# LENDING CLUB CASE STUDY ANALYSIS


- Submitted by
  Aashna Behl

# Problem Statement

We are provided with real world data of the largest online marketplace facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

The goal is to identify the risky applicants based on the EDA techniques and share the recommendations for the same.

# Analysis Approach

- - Clean the data
- - Understand Patterns of each variable individually.
- - Understand Patterns of multiple variables with each other.
- - Give recommendations based on the patterns.

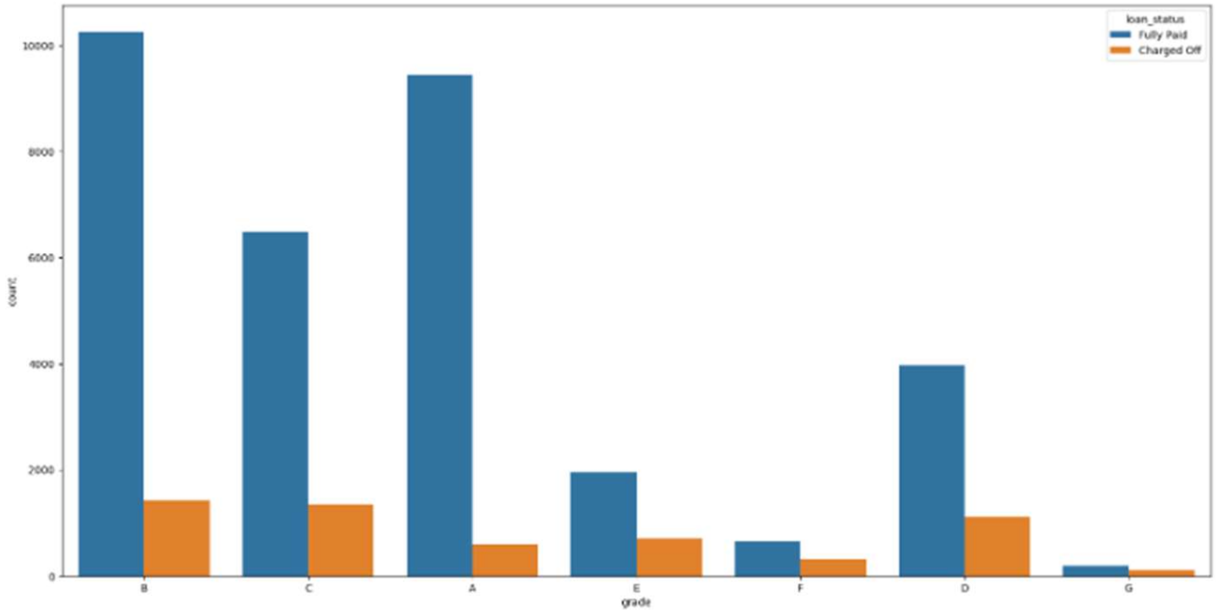
# Univariate Analysis

Analysis done for both categorical and numerical features.

- - Higher loan amounts have higher defaulters.
- - Higher grade and Sub Grade loans have higher number of defaulters.
- - Verification Process does not seem to be effective, since unverified borrowers are the highest defaulters.
- - Educational , Renewable energy and Small Business Loans are risky
- - Borrowers Public Record Bankruptcies and/or derogatory Public records are risky
- Any sort of delinquency is risky.
- More enquiries are risky.

## Grade

In [284]:
```
uni_cat_analysis('grade')
```

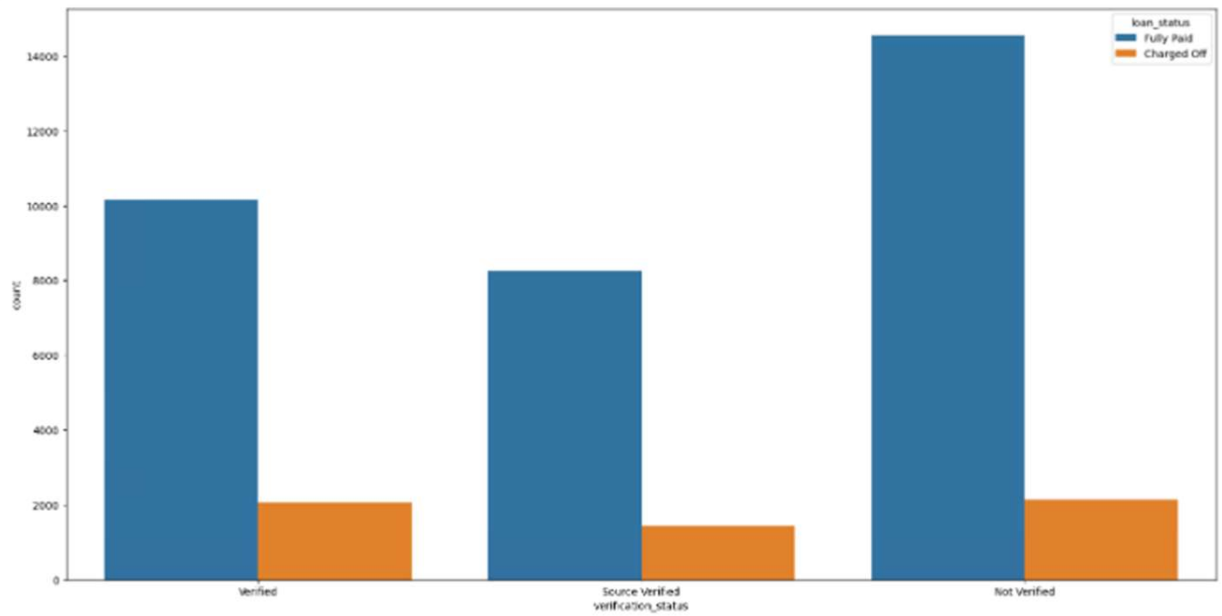



| | grade | Defaulter% | tot_count |
|---|---|---|---|
| 0 | G | 33.779264 | 299 |
| 1 | F | 32.684426 | 976 |
| 2 | E | 26.849418 | 2663 |
| 3 | D | 21.986234 | 5085 |
| 4 | C | 17.194281 | 7834 |
| 5 | B | 12.205567 | 11675 |
| 6 | A | 5.993031 | 10045 |

Observation: Grade of Loan is certainly linked with higher defaulters. Defaulter % continuously increases from A to G with D,E,F,G being risky grades. Lower grade loan is less risk

## Verification Status

In [289]:
```
uni_cat_analysis('verification_status')
```

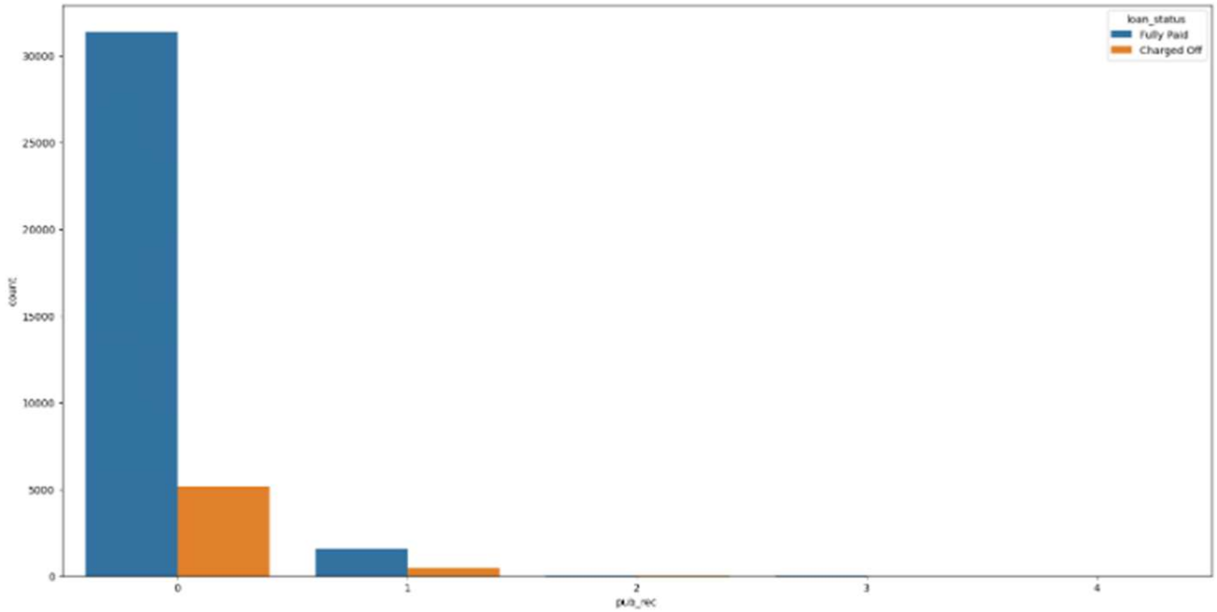



| | verification_status | Defaulter% | tot_count |
|---|---|---|---|
| 0 | Verified | 16.803212 | 12206 |
| 1 | Source Verified | 14.818642 | 9677 |
| 2 | Not Verified | 12.830957 | 16694 |

Observation: Surprisingly a 'Verfied' status has more defaulters than non-verified. Requested verification documents either not vetted or can be forged easily.

## Derogatory Public Records

In [292]:
```
uni_cat_analysis('pub_rec')
```

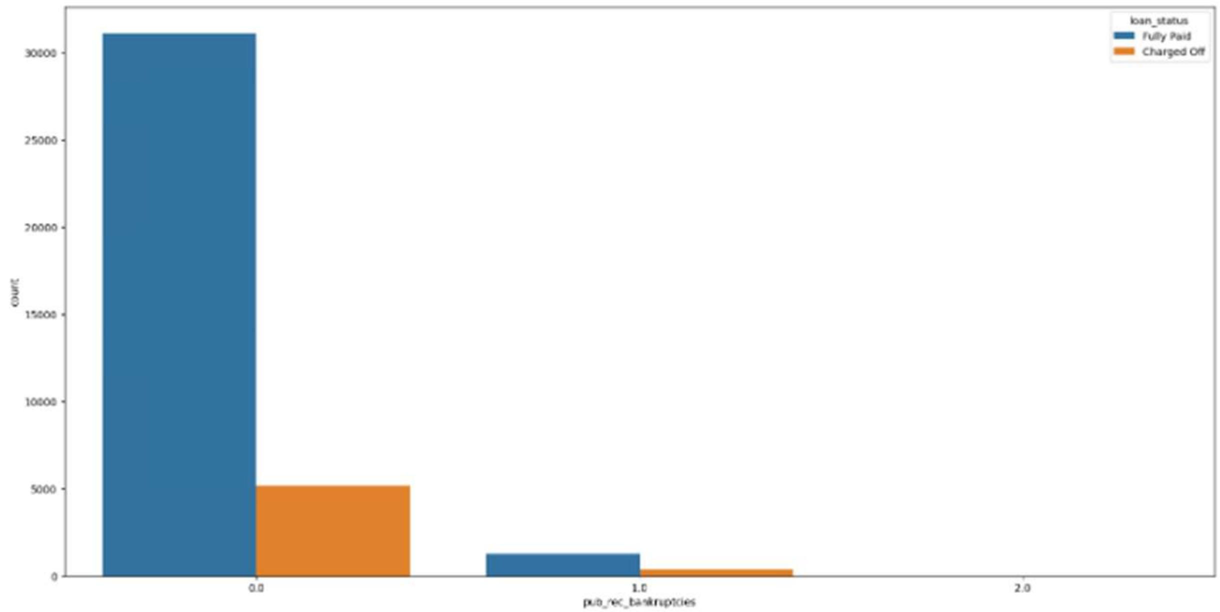



| | pub_rec | Defaulter% | tot_count |
|---|---|---|---|
| 0 | 1 | 22.702434 | 2013 |
| 1 | 2 | 20.833333 | 48 |
| 2 | 0 | 14.134276 | 36507 |

Observation: Any amount of public derogatory records has higher chance of defaulters.

## Public record Bankruptcies

In [293]:
```
uni_cat_analysis('pub_rec_bankruptcies')
```

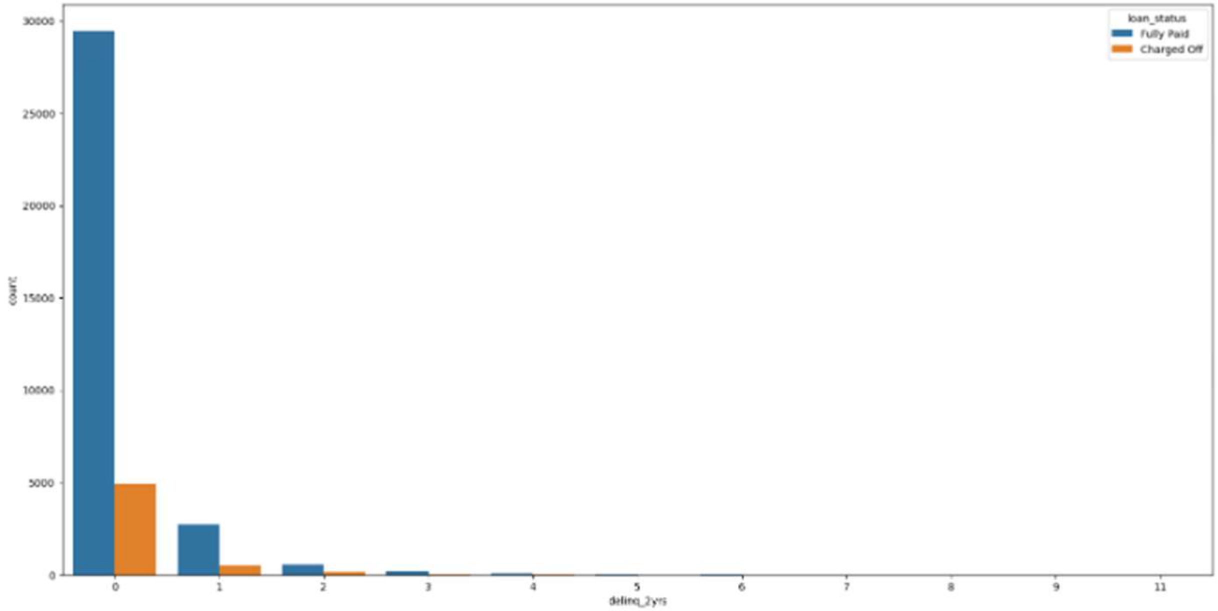



| | pub_rec_bankruptcies | Defaulter% | tot_count |
|---|---|---|---|
| 0 | 2.0 | 40.000000 | 5 |
| 1 | 1.0 | 22.357972 | 1637 |
| 2 | 0.0 | 14.186765 | 36238 |

Observation: As expoected, Higher the banckruptcies recorded higher the chance of defaulters.

## 30+ Days Overdue in last 2 years

In [332]:
```
#Even though numeric data, can be analysed as categorical for the kind of data is present.
uni_cat_analysis('delinq_2yrs')
```

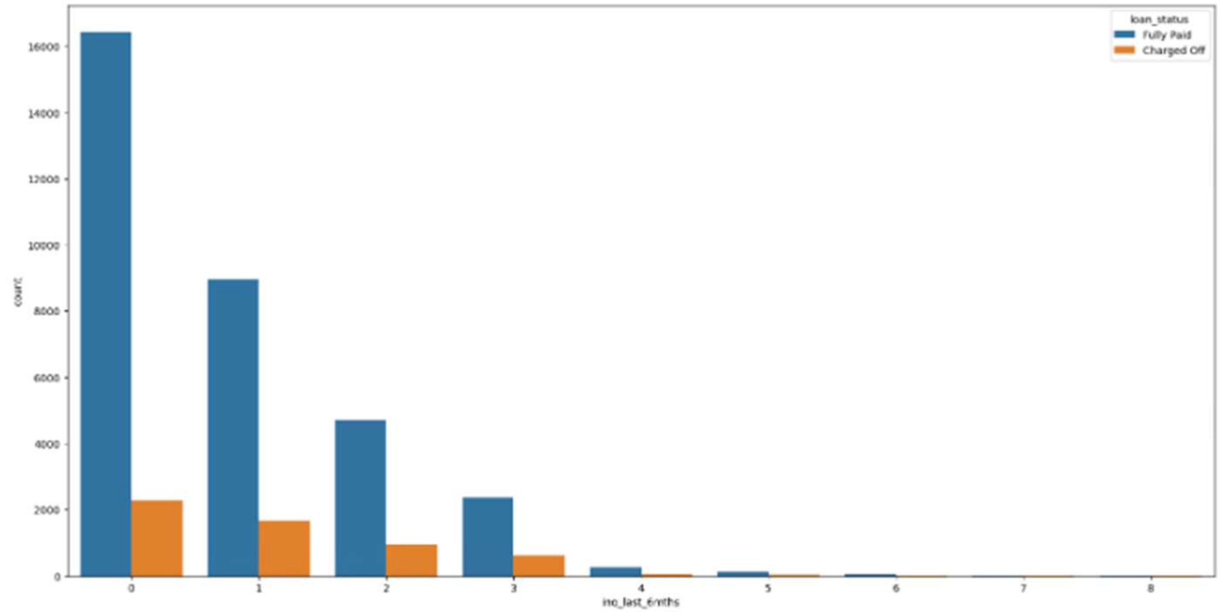



| | delinq_2yrs | Defaulter% | tot_count |
|---|---|---|---|
| 0 | 8 | 50.000000 | 2 |
| 1 | 7 | 25.000000 | 4 |
| 2 | 4 | 21.666667 | 60 |
| 3 | 2 | 19.316493 | 673 |
| 4 | 3 | 16.509434 | 212 |
| 5 | 1 | 15.809167 | 3207 |
| 6 | 0 | 14.354679 | 34386 |
| 7 | 5 | 14.285714 | 21 |
| 8 | 6 | 10.000000 | 10 |

Observation: Any sort of deliqunecy is definitely an indicator of risk. No past due has lesser defaulters in comparison to loans. There are exceptions but very few. Higher delinq_2yrs should be avoided for loan.

## Inquiry in Last 6 months

In [337]:
```
#Even though numeric data, can be analysed as categorical for the kind of data is present.
uni_cat_analysis('inq_last_6mths')
```




| | inq_last_6mths | Defaulter% | tot_count |
|---|---|---|---|
| 0 | 7 | 29.411765 | 34 |
| 1 | 6 | 25.396825 | 63 |
| 2 | 8 | 21.428571 | 14 |
| 3 | 3 | 20.750922 | 2983 |
| 4 | 5 | 19.444444 | 144 |
| 5 | 2 | 16.678458 | 5654 |
| 6 | 4 | 16.139241 | 316 |
| 7 | 1 | 15.731707 | 10660 |
| 8 | 0 | 12.186648 | 18709 |

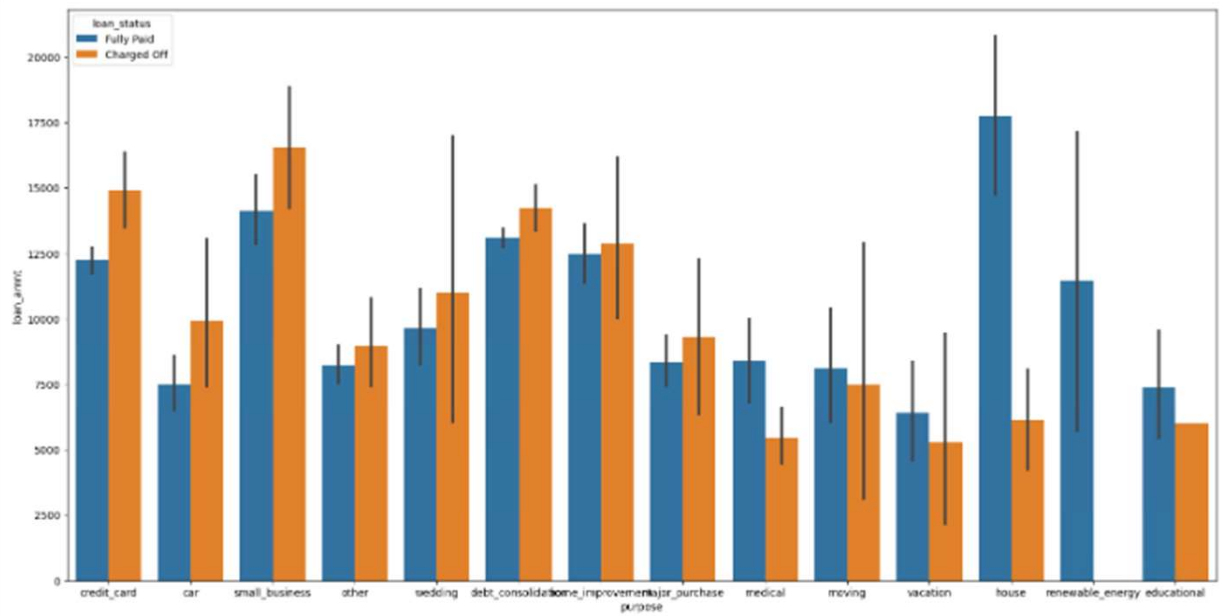
Observation: Higher enquiries mean higher risk. Most people who have taken loan have taken withing 2-3 enquiries. Individuals who have enquired a lot of times have also defaulted more.

# Bivariate Analysis

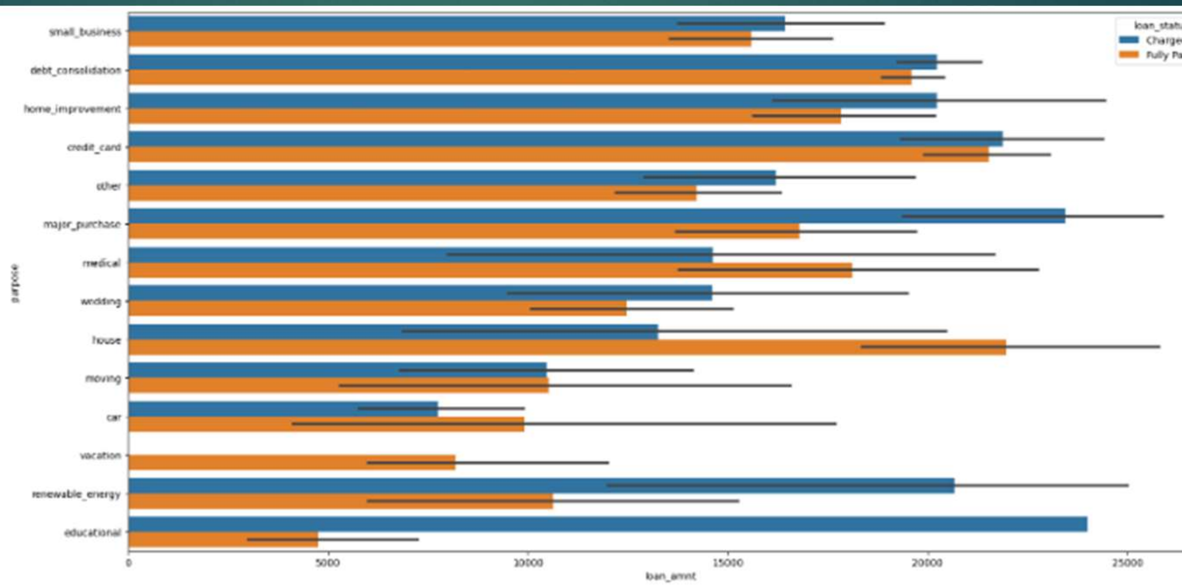
- Home improvement , Renewable energy loans and house loans are higher in higher income groups.
- House loans are safer in higher grades but riskier in lower grades.
- Car Loans, Credit card loans and small business loans are risky in December.
- 4 states WT, MT, UT and TN are risky. Credit card loans have the highest risk in these states, and higher income groups interestingly are more risky in these states.

**Loan Amount against Issue Month Dec to see the pattern**

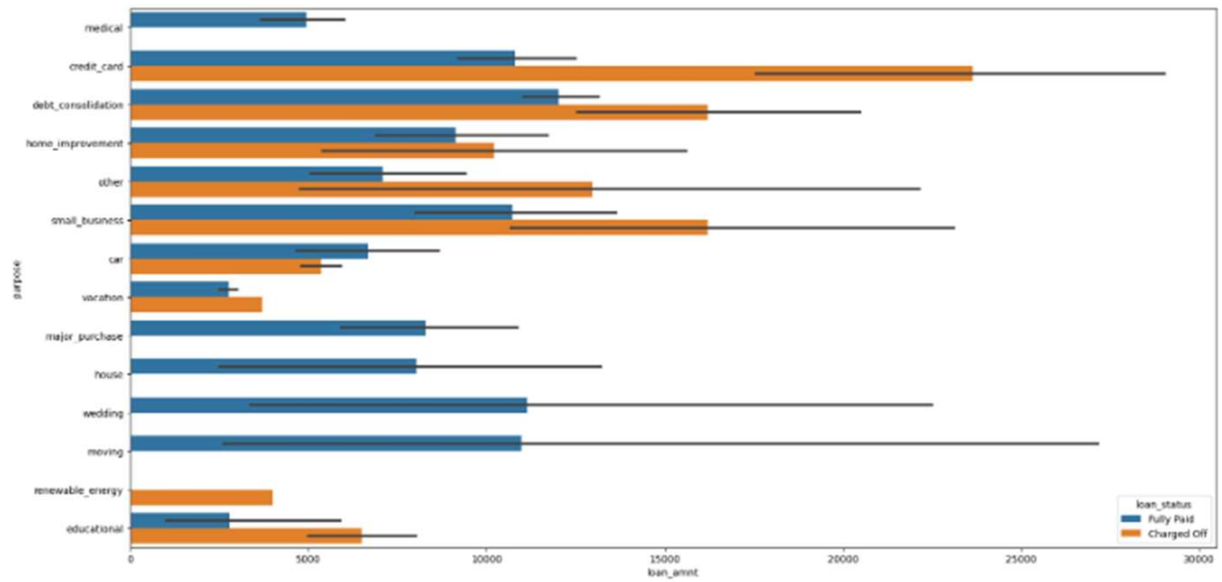
In [542]:
```
fig, ax = plt.subplots(figsize=(20, 10))
df_loan_closed_purpose = df_loan_closed.loc[(df_loan_closed.issue_month_c == 'Dec')]
sns.barplot(x = 'purpose',y ='loan_amnt',hue = 'loan_status', data = df_loan_closed_purpose)
plt.show()
```




Observation: Small Business loan continues to stay the most risky in Dec. Credit Card Loans and Car Loans in Dec are risky.




1 Observation: Educational loans and renewable energy loans and major purchase loans of higher grade(F,G) are risky. House loans and medical loans of higher grade are safer.




Observation: Maximum loans in these 4 states are credit card with highest default ratio. Small business and debt consolidation loans are also higher amount and high defaut ratio.

# Summary

- Make verification status stringent.
- Check for public record bankruptcies and derogatory public records
- High debt to income ratio risky.
- Check credit card loans in WY, MT, UT, TN
- Missing employment records risky.
- Any delinquency is high risk.
- High number of inquiries are high risk.