

WQD 7009 BIG DATA APPLICATIONS AND ANALYTICS

1/2022/2023

OBESITY DATASET – HBASE QUERYING

INDIVIDUAL PROJECT

Matric Number:	S2151665
Name	Kelviin Rajj Karupaya

Introduction.

Obesity is a medical condition that is occasionally referred to as a disease in which excess body fat has built up to the point that it may be harmful to health. People are classified as obese when their body mass index (BMI)—a person's weight divided by the square of the person's height—is over 30 kg/m²; the range 25–30 kg/m² is defined as overweight. Some East Asian countries use lower values to calculate obesity. Obesity is a major contributor to disability and is linked to a number of illnesses and ailments, including osteoarthritis, type 2 diabetes, obstructive sleep apnea, some forms of cancer, and cardiovascular diseases.

Individual, social, and environmental factors all contribute to obesity. Diet, physical activity, mechanisation, urbanisation, genetic predisposition, drugs, mental illnesses, economic policies, endocrine problems, and exposure to endocrine-disrupting substances are a few examples of recognised causes.

While most obese people at any given moment are trying to reduce weight and frequently succeeding, sustaining weight reduction over the long run is uncommon. No intervention that is efficient, clear-cut, and supported by research exists to prevent obesity. A multifaceted strategy, involving interventions at the social, local, family, and individual levels, is needed to prevent obesity. The major therapies suggested by medical specialists are diet modifications and exercise. If these dietary options are available, inexpensive, and accessible, diet quality can be improved by lowering consumption of energy-dense foods such those high in fat or sugar and increasing intake of dietary fibre. To suppress the appetite or lower fat absorption, medications can be used in conjunction with a healthy diet.

Obesity is a prominent global cause of mortality that is avoidable, with both adults and children are becoming more and more obese. In 195 countries in 2015, there were 600 million obese adults (12%) and 100 million obese children. Women are more likely than males to be obese. Today, obesity is stigmatized in most of the world. Conversely, some cultures, past and present, have a favourable view of obesity, seeing it as a symbol of wealth and fertility. However, a number of medical organisations, including the American Medical Association and the American Heart Association, defined obesity as a disease in 2013.

About the Dataset.

Data source: <https://www.kaggle.com/datasets/fabinmndez/obesitydata>

This paper contains data for the estimation of obesity levels in people from the countries of Mexico, Peru and Colombia, with ages between 14 and 61 and diverse eating habits and physical condition. The data was collected using a web platform with a survey where anonymous users answered each question, then the information was processed obtaining 17 attributes and 2111 records.

The attributes related with eating habits are: Frequent consumption of high caloric food (FAVC), Frequency of consumption of vegetables (FCVC), Number of main meals (NCP), Consumption of food between meals (CAEC), Consumption of water daily (CH20), and Consumption of alcohol (CALC). The attributes related with the physical condition are: Calories consumption monitoring (SCC), Physical activity frequency (FAF), Time using technology devices (TUE), Transportation used (MTRANS), other variables obtained were: Gender, Age, Height and Weight. Finally, all data was labelled and the class variable (NObesidad) was created with the values of: Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III. Based on the dataset, the attributes for every individual can be divided into 5 categories as follows:

1. General information

Data	Data type
Gender 1. Male 2. Female	String
Age 1. Young adults (<30) 2. Adults (30-50) 3. Senior Citizens (50<)	Integer
Height	Float
Weight	Integer
family_history_with_overweight (Individual with family history of obesity) 1. Yes 2. No	String

2. Eating habits

Data	Data type
FAVC (Consumption frequency of high caloric food) 1. Yes 2. No	String
FCVC (Consumption frequency of vegetables) 1. Never 2. Sometimes 3. Always	Integer
NCP (Number of main meals) 1. (<1.5) 2. (1.5 - 2.5) 3. (2.5 - 3.5) 4. (>3.5)	Integer
CAEC (Consumption of food between meals) 1. No 2. Sometimes 3. Frequently 4. Always	String
CH20 (Consumption of water daily) - Litres 1. (<1) 2. (1 - 2) 3. (>2)	Integer
CALC (Consumption frequency of alcohol) 1. No 2. Sometimes 3. Frequently 4. Always	String

3. Physical conditions

Data	Data type
SCC (Calories intake monitoring) 1. Yes 2. No	String
FAF (Physical activity frequency) 1. (<0.75) – less than a day 2. (0.75 – 1.5) – 1 to 2 days 3. (1.5 – 2.25) – 2 to 4 days 4. (2.25-3) – 4 to 5 days	Float
TUE (Time spent on technology devices) 1. (<0.66) – 0 to 2 hours 2. (0.66-1.32) – 3 to 5 hours 3. (>1.32) – more than 5 hours	Float
MTRANS (Transportation used) 1. Automobile 2. Bike 3. Motorbike 4. Public transportation	String

5. Walking	
------------	--

4. Does the person smoke

Data	Data type
SMOKE (Does the person smoke?) 1. Yes 2. No	String

5. Level of obesity (**Target Variable**)

Data	Data type
NObeyesdad (Level of obesity) 1. Insufficient_Weight 2. Normal_Weight 3. Obesity_Type_I 4. Obesity_Type_II 5. Obesity_Type_III 6. Overweight_Level_I 7. Overweight_Level_II	String

HBase queries and results.

3 shell commands to obtain the status, version and username of the HBase.

```
hbase(main):002:0> status
1 active master, 0 backup masters, 1 servers, 0 dead, 3.0000 average load

hbase(main):003:0> version
1.2.0-cdh5.10.0, rUnknown, Fri Jan 20 12:13:18 PST 2017

hbase(main):004:0> whoami
cloudera (auth:SIMPLE)
  groups: cloudera, default

hbase(main):005:0> table_help
Help for table-reference commands.

You can either create a table via 'create' and then manipulate the table via commands like 'put', 'get', etc.
See the standard help information for how to use each of these commands.

However, as of 0.96, you can also get a reference to a table, on which you can invoke commands.
For instance, you can get create a table and keep around a reference to it via:

  hbase> t = create 't', 'cf'

Or, if you have already created the table, you can get a reference to it:

  hbase> t = get_table 't'

You can do things like call 'put' on the table:

  hbase> t.put 'r', 'cf:q', 'v'

which puts a row 'r' with column family 'cf', qualifier 'q' and value 'v' into table t.

To read the data out, you can scan the table:

  hbase> t.scan

which will read all the rows in table 't'.

Essentially, any command that takes a table name can also be done via table reference.
Other commands include things like: get, delete, deleteall,
get_all_columns, get_counter, count, incr. These functions, along with
the standard JRuby object methods are also available via tab completion.

For more information on how to use each of these commands, you can also just type:

  hbase> t.help 'scan'

which will output more information on how to use that command.
```

DDL COMMANDS.

The obesity table is created in HBase with the column family (“cf”). The csv file in Hadoop Distributed File System is imported to HBase with the respective attributes assigned to the column family. Once the file import is completed, queries can be performed on the table to obtain insights.

[illegible]

```
hbase(main):007:0> describe 'obesity'
Table obesity is ENABLED
obesity
COLUMN FAMILIES DESCRIPTION
(NAME => 'cf', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true')
1 row(s) in 0.2210 seconds
```

```
hbase(main):008:0>
```

```
hbase(main):009:0> show_filters
ColumnPrefixFilter
TimestampsFilter
PageFilter
MultipleColumnPrefixFilter
FamilyFilter
ColumnPaginationFilter
SingleColumnValueFilter
RowFilter
QualifierFilter
ColumnRangeFilter
ValueFilter
PrefixFilter
SingleColumnValueExcludeFilter
ColumnCountGetFilter
InclusiveStopFilter
DependentColumnFilter
FirstKeyOnlyFilter
KeyOnlyFilter
```

```
hbase(main):010:0> disable 'obesity'
0 row(s) in 2.4860 seconds

hbase(main):011:0> list
TABLE
obesity
1 row(s) in 0.0360 seconds

=> ["obesity"]
hbase(main):012:0> is_enabled 'obesity'
false
0 row(s) in 0.0410 seconds
```

DML COMMANDS.

1. General information on attributes:

```
hbase(main):003:0> scan 'obesity'

999          COLUMNS: LAEL, timestamp=1669481071095, value=SomeLines
999          column=cf:CALC, timestamp=1669481071095, value=Sometimes
999          column=cf:CN20, timestamp=1669481071095, value=2
999          column=cf:FAF, timestamp=1669481071095, value=0
999          column=cf:FAVC, timestamp=1669481071095, value=yes
999          column=cf:FCVC, timestamp=1669481071095, value=2
999          column=cf:Gender, timestamp=1669481071095, value=Male
999          column=cf:Height, timestamp=1669481071095, value=1.7
999          column=cf:MTRANS, timestamp=1669481071095, value=Public_Transportation
999          column=cf:MCP, timestamp=1669481071095, value=3
999          column=cf:NOBeyesdad, timestamp=1669481071095, value=Overweight_Level_II
999          column=cf:SCC, timestamp=1669481071095, value=no
999          column=cf:SMOKE, timestamp=1669481071095, value=no
999          column=cf:TUE, timestamp=1669481071095, value=1
999          column=cf:Weight, timestamp=1669481071095, value=85
999          column=cf:family_history_with_overweight, timestamp=1669481071095, value=yes
2111 row(s) in 7.7950 seconds

272 row(s) in 0.9760 seconds
hbase(main):019:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('NOBeyesdad'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Insufficient_Weight')))}

287 row(s) in 0.7300 seconds
hbase(main):018:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('NOBeyesdad'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Normal_Weight')))}

290 row(s) in 1.0870 seconds
hbase(main):020:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('NOBeyesdad'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Overweight_Level_I')))}

290 row(s) in 0.6470 seconds
hbase(main):021:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('NOBeyesdad'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Overweight_Level_II')))}

351 row(s) in 0.8630 seconds
hbase(main):010:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('NOBeyesdad'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Obesity_Type_I')))}

297 row(s) in 0.5240 seconds
hbase(main):011:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('NOBeyesdad'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Obesity_Type_II')))}

324 row(s) in 0.8420 seconds
hbase(main):012:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('NOBeyesdad'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Obesity_Type_III')))}

1726 row(s) in 4.3140 seconds
hbase(main):024:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('family_history_with_overweight'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('yes')))}

1866 row(s) in 4.0490 seconds
hbase(main):025:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('FAVC'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('yes')))}

639 row(s) in 1.6500 seconds
hbase(main):026:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('CALC'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('no')))}

96 row(s) in 0.3400 seconds
hbase(main):027:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('SCC'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('yes')))}

56 row(s) in 0.4770 seconds
hbase(main):028:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('MTRANS'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('Walking')))}

44 row(s) in 0.3210 seconds
hbase(main):031:0> scan 'obesity', { FILTER => SingleColumnValueFilter.new(Bytes.toBytes('cf'), Bytes.toBytes('SMOKE'), CompareFilter::CompareOp.valueOf('EQUAL'), BinaryComparator.new(Bytes.toBytes('yes')))}

```


In this assignment, the analysis is focused on the various factors involved in affecting the level of obesity among the individuals who are Normal Weight, Overweight Level 1, Overweight Level 2, Obesity Type 1, Obesity Type 2 and Obesity Type 3. It is observed from above that there are more individuals affected with Obesity Type 1 (351), Obesity Type 3 (324) and Obesity Type 2 (297), followed by Overweight Level 1 (290) and Overweight Level 2 (290) and finally, Normal Weight (287) and Insufficient Weight (272). Based on query results from the collected dataset overall, it is observed that most of the individuals:

- Have a family history of overweight
- Practise high consumption of food with increased calories
- Practise consumption of alcohol
- Do not monitor their calories consumption
- Travel with an available transport
- Do not smoke

2. Gender classification.

```
146 row(s) in 0.6550 seconds
hbase(main):004:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','Gender',=,'binary:Male') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

145 row(s) in 0.6480 seconds
hbase(main):005:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','Gender',=,'binary:Male') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

187 row(s) in 0.7200 seconds
hbase(main):006:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','Gender',=,'binary:Male') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}

195 row(s) in 0.7660 seconds
hbase(main):007:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','Gender',=,'binary:Male') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}

295 row(s) in 0.9470 seconds
hbase(main):008:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','Gender',=,'binary:Male') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

323 row(s) in 1.1600 seconds
hbase(main):010:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','Gender',=,'binary:Female') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}
```

Based on the query results it is observed that number of males is the highest for those who are Normal Weight, Overweight Level 2, Obesity Type 1 and Obesity Type 2. The number of male and female who are Overweight Level 1 is equal. The number of females is the highest for those who are Obesity Type 3.

3. Normal Weight vs Overweight Level 1 and 2

2.1 Family history with overweight

```
155 row(s) in 0.4210 seconds
hbase(main):029:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','family_history_with_overweight',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

209 row(s) in 0.5740 seconds
hbase(main):027:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','family_history_with_overweight',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

272 row(s) in 0.9510 seconds
hbase(main):028:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','family_history_with_overweight',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}
```

2.2 Consumption frequency of high caloric food

```
208 row(s) in 0.8930 seconds
hbase(main):030:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FAVC',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

268 row(s) in 1.1710 seconds
hbase(main):031:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FAVC',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

216 row(s) in 0.7780 seconds
hbase(main):032:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FAVC',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}
```

2.3 Consumption frequency of vegetables

```
155 row(s) in 0.6040 seconds
hbase(main):051:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

186 row(s) in 0.6950 seconds
hbase(main):052:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

192 row(s) in 0.4480 seconds
hbase(main):053:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}
```

2.4 Consumption of food between meals

```
159 row(s) in 0.6720 seconds
hbase(main):038:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CAEC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

236 row(s) in 1.0290 seconds
hbase(main):039:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CAEC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

270 row(s) in 0.8720 seconds
hbase(main):040:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CAEC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}
```

2.5 Does the person smoke?

```
274 row(s) in 2.2200 seconds
hbase(main):015:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SMOKE',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

287 row(s) in 0.9860 seconds
hbase(main):023:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SMOKE',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

285 row(s) in 1.0760 seconds
hbase(main):024:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SMOKE',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}
```

2.6 Water consumption amount

```
186 row(s) in 0.5720 seconds
hbase(main):042:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CH20',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}

154 row(s) in 0.5700 seconds
hbase(main):043:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CH20',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

164 row(s) in 0.6350 seconds
hbase(main):044:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CH20',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}
```

2.7 Calories intake monitoring

```
257 row(s) in 0.8470 seconds
hbase(main):022:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SCC',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

253 row(s) in 1.0450 seconds
hbase(main):055:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SCC',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

286 row(s) in 1.5370 seconds
hbase(main):058:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SCC',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}
```

2.8 Consumption frequency of alcohol

```
143 row(s) in 0.4990 seconds
hbase(main):034:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CALC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}

224 row(s) in 0.6810 seconds
hbase(main):035:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CALC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

161 row(s) in 0.6990 seconds
hbase(main):036:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CALC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}
```

2.9 Mode of transportation used

```
189 row(s) in 0.8070 seconds
hbase(main):026:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','MTRANS',=,'binary:Public_Transportation') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}

212 row(s) in 0.7010 seconds
hbase(main):027:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','MTRANS',=,'binary:Public_Transportation') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

200 row(s) in 0.4290 seconds
hbase(main):059:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','MTRANS',=,'binary:Public_Transportation') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}
```

2.10 Time spent on technology devices

```
172 row(s) in 0.8070 seconds
hbase(main):007:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','TUE',<,'binary:0.66') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_I')",COLUMNS=>['cf']}

133 row(s) in 0.2990 seconds
hbase(main):011:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','TUE',<,'binary:0.66') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Overweight_Level_II')",COLUMNS=>['cf']}

129 row(s) in 0.2750 seconds
hbase(main):014:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','TUE',<,'binary:0.66') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}
```

From the querying results above, it is observed that majority of the individuals in all 3 categories:

- have family members that suffered or suffers with overweight
- consume food with high calorific value
- consume vegetables at a moderate amount with their meals
- sometimes take food in between their meals
- do not smoke
- consume 1-2 litres of water daily
- do not monitor their calories consumption on a daily basis
- sometimes consume alcohol
- usually use public transport as a mode of transportation
- spend 0 to 2 hours in technological devices such as smartphone, television, computer, etc.

4. Normal Weight vs Obesity Type 1,2 and 3

3.1 Family history with overweight

```
155 row(s) in 0.4210 seconds
hbase(main):029:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','family_history_with_overweight',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

344 row(s) in 1.6360 seconds
hbase(main):003:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','family_history_with_overweight',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}

296 row(s) in 1.2840 seconds
hbase(main):004:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','family_history_with_overweight',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

324 row(s) in 0.8270 seconds
hbase(main):005:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','family_history_with_overweight',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}
```

3.2 Consumption frequency of high caloric food

```
208 row(s) in 0.8930 seconds
hbase(main):030:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FAVC',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

323 row(s) in 0.8780 seconds
hbase(main):013:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FAVC',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}

290 row(s) in 0.6450 seconds
hbase(main):014:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FAVC',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

340 row(s) in 0.6170 seconds
hbase(main):015:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FAVC',=,'binary:yes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}
```

3.3 Consumption frequency of vegetables

```
155 row(s) in 0.6040 seconds
hbase(main):051:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

256 row(s) in 0.6870 seconds
hbase(main):016:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}

138 row(s) in 0.3200 seconds
hbase(main):021:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

138 row(s) in 0.4770 seconds
hbase(main):022:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:3') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

324 row(s) in 0.5980 seconds
hbase(main):023:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','FCVC',=,'binary:3') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}
```

3.4 Consumption of food between meals

```
159 row(s) in 0.6720 seconds
hbase(main):038:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CAEC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

323 row(s) in 0.4870 seconds
hbase(main):024:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CAEC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}

293 row(s) in 0.8670 seconds
hbase(main):025:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CAEC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

338 row(s) in 0.7430 seconds
hbase(main):026:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CAEC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}
```

3.5 Does the person smoke?

```
274 row(s) in 2.2200 seconds
hbase(main):015:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SMOKE',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

345 row(s) in 0.7620 seconds
hbase(main):027:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SMOKE',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}

282 row(s) in 0.7060 seconds
hbase(main):028:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SMOKE',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

323 row(s) in 0.9760 seconds
hbase(main):029:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SMOKE',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}
```

3.6 Water consumption amount

```
164 row(s) in 0.6350 seconds
hbase(main):044:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CH20',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

149 row(s) in 0.2150 seconds
hbase(main):034:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CH20',=,'binary:3') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}

177 row(s) in 0.5600 seconds
hbase(main):038:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CH20',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

173 row(s) in 0.4290 seconds
hbase(main):042:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CH20',=,'binary:2') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}
```

3.7 Calories intake monitoring

```
257 row(s) in 0.8470 seconds
hbase(main):022:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SCC',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

349 row(s) in 1.0100 seconds
hbase(main):043:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SCC',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}

296 row(s) in 0.7230 seconds
hbase(main):044:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SCC',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

324 row(s) in 0.7800 seconds
hbase(main):045:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','SCC',=,'binary:no') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}
```

3.8 Consumption frequency of alcohol

```
161 row(s) in 0.6990 seconds
hbase(main):036:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CALC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']}

323 row(s) in 0.8540 seconds
hbase(main):047:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CALC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']}

224 row(s) in 0.7030 seconds
hbase(main):048:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CALC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']}

172 row(s) in 0.7250 seconds
hbase(main):049:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','CALC',=,'binary:Sometimes') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']}
```

3.9 Mode of transportation used

```
200 row(s) in 0.4290 seconds
hbase(main):059:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','MTRANS',=,'binary:Public_Transportation') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']]

236 row(s) in 0.6120 seconds
hbase(main):050:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','MTRANS',=,'binary:Public_Transportation') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']]

200 row(s) in 1.1120 seconds
hbase(main):051:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','MTRANS',=,'binary:Public_Transportation') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']]

323 row(s) in 0.6430 seconds
hbase(main):052:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','MTRANS',=,'binary:Public_Transportation') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']]
```

3.10 Time spent on technology devices

```
129 row(s) in 0.2750 seconds
hbase(main):014:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','TUE',<,'binary:0.66') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Normal_Weight')",COLUMNS=>['cf']]

163 row(s) in 0.5110 seconds
hbase(main):055:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','TUE',<,'binary:0.66') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_III')",COLUMNS=>['cf']]

192 row(s) in 0.5060 seconds
hbase(main):056:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','TUE',<,'binary:0.66') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_II')",COLUMNS=>['cf']]

191 row(s) in 0.3240 seconds
hbase(main):057:0> scan 'obesity',{ FILTER => "SingleColumnValueFilter('cf','TUE',<,'binary:0.66') AND SingleColumnValueFilter('cf','NObeyesdad',=,'binary:Obesity_Type_I')",COLUMNS=>['cf']]
```

From the querying results above, it is observed that majority of the individuals in all 4 categories:

- have family members that suffered or suffers with overweight
- consume food with high calorific value
- sometimes take food in between their meals
- do not smoke
- do not monitor their calories consumption on a daily basis
- sometimes consume alcohol
- usually use public transport as a mode of transportation
- spend 0 to 2 hours in technological devices such as smartphone, television, computer, etc.

There were 2 attributes that differentiated the individuals with Obesity from those who are Normal weight:

- Consumption of water

Majority of the individuals with Obesity Type 3 consume water more than 2 litres while the Normal Weight, Obesity Type 1 and Obesity Type 2 consume between 1-2 litres of water on a daily basis.

- Consumption frequency of vegetables

Normal Weight and Obesity Type 1 consume vegetables sometimes. There is an equal number of individuals in Obesity Type 2 who sometimes and always consume vegetables. Majority of the individuals in Obesity Type 3 always consume vegetables with their meals.