

ReadMe.txt colin_wirth_cc_approval.ipynb

colin_wirth_cc_approval.ipynb

+ Cell ▾ ▶ Run all Kernel SQL Attach to Select Connection ▾

🔍 🔄 🛡️ 📄 ↩️

First, let's learn about our dataset, or the people who applied for credit cards...

See 1_chart_SandDance.png

We can learn a few things from this chart, mainly that the most likely group to be approved for a credit card is 30-35 years old.

****Let's group the credit card applicants into bins based on their total income, we to reduce the number of digits in total income so it appears better on visualizations. We do this in two steps:**

Step 1:**

```
[98] 1 alter table dataset
      2 add IncomeReduced decimal(7,3);
      3
```

SQL

Commands completed successfully.

Total execution time: 00:00:00.117

Step 2:

```
▶ 1
   2 update dataset
   3 set IncomeReduced = Total_income / 1000;
```

SQL

(9709 rows affected)

Total execution time: 00:00:00.039

Now Total_income has been reduced by \$1,000.

Now we group applicants into income brackets of \$20,000 and show what percentage from each group were approved for credit cards:

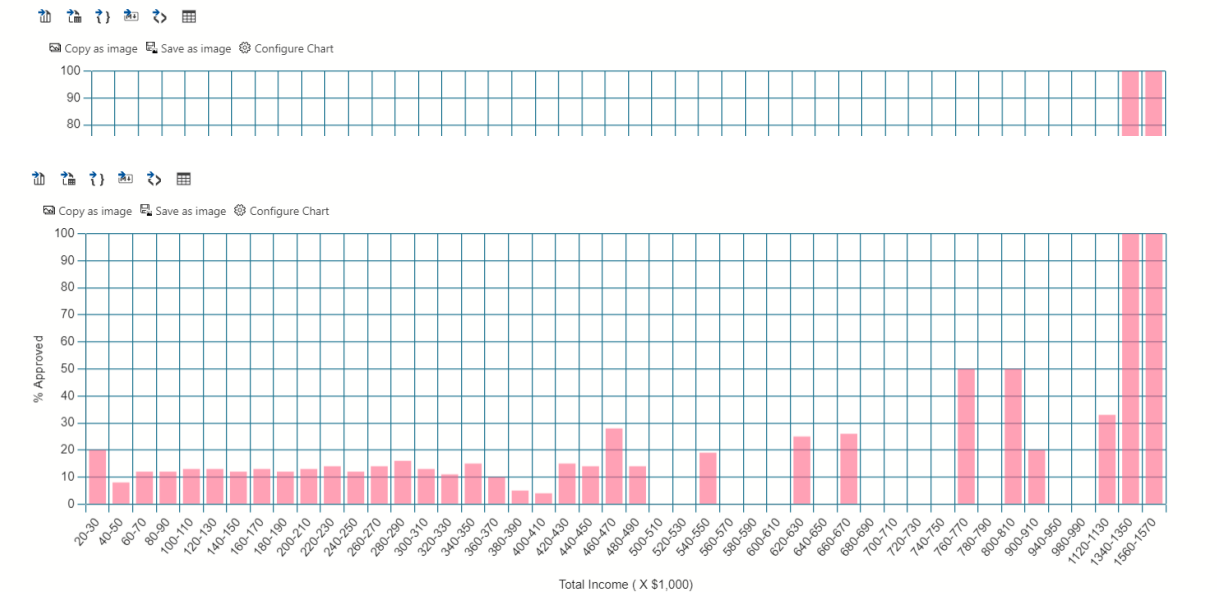
```
[118] 1 -- Income bins ($20k) vs pct approved
      2 with newTable as
      3 (select
      4   ID,
      5   concat(
      6     floor(IncomeReduced / 20) * 20,
      7     '-',
      8     floor(IncomeReduced / 20) * 20 + 10)
      9   as Income_group,
     10   cast([Target] as int) as outcome
     11 from dataset)
     12
```

Now we group applicants into income brackets of \$20,000 and show what percentage from each group were approved for credit cards:

```
1  -- Income bins ($20k) vs pct approved
2  with newTable as
3  (select
4    ID,
5    concat(
6      floor(IncomeReduced / 20) * 20,
7      '-',
8      floor(IncomeReduced / 20) * 20 + 10)
9    as Income_group,
10   cast([Target] as int) as outcome
11  from dataset)
12
13  select
14    Income_group,
15    100 * (sum(outcome)) / (count(ID)) as pct_Approved
16  from newTable
17  group by Income_group
18  order by cast( SUBSTRING(Income_group, 1, CHARINDEX('-', Income_Group) - 1) as int);
```

(46 rows affected)

Total execution time: 00:00:00.025



Let's examine all approved applicants and compare their housing type.

_3_chart_SandDance.png

The majority of approved applicants (88%) live in their owned house or apartment.

And a similar visualization for education type.

4_chart_SandDance.png

The majority of applicants approved for a credit card have a secondary degree. Fewer have any higher education.

Next, let's compare the income type of all approved applicants.

Double-click to edit

```
[13] 1 select
      2     Income_type,
      3     --count(ID) as totalApplied,
      4     --sum(cast([Target] as int)) as totalApproved,
      5     ( 100 * sum(cast([Target] as int)) / count(ID) ) as pctApproved
```

```
[13] 1 select
      2     Income_type,
      3     --count(ID) as totalApplied,
      4     --sum(cast([Target] as int)) as totalApproved,
      5     ( 100 * sum(cast([Target] as int)) / count(ID) ) as pctApproved
      6 from dataset
      7 group by Income_type;
```

SQL

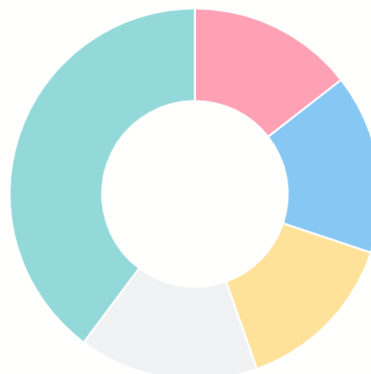
(5 rows affected)

Total execution time: 00:00:00.009



Copy as image Save as image Configure Chart

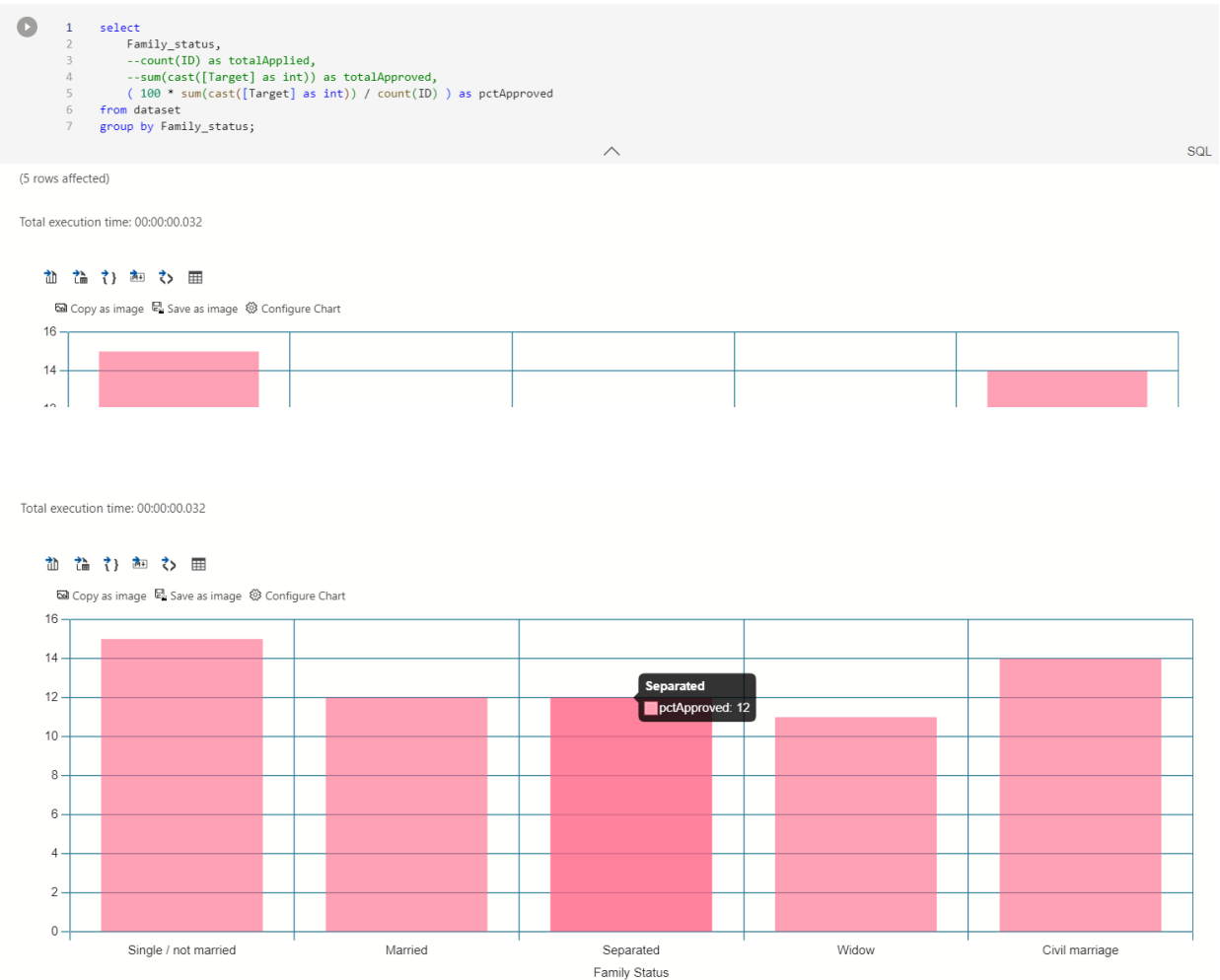
State servant
Commercial associate
Pensioner
Working
Student



5_chart.png

Students are much more likely to be approved (33%) than other applicants. However, only 3 students applied.

There's one more variable we haven't explore, family status:



The most likely group to be approved for a credit card were Single.