OVERVIEW

The data as described by Myles O'Neill .This data set includes 721 Pokemon, including their number, name, first and second type, and basic stats: HP, Attack, Defense, Special Attack, Special Defense, and Speed. It has been of great use when teaching statistics to kids. With certain types you can also give a geeky introduction to machine learning.

This are the raw attributes that are used for calculating how much damage an attack will do in the games. This dataset is about the pokemon games (NOT pokemon cards or Pokemon Go).

OBJECTIVES

- To provide data of pokemons
- To describe different types of pokemons
- To know the strongest and weakest pokemons

COLUMNS AND DATA TYPE

Number	Int
Name	String
Type 1	String
Type 2	String
Total	Int
Hp	Int
Attack	Int
Defense	Int
Sp_Atk	Int
Sp_Def	Int
Speed	Int
Generation	Int
Legendary	String

PROBLEM STATEMENTS:

1)Find out the average HP (Hit points) of all the Pokémon

2)Find the powerful and moderate pokemons (hp|avg hp(powerul),hp<avg hp(moderate))

3)Find out the top 10

-|Pokémons according to their HP's

-Pokémons based on their Attack stat

-Pokémons based on their Defense stat,total power

-Pokémons having a drastic change in their attack and sp.attack

-Pokémons based on their defense and sp.defense

-Pokémons according to their fastest Pokémons

4)The attack distribution of pokemons across all generations

5)All stat analysis of pokemons

6)Compare Charmander and Mega Venusaur

7)Distribution of various pokemons by type

HIVE QUERIES

- **Using Database and Creating New Table to upload the data**

```
hive> use poc;
OK
Time taken: 0.334 seconds
hive> CREATE TABLE pokemon (Number Int,Name String,Type1 String,Type2 String,Total Int,HP Int,Attack Int,Defens
 Int,Sp_Atk Int,Sp_Def Int,Speed Int,Generation Int,Legendary String) ROW FORMAT DELIMITED
    > fields terminated BY ','
    > lines terminated BY '\n'
    > tblproperties("skip.header.line.count"="1");
OK
Time taken: 1.77 seconds
hive>
```

```
hive> LOAD DATA LOCAL INPATH '/home/training/Desktop/Pokemon.csv' INTO table pokemon;
Copying data from file:/home/training/Desktop/Pokemon.csv
Copying file: file:/home/training/Desktop/Pokemon.csv
Loading data to table poc.pokemon
chgrp: changing ownership of '/user/hive/warehouse/poc.db/pokemon/Pokemon.csv': User does not be
long to hive
Table poc.pokemon stats: [num_partitions: 0, num_files: 1, num_rows: 0, total_size: 44028, raw_d
ata_size: 0]
OK
Time taken: 2.283 seconds
hive>
```

- **Description of table pokemon**

  Command:describe pokemon;

```
hive> describe pokemon;
OK
number  int
name    string
type1   string
type2   string
total   int
hp      int
attack  int
defense int
sp_atk  int
sp_def  int
speed   int
generation      int
legendary       string
Time taken: 0.4 seconds
hive>
```

PROBLEM STATEMENT 1: Find out the average HP (Hit points) of all the Pokémon

Command: Select avg(HP) from pokemon;

```
hive> Select avg(HP) from pokemon;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0001, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job
_202012010916_0001
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 10:24:42,751 Stage-1 map = 0%,  reduce = 0%
2020-12-01 10:24:47,851 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.81 sec
2020-12-01 10:24:48,864 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.81 sec
2020-12-01 10:24:49,873 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.81 sec
2020-12-01 10:24:50,881 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.81 sec
2020-12-01 10:24:51,891 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.81 sec
2020-12-01 10:24:52,902 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.81 sec
2020-12-01 10:24:53,910 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.95 sec
2020-12-01 10:24:54,919 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.95 sec
2020-12-01 10:24:55,929 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.95 sec
2020-12-01 10:24:56,939 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.95 sec
MapReduce Total cumulative CPU time: 1 seconds 950 msec
Ended Job = job_202012010916_0001
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 1.95 sec   HDFS Read: 44262 HDFS Write: 9 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 950 msec
OK
69.25875
Time taken: 29.201 seconds
```

In the above screenshot, you can see that the average Hit point of the Pokémon is **69.25875**

PROBLEM STATEMENT 2: Find the powerful and moderate pokemons **condition**:(hp|avg then hp(powerul),hp<avg hp then (moderate))

To count the number of powerful and moderate we divide them from table and put them in new table as per condition

Command:create table pokemon1 as select *, IF(HP|69.25875, 'powerful', IF(HP<69.25875, 'Moderate','powerless')) AS power_rate from pokemon;

```
hive> create table pokemon1 as select *, IF(HP>69.25875, 'powerful', IF(HP<69.25875, 'Moderate','powerless')) AS power_rate from pokemon;
Total MapReduce jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_202012010916_0002, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job_202012010916_0002
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2020-12-01 10:33:05,005 Stage-1 map = 0%,  reduce = 0%
2020-12-01 10:33:09,023 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.21 sec
2020-12-01 10:33:10,029 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.21 sec
2020-12-01 10:33:11,041 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.21 sec
2020-12-01 10:33:12,048 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.21 sec
MapReduce Total cumulative CPU time: 1 seconds 210 msec
Ended Job = job_202012010916_0002
Ended Job = -253322857, job is filtered out (removed at runtime).
Ended Job = 924022080, job is filtered out (removed at runtime).
Moving data to: hdfs://localhost.localdomain:8020/tmp/hive-training/hive_2020-12-01_10-32-56_465_3833999094079513794-1/-ext-10001
Moving data to: hdfs://localhost.localdomain:8020/user/hive/warehouse/poc.db/pokemon1
chgrp: changing ownership of 'hdfs://localhost.localdomain:8020/user/hive/warehouse/poc.db/pokemon1': User does not belong to hive
Table poc.pokemon1 stats: [num_partitions: 0, num_files: 1, num_rows: 0, total_size: 51206, raw_data_size: 0]
801 Rows loaded to hdfs://localhost.localdomain:8020/tmp/hive-training/hive_2020-12-01_10-32-56_465_3833999094079513794-1/-ext-10000
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 1.21 sec   HDFS Read: 44262 HDFS Write: 51206 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 210 msec
OK
Time taken: 15.916 seconds
hive>
```

As new table is created now we can count names

Command:select COUNT(name),power_rate from pokemon1 group by power_rate;

```
hive> select COUNT(name),power_rate from pokemon1 group by power_rate;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0003, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job
_202012010916_0003
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 10:39:00,049 Stage-1 map = 0%,  reduce = 0%
2020-12-01 10:39:04,063 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:05,076 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:06,085 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:07,097 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:08,106 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:09,111 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:10,116 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:11,121 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:12,126 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:13,133 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:14,166 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.77 sec
2020-12-01 10:39:15,175 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.9 sec
2020-12-01 10:39:16,181 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.9 sec
2020-12-01 10:39:17,188 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.9 sec
MapReduce Total cumulative CPU time: 1 seconds 900 msec
Ended Job = job_202012010916_0003
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 1.9 sec   HDFS Read: 51438 HDFS Write: 38 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 900 msec
OK
422     Moderate
378     powerful
1       powerless
Time taken: 22.059 seconds
hive>
```

In the above screenshot, you can see the count of

Moderate-422

Powerful-378

powerless-1

PROBLEM STATEMENT 3: Find out the top 10

-Pokémons according to their HP's

-|Pokémons based on their Attack stat

-|Pokémons based on their Defense stat

-|Pokémons having a drastic change in their attack and sp.attack

-|Pokémons based on their defense and sp.defense

-|Pokémons according to their fastest Pokémons

- **Pokémons according to their HP's**

  Command:select name,hp from pokemon1 order by hp desc limit 10;

```
hive> select name,hp from pokemon1 order by hp desc limit 10;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0004, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job_202012010916_000
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 11:02:50,626 Stage-1 map = 0%,  reduce = 0%
2020-12-01 11:03:04,224 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:05,257 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:06,289 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:07,316 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:08,380 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:09,429 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:10,466 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:11,491 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:12,518 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:13,537 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:14,557 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:15,575 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:16,586 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:17,605 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:18,635 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2020-12-01 11:03:19,674 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
2020-12-01 11:03:20,692 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
2020-12-01 11:03:21,713 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
2020-12-01 11:03:22,734 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
```

```
2020-12-01 11:03:23,765 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
2020-12-01 11:03:24,788 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
2020-12-01 11:03:25,803 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
MapReduce Total cumulative CPU time: 5 seconds 330 msec
Ended Job = job_202012010916_0004
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 5.33 sec   HDFS Read: 51438 HDFS Write: 152 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 330 msec
OK
Blissey 255
Chansey 250
Wobbuffet       190
Wailord 170
Alomomola       165
Snorlax 160
Slaking 150
GiratinaOrigin Forme    150
GiratinaAltered Forme   150
Drifblim        150
Time taken: 45.104 seconds
hive>
```

From the above screen shot, we can see the top 10 according to HP .

- Pokémons based on their Attack stat

  Command:select name,attack from pokemon1 order by attack desc limit 10;

```
hive> select name,attack from pokemon order by attack desc limit 10;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0005, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job_202012010916_0005
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 11:12:30,681 Stage-1 map = 0%,  reduce = 0%
2020-12-01 11:12:42,999 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:44,024 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:45,037 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:46,058 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:47,070 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:48,085 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:49,099 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:50,110 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:51,126 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 sec
2020-12-01 11:12:52,155 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
2020-12-01 11:12:53,171 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
2020-12-01 11:12:54,186 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
2020-12-01 11:12:55,198 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
2020-12-01 11:12:56,215 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
2020-12-01 11:12:57,225 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
2020-12-01 11:12:58,245 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
MapReduce Total cumulative CPU time: 4 seconds 760 msec
Ended Job = job_202012010916_0005
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 4.76 sec   HDFS Read: 44262 HDFS Write: 238 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 760 msec
OK
MewtwoMega Mewtwo X     190
HeracrossMega Heracross 185
GroudonPrimal Groudon   180
DeoxysAttack Forme      180
RayquazaMega Rayquaza   180
GarchompMega Garchomp   170
KyuremBlack Kyurem      170
GalladeMega Gallade     165
BanetteMega Banette     165
Rampardos       165
Time taken: 37.291 seconds
hive>
```

From the above screen shot, we can see the top 10 according to Attack stat .

- Pokémons based on their Defense stat
  command:select name,defense from pokemon order by defense desc limit

```
hive> select name,defense from pokemon order by defense desc limit 10;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0006, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job_202012010916_0006
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 11:21:44,596 Stage-1 map = 0%,  reduce = 0%
2020-12-01 11:21:52,645 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:21:53,659 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:21:54,670 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:21:55,682 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:21:56,696 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:21:57,712 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:21:58,738 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:21:59,754 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:22:00,766 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2020-12-01 11:22:01,775 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
2020-12-01 11:22:02,791 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
2020-12-01 11:22:03,802 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
2020-12-01 11:22:04,812 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
2020-12-01 11:22:05,822 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
2020-12-01 11:22:06,835 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
2020-12-01 11:22:07,859 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_202012010916_0006
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 4.29 sec   HDFS Read: 44262 HDFS Write: 157 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 290 msec
OK
SteelixMega Steelix     230
AggronMega Aggron       230
Shuckle 230
Steelix 200
Regirock        200
Avalugg 184
SlowbroMega Slowbro     180
Aggron  180
Cloyster        180
Bastiodon       168
Time taken: 32.325 seconds
hive>
```

From the above screen shot, we can see the top 10 according to Defense stat .

- **Pokémons having a drastic change in their attack and sp.attack**

Command:select name,(attack-sp_atk) as atk_diff from pokemon order by atk_diff limit 10;

```
hive> select name,(attack-sp_atk) as atk_diff from pokemon order by atk_diff limit 10;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0007, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job_202012010916_0007
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 11:30:00,333 Stage-1 map = 0%,  reduce = 0%
2020-12-01 11:30:10,431 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:11,471 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:12,490 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:13,519 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:14,544 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:15,566 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:16,610 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:17,656 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:18,712 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:19,757 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:20,780 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:21,802 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:22,845 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:23,889 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:24,936 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:25,988 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:27,041 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2020-12-01 11:30:28,092 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.46 sec
2020-12-01 11:30:29,181 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.46 sec
2020-12-01 11:30:30,217 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
2020-12-01 11:30:31,240 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
2020-12-01 11:30:32,249 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
2020-12-01 11:30:33,266 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
2020-12-01 11:30:34,283 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
2020-12-01 11:30:35,298 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
2020-12-01 11:30:36,339 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
MapReduce Total cumulative CPU time: 8 seconds 940 msec
Ended Job = job_202012010916_0007
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 8.94 sec   HDFS Read: 44262 HDFS Write: 171 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 940 msec
OK
Name    NULL
AlakazamMega Alakazam   -125
DarmanitanZen Mode      -110
GengarMega Gengar       -105
Chandelure      -90
Abra    -85
Alakazam        -85
Kadabra -85
Duosion -85
GardevoirMega Gardevoir -80
Time taken: 49.925 seconds
hive>
```

From the above screen shot ,we can seePokémons having a drastic change in their attack and sp.attack

- Pokémons based on their defense and sp.defense

command:select name,(defense-sp_def) as def_diff from pokemon order by def_diff limit 10;

```
hive> select name,(defense-sp_def) as def_diff from pokemon order by def_diff limit 10;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0008, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job_202012010916_0008
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 11:37:36,538 Stage-1 map = 0%,  reduce = 0%
2020-12-01 11:37:45,589 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:46,601 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:47,616 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:48,627 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:49,640 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:50,653 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:51,663 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:52,675 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:53,687 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2020-12-01 11:37:54,695 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.72 sec
2020-12-01 11:37:55,706 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.72 sec
2020-12-01 11:37:56,715 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.72 sec
2020-12-01 11:37:57,723 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.72 sec
2020-12-01 11:37:58,731 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.72 sec
2020-12-01 11:37:59,739 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.72 sec
2020-12-01 11:38:00,757 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.72 sec
MapReduce Total cumulative CPU time: 4 seconds 720 msec
Ended Job = job_202012010916_0008
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 4.72 sec   HDFS Read: 44262 HDFS Write: 148 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 720 msec
OK
Name    NULL
Blissey -125
Cryogonal       -105
Chansey -100
Regice  -100
Florges -86
Goodra  -80
Mantyke -70
KyogrePrimal Kyogre     -70
GardevoirMega Gardevoir -70
Time taken: 31.836 seconds
hive>
```

From the above screen shot ,we can seePokémons having a drastic change in their defense and sp.defense

- Pokémons according to their fastest Pokémons

  Command:Select name, speed from pokemon order by speed desc limit 10;

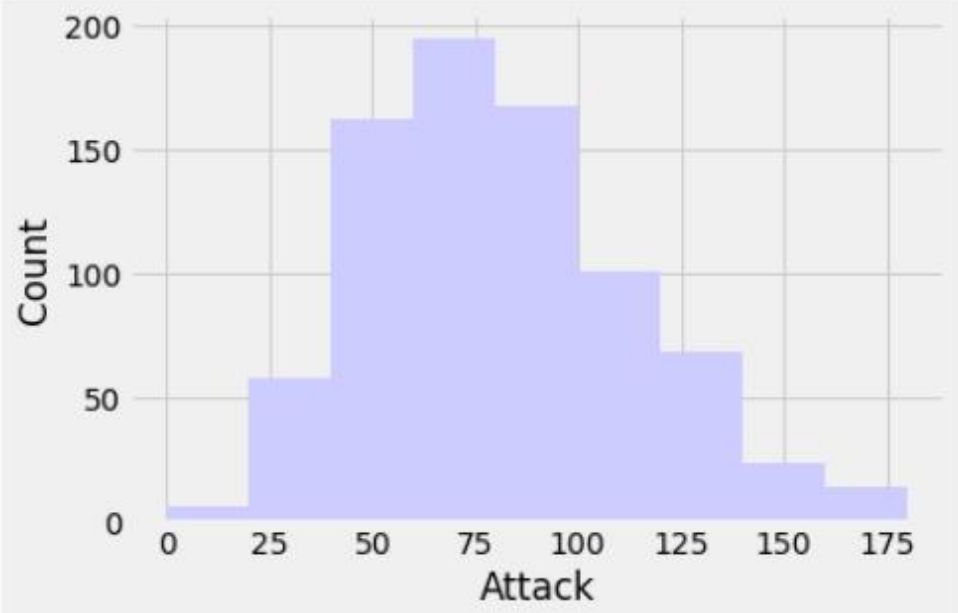
```
hive> Select name, speed from pokemon order by speed desc limit 10;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202012010916_0009, Tracking URL = http://localhost.localdomain:50030/jobdetails.jsp?jobid=job_202012010916_0009
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_202012010916_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-12-01 11:43:31,690 Stage-1 map = 0%,  reduce = 0%
2020-12-01 11:43:38,730 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:39,755 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:40,764 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:41,773 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:42,781 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:43,791 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:44,798 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:45,808 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.63 sec
2020-12-01 11:43:46,830 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.75 sec
2020-12-01 11:43:47,839 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.75 sec
2020-12-01 11:43:48,850 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.75 sec
2020-12-01 11:43:49,861 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.75 sec
2020-12-01 11:43:50,869 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.75 sec
2020-12-01 11:43:51,881 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.75 sec
2020-12-01 11:43:52,893 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.75 sec
MapReduce Total cumulative CPU time: 3 seconds 750 msec
Ended Job = job_202012010916_0009
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 3.75 sec   HDFS Read: 44262 HDFS Write: 225 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 750 msec
OK
DeoxysSpeed Forme       180
Ninjask 160
AerodactylMega Aerodactyl       150
DeoxysNormal Forme      150
DeoxysAttack Forme      150
AlakazamMega Alakazam   150
SceptileMega Sceptile   145
BeedrillMega Beedrill   145
Accelgor        145
MewtwoMega Mewtwo Y     140
Time taken: 31.921 seconds
hive>
```

From the above screen shot ,we can seePokémons which are fastest based on speed

PROBLEM STATEMENT 4: The attack distribution of pokemons across all generations

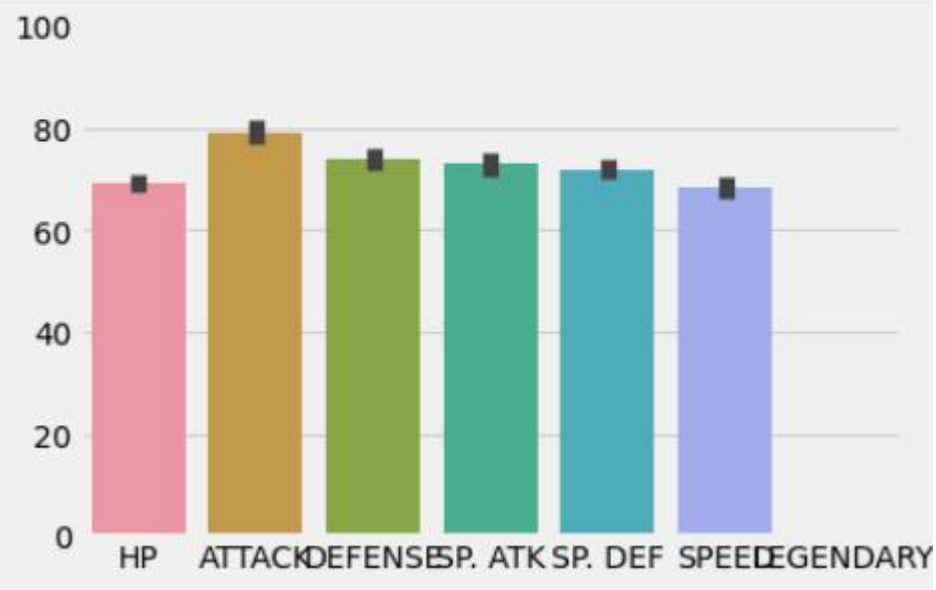
```
bins=range(0,200,20)
plt.hist(df["ATTACK"],bins,histtype="bar",rwidth=1.2,color='#CCCCFF'
plt.xlabel('Attack')
plt.ylabel('Count')
plt.plot()

plt.show()
```



PROBLEM STATEMENT 5: All stat analysis of pokemons

```
df2=df.drop(['GENERATION','TOTAL'],axis=1)
sns.barplot(data=df2)
plt.ylim(0,100)
plt.show()
```

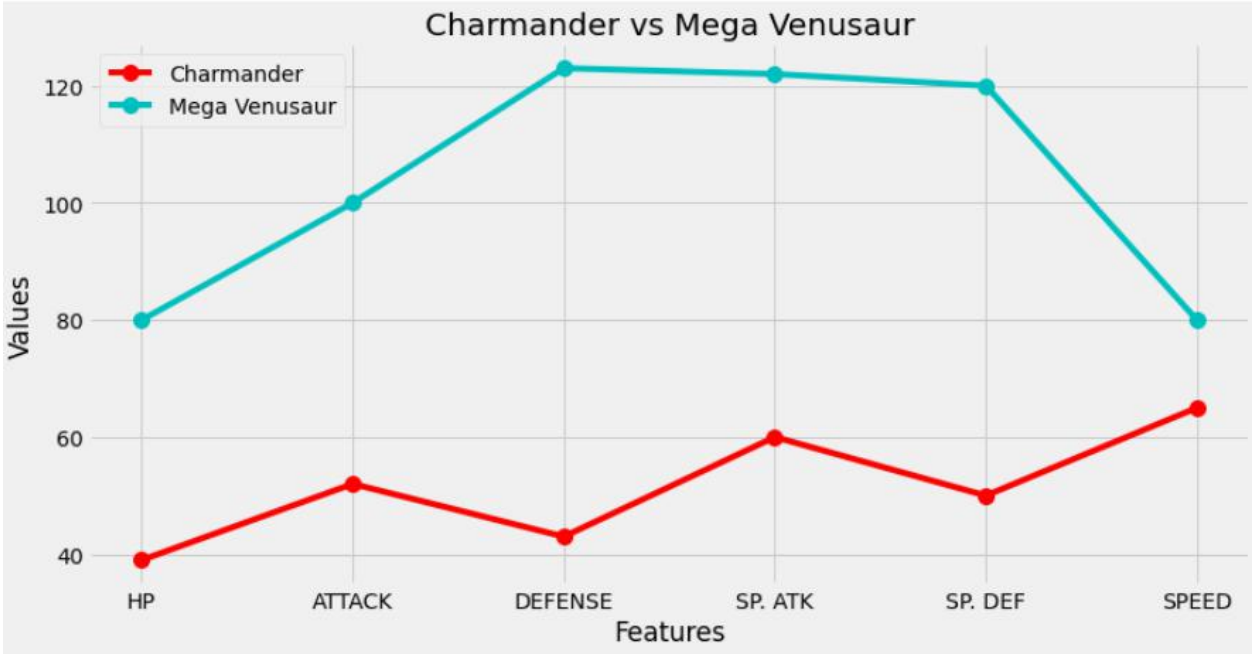


PROBLEM STATEMENT 6:Compare Charmander and Mega Venusaur

```
df_test = df.drop(["TYPE 1", "TYPE 2", "TOTAL", "GENERATION", "LEGENDARY"], axis=1)

def poke_compare(poke1, poke2):
    try:
        df_test.loc[poke1].plot(color="r", marker="o", markersize=10 , label=poke1)
        df_test.loc[poke2].plot(color="c", marker="o", markersize=10, label=poke2)
        plt.xlabel("Features")
        plt.ylabel("Values")
        plt.title(poke1 + " vs " + poke2)
        fig = plt.gcf()
        fig.set_size_inches(12, 6)
        plt.legend()
        plt.show()
    except Exception:
        print("Please enter the correct names of the Pokemons")

poke_compare("Charmander", "Mega Venusaur")
```



PROBLEM STATEMENT 7: Distribution of various types of pokemons

```
labels = 'Water', 'Normal', 'Grass', 'Bug', 'Psychic', 'Fire', 'Electric', 'Rock', 'Other'
sizes = [112, 98, 70, 69, 57, 52, 44, 44, 175]
colors = ['#0040ff', '#ffff00', '#00ff00', '#ff00bf', '#ff0000', '#006400', '#C0C0C0', '#FFFFFF', '#FF00FF']
explode = (0, 0, 0.1, 0, 0, 0, 0, 0, 0)
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=90)
plt.axis('equal')
plt.title("Percentage of Different Types of Pokemon")
plt.plot()
fig=plt.gcf()
fig.set_size_inches(7,7)
plt.show()
```