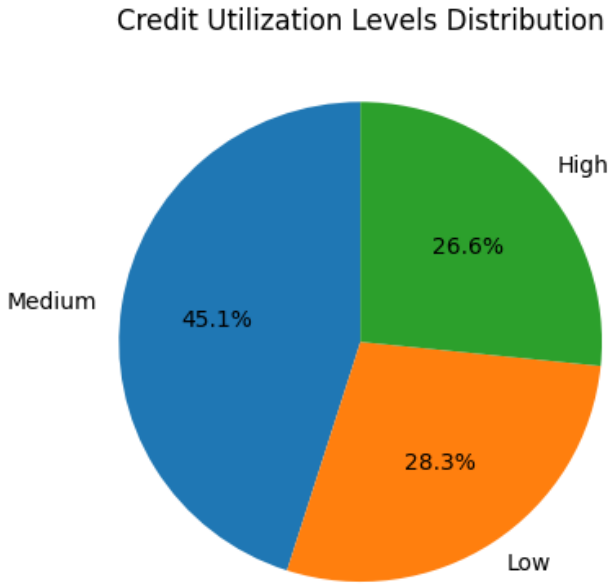
GREENSOL DATA SCIENCE INTERN ASSESSMENT REPORT

The credit scoring assessment project explores the use of data science techniques to evaluate an individual's credit worthiness based on historical customer data.

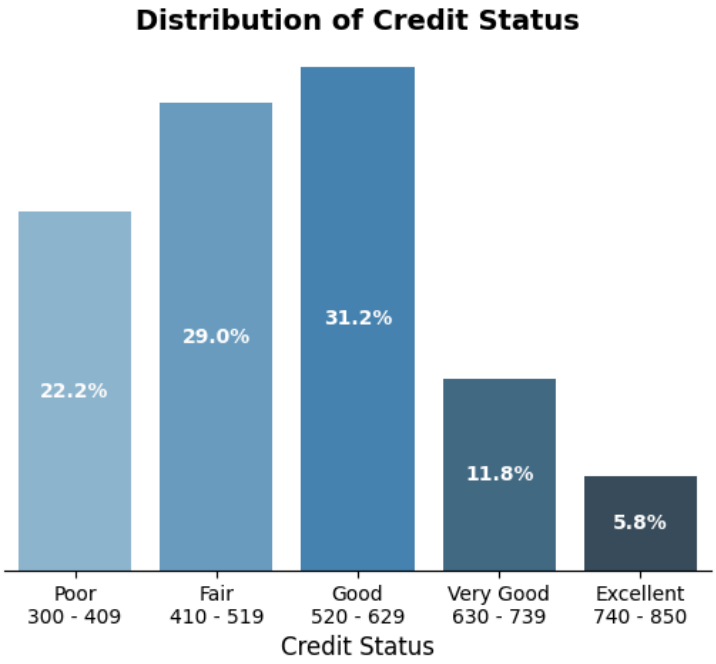
CREDIT UTILIZATION LEVELS

The pie chart presents a proportional view of customers by their credit utilization level. Higher credit utilization level suggests over-reliance on credit with high risk of not paying back. The chart reveals that a significant portion of the customers Falls within "Medium" category, suggesting a moderate credit risk profile. The chart reveals that a significant portion of the customer falls within the low to medium risk, collectively accounting for 73.4% of the population. This indicates that while the majority of customers are 'creditworthy', there is still a notable segment classified as "High", which raises risk considerations.



DISTRIBUTIION OF CREDIT STATUS

The bar chart, enhanced with percentage labels inside each bar. It confirms the dominance of "Good" (30.4%) and "Fair" (28.2%) credit statuses, followed by "Poor" (21.6%). "Very Good" and "Excellent" segments are comparatively smaller. Indicating a limited pool of highly creditworthy individuals. The bar chart shows that 48.8% of the users (falling into the "Good", "Very Good" and "Excellent" category) can be trusted to use the credits.



Assumptions made:

- I assumed the *transaction_date* to be the date the user made the payment or transaction.
- I assumed the *due_date* to be the date when the payment was expected.

Challenges encountered:

- I faced the challenge of understanding the context of some features - a metadata would have helped eliminate this.

Improvements if going into production:

- I would suggest more dta - to help generalize better.
- I would suggest a machine learning model be used rather than a scoring formula - a machine learning model would be able to learn more patterns in the dataset.