Create Table

CREATE EXTERNAL TABLE finalproject_table(Housing INT, Age INT, Sex INT, Employment INT, Monthly_income FLOAT, Credit INT)ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/yl5287/finalproject/input';

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> CREATE EXTERNAL TABLE finalproject_table(Housing INT, Age INT, Sex INT, Employment INT, Monthly_income FLOAT, Credit INT)ROW FORMAT DELIMITED FIEL
DS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/yl5287/finalproject/input';
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select * from finalproject_table
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> ;
```

finalproject_table.housing	finalproject_table.age	finalproject_table.sex	finalproject_table.employment	finalproject_table.monthly_income	finalproject_table.credit
1	3	1	1	0.023948882	1
1	3	1	1	0.019320725	1
1	7	1	1	0.0022773477	1
1	6	1	1	0.013223309	1
1	4	0	1	0.0057301004	0
1	4	1	1	0.01013787	1
1	7	1	1	0.0018365707	1
1	5	0	1	0.0052893236	1
1	3	0	1	0.007933985	0
1	3	0	1	0.015427194	1
1	5	1	1	0.0069789686	0
1	4	1	1	0.008154374	1
1	3	1	1	0.010725573	1
1	3	1	1	0.011386738	1
1	3	1	1	0.01946765	1
1	5	0	1	0.004113918	1
1	3	1	1	0.019320725	1
1	3	1	1	0.020128815	1
1	5	1	1	0.03166248	0
1	3	1	1	0.0070524313	1
1	6	0	1	0.0049954723	1
1	4	1	1	0.010358259	1
1	4	1	1	0.0012488681	0
1	5	0	1	0.0069789686	0

Credit Good Person Attributes Distribution:

Sex:

select sex, count(sex) as gender_count from finalproject_table where credit = 1 group by sex;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select sex, count(sex) as gender_count from finalproject_table where credit = 1 group by sex;
+------+--------------+
| sex  | gender_count |
+------+--------------+
| 0    | 81308        |
| 1    | 124688       |
+------+--------------+
```

Housing:

select housing, count(housing) as housing_count from finalproject_table where credit = 1 group by housing;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select housing, count(housing) as housing_count from finalproject_table where credit = 1 group by housing;
+---------+---------------+
| housing | housing_count |
+---------+---------------+
| 0       | 6500          |
| 1       | 199496        |
+---------+---------------+
```

Age:

select age, count(age) as age_group from finalproject_table where credit = 1 group by age;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select age, count(age) as age_group from finalproject_table where credit = 1 group by age;
+------+-----------+
| age  | age_group |
+------+-----------+
| 1    | 9000      |
| 2    | 45923     |
| 3    | 52643     |
| 4    | 43203     |
| 5    | 31806     |
| 6    | 17260     |
| 7    | 6161      |
+------+-----------+
```

Employment:

select employment, count(employment) as jobs from finalproject_table where credit = 1 group by employment;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select employment, count(employment) as jobs from finalproject_table where credit = 1 group by employment;
+------------+--------+
| employment | jobs   |
+------------+--------+
| 0          | 19357  |
| 1          | 186639 |
+------------+--------+
```

Multiples of Avg Income:

select round(monthly_income,1) as round_income, count(round(monthly_income,1)) as income_group from finalproject_table where credit = 1 group by round(monthly_income,1);

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select round(monthly_income,1) as round_income, count(round(monthly_income,1)) as income_group from finalproject_table where credit = 1 group by r
ound(monthly_income,1);
+--------------+--------------+
| round_income | income_group |
+--------------+--------------+
| 0.0          | 168361       |
| 0.1          | 977          |
| 0.2          | 607          |
| 0.3          | 2323         |
| 0.4          | 9647         |
| 0.5          | 5051         |
| 0.6          | 3898         |
| 0.7          | 2894         |
| 0.8          | 2128         |
| 0.9          | 1673         |
| 1.0          | 1183         |
| 1.1          | 1217         |
| 1.2          | 710          |
| 1.3          | 510          |
| 1.4          | 767          |
| 1.5          | 449          |
| 1.6          | 408          |
| 1.7          | 476          |
| 1.8          | 279          |
| 1.9          | 189          |
| 2.0          | 263          |
| 2.1          | 168          |
| 2.2          | 159          |
| 2.3          | 219          |
| 2.4          | 120          |
| 2.5          | 97           |
| 2.6          | 97           |
| 2.7          | 94           |
| 2.8          | 101          |
| 2.9          | 66           |
| 3.0          | 54           |
```

Credit Bad Person Attributes Distribution:
Sex:
select sex, count(sex) as gender_count from finalproject_table where credit = 0 group by sex;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select sex, count(sex) as gender_count from finalproject_table where credit = 0 group by sex;
+------+---------------+
| sex  | gender_count  |
+------+---------------+
| 0    | 16698         |
| 1    | 20182         |
+------+---------------+
```

Housing:
select housing, count(housing) as housing_count from finalproject_table where credit = 0 group by housing;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select housing, count(housing) as housing_count from finalproject_table where credit = 0 group by housing;
+----------+----------------+
| housing  | housing_count  |
+----------+----------------+
| 0        | 3582           |
| 1        | 33298          |
+----------+----------------+
```
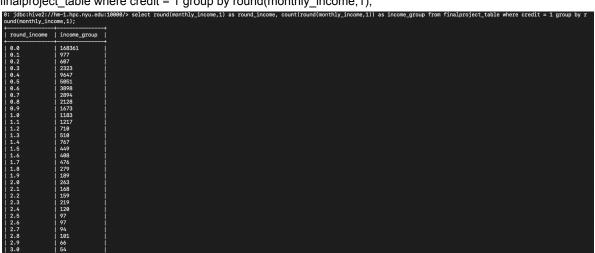
Age:
select age, count(age) as age_group from finalproject_table where credit = 0 group by age;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/>  select age, count(age) as age_group from finalproject_table where credit = 0 group by age;
+------+------------+
| age  | age_group  |
+------+------------+
| 1    | 2385       |
| 2    | 6969       |
| 3    | 8916       |
| 4    | 7968       |
| 5    | 5769       |
| 6    | 3173       |
| 7    | 1700       |
+------+------------+
```

Employment:
select employment, count(employment) as jobs from finalproject_table where credit = 0 group by employment;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select employment, count(employment) as jobs from finalproject_table where credit = 0 group by employment;
+-------------+-------+
| employment  | jobs  |
+-------------+-------+
| 0           | 6542  |
| 1           | 30338 |
+-------------+-------+
```

Multiples of Avg Income:
select round(monthly_income,1) as round_income, count(round(monthly_income,1)) as income_group from
finalproject_table where credit = 0 group by round(monthly_income,1);

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select round(monthly_income,1) as round_income, count(round(monthly_income,1)) as income_group from finalproject_table where credit = 0 group by r
ound(monthly_income,1);
+---------------+---------------+
| round_income  | income_group  |
+---------------+---------------+
| -0.4          | 1             |
| -0.3          | 1             |
| -0.1          | 4             |
| 0.0           | 23676         |
| 0.1           | 220           |
| 0.2           | 228           |
| 0.3           | 822           |
| 0.4           | 3232          |
| 0.5           | 1879          |
| 0.6           | 1471          |
| 0.7           | 1071          |
| 0.8           | 781           |
| 0.9           | 610           |
| 1.0           | 459           |
| 1.1           | 456           |
| 1.2           | 239           |
| 1.3           | 154           |
| 1.4           | 308           |
| 1.5           | 144           |
| 1.6           | 151           |
| 1.7           | 178           |
| 1.8           | 87            |
| 1.9           | 72            |
| 2.0           | 90            |
| 2.1           | 51            |
| 2.2           | 37            |
| 2.3           | 78            |
| 2.4           | 28            |
| 2.5           | 36            |
| 2.6           | 27            |
| 2.7           | 27            |
| 2.8           | 38            |
| 2.9           | 21            |
```

People with negative income:

select * from finalproject_table where monthly_income < 0;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select * from finalproject_table where monthly_income < 0;

| finalproject_table.housing | finalproject_table.age | finalproject_table.sex | finalproject_table.employment | finalproject_table.monthly_income | finalproject_table.credit |
| 1                          | 5                      | 0                      | 0                             | -5.142398E-4                      | 0                         |
| 1                          | 5                      | 1                      | 0                             | -0.0023508104                     | 0                         |
| 1                          | 6                      | 0                      | 0                             | -0.002571199                      | 0                         |
| 1                          | 4                      | 1                      | 0                             | -3.6731415E-4                     | 0                         |
| 1                          | 4                      | 1                      | 0                             | -0.01065211                       | 0                         |
| 1                          | 4                      | 1                      | 0                             | -0.007933985                      | 0                         |
| 1                          | 4                      | 1                      | 0                             | -7.346283E-4                      | 0                         |
| 1                          | 4                      | 1                      | 0                             | -0.029238205                      | 0                         |
| 1                          | 3                      | 1                      | 0                             | -2.2038848E-4                     | 0                         |
| 1                          | 3                      | 1                      | 0                             | -4.4077696E-4                     | 0                         |
| 1                          | 3                      | 1                      | 0                             | -4.4077696E-4                     | 0                         |
| 1                          | 6                      | 1                      | 0                             | -0.0033058273                     | 0                         |
| 1                          | 5                      | 1                      | 0                             | -0.004848547                      | 0                         |
| 1                          | 5                      | 1                      | 0                             | -2.2038848E-4                     | 0                         |
| 1                          | 3                      | 1                      | 0                             | -0.029311668                      | 0                         |
| 1                          | 5                      | 1                      | 0                             | -0.0049220095                     | 0                         |
| 1                          | 4                      | 0                      | 0                             | -1.4692565E-4                     | 0                         |
| 1                          | 3                      | 1                      | 0                             | -0.024977362                      | 0                         |
| 1                          | 6                      | 0                      | 0                             | -0.0052893236                     | 0                         |
| 1                          | 6                      | 1                      | 0                             | -0.0035996786                     | 0                         |
| 1                          | 6                      | 1                      | 0                             | -0.0019100335                     | 0                         |
| 1                          | 3                      | 1                      | 0                             | -0.004628158                      | 0                         |
| 1                          | 6                      | 1                      | 0                             | -0.001689645                      | 0                         |
| 1                          | 4                      | 0                      | 0                             | -8.0809114E-4                     | 0                         |
| 1                          | 5                      | 0                      | 0                             | -0.0018365707                     | 0                         |
| 1                          | 3                      | 1                      | 0                             | -5.877026E-4                      | 0                         |
| 1                          | 5                      | 1                      | 0                             | -0.34909537                       | 0                         |
| 1                          | 4                      | 1                      | 0                             | -0.019908426                      | 0                         |
| 1                          | 2                      | 1                      | 0                             | -0.0011019424                     | 0                         |
| 1                          | 6                      | 1                      | 0                             | -0.012488681                      | 0                         |
| 1                          | 7                      | 1                      | 0                             | -4.4077696E-4                     | 0                         |
| 1                          | 6                      | 1                      | 0                             | -0.012488681                      | 0                         |
| 1                          | 3                      | 1                      | 0                             | -0.016749525                      | 0                         |
| 1                          | 6                      | 1                      | 0                             | -0.016088359                      | 0                         |
```

Count the number of good and bad credit for male who is employed and own housing:

select credit, count(credit) from finalproject_table where (housing = 0 and sex = 0 and employment = 0) group by credit;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select credit, count(credit) from finalproject_table where (housing = 0 and sex = 0 and employment = 0) group by credit;

| credit | _c1  |
| 0      | 539  |
| 1      | 1312 |
```

Count the number of good and bad credit for female who is unemployed with no owned housing:

select credit, count(credit) from finalproject_table where (housing = 1 and sex = 1 and employment = 1) group by credit;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select credit, count(credit) from finalproject_table where (housing = 1 and sex = 1 and employment = 1) group by credit;

| credit | _c1    |
| 0      | 15812  |
| 1      | 113418 |
```

Count the number of good and bad credit for male who is employed:

select credit, count(credit) from finalproject_table where (sex = 1 and employment = 1) group by credit;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select credit, count(credit) from finalproject_table where (sex = 1 and employment = 1) group by credit;

| credit | _c1    |
| 0      | 16970  |
| 1      | 115197 |
```

Count the number of good and bad credit for female who is unemployed:

select credit, count(credit) from finalproject_table where (sex = 0 and employment = 0) group by credit;

```
0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select credit, count(credit) from finalproject_table where (sex = 0 and employment = 0) group by credit;

| credit | _c1  |
| 0      | 3330 |
| 1      | 9866 |
```