Communicate Data Findings

Factors that impact the optimum loan for an applicant.

Introduction:

This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others.

The aim of the data analysis is to provide a graphical summary of important features of the data set and predicting the Loan outcome in the dataset.

Import of the needed libraries:

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [2]:

```
df=pd.read_csv("prosperLoanData.csv")
df.head()
```

Out[2]:

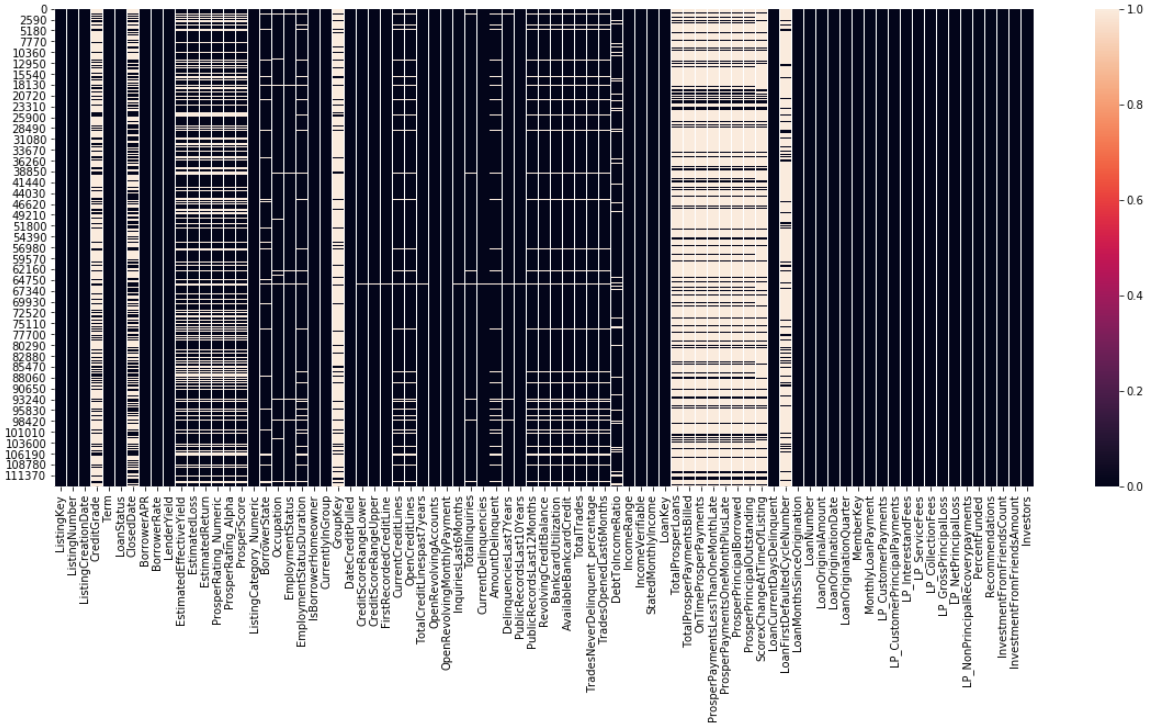
	ListingKey	ListingNumber	ListingCreationDate	CreditGrade	Ter
0	1021339766868145413AB3B	193129	2007-08-26 19:09:29.263000000	C	36
1	1027360249950330 8B223C1	1209647	2014-02-27 08:28:07.900000000	NaN	36
2	0EE9337825851032864889A	81716	2007-01-05 15:00:47.090000000	HR	36
3	0EF5356002482715299901A	658116	2012-10-22 11:02:35.010000000	NaN	36
4	0F023589499656230C5E3E2	909464	2013-09-14 18:38:39.097000000	NaN	36

5 rows × 81 columns

Null Values in dataset

In [8]:

```
fig, ax = plt.subplots(figsize = (20,8))
ax = sns.heatmap(df.isnull(), vmin=0, vmax = 1)
```



Checking for duplicate values

Notes: The dataset contains alot of missing data.

In [9]:

```
df[df.duplicated()]
```

Out[9]:

	ListingKey	ListingNumber	ListingCreationDate	CreditGrade	Term	LoanStatus	Cl

0 rows × 81 columns

Notes: No duplicates present

In [10]:

```
df.rename(columns={'ListingCategory (numeric)' : 'ListingCategory_Numeric','ProsperRati
ng (Alpha)': 'ProsperRating_Alpha', 'ProsperRating (numeric)': 'ProsperRating_Numeric',
'TradesNeverDelinquent (percentage)':'TradesNeverDelinquent_percentage'}, inplace=True)
# Due to many variables in this dataset, I only use few variables. Those are:
column = ['ListingKey' ,'ListingNumber' ,'Term' , 'LoanStatus', 'EstimatedEffectiveYiel
d' ,  'BorrowerAPR' ,
        'BorrowerRate' , 'ProsperRating_Numeric', 'ProsperRating_Alpha', 'ProsperScore'
, 'ListingCategory_Numeric',
        'EmploymentStatus', 'Occupation',  'EmploymentStatusDuration',  'IsBorrowerHome
owner',  'IncomeVerifiable',
        'StatedMonthlyIncome',  'MonthlyLoanPayment', 'Recommendations',  'DebtToIncome
Ratio', 'LoanOriginalAmount' ,
        'PercentFunded', 'IncomeRange', 'Investors', 'BorrowerState']
loan=df[column]
#### Removing null Value Data for better visualisation experience
loan_clean=loan.dropna()
```

In [11]:

```
loan.head()
```

Out[11]:

	ListingKey	ListingNumber	Term	LoanStatus	EstimatedEffective
0	10213397668681454 13AB3B	193129	36	Completed	NaN
1	10273602499503308B223C1	1209647	36	Current	0.07960
2	0EE9337825851032864889A	81716	36	Completed	NaN
3	0EF5356002482715299901A	658116	36	Current	0.08490
4	0F023589499656230C5E3E2	909464	36	Current	0.18316

5 rows × 25 columns

Summary of Exploratory Data

1. Univariate Exploration

.Original Loan amount is right skewed. Most of the loan amount are in range of 0-5000.

.Most of the Loan status is Current. Apart from it we can see alot of completed loan case.

.54.4% of the borrower's are homeowner.

.Most of the borrowers are employed amd full-time worker.

.There are very few people(4.7%) with income less than 24,999.

.Loans with "AA" category are with lesser count

.There are alot of borrower's with prosperScore less than 6.0. Based on this we can say loan was even provided to people .with high risk score.

.Most of the borrower's didn't specify there occupation that was given in form, they wrote Other.

.Apart from that "Professional", "Executive", "Computer Programmer", "Teacher", and "Analyst" are the top 5 most common occupations of the borrower's.

.There are 0.1% borrowers without verfied income.

.50% of borrowerARP is in range of 0.161570 - 0.287800

.The borrowers rate follow an approximately unimodal distribution, with the peak around 0.16

.The monthly income the borrower stated at the time the listing was created. Monthly Income will definitely be right skewed .since very few people will be having a higher salary.

2. Bivariate Exploration

Negative correlation between BorrowerRate and Prosper Score, means lower the prosper rate, higher will be the borrower rate

Higher loan amount is associated with high monthly payment, as they are have strong positive correlation.

Loan original amount and monthly loan payment is highly positive correlated,

Borrowers interest rate and proper score are highly negatively correlated,

Borrower interest rate and Loan original amount are negatively correlated.

Full-time employs have generally high prosper rating than others.

Self-employed people have low prosper rating

Executives, Pharmacist, Doctors and Judge have large loan amount in comparision to other occupants

People with high Income range took high loan amount. Prosper Rating B and C have most number of people who are employed.

Borrowers with salary range are between 50,000− 74,999 are mostly in prosper rating of A, B, C or D.

3. Multivariate Exploration

Based on the Prosper Rating we can conclude that large amount of amount of loan is easily provided to those who have high monthly salary and are employed or full-time worker.

Employment Status And Monthly Salary can be considered few parameter that affected Prosper Rating.

Employed, Self-employed and full-time borrowers generally have high monthly amount who takes loan.

Employed people tends to take larger loan amount in comparision to other Employment status irrespective of their Income Range.

People who are self-employed and have decent salary between 75000− 99999 apply for loan to fulfill their needs.

The non-homeowner tend to have a higher interest rate, and thus lower prosper rating. On the other hand homeowner tends to have lower interest rate and higher prosper rating. So we can conclude that homeowner are the safest bet when gving a loan. We can see that HR prosper rating applicants have higher interest rates

In [12]:

```
loan_clean.describe()
```

Out[12]:

	ListingNumber	Term	EstimatedEffectiveYield	BorrowerAPR	Borrowe
count	7.621600e+04	76216.000000	76216.000000	76216.000000	76216.0(
mean	7.737463e+05	42.760470	0.166738	0.223901	0.19362
std	2.344166e+05	11.744005	0.067595	0.079372	0.07408
min	4.162750e+05	12.000000	-0.181600	0.045830	0.04000(
25%	5.603650e+05	36.000000	0.114800	0.161570	0.13490(
50%	7.425500e+05	36.000000	0.157670	0.215660	0.18450(
75%	9.747142e+05	60.000000	0.219000	0.287800	0.25490(
max	1.255149e+06	60.000000	0.319900	0.423950	0.36000(

Explanatory Data Visualization

Question : Does loan amount impact the borrower interest rate?

In [13]:

```
plt.figure(figsize = [8, 6])
sns.regplot(data = loan_clean, x = 'BorrowerRate', y ="LoanOriginalAmount",scatter_kws
= {'alpha' : 1/100})
plt.title("Loan Amount vs Borrower Interest Rate");
```

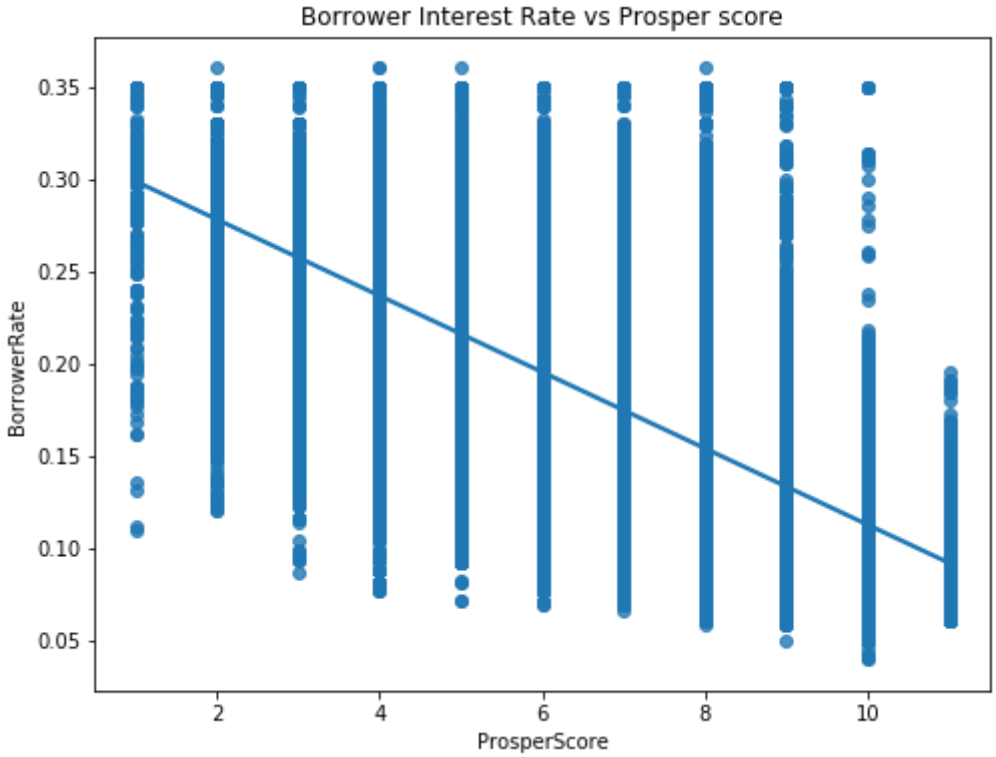
Conclusion:

Borrower interest rate and Loan original amount have moderate negative correlation. Thus we expect that for higher loan amount, the borrower interest rate will be lesser.

Question : What is the impact of prosper score on borrower interest rate?

In [14]:

```
plt.figure(figsize = [8, 6])
sns.regplot(data = loan_clean, x = 'ProsperScore', y ="BorrowerRate")
plt.title("Borrower Interest Rate vs Prosper score");
```



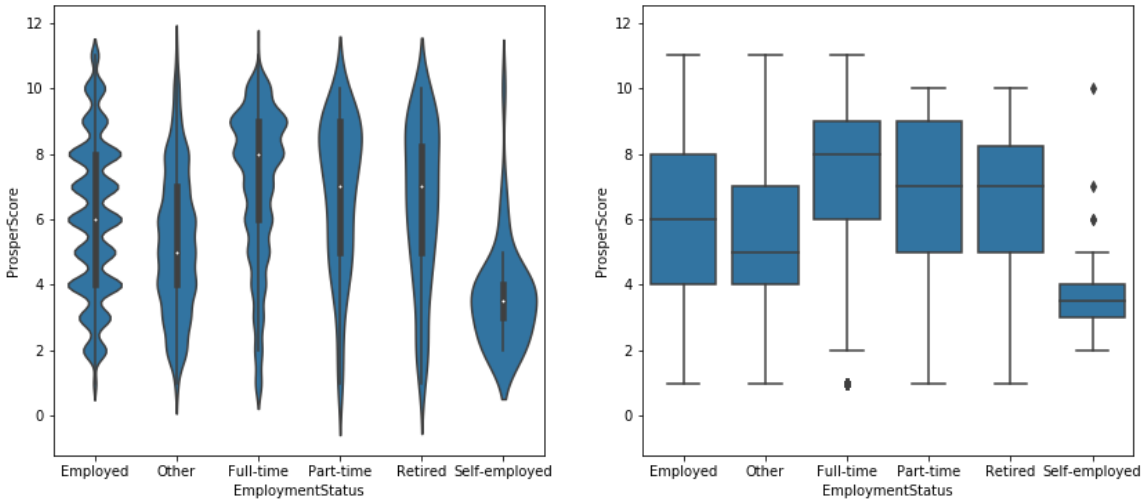
Conclusion:

Borrowers interest rate and proper score are negatively correlated. With lower prosper score, the borrower interest rate tends to increase. Therefore, more the prosper score will be, lesser will be the interest rate.

Question: Can we conclude anything regarding the relationship between ProsperScore and Employment Status?? Does certain Employment status lead to better chance of getting optimum loan ?

In [15]:

```
loan_df= loan_clean[loan_clean['IncomeRange'] != 'Not employed']
plt.figure(figsize = [14, 6])
base_color = sns.color_palette()[0]

# left plot: violin plot
plt.subplot(1, 2, 1)
ax1 = sns.violinplot(data = loan_df, x = 'EmploymentStatus', y = 'ProsperScore', color
= base_color)

# right plot: box plot
plt.subplot(1, 2, 2)
sns.boxplot(data = loan_df, x = 'EmploymentStatus', y = 'ProsperScore', color = base_co
lor)
plt.ylim(ax1.get_ylim()); # set y-axis limits to be same as left plot
```



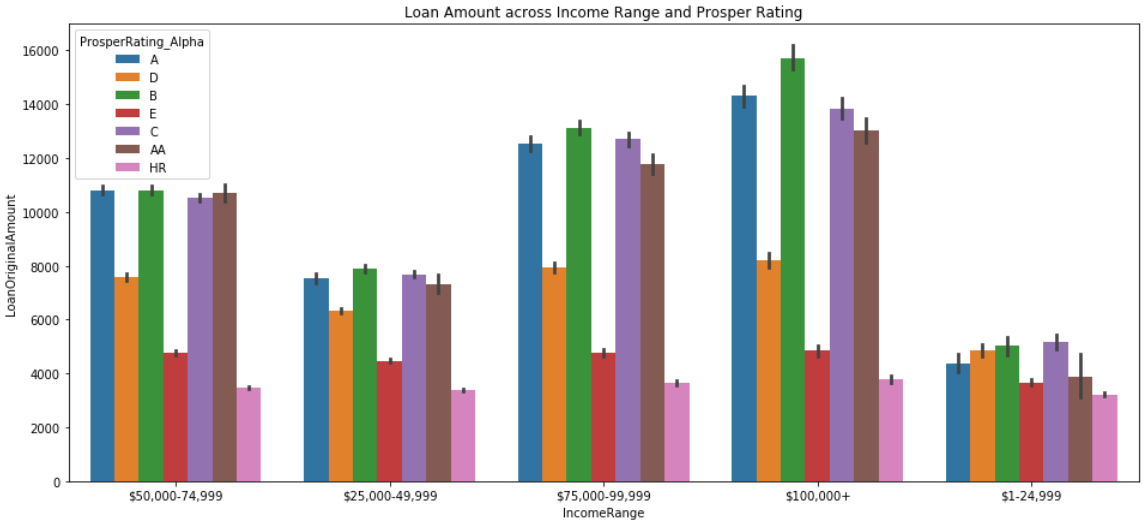
Conslusion:

Full-time employs have generally high prosper rating than others, and have better chance of getting loan easily with lesser borrower interest rate. Self-employed people have low prosper rating, thus it's hard for them to get high loan amount with lesser borrower interest rate.

Question: Can we find any conclusion using prosper rating, loan amount and income range?

In [16]:

```
loan_clean_sub = loan_clean[(loan_clean['StatedMonthlyIncome'] <= 10000) ]
loan_clean_sub = loan_clean_sub[loan_clean_sub['IncomeRange'] != 'Not employed']
loan_clean_sub = loan_clean_sub[loan_clean_sub.EmploymentStatus != 'Not employed']
plt.figure(figsize = [16, 7])
loan_clean_sub = loan_clean_sub[loan_clean_sub['IncomeRange'] != 'Not employed']
ax = sns.barplot(data = loan_clean_sub, x = 'IncomeRange', y = 'LoanOriginalAmount', hu
e = 'ProsperRating_Alpha',order=loan_clean_sub.IncomeRange.value_counts().index)
ax.legend(loc=2,  title = 'ProsperRating_Alpha')
plt.title('Loan Amount across Income Range and Prosper Rating');
```



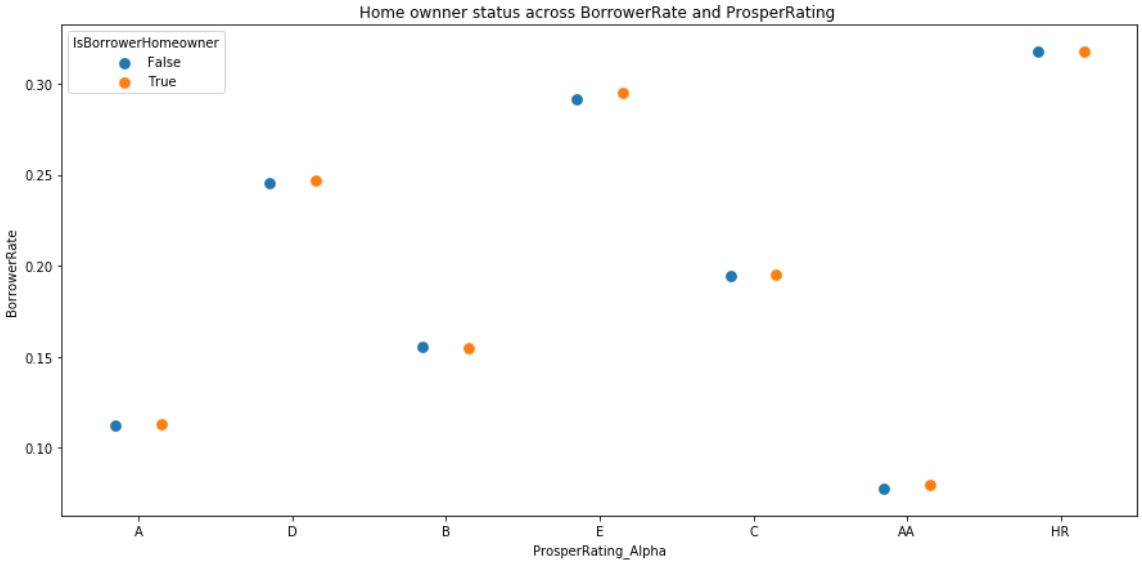
Conclusion:

Large amount of loan is easily provideed to those will high income range and they have generally better prosper rating than other, thus their borrower interest rate would alse be lesser than others.

Question : Does being a Homeownner any impact on Borrower Rate and Prosper Rating ?

In [17]:

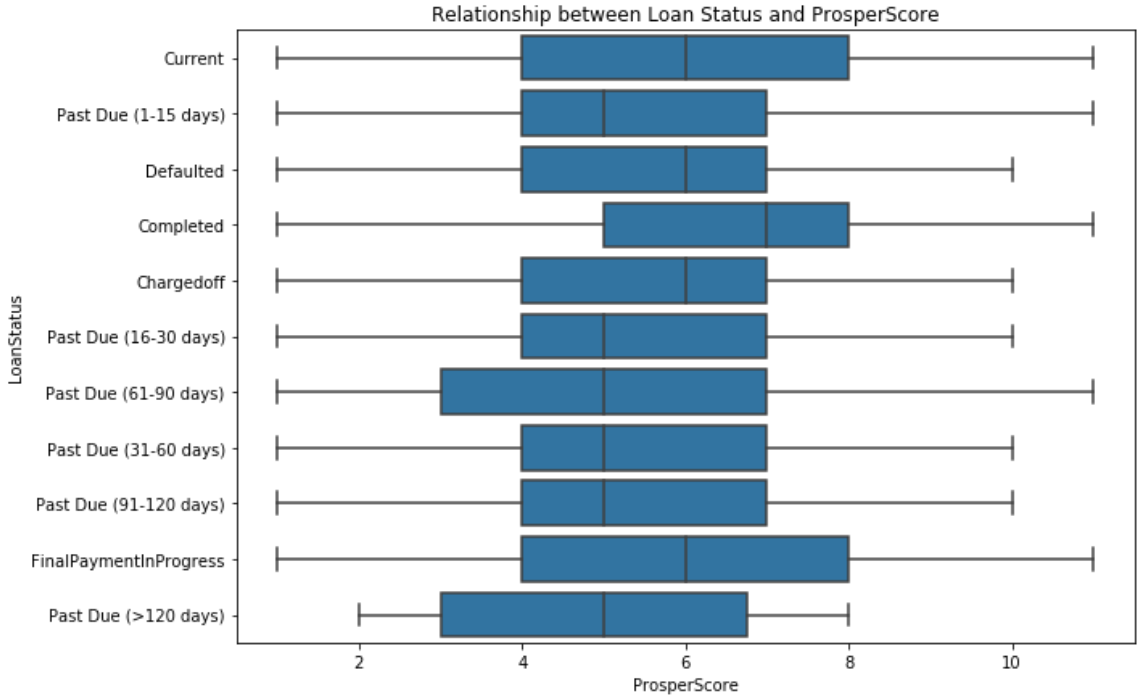
```
plt.figure(figsize = [15, 7])
ax = sns.pointplot(data = loan_clean_sub, x = 'ProsperRating_Alpha', y = 'BorrowerRate'
, hue = 'IsBorrowerHomeowner',dodge = 0.3, linestyles = "");
plt.title('Home ownner status across BorrowerRate and ProsperRating ');
```



Relationship between Loan Status and Prosper Score

In [18]:

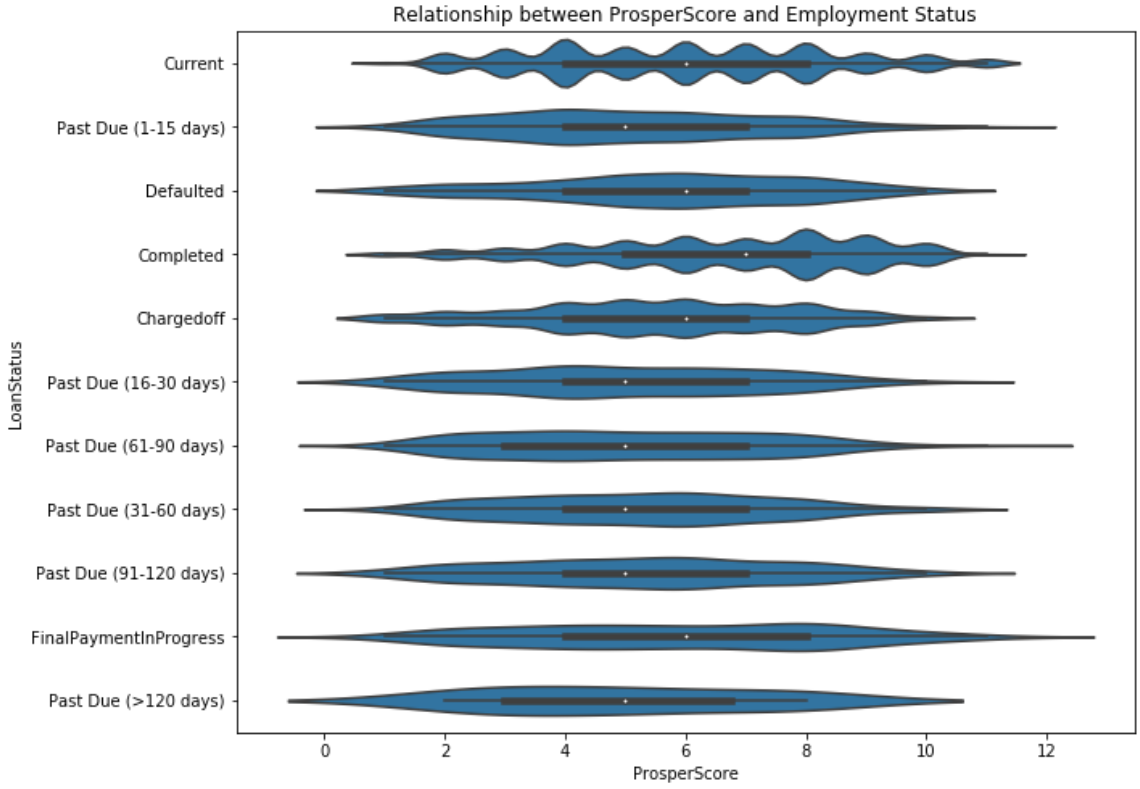
```
plt.figure(figsize = [10, 7])
base_color = sns.color_palette()[0]

sns.boxplot(data = loan_clean, y = 'LoanStatus', x = 'ProsperScore', color = base_color
)
plt.title("Relationship between Loan Status and ProsperScore");
```



Relationship between ProsperScore and Employment Status

In [19]:

```
plt.figure(figsize = [10, 8])
base_color = sns.color_palette()[0]
sns.violinplot(data = loan_clean, y = 'LoanStatus', x = 'ProsperScore', color = base_co
lor)
plt.title("Relationship between ProsperScore and Employment Status");
```

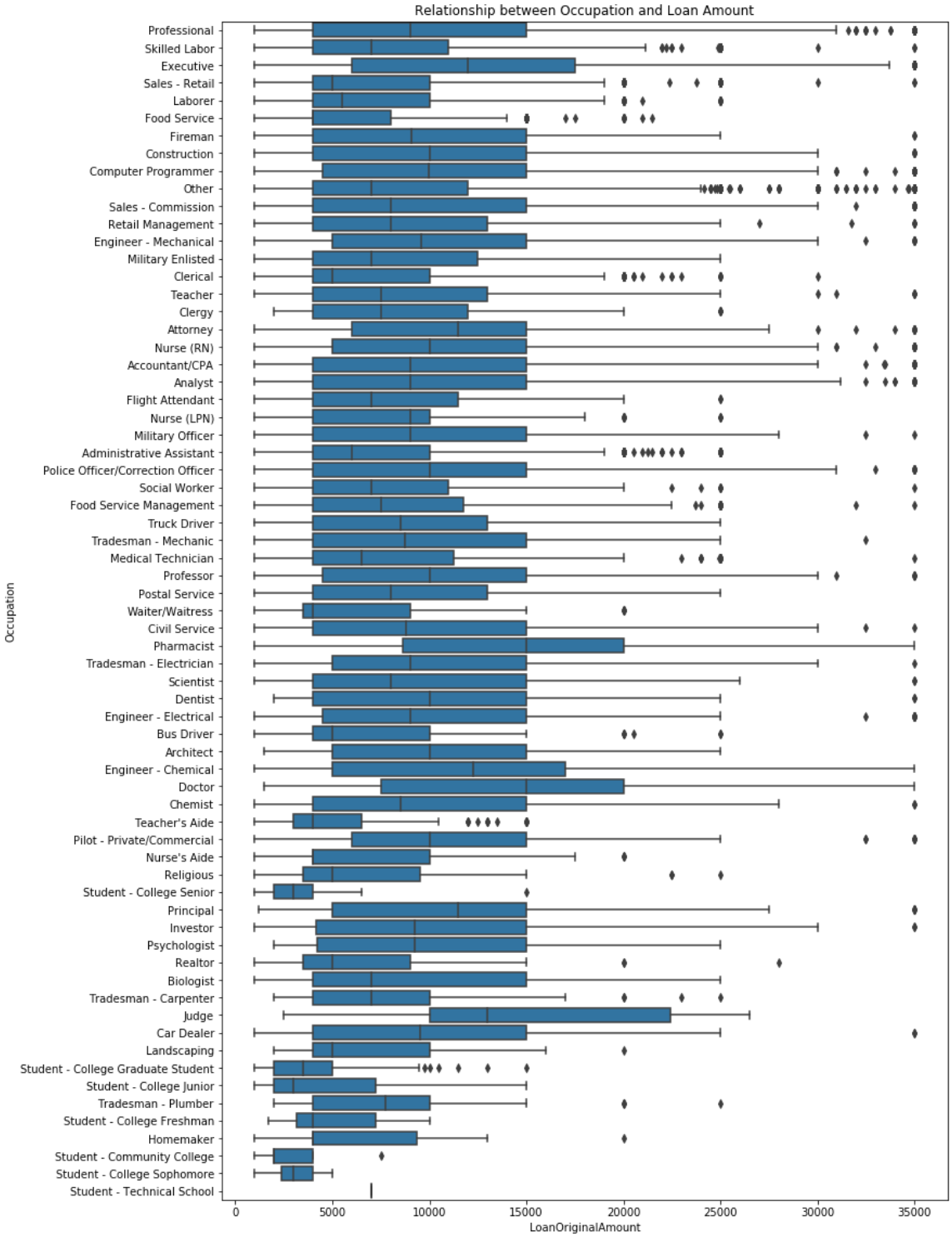


Relationship between Occupation and Loan Amount

Question : Which Occupants generally have high loan amount than others??

In [20]:

```
plt.figure(figsize = [12, 20])
base_color = sns.color_palette()[0]
sns.boxplot(data = loan_clean, y = 'Occupation', x = 'LoanOriginalAmount', color = base
_color) # set y-axis limits to be same as left plot
plt.title("Relationship between Occupation and Loan Amount");
```



Conclusion:

Executives, Pharmacist, Doctors and Judge have large loan amount in comparision to other occupants

Conclusion:

We observe that non-homeowner tend to have a higher interest rate, and thus lower prosper rating. On the other hand homeowner tends to have lower interest rate and higher prosper rating. So we can conclude that homeowner are the safest bet when gving a loan. We can see that HR prosper rating applicants have higher interest rates

Key Insights:

Large amount of loan is easily provideed to those will high income range and they have generally better prosper rating than other, thus their borrower interest rate would alse be lesser than others.

Full-time employs have generally high prosper rating than others, and have better chance of getting loan easily with lesser borrower interest rate.

Self-employed people have low prosper rating, thus it's hard for them to get high loan amount with lesser borrower interest rate. Employment Status And Income Range can be considered few parameter that affected Prosper Rating. We observe that non-homeowner tend to have a higher interest rate, and thus lower prosper rating. On the other hand homeowner tends to have lower interest rate and higher prosper rating. So we can conclude that homeowner are the safest bet when gving a loan. We can see that HR prosper rating applicants have higher interest rates

Borrowers interest rate and proper score are negatively correlated. With lower prosper score, the borrower interest rate tends to increase. Therefore, more the prosper score will be, lesser will be the interest rate.

Borrower interest rate and Loan original amount have moderate negative correlation. Thus we expect that for higher loan amount, the borrower interest rate will be lesser.