In [1]:
```python
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
#import pyodbc
import seaborn as sns
%matplotlib inline

#import sqlserver.connector as sq
```

In [ ]:

In [2]:
```python
#df = pd.read_excel("Z:\Godswill\Project\Python\Casualty Project.xlsx")
df = pd.read_excel("C:\Python\Casualty Project.xlsx")
```
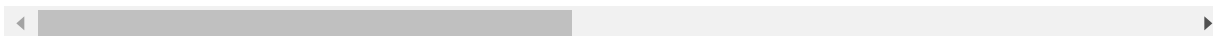
In [91]:
```python
#print(df)

df.head() #first 5 rows
```

Out[91]:

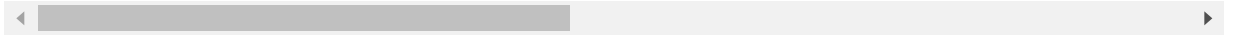|	ExposureID	ClaimID	PolicyID	Claim Number	Exposure Counts	State	Claim Closed Date	Date Of Loss	Loss YY	Loss MM
0	307	406	406	TX-200015	1	TX	2008-05-06 16:45:19.433	2008-01-08 11:10:00	2008	1
1	309	407	407	TX-200017	1	TX	2008-03-29 14:45:10.163	2008-01-18 12:34:00	2008	1
2	396	365	367	TX-200125	1	TX	2008-08-18 10:45:10.383	2008-01-28 16:18:00	2008	1
3	404	301	301	TX-200007	1	TX	2008-05-29 15:15:28.067	2007-12-14 17:34:00	2007	12
4	464	365	367	TX-200125	1	TX	2008-08-18 10:45:10.383	2008-01-28 16:18:00	2008	1

5 rows × 62 columns

In [92]: `df.tail() #last 5 rows`

Out[92]:

	ExposureID	ClaimID	PolicyID	Claim Number	Exposure Counts	State	Claim Closed Date	Date Of Loss	Loss YY	Loss MM
96122	932779	582732	613954	AZ-662690	1	AZ	NaT	2021-04-28 17:00:00	2021	4
96123	932815	582749	614274	GA-662816	1	GA	NaT	2021-04-27 18:43:00	2021	4
96124	932867	582192	613800	OK-662673	1	OK	NaT	2021-04-28 18:50:00	2021	4
96125	932868	582192	613800	OK-662673	1	OK	NaT	2021-04-28 18:50:00	2021	4
96126	932899	582819	613829	SC-662811	1	SC	NaT	2021-04-29 17:18:00	2021	4

5 rows × 62 columns

In [93]:
```python
for col in df.columns:
    print(col)
```

ExposureID
ClaimID
PolicyID
Claim Number
Exposure Counts
State
Claim Closed Date
Date Of Loss
Loss YY
Loss MM
Loss DD
Loss Qtr
Claim Reported Date
Policy Deductible
Exposure Limit
LOB
LOB Category
Exposure
Exposure Reported Date
Exposure Closing Reason
Exposure Status
Casualty General Damages High
Casualty General Damages Low
Casualty Initial Demand
Casualty Initial Demand Date
Casualty Initial Offer Amount
Casualty Initial Offer Date
Casualty Settlement Amount
Casualty Last Demand Amount
Casualty Last Offer Amount
Agent City
Casualty Total Settlement Value High
Casualty Total Settlement Value Low
Casualty Medical Billing Submitted
Casualty Medical Billing Allowed
Insured Physical City
Loss City
ExposureCity
Claim Cost CWP Attorney Repd
Claim Cost No Attorney Repd
Claim Cost CWP No Attorney Repd
Total Payments Attorney Repd
Total Payments No Attorney Repd
Total Payments CWP Attorney Repd
Total Paymennts CWP No Attorney Repd
Total Payments CWP
Claim Cost Attorney repd
Exposure Type
Loss Cause(Formula)
LOB(No Covs)
InjuryType1: Duplicate
General Injury Type
InjuryType1ID
InjuryType2ID
InjuryType3ID
Injury Type2
Injury Type3

```
                    Injury Type
                    Soft Tissue Indicator
                    Reserves Loss Range
                    Casualty Medical Billing Submitted-CAP
                    Casualty Medical Billing Allowed-CAP
```

In [94]: `df.dtypes   #datatypes`

Out[94]:
```
         ExposureID                                int64
         ClaimID                                   int64
         PolicyID                                  int64
         Claim Number                              object
         Exposure Counts                           int64
                                                   ...
         Injury Type                               object
         Soft Tissue Indicator                     object
         Reserves Loss Range                       object
         Casualty Medical Billing Submitted-CAP    float64
         Casualty Medical Billing Allowed-CAP      float64
         Length: 62, dtype: object
```

In [95]: `df.info()` *#display number of entries of datatypes in columns*

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96127 entries, 0 to 96126
Data columns (total 62 columns):
 #   Column                                 Non-Null Count   Dtype
---  ------                                 --------------   -----
 0   ExposureID                             96127 non-null   int64
 1   ClaimID                                96127 non-null   int64
 2   PolicyID                               96127 non-null   int64
 3   Claim Number                           96127 non-null   object
 4   Exposure Counts                        96127 non-null   int64
 5   State                                  96127 non-null   object
 6   Claim Closed Date                      88925 non-null   datetime64[ns]
 7   Date Of Loss                           96127 non-null   datetime64[ns]
 8   Loss YY                                96127 non-null   int64
 9   Loss MM                                96127 non-null   int64
 10  Loss DD                                96127 non-null   int64
 11  Loss Qtr                               96127 non-null   object
 12  Claim Reported Date                    96127 non-null   datetime64[ns]
 13  Policy Deductible                      96118 non-null   float64
 14  Exposure Limit                         96118 non-null   float64
 15  LOB                                    96127 non-null   int64
 16  LOB Category                           96120 non-null   object
 17  Exposure                               96127 non-null   int64
 18  Exposure Reported Date                 96127 non-null   datetime64[ns]
 19  Exposure Closing Reason                89212 non-null   object
 20  Exposure Status                        96127 non-null   object
 21  Casualty General Damages High          94714 non-null   float64
 22  Casualty General Damages Low           94715 non-null   float64
 23  Casualty Initial Demand                33934 non-null   float64
 24  Casualty Initial Demand Date           33934 non-null   datetime64[ns]
 25  Casualty Initial Offer Amount          61310 non-null   float64
 26  Casualty Initial Offer Date            61310 non-null   datetime64[ns]
 27  Casualty Settlement Amount             52585 non-null   float64
 28  Casualty Last Demand Amount            33934 non-null   float64
 29  Casualty Last Offer Amount             61311 non-null   float64
 30  Agent City                             96063 non-null   object
 31  Casualty Total Settlement Value High   94796 non-null   float64
 32  Casualty Total Settlement Value Low    94796 non-null   float64
 33  Casualty Medical Billing Submitted     46342 non-null   float64
 34  Casualty Medical Billing Allowed       45541 non-null   float64
 35  Insured Physical City                  94825 non-null   object
 36  Loss City                              96126 non-null   object
 37  ExposureCity                           95701 non-null   object
 38  Claim Cost CWP Attorney Repd           32112 non-null   float64
 39  Claim Cost No Attorney Repd            52450 non-null   float64
 40  Claim Cost CWP No Attorney Repd        28905 non-null   float64
 41  Total Payments Attorney Repd           40500 non-null   float64
 42  Total Payments No Attorney Repd        52450 non-null   float64
 43  Total Payments CWP Attorney Repd       32504 non-null   float64
 44  Total Paymennts CWP No Attorney Repd   29173 non-null   float64
 45  Total Payments CWP                     64853 non-null   float64
 46  Claim Cost Attorney repd               40500 non-null   float64
 47  Exposure Type                          96127 non-null   object
 48  Loss Cause(Formula)                    96127 non-null   object
 49  LOB(No Covs)                           96127 non-null   int64
 50  InjuryType1: Duplicate                 0 non-null       float64
 51  General Injury Type                    96126 non-null   object
```

```
52  InjuryType1ID                            24468 non-null  float64
53  InjuryType2ID                              438 non-null  float64
54  InjuryType3ID                              105 non-null  float64
55  Injury Type2                               438 non-null  object
56  Injury Type3                               105 non-null  object
57  Injury Type                              24468 non-null  object
58  Soft Tissue Indicator                    96127 non-null  object
59  Reserves Loss Range                      96127 non-null  object
60  Casualty Medical Billing Submitted-CAP   46343 non-null  float64
61  Casualty Medical Billing Allowed-CAP     45542 non-null  float64
dtypes: datetime64[ns](6), float64(28), int64(10), object(18)
memory usage: 45.5+ MB
```

In [17]: `df['Casualty Settlement Amount']`

Out[17]:
```
0           NaN
1           NaN
2         500.0
3           NaN
4         500.0
          ...
96123       NaN
96124       NaN
96125       NaN
96126       NaN
96127       NaN
Name: Casualty Settlement Amount, Length: 96128, dtype: float64
```

In [96]: `df.describe()`

Out[96]:

	ExposureID	ClaimID	PolicyID	Exposure Counts	Loss YY	Loss MM	
count	96127.000000	96127.000000	96127.000000	96127.0	96127.000000	96127.000000	961
mean	492390.604086	297455.787188	310630.258772	1.0	2014.874395	6.532036	
std	274850.059129	169850.630782	178103.605913	0.0	3.923739	3.424175	
min	307.000000	301.000000	301.000000	1.0	2003.000000	1.000000	
25%	249607.500000	145313.500000	151779.500000	1.0	2011.000000	4.000000	
50%	512566.000000	301928.000000	313688.000000	1.0	2015.000000	6.000000	
75%	735411.500000	447516.000000	466682.000000	1.0	2018.000000	10.000000	
max	932961.000000	582819.000000	614274.000000	1.0	2021.000000	12.000000	

8 rows × 38 columns

Graphs

```
In [22]: df['Casualty General Damages High'].value_counts()
         #sns.countplot(x='Casualty General Damages Low', data=df)
```

```
Out[22]: 0.00          45127
         1500.00        5071
         1000.00        4495
         2000.00        4117
         2500.00        2928
                        ...
         4750.75           1
         19754.26          1
         4751.00           1
         2140.00           1
         5831.76           1
         Name: Casualty General Damages High, Length: 7369, dtype: int64
```

Casualty Settlement Amount within GAINSCO States

```
In [97]: sns.barplot(x= 'State', y= 'Casualty Settlement Amount', data=df)
```

```
Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0x2648904d670>
```
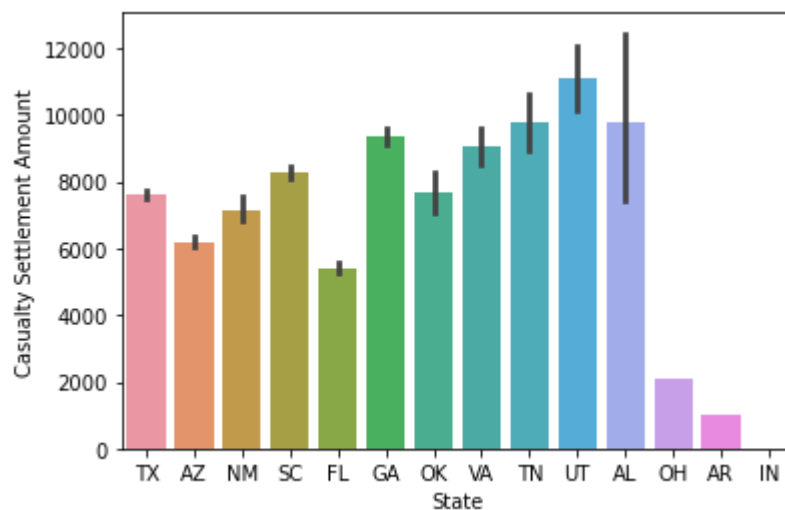

Relationship between Casualty settlement amounts and Soft Tissue Indicator

In [98]:
```
sns.barplot(x= 'Soft Tissue Indicator', y= 'Casualty Settlement Amount', data=
df)
```

Out[98]: <matplotlib.axes._subplots.AxesSubplot at 0x2648f9a83d0>
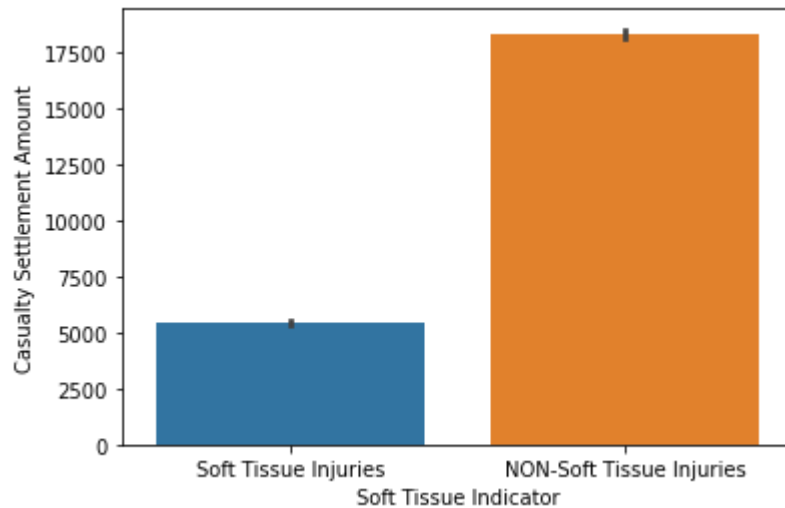
Casualty Settlement Amount by Loss Year

In [102]:
```
fig_dims = (14, 6)
fig, ax = plt.subplots(figsize=fig_dims)
sns.barplot(x = 'Loss YY', y= 'Casualty Settlement Amount', ax=ax, data=df)
```

Out[102]: <matplotlib.axes._subplots.AxesSubplot at 0x26574d787f0>
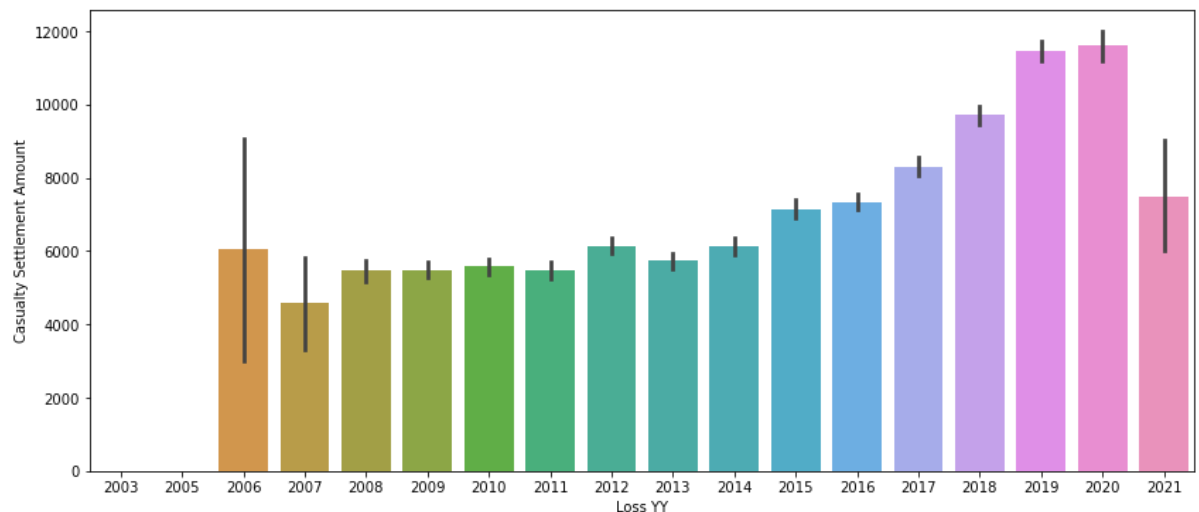
General Injury Type Vs Casualty Settlement Amount

In [12]:
```
fig_dims = (16, 8)
fig, ax = plt.subplots(figsize=fig_dims)
sns.barplot(x = 'Casualty Settlement Amount', y= 'General Injury Type', ax=ax,
data=df)
```

Out[12]:   <matplotlib.axes._subplots.AxesSubplot at 0x22888aee2b0>
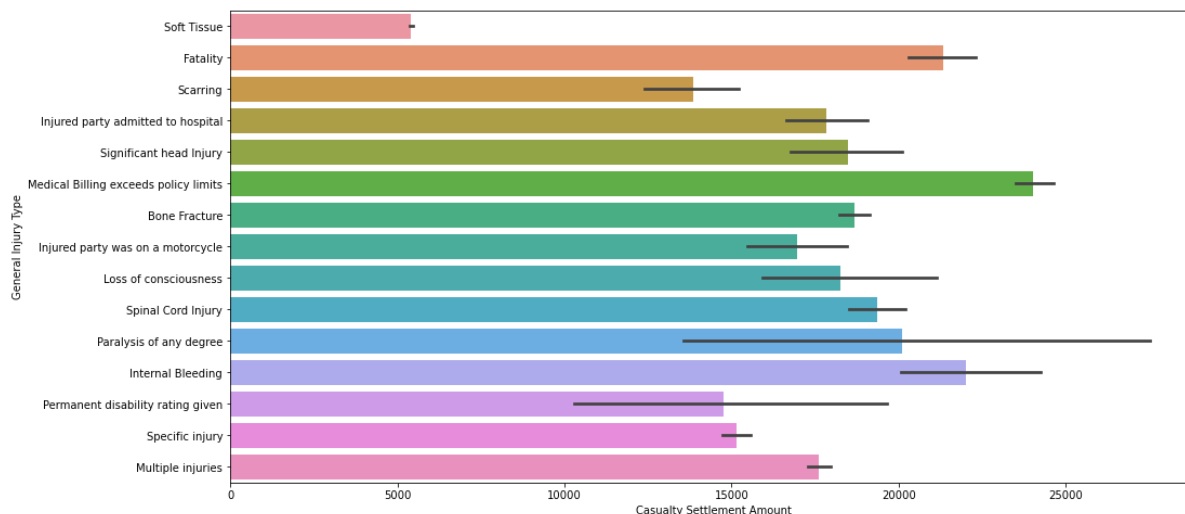
In [ ]:

Total Payments Attorney Repd vs Loss YY

In [ ]:
```
fig_dims = (14, 6)
fig, ax = plt.subplots(figsize=fig_dims)
sns.barplot(x= 'Loss YY', y= 'Total Payments Attorney Repd', data=df)
```

Total Payments No Attorney Repd vs Loss YY

In [ ]:
```
fig_dims = (14, 6)
fig, ax = plt.subplots(figsize=fig_dims)
sns.barplot(x= 'Loss YY', y= 'Total Payments No Attorney Repd', data=df)
```

Exposure Closing Reason vs Casualty Last Offer Amount

In [104]:
```
fig_dims = (14, 6)
fig, ax = plt.subplots(figsize=fig_dims)
sns.barplot(x= 'Casualty Last Offer Amount', y= 'Exposure Closing Reason', dat
a=df.head(65000))
```

Out[104]:   `<matplotlib.axes._subplots.AxesSubplot at 0x2657efdeaf0>`
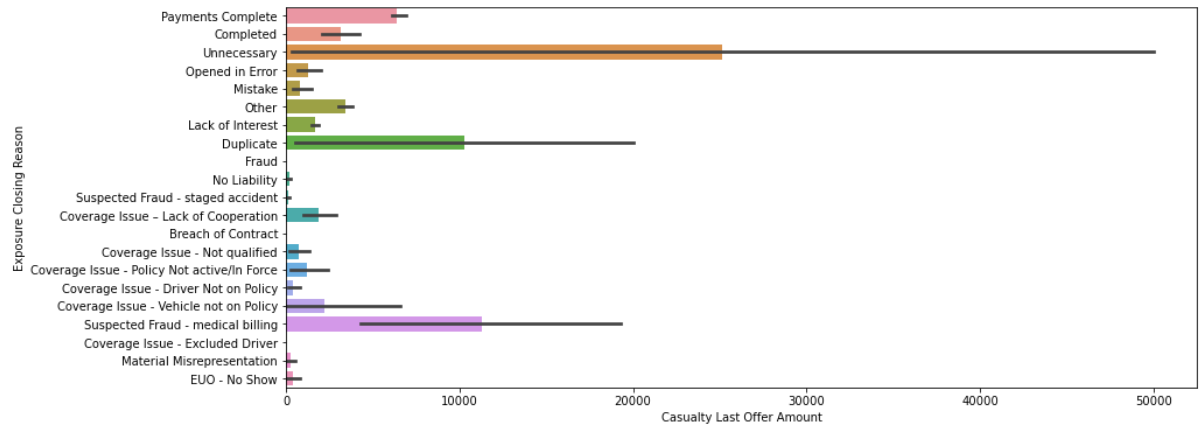
In [ ]:

In [ ]:

In [ ]:
```
#plt.figure(figsize = (7,5))
fig, ax = plt.subplots(figsize=fig_dims)
plt.subplot(1,2,1)
sns.countplot('General Injury Type', hue = 'Casualty Last Offer Amount', data=
df)
```

```
In [25]: plt.figure(figsize = (10,5))
         plt.subplot(1,2,1)
         sns.barplot('Casualty Last Offer Amount', hue = 'General Injury Type', data=df
         )
```

Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x2647f915ee0>
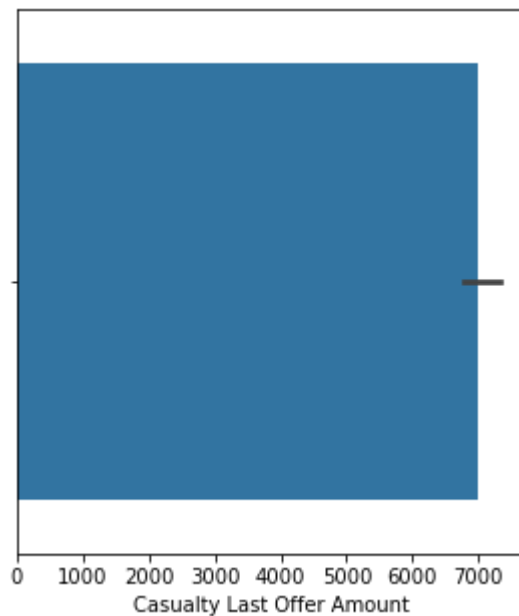
Categorical and Quantitative Variables

```
In [50]:  #Categorical #Quantitative Variables
          x = [5,7,8,7,2,17,2,9,4,11,12,9,6]
          y = [99,86,87,88,111,86,103,87,94,78,77,85,86]
          fig_dims = (10, 6)
          fig, ax = plt.subplots(figsize=fig_dims)
          plt.scatter(x, y)
          plt.xlabel('Total Payments CWP Attorney Repd')
          plt.ylabel('Total Payments CWP No Attorney Repd')
          plt.title('Total Payments CWP')
          #plt.legend(loc='upper right')
          plt.show()
```
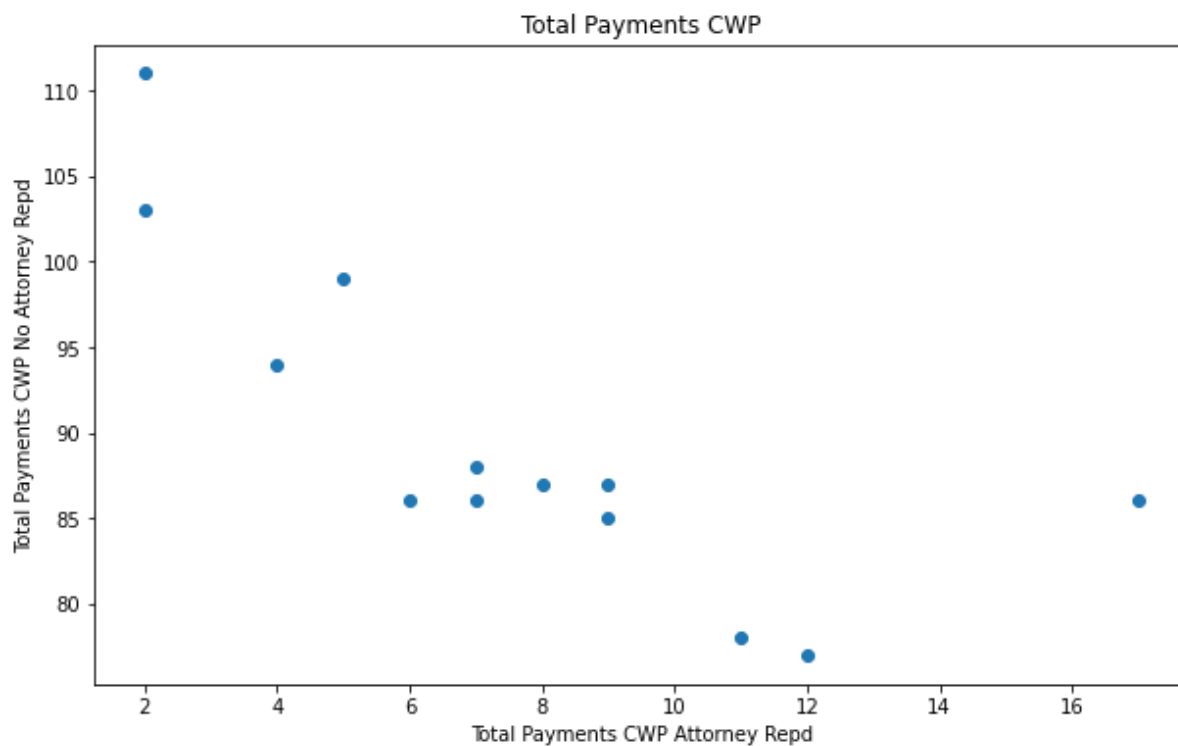

Effect of Loss Cause(Formula) on Casualty Settlement Amount

```
In [129]:  MD=df.groupby('Loss Cause(Formula)',as_index=False).agg({ 'Casualty Settlement
           Amount': 'sum'})
```

```
In [126]:  MD['Casualty_Settlement Amount'] = np.log(MD['Casualty Settlement Amount'])
```

In [128]:
```python
MD.plot.barh(x='Loss Cause(Formula)', y='Casualty_Settlement Amount', rot=0)
#plt.xlabel('Casualty Settlement Amount')
#plt.ylabel('natural_log_Claim Cost')
```
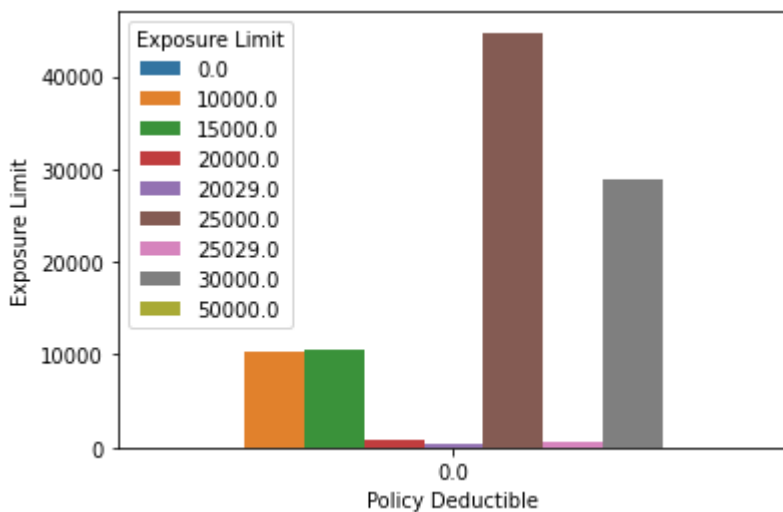
Out[128]: `<matplotlib.axes._subplots.AxesSubplot at 0x1d870c3b790>`
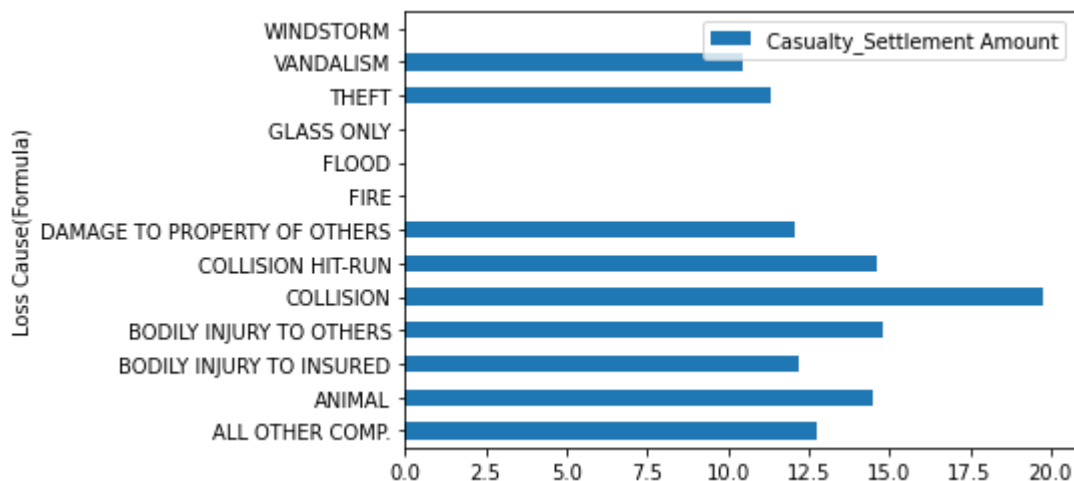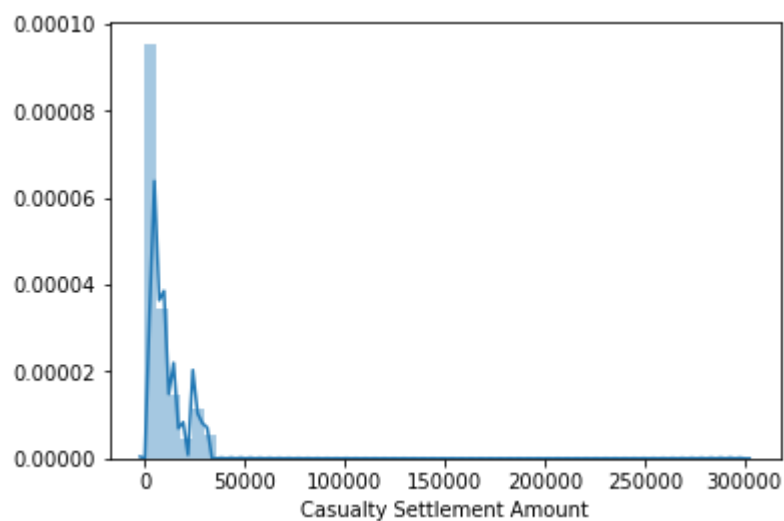

In [ ]:

Exposure Limit vs Policy Deductible

In [130]:
```python
sns.countplot(x='Policy Deductible', hue= 'Exposure Limit', data=df)
plt.xlabel('Policy Deductible')
plt.ylabel('Exposure Limit')
```
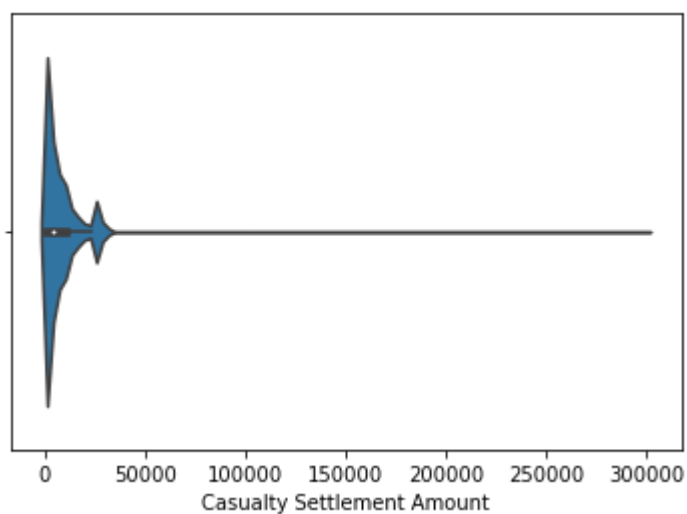
Out[130]: `Text(0, 0.5, 'Exposure Limit')`

In [21]: `sns.distplot(df['Casualty Settlement Amount'])`

Out[21]: `<matplotlib.axes._subplots.AxesSubplot at 0x1d218f59d30>`

In [22]: `sns.violinplot(df['Casualty Settlement Amount'])`

Out[22]: `<matplotlib.axes._subplots.AxesSubplot at 0x1d20b2cf550>`

Demographics

In [34]: `df['ExposureCity'].value_counts()`

Out[34]:
```
Houston             4145
Phoenix             2832
unk                 2328
Dallas              2108
Tucson              1452
                     ...
BENNESTTSVILLE         1
ELBERTON               1
las cruces             1
Columia                1
West Hollywood         1
Name: ExposureCity, Length: 6982, dtype: int64
```
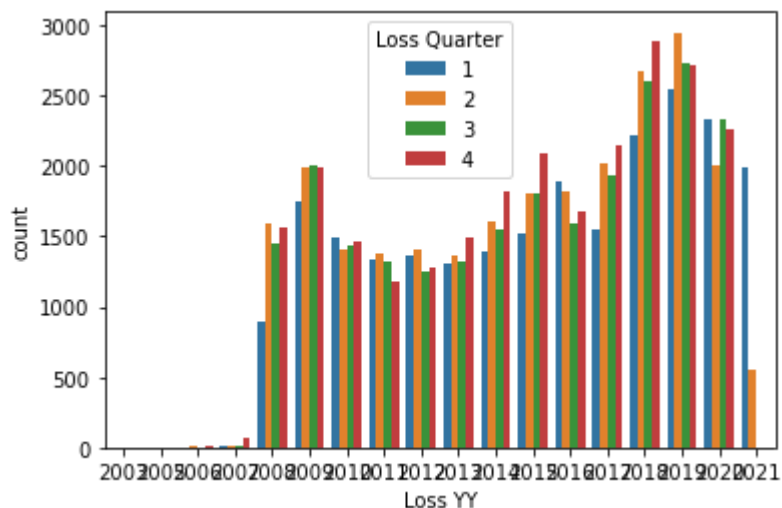
In [29]: `df["State"].unique`

Out[29]:
```
<bound method Series.unique of 0          TX
1          TX
2          TX
3          TX
4          TX
           ..
96122      AZ
96123      GA
96124      OK
96125      OK
96126      SC
Name: State, Length: 96127, dtype: object>
```

Year

In [54]:
```
sns.countplot('Loss YY', hue= 'Loss Quarter', data=df)
#plt.xlabel('exposure city')
#plt.ylabel('state')
```

Out[54]: `<matplotlib.axes._subplots.AxesSubplot at 0x1d2ad599a30>`

In [48]: `df['Loss Quarter']`

Out[48]:
```
0          1
1          1
2          1
3          4
4          1
          ..
96122      2
96123      2
96124      2
96125      2
96126      2
Name: Loss Quarter, Length: 96127, dtype: int64
```

In [44]:
```python
df.columns = [col.strip() for col in df.columns]  #strips out wide spaces in c
olumns names as regards to List comprehension
df.columns
```

Out[44]:
```
Index(['ExposureID', 'ClaimID', 'PolicyID', 'Claim Number', 'Exposure Count
s',
       'State', 'Claim Closed Date', 'Date Of Loss', 'Loss YY', 'Loss MM',
       'Loss DD', 'Loss Qtr', 'Claim Reported Date', 'Policy Deductible',
       'Exposure Limit', 'LOB', 'LOB Category', 'Exposure',
       'Exposure Closed prYY', 'Exposure Closed prYY-1',
       'Exposure Reported Date', 'Exposure Reported prYY',
       'Exposure Reported prYY-1', 'Casualty General Damages High',
       'Casualty General Damages Low', 'Casualty Initial Demand',
       'Casualty Initial Demand Date', 'Casualty Initial Offer Amount',
       'Casualty Initial Offer Date', 'Casualty Settlement Amount',
       'Casualty Last Demand Amount', 'Casualty Last Offer Amount',
       'Casualty Total Settlement Value High',
       'Casualty Total Settlement Value Low',
       'Casualty Medical Billing Submitted',
       'Casualty Medical Billing Allowed', 'Loss Cause(Formula)',
       'Reserves Loss Range', 'Casualty Medical Billing Submitted-CAP',
       'Casualty Medical Billing Allowed-CAP'],
      dtype='object')
```

In [55]: `df.loc[1:5,'Casualty Settlement Amount'] #display columns 1-5 of veh damage...`
`column`

Out[55]:
```
1        NaN
2      500.0
3        NaN
4      500.0
5        NaN
Name: Casualty Settlement Amount, dtype: float64
```

In [ ]:

In [ ]:

In [ ]: