Home Credit Default Risk

Angelyn Arthur

Background

Home Credit aims to provide loans for those who have insufficient or non-existent credit histories.
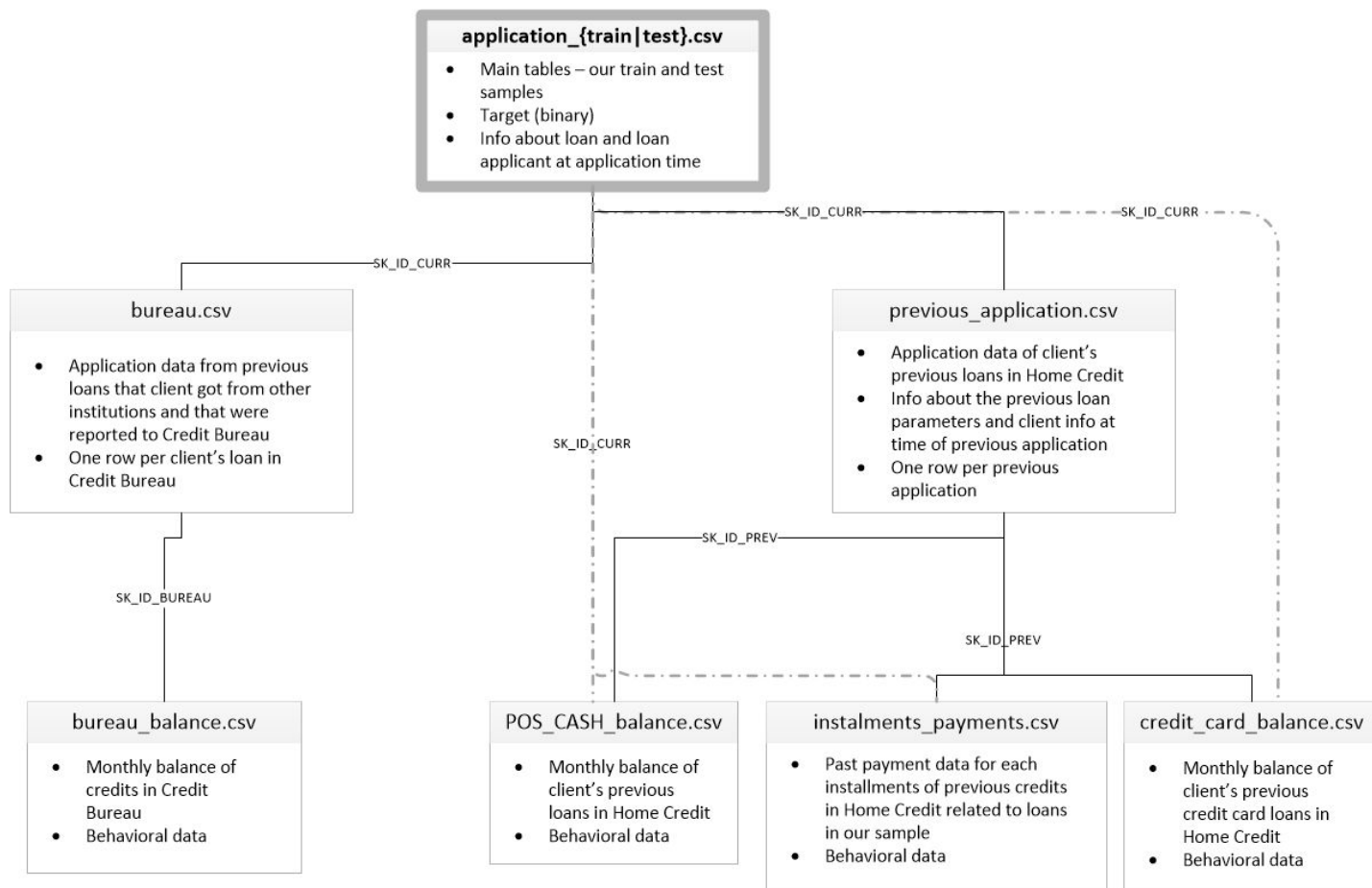
➜   Focused on markets that have underserved consumer finance needs.

Loan type during application time*:*
- **Cash loans** *are typically offered to customers for consumer goods or services.*
- **Revolving loans***, including credit cards, are typically offered to existing customers for their purchases of goods or services up to individual credit limits on a revolving basis.*

Schema

application_{train|test}.csv
- Main tables – our train and test samples
- Target (binary)
- Info about loan and loan applicant at application time

SK_ID_CURR
SK_ID_CURR
SK_ID_CURR
SK_ID_CURR

bureau.csv
- Application data from previous loans that client got from other institutions and that were reported to Credit Bureau
- One row per client's loan in Credit Bureau

previous_application.csv
- Application data of client's previous loans in Home Credit
- Info about the previous loan parameters and client info at time of previous application
- One row per previous application

SK_ID_PREV

SK_ID_BUREAU

SK_ID_PREV

bureau_balance.csv
- Monthly balance of credits in Credit Bureau
- Behavioral data

POS_CASH_balance.csv
- Monthly balance of client's previous loans in Home Credit
- Behavioral data

instalments_payments.csv
- Past payment data for each installments of previous credits in Home Credit related to loans in our sample
- Behavioral data

credit_card_balance.csv
- Monthly balance of client's previous credit card loans in Home Credit
- Behavioral data

Executive Summary

Tools Used: Python, Tableau, Excel
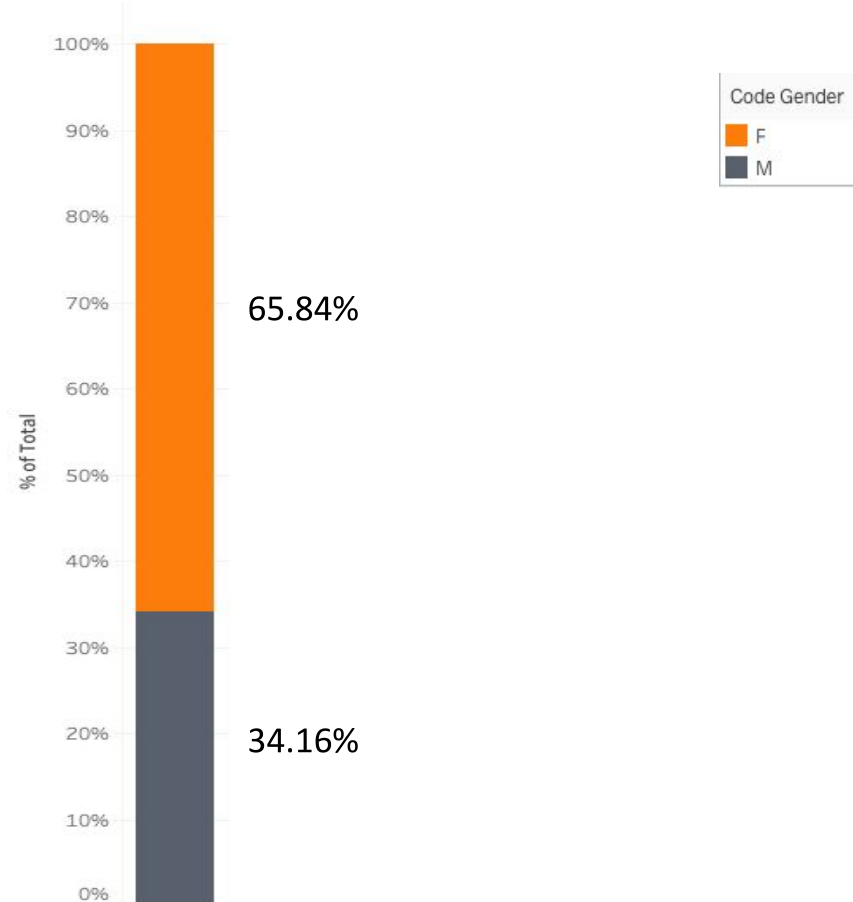Dataset: https://www.kaggle.com/c/home-credit-default-risk/data

Repayment by Demographics

- Gender
- Age
- Education
- Loan Type

- Hypothesis Testing
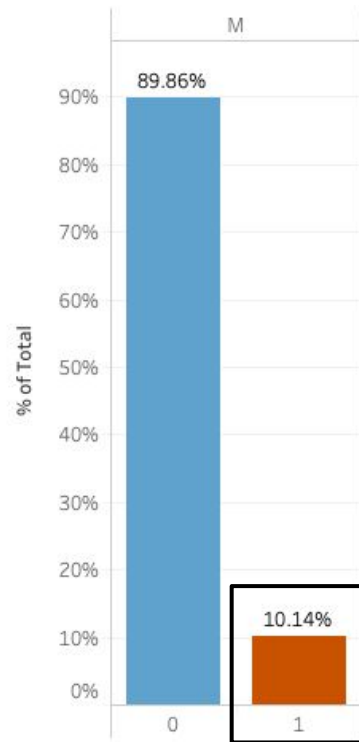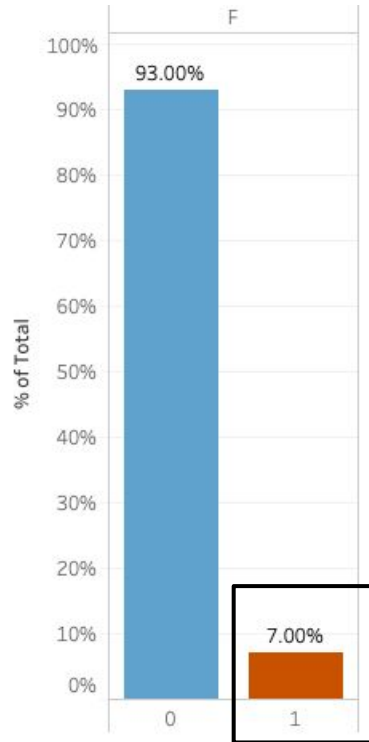- Takeaways
- Next Steps

How do these demographics affect the ability to repay a loan?

Gender

There are nearly twice as many females than males who have a loan.

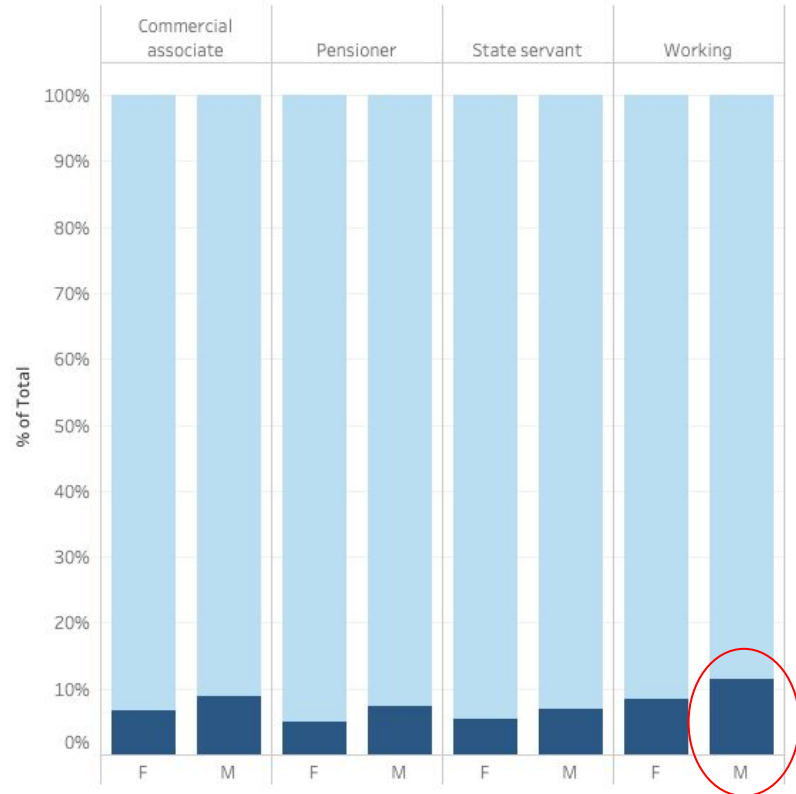Males have a higher chance of not repaying a loan despite being significantly outnumbered.

0 - Repaid
1 - Default

Regardless of income type or family status, across the board, males tend to have a higher chance of defaulting on a loan.
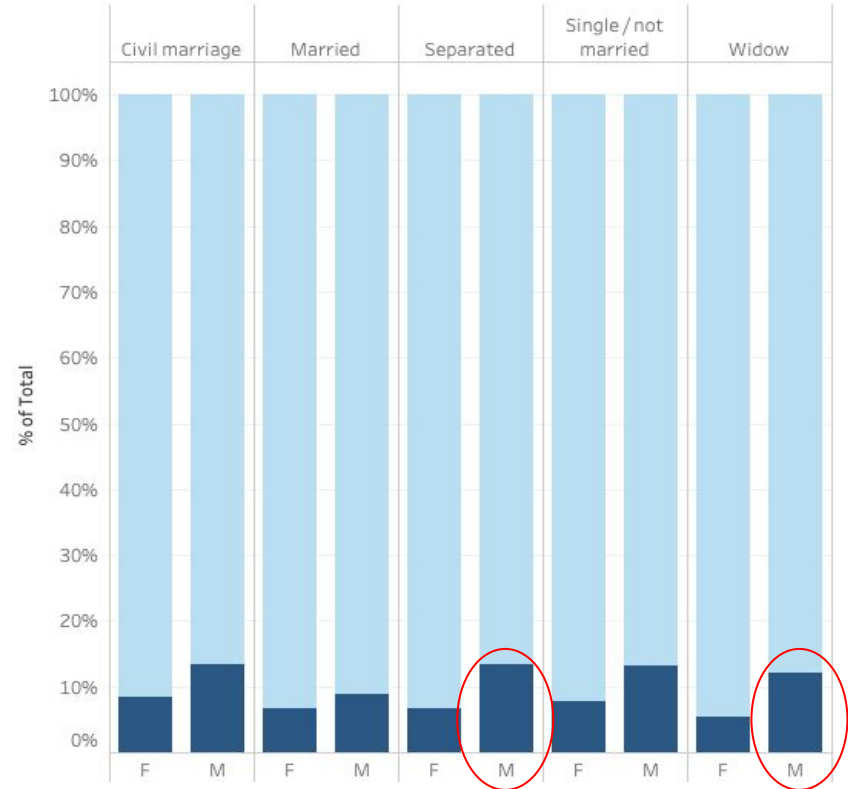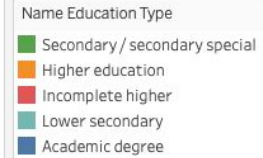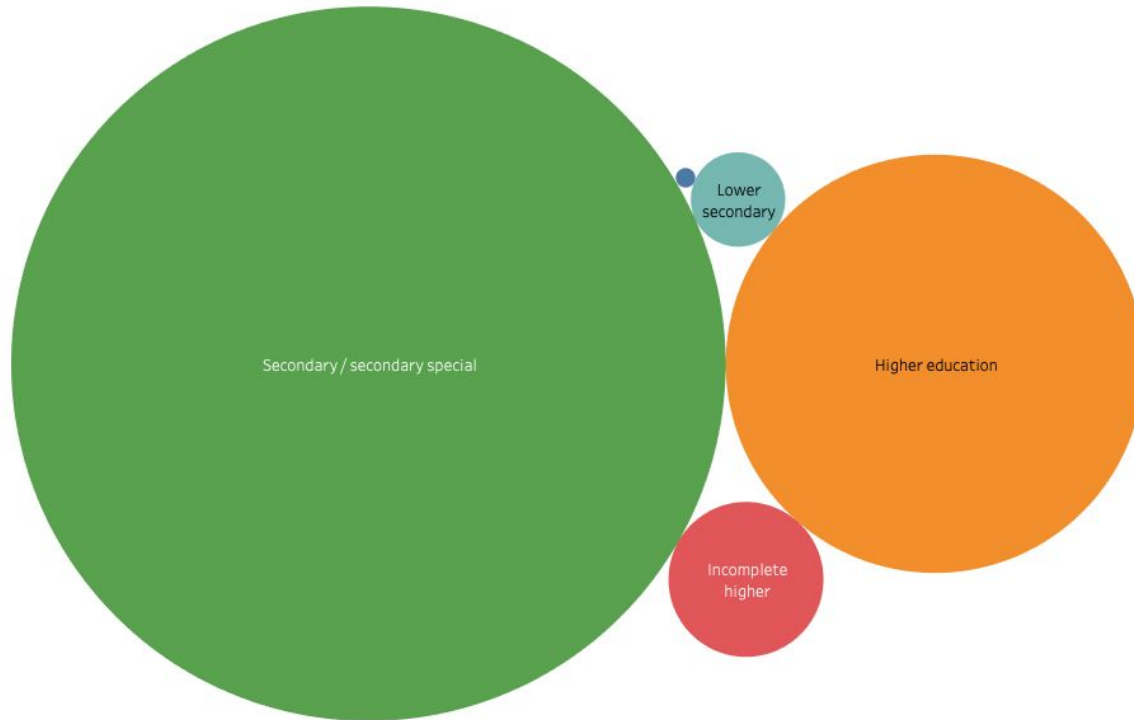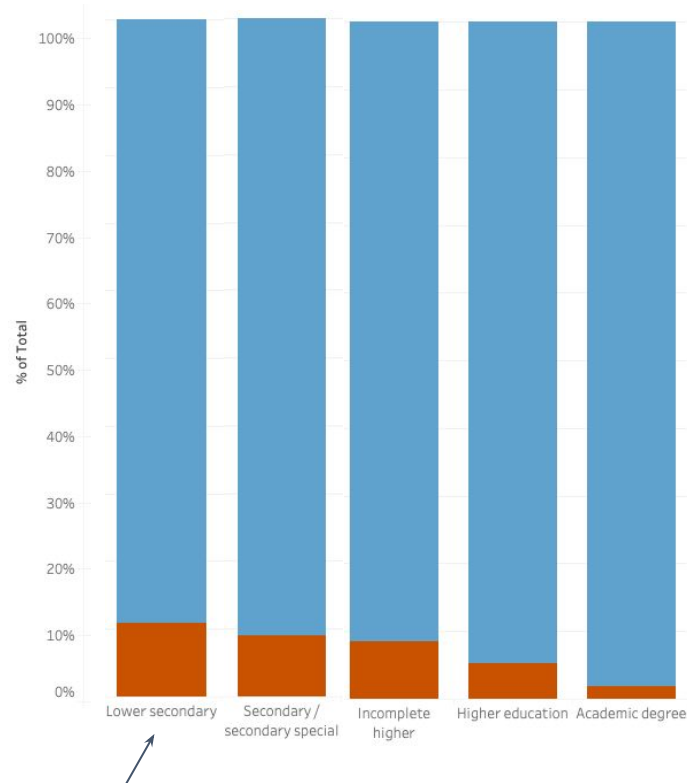
Education

Most clients have either secondary/secondary special or higher education.

Typically, people repay their loan if they have a more advanced education level.

Lower secondary education clients have the highest chance of not repaying their loan despite only accounting for 1.24% of education types.
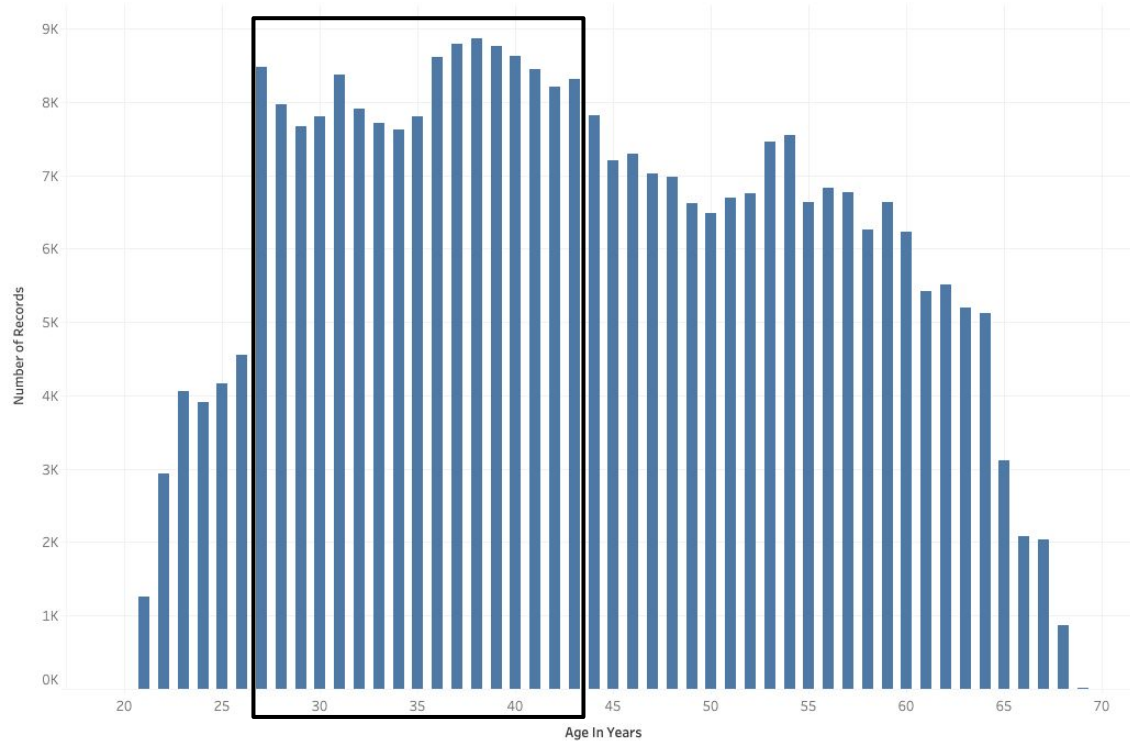

Target
■ 0
■ 1

% of Total

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

Lower secondary   Secondary /   Incomplete   Higher education   Academic degree
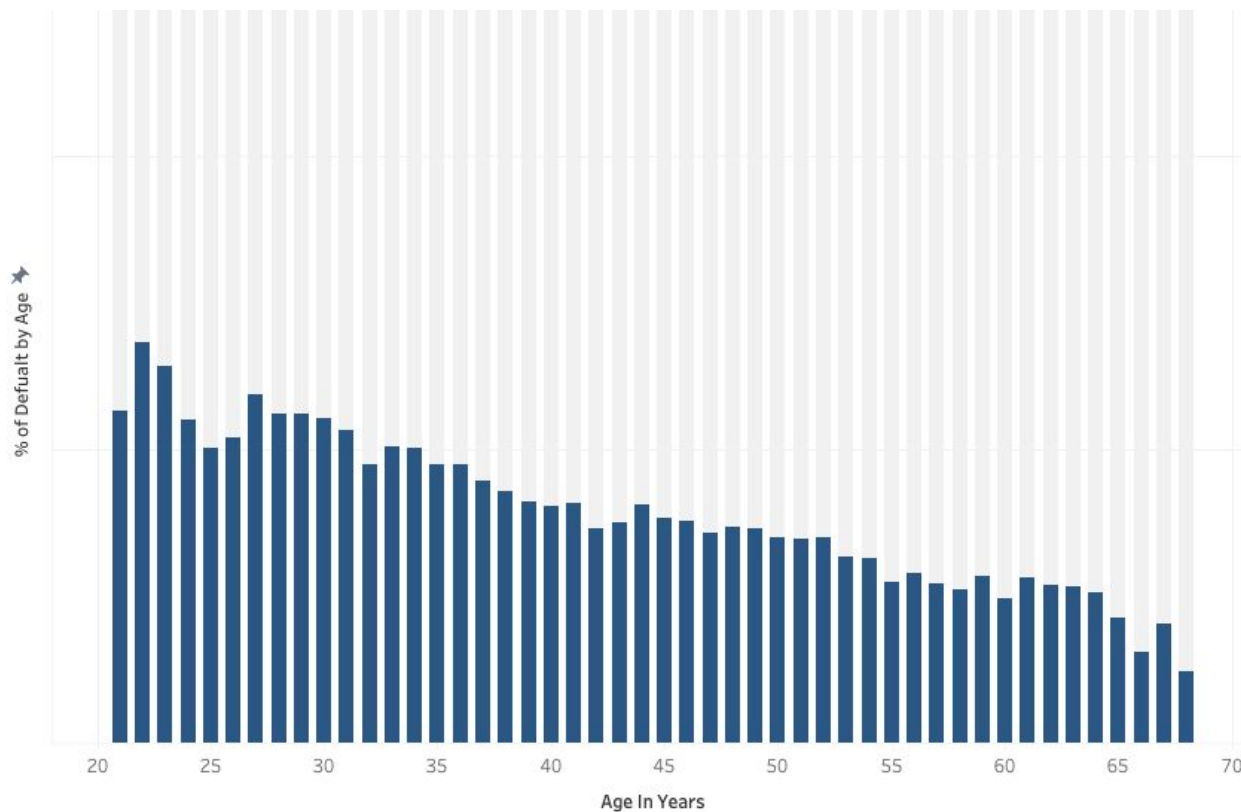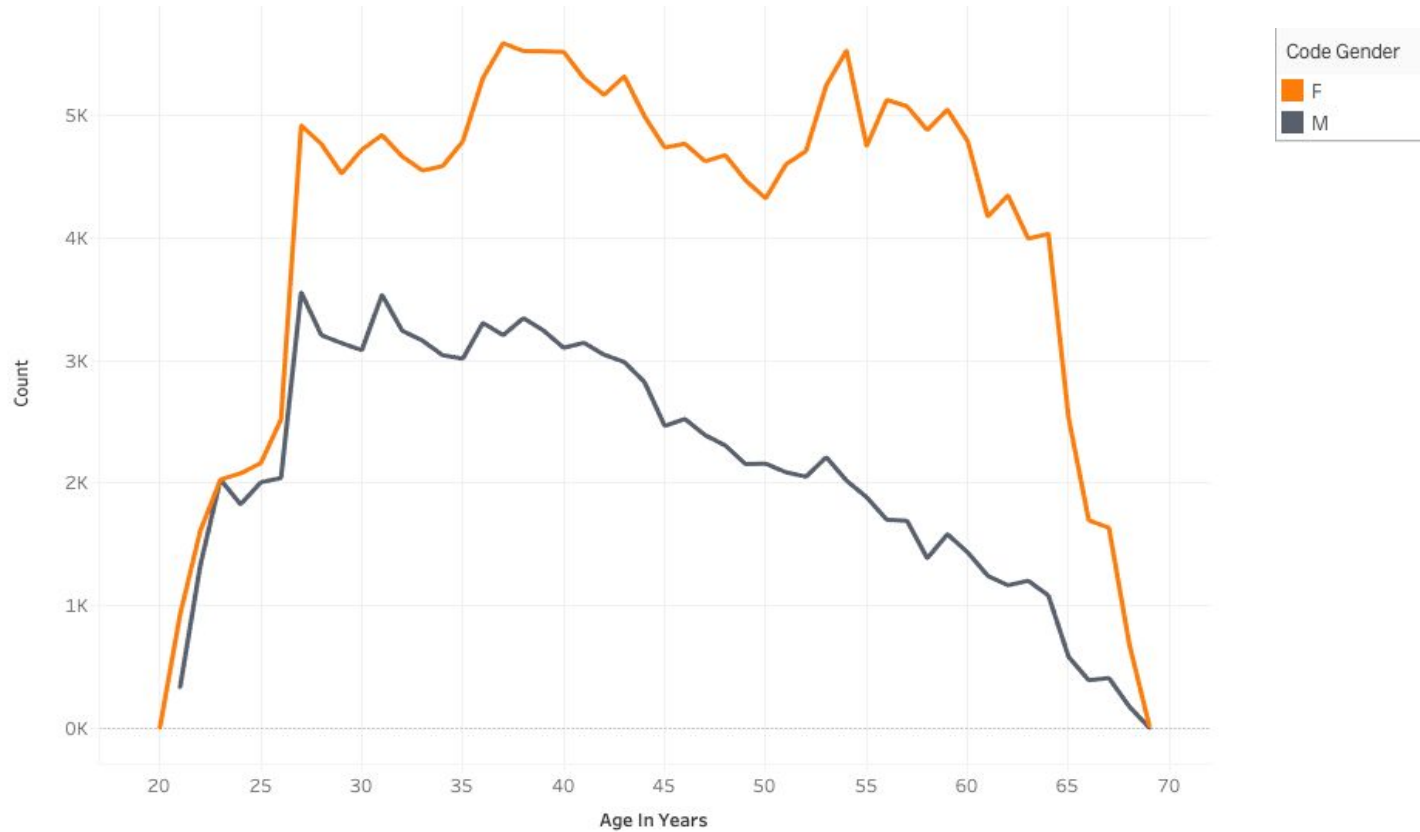secondary special   higher

Age

Higher distribution for loans in ages ranging from the late 20s to mid 40s.

The chance of a client defaulting on a loan increases by 7% from the oldest age bracket to the youngest.
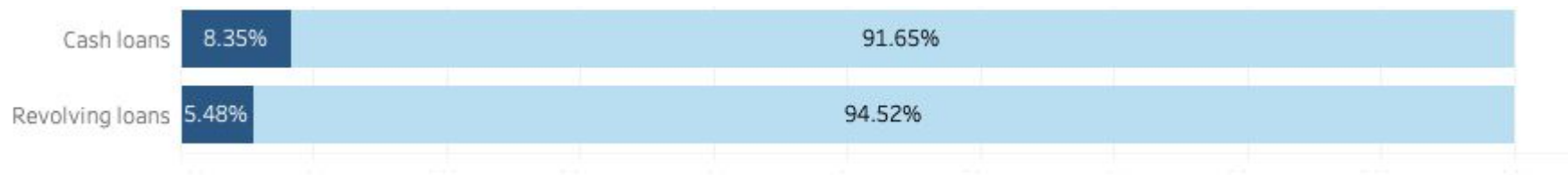
Men tend to skew younger in age and women older.

Loan Type

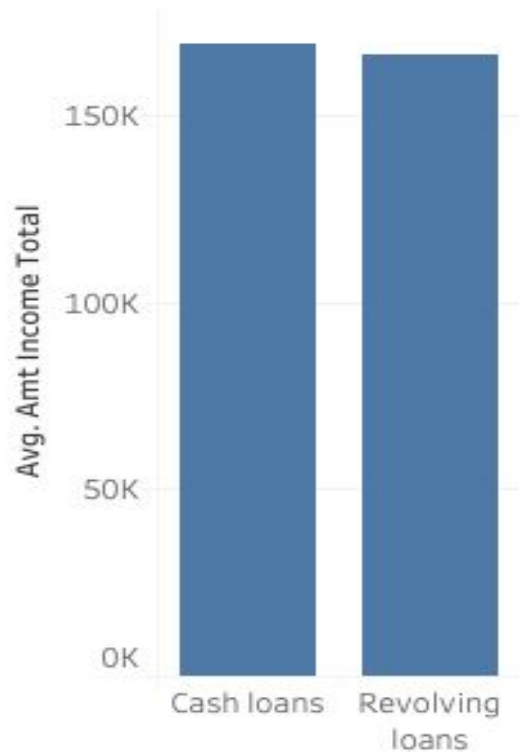# Approximately 90% of all loans are cash loans.

Cash loans	8.35%	91.65%
Revolving loans	5.48%	94.52%
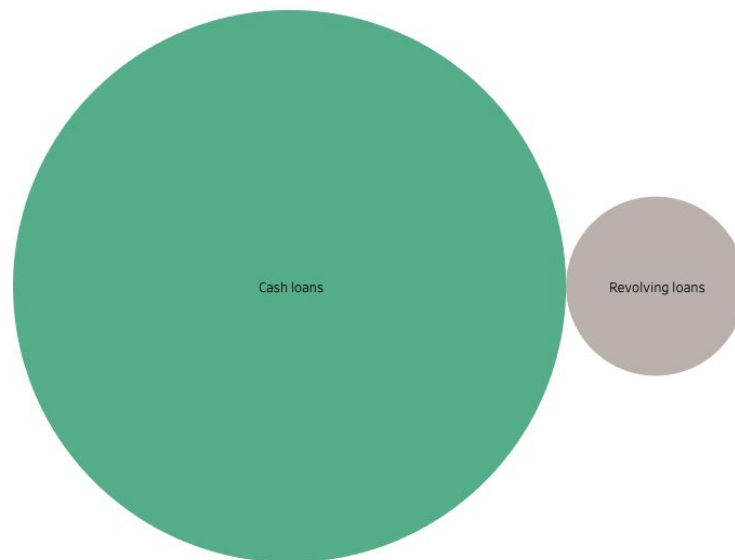
Clients with cash loans have a slightly higher income on average than those with revolving loans.

Hypothesis Testing

Is age statistically significant in regard to gender?

- The p-value (0) is less than the alpha (0.05)
- We reject the null hypothesis and we can see that there is a significant difference between the average female age and the average male age.

Furthermore, we can conclude that males are more prone to default on a loan since the age of a male client skews younger, and there's more of a likelihood to default the younger you are.

Female age vs Male age				
2 population (un-paired)	Female age avg	Male age avg	sample mean (difference)	standard error difference
sample mean	44.7	41.0	3.731	0.045
sample size	202448	105059		
population std	11.388	12.042		
alpha	0.05	0.05	0 =p-value	
HO: (no significant difference between male and female ages)				
HA: (a significant difference between male and female ages)				

Is income statistically significant in regard to loan type?

- The p-value (0.0005) is less than the alpha (0.05)
- We reject the null hypothesis and we can see that income is statistically significant and there is a higher income level for those who have cash loans as opposed to those with revolving loans.

Despite having a higher income on average, people with cash loans tend to not repay their loan almost 3% more than people with revolving loans.

Cash loans vs Revolving Loans				
2 population (un-paired)	Cash loans income	Revolving loans income	sample mean (difference)	standard error difference
sample mean	169069.5	166217.0	2852.496	871.678
sample size	278232	29279		
population std	245910.951	126028.601		
alpha	0.05	0.05	0.000533156 =p-value	
HO: (no significant difference between cash loans vs revolving loans regarding income)				
HA: (a significant difference between cash loans vs revolving loans regarding income)				

Takeaways

Business Insights for Home Credit

Gender
- Since males have a higher chance of not repaying a loan, provide additional assistance for the male demographic, particularly for separated and widowed males who have over a 6% default rate.

Education
- Since default risk increases as education level decreases, implement courses on financial responsibility for prospective clients who have a lower education level before they are approved for a loan to lower default risk.

Age
- Implement courses for younger clients since they may not be as equipped with money management skills as opposed to older clients.

Next Steps

- Look into other demographics that can provide insights into who is more prone to default on a loan.
  - How long a client has been employed.
  - Default risk in regards to a client owning a car or owning realty.

- Look into past behavioral data and application data of client's previous loans.

End