Readme

## Udacity Communicate data findings

# Léon de Jong

# Dataset

The dataset consists of 113937 entries of loan-listings from the website [www.prosper.com](http://www.prosper.com)

The dataset is obtained from

<https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv>

and comes with an explanatory variable definitions spreadsheet that can be accessed through

<https://docs.google.com/spreadsheets/d/1gDyi_L4UvIrLTEC6Wri5nbaMmkGmLQBk-Yx3z0XDEtI/edit#gid=0>

## Findings

The dataset turned out to contain a vast nr of 81 variables. A selection of 25 was chosen to further investigate the relation between two ratings systems used and a set of potential influencing variables.

Initial investigation showed differences between the rating systems when compared to the potential influencing variables. Some influencing variable like Estimated loss had a large effect where others seem to influence the score less.

## Recourses

During the process I have made use of stackoverflow and the seaborn and matplotlib documentation

Such as:

<https://matplotlib.org/3.1.1/gallery/subplots_axes_and_figures/figure_title.html>

<https://seaborn.pydata.org/tutorial/color_palettes.html>

<https://stackoverflow.com/questions/44654421/getting-the-same-subplot-size-using-matplotlib-imshow-and-scatter>

<https://www.mikulskibartosz.name/how-to-remove-outliers-from-seaborn-boxplot-charts/>

<https://stackoverflow.com/questions/33005872/seaborn-regplot-partially-see-through-alpha>

<https://stackoverflow.com/questions/36921550/how-to-change-the-point-size-for-regplot-seaborns-scatter-plot-function-pyt>

[python - seaborn displot() is not plotting within defined subplots - Stack Overflow](https://stackoverflow.com/questions/63895392/seaborn-displot-is-not-plotting-within-defined-subplots)

[python - Change x ticks in Regplot Seaborn - Stack Overflow](https://stackoverflow.com/questions/61410812/change-x-ticks-in-regplot-seaborn)

[python - reverse the order of CategoricalDtype in Pandas - Stack Overflow](https://stackoverflow.com/questions/52746012/reverse-the-order-of-categoricaldtype-in-pandas)

## Presentation

The presentation focusses on the key variables and how the relate to the rating systems.

The skewedness of a particular variable `Amounts Delinquent` is highlighted to show axis transformation in a log scale

Further plots show the different relations between the key variables and the rating system plus the conclusion

## Conclusion

The Alpha rating system seems to be the most in step with the viewed explanatory variables.