DATA VISUALIZATION USING TABLEAU

Submitted by: Sai Saketh Boyanapalli

Created a tableau story to show the relationship between loan status and other variables of loan prosper data, Improved the visualization by taking feedback from 3 users.

Final visualization:

https://public.tableau.com/profile/sai.saketh#!/vizhome/LoanProsperProject_final/Story1?publish=yes

Initial visualization:

https://public.tableau.com/profile/sai.saketh#!/vizhome/LoanProsperProject_final/Story1?publish=yes

Summary

This graph shows the relation between the Income of the borrower, No of Credit lines, Interest Rate on loan per year, Status of the loan and employment status of the borrower at the time of listing and when hovered on a point it shows the amount loaned, no of records etc.,

Design

Conducted an Exploratory data analysis to initially explore data and to see if something interesting can be found.
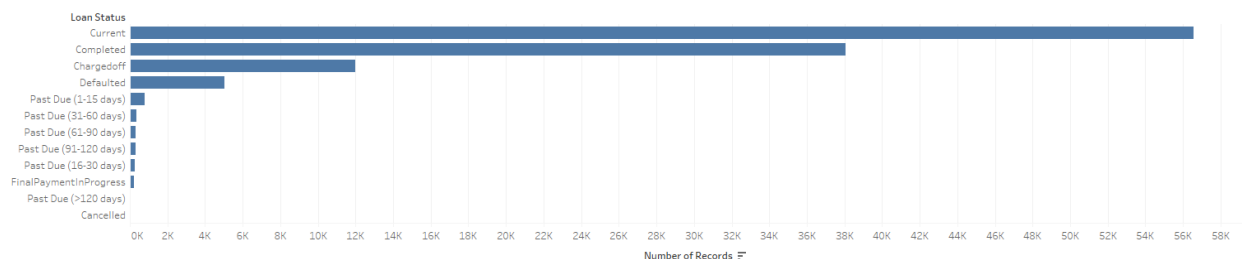
I have chosen a bar chart to see no of records across discrete categories of loan status.

From this bar chart we can see that most of the data is available for current, completed, charge off and defaulted loans

From this narrowed down the question to deal with completed, charged off and defaulted loans

I wanted to see if there are any trends that emerged from this.

As per variable to compare to I have looked at all the variables and noted down few variables of interest.

Loan original Amount

Monthly Loan payment

Income Range

Debt to Income Ratio

Current Credit Lines

Open Credit Lines

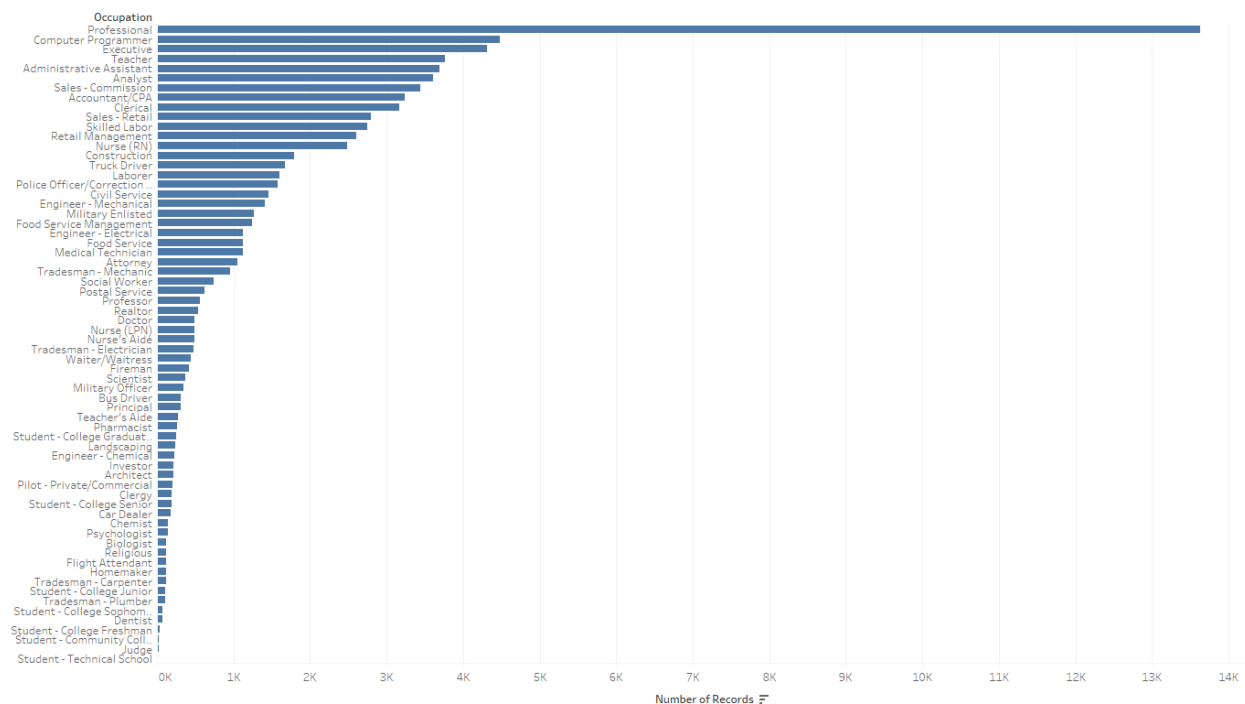Total Inquiries

Bank Card Utilization

Occupation

Employment Status

Term

From this, I was interested in no of loans taken with respect to occupation
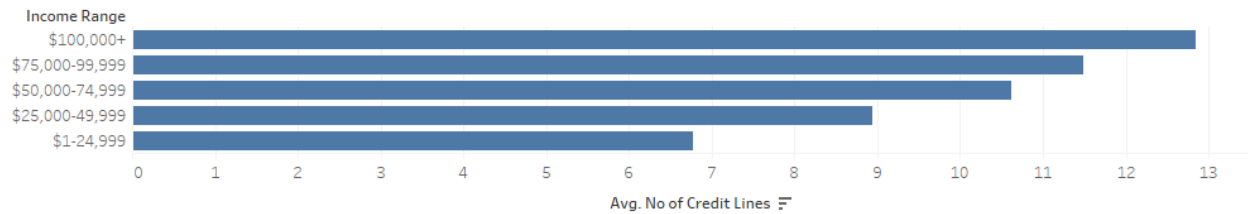

EDA Ocupation

Here we can see that Professionals have taken more loans than any other occupation, but all the other occupations listed below are professionals.
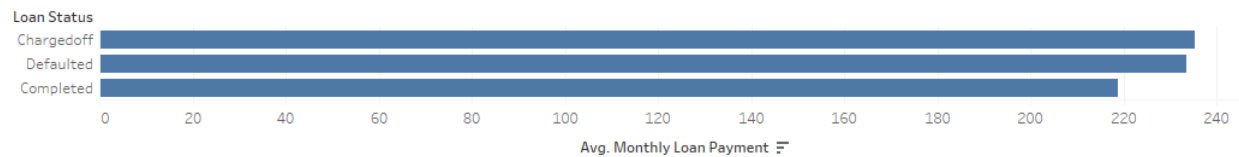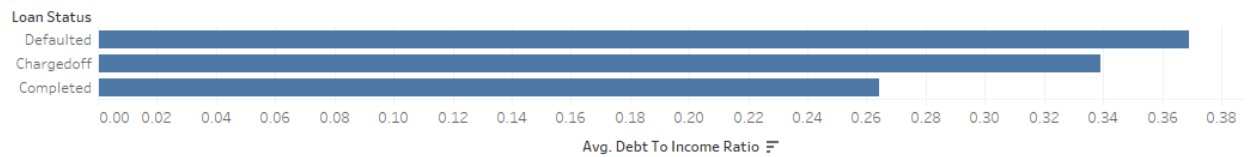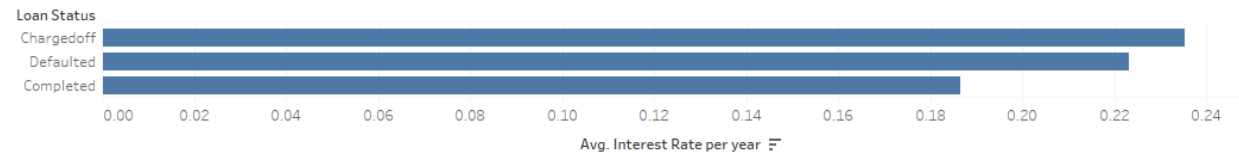
That removed occupation from the variables of interest.

Now I wanted to see a relationship between completed loans and defaulted/charged off loans

Created bar plots to see this trend across other measures.

Income Range

$100,000+	(bar to ~12.8)
$75,000-99,999	(bar to ~11.5)
$50,000-74,999	(bar to ~10.5)
$25,000-49,999	(bar to ~8.9)
$1-24,999	(bar to ~6.8)

Avg. No of Credit Lines

Interest rate/loan status

Loan Status

Chargedoff	(bar to ~0.235)
Defaulted	(bar to ~0.225)
Completed	(bar to ~0.185)

Avg. Interest Rate per year

Loan Status

Defaulted	(bar to ~0.37)
Chargedoff	(bar to ~0.345)
Completed	(bar to ~0.265)

Avg. Debt To Income Ratio

Loan Status

Chargedoff	(bar to ~235)
Defaulted	(bar to ~233)
Completed	(bar to ~220)

Avg. Monthly Loan Payment

Comparing discrete categories with an average measure

All the above graphs follow a specific trend.

To show this I have created a multivariate plot

Created a scatterplot to see this relation

Filtered loan status to show only completed, defaulted, charged off loans.

Thought that it would be best to showcase this using color of the point as we have 3 discrete statuses.

In case of income range size of point looked appropriate as that would seem that increase in size indicates increase in income.

Initially thought of showing employment status as a drop down but as we have few categories in this it was better to show everything at once.

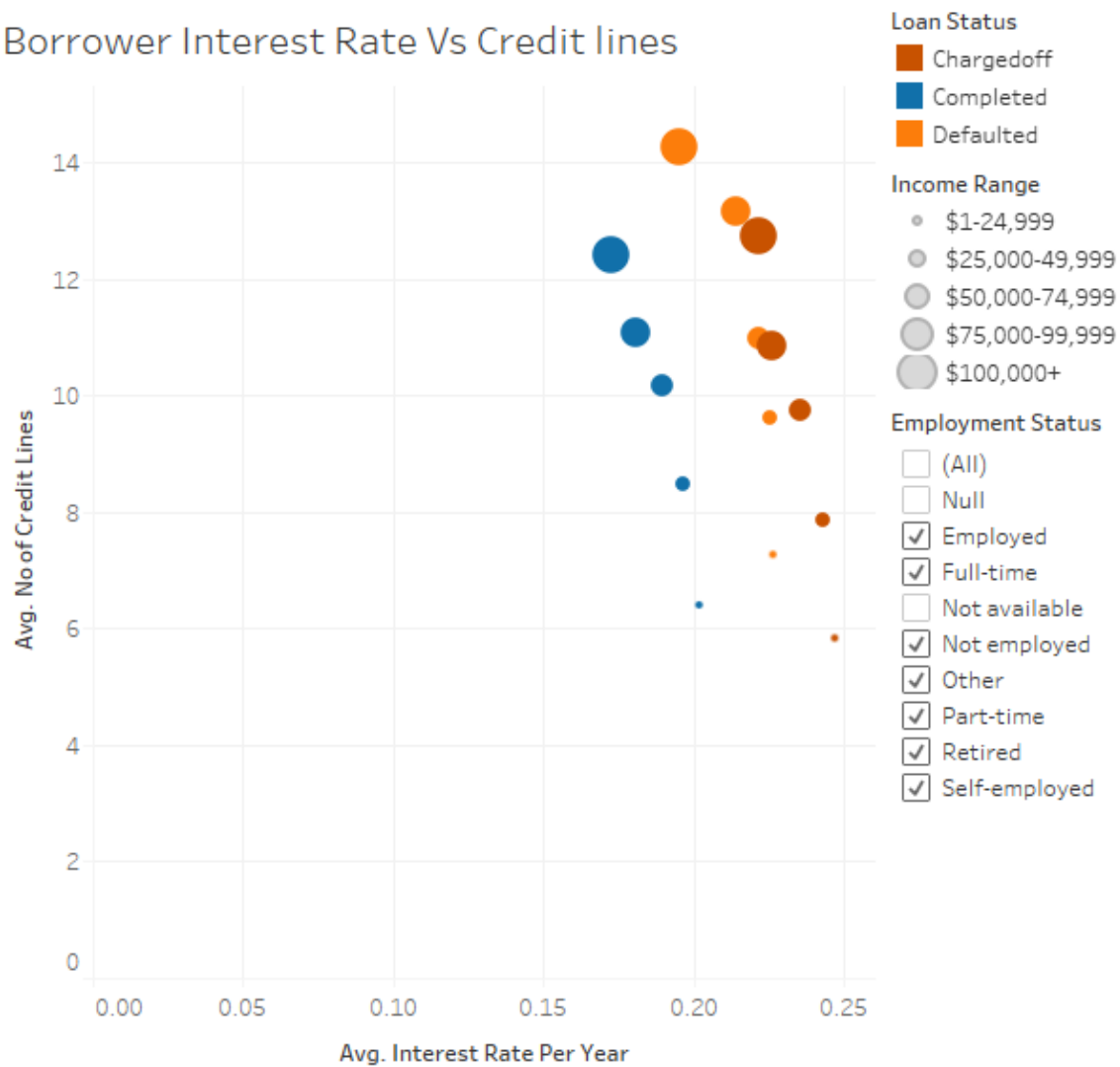**Initial visualization**

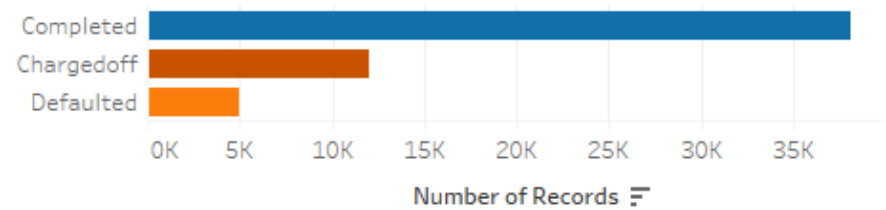## Initial

From the feedback received

1. Added a comprehensive title.
2. Changed the colors so, that they reflect what they are, and users can easily understand from it.
3. Chosen a color palette taking color blindness into account.
4. Added employment status tab so, that the reader understands it's an interactive graph
5. Changed the labels for axis to a more prominent one
6. Added a bar chart of no of records so that reader can see that there are more completed records than that of defaulted and charged off.
7. Added a link between two charts so, that changing the employment status in one graph reflects on other.
8. Added amount loaned in the details tab
9. Added no of records in the details tab
10. Changed tooltip labels

Final Visualization

Borrower Interest Rate Vs Credit lines

Loan Status
- Chargedoff
- Completed
- Defaulted

Income Range
- $1-24,999
- $25,000-49,999
- $50,000-74,999
- $75,000-99,999
- $100,000+

Employment Status
- (All)
- Null
- ☑ Employed
- ☑ Full-time
- Not available
- ☑ Not employed
- ☑ Other
- ☑ Part-time
- ☑ Retired
- ☑ Self-employed

Avg. No of Credit Lines (y-axis: 0, 2, 4, 6, 8, 10, 12, 14)

Avg. Interest Rate Per Year (x-axis: 0.00, 0.05, 0.10, 0.15, 0.20, 0.25)

Loan Status
- Completed
- Chargedoff
- Defaulted

(0K, 5K, 10K, 15K, 20K, 25K, 30K, 35K)

Number of Records

FEEDBACK

Feedback 1

What do you notice in the visualization?

More money you make more credit lines you have, higher the interest rate more likely the people are going to default!

What questions do you have about the data?

Couldn't notice that the graph is interactive, how many loans they took!

What relationships do you notice?

higher the interest rate more likely the people are going to default!

What do you think is the main takeaway from this visualization?

People are not fully understanding the amount of money they are borrowing and the reality of paying it back. Therefore, people are getting in over their head and defaulting of failing to pay back loans across socio-economic status.

Is there something you don't understand the graphic?

Are their different "requirement" that people must fulfill prior to opening a credit line. Individual across economic standing are receiving excessive financial loans, but continuing to default. What policies are banks/credit lines using?

Feedback 2

What do you notice in this visualization?

A relation between income and credit lines and Borrower rate increases people tend to default loan

What relationships do you notice?

That borrower rate plays the main role where you tend to complete or default loan

Is there something you don't understand in the graphic?

What borrower rate means

Colors were bit misleading

No title

Was not able to know that the graph was interactive

Feedback 3

What do you notice in this visualization?

As the income level increases, credit line increases

What relationships do you notice?

What do you think is the main takeaway from this visualization?

Increase borrower rate more loan defaulting, charged off

Credit line increase

Is there something you don't understand in the graphic?

Could not understand the borrower rate

Could not understand why there isn't significant difference between number of completed records and defaulters, while the plot shows otherwise.

Resources:

https://www.tableau.com/