Pig Use Case Pokemon Data Analysis

Problem Statement 1

Find the list of players that have been selected in the qualifying round (DEFENCE>55)..

Steps to find the list of players that have been selected in the qualifying round.

- In line1, we are loading the pokemon dataset, using PigStorage as Load_Data .
- In line 2, we are Filtering the loaded data having Defense>55
- Finally, using dump, we are printing the selected_list.

Pig Scripts

```
Load_Data = LOAD '/home/acadgild/pig/assignment_5.3_pokemon/Pokemon.csv' USING
PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defens
e:int,Sp_Atk:int,Sp_Def:int,Speed:int);

selected_list = FILTER Load_Data BY Defense>55;

dump selected_list;
```

Result:

```
2017-10-28 18:56:34,458 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
(2,Ivysaur,Grass,Poison,405,60,62,63,80,80,60)
(3,Venusaur,Grass,Poison,525,80,82,83,100,100,80)
(3,VenusaurMega Venusaur,Grass,Poison,625,80,100,123,122,120,80)
(5,Charmeleon,Fire,,405,58,64,58,80,65,80)
(6,Charizard,Fire,Flying,534,78,84,78,109,85,100)
(6,CharizardMega Charizard X,Fire,Dragon,634,78,130,111,130,85,100)
(6,CharizardMega Charizard Y,Fire,Flying,634,78,104,78,159,115,100)
(7,Squirtle,Water,,314,44,48,65,50,64,43)
(8,Wartortle,Water,,405,59,63,80,65,80,58)
(9,Blastoise,Water,,530,79,83,100,85,105,78)
(9,BlastoiseMega Blastoise,Water,,630,79,103,120,135,115,78)
(18,Pidgeot,Normal,Flying,479,83,80,75,70,70,101)
(18,PidgeotMega Pidgeot,Normal,Flying,579,83,80,80,135,80,121)
(20,Raticate,Normal,,413,55,81,60,50,70,97)
(22,Fearow,Normal,Flying,442,65,90,65,61,61,100)
(24,Arbok,Poison,,438,60,85,69,65,79,80)
(27,Sandshrew,Ground,,300,50,75,85,20,30,40)
(28,Sandslash,Ground,,450,75,100,110,45,55,65)
(30,Nidorina,Poison,,365,70,62,67,55,55,56)
(31,Nidoqueen,Poison,Ground,505,90,92,87,75,85,76)
```

Sample Output:

(2,Ivysaur,Grass,Poison,405,60,62,63,80,80,60)
(3,Venusaur,Grass,Poison,525,80,82,83,100,100,80)
(3,VenusaurMega Venusaur,Grass,Poison,625,80,100,123,122,120,80)
(5,Charmeleon,Fire,,405,58,64,58,80,65,80)
(6,Charizard,Fire,Flying,534,78,84,78,109,85,100)
(6,CharizardMega Charizard X,Fire,Dragon,634,78,130,111,130,85,100)
(6,CharizardMega Charizard Y,Fire,Flying,634,78,104,78,159,115,100)

Problem Statement 2

State the number of players taking part in the competition after getting selected in the qualifying round.

Steps to find the number of players taking part in the competition after getting selected in the qualifying round.

- In line1, We are counting the number of items in the selected_list , for that, we are grouping the selected_list;
- In line 2, For the grouped statement we are counting the selected_list.

Pig Scripts

```
group_selcted_list = Group selected_list All;

count_selcted_list = foreach group_selcted_list GENERATE COUNT(selected_list);

DUMP;
```

Result:

```
2017-10-28 20:01:15,085 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
(544)
```

Output:

(544)

Problem Statement 3

Using random() generate random numbers for each Pokémon on the selected list

Steps to generate random numbers for each Pokémon on the selected list.

- We are generating the random number for each pokemon item in the selected_list and including it column0, by giving the name as random_include1
- Dumping the result.

Pig Scripts

```
random_include1 = foreach selected_list GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,Sp_Atk,Sp_Def,Speed;

DUMP;
```

Output:

```
2017-10-28 19:05:59,431 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
(0.13369528499292804,Ivysaur,Grass,Poison,405,60,62,63,80,80,60)
(0.6063276492728455,Venusaur,Grass,Poison,525,80,82,83,100,100,80)
(0.5954552281351726,VenusaurMega Venusaur,Grass,Poison,625,80,100,123,122,120,80)
(0.5308450188875791,Charmeleon,Fire,,405,58,64,58,80,65,80)
(0.09018812658453357,Charizard,Fire,Flying,534,78,84,78,109,85,100)
(0.4140194442865689,CharizardMega Charizard X,Fire,Dragon,634,78,130,111,130,85,100)
(0.8385085365228895,CharizardMega Charizard Y,Fire,Flying,634,78,104,78,159,115,100)
(0.797599106520715,Squirtle,Water,,314,44,48,65,50,64,43)
(0.24937485047840424,Wartortle,Water,,405,59,63,80,65,80,58)
(0.013108982810311254,Blastoise,Water,,530,79,83,100,85,105,78)
(0.780802814991062,BlastoiseMega Blastoise,Water,,630,79,103,120,135,115,78)
(0.16887917957919474,Pidgeot,Normal,Flying,479,83,80,75,70,70,101)
(0.8526220968533081,PidgeotMega Pidgeot,Normal,Flying,579,83,80,80,135,80,121)
(0.3005130743489165,Raticate,Normal,,413,55,81,60,50,70,97)
(0.5454345737195112,Fearow,Normal,Flying,442,65,90,65,61,61,100)
```

Sample_Result:

(0.13369528499292804,Ivysaur,Grass,Poison,405,60,62,63,80,80,60)
(0.6063276492728455,Venusaur,Grass,Poison,525,80,82,83,100,100,80)
(0.5954552281351726,VenusaurMega Venusaur,Grass,Poison,625,80,100,123,122,120,80)
(0.5308450188875791,Charmeleon,Fire,,405,58,64,58,80,65,80)
(0.09018812658453357,Charizard,Fire,Flying,534,78,84,78,109,85,100)
(0.4140194442865689,CharizardMega Charizard X,Fire,Dragon,634,78,130,111,130,85,100)
(0.8385085365228895,CharizardMega Charizard Y,Fire,Flying,634,78,104,78,159,115,100)
(0.797599106520715,Squirtle,Water,,314,44,48,65,50,64,43)
(0.24937485047840424,Wartortle,Water,,405,59,63,80,65,80,58)
(0.013108982810311254,Blastoise,Water,,530,79,83,100,85,105,78)
(0.780802814991062,BlastoiseMega Blastoise,Water,,630,79,103,120,135,115,78)
(0.16887917957919474,Pidgeot,Normal,Flying,479,83,80,75,70,70,101)
(0.8526220968533081,PidgeotMega Pidgeot,Normal,Flying,579,83,80,80,135,80,121)
(0.3005130743489165,Raticate,Normal,,413,55,81,60,50,70,97)
(0.5454345737195112,Fearow,Normal,Flying,442,65,90,65,61,61,100)
(0.7178758788192388,Arbok,Poison,,438,60,85,69,65,79,80)
(0.7448790741754444,Sandshrew,Ground,,300,50,75,85,20,30,40)
(0.5623240954539279,Sandslash,Ground,,450,75,100,110,45,55,65)
(0.7970032931912423,Nidorina,Poison,,365,70,62,67,55,55,5

Problem Statement 4

Arrange the new list in a descending order according to a column randomly.

Steps to Arrange the new list in a descending order according to a column randomly.

- Once the random number is generated in random_include1, we are arranging the items in random_include1, based on the descending value of random number and giving the name as random1_desending
- Dumping the result

Pig Scripts

```
random1_desending = ORDER random_include1 BY $0 DESC;

DUMP;
```

Output:

```
2017-10-28 19:10:44,496 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
(0.9969461722246384,Zebstrika,Electric,,497,75,100,63,80,63,116)
(0.9948256843515454,Dusknoir,Ghost,,525,45,100,135,65,135,45)
(0.9940171970059921,Walrein,Ice,Water,530,110,80,90,95,90,65)
(0.9887018422083367,Emolga,Electric,Flying,428,55,75,60,75,60,103)
(0.9860437928530448,DarmanitanZen Mode,Fire,Psychic,540,105,30,105,140,105,55)
(0.9855190135674572,Primeape,Fighting,,455,65,105,60,60,70,95)
(0.9850884868782851,LucarioMega Lucario,Fighting,Steel,625,70,145,88,140,70,112)
(0.9833586043868916,Claydol,Ground,Psychic,500,60,70,105,70,120,75)
(0.9830995758672981,Seviper,Poison,,458,73,100,60,100,60,65)
(0.981050702675742,KangaskhanMega Kangaskhan,Normal,,590,105,125,100,60,100,100)
(0.9797284283196244,Leafeon,Grass,,525,65,110,130,60,65,95)
(0.9694229468071048,GardevoirMega Gardevoir,Psychic,Fairy,618,68,85,65,165,135,100)
(0.9686349425161301,Sawsbuck,Normal,Grass,475,80,100,70,60,70,95)
(0.9684490145428372,Purugly,Normal,,452,71,82,64,64,59,112)
```

Sample Result:

(0.9969461722246384,Zebstrika,Electric,,497,75,100,63,80,63,116)
(0.9948256843515454,Dusknoir,Ghost,,525,45,100,135,65,135,45)
(0.9940171970059921,Walrein,Ice,Water,530,110,80,90,95,90,65)
(0.9887018422083367,Emolga,Electric,Flying,428,55,75,60,75,60,103)
(0.9860437928530448,DarmanitanZen Mode,Fire,Psychic,540,105,30,105,140,105,55)
(0.9855190135674572,Primeape,Fighting,,455,65,105,60,60,70,95)
(0.9850884868782851,LucarioMega Lucario,Fighting,Steel,625,70,145,88,140,70,112)
(0.9833586043868916,Claydol,Ground,Psychic,500,60,70,105,70,120,75)
(0.9830995758672981,Seviper,Poison,,458,73,100,60,100,60,65)
(0.981050702675742,KangaskhanMega Kangaskhan,Normal,,590,105,125,100,60,100,100)
(0.9797284283196244,Leafeon,Grass,,525,65,110,130,60,65,95)
(0.9694229468071048,GardevoirMega Gardevoir,Psychic,Fairy,618,68,85,65,165,135,100)
(0.9686349425161301,Sawsbuck,Normal,Grass,475,80,100,70,60,70,95)
(0.9684490145428372,Purugly,Normal,,452,71,82,64,64,59,112)
(0.9639084340923253,Alomomola,Water,,470,165,75,80,40,45,65)
(0.9636109719259757,Swalot,Poison,,467,100,73,83,73,83,55)
(0.9633195818242594,Seadra,Water,,440,55,65,95,95,45,85)
(0.9629055995286754,Parasect,Bug,Grass,405,60,95,80,60,80,30)
(0.962615880169873,Golduck,Water,,500,80,82,78,95,80,85)
(0.962436008812189,Jirachi,Steel,Psychic,600,100,100,100,100,100,100)
(0.9621594910841437,Simipour,Water,,498,75,98,63,98,63,101)

Problem Statement 5

Now on a new relation again associate random numbers for each Pokemon and arrange in descending order according to column random.

Steps to associate random numbers for each Pokemon and arrange in descending order according to column random...

- Once again,We are generating the random number for each pokemon item in the selected_list and including it column0 by giving the name as random_include2.

- Once the random number is generated in random_include2, we are arranging the items in random_include2 based on the descending value of random number and giving the name as random2_desending
- Dumping the result.

Pig Scripts

```
random_include2 = foreach selected_list GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,Sp_Atk,Sp_Def,Speed;

random2_desending = ORDER random_include2 BY $0 DESC;

DUMP;
```

Output:

```
2017-10-28 19:14:39,094 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
(0.9943381731183237,ManectricMega Manectric,Electric,,575,70,75,80,135,80,135)
(0.9941041170975857,Pancham,Fighting,,348,67,82,62,46,48,43)
(0.993327037109632,Torterra,Grass,Ground,525,95,109,105,75,85,56)
(0.9917558906153604,Marowak,Ground,,425,60,80,110,50,80,45)
(0.9897832453789286,Kingdra,Water,Dragon,540,75,95,95,95,95,85)
(0.9884875586428901,Trevenant,Ghost,Grass,474,85,110,76,65,82,56)
(0.9857481645075057,Goodra,Dragon,,600,90,100,70,110,150,80)
(0.9842736619882175,Jellicent,Water,Ghost,480,100,60,70,85,105,60)
(0.9842274012802997,HoopaHoopa Confined,Psychic,Ghost,600,80,110,60,150,130,70)
(0.9816497581257257,Pachirisu,Electric,,405,60,45,70,45,90,95)
(0.9801039925224295,Crawdaunt,Water,Dark,468,63,120,85,90,55,55)
(0.9773540288351539,Rayquaza,Dragon,Flying,680,105,150,90,150,90,95)
(0.9740791770119602,Sigilyph,Psychic,Flying,490,72,58,80,103,80,97)
(0.9740545626906792,Slowpoke,Water,Psychic,315,90,65,65,40,40,15)
(0.9739382003305058,Nidoking,Poison,Ground,505,81,102,77,85,75,85)
```

Sample Result:

```
(0.9943381731183237,ManectricMega Manectric,Electric,,575,70,75,80,135,80,135)
(0.9941041170975857,Pancham,Fighting,,348,67,82,62,46,48,43)
(0.993327037109632,Torterra,Grass,Ground,525,95,109,105,75,85,56)
(0.9917558906153604,Marowak,Ground,,425,60,80,110,50,80,45)
(0.9897832453789286,Kingdra,Water,Dragon,540,75,95,95,95,95,85)
(0.9884875586428901,Trevenant,Ghost,Grass,474,85,110,76,65,82,56)
(0.9857481645075057,Goodra,Dragon,,600,90,100,70,110,150,80)
(0.9842736619882175,Jellicent,Water,Ghost,480,100,60,70,85,105,60)
(0.9842274012802997,HoopaHoopa Confined,Psychic,Ghost,600,80,110,60,150,130,70)
(0.9816497581257257,Pachirisu,Electric,,405,60,45,70,45,90,95)
(0.9801039925224295,Crawdaunt,Water,Dark,468,63,120,85,90,55,55)
(0.9773540288351539,Rayquaza,Dragon,Flying,680,105,150,90,150,90,95)
(0.9740791770119602,Sigilyph,Psychic,Flying,490,72,58,80,103,80,97)
(0.9740545626906792,Slowpoke,Water,Psychic,315,90,65,65,40,40,15)
(0.9739382003305058,Nidoking,Poison,Ground,505,81,102,77,85,75,85)
(0.969485502580271,Conkeldurr,Fighting,,505,105,140,95,55,65,45)
```

Problem Statement 6

From the two different descending lists of random Pokemons, select the top 5 Pokemons for 2 different players.

Steps select the top 5 Pokemons for 2 different random players list.

- We are going to show only the top 5 items of random1_desending as limit_data_random1_desending
- Similarly,We are going to show only the top 5 items of random2_desending as limit_data_random2_desending

Pig Scripts

limit_data_random1_desending = LIMIT random1_desending 5 ;

DUMP;

Output for limit_data_random1_desending:

```
2017-10-28 19:22:09,762 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
(0.998436136580118,Togetic,Fairy,Flying,405,55,40,85,80,105,40)
(0.9968148125001002,Palkia,Water,Dragon,680,90,120,100,150,120,100)
(0.9933558686425085,Aurorus,Rock,Ice,521,123,77,72,99,92,58)
(0.9911183527202121,Jumpluff,Grass,Flying,460,75,55,70,55,95,110)
(0.9882708967236125,Lucario,Fighting,Steel,525,70,110,70,115,70,90)
```

Result for limit_data_random1_desending:

(0.998436136580118,Togetic,Fairy,Flying,405,55,40,85,80,105,40)
(0.9968148125001002,Palkia,Water,Dragon,680,90,120,100,150,120,100)
(0.9933558686425085,Aurorus,Rock,Ice,521,123,77,72,99,92,58)
(0.9911183527202121,Jumpluff,Grass,Flying,460,75,55,70,55,95,110)
(0.9882708967236125,Lucario,Fighting,Steel,525,70,110,70,115,70,90)

Pig Scripts

limit_data_random2_desending = LIMIT random2_desending 5 ;

DUMP;

Output for limit_data_random2_desending:

```
2017-10-28 19:23:23,660 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
(0.9985947507813646,Rhyperior,Ground,Rock,535,115,140,130,55,55,40)
(0.9949539238382045,Malamar,Dark,Psychic,482,86,92,88,68,75,73)
(0.9944653637119645,MetagrossMega Metagross,Steel,Psychic,700,80,145,150,105,110,110)
(0.990429785416758,Axew,Dragon,,320,46,87,60,30,40,57)
(0.9856652882543068,Piloswine,Ice,Ground,450,100,100,80,60,60,50)
```

Result for limit_data_random2_desending:

(0.9985947507813646,Rhyperior,Ground,Rock,535,115,140,130,55,55,40)

(0.9949539238382045,Malamar,Dark,Psychic,482,86,92,88,68,75,73)
(0.9944653637119645,MetagrossMega Metagross,Steel,Psychic,700,80,145,150,105,110,110)
(0.990429785416758,Axew,Dragon,,320,46,87,60,30,40,57)
(0.9856652882543068,Piloswine,Ice,Ground,450,100,100,80,60,60,50)

Problem Statement 7

Store the data on a local drive to announce for the final match. By the name player1 and player2 (only show the NAME and HP).

Steps to store the data on a local drive to announce for the final match. By the name player1 and player2 (by showing the NAME and HP).

- We only need to store the name and HP details of each player from limit_data_random1_desending, so we are extracting those two details as filter_only_name1
- We are storing the contents of filter_only_name1 in player1.

Pig Scripts

```
filter_only_name1 = foreach limit_data_random1_desending Generate ($1,HP);

DUMP;

STORE filter_only_name1 INTO '/home/acadgild/pig/assignment_5.3_pokemon/player1;
```

Output for limit_data_random1_desending:

```
2017-10-28 19:34:51,780 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
((Yveltal,126))
((AggronMega Aggron,70))
((Electrode,60))
((GlalieMega Glalie,80))
((AegislashShield Forme,60))
```

Result for limit_data_random1_desending :

((Yveltal,126))
((AggronMega Aggron,70))
((Electrode,60))
((GlalieMega Glalie,80))
((AegislashShield Forme,60))

- We only need to store the name and HP details of each player from limit_data_random2_desending, so we are extracting those two details as filter_only_name2
- We are storing the contents of filter_only_name1 in player2.

Pig Scripts

filter_only_name2 = foreach limit_data_random2_desending Generate ($1,HP);

DUMP;

STORE filter_only_name2 INTO '/home/acadgild/pig/assignment_5.3_pokemon/player2';

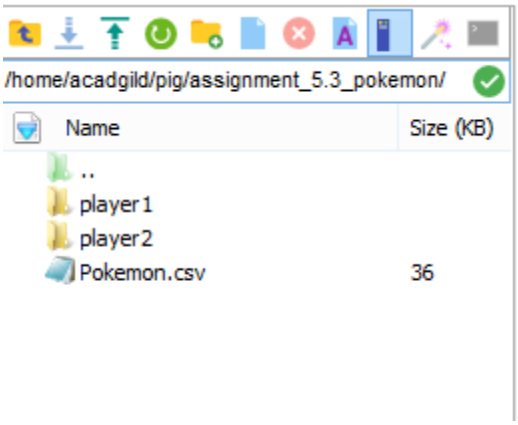
Output for limit_data_random2_desending:

```
2017-10-28 19:35:42,282 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
((Conkeldurr,105))
((Rampardos,97))
((Ninetales,73))
((Hydreigon,92))
((Victini,100))
```

Result for limit_data_random2_desending:

((Conkeldurr,105))
((Rampardos,97))
((Ninetales,73))
((Hydreigon,92))
((Victini,100))

We can see the ouputs stored in the below folder.



Content of Player1.txt
(Yveltal,126) (AggronMega Aggron,70)(Electrode,60)(GlalieMega Glalie,80) (AegislashShield Forme,60)

Contents of Player2.txt
(Conkeldurr,105) (Rampardos,97) (Ninetales,73) (Hydreigon,92) (Victini,100)