PROPER LOAN DATASEST

GENERAL LOAN INFORMATION

There is a total of 81 columns and 113937 observations. The loan data is obtained from Prosper, a P2P lending company in the USA. While some of the loans appear to be closed loans, some are still open (meaning that there is still some form of transaction going on). The loan data shows lending transactions that has taken place over time and the outcome of the transactions depicted by the loan status. The loan status appears to be the dependent variable which can be predicted using other available variables in the dataset. The loan status reveal borrowers whose loans are still current, charged off, defaulted, completed payment and past due.

Priority of this visualization

My major interest is to find out factors that lead to increased loan default, which Prosper must avoid in order to remain in business and not incur heavy losses.

Overall, knowing which class (with respect to their credit grade, occupation, employment status, borrower state, proposer score (alpha) and other credit information) of borrowers default' on their loans would be the priority of this visualization so as to save the Lending company 'prosper' from accumulating huge losses and losing customer trust.

MAIN FINDINGS

The distribution of the loan amount is right-skewed. The loan distribution is Bi-modal with majority of the loan's distribution around 4000, 10000, 16000, while the smaller portion of loans are distributed around 20000 and 26000.

From Graph 3, borrowers with income range of 50000 - 74999 were the highest borrowers, followed by borrowers with income range of 25000 – 49999. This highest group of income earners (100000 + and 75000 – 99999) were 3rd and 4th highest borrowers respectively.

From Graph 9, as already stated initially, the management policy of Prosper is well to do given that they do not grant as much loans to the borrowers with employment status of 'retired', 'other', 'part-time', 'self-employed', 'Not-employed'. This serves as a preventive measure against 'potential borrowers' who cannot repay their loans. It shows the people who are 'employed' have lesser financial risks associated with them as they default least.

Graph 11 – Distribution of borrower states by current_default ratio

From the graph, it shows that states which attracts less risks of defaults, includes:

1. Mt – Montana
2. HI – Hawaii
3. CO – Colorado
4. WY – Wyoming
5. NM - New Mexico
6. NH - New Hampshire
7. AZ – Arizona

8. NE - Nebraska
9. MA – Massachusetts
10. TX – Texas
11. NY – New York City
12. CA – California
13. AR – Arkansas
14. OR – Orlando
15. NJ - New Jersey
16. DC – Washington DC
17. VT - Vermont
18. IL – Illinois
19. IN – Indiana

NOTE: The above states have high current_default ratio

States which pose a high default risk, in order of severity includes:

1. RI – Rhode Island
2. SD – South Dakota
3. NC – North Carolina
4. MO – Missouri
5. TN – Tennessee
6. KY – Kentucky
7. MS – Mississippi
8. ID – Idaho
9. AL – Alabama
10. MD – Maryland
11. LA – Los Angeles
12. FL – Florida

NOTE: Unfortunately, 'the reason for borrowing the loan' is not stated. That would have been a very useful information.

From the graph 12, borrowers with loan status of 36months perform better than borrowers with loan status of 60 months. In this case, it is difficult to say if there are more individual borrowers compared to borrowers who seeks 60 months loan (business owners), or individual borrowers are more favoured

From the graph 14, it shows that the management of Prosper displayed good management practices. Because the ratio shows that the algorithm behind the 'prosper rating' has a high precision and accuracy. The credit risk does indeed decrease down the ratings and vice versa.

Graph 15 shows the Occupations to which borrowers are less likely to default. From the graph, we can see that the best professions in order of less financial risks are:

1. Pilots
2. Student – college graduates
3. Attorney
4. Tradesman/Carpenter
5. Psychologist
6. Engineer – Chemical

7. Scientist
8. Tradesman – Electrician
9. Doctor
10. Architect
11. Analyst
12. Investor
13. Executive
14. Military officer
15. Nurse
16. Computer Programmer
17. Engineer – mechanical

Most risky occupations with a strong likelihood to default include:

1. Students – community college
2. Students – college freshman
3. Students – college junior
4. Tradesman – Plumber
5. Homemaker
6. Waiter/waitress
7. Food service
8. Religious
9. Labourer
10. Student – college senior
11. Military enlisted
12. Sales – retail
13. Nurse's Aide

Graph 17 - Plot of Loan status against Borrower annual percentage rate and monthly income

Once again, the major interest is the default rate. One significant observation shows that 'all loan amount greater 25, 000' has no defaults. So, it could mean that borrowers who borrow in this range do not default. (this required further investigation)