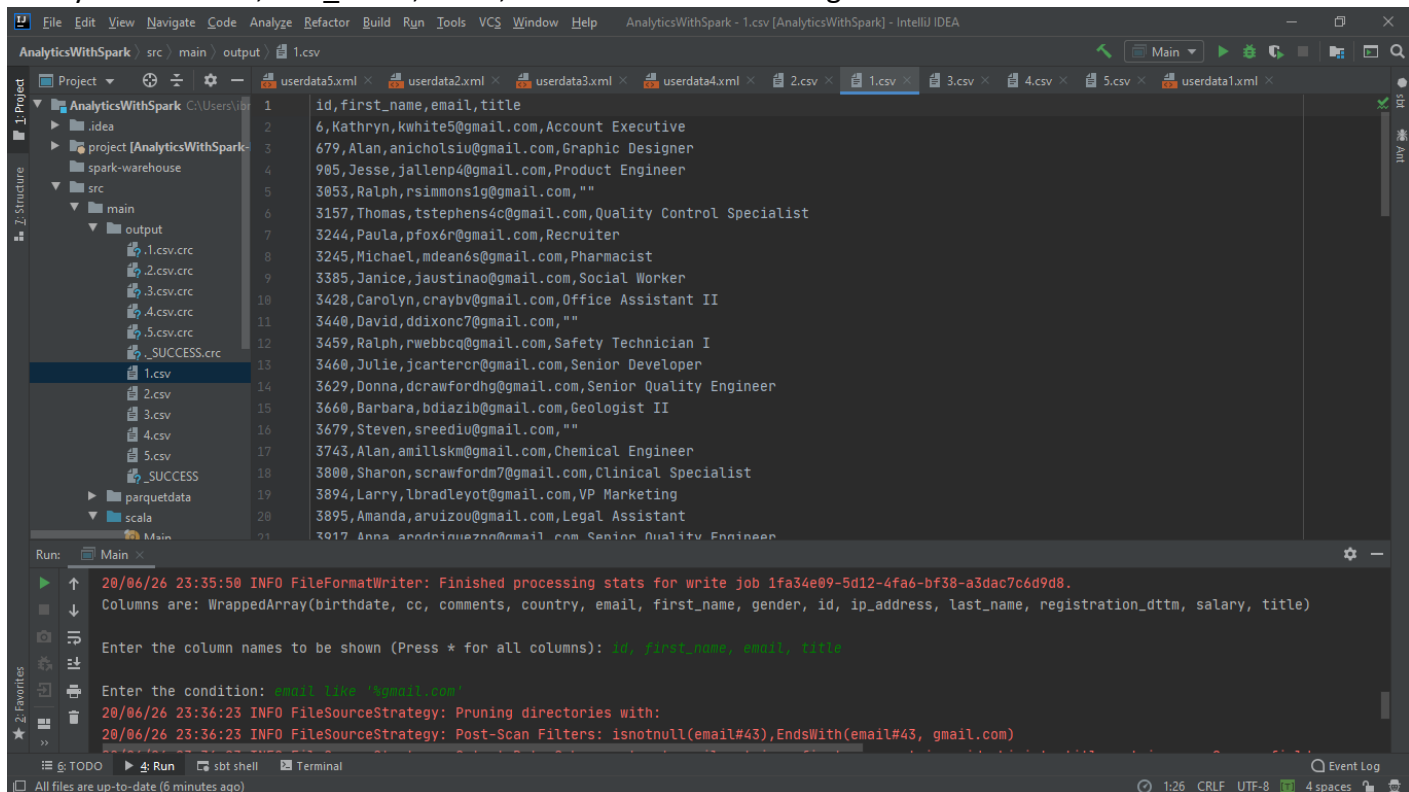
AnalyticsWithSpark src main output 2.csv

Project AnalyticsWithSpark
  .idea
  project [AnalyticsWithSpark]
  spark-warehouse
  src
  main
  output
    1.csv.crc
    2.csv.crc
    3.csv.crc
    4.csv.crc
    5.csv.crc
    SUCCESS.crc
    1.csv
    2.csv
    3.csv
    4.csv
    5.csv
    SUCCESS
  parquetdata
  scala
  Main

Run: Main
20/06/26 23:35:33 INFO FileFormatWriter: Finished processing stats for write job 8b3e1bb2-80fa-4048-8d48-65a049c71b2a.
Columns are: WrappedArray(birthdate, cc, comments, country, email, first_name, gender, id, ip_address, last_name, registration_dttm, salary, title)
Enter the column names to be shown (Press * for all columns): id, salary
Enter the condition: salary > 40000
20/06/26 23:35:48 INFO FileSourceStrategy: Pruning directories with:
20/06/26 23:35:48 INFO FileSourceStrategy: Post-Scan Filters: isnotnull(salary#50), (cast(salary#50 as int) > 40000)

1 id,salary
2 1,49756.53
3 2,150280.17
4 3,144972.51
5 4,90263.05
6 6,69227.11
7 8,186469.43
8 9,231067.84
9 11,210001.95
10 13,44165.46
11 14,286592.99
12 15,157099.71
13 16,239100.65
14 17,116800.65
15 18,248877.99
16 19,177122.99
17 20,137251.19
18 21,87978.22
19 22,131283.64
20 23,182233.49
21 24,278001.44
```

Query #3: select id, first_name, email, title where email like '%gmail.com'



```
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help AnalyticsWithSpark - 1.csv [AnalyticsWithSpark] - IntelliJ IDEA

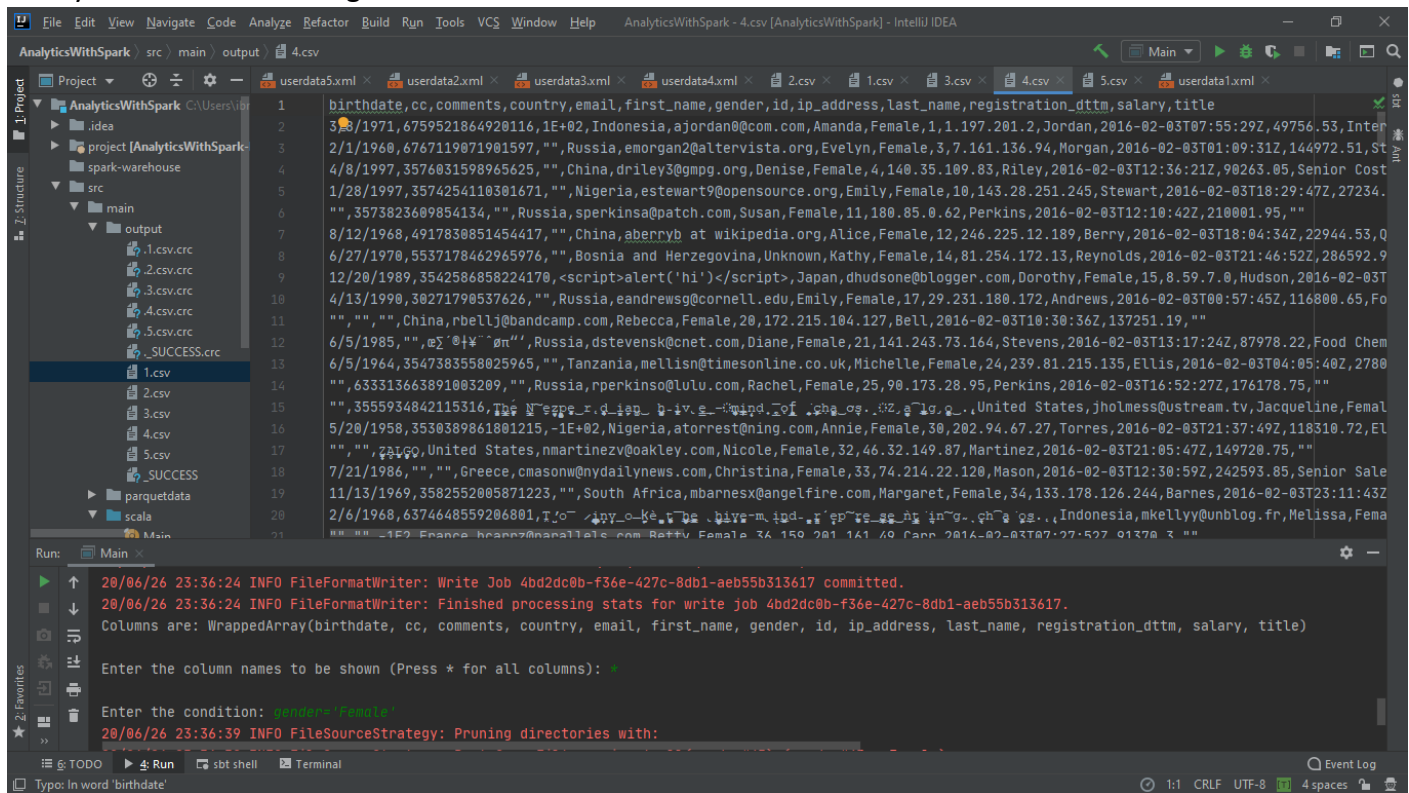
AnalyticsWithSpark src main output 1.csv

Project AnalyticsWithSpark
  .idea
  project [AnalyticsWithSpark]
  spark-warehouse
  src
  main
  output
    1.csv.crc
    2.csv.crc
    3.csv.crc
    4.csv.crc
    5.csv.crc
    SUCCESS.crc
    1.csv
    2.csv
    3.csv
    4.csv
    5.csv
    SUCCESS
  parquetdata
  scala
  Main

Run: Main
20/06/26 23:35:50 INFO FileFormatWriter: Finished processing stats for write job 1fa34e09-5d12-4fa6-bf38-a3dac7c6d9d8.
Columns are: WrappedArray(birthdate, cc, comments, country, email, first_name, gender, id, ip_address, last_name, registration_dttm, salary, title)
Enter the column names to be shown (Press * for all columns): id, first_name, email, title
Enter the condition: email like '%gmail.com'
20/06/26 23:36:23 INFO FileSourceStrategy: Pruning directories with:
20/06/26 23:36:23 INFO FileSourceStrategy: Post-Scan Filters: isnotnull(email#43), EndsWith(email#43, gmail.com)

1 id,first_name,email,title
2 6,Kathryn,kwhite5@gmail.com,Account Executive
3 679,Alan,anicholsiv@gmail.com,Graphic Designer
4 905,Jesse,jallenp4@gmail.com,Product Engineer
5 3053,Ralph,rsimmons1g@gmail.com,""
6 3157,Thomas,tstephens4c@gmail.com,Quality Control Specialist
7 3244,Paula,pfox6r@gmail.com,Recruiter
8 3245,Michael,mdean6s@gmail.com,Pharmacist
9 3385,Janice,jaustinao@gmail.com,Social Worker
10 3428,Carolyn,craybv@gmail.com,Office Assistant II
11 3440,David,ddixonc7@gmail.com,""
12 3459,Ralph,rwebbcq@gmail.com,Safety Technician I
13 3460,Julie,jcartercr@gmail.com,Senior Developer
14 3629,Donna,dcrawfordhg@gmail.com,Senior Quality Engineer
15 3660,Barbara,bdiazib@gmail.com,Geologist II
16 3679,Steven,sreediv@gmail.com,""
17 3743,Alan,amillskm@gmail.com,Chemical Engineer
18 3800,Sharon,scrawfordm7@gmail.com,Clinical Specialist
19 3894,Larry,lbradleyot@gmail.com,VP Marketing
20 3895,Amanda,arvizou@gmail.com,Legal Assistant
21 3017,Anna,aprodiguezn@gmail.com,Senior Quality Engineer
```

Query #4: select * where gender='Female'

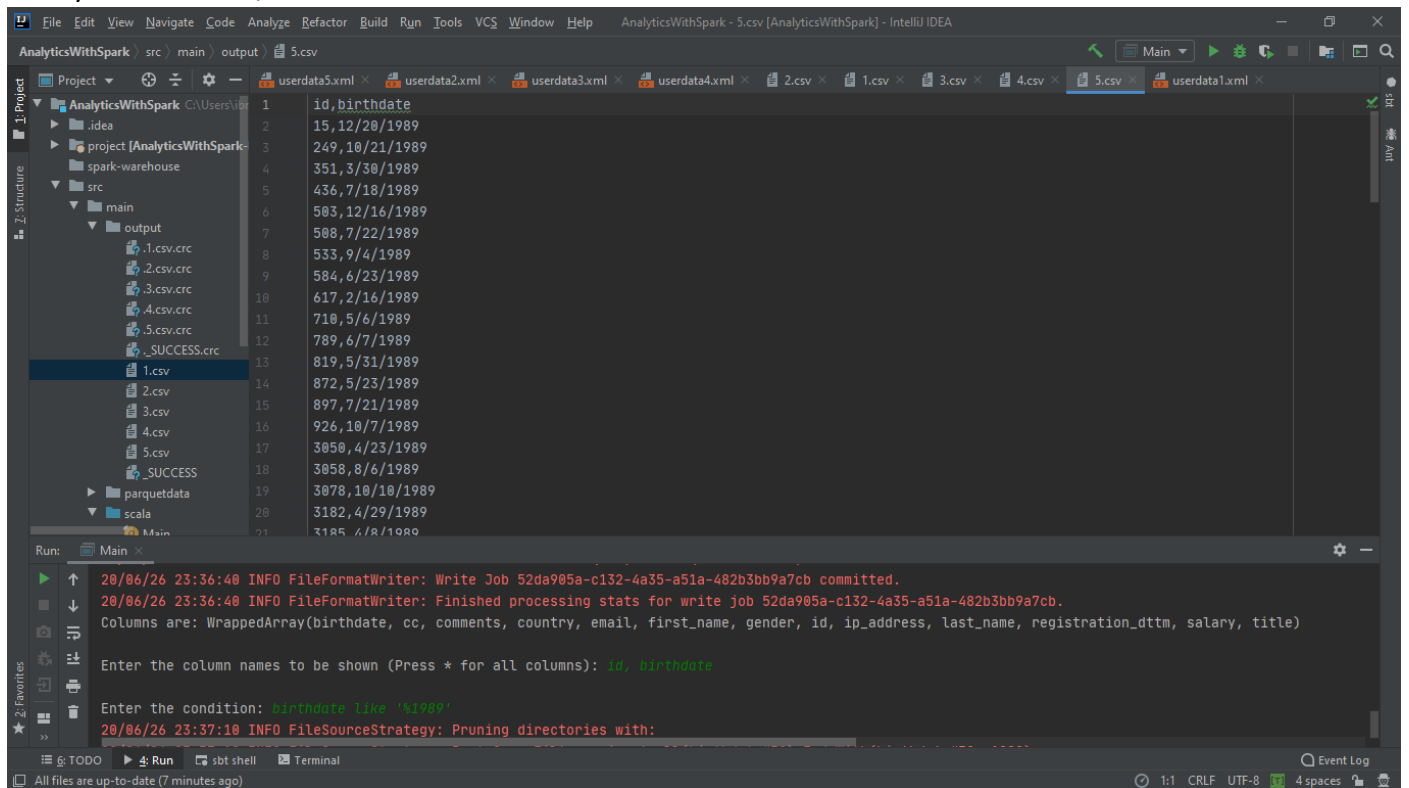


The screenshot shows the IntelliJ IDEA interface with a Spark SQL query being executed. The query is `select * where gender='Female'`. The results are displayed in a table with columns: birthdate, cc, comments, country, email, first_name, gender, id, ip_address, last_name, registration_dttm, salary, title. The results show a list of female users with their personal and professional details.

```
select * where gender='Female'
```

birthdate	cc	comments	country	email	first_name	gender	id	ip_address	last_name	registration_dttm	salary	title
3/8/1971	6759521864920116	1E+02	Indonesia	ajordan@com.com	Amanda	Female	1	1.197.201.2	Jordan	2016-02-03T07:55:29Z	49756.53	Inter
2/1/1960	6767119071901597	""	Russia	emorgan2@altervista.org	Evelyn	Female	3	7.161.136.94	Morgan	2016-02-03T01:09:31Z	144972.51	St
4/8/1997	357603159865625	""	China	driley3@gmpg.org	Denise	Female	4	140.35.109.83	Riley	2016-02-03T12:36:21Z	90263.05	Senior Cost
1/28/1997	3574254110301671	""	Nigeria	estewart9@opensource.org	Emily	Female	10	143.28.251.245	Stewart	2016-02-03T18:29:47Z	27234.	
""	3573823609854134	""	Russia	sperkinsa@patch.com	Susan	Female	11	180.85.0.62	Perkins	2016-02-03T12:10:42Z	210001.95	""
8/12/1968	4917830851454417	""	China	aberryb@wikipedia.org	Alice	Female	12	246.225.12.189	Berry	2016-02-03T18:04:34Z	22944.53	Q
6/27/1970	5537178462965976	""	Bosnia and Herzegovina	Unknown	Kathy	Female	14	81.254.172.13	Reynolds	2016-02-03T21:46:52Z	286592.9	
12/20/1989	3542586858224170	<script>alert('hi')</script>	Japan	dhudson@blogger.com	Dorothy	Female	15	8.59.7.0	Hudson	2016-02-03T		
4/13/1990	30271790537626	""	Russia	eandrewsg@cornell.edu	Emily	Female	17	29.231.180.172	Andrews	2016-02-03T00:57:45Z	116800.65	Fo
""	""	""	China	rbell@bandcamp.com	Rebecca	Female	20	172.215.104.127	Bell	2016-02-03T10:30:36Z	137251.19	""
6/5/1985	""	""	Russia	dstevensk@cnet.com	Diane	Female	21	141.243.73.164	Stevens	2016-02-03T13:17:24Z	87978.22	Food Chem
6/5/1964	3547383558025965	""	Tanzania	mellism@timesonline.co.uk	Michelle	Female	24	239.81.215.135	Ellis	2016-02-03T04:05:40Z	2780	
""	633313663891003209	""	Russia	rperkins@lulu.com	Rachel	Female	25	90.173.28.95	Perkins	2016-02-03T16:52:27Z	176178.75	""
""	3555934842115316	""	United States	jholmess@ustream.tv	Jacqueline	Fema						
5/20/1958	3530389861801215	-1E+02	Nigeria	atorrest@ning.com	Annie	Female	30	202.94.67.27	Torres	2016-02-03T21:37:49Z	118310.72	El
""	""	""	United States	nmartinezv@oakley.com	Nicole	Female	32	46.32.149.87	Martinez	2016-02-03T21:05:47Z	149720.75	""
7/21/1986	""	""	Greece	cmasonw@nydailynews.com	Christina	Female	33	74.214.22.120	Mason	2016-02-03T12:30:59Z	242593.85	Senior Sale
11/13/1969	3582552005871223	""	South Africa	mbarnesx@angelfire.com	Margaret	Female	34	133.178.126.244	Barnes	2016-02-03T23:11:43Z		
2/6/1968	6374648559206801	""	Indonesia	mkellyy@unblog.fr	Melissa	Fema						

Query #5: select id, birthdate where birthdate like '%1989'



The screenshot shows the IntelliJ IDEA interface with a Spark SQL query being executed. The query is `select id, birthdate where birthdate like '%1989'`. The results are displayed in a table with columns: id, birthdate. The results show a list of users born in 1989.

```
select id, birthdate where birthdate like '%1989'
```

id	birthdate
15	12/20/1989
249	10/21/1989
351	3/30/1989
436	7/18/1989
503	12/16/1989
508	7/22/1989
533	9/4/1989
584	6/23/1989
617	2/16/1989
710	5/6/1989
789	6/7/1989
819	5/31/1989
872	5/23/1989
897	7/21/1989
926	10/7/1989
3050	4/23/1989
3058	8/6/1989
3078	10/10/1989
3182	4/29/1989
3185	4/8/1989