In [2]:

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

In [3]:

```
df=pd.read_csv(r'C:\Users\Lenovo\Downloads\new clean')
desc=pd.read_csv(r'C:\Users\Lenovo\Documents\LCDataDictionary.csv')
```

In [4]:

```
df.head()
```

Out[4]:

| | Unnamed: 0 | id | member_id | loan_amnt | funded_amnt | funded_amnt_inv | term | int_rate |
|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 1077501 | 1296599.0 | 5000.0 | 5000.0 | 4975.0 | 36 months | 10.65% |
| **1** | 1 | 1077430 | 1314167.0 | 2500.0 | 2500.0 | 2500.0 | 60 months | 15.27% |
| **2** | 2 | 1077175 | 1313524.0 | 2400.0 | 2400.0 | 2400.0 | 36 months | 15.96% |
| **3** | 3 | 1076863 | 1277178.0 | 10000.0 | 10000.0 | 10000.0 | 36 months | 13.49% |
| **4** | 4 | 1075358 | 1311748.0 | 3000.0 | 3000.0 | 3000.0 | 60 months | 12.69% |

5 rows × 48 columns

In [5]:

```
df.drop('Unnamed: 0',axis=1,inplace=True)
```

In [6]:

```
for i in range(desc.shape[0]):
    if desc.loc[i]['LoanStatNew'] not in df:
        desc.drop(index=i, inplace=True)
```

In [7]:

```
desc.reset_index(drop=True, inplace=True)
```

In [8]:

```
1 desc
```

Out[8]:

| | LoanStatNew | Description |
|---|---|---|
| **0** | addr_state | The state provided by the borrower in the loan... |
| **1** | annual_inc | The self-reported annual income provided by th... |
| **2** | collection_recovery_fee | post charge off collection fee |
| **3** | delinq_2yrs | The number of 30+ days past-due incidences of ... |
| **4** | dti | A ratio calculated using the borrower’s total ... |
| **5** | earliest_cr_line | The month the borrower's earliest reported cre... |
| **6** | emp_length | Employment length in years. Possible values ar... |
| **7** | emp_title | The job title supplied by the Borrower when ap... |
| **8** | fico_range_high | The upper boundary range the borrower’s FICO a... |
| **9** | fico_range_low | The lower boundary range the borrower’s FICO a... |

In [9]:

```
1 df.head()
```

Out[9]:

| | id | member_id | loan_amnt | funded_amnt | funded_amnt_inv | term | int_rate | installmer |
|---|---|---|---|---|---|---|---|---|
| **0** | 1077501 | 1296599.0 | 5000.0 | 5000.0 | 4975.0 | 36 months | 10.65% | 162.8 |
| **1** | 1077430 | 1314167.0 | 2500.0 | 2500.0 | 2500.0 | 60 months | 15.27% | 59.8 |
| **2** | 1077175 | 1313524.0 | 2400.0 | 2400.0 | 2400.0 | 36 months | 15.96% | 84.3 |
| **3** | 1076863 | 1277178.0 | 10000.0 | 10000.0 | 10000.0 | 36 months | 13.49% | 339.3 |
| **4** | 1075358 | 1311748.0 | 3000.0 | 3000.0 | 3000.0 | 60 months | 12.69% | 67.7 |

5 rows × 47 columns

In [10]:

```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42350 entries, 0 to 42349
Data columns (total 47 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   id                        42350 non-null  int64  
 1   member_id                 42350 non-null  float64
 2   loan_amnt                 42350 non-null  float64
 3   funded_amnt               42350 non-null  float64
 4   funded_amnt_inv           42350 non-null  float64
 5   term                      42350 non-null  object 
 6   int_rate                  42350 non-null  object 
 7   installment               42350 non-null  float64
 8   grade                     42350 non-null  object 
 9   sub_grade                 42350 non-null  object 
 10  emp_title                 42350 non-null  object 
 11  emp_length                42350 non-null  object 
 12  home_ownership            42350 non-null  object 
 13  annual_inc                42350 non-null  float64
 14  verification_status       42350 non-null  object 
 15  issue_d                   42350 non-null  object 
 16  loan_status               42350 non-null  object 
 17  purpose                   42350 non-null  object 
 18  title                     42350 non-null  object 
 19  zip_code                  42350 non-null  object 
 20  addr_state                42350 non-null  object 
 21  dti                       42350 non-null  float64
 22  delinq_2yrs               42350 non-null  float64
 23  earliest_cr_line          42350 non-null  object 
 24  fico_range_low            42350 non-null  float64
 25  fico_range_high           42350 non-null  float64
 26  inq_last_6mths            42350 non-null  float64
 27  open_acc                  42350 non-null  float64
 28  pub_rec                   42350 non-null  float64
 29  revol_bal                 42350 non-null  float64
 30  revol_util                42350 non-null  object 
 31  total_acc                 42350 non-null  float64
 32  out_prncp                 42350 non-null  float64
 33  out_prncp_inv             42350 non-null  float64
 34  total_pymnt               42350 non-null  float64
 35  total_pymnt_inv           42350 non-null  float64
 36  total_rec_prncp           42350 non-null  float64
 37  total_rec_int             42350 non-null  float64
 38  total_rec_late_fee        42350 non-null  float64
 39  recoveries                42350 non-null  float64
 40  collection_recovery_fee   42350 non-null  float64
 41  last_pymnt_d              42350 non-null  object 
 42  last_pymnt_amnt           42350 non-null  float64
 43  last_credit_pull_d        42350 non-null  object 
 44  last_fico_range_high      42350 non-null  float64
 45  last_fico_range_low       42350 non-null  float64
 46  pub_rec_bankruptcies      42350 non-null  float64
dtypes: float64(28), int64(1), object(18)
memory usage: 15.2+ MB
```

In [ ]:

```
1
```

In [11]:

```
1 df['id'].dtype
```

Out[11]:

```
dtype('int64')
```

In [12]:

```
1 def int_rate_clean(i):
2     return (i[:-1])
```

In [13]:

```
1 df['int_rate']=df['int_rate'].apply(int_rate_clean).astype(float)
```

In [14]:

```
1 def term_clean(i):
2     return (i[0:3])
```

In [15]:

```
1 df['term']=df['term'].apply(term_clean).astype(int)
```

In [16]:

```
1 def revol_util_clean(i):
2     return (i[:-1])
```

In [17]:

```
1 df['revol_util']=df['revol_util'].apply(revol_util_clean).astype(float)
```

In [ ]:

```
1
```

In [18]:

```
1 df.describe()
```

Out[18]:

| | id | member_id | loan_amnt | funded_amnt | funded_amnt_inv | term |
|---|---|---|---|---|---|---|
| **count** | 4.235000e+04 | 4.235000e+04 | 42350.000000 | 42350.000000 | 42350.000000 | 42350.000000 |
| **mean** | 6.653378e+05 | 8.266770e+05 | 11109.211924 | 10840.211334 | 10162.262400 | 42.220165 |
| **std** | 2.188062e+05 | 2.788685e+05 | 7409.408311 | 7145.056871 | 7130.080610 | 10.516469 |
| **min** | 5.473400e+04 | 7.069400e+04 | 500.000000 | 500.000000 | 0.000000 | 36.000000 |
| **25%** | 4.989700e+05 | 6.395432e+05 | 5200.000000 | 5100.000000 | 4964.045719 | 36.000000 |
| **50%** | 6.450115e+05 | 8.249430e+05 | 9800.000000 | 9600.000000 | 8500.000000 | 36.000000 |
| **75%** | 8.263058e+05 | 1.034498e+06 | 15000.000000 | 15000.000000 | 14000.000000 | 60.000000 |
| **max** | 1.077501e+06 | 1.314167e+06 | 35000.000000 | 35000.000000 | 35000.000000 | 60.000000 |

8 rows × 32 columns

In [19]:

```
1 for i in df.columns:
2     if df[i].dtype=='object':
3         print(i)
```

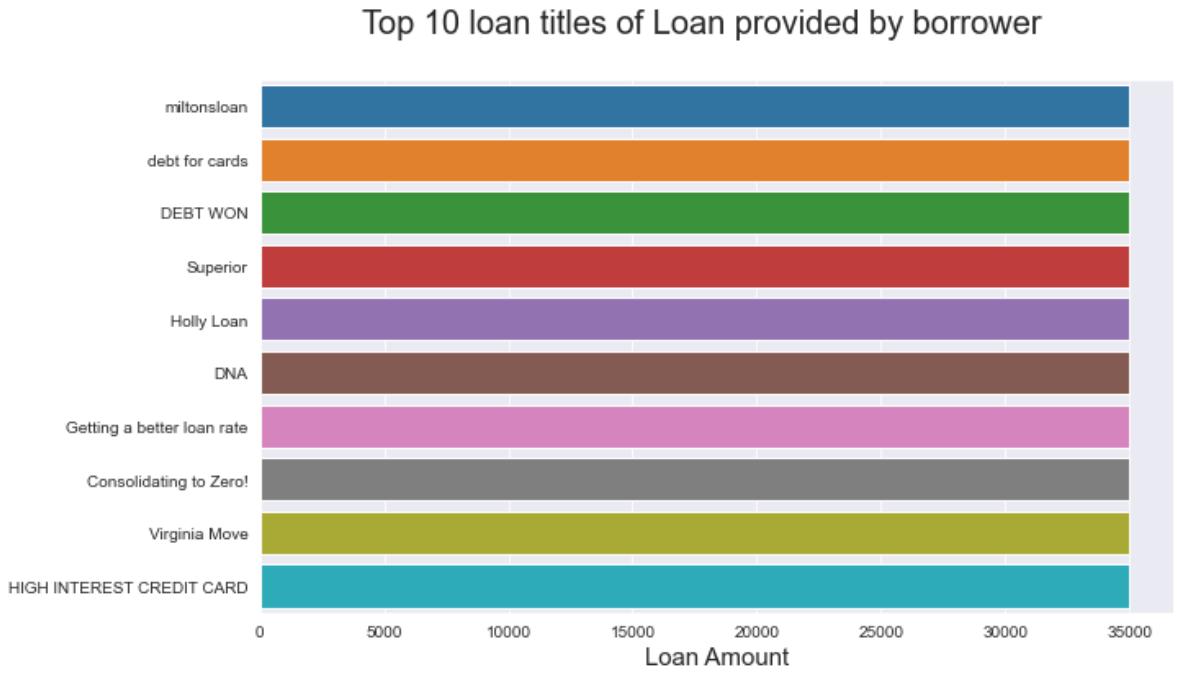
```
grade
sub_grade
emp_title
emp_length
home_ownership
verification_status
issue_d
loan_status
purpose
title
zip_code
addr_state
earliest_cr_line
last_pymnt_d
last_credit_pull_d
```

In [ ]:

```
1
```

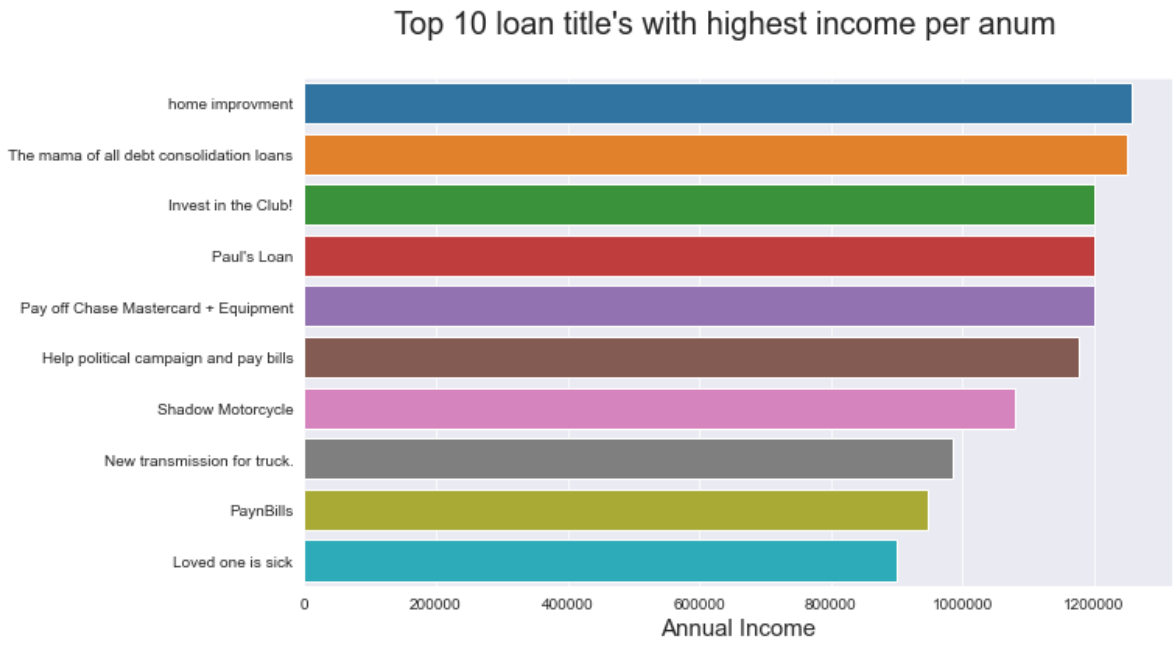
In [20]:

```
plt.figure(figsize=(10,6))
sns.set_style('darkgrid')
sns.barplot(y='title',x='loan_amnt',data=df.groupby('title')['loan_amnt'].mean().sort_v
plt.suptitle("Top 10 loan titles of Loan provided by borrower",fontsize=20)
plt.xlabel('Loan Amount',fontsize=15)
plt.ylabel('')
plt.show()
```



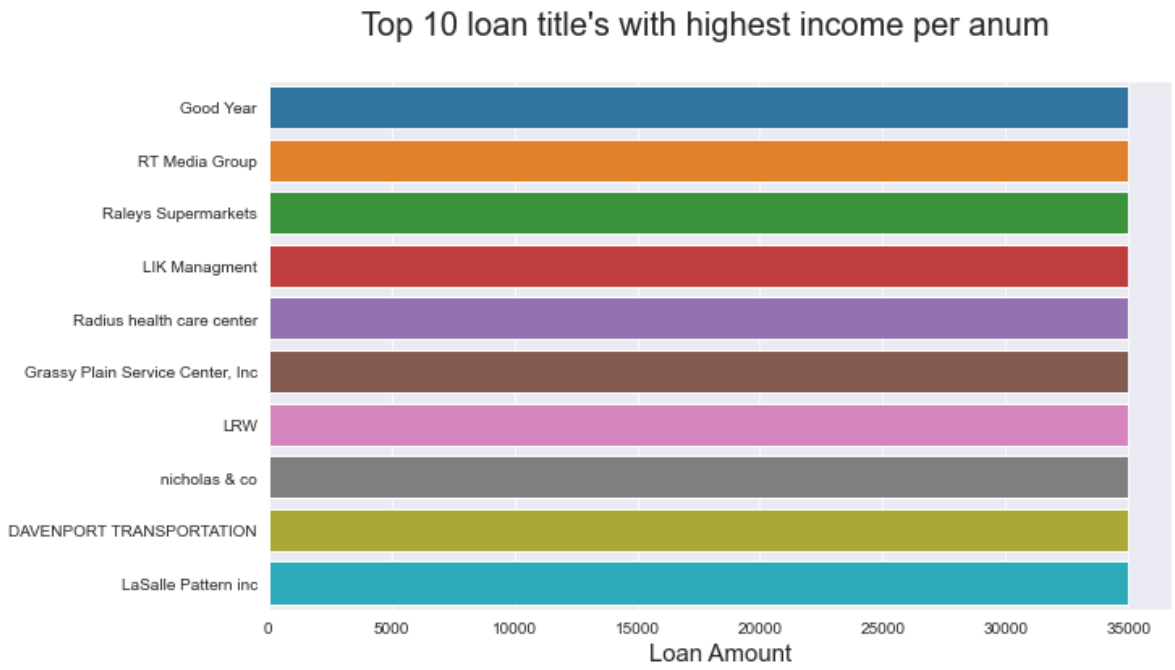
In [21]:

```
plt.figure(figsize=(10,6))
sns.barplot(y='title',x='annual_inc',data=df.groupby('title')['annual_inc'].mean().sort
plt.suptitle("Top 10 loan title's with highest income per anum",fontsize=20)
plt.xlabel('Annual Income',fontsize=15)
plt.ticklabel_format(style='plain',axis='x')
plt.ylabel('')
plt.show()
```



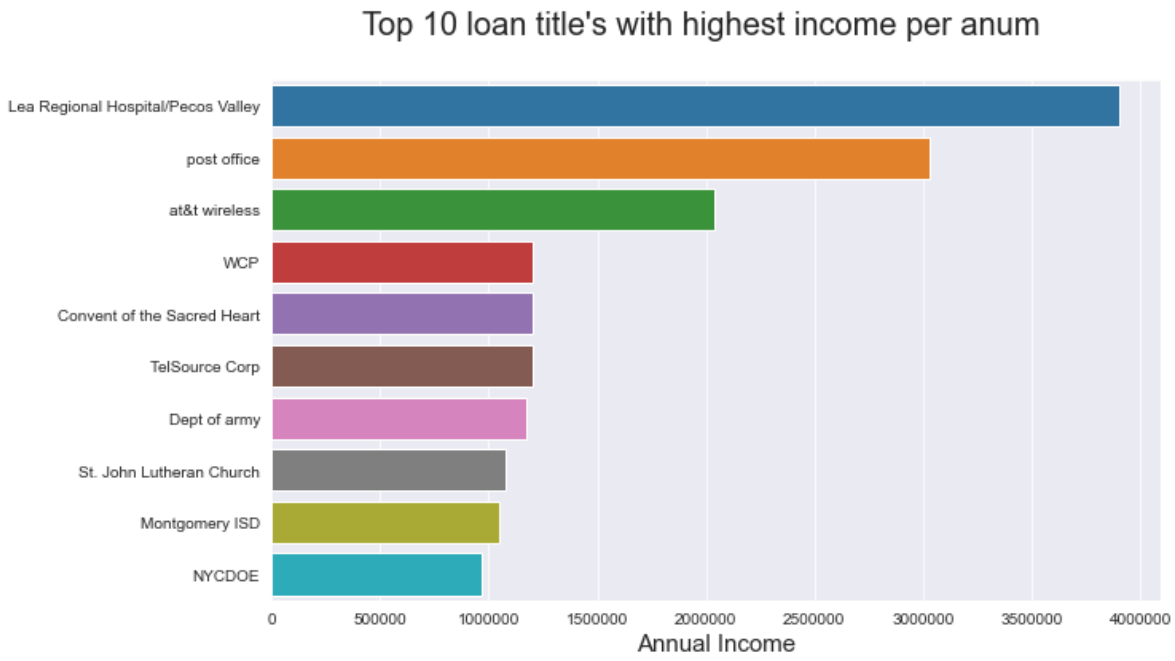
In [22]:

```
plt.figure(figsize=(10,6))
sns.barplot(y='emp_title',x='loan_amnt',data=df.groupby('emp_title')['loan_amnt'].mean(
plt.suptitle("Top 10 loan title's with highest income per anum",fontsize=20)
plt.xlabel('Loan Amount',fontsize=15)
plt.ticklabel_format(style='plain',axis='x')
plt.ylabel('')
plt.show()
```



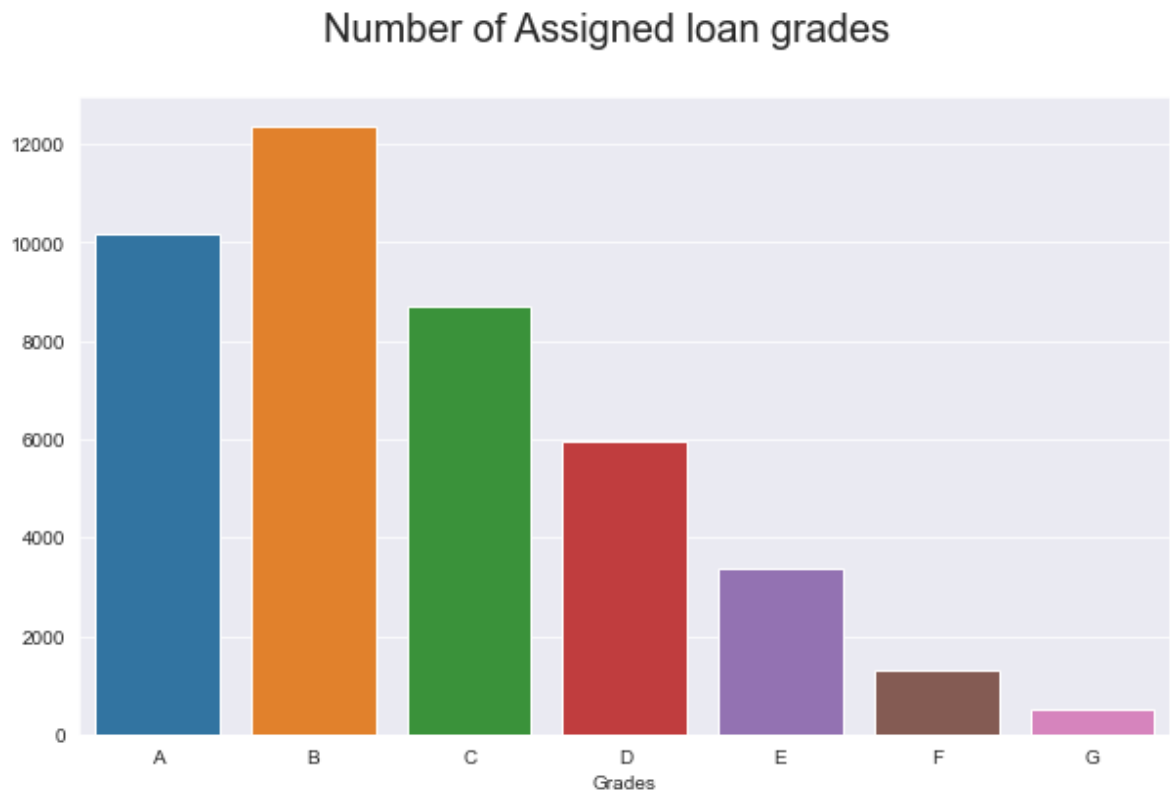
In [23]:

```
plt.figure(figsize=(10,6))
sns.barplot(y='emp_title',x='annual_inc',data=df.groupby('emp_title')['annual_inc'].mea
plt.suptitle("Top 10 loan title's with highest income per anum",fontsize=20)
plt.xlabel('Annual Income',fontsize=15)
plt.ticklabel_format(style='plain',axis='x')
plt.ylabel('')
plt.show()
```



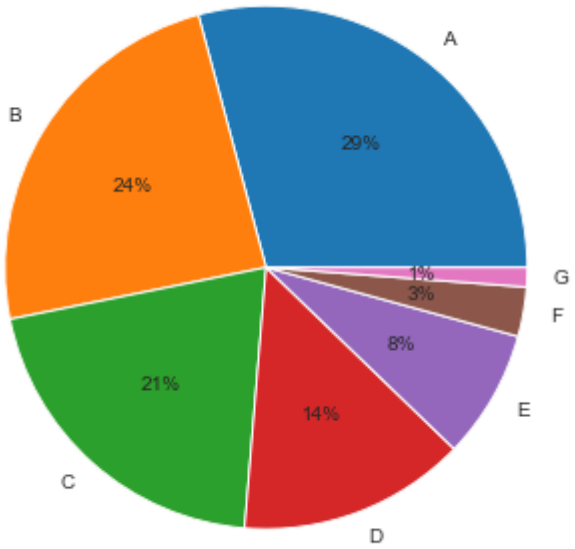
In [24]:

```
plt.figure(figsize=(10,6))
sns.countplot(x='grade', data=df.sort_values('grade'))
plt.suptitle("Number of Assigned loan grades",fontsize=20)
plt.xlabel('Grades')
plt.ylabel('')
plt.show()

```



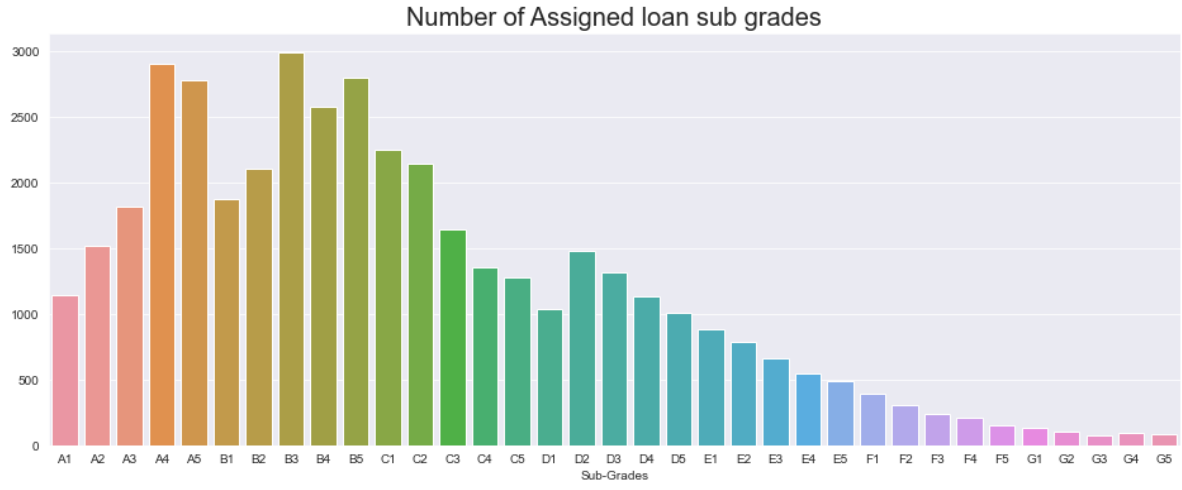
In [25]:

```
plt.figure(figsize=(12,6))
plt.pie(df['grade'].value_counts(),
        autopct='%.0f%%',
        labels=['A','B','C','D','E','F','G'])
plt.show()

```




In [26]:

```
plt.figure(figsize=(16,6))
sns.countplot(x='sub_grade', data=df.sort_values('sub_grade'))
plt.title("Number of Assigned loan sub grades",size=20)
plt.xlabel('Sub-Grades')
plt.ylabel('')
plt.show()
```

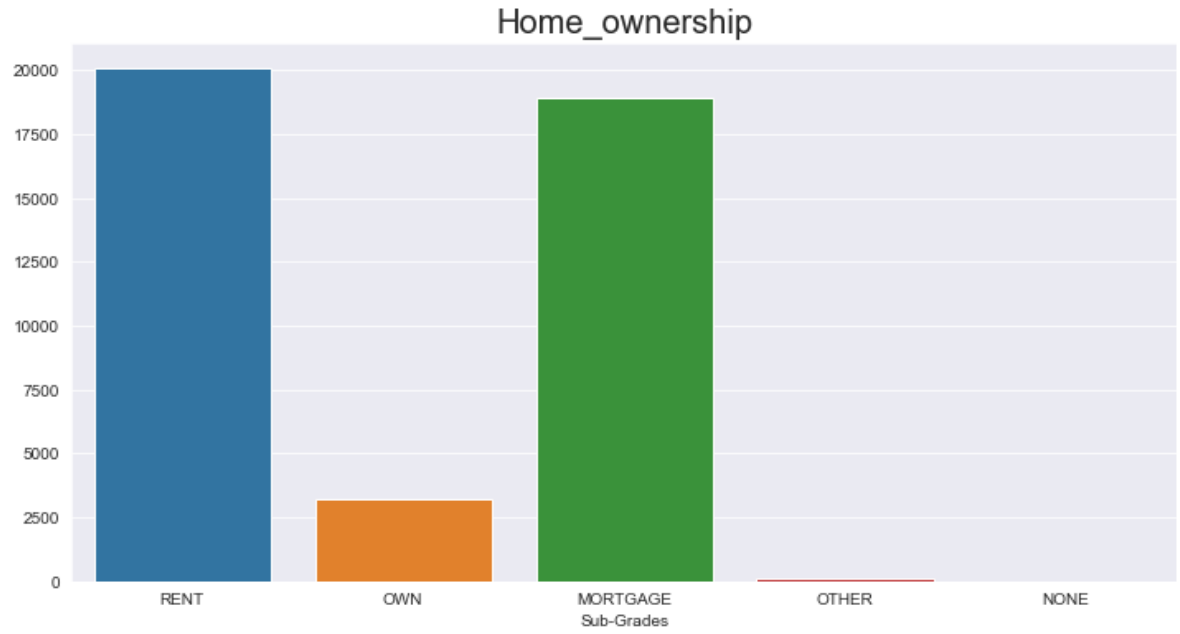


In [27]:

```
plt.figure(figsize=(16,6))
sns.countplot(x='emp_length', data=df.sort_values('emp_length'))
plt.title("EMPLOYEE LENGH",size=20)
plt.xlabel('No. of Years')
plt.ylabel('')
plt.show()
#HOW MUCH TIME
```

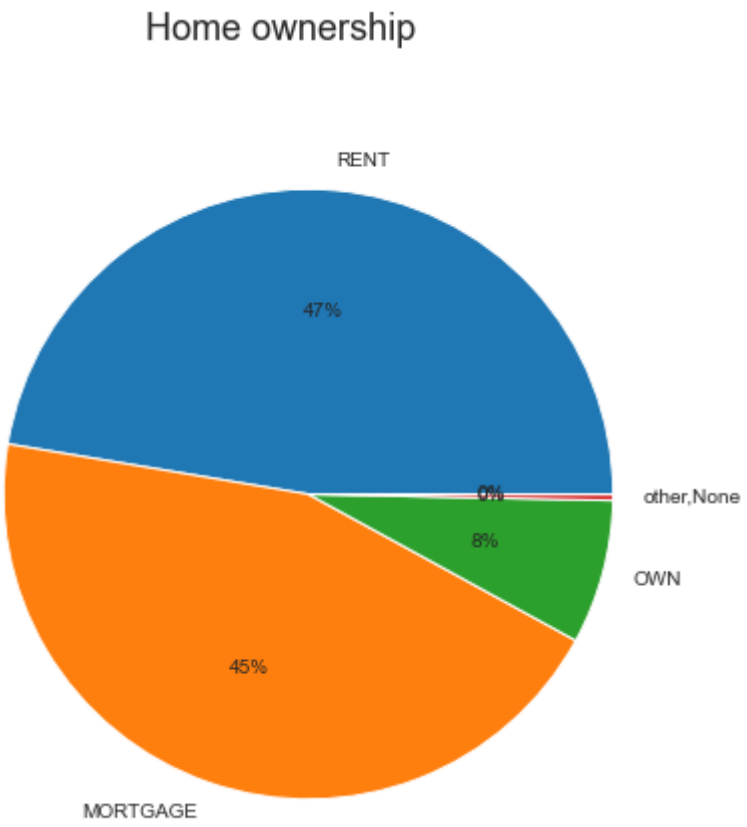
In [28]:

```
plt.figure(figsize=(12,6))
sns.countplot(x='home_ownership', data=df)
plt.title("Home_ownership",size=20)
plt.xlabel('Sub-Grades')
plt.ylabel('')
plt.show()
```



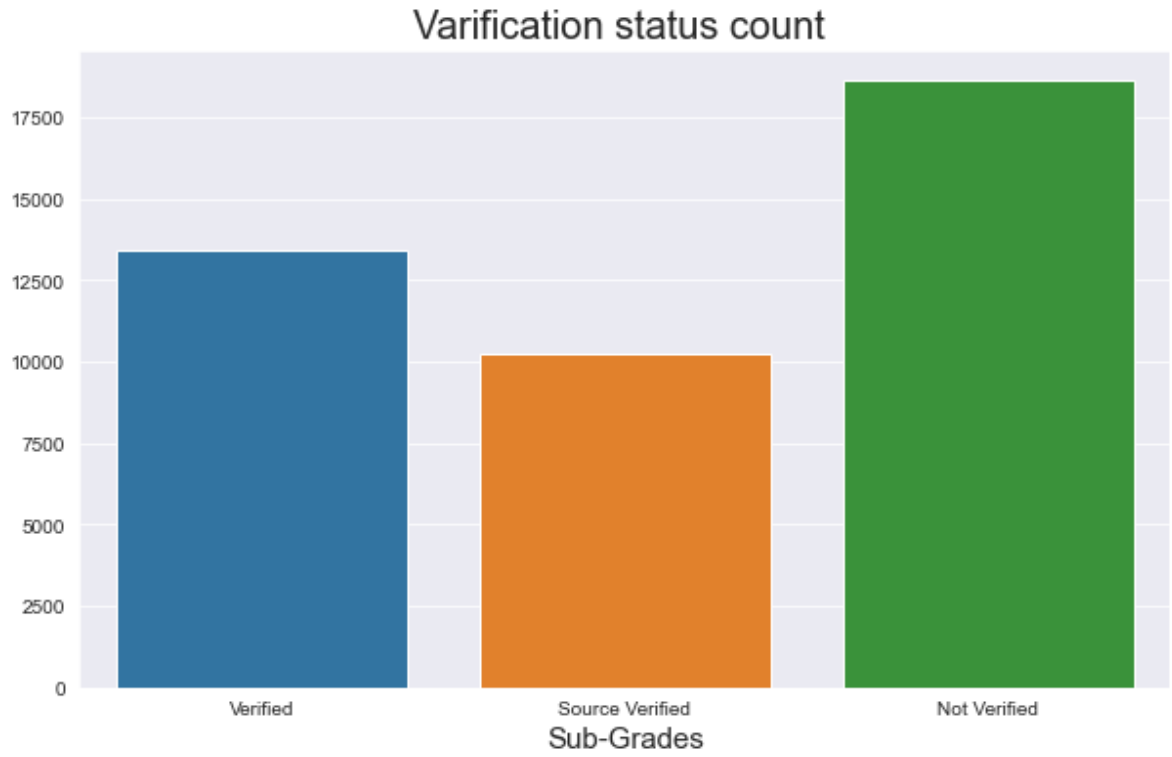
In [29]:

```
plt.figure(figsize=(15,7))
plt.pie(df['home_ownership'].value_counts(),
        autopct='%.0f%%',
        labels=['RENT','MORTGAGE','OWN','other,None',''])
plt.suptitle('Home ownership',fontsize=18)
plt.show()
```



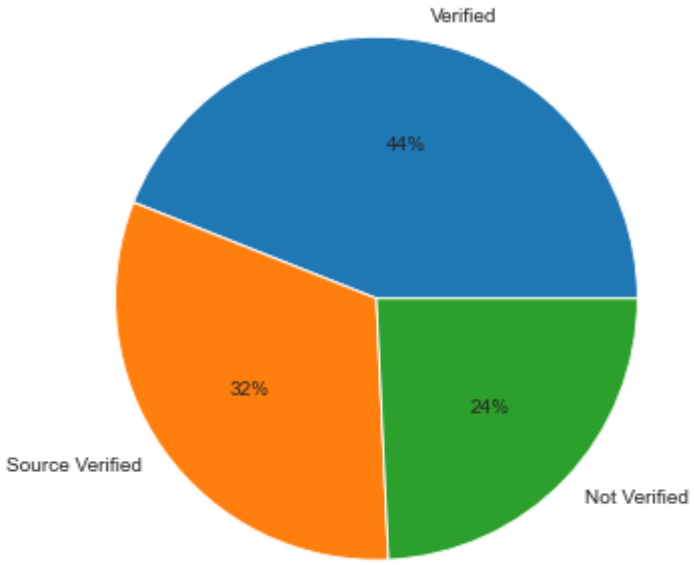
In [30]:

```
plt.figure(figsize=(10,6))
sns.countplot(x='verification_status', data=df)
plt.title("Varification status count ",size=20)
plt.xlabel('Sub-Grades',size=15)
plt.ylabel('')
plt.show()
```



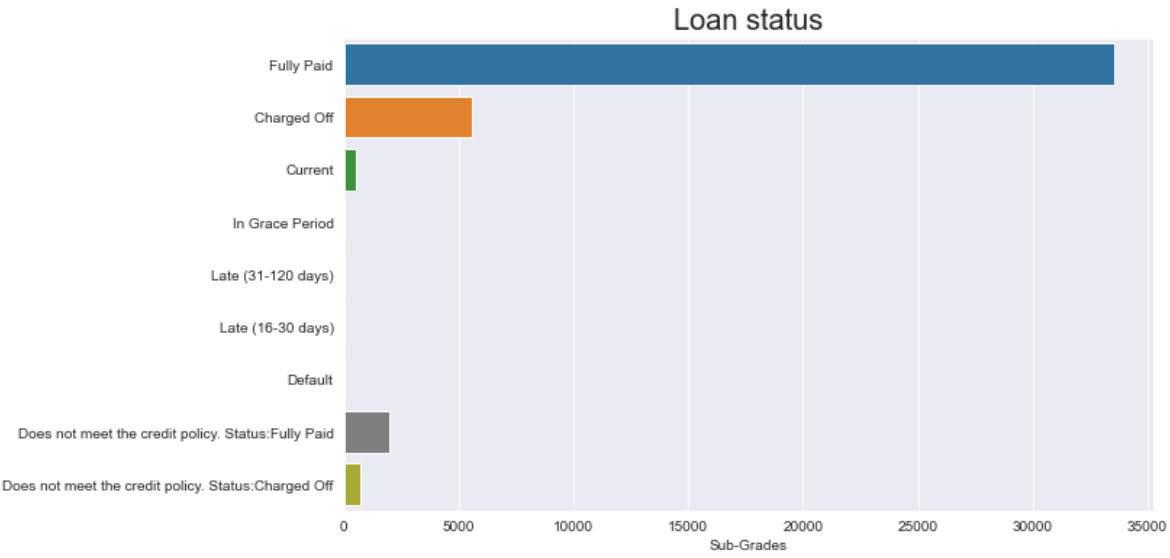
In [31]:

```
plt.figure(figsize=(12,6))
plt.pie(df['verification_status'].value_counts(),
        autopct='%.0f%%',
        labels=['Verified','Source Verified','Not Verified'])
plt.show()
```



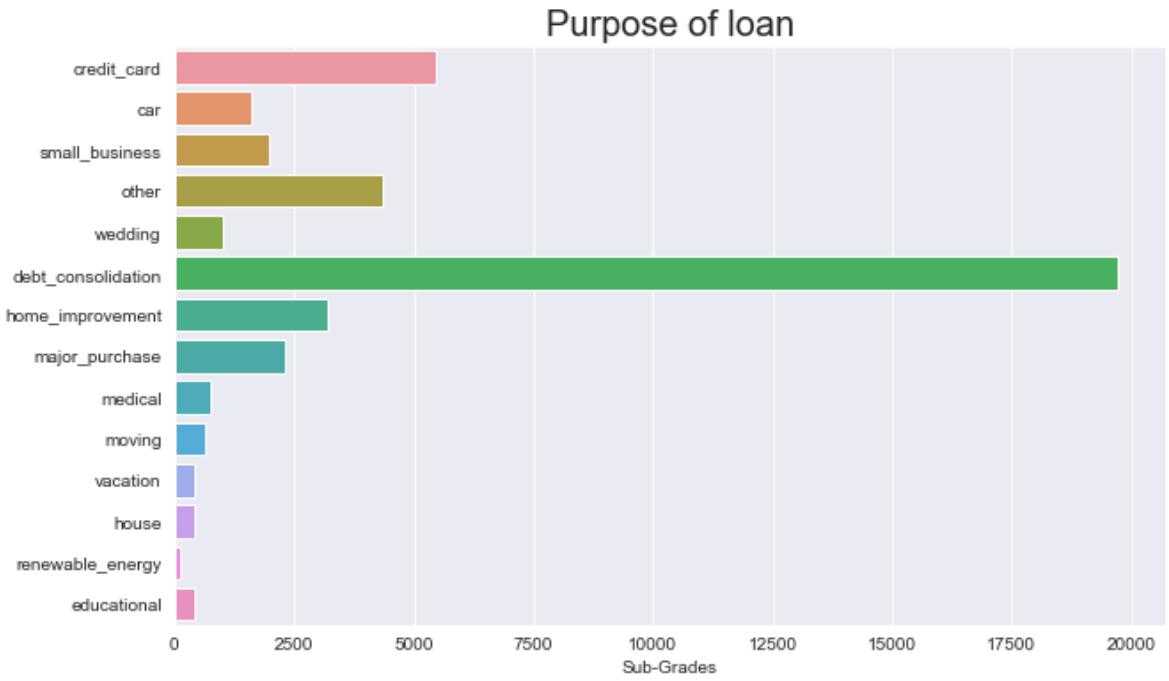

In [32]:

```
plt.figure(figsize=(10,6))
sns.countplot(y='loan_status', data=df)
plt.title("Loan status",size=20)
plt.xlabel('Sub-Grades')
plt.ylabel('')
plt.show()
```



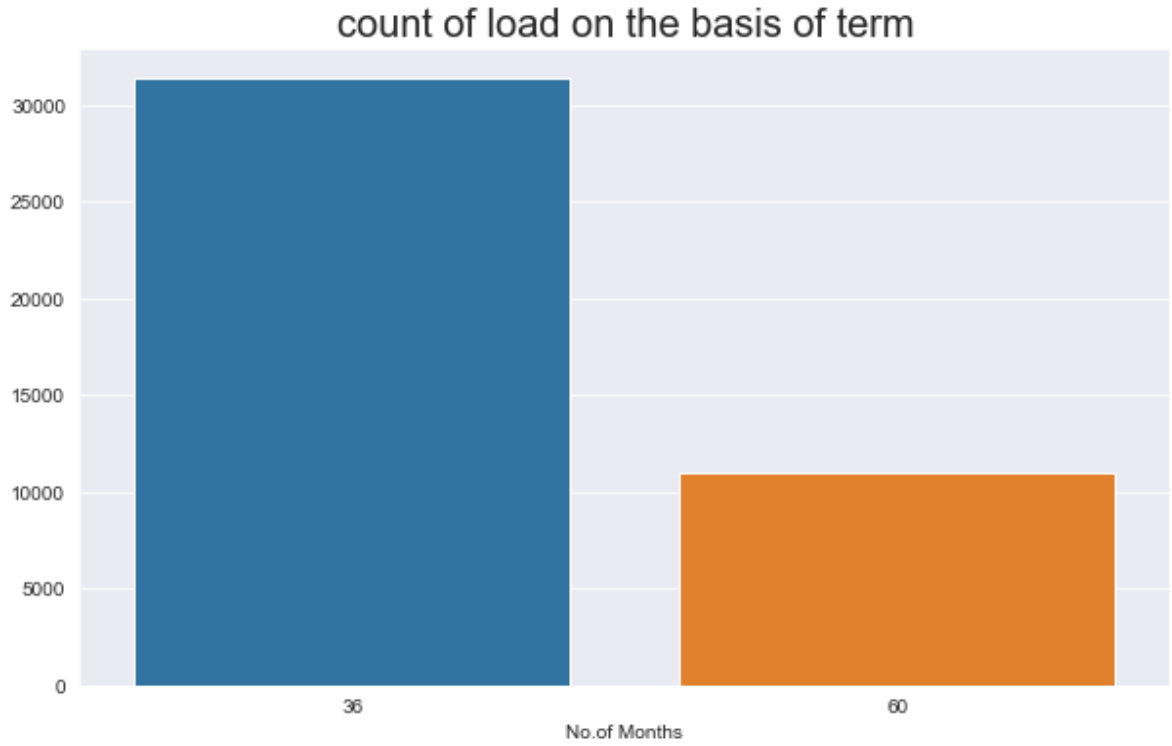
In [33]:

```
plt.figure(figsize=(10,6))
sns.countplot(y='purpose', data=df)
plt.title("Purpose of loan",size=20)
plt.xlabel('Sub-Grades')
plt.ylabel('')
plt.show()
```



In [34]:

```
plt.figure(figsize=(10,6))
sns.countplot(x='term', data=df)
plt.title("count of load on the basis of term",size=20)
plt.xlabel('No.of Months')
plt.ylabel('')
plt.show()
```




In [ ]:

```
1
```

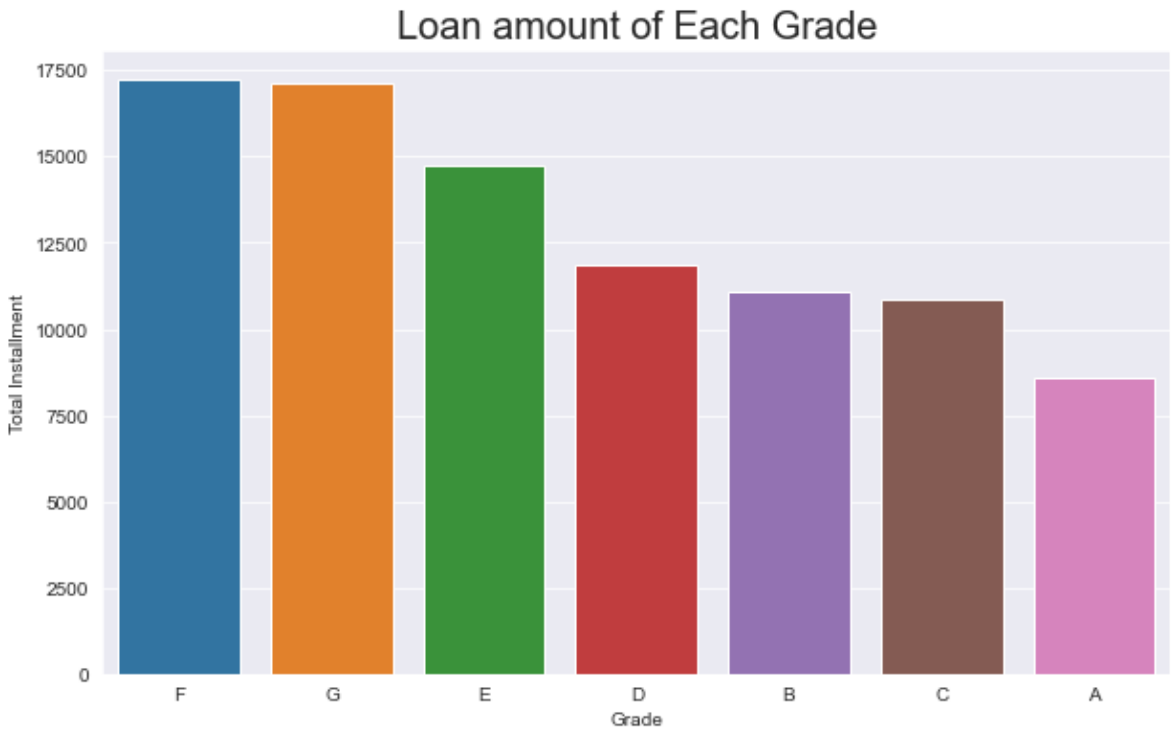
In [35]:

```
for i in df.columns:
    if df[i].dtype!='object':
        print(i)
```

```
id
member_id
loan_amnt
funded_amnt
funded_amnt_inv
term
int_rate
installment
annual_inc
dti
delinq_2yrs
fico_range_low
fico_range_high
inq_last_6mths
open_acc
pub_rec
revol_bal
revol_util
total_acc
out_prncp
out_prncp_inv
total_pymnt
total_pymnt_inv
total_rec_prncp
total_rec_int
total_rec_late_fee
recoveries
collection_recovery_fee
last_pymnt_amnt
last_fico_range_high
last_fico_range_low
pub_rec_bankruptcies
```

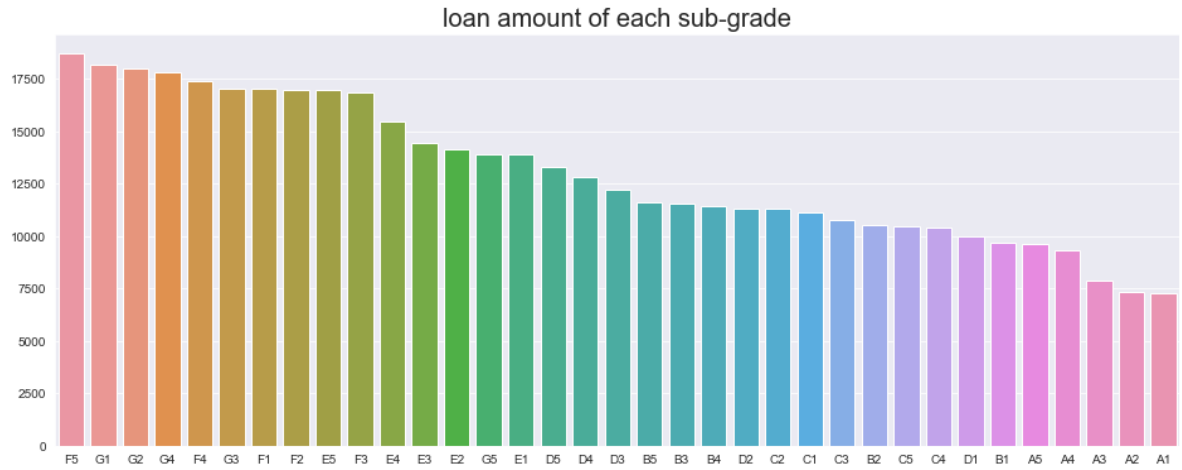
In [36]:

```
plt.figure(figsize=(10,6))
sns.barplot(x='grade', y='loan_amnt', data=df.groupby('grade')['loan_amnt'].mean().sort
plt.title('Loan amount of Each Grade',size=20)
plt.xlabel('Grade')
plt.ylabel('Total Installment')
plt.show()
```



In [37]:

```
plt.figure(figsize=(16,6))
sns.barplot(x='sub_grade', y='loan_amnt', data=df.groupby('sub_grade')['loan_amnt'].mea
plt.title('loan amount of each sub-grade',size=20)
plt.xlabel('')
plt.ylabel('')
plt.show()
```

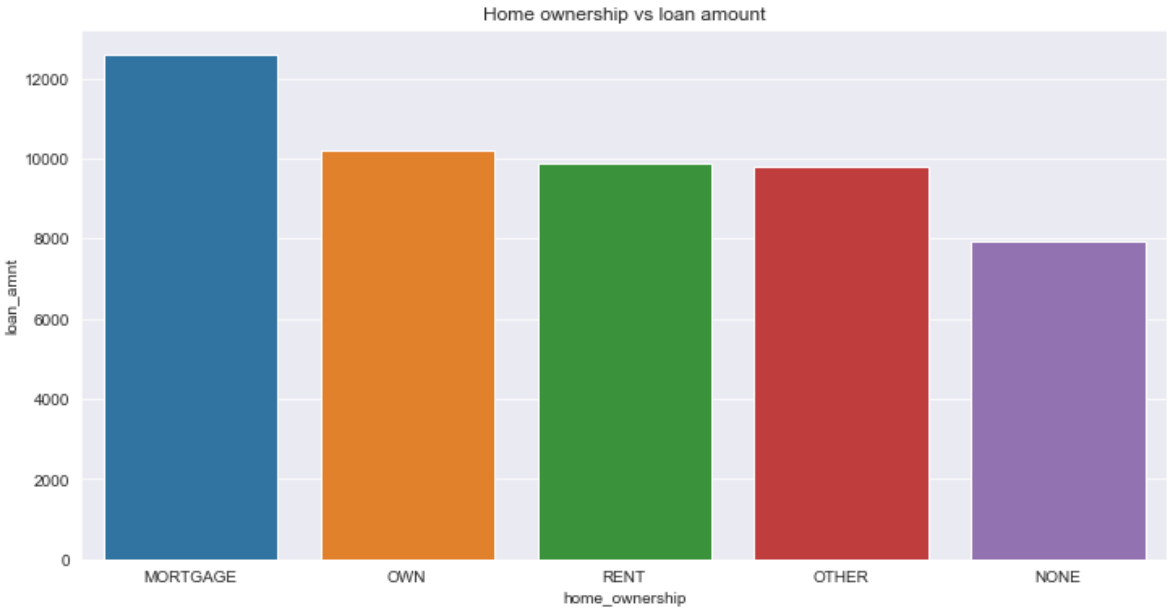



In [ ]:

```
1
```

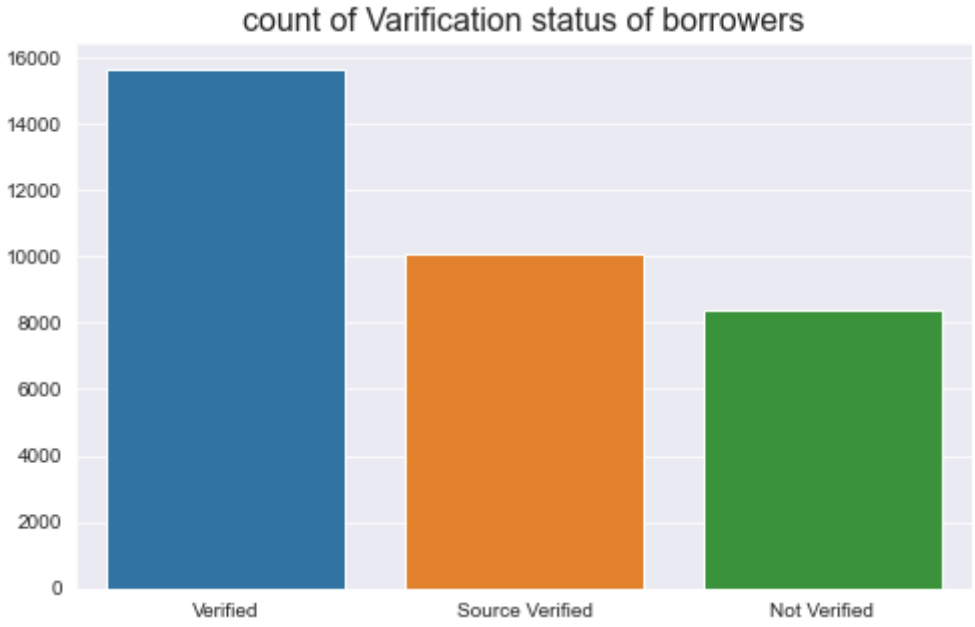
In [38]:

```
plt.figure(figsize=(12,6))
sns.barplot(x='home_ownership', y='loan_amnt', data=df.groupby('home_ownership')['loan_
plt.title('Home ownership vs loan amount')
plt.show()
```



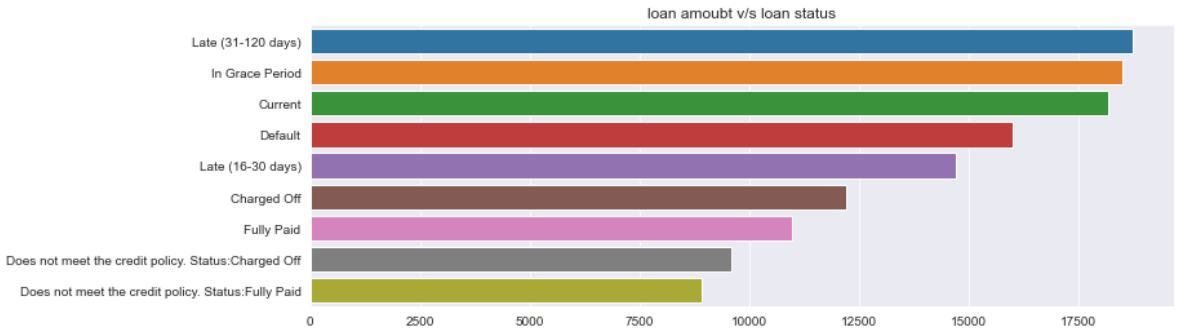

In [39]:

```
plt.figure(figsize=(8,5))
sns.barplot(x='verification_status', y='loan_amnt', data=df.groupby('verification_statu
plt.title('count of Varification status of borrowers',size=16)
plt.xlabel('')
plt.ylabel('')
plt.show()
```



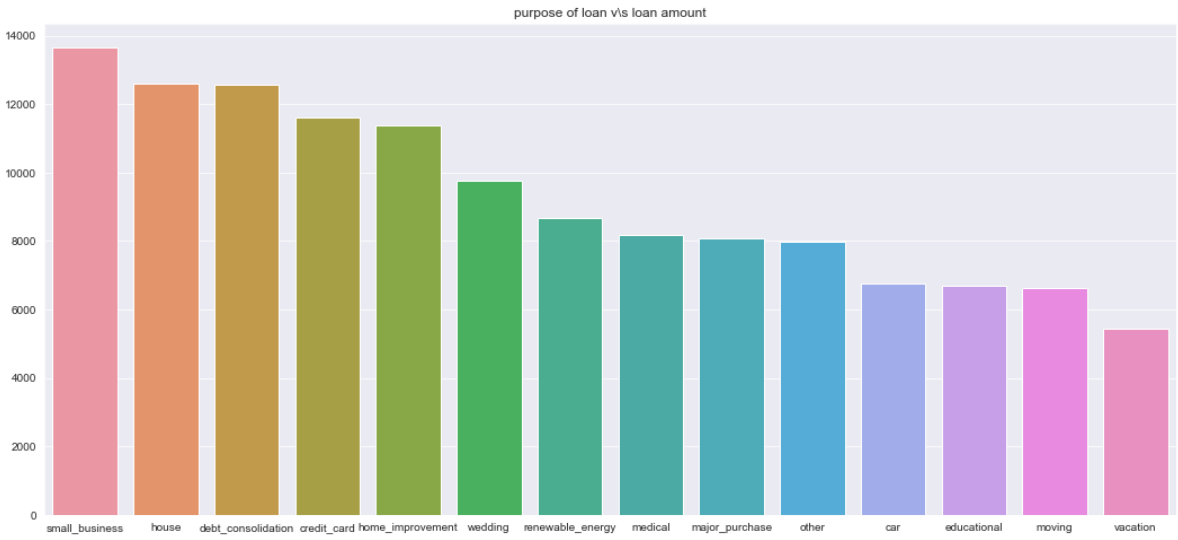
In [40]:

```
plt.figure(figsize=(12,4))
sns.barplot(y='loan_status', x='loan_amnt', data=df.groupby('loan_status')['loan_amnt']
plt.title('loan amoubt v/s loan status')
plt.xlabel('')
plt.ylabel('')
plt.show()
```




In [41]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='purpose', y='loan_amnt', data=df.groupby('purpose')['loan_amnt'].mean().
plt.title('purpose of loan v\s loan amount')
plt.xlabel('')
plt.ylabel('')
plt.show()
```




In [ ]:

```

```

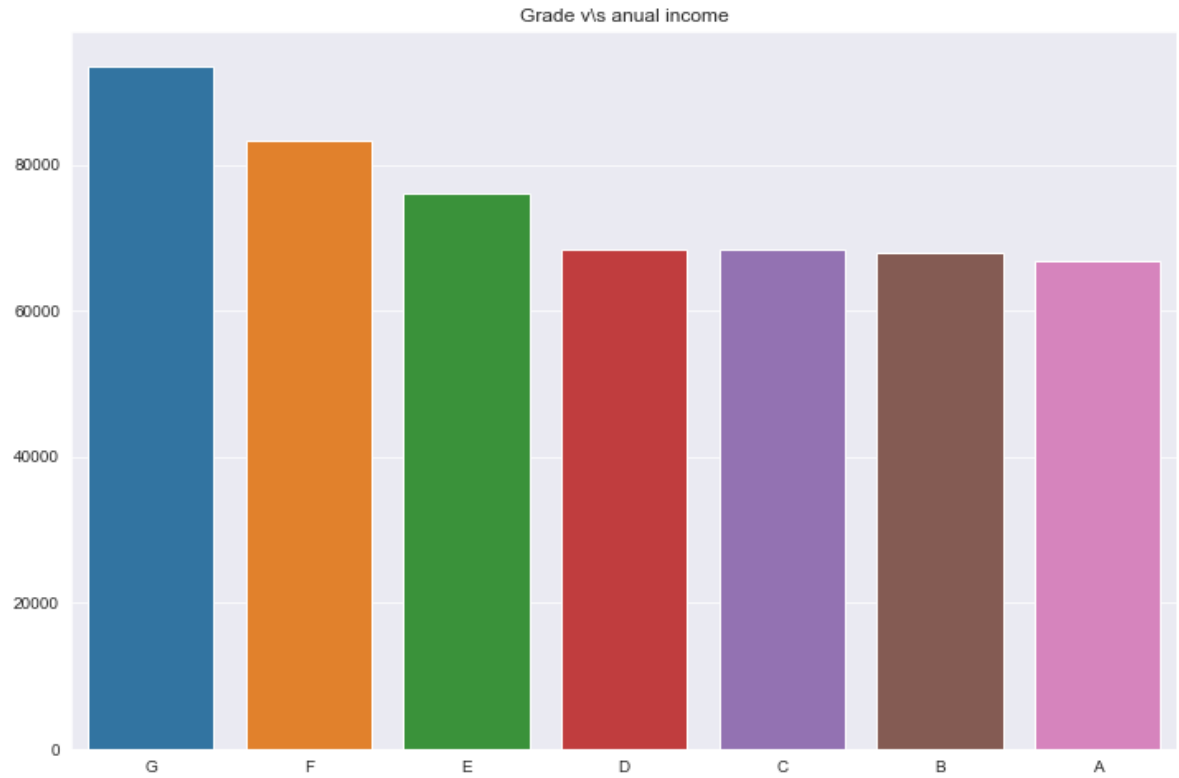
In [42]:

```
for i in df.columns:
    if df[i].dtype=='object':
        print(i)
```

```
grade
sub_grade
emp_title
emp_length
home_ownership
verification_status
issue_d
loan_status
purpose
title
zip_code
addr_state
earliest_cr_line
last_pymnt_d
last_credit_pull_d
```

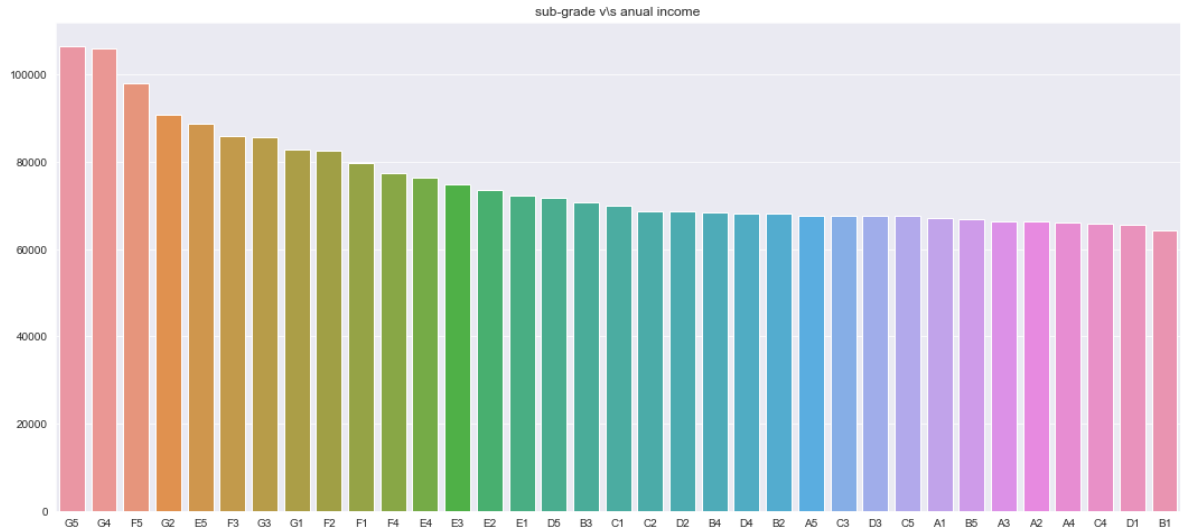
In [43]:

```
plt.figure(figsize=(12,8))
sns.barplot(x='grade', y='annual_inc', data=df.groupby('grade')['annual_inc'].mean().sc
plt.title('Grade v\s anual income')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



In [44]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='sub_grade', y='annual_inc', data=df.groupby('sub_grade')['annual_inc'].m
plt.title('sub-grade v\s anual income')
plt.xlabel('')
plt.ylabel('')
plt.show()
```

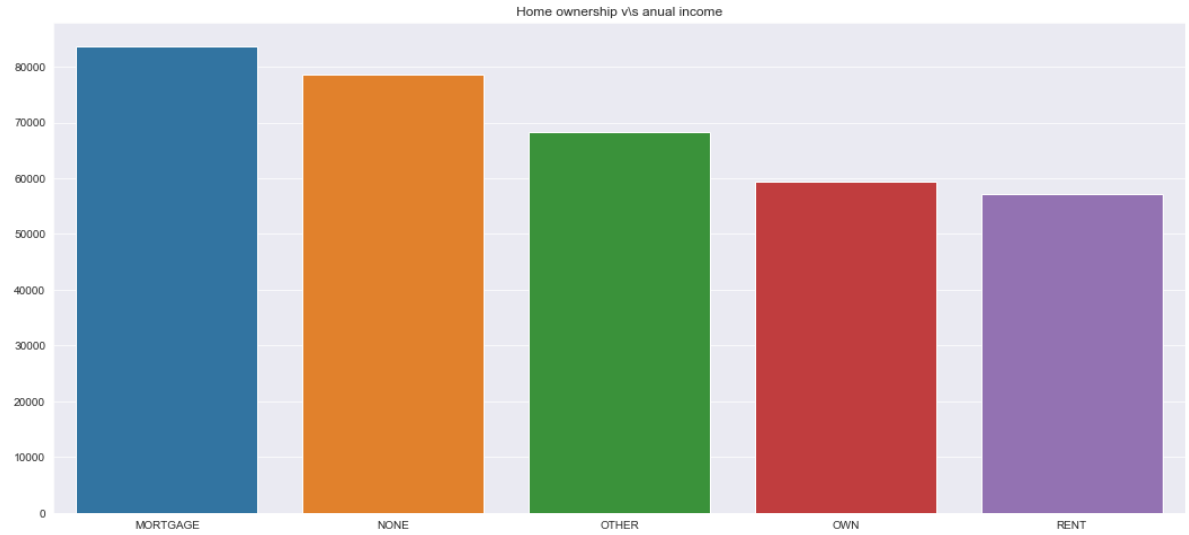



In [ ]:

```
1
```

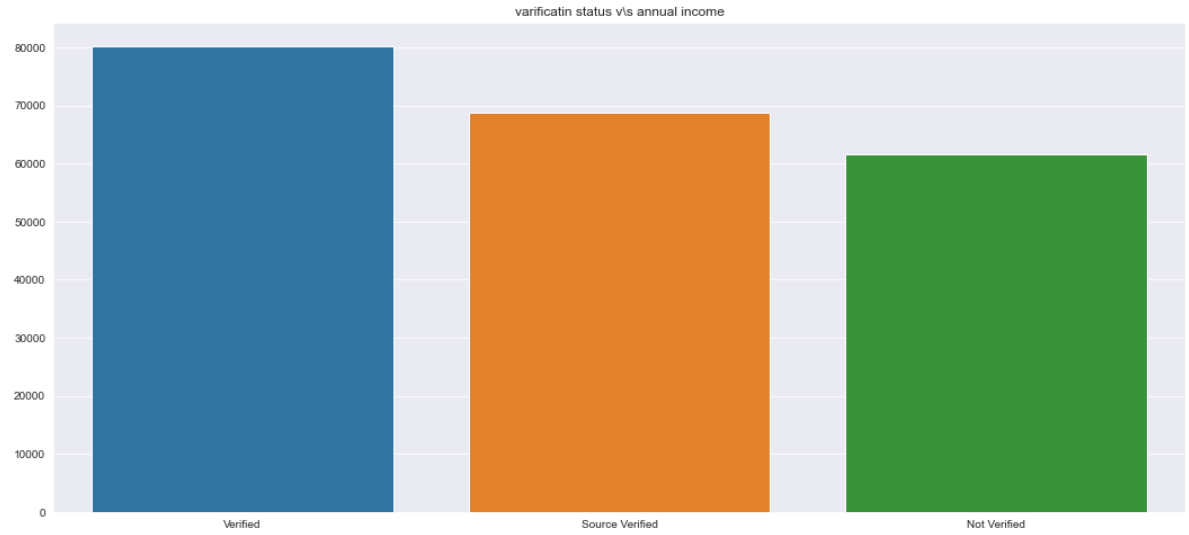
In [45]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='home_ownership', y='annual_inc', data=df.groupby('home_ownership')['annu
plt.title('Home ownership v\s anual income')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



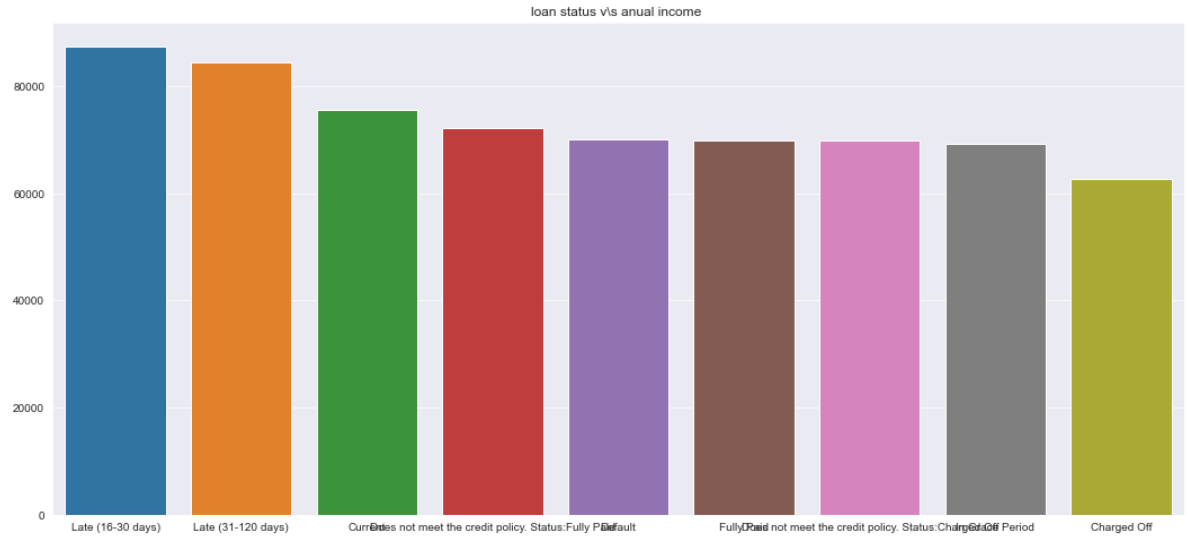

In [46]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='verification_status', y='annual_inc', data=df.groupby('verification_stat
plt.title('varificatin status v\s annual income')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



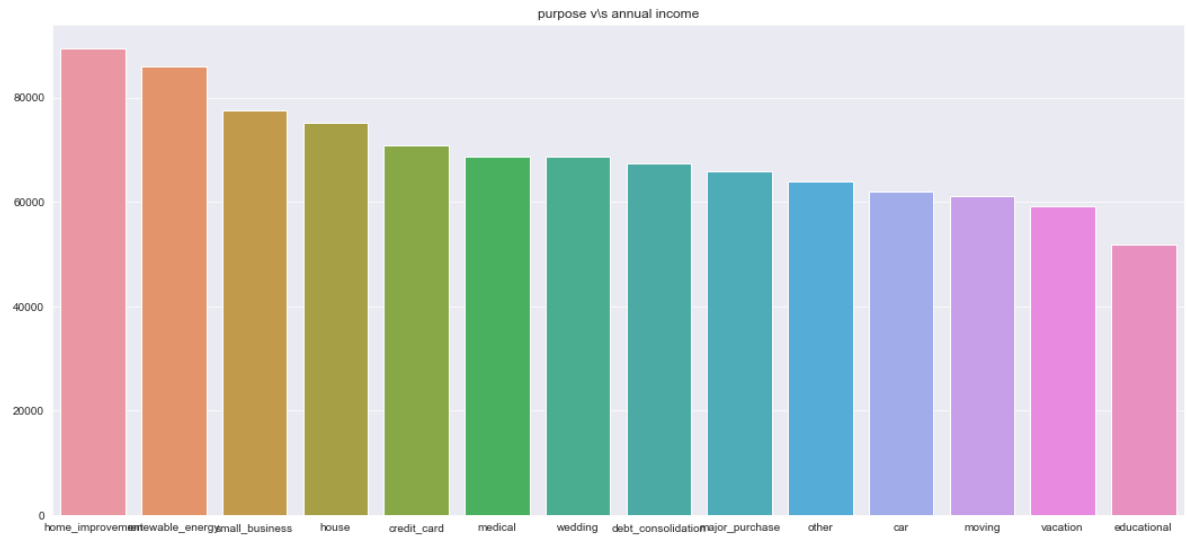
In [47]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='loan_status', y='annual_inc', data=df.groupby('loan_status')['annual_inc
plt.title('loan status v\s anual income ')
plt.xlabel('')
plt.ylabel('')
plt.show()
```




In [48]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='purpose', y='annual_inc', data=df.groupby('purpose')['annual_inc'].mean(
plt.title('purpose v\s annual income')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



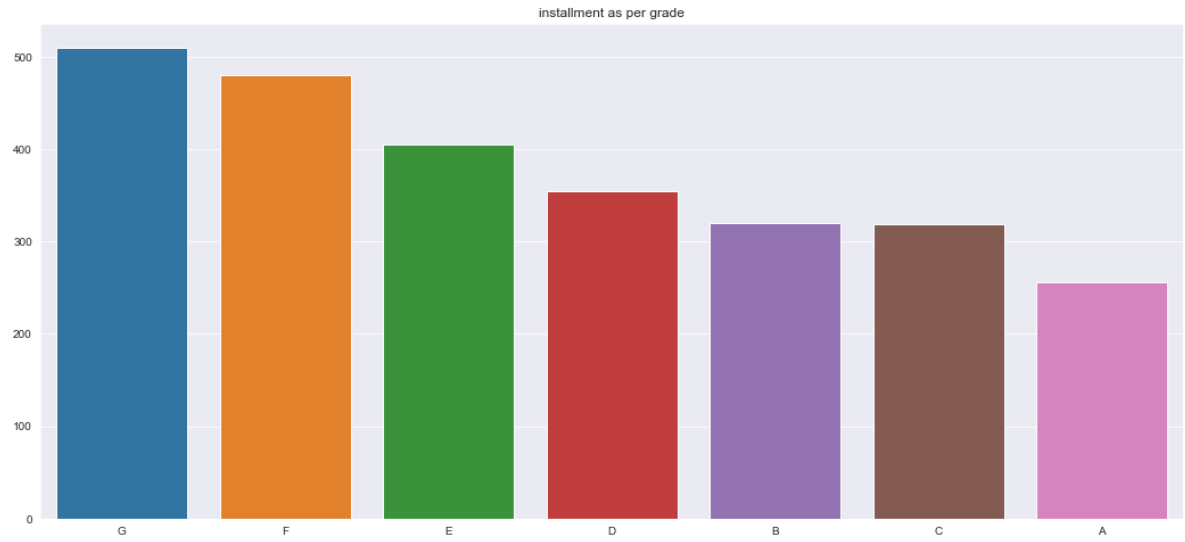
In [49]:

```
for i in df.columns:
    if df[i].dtype=='object':
        print(i)
```

```
grade
sub_grade
emp_title
emp_length
home_ownership
verification_status
issue_d
loan_status
purpose
title
zip_code
addr_state
earliest_cr_line
last_pymnt_d
last_credit_pull_d
```

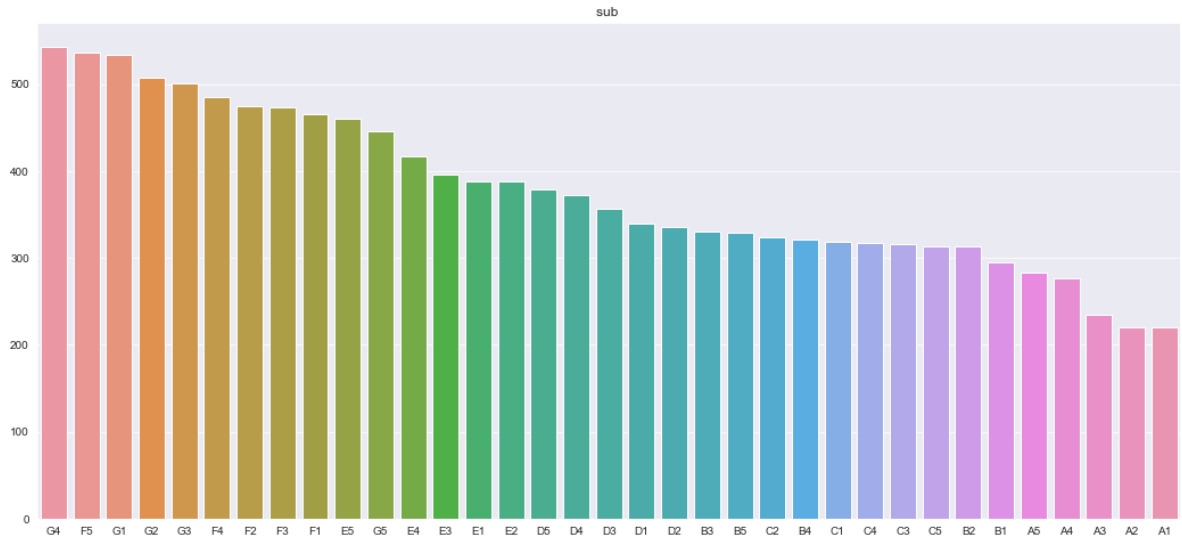
In [50]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='grade', y='installment', data=df.groupby('grade')['installment'].mean().
plt.title('installment as per grade')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



In [51]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='sub_grade', y='installment', data=df.groupby('sub_grade')['installment']
plt.title('sub ')
plt.xlabel('')
plt.ylabel('')
plt.show()
```

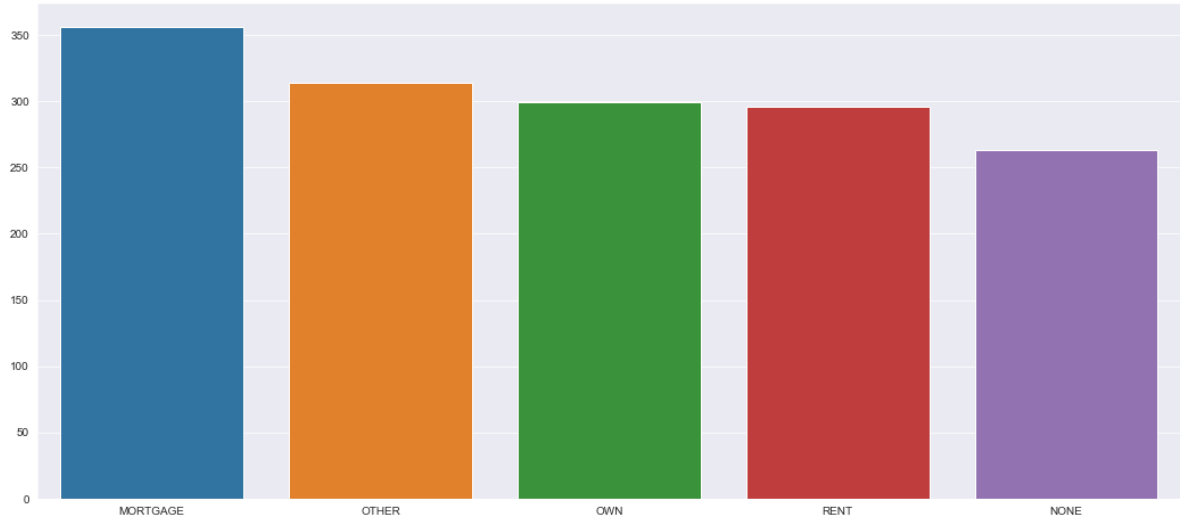



In [ ]:

```
1
```

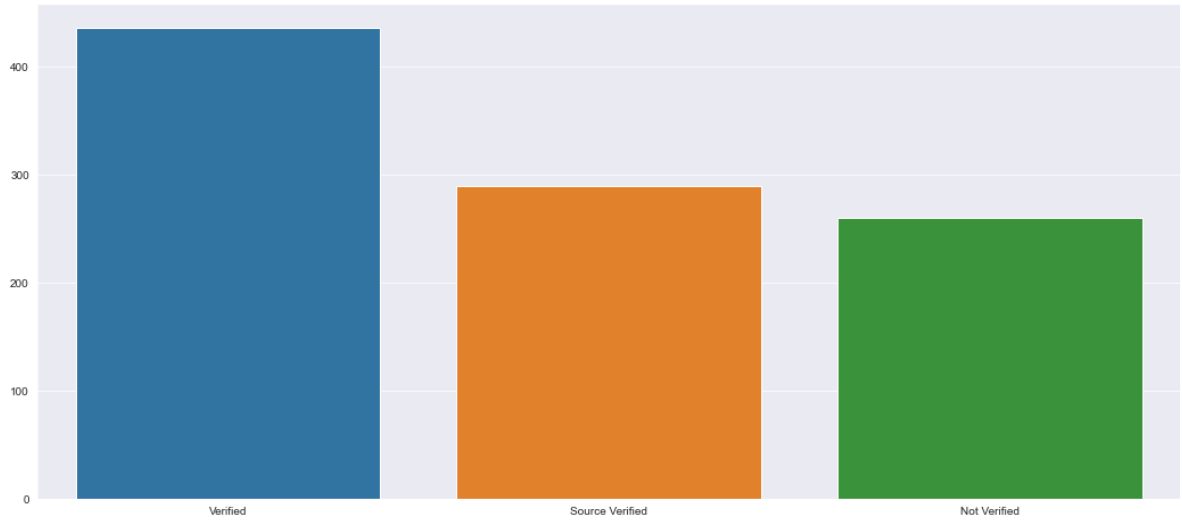
In [52]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='home_ownership', y='installment', data=df.groupby('home_ownership')['ins
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



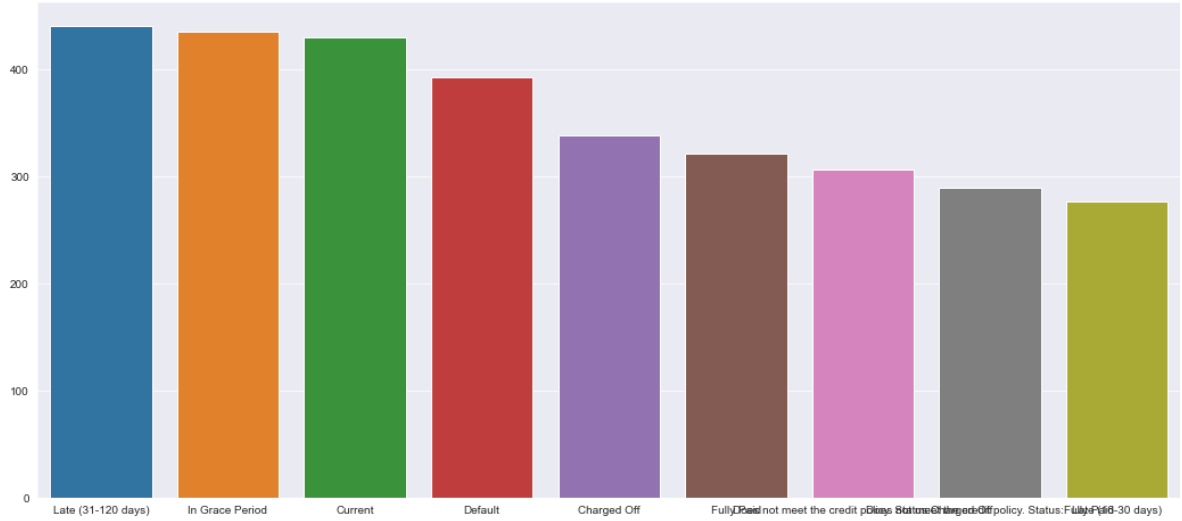

In [53]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='verification_status', y='installment', data=df.groupby('verification_sta
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



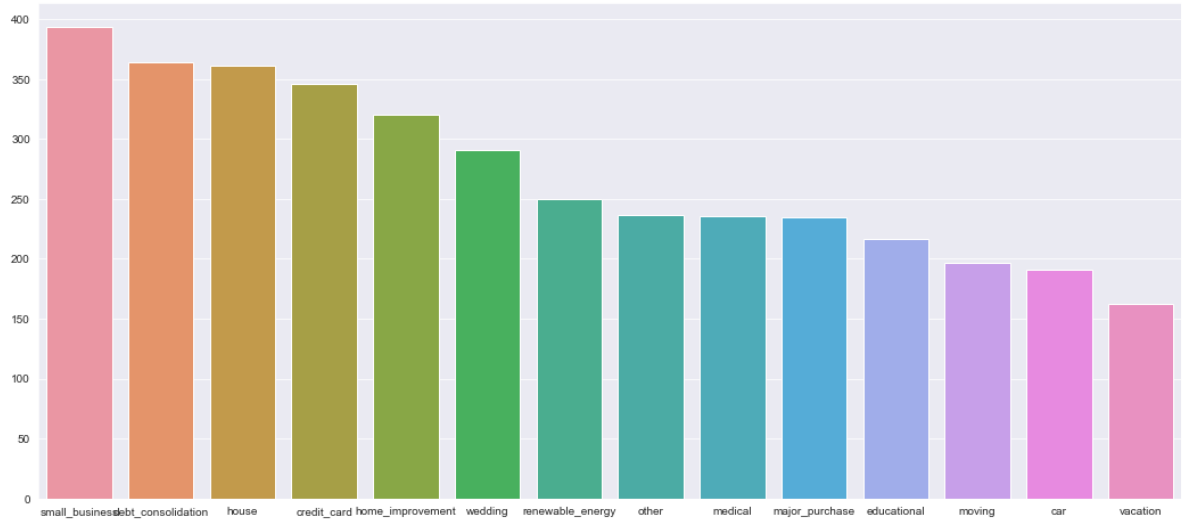
In [54]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='loan_status', y='installment', data=df.groupby('loan_status')['installme
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



In [55]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='purpose', y='installment', data=df.groupby('purpose')['installment'].mea
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



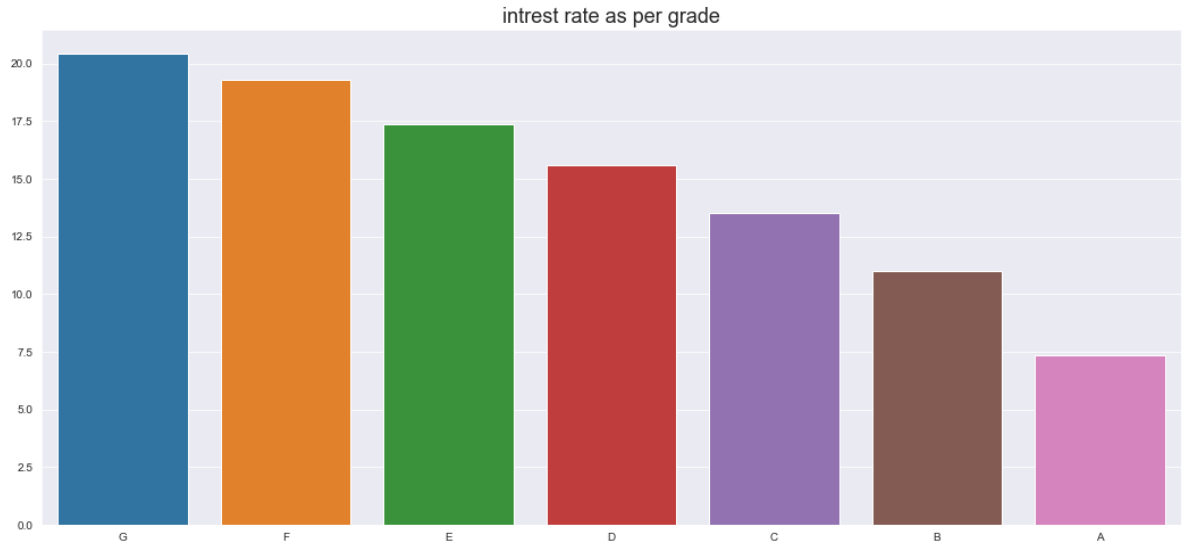

In [56]:

```
for i in df.columns:
    if df[i].dtype=='object':
        print(i)
```

```
grade
sub_grade
emp_title
emp_length
home_ownership
verification_status
issue_d
loan_status
purpose
title
zip_code
addr_state
earliest_cr_line
last_pymnt_d
last_credit_pull_d
```

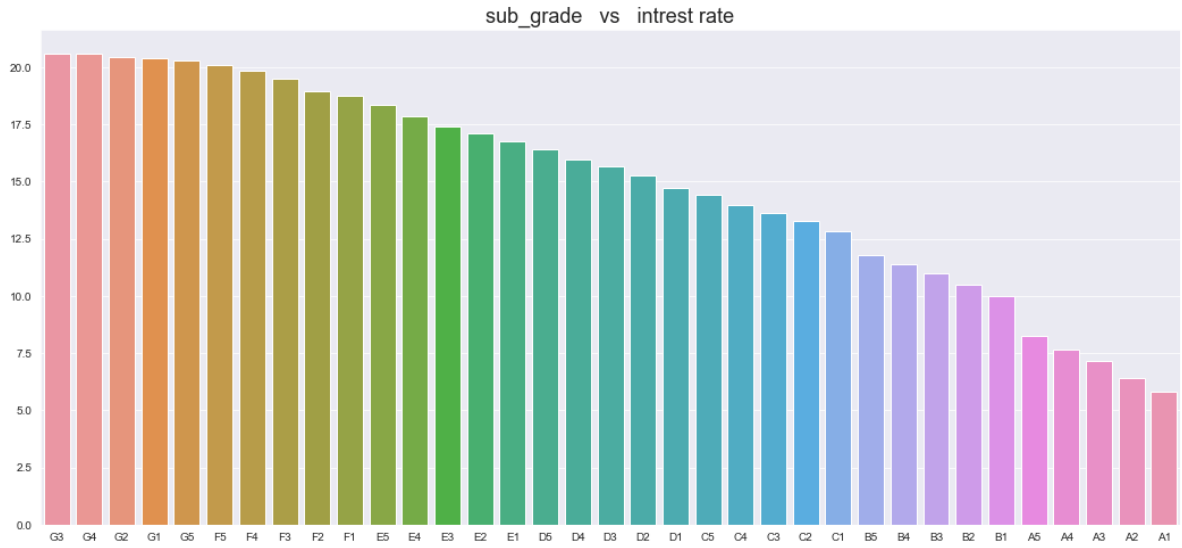
In [99]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='grade', y='int_rate', data=df.groupby('grade')['int_rate'].mean().sort_v
plt.title('intrest rate as per grade',size=18)
plt.xlabel('')
plt.ylabel('')
plt.show()
```



In [98]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='sub_grade', y='int_rate', data=df.groupby('sub_grade')['int_rate'].mean(
plt.title('sub_grade   vs   intrest rate',size=18)
plt.xlabel('')
plt.ylabel('')
plt.show()
```

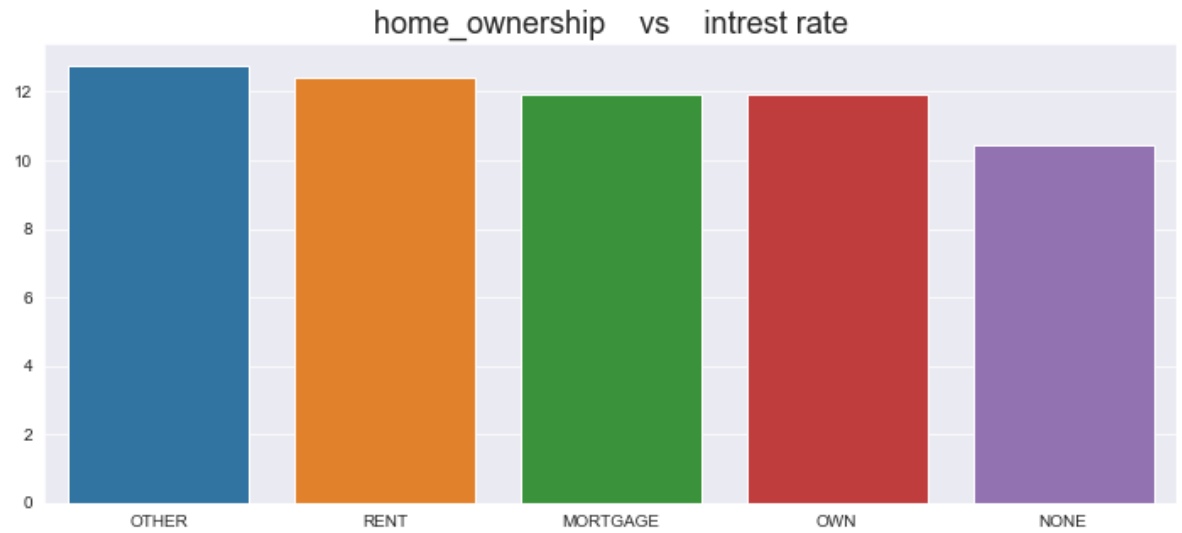



In [ ]:

```
1
```

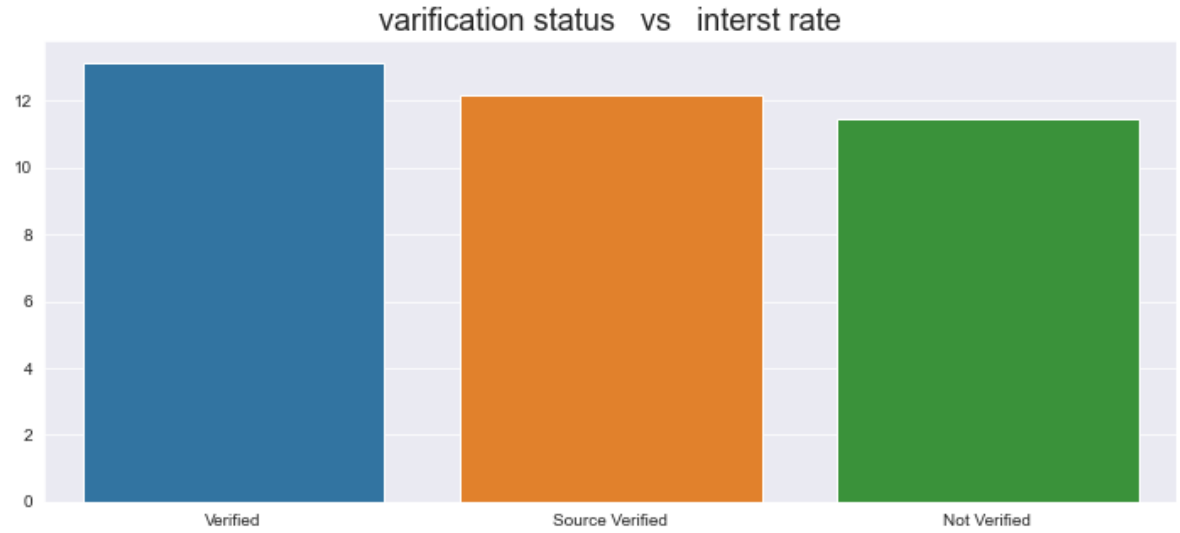
In [97]:

```
plt.figure(figsize=(12,5))
sns.barplot(x='home_ownership', y='int_rate', data=df.groupby('home_ownership')['int_ra
plt.title('home_ownership    vs    intrest rate',size=18)
plt.xlabel('')
plt.ylabel('')
plt.show()
```



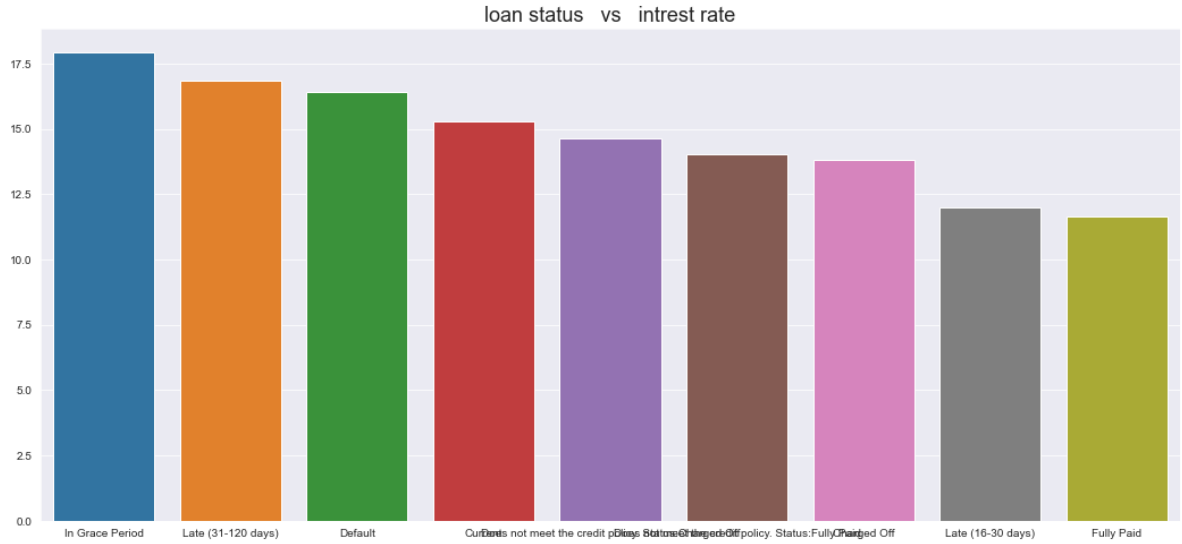
In [96]:

```
plt.figure(figsize=(12,5))
sns.barplot(x='verification_status', y='int_rate', data=df.groupby('verification_status
plt.title('varification status   vs   interst rate',size=18)
plt.xlabel('')
plt.ylabel('')
plt.show()
```



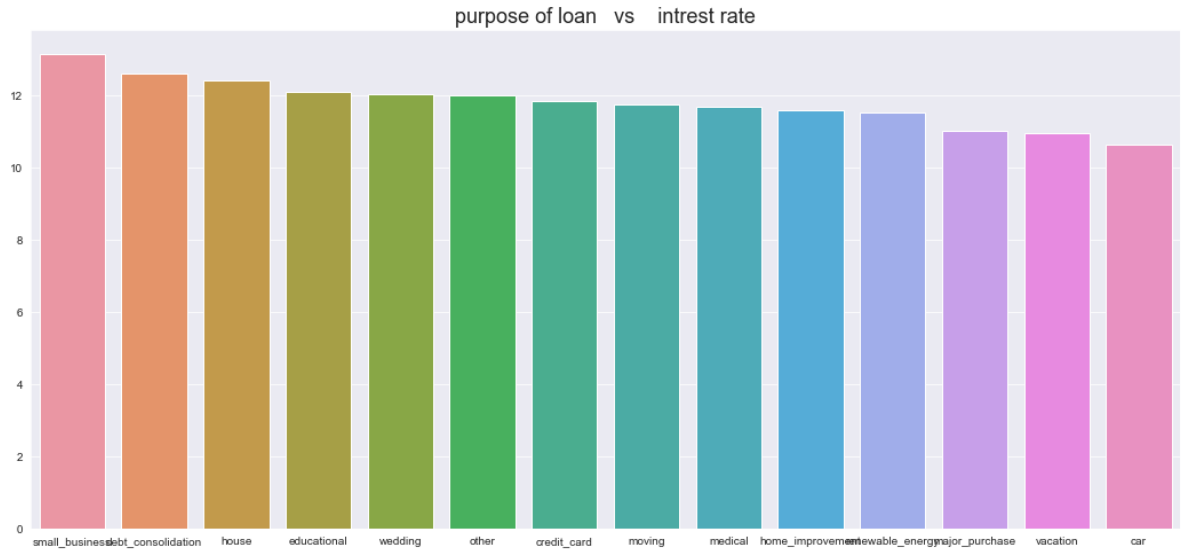

In [95]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='loan_status', y='int_rate', data=df.groupby('loan_status')['int_rate'].r
plt.title('loan status   vs   intrest rate',size=18)
plt.xlabel('')
plt.ylabel('')
plt.show()
```



In [94]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='purpose', y='int_rate', data=df.groupby('purpose')['int_rate'].mean().sc
plt.title('purpose of loan   vs    intrest rate',size=18)
plt.xlabel('')
plt.ylabel('')
plt.show()
```



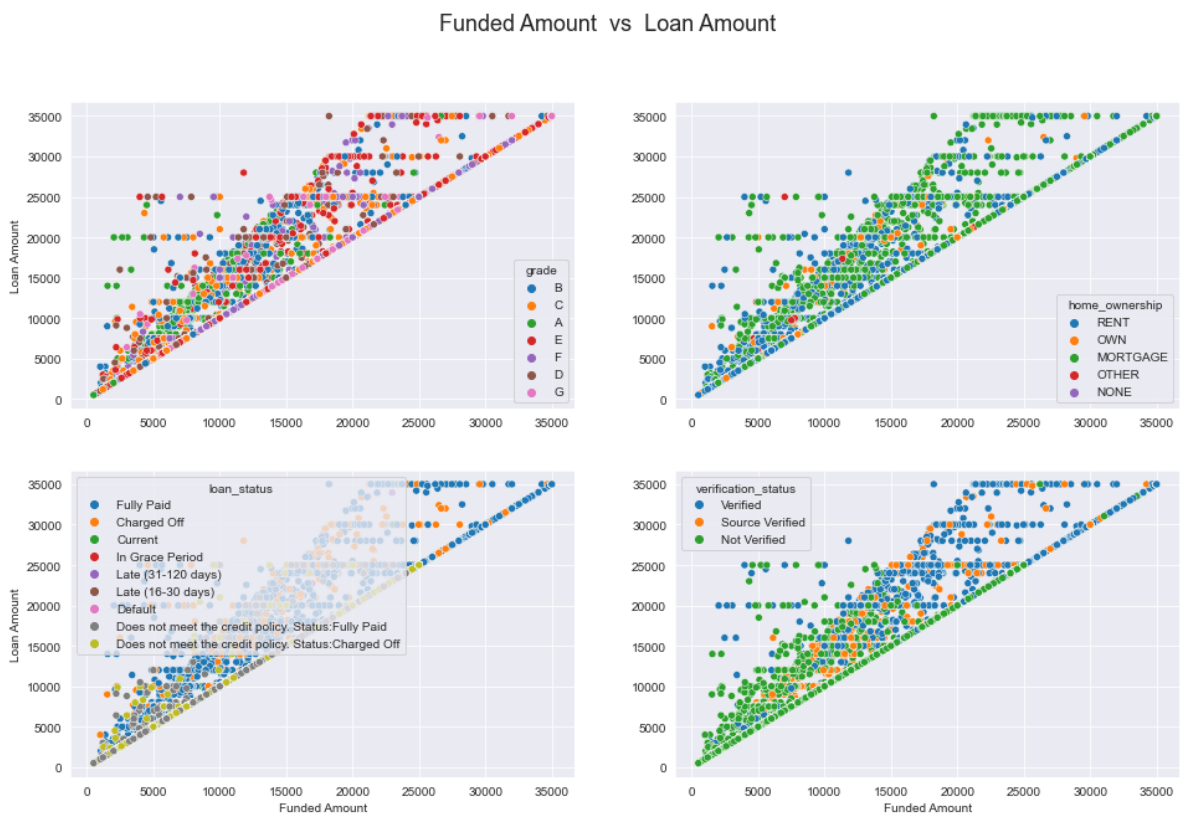
In [64]:

```
for i in df.columns:
    if df[i].dtype!='object':
        print(i)
```

```
id
member_id
loan_amnt
funded_amnt
funded_amnt_inv
term
int_rate
installment
annual_inc
dti
delinq_2yrs
fico_range_low
fico_range_high
inq_last_6mths
open_acc
pub_rec
revol_bal
revol_util
total_acc
out_prncp
out_prncp_inv
total_pymnt
total_pymnt_inv
total_rec_prncp
total_rec_int
total_rec_late_fee
recoveries
collection_recovery_fee
last_pymnt_amnt
last_fico_range_high
last_fico_range_low
pub_rec_bankruptcies
```

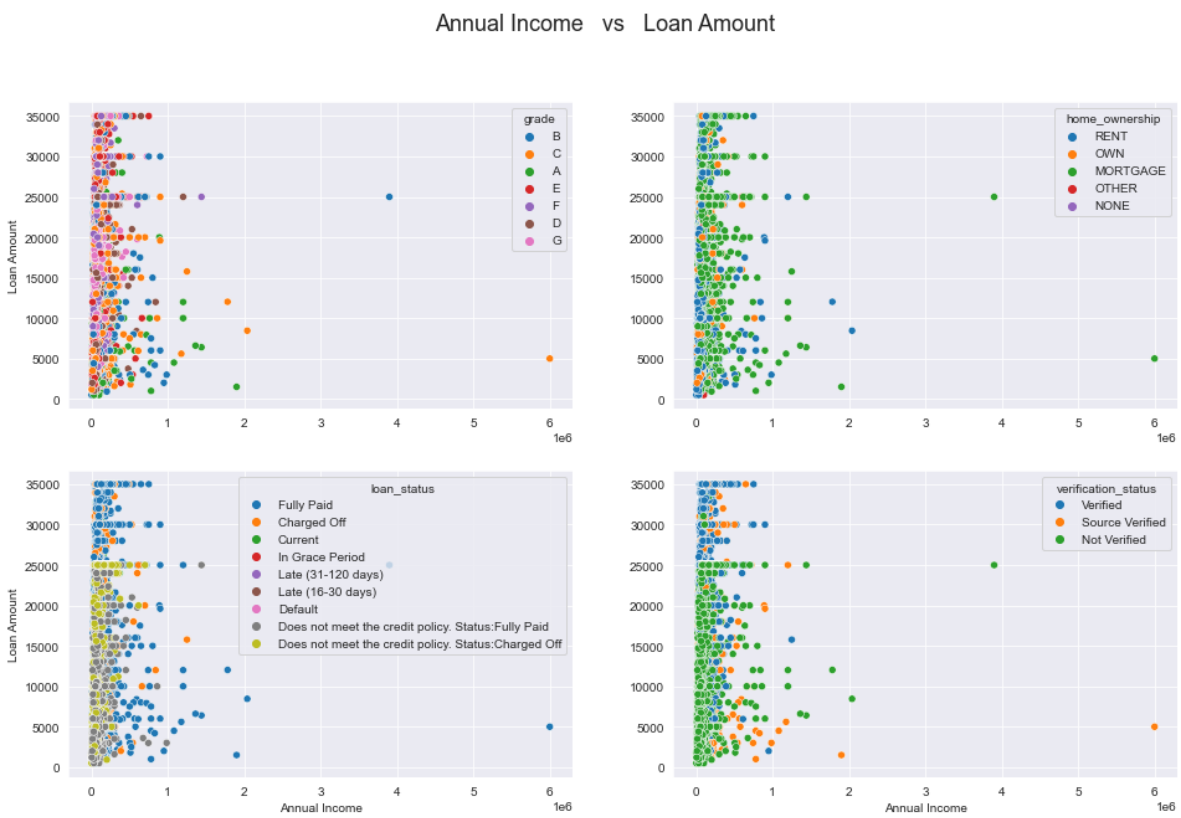
In [101]:

```
plt.figure(figsize=(16,10))

plt.subplot(2,2,1)
plt.suptitle('Funded Amount  vs  Loan Amount',fontsize=18)
sns.scatterplot(x='funded_amnt', y='loan_amnt', hue='grade', data=df)
plt.xlabel('')
plt.ylabel('Loan Amount')

plt.subplot(2,2,2)
sns.scatterplot(x='funded_amnt', y='loan_amnt', hue='home_ownership', data=df)
plt.xlabel('')
plt.ylabel('')

plt.subplot(2,2,3)
sns.scatterplot(x='funded_amnt', y='loan_amnt', hue='loan_status', data=df)
plt.xlabel('Funded Amount')
plt.ylabel('Loan Amount')

plt.subplot(2,2,4)
sns.scatterplot(x='funded_amnt', y='loan_amnt', hue='verification_status', data=df)
plt.xlabel('Funded Amount')
plt.ylabel('')

plt.show()
```



In [102]:

```
plt.figure(figsize=(16,10))

plt.subplot(2,2,1)
plt.suptitle('Annual Income   vs   Loan Amount',fontsize=18)
sns.scatterplot(x='annual_inc', y='loan_amnt', hue='grade', data=df)
plt.xlabel('')
plt.ylabel('Loan Amount')

plt.subplot(2,2,2)
sns.scatterplot(x='annual_inc', y='loan_amnt', hue='home_ownership', data=df)
plt.xlabel('')
plt.ylabel('')

plt.subplot(2,2,3)
sns.scatterplot(x='annual_inc', y='loan_amnt', hue='loan_status', data=df)
plt.xlabel('Annual Income')
plt.ylabel('Loan Amount')

plt.subplot(2,2,4)
sns.scatterplot(x='annual_inc', y='loan_amnt', hue='verification_status', data=df)
plt.xlabel('Annual Income')
plt.ylabel('')

plt.show()
```



In [103]:

```
plt.figure(figsize=(16,10))

plt.subplot(2,2,1)
plt.suptitle('Total payment   vs   Loan Amount',fontsize=18)
sns.scatterplot(x='total_pymnt', y='loan_amnt', hue='grade', data=df)
plt.xlabel('')
plt.ylabel('Loan Amount')

plt.subplot(2,2,2)
sns.scatterplot(x='total_pymnt', y='loan_amnt', hue='home_ownership', data=df)
plt.xlabel('')
plt.ylabel('')

plt.subplot(2,2,3)
sns.scatterplot(x='total_pymnt', y='loan_amnt', hue='loan_status', data=df)
plt.xlabel('Total payment')
plt.ylabel('Loan Amount')

plt.subplot(2,2,4)
sns.scatterplot(x='total_pymnt', y='loan_amnt', hue='verification_status', data=df)
plt.xlabel('Total payment')
plt.ylabel('')

plt.show()
```

In [104]:

```
plt.figure(figsize=(16,10))

plt.subplot(2,2,1)
plt.suptitle('Interest Rate  vs  Loan Amount',fontsize=18)
sns.scatterplot(x='int_rate', y='loan_amnt', hue='grade', data=df)
plt.xlabel('')
plt.ylabel('Loan Amount')

plt.subplot(2,2,2)
sns.scatterplot(x='int_rate', y='loan_amnt', hue='home_ownership', data=df)
plt.xlabel('')
plt.ylabel('')

plt.subplot(2,2,3)
sns.scatterplot(x='int_rate', y='loan_amnt', hue='loan_status', data=df)
plt.xlabel('Interest Rate')
plt.ylabel('Loan Amount')

plt.subplot(2,2,4)
sns.scatterplot(x='int_rate', y='loan_amnt', hue='verification_status', data=df)
plt.xlabel('Interest Rate')
plt.ylabel('')

plt.show()
```

In [105]:

```
plt.figure(figsize=(16,10))

plt.subplot(2,2,1)
plt.suptitle('Total payment  vs  Recovery',fontsize=18)
sns.scatterplot(x='total_pymnt', y='recoveries', hue='grade', data=df)
plt.xlabel('')
plt.ylabel('Recovery')

plt.subplot(2,2,2)
sns.scatterplot(x='total_pymnt', y='recoveries', hue='home_ownership', data=df)
plt.xlabel('')
plt.ylabel('')

plt.subplot(2,2,3)
sns.scatterplot(x='total_pymnt', y='recoveries', hue='loan_status', data=df)
plt.xlabel('Total payment')
plt.ylabel('Recovery')

plt.subplot(2,2,4)
sns.scatterplot(x='total_pymnt', y='recoveries', hue='verification_status', data=df)
plt.xlabel('Total payment')
plt.ylabel('')

plt.show()
```

In [70]:

```
for i in df.columns:
    if df[i].dtype=='object':
        print(i)
```

```
grade
sub_grade
emp_title
emp_length
home_ownership
verification_status
issue_d
loan_status
purpose
title
zip_code
addr_state
earliest_cr_line
last_pymnt_d
last_credit_pull_d
```

In [71]:

```
date_col=['issue_d','last_pymnt_d','last_credit_pull_d']
```

In [72]:

```
for i in date_col:
    df[i] = pd.to_datetime(df[i])
```

In [73]:

```
for i in df.columns:
    if df[i].dtype=='datetime64[ns]':
        print(i)
```

```
issue_d
last_pymnt_d
last_credit_pull_d
```

In [ ]:

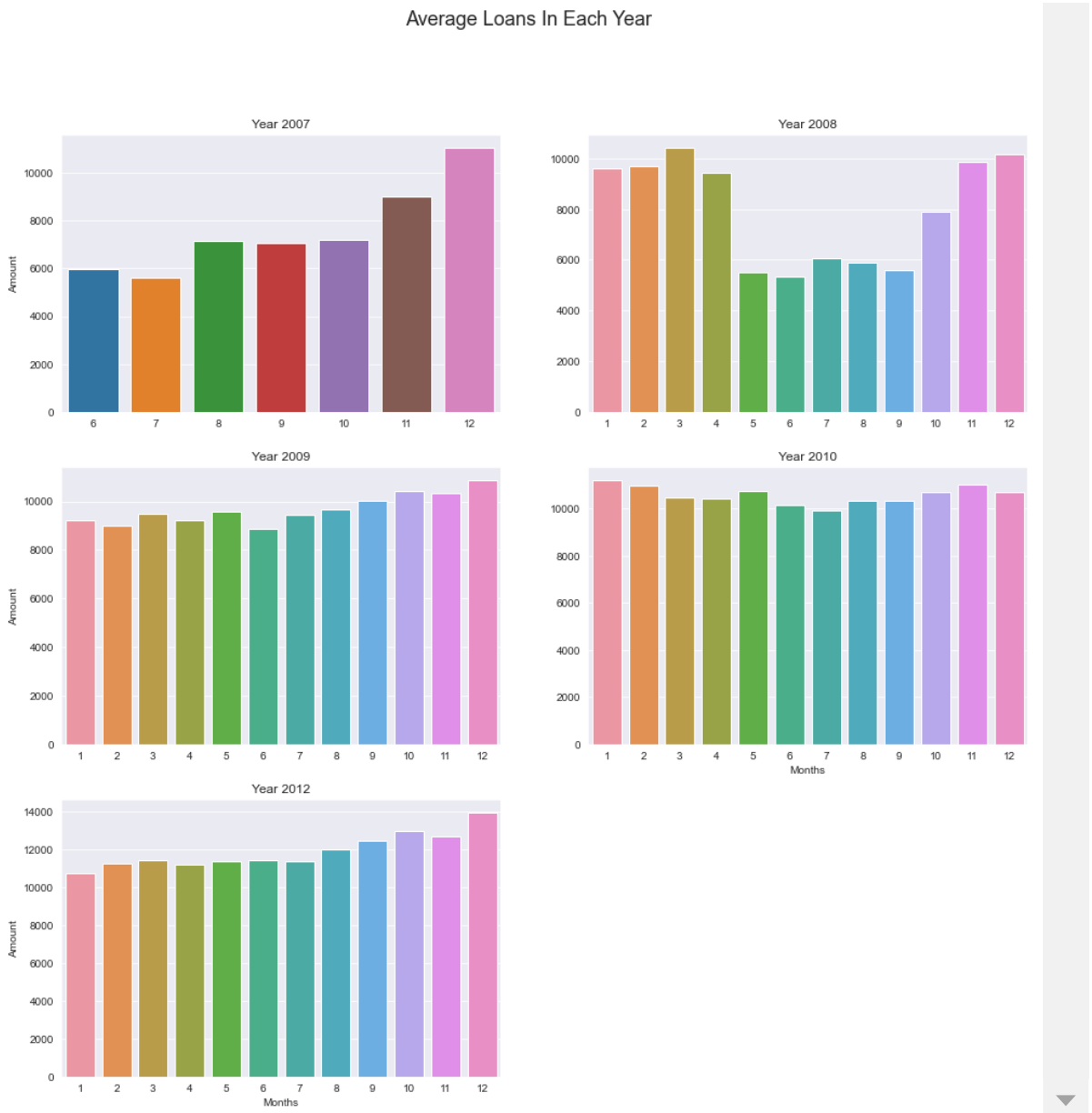
```

```

In [ ]:

```

```

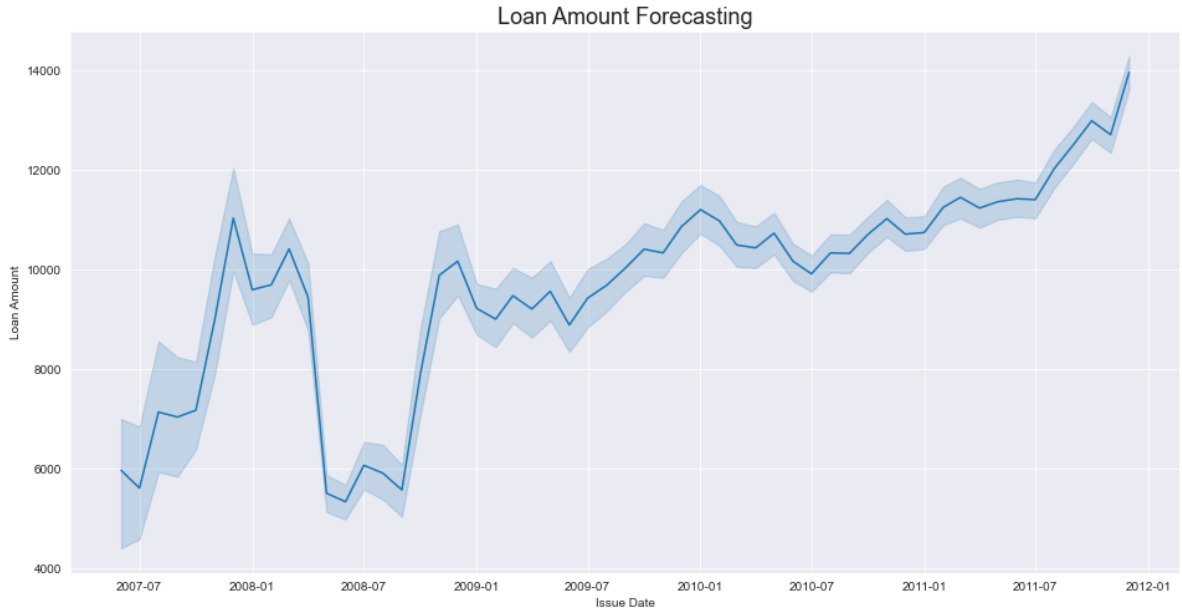
In [106]:

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

plt.subplot(3,2,1)
plt.suptitle('Average Loans In Each Year',fontsize=18)
sns.barplot(x='issue_d', y='loan_amnt', data=df[df['issue_d'].dt.year==2007].groupby(df
plt.title('Year 2007')
plt.xlabel('')
plt.ylabel('Amount')

plt.subplot(3,2,2)
sns.barplot(x='issue_d', y='loan_amnt', data=df[df['issue_d'].dt.year==2008].groupby(df
plt.title('Year 2008')
plt.xlabel('')
plt.ylabel('')

plt.subplot(3,2,3)
sns.barplot(x='issue_d', y='loan_amnt', data=df[df['issue_d'].dt.year==2009].groupby(df
plt.title('Year 2009')
plt.xlabel('')
plt.ylabel('Amount')

plt.subplot(3,2,4)
sns.barplot(x='issue_d', y='loan_amnt', data=df[df['issue_d'].dt.year==2010].groupby(df
plt.title('Year 2010')
plt.xlabel('Months')
plt.ylabel('')

plt.subplot(3,2,5)
sns.barplot(x='issue_d', y='loan_amnt', data=df[df['issue_d'].dt.year==2011].groupby(df
plt.title('Year 2012')
plt.xlabel('Months')
plt.ylabel('Amount')

plt.show()
```


Average Loans In Each Year
Year 2007
Year 2008
Year 2009
Year 2010
Year 2012
Amount
Months

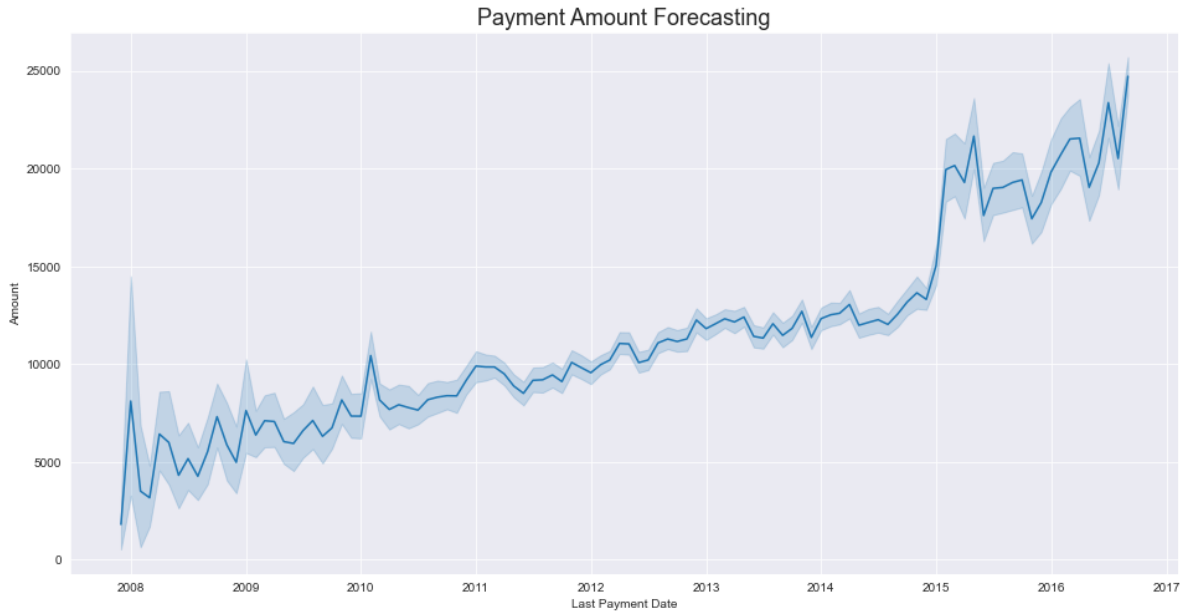
In [107]:

```
plt.figure(figsize=(16,8))
sns.lineplot(x='issue_d', y='loan_amnt', data=df)
plt.title('Loan Amount Forecasting',size=18)
plt.xlabel('Issue Date')
plt.ylabel('Loan Amount')
plt.show()
```



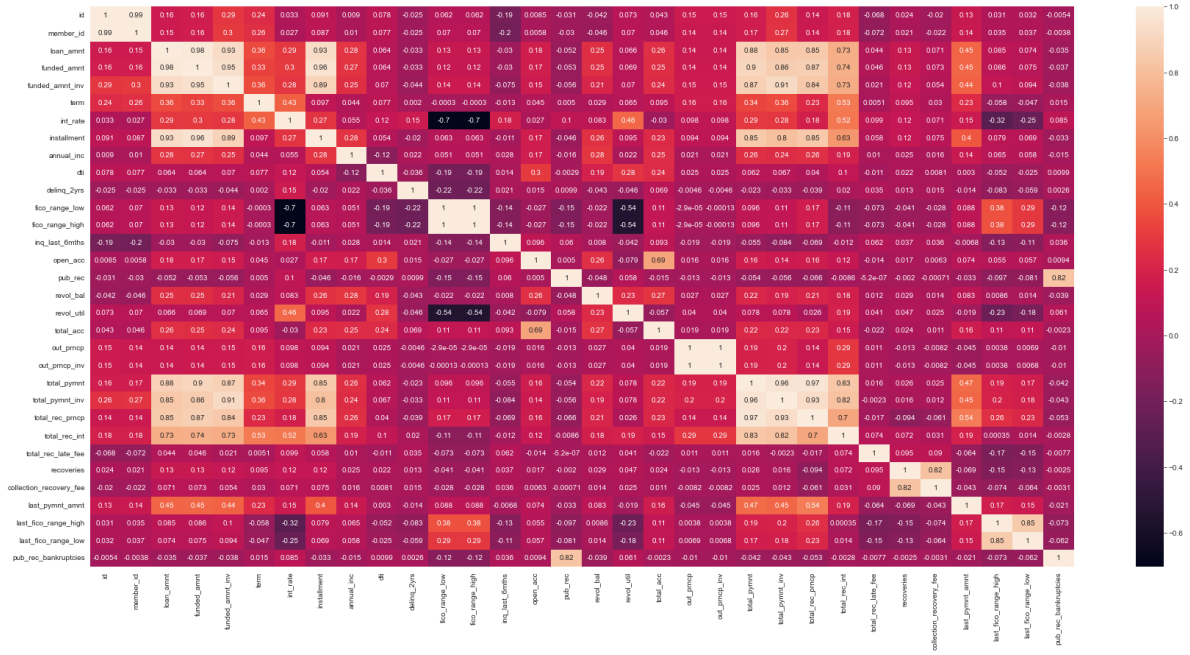

In [108]:

```
plt.figure(figsize=(16,8))
sns.lineplot(x='last_pymnt_d', y='total_pymnt', data=df)
plt.title('Payment Amount Forecasting',size=18)
plt.xlabel('Last Payment Date')
plt.ylabel('Amount')
plt.show()
```



In [77]:

```
plt.figure(figsize=(30,14))
sns.heatmap(df.corr(),annot=True)
plt.show()
```



In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```

```