

Census Data Analysis for Development

Objective:

Census Data Analysis for Education status, Socio-Economic, Employment generation and Product selling. This project would help in analyzing socio-economic status, employment and education on various factors like age group, citizenship, income, etc. All the above would help in development of both individual and nation.

Given Data:

1. Census_Records.json:

Age	Education	Marital Status	Gender	Tax Filer Status	Income	Parents	Country of Birth	Citizenship	Weeks Worked

2. AgeGroup.dat:

Fields: Age and Group

Census Analysis on following factors:

1. Education
2. Finance
3. Social
4. Citizenship and Immigrants
5. Employability
6. Product Selling & Others

Education:

Education means a lot in everyone's life as it facilitates our learning, knowledge and skill. Education is must for both men and women equally as both together make a healthy and educated society. An educated member certainly has a greater chance to contribute to his community. Education helps you become an active member of the society and participate in the ongoing changes and developments.

For Example: Education status for each category of female and male. How many female adults have Bachelor's degree.

We have achieved the same in our project in various Tasks mentioned below.

1. Total count of male/female based on education.
2. Total count of employed/unemployed based on education.
3. Total count for people in age range of 18-25 based on education.

Finance:

Taxes are like income for the government so that they can pay for socialized services within your country for service such as health care, pension funds, welfare, homeland security, war, etc. The countries with extremely prosperous and the citizens have a high per capita income wants to pay the tax regularly.

For example, Count of Tax filer status for each category of citizenship, what is the count of non-filer of middle-aged female who is a Foreign born- Not a citizen of U S.

We have achieved the same in our project in various Tasks mentioned below.

1. Tax analysis total and gender wise
2. Per Capita Income (PCI) analysis consolidated, gender wise and category wise

Socio-economic status:

The government wants establish Pension schemes, Scholarship and special employment for the persons who were orphans, elderly, and female widowed.

Analyzing socio-economic status in for each age group and gender based on citizenship- Average income: Average hours an individual works.

We have achieved the same in our project in various Tasks mentioned below.

1. Total amount dispensed on pension in x year(s)
2. Total amount dispensed on scholarship in current year
3. For given age range employable female widowed and divorced count

Citizenship and Immigrants

Analyzing various factors on both US citizenship and immigrants on various factors.

For example, how many non - United States born has acquired U S citizenship.

Count the total number of people whose Country of Birth is not United States and has citizenship as Foreign born- U S citizen by naturalization

We have achieved the same in our project in various Tasks mentioned below.

1. Citizens and immigrants count for employed lot
2. Non-US citizen(s) tax filer status
3. Country of birth wise count for US citizenship by naturalization

Employment Generation:

Analyzing qualification and gender in each category for employment.

For example, Count how many adult male have an education as some college but no degree and their average salary.

We have achieved the same in our project in various Tasks mentioned below.

1. For given age range employable female widowed and divorced count
2. Degree wise count for employability
3. Total count of employed/unemployed based on education.

Product Selling & Others:

Based on given data we can analysis selling of new product in the market also we can analysis factors like babies whose age less than 5 for vaccination, Orphan count with no education. Counting orphan who are not studying and based on their average income.

For example, count total infant whose parents are not in universe, have no education and their average income.

We have achieved the same in our project in various Tasks mentioned below.

1. Voter(s) count in x year(s)
2. Customer base analysis

Technology used:

Apache Hadoop:

Apache Hadoop an open-source software framework used for distributed storage and processing of very large data sets. It consists of computer clusters built from commodity hardware.

Characteristics of Hadoop:

- Reliable
- Flexible
- Economical
- Scalable

Map reduce:

Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner.

A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks.

Apache Pig:

Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.

Features:

- Ease of programming
- Optimization opportunities
- Extensibility.

Apache Hive:

The Apache Hive data warehouse software facilitates reading, writing, and managing large datasets residing in distributed storage using SQL. Structure can be projected onto data already in storage

HARDWARE & SOFTWARE REQUIREMENTS:

Operating System	:	LINUX
RAM	:	8GB
System Type	:	64 bit OS
Development Tool	:	Eclipse
Language Used	:	MapReduce Java, Hive, Pig

Education:

Task1: Total count of male/female based on education.

Output: Hive:

```
hive> 1274 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 610 msec
OK
10th grade      Female 12187
10th grade      Male   10384
11th grade      Female 10815
11th grade      Male    9690
12th grade no diploma  Female 2970
12th grade no diploma  Male   3304
1st 2nd 3rd or 4th grade      Female 2764
1st 2nd 3rd or 4th grade      Male   2591
5th or 6th grade      Female 4992
5th or 6th grade      Male   4761
7th and 8th grade      Female 12609
7th and 8th grade      Male   11518
9th grade         Female 9780
9th grade         Male   8755
Associates degree-academic program      Female 7684
Associates degree-academic program      Male   5266
Associates degree-occup /vocational      Female 9225
Associates degree-occup /vocational      Male   6733
Bachelors degree(BA AB BS)      Female 29557
Bachelors degree(BA AB BS)      Male   29680
Children          Female 69827
Children          Male   71669
Doctorate degree(PhD EdD)      Female 1099
Doctorate degree(PhD EdD)      Male   2714
High school graduate      Female 80977
High school graduate      Male   63857
Less than 1st grade      Female 1279
Less than 1st grade      Male   1133
Masters degree(MA MS MEng MEd MSW MBA)  Female 9493
Masters degree(MA MS MEng MEd MSW MBA)  Male   10150
Prof school degree (MD DDS DVM LLB JD)  Female 1530
Prof school degree (MD DDS DVM LLB JD)  Male   3828
Some college but no degree      Female 45012
Some college but no degree      Male   38690
Time taken: 156.749 seconds
hive>
```

Output:Pig

```
(( Children, Male),71669)
(( Children, Female),69827)
(( 9th grade, Male),8755)
(( 9th grade, Female),9780)
(( 10th grade, Male),10384)
(( 10th grade, Female),12187)
(( 11th grade, Male),9690)
(( 11th grade, Female),10815)
(( 5th or 6th grade, Male),4761)
(( 5th or 6th grade, Female),4992)
(( 7th and 8th grade, Male),11518)
(( 7th and 8th grade, Female),12609)
(( Less than 1st grade, Male),1133)
(( Less than 1st grade, Female),1279)
(( High school graduate, Male),63857)
(( High school graduate, Female),80977)
(( 12th grade no diploma, Male),3304)
(( 12th grade no diploma, Female),2970)
(( 1st 2nd 3rd or 4th grade, Male),2591)
(( 1st 2nd 3rd or 4th grade, Female),2764)
(( Doctorate degree(PhD EdD), Male),2714)
(( Doctorate degree(PhD EdD), Female),1099)
(( Bachelors degree(BA AB BS), Male),29680)
(( Bachelors degree(BA AB BS), Female),29557)
(( Some college but no degree, Male),38690)
(( Some college but no degree, Female),45012)
(( Associates degree-academic program, Male),5266)
(( Associates degree-academic program, Female),7684)
(( Associates degree-occup /vocational, Male),6733)
(( Associates degree-occup /vocational, Female),9225)
(( Masters degree(MA MS MEng MEd MSW MBA), Male),10150)
(( Masters degree(MA MS MEng MEd MSW MBA), Female),9493)
(( Prof school degree (MD DDS DVM LLB JD), Male),3828)
(( Prof school degree (MD DDS DVM LLB JD), Female),1530)
icloudera@localhost ~$
```

*t1 (~/Desktop/mydata) - gedit

Task2: Total count of employed/unemployed based on education.

Output: Hive

```
IK
10th grade      Employed-->    12044.0 UnEMPloyed-->   10527.0
11th grade      Employed-->    8798.0 UnEMPloyed-->   11707.0
12th grade no diploma Employed-->    2681.0 UnEMPloyed-->   3593.0
1st 2nd 3rd or 4th grade Employed-->    3339.0 UnEMPloyed-->   2016.0
5th or 6th grade      Employed-->    5511.0 UnEMPloyed-->   4242.0
7th and 8th grade      Employed-->   17234.0 UnEMPloyed-->   6893.0
9th grade      Employed-->   11430.0 UnEMPloyed-->   7105.0
Associates degree-academic program Employed-->    2094.0 UnEMPloyed-->    1
1856.0
Associates degree-occup /vocational Employed-->    2820.0 UnEMPloyed-->    1
138.0
Bachelors degree(BA AB BS) Employed-->    9615.0 UnEMPloyed-->  49622.0
Children      Employed-->   141496.0 UnEMPloyed-->   NULL
Doctorate degree(PhD EdD) Employed-->    530.0 UnEMPloyed-->   3283.0
High school graduate Employed-->   44342.0 UnEMPloyed-->  100492.0
Less than 1st grade      Employed-->    1678.0 UnEMPloyed-->    734.0
Masters degree(MA MS MEng MEd MSW MBA) Employed-->    2937.0 UnEMPloyed-->    1
1706.0
Prof school degree (MD DDS DVM LLB JD) Employed-->    666.0 UnEMPloyed-->    4
92.0
Some college but no degree      Employed-->   19037.0 UnEMPloyed-->   64665.0
Time taken: 135.667 seconds
hive>
```

Output:Pig-Employed

```
2010-11-20 22:20:41,473 [main] INFO org.apache.pig.backend.hadoop
( 9th grade,7105)
( 10th grade,10527)
( 11th grade,11707)
( 5th or 6th grade,4242)
( 7th and 8th grade,6893)
( Less than 1st grade,734)
( High school graduate,100492)
( 12th grade no diploma,3593)
( 1st 2nd 3rd or 4th grade,2016)
( Doctorate degree(PhD EdD),3283)
( Bachelors degree(BA AB BS),49622)
( Some college but no degree,64665)
( Associates degree-academic program,10856)
( Associates degree-occup /vocational,13138)
( Masters degree(MA MS MEng MEd MSW MBA),16706)
( Prof school degree (MD DDS DVM LLB JD),4692)
[cloudera@localhost ~]$
```

Output:Pig-Unemployed:

```
( Children,141496)
( 9th grade,11430)
( 10th grade,12044)
( 11th grade,8798)
( 5th or 6th grade,5511)
( 7th and 8th grade,17234)
( Less than 1st grade,1678)
( High school graduate,44342)
( 12th grade no diploma,2681)
( 1st 2nd 3rd or 4th grade,3339)
( Doctorate degree(PhD EdD),530)
( Bachelors degree(BA AB BS),9615)
( Some college but no degree,19037)
( Associates degree-academic program,2094)
( Associates degree-occup /vocational,2820)
( Masters degree(MA MS MEng MEd MSW MBA),2937)
( Prof school degree (MD DDS DVM LLB JD),666)
[cloudera@localhost ~]$
```

Task 3.Total count for people in age range of 18-25 based on education.

Output: Hive:

```
OK
Education--> 10th grade      Total Count--> 2411
Education--> 11th grade      Total Count--> 5310
Education--> 12th grade no diploma Total Count--> 1824
Education--> 1st 2nd 3rd or 4th grade      Total Count--> 275
Education--> 5th or 6th grade      Total Count--> 871
Education--> 7th and 8th grade      Total Count--> 989
Education--> 9th grade      Total Count--> 1486
Education--> Associates degree-academic program      Total Count--> 1414
Education--> Associates degree-occup /vocational      Total Count--> 1558
Education--> Bachelors degree(BA AB BS)      Total Count--> 5714
Education--> Doctorate degree(PhD EdD)      Total Count--> 15
Education--> High school graduate      Total Count--> 18966
Education--> Less than 1st grade      Total Count--> 187
Education--> Masters degree(MA MS MEng MEd MSW MBA) Total Count--> 358
Education--> Prof school degree (MD DDS DVM LLB JD) Total Count--> 27
Education--> Some college but no degree      Total Count--> 20311
Time taken: 29.134 seconds
hive> █
```

Finance

Task1: Tax analysis total and gender wise

Output:Hive

```
Female 1710.1663736369826
Male 1772.7254616592884
Time taken: 28.998 seconds
hive> █
```

Task2: Per Capita Income(PCI) analysis consolidated,gender wise and category wise

Output: Hive: Category wise

```
age group--> Teenager      sum of income--> 1689.5446269570016
age group--> adult      sum of income--> 1813.7500828047719
age group--> elderly      sum of income--> 1662.5739941670317
age group--> infants      sum of income--> 1667.2678898605448
age group--> middle-aged      sum of income--> 1737.4900611355397
age group--> senior citizen      sum of income--> 1708.379683926455
Time taken: 66.15 seconds
hive> █
```

Output:Hive :Total PCI:

```
TotalPCI--> 1740.0260960934236
Time taken: 29.013 seconds
hive> █
```


Planning

Task 1.Voter(s) count in x year(s)

Input: Hive: Data from user: Enter year

```
hive> set year=2017;
hive> select 'Total_Voters Count-->', COUNT(*) from final_census where age+(${hiveconf:year}-YEAR(from_unixtime(unix_timestamp())))>=18;
Total MapReduce jobs = 1
```

Output: Hive

```
OK
Total_Voters Count-->      437549
Time taken: 31.156 seconds
hive> █
```

Task 2. Senior Citizen(s) count in x year(s)

Input: Hive: Data from user: Enter year

```
hive> set year=2017;
hive> select 'Total_Senior_Citizen in given year-->',count(*) from final_census where age+(${hiveconf:year}-YEAR(from_unixtime(unix_timestamp())))>=60;
Total MapReduce jobs = 1
Launching Job 1 out of 1
```

Output:Hive

```
OK
Total_Senior_Citizen in given year--> 100079
Time taken: 30.949 seconds
hive> █
```

Task 3.Total number of Male/Female

```
Total MapReduce jobs = 1
Time taken: 29.985 seconds
OK
gender-->      Female Total count-->  311800
gender-->      Male   Total count-->  284723
Time taken: 29.985 seconds
hive> █
```

Task 4.Citizens and immigrants count for employed lot

1.11.16 10:00:00 AM 47.200.300.100

```
hive> select 'CitizenShip-->',citizen,'Total Count-->', COUNT(*) from ( select CASE citizen when ' Native- Born in the United States' then 'Native Born United States' else 'Immigrants' END citizen fr  
om final_census) a group by citizen;
```

Total MapReduce jobs = 1

Output: Hive

```
OK
CitizenShip--> Immigrants      Total Count--> 67265
CitizenShip--> Native Born United States      Total Count--> 529258
Time taken: 26.96 seconds
hive> _
```

Miscellaneous

Task1: .Degree wise count for Employability

Output:MapReduce:

```
hduser@ubuntu64server:~$ hadoop fs -cat /kk1/p*
10th grade      10527
11th grade      11707
12th grade no diploma  3593
1st 2nd 3rd or 4th grade      2016
5th or 6th grade      4242
7th and 8th grade      6893
9th grade        7105
Associates degree-academic program      10856
Associates degree-occup /vocational      13138
Bachelors degree(BA AB BS)      49622
Children         0
Doctorate degree(PhD EdD)      3283
High school graduate      100492
Less than 1st grade      734
Masters degree(MA MS MEng MEd MSW MBA)  16706
Prof school degree (MD DDS DVM LLB JD)  4692
Some college but no degree      64665
hduser@ubuntu64server:~$
```

```
hive>
> select 'Education-->',edu,'Total Count-->',count(*) from final_census where ww=0 group by edu;
```

Output: Hive

```
OK
Education--> 10th grade      Total Count--> 12044
Education--> 11th grade      Total Count--> 8798
Education--> 12th grade no diploma Total Count--> 2681
Education--> 1st 2nd 3rd or 4th grade Total Count--> 3339
Education--> 5th or 6th grade    Total Count--> 5511
Education--> 7th and 8th grade    Total Count--> 17234
Education--> 9th grade        Total Count--> 11430
Education--> Associates degree-academic program Total Count--> 2094
Education--> Associates degree-occup /vocational Total Count--> 2820
Education--> Bachelors degree(BA AB BS) Total Count--> 9615
Education--> Children            Total Count--> 141496
Education--> Doctorate degree(PhD EdD) Total Count--> 530
Education--> High school graduate Total Count--> 44342
Education--> Less than 1st grade    Total Count--> 1678
Education--> Masters degree(MA MS MEng MEd MSW MBA) Total Count--> 2937
Education--> Prof school degree (MD DDS DVM LLB JD) Total Count--> 666
Education--> Some college but no degree Total Count--> 19037
Time taken: 28.947 seconds
hive>
```

t1 (~/Desktop/mydata) - gedit

Output: Adv Map Reduce:

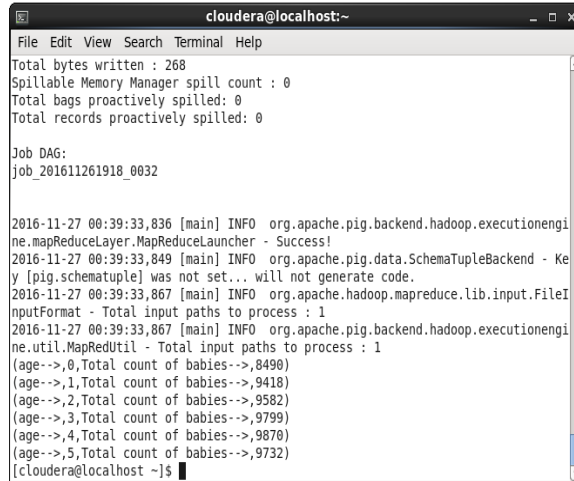
```
hduser@ubuntu64server:~$ hadoop fs -cat /kk1/p*
10th grade      10527
11th grade      11707
12th grade no diploma 3593
1st 2nd 3rd or 4th grade 2016
5th or 6th grade 4242
7th and 8th grade 6893
9th grade       7105
Associates degree-academic program 10856
Associates degree-occup /vocational 13138
Bachelors degree(BA AB BS) 49622
Children        0
Doctorate degree(PhD EdD) 3283
High school graduate 100492
Less than 1st grade 734
Masters degree(MA MS MEng MEd MSW MBA) 16706
Prof school degree (MD DDS DVM LLB JD) 4692
Some college but no degree 64665
hduser@ubuntu64server:~$
```

Task 2.Customer base analysis

Analyzing number of babies for vaccination whose age <=5:

Input: Output: Pig

```
a = load '/user/cloudera/Census Records.json' using JsonLoader
('age:int,edu:chararray,mar:chararray,gen:chararray,tax:chararray,income:long,parent:chararray,country:chararray,citizen:chararray,ww:int');
b = foreach a generate age,gen,income;
c = filter b by age <=5;
d = group c by age;
e = foreach d generate 'age-->', group,'Total count of babies-->',COUNT(c.age);
dump e;
```



```
cloudera@localhost:~
File Edit View Search Terminal Help

Total bytes written : 268
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_201611261918_0032

2016-11-27 00:39:33,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2016-11-27 00:39:33,849 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2016-11-27 00:39:33,867 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2016-11-27 00:39:33,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(age-->,0,Total count of babies-->,8490)
(age-->,1,Total count of babies-->,9418)
(age-->,2,Total count of babies-->,9582)
(age-->,3,Total count of babies-->,9799)
(age-->,4,Total count of babies-->,9870)
(age-->,5,Total count of babies-->,9732)
[cloudera@localhost ~]$
```

Task 3.Non-US citizen(s) tax filer status

```
hive> select age,edu,gen,'TaxFilerStatus-->', tax,'Citizenship-->',citizen ,income,ww from final_census where citizen not in(' Native- Born in the United States');
total MapReduce jobs = 1
```

Output:Hive

```
37 Bachelors degree(BA AB BS) Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 624.68 0
3 Children Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 1205.3 0
25 Some college but no degree Male TaxFilerStatus--> Single Citizenship--> Foreign born- Not a citizen of U S 3442.4 52
28 Bachelors degree(BA AB BS) Female TaxFilerStatus--> Single Citizenship--> Foreign born- U S citizen by naturalization 1741.48 52
53 1st 2nd 3rd or 4th grade Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 1370.33 28
38 7th and 8th grade Female TaxFilerStatus--> Head of household Citizenship--> Foreign born- Not a citizen of U S 1219.11 0
14 Children Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 677.03 0
2 Children Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 862.44 0
3 Children Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 1804.74 0
46 5th or 6th grade Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- U S citizen by naturalization 2688.61 52
43 Some college but no degree Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 2239.83 52
41 5th or 6th grade Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 938.99 52
26 11th grade Male TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 1199.34 52
22 Some college but no degree Male TaxFilerStatus--> Joint both under 65 Citizenship--> Native- Born abroad of American Parent(s) 1900.14 52
12 Children Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- U S citizen by naturalization 1532.61 0
52 12th grade no diploma Male TaxFilerStatus--> Nonfiler Citizenship--> Native- Born in Puerto Rico or U S Outlying 1140.64 0
25 Some college but no degree Male TaxFilerStatus--> Single Citizenship--> Foreign born- Not a citizen of U S 1740.9 52
46 Some college but no degree Male TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 1031.19 52
48 High school graduate Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- U S citizen by naturalization 740 52
35 High school graduate Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 1504.92 0
26 9th grade Male TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 1171.52 52
28 12th grade no diploma Male TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 1140.07 52
43 Some college but no degree Male TaxFilerStatus--> Single Citizenship--> Native- Born abroad of American Parent(s) 1019.25 36
24 High school graduate Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- U S citizen by naturalization 852.49 52
31 High school graduate Male TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- U S citizen by naturalization 648.87 26
39 12th grade no diploma Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 1432.86 0
53 High school graduate Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- U S citizen by naturalization 2590.42 26
19 5th or 6th grade Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 1329.61 0
49 High school graduate Female TaxFilerStatus--> Single Citizenship--> Native- Born in Puerto Rico or U S Outlying 1198.34 52
23 High school graduate Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- Not a citizen of U S 2632.78 52
38 Some college but no degree Female TaxFilerStatus--> Joint both under 65 Citizenship--> Foreign born- U S citizen by naturalization 1386.91 52
32 Some college but no degree Male TaxFilerStatus--> Single Citizenship--> Foreign born- Not a citizen of U S 1230.37 0
46 1st 2nd 3rd or 4th grade Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 762.63 0
37 7th and 8th grade Male TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 1195.92 0
24 High school graduate Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 1761.07 0
24 7th and 8th grade Male TaxFilerStatus--> Single Citizenship--> Foreign born- Not a citizen of U S 2938.54 52
51 Masters degree(MA MS MEng MEd MSW MBA) Male TaxFilerStatus--> Single Citizenship--> Foreign born- U S citizen by naturalization 672.59 52
3 Children Male TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 649.07 0
26 5th or 6th grade Female TaxFilerStatus--> Nonfiler Citizenship--> Foreign born- Not a citizen of U S 1287.85 0
Time taken: 38.064 seconds
hive>
```

Task 4. Country of birth wise count for US citizenship by naturalization

```
Time taken: 0.1200 seconds  
hive> select cntry, count(citizen) from final_census where citizen  
      > = ' Foreign born- U S citizen by naturalization' group by cntry;
```

Output:Hive:

```
OK  
?      3113  
Cambodia      75  
Canada 770  
China  430  
Columbia      397  
Cuba  1251  
Dominican-Republic      379  
Ecuador      192  
El-Salvador      227  
England      496  
France  87  
Germany      1054  
Greece 300  
Guatemala      98  
Haiti  144  
Holand-Netherlands      28  
Honduras      87  
Hong Kong      99  
Hungary      187  
India  384  
Iran  141  
Ireland      206  
Italy  793  
Jamaica      342  
Japan  152  
Laos  82  
Mexico 2218  
Nicaragua      110  
Panama  38  
Peru  202  
Philippines      1220  
Poland 577  
Portugal      248  
Scotland      106  
South Korea      472  
Taiwan 283  
Thailand      53  
Trinidad&Tobago      62  
Vietnam      371  
Yugoslavia      141  
Time taken: 27.363 seconds  
hive> █
```

Social:

Task 1.Total amount dispensed on pension in x year(s)

Input:Adv Map Reduce

Data from user: Enter year:

Data Validation : Yes

```
hduser@ubuntu64server:~$ hadoop jar c6.jar /censustext.txt /kk6;  
Pension in Year : Enter Year  
2020
```

Output: MR

```
Bytes Written: 32  
hduser@ubuntu64server:~$ hadoop fs -cat /kk6/p*;  
Total Pension amount for the given year--> 21405000
```

Task 2.Total amount dispensed on scholarship in current year

Input: Secondary table:Scholar2

Father only present,	500
Mother only present,	700
Neither parent present,	700
Not in universe,	1000

Output: Pig:

```

a = load '/user/cloudera/Census_Records.json' using JsonLoader
('Age:int,Education:chararray,MartialStatus:chararray,Gender:chararray,TaxFilerStatus:chararray,Income:float,Parents:chararray,
CountryOfBirth:chararray,Citizenship:chararray,WeeksWorked:chararray');
b = load '/user/cloudera/scholar2' using PigStorage(',') as (status:chararray,schamt:int);
c = join a by Parents,b by status;
d = foreach c generate $6 as parent,$11 as Schamt;
e = group d by $0;
f = foreach e generate group,SUM(d.Schamt);
dump f;

```

```

cloudera@localhost:~
File Edit View Search Terminal Help
2016-11-27 02:33:02,425 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2016-11-27 02:33:02,429 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2016-11-27 02:33:02,436 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2016-11-27 02:33:02,436 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
( Not in universe,431452000)
( Father only present,2781500)
( Mother only present,26821900)
( Neither parent present,3411100)
[cloudera@localhost ~]$ d

```

Task3: Input: Map Reduce: Data From user: Entering min and Max age:

Data Validation: Yes

```

hduser@ubuntu64server:~$ hadoop jar c4.jar /Census_Records.json /jj15
Enter Min age
22
Enter Max age
30

```

Output:Map reduce:

```

hduser@ubuntu64server:~$ hadoop fs -cat /jj15/p*
Employed female widowed and Divorced in the given age is--> 1901
hduser@ubuntu64server:~$

```

Additional Tasks

Task: Total Number of babies between age 0 to 5.

```

hive> select 'Total no of babies between 0 to 5 age-->', count(*) from final_census where age<=5;

```

```

Total no of babies between 0 to 5 age--> 56891
Time taken: 26.067 seconds

```

Conclusion:

Census Data Analysis can enhance various factors in terms of Education, Social challenges, improve status of Economy, product selling. The above will helps in to promote common economic and helps to ensure the balanced and rapid development of all parts of the country.