Table of Contents

Prosper Loan Dataset

by Uju Chinedum

Introduction

Notebook Imports and Settings

In [1]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sb

%matplotlib inline
```

In [2]:
```
pd.options.display.max_columns = 81
pd.options.display.max_rows = 81
sb.set_style("dark")
```

Function Definitions

In [3]:
```
def summary(df):
    df.info()
    return df.sample(10)
```

In [4]:
```
def dtype(col, dtype):
    """
    Returns a series with the data type `dtype`
    """

    if dtype == "date":
        if type(col) == list:
            for i in col:
                df[i] = pd.to_datetime(df[i])
        else:
            df[col] = pd.to_datetime(df[col])

    else:
        if type(col) == list:
            for i in col:
                df[i] = df[i].astype(dtype)
        else:
            df[col] = df[col].astype(dtype)
```

Data Wrangling

Data Gathering

In [5]:
```
data = pd.read_csv("Prosper Loan Data.csv")
```

Data Assessing

In [6]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113937 entries, 0 to 113936
Data columns (total 81 columns):
 #   Column                               Non-Null Count   Dtype  
---  ------                               --------------   -----  
 0   ListingKey                           113937 non-null  object 
 1   ListingNumber                        113937 non-null  int64  
 2   ListingCreationDate                  113937 non-null  object 
 3   CreditGrade                          28953 non-null   object 
 4   Term                                 113937 non-null  int64  
 5   LoanStatus                           113937 non-null  object 
 6   ClosedDate                           55089 non-null   object 
 7   BorrowerAPR                          113912 non-null  float64
 8   BorrowerRate                         113937 non-null  float64
 9   LenderYield                          113937 non-null  float64
 10  EstimatedEffectiveYield              84853 non-null   float64
 11  EstimatedLoss                        84853 non-null   float64
 12  EstimatedReturn                      84853 non-null   float64
 13  ProsperRating (numeric)              84853 non-null   float64
 14  ProsperRating (Alpha)                84853 non-null   object 
 15  ProsperScore                         84853 non-null   float64
 16  ListingCategory (numeric)            113937 non-null  int64  
 17  BorrowerState                        108422 non-null  object 
 18  Occupation                           110349 non-null  object 
 19  EmploymentStatus                     111682 non-null  object 
 20  EmploymentStatusDuration             106312 non-null  float64
 21  IsBorrowerHomeowner                  113937 non-null  bool   
 22  CurrentlyInGroup                     113937 non-null  bool   
 23  GroupKey                             13341 non-null   object 
 24  DateCreditPulled                     113937 non-null  object 
 25  CreditScoreRangeLower                113346 non-null  float64
```

```
 26  CreditScoreRangeUpper                113346 non-null  float64
 27  FirstRecordedCreditLine              113240 non-null  object
 28  CurrentCreditLines                   106333 non-null  float64
 29  OpenCreditLines                      106333 non-null  float64
 30  TotalCreditLinespast7years           113240 non-null  float64
 31  OpenRevolvingAccounts                113937 non-null  int64
 32  OpenRevolvingMonthlyPayment          113937 non-null  float64
 33  InquiriesLast6Months                 113240 non-null  float64
 34  TotalInquiries                       112778 non-null  float64
 35  CurrentDelinquencies                 113240 non-null  float64
 36  AmountDelinquent                     106315 non-null  float64
 37  DelinquenciesLast7Years              112947 non-null  float64
 38  PublicRecordsLast10Years             113240 non-null  float64
 39  PublicRecordsLast12Months            106333 non-null  float64
 40  RevolvingCreditBalance               106333 non-null  float64
 41  BankcardUtilization                  106333 non-null  float64
 42  AvailableBankcardCredit              106393 non-null  float64
 43  TotalTrades                          106393 non-null  float64
 44  TradesNeverDelinquent (percentage)   106393 non-null  float64
 45  TradesOpenedLast6Months              106393 non-null  float64
 46  DebtToIncomeRatio                    105383 non-null  float64
 47  IncomeRange                          113937 non-null  object
 48  IncomeVerifiable                     113937 non-null  bool
 49  StatedMonthlyIncome                  113937 non-null  float64
 50  LoanKey                              113937 non-null  object
 51  TotalProsperLoans                    22085 non-null   float64
 52  TotalProsperPaymentsBilled           22085 non-null   float64
 53  OnTimeProsperPayments                22085 non-null   float64
 54  ProsperPaymentsLessThanOneMonthLate  22085 non-null   float64
 55  ProsperPaymentsOneMonthPlusLate      22085 non-null   float64
 56  ProsperPrincipalBorrowed             22085 non-null   float64
 57  ProsperPrincipalOutstanding          22085 non-null   float64
 58  ScorexChangeAtTimeOfListing          18928 non-null   float64
 59  LoanCurrentDaysDelinquent            113937 non-null  int64
 60  LoanFirstDefaultedCycleNumber        16952 non-null   float64
 61  LoanMonthsSinceOrigination           113937 non-null  int64
 62  LoanNumber                           113937 non-null  int64
 63  LoanOriginalAmount                   113937 non-null  int64
 64  LoanOriginationDate                  113937 non-null  object
 65  LoanOriginationQuarter               113937 non-null  object
 66  MemberKey                            113937 non-null  object
 67  MonthlyLoanPayment                   113937 non-null  float64
 68  LP_CustomerPayments                  113937 non-null  float64
 69  LP_CustomerPrincipalPayments         113937 non-null  float64
 70  LP_InterestandFees                   113937 non-null  float64
 71  LP_ServiceFees                       113937 non-null  float64
 72  LP_CollectionFees                    113937 non-null  float64
 73  LP_GrossPrincipalLoss                113937 non-null  float64
 74  LP_NetPrincipalLoss                  113937 non-null  float64
 75  LP_NonPrincipalRecoverypayments      113937 non-null  float64
 76  PercentFunded                        113937 non-null  float64
 77  Recommendations                      113937 non-null  int64
 78  InvestmentFromFriendsCount           113937 non-null  int64
 79  InvestmentFromFriendsAmount          113937 non-null  float64
 80  Investors                            113937 non-null  int64
dtypes: bool(3), float64(50), int64(11), object(17)
memory usage: 68.1+ MB
```

In [7]:
```
data.sample(10)
```

Out[7]:

	ListingKey	ListingNumber	ListingCreationDate	CreditGrade	Term	LoanStatus	ClosedDat
4938	4B5A35944694837292AEE4F	1018308	2013-11-05 08:35:05.317000000	NaN	36	Current	Nal

52843	66693380122035131DA65D2	85599	2007-01-16 16:09:28.267000000	HR	36	Completed	2010-01-2 00:00:0
49374	F19835453901460088DAC06	581040	2012-04-21 21:47:55.883000000	NaN	60	Current	NaN
86450	D1B135904961755452E804F	904618	2013-09-17 14:59:22.417000000	NaN	36	Current	NaN
57449	AA963408420273796DFCC28	259816	2008-01-04 03:36:19.833000000	C	36	Completed	2010-11-1 00:00:0
64720	CB2A3604389785459F86BAB	1223054	2014-03-04 06:07:25.433000000	NaN	36	Current	NaN
75625	65D23593827405001E3DF6F	993064	2013-11-13 15:03:11.413000000	NaN	60	Current	NaN
78189	860135814904863476FFD70	803089	2013-06-09 14:43:39.137000000	NaN	60	Current	NaN
31176	18C535675611806745DA84E	693577	2013-01-02 19:36:51.577000000	NaN	12	Completed	2014-01-1 00:00:0
32704	54B13600143971062466FF4	1133030	2014-01-13 19:50:36.300000000	NaN	36	Current	NaN

In [8]:
```
data.isna().sum()
```

Out[8]:
```
ListingKey                          0
ListingNumber                       0
ListingCreationDate                 0
CreditGrade                     84984
Term                                0
LoanStatus                          0
ClosedDate                      58848
BorrowerAPR                        25
BorrowerRate                        0
LenderYield                         0
EstimatedEffectiveYield         29084
EstimatedLoss                   29084
EstimatedReturn                 29084
ProsperRating (numeric)         29084
ProsperRating (Alpha)           29084
ProsperScore                    29084
ListingCategory (numeric)           0
BorrowerState                    5515
Occupation                       3588
EmploymentStatus                 2255
EmploymentStatusDuration         7625
IsBorrowerHomeowner                 0
CurrentlyInGroup                    0
GroupKey                       100596
DateCreditPulled                    0
CreditScoreRangeLower             591
CreditScoreRangeUpper             591
FirstRecordedCreditLine           697
CurrentCreditLines               7604
OpenCreditLines                  7604
TotalCreditLinespast7years        697
OpenRevolvingAccounts               0
OpenRevolvingMonthlyPayment         0
InquiriesLast6Months              697
TotalInquiries                   1159
CurrentDelinquencies              697
AmountDelinquent                 7622
DelinquenciesLast7Years           990
```

```
PublicRecordsLast10Years                697
PublicRecordsLast12Months              7604
RevolvingCreditBalance                 7604
BankcardUtilization                    7604
AvailableBankcardCredit                7544
TotalTrades                            7544
TradesNeverDelinquent (percentage)     7544
TradesOpenedLast6Months                7544
DebtToIncomeRatio                      8554
IncomeRange                               0
IncomeVerifiable                          0
StatedMonthlyIncome                       0
LoanKey                                   0
TotalProsperLoans                     91852
TotalProsperPaymentsBilled            91852
OnTimeProsperPayments                 91852
ProsperPaymentsLessThanOneMonthLate   91852
ProsperPaymentsOneMonthPlusLate       91852
ProsperPrincipalBorrowed              91852
ProsperPrincipalOutstanding           91852
ScorexChangeAtTimeOfListing           95009
LoanCurrentDaysDelinquent                 0
LoanFirstDefaultedCycleNumber         96985
LoanMonthsSinceOrigination                0
LoanNumber                                0
LoanOriginalAmount                        0
LoanOriginationDate                       0
LoanOriginationQuarter                    0
MemberKey                                 0
MonthlyLoanPayment                        0
LP_CustomerPayments                       0
LP_CustomerPrincipalPayments              0
LP_InterestandFees                        0
LP_ServiceFees                            0
LP_CollectionFees                         0
LP_GrossPrincipalLoss                     0
LP_NetPrincipalLoss                       0
LP_NonPrincipalRecoverypayments           0
PercentFunded                             0
Recommendations                           0
InvestmentFromFriendsCount                0
InvestmentFromFriendsAmount               0
Investors                                 0
dtype: int64
```

Data Cleaning

In [9]:
```
df = data.copy()
```

In [10]:
```
extreme_na_cols = [col for col in df.columns if df[col].isna().sum() >= 58848]
extreme_na_cols
```

Out[10]:
```
['CreditGrade',
 'ClosedDate',
 'GroupKey',
 'TotalProsperLoans',
 'TotalProsperPaymentsBilled',
 'OnTimeProsperPayments',
 'ProsperPaymentsLessThanOneMonthLate',
 'ProsperPaymentsOneMonthPlusLate',
 'ProsperPrincipalBorrowed',
 'ProsperPrincipalOutstanding',
 'ScorexChangeAtTimeOfListing',
 'LoanFirstDefaultedCycleNumber']
```

In [11]:
```
df.drop(extreme_na_cols, axis = 1, inplace = True)
```

In [12]:
```
summary(df)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113937 entries, 0 to 113936
Data columns (total 69 columns):
 #   Column                                Non-Null Count   Dtype
---  ------                                --------------   -----
 0   ListingKey                            113937 non-null  object
 1   ListingNumber                         113937 non-null  int64
 2   ListingCreationDate                   113937 non-null  object
 3   Term                                  113937 non-null  int64
 4   LoanStatus                            113937 non-null  object
 5   BorrowerAPR                           113912 non-null  float64
 6   BorrowerRate                          113937 non-null  float64
 7   LenderYield                           113937 non-null  float64
 8   EstimatedEffectiveYield               84853 non-null   float64
 9   EstimatedLoss                         84853 non-null   float64
 10  EstimatedReturn                       84853 non-null   float64
 11  ProsperRating (numeric)               84853 non-null   float64
 12  ProsperRating (Alpha)                 84853 non-null   object
 13  ProsperScore                          84853 non-null   float64
 14  ListingCategory (numeric)             113937 non-null  int64
 15  BorrowerState                         108422 non-null  object
 16  Occupation                            110349 non-null  object
 17  EmploymentStatus                      111682 non-null  object
 18  EmploymentStatusDuration              106312 non-null  float64
 19  IsBorrowerHomeowner                   113937 non-null  bool
 20  CurrentlyInGroup                      113937 non-null  bool
 21  DateCreditPulled                      113937 non-null  object
 22  CreditScoreRangeLower                 113346 non-null  float64
 23  CreditScoreRangeUpper                 113346 non-null  float64
 24  FirstRecordedCreditLine               113240 non-null  object
 25  CurrentCreditLines                    106333 non-null  float64
 26  OpenCreditLines                       106333 non-null  float64
 27  TotalCreditLinespast7years            113240 non-null  float64
 28  OpenRevolvingAccounts                 113937 non-null  int64
 29  OpenRevolvingMonthlyPayment           113937 non-null  float64
 30  InquiriesLast6Months                  113240 non-null  float64
 31  TotalInquiries                        112778 non-null  float64
 32  CurrentDelinquencies                  113240 non-null  float64
 33  AmountDelinquent                      106315 non-null  float64
 34  DelinquenciesLast7Years               112947 non-null  float64
 35  PublicRecordsLast10Years              113240 non-null  float64
 36  PublicRecordsLast12Months             106333 non-null  float64
 37  RevolvingCreditBalance                106333 non-null  float64
 38  BankcardUtilization                   106333 non-null  float64
 39  AvailableBankcardCredit               106393 non-null  float64
 40  TotalTrades                           106393 non-null  float64
 41  TradesNeverDelinquent (percentage)    106393 non-null  float64
 42  TradesOpenedLast6Months               106393 non-null  float64
 43  DebtToIncomeRatio                     105383 non-null  float64
 44  IncomeRange                           113937 non-null  object
 45  IncomeVerifiable                      113937 non-null  bool
 46  StatedMonthlyIncome                   113937 non-null  float64
 47  LoanKey                               113937 non-null  object
 48  LoanCurrentDaysDelinquent             113937 non-null  int64
 49  LoanMonthsSinceOrigination            113937 non-null  int64
 50  LoanNumber                            113937 non-null  int64
 51  LoanOriginalAmount                    113937 non-null  int64
 52  LoanOriginationDate                   113937 non-null  object
 53  LoanOriginationQuarter                113937 non-null  object
 54  MemberKey                             113937 non-null  object
 55  MonthlyLoanPayment                    113937 non-null  float64
```

```
 56  LP_CustomerPayments                  113937 non-null  float64
 57  LP_CustomerPrincipalPayments         113937 non-null  float64
 58  LP_InterestandFees                   113937 non-null  float64
 59  LP_ServiceFees                       113937 non-null  float64
 60  LP_CollectionFees                    113937 non-null  float64
 61  LP_GrossPrincipalLoss                113937 non-null  float64
 62  LP_NetPrincipalLoss                  113937 non-null  float64
 63  LP_NonPrincipalRecoverypayments      113937 non-null  float64
 64  PercentFunded                        113937 non-null  float64
 65  Recommendations                      113937 non-null  int64
 66  InvestmentFromFriendsCount           113937 non-null  int64
 67  InvestmentFromFriendsAmount          113937 non-null  float64
 68  Investors                            113937 non-null  int64
dtypes: bool(3), float64(41), int64(11), object(14)
memory usage: 57.7+ MB
```

Out[12]:

	ListingKey	ListingNumber	ListingCreationDate	Term	LoanStatus	BorrowerAPR	Borrow
40742	4DDC35284087468736A7C12	532281	2011-10-13 06:59:34.123000000	36	Current	0.29254	
34166	CABC35025679023563DFF00	488497	2010-12-20 10:03:06.987000000	36	Completed	0.35858	
30843	33963600316843855B29FAA	1163721	2014-01-27 12:33:57.590000000	36	Current	0.23847	
38246	95B3336522781772549AE49	23730	2006-07-09 11:56:57.800000000	36	Completed	0.22744	
27992	012C3595176392133BD8DD4	1033886	2013-11-25 13:43:30.753000000	60	Current	0.16662	
56963	A8383507680272738CD72B7	492731	2011-02-02 12:05:47.010000000	36	Completed	0.30532	
6436	0873349789238564038 7B03	482624	2010-11-02 15:12:43.773000000	36	Completed	0.07990	
68489	3C953559166100890031F96	642670	2012-09-19 11:26:10.700000000	36	Current	0.35797	
97718	2B3B3583977089847955AAF	832500	2013-07-09 11:59:38.117000000	36	Current	0.21434	
100985	6D2035433087077253BA1B4	571340	2012-03-23 11:11:37.543000000	36	Completed	0.35797	

In [13]:
```
na_cols = [col for col in df.columns if df[col].isna().any() == True]
na_cols
```

Out[13]:
```
['BorrowerAPR',
 'EstimatedEffectiveYield',
 'EstimatedLoss',
 'EstimatedReturn',
 'ProsperRating (numeric)',
 'ProsperRating (Alpha)',
 'ProsperScore',
 'BorrowerState',
 'Occupation',
 'EmploymentStatus',
 'EmploymentStatusDuration',
 'CreditScoreRangeLower',
 'CreditScoreRangeUpper',
 'FirstRecordedCreditLine',
 'CurrentCreditLines',
 'OpenCreditLines',
```

```
 'TotalCreditLinespast7years',
 'InquiriesLast6Months',
 'TotalInquiries',
 'CurrentDelinquencies',
 'AmountDelinquent',
 'DelinquenciesLast7Years',
 'PublicRecordsLast10Years',
 'PublicRecordsLast12Months',
 'RevolvingCreditBalance',
 'BankcardUtilization',
 'AvailableBankcardCredit',
 'TotalTrades',
 'TradesNeverDelinquent (percentage)',
 'TradesOpenedLast6Months',
 'DebtToIncomeRatio']
```

In [14]:
```
for col in na_cols:
    if df[col].dtypes in ["int64", "float64"]:
        print(f"{col}: {df[col].mean()}")
    else:
        print(f"{col}: {df[col].mode()[0]}")
```

```
BorrowerAPR: 0.218827655909788
EstimatedEffectiveYield: 0.16866147490365632
EstimatedLoss: 0.08030585836682703
EstimatedReturn: 0.09606829611209916
ProsperRating (numeric): 4.07224258423391
ProsperRating (Alpha): C
ProsperScore: 5.950066585742402
BorrowerState: CA
Occupation: Other
EmploymentStatus: Employed
EmploymentStatusDuration: 96.07158175934984
CreditScoreRangeLower: 685.5677306653962
CreditScoreRangeUpper: 704.5677306653962
FirstRecordedCreditLine: 1993-12-01 00:00:00
CurrentCreditLines: 10.317192216903502
OpenCreditLines: 9.260163824964968
TotalCreditLinespast7years: 26.75453903214412
InquiriesLast6Months: 1.4350847756976333
TotalInquiries: 5.584404759793577
CurrentDelinquencies: 0.5920522783468739
AmountDelinquent: 984.5070592108357
DelinquenciesLast7Years: 4.154984196127387
PublicRecordsLast10Years: 0.3126457082303073
PublicRecordsLast12Months: 0.015094091204047661
RevolvingCreditBalance: 17598.706751431822
BankcardUtilization: 0.5613086247919373
AvailableBankcardCredit: 11210.225447162877
TotalTrades: 23.230033930803717
TradesNeverDelinquent (percentage): 0.8858971924845054
TradesOpenedLast6Months: 0.8023272207758029
DebtToIncomeRatio: 0.2759466040063403
```

In [15]:
```
for col in na_cols:
    if df[col].dtypes in ["int64", "float64"]:
        fill = df[col].mean()
        df[col].fillna(fill, inplace = True)
    else:
        fill = df[col].mode()[0]
        df[col].fillna(fill, inplace = True)
```

In [16]:
```
df.isna().sum()
```

Out[16]:
```
ListingKey                          0
ListingNumber                       0
```

ListingCreationDate	0
Term	0
LoanStatus	0
BorrowerAPR	0
BorrowerRate	0
LenderYield	0
EstimatedEffectiveYield	0
EstimatedLoss	0
EstimatedReturn	0
ProsperRating (numeric)	0
ProsperRating (Alpha)	0
ProsperScore	0
ListingCategory (numeric)	0
BorrowerState	0
Occupation	0
EmploymentStatus	0
EmploymentStatusDuration	0
IsBorrowerHomeowner	0
CurrentlyInGroup	0
DateCreditPulled	0
CreditScoreRangeLower	0
CreditScoreRangeUpper	0
FirstRecordedCreditLine	0
CurrentCreditLines	0
OpenCreditLines	0
TotalCreditLinespast7years	0
OpenRevolvingAccounts	0
OpenRevolvingMonthlyPayment	0
InquiriesLast6Months	0
TotalInquiries	0
CurrentDelinquencies	0
AmountDelinquent	0
DelinquenciesLast7Years	0
PublicRecordsLast10Years	0
PublicRecordsLast12Months	0
RevolvingCreditBalance	0
BankcardUtilization	0
AvailableBankcardCredit	0
TotalTrades	0
TradesNeverDelinquent (percentage)	0
TradesOpenedLast6Months	0
DebtToIncomeRatio	0
IncomeRange	0
IncomeVerifiable	0
StatedMonthlyIncome	0
LoanKey	0
LoanCurrentDaysDelinquent	0
LoanMonthsSinceOrigination	0
LoanNumber	0
LoanOriginalAmount	0
LoanOriginationDate	0
LoanOriginationQuarter	0
MemberKey	0
MonthlyLoanPayment	0
LP_CustomerPayments	0
LP_CustomerPrincipalPayments	0
LP_InterestandFees	0
LP_ServiceFees	0
LP_CollectionFees	0
LP_GrossPrincipalLoss	0
LP_NetPrincipalLoss	0
LP_NonPrincipalRecoverypayments	0
PercentFunded	0
Recommendations	0
InvestmentFromFriendsCount	0
InvestmentFromFriendsAmount	0

```
Investors                            0
dtype: int64
```

In [17]:
```
summary(df)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113937 entries, 0 to 113936
Data columns (total 69 columns):
 #   Column                               Non-Null Count   Dtype
---  ------                               --------------   -----
 0   ListingKey                           113937 non-null  object
 1   ListingNumber                        113937 non-null  int64
 2   ListingCreationDate                  113937 non-null  object
 3   Term                                 113937 non-null  int64
 4   LoanStatus                           113937 non-null  object
 5   BorrowerAPR                          113937 non-null  float64
 6   BorrowerRate                         113937 non-null  float64
 7   LenderYield                          113937 non-null  float64
 8   EstimatedEffectiveYield              113937 non-null  float64
 9   EstimatedLoss                        113937 non-null  float64
 10  EstimatedReturn                      113937 non-null  float64
 11  ProsperRating (numeric)              113937 non-null  float64
 12  ProsperRating (Alpha)                113937 non-null  object
 13  ProsperScore                         113937 non-null  float64
 14  ListingCategory (numeric)            113937 non-null  int64
 15  BorrowerState                        113937 non-null  object
 16  Occupation                           113937 non-null  object
 17  EmploymentStatus                     113937 non-null  object
 18  EmploymentStatusDuration             113937 non-null  float64
 19  IsBorrowerHomeowner                  113937 non-null  bool
 20  CurrentlyInGroup                     113937 non-null  bool
 21  DateCreditPulled                     113937 non-null  object
 22  CreditScoreRangeLower                113937 non-null  float64
 23  CreditScoreRangeUpper                113937 non-null  float64
 24  FirstRecordedCreditLine              113937 non-null  object
 25  CurrentCreditLines                   113937 non-null  float64
 26  OpenCreditLines                      113937 non-null  float64
 27  TotalCreditLinespast7years           113937 non-null  float64
 28  OpenRevolvingAccounts                113937 non-null  int64
 29  OpenRevolvingMonthlyPayment          113937 non-null  float64
 30  InquiriesLast6Months                 113937 non-null  float64
 31  TotalInquiries                       113937 non-null  float64
 32  CurrentDelinquencies                 113937 non-null  float64
 33  AmountDelinquent                     113937 non-null  float64
 34  DelinquenciesLast7Years              113937 non-null  float64
 35  PublicRecordsLast10Years             113937 non-null  float64
 36  PublicRecordsLast12Months            113937 non-null  float64
 37  RevolvingCreditBalance               113937 non-null  float64
 38  BankcardUtilization                  113937 non-null  float64
 39  AvailableBankcardCredit              113937 non-null  float64
 40  TotalTrades                          113937 non-null  float64
 41  TradesNeverDelinquent (percentage)   113937 non-null  float64
 42  TradesOpenedLast6Months              113937 non-null  float64
 43  DebtToIncomeRatio                    113937 non-null  float64
 44  IncomeRange                          113937 non-null  object
 45  IncomeVerifiable                     113937 non-null  bool
 46  StatedMonthlyIncome                  113937 non-null  float64
 47  LoanKey                              113937 non-null  object
 48  LoanCurrentDaysDelinquent            113937 non-null  int64
 49  LoanMonthsSinceOrigination           113937 non-null  int64
 50  LoanNumber                           113937 non-null  int64
 51  LoanOriginalAmount                   113937 non-null  int64
 52  LoanOriginationDate                  113937 non-null  object
 53  LoanOriginationQuarter               113937 non-null  object
 54  MemberKey                            113937 non-null  object
 55  MonthlyLoanPayment                   113937 non-null  float64
```

```
 56  LP_CustomerPayments                  113937 non-null  float64
 57  LP_CustomerPrincipalPayments         113937 non-null  float64
 58  LP_InterestandFees                   113937 non-null  float64
 59  LP_ServiceFees                       113937 non-null  float64
 60  LP_CollectionFees                    113937 non-null  float64
 61  LP_GrossPrincipalLoss                113937 non-null  float64
 62  LP_NetPrincipalLoss                  113937 non-null  float64
 63  LP_NonPrincipalRecoverypayments      113937 non-null  float64
 64  PercentFunded                        113937 non-null  float64
 65  Recommendations                      113937 non-null  int64
 66  InvestmentFromFriendsCount           113937 non-null  int64
 67  InvestmentFromFriendsAmount          113937 non-null  float64
 68  Investors                            113937 non-null  int64
dtypes: bool(3), float64(41), int64(11), object(14)
memory usage: 57.7+ MB
```

Out[17]:

	ListingKey	ListingNumber	ListingCreationDate	Term	LoanStatus	BorrowerAPR	Borrow
93789	651D3474085549777F649DC	444720	2010-01-29 02:32:31.687000000	36	Completed	0.23129	
113330	F2EA342008468883025A1AA	331074	2008-05-12 12:50:13.067000000	36	Completed	0.12453	
19998	8AA83583864774354BCDC8E	843139	2013-07-17 12:55:49.563000000	60	Current	0.18136	
15053	98823365845254224EFD03C	18585	2006-06-08 20:57:28.357000000	36	Chargedoff	0.23748	
86955	64CF3528737615098345O8B	535671	2011-10-26 12:25:48.233000000	36	Current	0.35132	
103813	930A3602469426701913500	1156993	2014-02-11 18:06:05.190000000	60	Current	0.20833	
12689	55A03510276261562A73695	499531	2011-03-26 03:15:48.617000000	36	Completed	0.35643	
1046	2E3A3594728148214D8C338	1049019	2013-11-25 12:06:16.250000000	36	Current	0.15223	
63825	C99235311946381591F884F	537841	2011-11-07 12:49:11.427000000	36	Defaulted	0.20200	
63245	D0E335382290039785 4A653	554759	2012-01-27 11:47:14.570000000	36	Current	0.33973	

In [18]:
```
date = ["ListingCreationDate", "DateCreditPulled", "FirstRecordedCreditLine", "LoanOrigi
dtype(date, "date")
```

In [19]:
```
string = ["ProsperRating (numeric)", "ProsperScore", "ListingCategory (numeric)", "LoanN
dtype(string, int)
dtype(string, object)
```

In [20]:
```
integer = ["EmploymentStatusDuration", "CreditScoreRangeLower", "CreditScoreRangeUpper",
           "OpenCreditLines", "TotalCreditLinespast7years", "InquiriesLast6Months", "Tot
           "AmountDelinquent", "DelinquenciesLast7Years", "PublicRecordsLast10Years", "P
           "TradesOpenedLast6Months", ]
dtype(integer, "int64")
```

In [21]:
```
summary(df)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113937 entries, 0 to 113936
Data columns (total 69 columns):
```

```
 #   Column                                Non-Null Count   Dtype
---  ------                                --------------   -----
 0   ListingKey                            113937 non-null  object
 1   ListingNumber                         113937 non-null  int64
 2   ListingCreationDate                   113937 non-null  datetime64[ns]
 3   Term                                  113937 non-null  int64
 4   LoanStatus                            113937 non-null  object
 5   BorrowerAPR                           113937 non-null  float64
 6   BorrowerRate                          113937 non-null  float64
 7   LenderYield                           113937 non-null  float64
 8   EstimatedEffectiveYield               113937 non-null  float64
 9   EstimatedLoss                         113937 non-null  float64
 10  EstimatedReturn                       113937 non-null  float64
 11  ProsperRating (numeric)               113937 non-null  object
 12  ProsperRating (Alpha)                 113937 non-null  object
 13  ProsperScore                          113937 non-null  object
 14  ListingCategory (numeric)             113937 non-null  object
 15  BorrowerState                         113937 non-null  object
 16  Occupation                            113937 non-null  object
 17  EmploymentStatus                      113937 non-null  object
 18  EmploymentStatusDuration              113937 non-null  int64
 19  IsBorrowerHomeowner                   113937 non-null  bool
 20  CurrentlyInGroup                      113937 non-null  bool
 21  DateCreditPulled                      113937 non-null  datetime64[ns]
 22  CreditScoreRangeLower                 113937 non-null  int64
 23  CreditScoreRangeUpper                 113937 non-null  int64
 24  FirstRecordedCreditLine               113937 non-null  datetime64[ns]
 25  CurrentCreditLines                    113937 non-null  int64
 26  OpenCreditLines                       113937 non-null  int64
 27  TotalCreditLinespast7years            113937 non-null  int64
 28  OpenRevolvingAccounts                 113937 non-null  int64
 29  OpenRevolvingMonthlyPayment           113937 non-null  float64
 30  InquiriesLast6Months                  113937 non-null  int64
 31  TotalInquiries                        113937 non-null  int64
 32  CurrentDelinquencies                  113937 non-null  int64
 33  AmountDelinquent                      113937 non-null  int64
 34  DelinquenciesLast7Years               113937 non-null  int64
 35  PublicRecordsLast10Years              113937 non-null  int64
 36  PublicRecordsLast12Months             113937 non-null  int64
 37  RevolvingCreditBalance                113937 non-null  float64
 38  BankcardUtilization                   113937 non-null  float64
 39  AvailableBankcardCredit               113937 non-null  float64
 40  TotalTrades                           113937 non-null  int64
 41  TradesNeverDelinquent (percentage)    113937 non-null  float64
 42  TradesOpenedLast6Months               113937 non-null  int64
 43  DebtToIncomeRatio                     113937 non-null  float64
 44  IncomeRange                           113937 non-null  object
 45  IncomeVerifiable                      113937 non-null  bool
 46  StatedMonthlyIncome                   113937 non-null  float64
 47  LoanKey                               113937 non-null  object
 48  LoanCurrentDaysDelinquent             113937 non-null  int64
 49  LoanMonthsSinceOrigination            113937 non-null  int64
 50  LoanNumber                            113937 non-null  object
 51  LoanOriginalAmount                    113937 non-null  int64
 52  LoanOriginationDate                   113937 non-null  datetime64[ns]
 53  LoanOriginationQuarter                113937 non-null  object
 54  MemberKey                             113937 non-null  object
 55  MonthlyLoanPayment                    113937 non-null  float64
 56  LP_CustomerPayments                   113937 non-null  float64
 57  LP_CustomerPrincipalPayments          113937 non-null  float64
 58  LP_InterestandFees                    113937 non-null  float64
 59  LP_ServiceFees                        113937 non-null  float64
 60  LP_CollectionFees                     113937 non-null  float64
 61  LP_GrossPrincipalLoss                 113937 non-null  float64
 62  LP_NetPrincipalLoss                   113937 non-null  float64
 63  LP_NonPrincipalRecoverypayments       113937 non-null  float64
```

```
 64  PercentFunded                        113937 non-null  float64
 65  Recommendations                      113937 non-null  int64
 66  InvestmentFromFriendsCount           113937 non-null  int64
 67  InvestmentFromFriendsAmount          113937 non-null  float64
 68  Investors                            113937 non-null  int64
dtypes: bool(3), datetime64[ns](4), float64(24), int64(24), object(14)
memory usage: 57.7+ MB
```

Out[21]:

	ListingKey	ListingNumber	ListingCreationDate	Term	LoanStatus	BorrowerAPR	Borrowe
13511	48283589814944764628 3FC	948438	2013-10-02 10:37:45.763	60	Current	0.21935	
91676	E8583593534478326243 1FB	962341	2013-10-25 19:26:12.750	60	Current	0.18965	
59252	73E435045158005448FEAA8	489211	2010-12-28 07:45:04.970	36	Completed	0.30532	
88015	221435529874669307B14E1	615762	2012-07-24 11:10:07.323	36	Completed	0.35797	
75641	94253552957062952CEF968	616566	2012-07-26 07:38:06.893	36	Current	0.23939	
6953	A33835968624159508C364B	1083420	2013-12-12 13:53:10.487	60	Current	0.16662	
99219	78873415204688165F74BEA	291397	2008-03-10 18:47:19.037	36	Completed	0.18292	
62594	30C63593899676835F736CF	995246	2013-11-04 09:09:21.770	36	Current	0.33215	
55098	5B5335254847543381E8249	526989	2011-09-13 10:29:58.620	36	Current	0.30532	
108750	EF1634317391165018 34B4C	393542	2008-09-05 12:08:46.880	36	Completed	0.15629	

In [22]:
```
df.describe(include = "all", datetime_is_numeric=True)
```

Out[22]:

	ListingKey	ListingNumber	ListingCreationDate	Term	LoanStatus	BorrowerAPR
count	113937	1.139370e+05	113937	113937.000000	113937	113937.000000
unique	113066	NaN	NaN	NaN	12	NaN
top	17A93590655669644DB4C06	NaN	NaN	NaN	Current	NaN
freq	6	NaN	NaN	NaN	56576	NaN
mean	NaN	6.278857e+05	2011-07-09 08:07:23.491206656	40.830248	NaN	0.218828
min	NaN	4.000000e+00	2005-11-09 20:44:28.847000	12.000000	NaN	0.006530
25%	NaN	4.009190e+05	2008-09-19 10:02:14.769999872	36.000000	NaN	0.156290
50%	NaN	6.005540e+05	2012-06-16 12:37:19.990000128	36.000000	NaN	0.209790
75%	NaN	8.926340e+05	2013-09-09 19:40:48.963000064	36.000000	NaN	0.283700

max	NaN	1.255725e+06	2014-03-10 12:20:53.760000	60.000000	NaN	0.512290
std	NaN	3.280762e+05	NaN	10.436212	NaN	0.080355

In [23]:
```
df.duplicated().sum()
```

Out[23]:
```
0
```

In [24]:
```
df.to_csv("modified.csv", index = False)
```

Project Info

Dataset Structure

> The dataset comprises of 113937 rows and 81 columns initially with 3 boolean columns, 61 numeric columns and 14 object columns but there are a lot of columns with too many missing values and incorrect data types. After investigatig those columns, I found that they were unique and needed but with no way to get them, I decided to drop those columns with extreme missing values (i.e >=58848 missing values) and fill the remaining. After that the new dataframe comprised of 113937 rows ands 69 columns with 3 boolean columns, 4 datetime columns, 48 numeric columns and 14 object columns.

Features of Interest

> - `BorrowerAPR`
> - `ProsperRating (numeric)`
> - `ProsperRating (Alpha)`
> - `ProsperScore`

Support Features

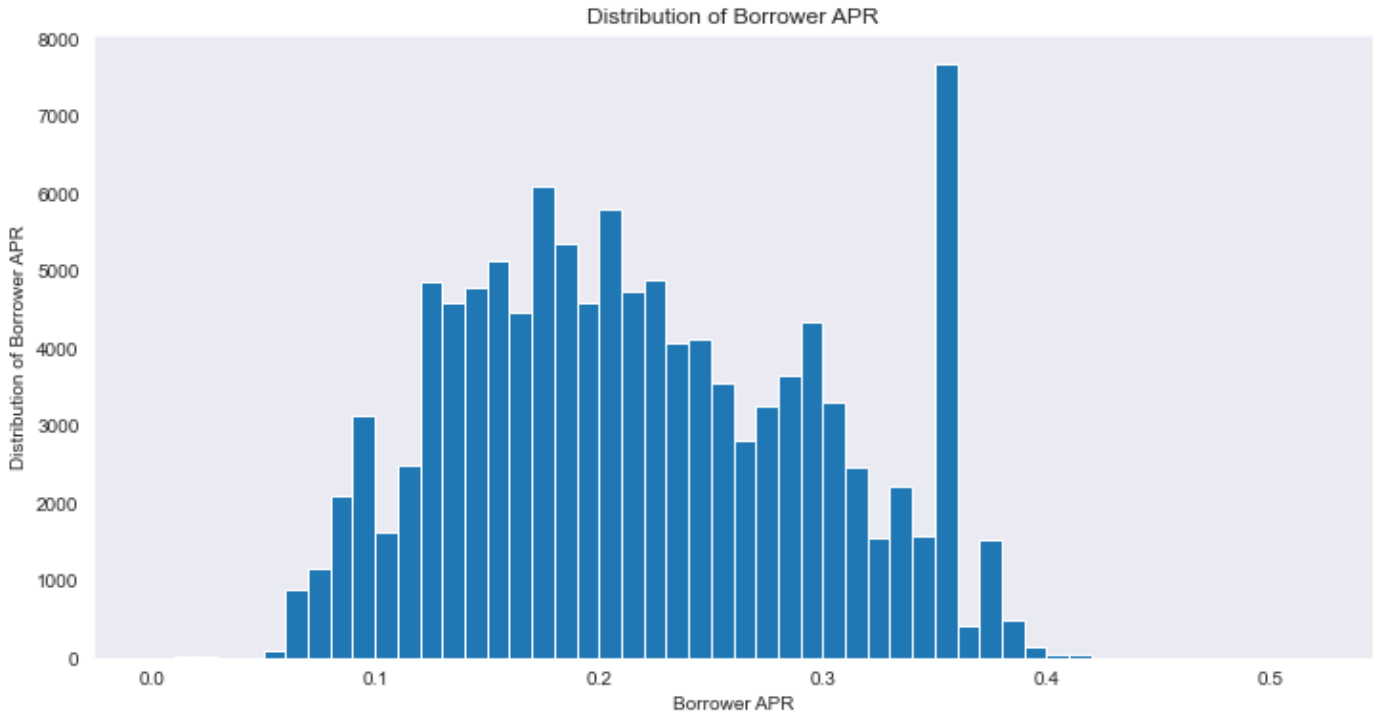
> - `EmploymentStatus`
> - `IsBorrowerHomeowner`
> - `IncomeVerifiable`
> - `LoanOriginalAmount`
> - `TotalTrades`

Univariate Exploration

In [25]:
```
e = np.arange(0, df["BorrowerAPR"].max() + 0.01, 0.01)
e
```

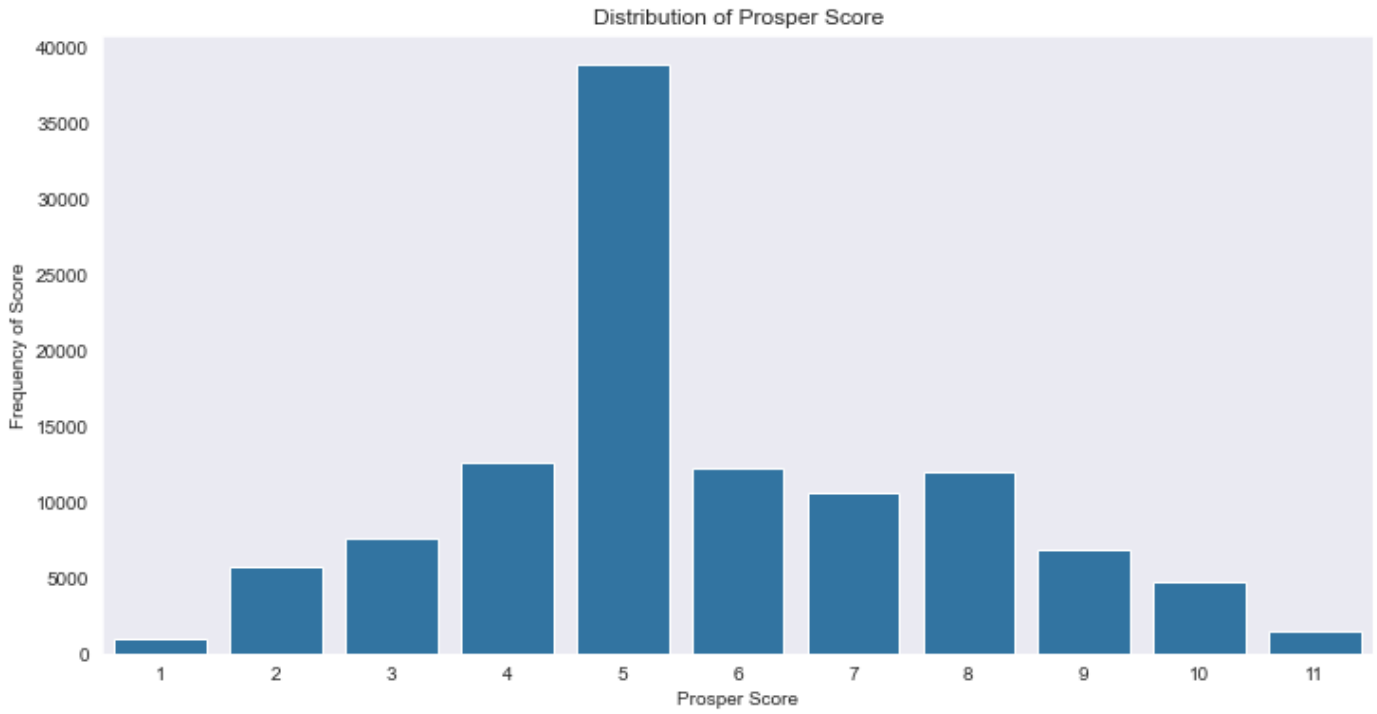
Out[25]:
```
array([0.  , 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1 ,
       0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2 , 0.21,
       0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3 , 0.31, 0.32,
       0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4 , 0.41, 0.42, 0.43,
       0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.5 , 0.51, 0.52])
```

In [26]:
```
plt.figure(figsize = (12, 6))
plt.hist(data = df, x = "BorrowerAPR", bins = e)

plt.title("Distribution of Borrower APR")
plt.xlabel("Borrower APR")
plt.ylabel("Distribution of Borrower APR")

plt.show()
```



> **The distribution of `BorrowerAPR` looks normal but with a spike in value at around 0.35 - 0.37**

In [27]:
```
plt.figure(figsize = (12, 6))
sb.countplot(data = df, x = "ProsperScore", color = sb.color_palette()[0])

plt.title("Distribution of Prosper Score")
plt.xlabel("Prosper Score")
plt.ylabel("Frequency of Score")

plt.show()
```

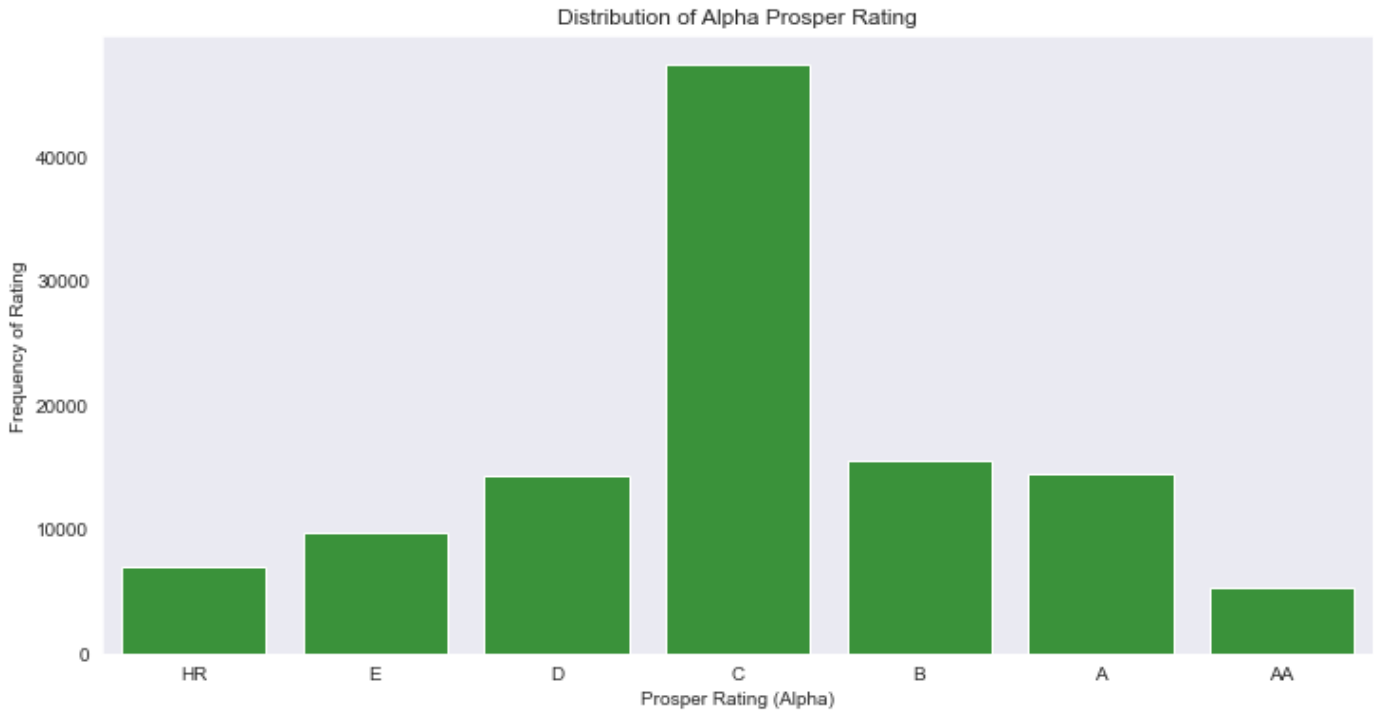


In [28]:
```
plt.figure(figsize = (12, 6))
sb.countplot(data = df, x = "ProsperRating (numeric)", color = sb.color_palette()[1])

plt.title("Distribution of Numeric Prosper Rating")
plt.xlabel("Prosper Rating (Numeric)")
plt.ylabel("Frequency of Rating")

plt.show()
```

In [29]:
```
order = ["HR", "E", "D", "C", "B", "A", "AA"]

# For reproducability
pd_ver = pd.__version__.split(".")

if (int(pd_ver[0]) > 0) or (int(pd_ver[1]) >= 21):
    rating = pd.api.types.CategoricalDtype(ordered = True, categories = order)
    df["ProsperRating (Alpha)"] = df["ProsperRating (Alpha)"].astype(rating)
```

```
else:
    df["ProsperRating (Alpha)"] = df["ProsperRating (Alpha)"].astype("category", ordered
```

In [30]:
```
plt.figure(figsize = (12, 6))
sb.countplot(data = df, x = "ProsperRating (Alpha)", color = sb.color_palette()[2])

plt.title("Distribution of Alpha Prosper Rating")
plt.xlabel("Prosper Rating (Alpha)")
plt.ylabel("Frequency of Rating")

plt.show()
```

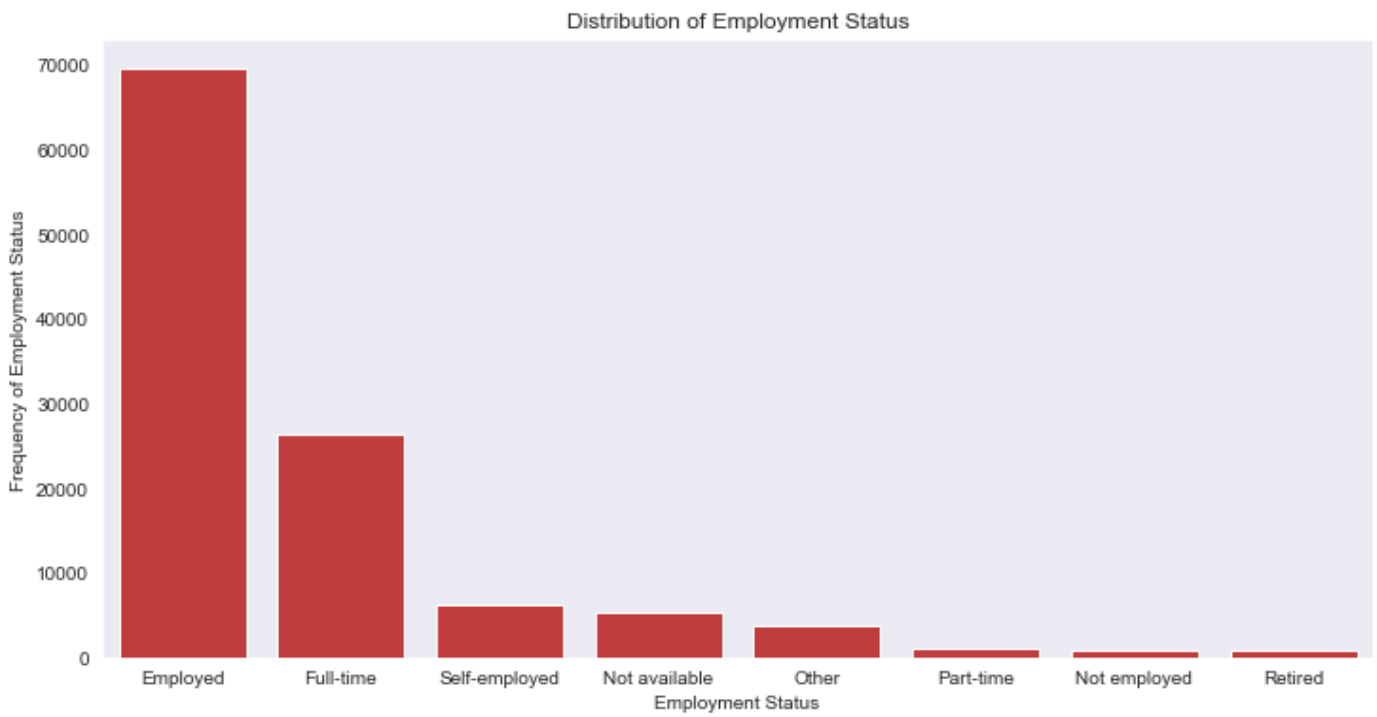


> **From the visualizations above, I observed that the `ProsperRating (numeric)` and `ProsperRating (Alpha)` are infact the same and the only difference was how they were represented. The frequency of these two increased as the ratings until they got to the modal rating and then began to fall. `ProsperScore` observed a similar pattern.**

In [31]:
```
arr = df["EmploymentStatus"].value_counts().index
df["EmploymentStatus"].value_counts()
```

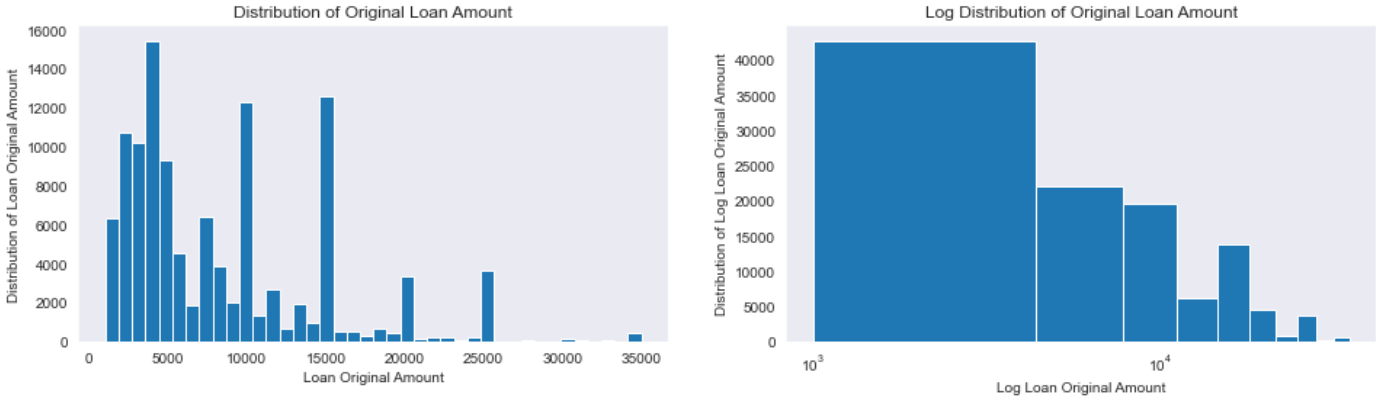
Out[31]:
```
Employed         69577
Full-time        26355
Self-employed     6134
Not available     5347
Other             3806
Part-time         1088
Not employed       835
Retired            795
Name: EmploymentStatus, dtype: int64
```

In [32]:
```
plt.figure(figsize = (12, 6))
sb.countplot(data = df, x = "EmploymentStatus", color = sb.color_palette()[3], order = a

plt.title("Distribution of Employment Status")
plt.xlabel("Employment Status")
plt.ylabel("Frequency of Employment Status")

plt.show()
```



> **The distribution of `EmploymentStatus` showed that Employed people tend to apply for loan more than others. It may be argued that this is because the missing values of those features were filled with their mode but that just implies that they were alreadng leading results in said features**

In [33]:
```
plt.figure(figsize = (16, 4))

plt.subplot(1, 2, 1)
plt.hist(data = df, x = "LoanOriginalAmount", bins = 40)

plt.title("Distribution of Original Loan Amount")
plt.xlabel("Loan Original Amount")
plt.ylabel("Distribution of Loan Original Amount")

plt.subplot(1, 2, 2)
plt.hist(data = df, x = "LoanOriginalAmount")

plt.title("Log Distribution of Original Loan Amount")
plt.xscale("log")
plt.xlabel("Log Loan Original Amount")
plt.ylabel("Distribution of Log Loan Original Amount")

plt.show()
```

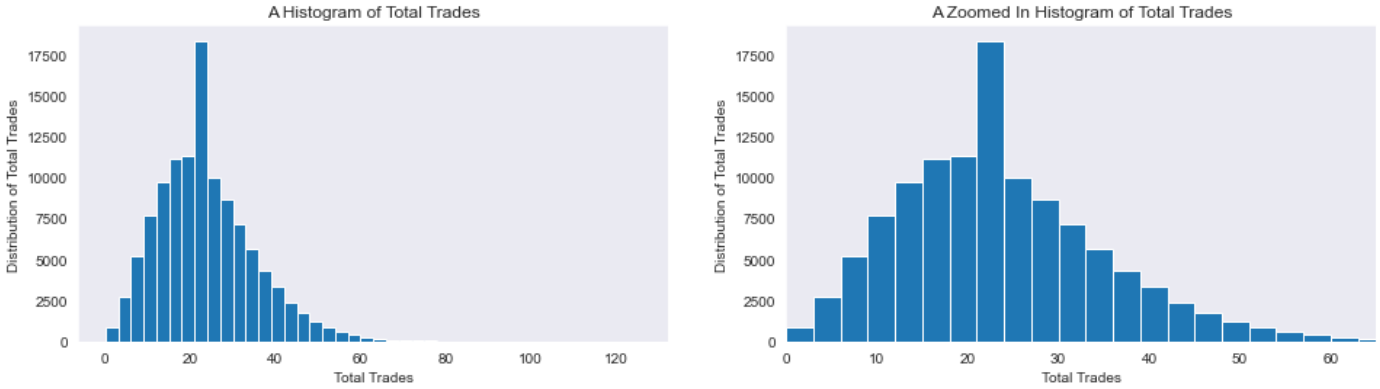


> **The distribution of `LoanOriginalAmount` is not skewed. Even with a log transformation, there is no**

change. But most people had an origination amount of about 4000 - 5000

In [34]:
```
bin_edge = np.arange(0, df["TotalTrades"].max() + 3, 3)
```

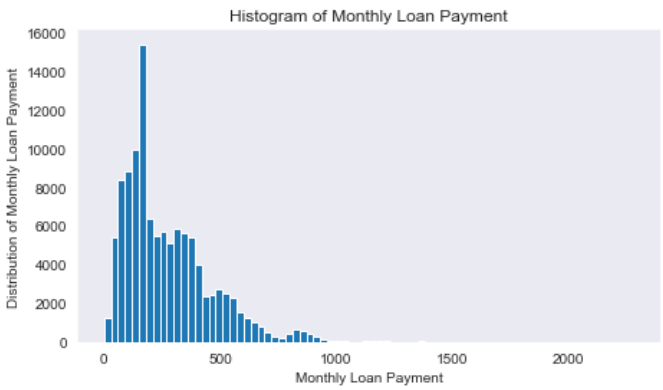
In [35]:
```
plt.figure(figsize = (16, 4))

plt.subplot(1, 2, 1)
plt.title("A Histogram of Total Trades")
plt.hist(data = df, x = "TotalTrades", bins = bin_edge)

plt.xlabel("Total Trades")
plt.ylabel("Distribution of Total Trades")

plt.subplot(1, 2, 2)
plt.title("A Zoomed In Histogram of Total Trades")
plt.hist(data = df, x = "TotalTrades", bins = bin_edge)

plt.xlabel("Total Trades")
plt.xlim(0, 65)
plt.ylabel("Distribution of Total Trades")

plt.show()
```

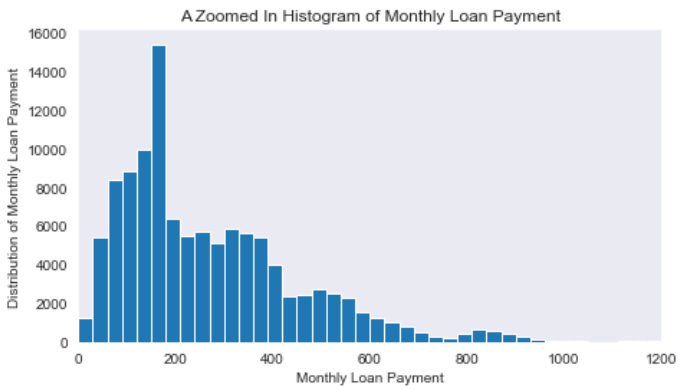


While the distribution of `TotalTrades` was skewed to the right, it still showed that most people tended to have trade lines of between 20 - 25.

In [36]:
```
edges = np.arange(0, df["MonthlyLoanPayment"].max() + 30, 30)
```

In [37]:
```
plt.figure(figsize = (16, 4))

plt.subplot(1, 2, 1)
plt.hist(data = df, x = "MonthlyLoanPayment", bins = edges)

plt.title("Histogram of Monthly Loan Payment")
plt.xlabel("Monthly Loan Payment")
plt.ylabel("Distribution of Monthly Loan Payment")

plt.subplot(1, 2, 2)
plt.title("A Zoomed In Histogram of Monthly Loan Payment")
plt.hist(data = df, x = "MonthlyLoanPayment", bins = edges)

plt.xlabel("Monthly Loan Payment")
plt.xlim(0, 1200)
plt.ylabel("Distribution of Monthly Loan Payment")

plt.show()
```





A closer look at the distribution of `MonthlyLoanPayment` showed that most people were expected to pay around 160 - 230 dollars.

Bivariate Exploration

In [38]:
```
plt.figure(figsize = (12, 6))
sb.countplot(data = df, x = "ProsperScore", hue = "IncomeVerifiable")

plt.title("Relationship Between Prosper Score and Income Verifiabe")
plt.xlabel("Prosper Score")
plt.ylabel("Frequency ")

plt.legend(title = "Income Verifiable")

plt.show()
```

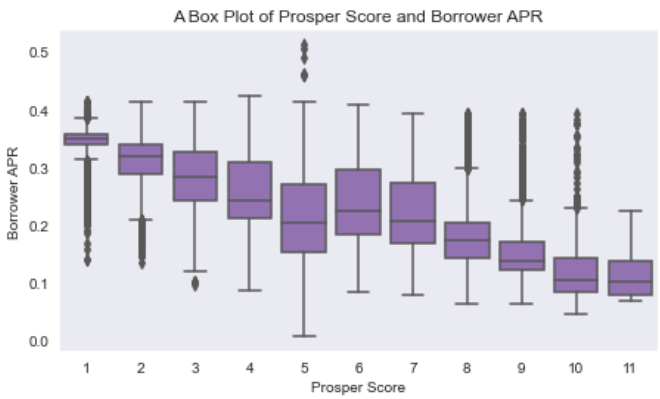
The above visulization shows the relationship between the Prosper Score and Income Verifiable. It shows that verification of income plays an important role in the Prosper Score. Prosper Score `10` and `11`, which are the highest, are only applied to those who have verifiable sources of income. But Prosper Score `5`, which is the modal Score, has the most people with verified sources of income

In [39]:
```
plt.figure(figsize = (16, 4))

plt.subplot(1, 2, 1)
```

```
sb.countplot(data = df, x = "ProsperRating (numeric)", hue = "EmploymentStatus", hue_ord

plt.title("Relationship Between Prosper Rating and Employment Status")
plt.xlabel("Numeric Prosper Rating")
plt.ylabel("Frequency ")

plt.legend(title = "Employment Status")

plt.subplot(1, 2, 2)
sb.countplot(data = df, x = "ProsperScore", hue = "EmploymentStatus", hue_order = arr)

plt.title("Relationship Between Prosper Score and Employment Status")
plt.xlabel("Prosper Score")
plt.ylabel("Frequency ")

plt.legend(title = "Employment Status", loc = "upper right")

plt.show()
```

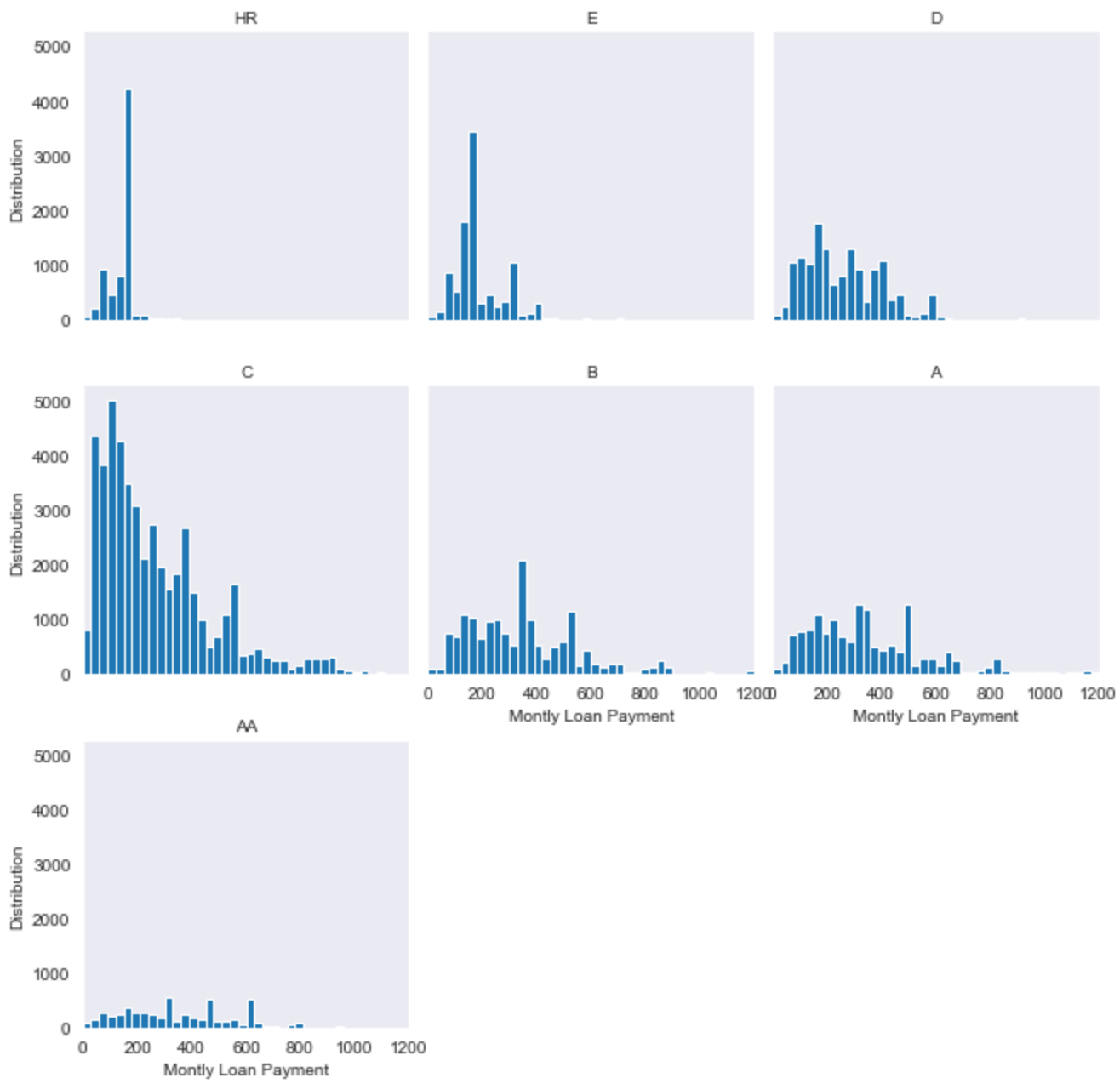
The above visulization shows the relationship between the Numeric Prosper Rating and Prosper Score against Employment Status. It shows that for every Rating and Score, `Employed` people made up most of the consideration. `Not available` and `Part-time` made up the least except in the Rating `4` and Score `5` where `Full-time` makes the highest consideration and `Not available` also made a considerable portion compared to others.

In [40]:
```
c = sb.color_palette()[4]

plt.figure(figsize = (16, 4))

plt.subplot(1, 2, 1)
plt.title("A Violin Plot of Prosper Score and Borrower APR")
sb.violinplot(data = df, x = "ProsperScore", y = "BorrowerAPR", inner = "quartile", colo

plt.xlabel("Prosper Score")
plt.ylabel("Borrower APR")

plt.subplot(1, 2, 2)
plt.title("A Box Plot of Prosper Score and Borrower APR")
sb.boxplot(data = df, x = "ProsperScore", y = "BorrowerAPR", color = c)

plt.xlabel("Prosper Score")
plt.ylabel("Borrower APR")

plt.show()
```



The above visulization shows the relationship between the Prosper Score and Borrower APR. It shows that Prosper Score `5` had a wider range of Borrower APR and the lowest Borrower APR of all the Score. Prosper Score `4` had the highest Borrower APR. Though there were a lot of outliers in Prosper Score `1` and it also had the lowest range of Borrower APR.

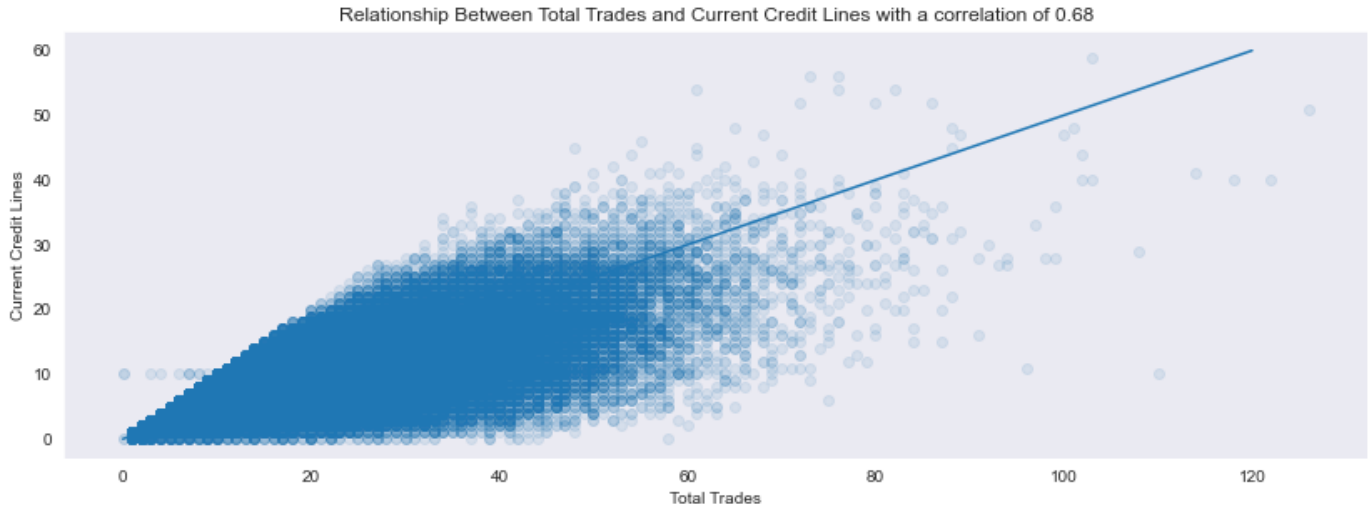
In [41]:
```
g = sb.FacetGrid(data = df, col = "ProsperRating (Alpha)", col_wrap = 3)
g.map(plt.hist, "MonthlyLoanPayment", bins = edges)
g.set_titles("{col_name}")

g.set(xlim = (0, 1200))
g.set_xlabels("Montly Loan Payment")
g.set_ylabels("Distribution")

plt.show()
```

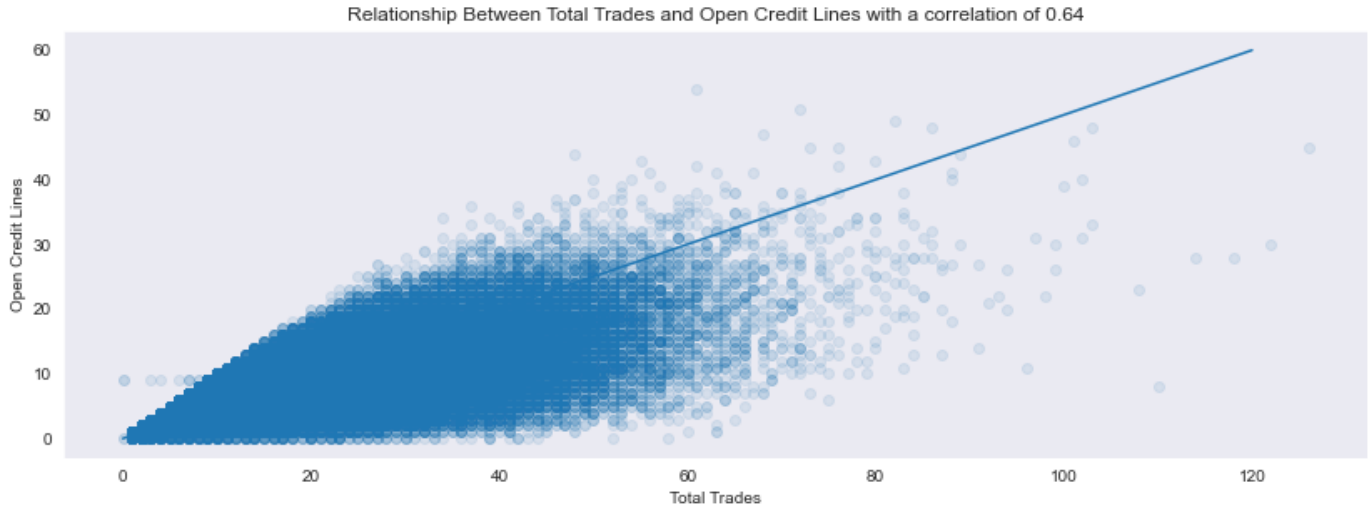


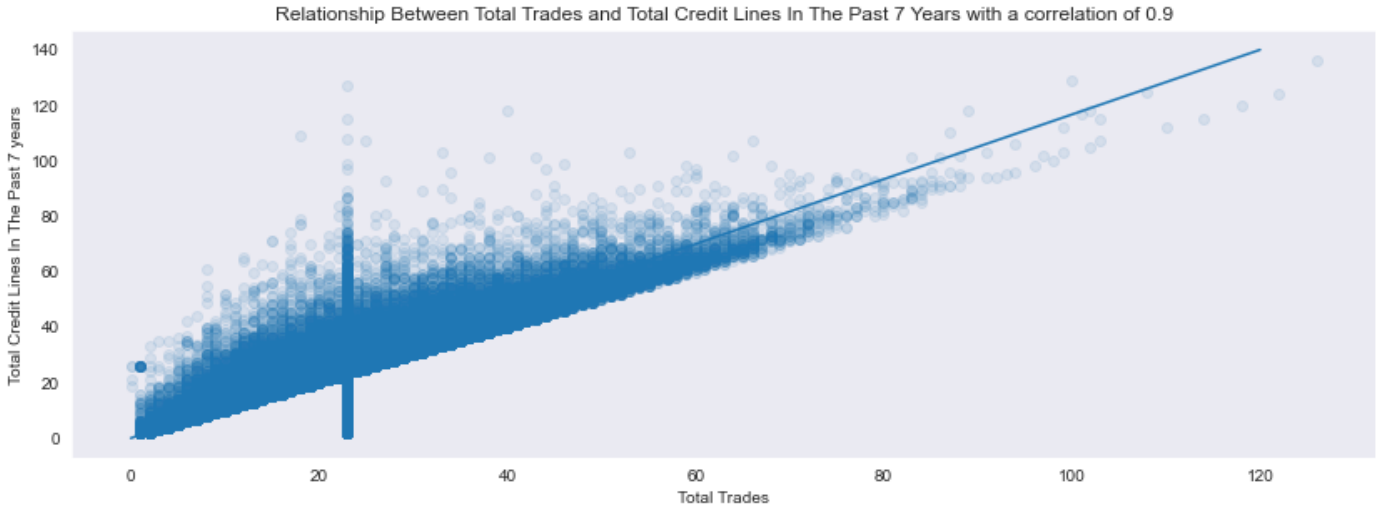
The above visulization shows the distribution of Monthly Loan Payment for each Alphabetic Prosper Rating. It shows that Rating `C` had the most occurence of Monthly Loan Payment.

In [42]:
```
plt.figure(figsize = (14, 16))

plt.subplot(3, 1, 1)
corr = df["TotalTrades"].corr(df["CurrentCreditLines"])

plt.scatter(data = df, x = "TotalTrades", y = "CurrentCreditLines", alpha = 0.1)
plt.plot([0, 120], [0, 60])

plt.title(f"Relationship Between Total Trades and Current Credit Lines with a correlatio
plt.xlabel("Total Trades")
plt.ylabel("Current Credit Lines")

plt.subplot(3, 1, 2)
corr = df["TotalTrades"].corr(df["OpenCreditLines"])

plt.scatter(data = df, x = "TotalTrades", y = "OpenCreditLines", alpha = 0.1)
plt.plot([0, 120], [0, 60])

plt.title(f"Relationship Between Total Trades and Open Credit Lines with a correlation o
plt.xlabel("Total Trades")
```

```
plt.ylabel("Open Credit Lines")

plt.subplot(3, 1, 3)
corr = df["TotalTrades"].corr(df["TotalCreditLinespast7years"])

plt.scatter(data = df, x = "TotalTrades", y = "TotalCreditLinespast7years", alpha = 0.1)
plt.plot([0, 120], [0, 140])

plt.title(f"Relationship Between Total Trades and Total Credit Lines In The Past 7 Years
with a correlation of {round(corr, 2)}")
plt.xlabel("Total Trades")
plt.ylabel("Total Credit Lines In The Past 7 years")

plt.show()
```



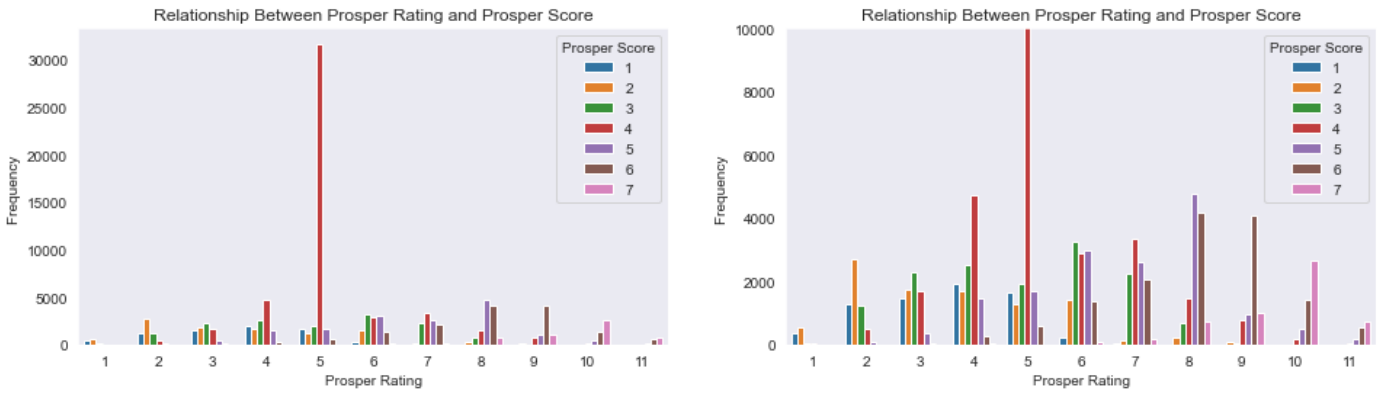




The above visulization shows the relationship between the Total Trades and Current Credit Lines,

Open Credit Lines and Total Credit Lines In The Past 7 Years. It shows that they are all positively correlated with Total Credit Lines In The Past 7 Years having the highest correlation.

In [43]:
```
plt.figure(figsize = (16, 4))

plt.subplot(1, 2, 1)
sb.countplot(data = df, x = "ProsperScore", hue = "ProsperRating (numeric)")

plt.title("Relationship Between Prosper Rating and Prosper Score")
plt.xlabel("Prosper Rating")
plt.ylabel("Frequency ")

plt.legend(title = "Prosper Score", loc = "upper right")

plt.subplot(1, 2, 2)
sb.countplot(data = df, x = "ProsperScore", hue = "ProsperRating (numeric)")

plt.title("Relationship Between Prosper Rating and Prosper Score")
plt.xlabel("Prosper Rating")
plt.ylabel("Frequency ")
plt.ylim(0, 10000)

plt.legend(title = "Prosper Score", loc = "upper right")

plt.show()
```



The above visulization shows the relationship between Prosper Rating and Prosper Score. Each Rating had varying levels of Score but Rating `5` had the highest Score of `4`. I also noticed that Scores `8`, `9`, `10`, `11` had no relationhip with the Ratings.

Relationship Between Main Features

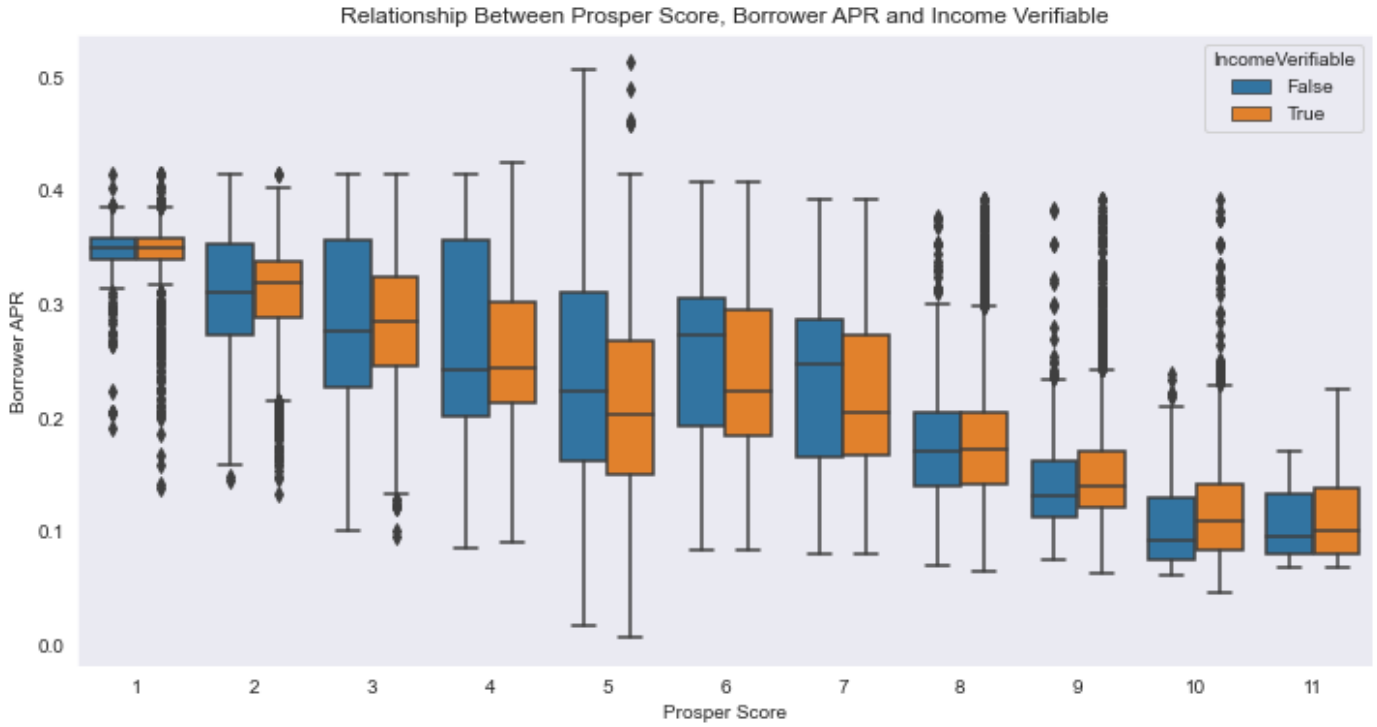
> My main features showed some interesting relationships with themselves and other features. I observed that Prosper Rating influenced one's Prosper Score. The max Borrower APR tended to decrease as Prosper Score increased.

Relationship Between Other Features

> Employed people made up the highest composition of Ratings and also had better Ratings than others as did Income Verification wih Prosper Score. I also observed that the number of Trade Lines a person has positively influences their Credit Lines.

Multivariate Exploration

In [44]:
```
plt.figure(figsize = (12, 6))
sb.boxplot(data = df, x = "ProsperScore", y = "BorrowerAPR", hue = "IncomeVerifiable")

plt.title("Relationship Between Prosper Score, Borrower APR and Income Verifiable")
plt.xlabel("Prosper Score")
plt.ylabel("Borrower APR")

plt.show()
```



Conclusions

- The distribution of BorrowerAPR looks normal but with a spike in value at around 0.35 - 0.37
- From the visualizations above, I observed that the ProsperRating (numeric) and ProsperRating (Alpha) are infact the same and the only difference was how they were represented. The frquency of these two increased as the ratings until they got to the modal rating and then began to fall. ProsperScore observed a similar pattern.
- The distribution of EmploymentStatus showed that Employed people tend to apply for loan more than others. It may be argued that this is because the missing values of those features were filled with their mode but that just implies that they were alreadng leading results in said features
- The distribution of LoanOriginalAmount is not skewed. Even with a log transformation, there is no change. But most people had an origination amount of about 4000 - 5000
- While the distribution of TotalTrades was skewed to the right, it still showed that most people tended to have trade lines of between 20 - 25.
- A closer look at the distribution of MonthlyLoanPayment showed that most people were expected to pay around 160 - 230 dollars.
- The relationship between the Prosper Score and Income Verifiable shows that verification of income plays an important role in the Prosper Score. Prosper Score 10 and 11, which are the highest, are only applied to those who have verifiable sources of income. But Prosper Score 5, which is the modal Score, has the most people with verified sources of income.

- The relationship between the Numeric Prosper Rating and Prosper Score against Employment Status. It shows that for every Rating and Score, Employed people made up most of the consideration. Not available and Part-time made up the least except in the Rating 4 and Score 5 where Full-time makes the highest consideration and Not available also made a considerable portion compared to others.
- The relationship between the Prosper Score and Borrower APR that Prosper Score 5 had a wider range of Borrower APR and the lowest Borrower APR of all the Score. Prosper Score 4 had the highest Borrower APR. Though there were a lot of outliers in Prosper Score 1 and it also had the lowest range of Borrower APR.
- The distribution of Monthly Loan Payment for each Alphabetic Prosper Rating that Rating C had the most occurence of Monthly Loan Payment.
- The relationship between the Total Trades and Current Credit Lines, Open Credit Lines and Total Credit Lines In The Past 7 Years. It shows that they are all positively correlated with Total Credit Lines In The Past 7 Years having the highest correlation.
- The relationship between Prosper Rating and Prosper Score. Each Rating had varying levels of Score but Rating 5 had the highest Score of 4. I also noticed that Scores 8, 9, 10, 11 had no relationhip with the Ratings.