Create table if not exists data\_table(

AAGE int,

ACLSWKR string,

ADTIND int,

ADTOCC int,

AHGA string,

AHRSPAY int,

AHSCOL string,

AMARITL string,

AMJIND string,

AMJOCC string,

ARACE string,

AREORGN string,

ASEX string,

AUNMEM string,

AUNTYPE string,

AWKSTAT string,

CAPGAIN int,

CAPLOSS int,

DIVVAL int,

FILESTAT string,

GRINREG string,

GRINST string,

HHDFMX string,

HHDREL string,

MARSUPWT double,

MIGMTR1 string,

MIGMTR3 string,

MIGMTR4 string,

MIGSAME string,

MIGSUN string,

NOEMP int,

PARENT string,

PEFNTVTY string,

PEMNTVTY string,

PENATVTY string,

PRCITSHP string,

PTOTVAL int,

SEOTR string,

TAXINC int,

VETQVA int,

VETYN int,

income int)

row format delimited

fields terminated by ‘,’

lines terminated by ‘\n’;

Count the number of rows with missing values in columns 26-28 (inclusive).

|  |  |  |
| --- | --- | --- |
|  | string | migration code-change in msa MIGMTR1 |
|  | string | migration code-change in reg MIGMTR3 |
|  | string | migration code-move within reg MIGMTR4 |

SELECT 列名称 FROM 表名称 WHERE 列 运算符 值

Select \* from data\_table where MIGMTR1 = ‘?’ and MIGMTR3 = ‘?’ and MIGMTR4 = ‘?’;

Select count(\*) from data\_table where MIGMTR1 = ‘?’ and MIGMTR3 = ‘?’ and MIGMTR4 = ‘?’;

Find all rows with ‘female’ as gender, and write them into a CSV file.

INSERT OVERWRITE LOCAL DIRECTORY '/tmp/pods\_comedy' ROW FORMAT  
DELIMITED FIELDS TERMINATED BY ','  
SELECT ASEX FROM data\_table WHERE keywords RLIKE 'female';

Calculate the minimum, maximum, mean, and median age. Look up the use of Hive’s  
built-in statistical functions.

select avg(salary) from employee group by age;

SELECT AVG(OrderPrice) AS OrderAverage FROM Orders

Select avg(age) as average\_age from data\_table;

Select min(age) as average\_age from data\_table;

Select max(age) as average\_age from data\_table;

select percentile(cast(age as BIGINT), 0.5) from data\_table;

Find out the distinct nominal values of the ‘education’ column, and a count of each.

SELECT education as educationxx, COUNT(\*) as num FROM data\_table GROUP BY education