Project Hive

1. Question 1
   a. **Creating a table and loading data into it.**
      create external table if not exists ayush_proj(time int,sex string,cause string,value int)
      row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
      stored as textfile
      location '/user/f201605673498/project_hive'
      tblproperties('skip.header.line.count'='1');
   b. **Creating a partition table.**
      create external table ayush_part(sex string,cause string,value int)
      partitioned by (time string)
      row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
      stored as textfile;
   c. **Inserting data dynamically into the partition table.**
      i. insert overwrite table ayush_part partition (time=2001) select sex,cause,value from ayush_proj where time=2001;
      ii. insert overwrite table ayush_part partition (time=2002) select sex,cause,value from ayush_proj where time=2002;
      iii. insert overwrite table ayush_part partition (time=2003) select sex,cause,value from ayush_proj where time=2003;
      iv. insert overwrite table ayush_part partition (time=2004) select sex,cause,value from ayush_proj where time=2004;
      v. insert overwrite table ayush_part partition (time=2005) select sex,cause,value from ayush_proj where time=2005;
      vi. insert overwrite table ayush_part partition (time=2006) select sex,cause,value from ayush_proj where time=2006;
      vii. insert overwrite table ayush_part partition (time=2007) select sex,cause,value from ayush_proj where time=2007;
      viii. insert overwrite table ayush_part partition (time=2008) select sex,cause,value from ayush_proj where time=2008;
   d. **Display partition.**
      i. show partitions ayush_part;

```
OK
time=2001
time=2002
time=2003
time=2004
time=2005
time=2006
time=2007
time=2008
Time taken: 0.57 seconds, Fetched: 8 row(s)
```

2. Question 2

a. Datapoint per year

    i.   select time, count(*) from ayush_part group by time;

```
OK
2001    127
2002    127
2003    127
2004    127
2005    127
2006    127
2007    127
2008    127
Time taken: 22.03 seconds, Fetched: 8 row(s)
```

3. Question 3

a. Different type of cause of death

i. select distinct cause from ayush_part;

```
OK
Accidental poisoning by and exposure to noxious substances
Accidents
All causes of death (A00-Y89) excluding S00-T98
Assault
Asthma and status asthmaticus
Cerebrovascular diseases
Certain conditions originating in the perinatal period (P00-P96)
Certain infectious and parasitic diseases (A00-B99)
Chronic liver disease
Chronic lower respiratory diseases
Congenital malformations of the circulatory system
Congenital malformations of the nervous system
Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)
Diabetes mellitus
Diseases of kidney and ureter
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
Diseases of the circulatory system (I00-I99)
Diseases of the digestive system (K00-K93)
Diseases of the genitourinary system  (N00-N99)
Diseases of the musculoskeletal system and connective tissue (M00-M99)
Diseases of the nervous system and the sense organs (G00-H95)
Diseases of the respiratory system (J00-J99)
Diseases of the skin and subcutaneous tissue (L00-L99)
Drug dependence, toxicomania (F11-F16, F18-F19)
Endocrine, nutritional and metabolic diseases (E00-E90)
Event of undetermined intent
External causes of morbidity and mortality (V01-Y89)
Falls
Human immunodeficiency virus [HIV] disease
Ill-defined and unknown causes of mortality
Influenza
Intentional self-harm
Ischaemic heart diseases
```

```
Malignant melanoma of skin
Malignant neoplasm of bladder
Malignant neoplasm of breast
Malignant neoplasm of cervix uteri
Malignant neoplasm of colon
Malignant neoplasm of kidney, except renal pelvis
Malignant neoplasm of larynx,  trachea, bronchus and lung
Malignant neoplasm of lip, oral cavity, pharynx
Malignant neoplasm of liver and intrahepatic bile ducts
Malignant neoplasm of oesophagus
Malignant neoplasm of other parts of uterus
Malignant neoplasm of ovary
Malignant neoplasm of pancreas
Malignant neoplasm of prostate
Malignant neoplasm of rectosigmoid junction, rectum, anus and anal canal
Malignant neoplasm of stomach
Malignant neoplasms (C00-C97)
Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue
Meningitis
Meningococcal infection
Mental and behavioural disorders (F00-F99)
Mental and behavioural disorders due to use of alcohol
Neoplasms
Other heart diseases (I30-I33, I39-I52)
Pneumonia
Pregnancy, childbirth and the puerperium (O00-O99)
Rheumatoid arthritis and arthrosis (M05-M06,M15-M19)
Sudden infant death syndrome
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)
Transport accidents (V01-V99)
Tuberculosis
Ulcer of stomach, duodenum and jejunum
Viral hepatitis
Time taken: 21.927 seconds, Fetched: 66 row(s)
```

4. Question 4
   a. **Total number of deaths.**
      i. select sum(value) from ayush_part;

```
OK
1.2745563E7
Time taken: 41.254 seconds, Fetched: 1 row(s)
```

   b. **Year having maximum death.**
      i. Select time,sum(value) as cnt from ayush_part group by time order by cnt desc limit 1;

```
OK
2003    1654533.0
Time taken: 64.827 seconds, Fetched: 1 row(s)
```

5. Question 5

a. Top 5 cause of death.

i. select cause,sum(value) as cnt from ayush_part group by cause order by cnt desc limit 5;

```
OK
All causes of death (A00-Y89) excluding S00-T98 4307271.0
Neoplasms        1253797.0
Diseases of the circulatory system (I00-I99)    1236278.0
Malignant neoplasms (C00-C97)    1201612.0
Other heart diseases (I30-I33, I39-I52) 361927.0
Time taken: 63.521 seconds, Fetched: 5 row(s)
```

b. Top 5 cause of death in Males.

i. select cause,sum(value) as cnt from ayush_part where sex rlike "Males" group by cause order by cnt desc limit 5;

```
OK
All causes of death (A00-Y89) excluding S00-T98 2208992.0
Neoplasms        743172.0
Malignant neoplasms (C00-C97)    716802.0
Diseases of the circulatory system (I00-I99)    576494.0
External causes of morbidity and mortality (V01-Y89)    190599.0
Time taken: 47.567 seconds, Fetched: 5 row(s)
```

c. Top 5 cause of death in Females.

i. select cause,sum(value) as cnt from ayush_part where sex rlike "Females" group by cause order by cnt desc limit 5;

```
OK
All causes of death (A00-Y89) excluding S00-T98 2098279.0
Diseases of the circulatory system (I00-I99)    659784.0
Neoplasms        510625.0
Malignant neoplasms (C00-C97)    484810.0
Other heart diseases (I30-I33, I39-I52) 206511.0
Time taken: 58.754 seconds, Fetched: 5 row(s)
```