

```
In [1]: 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 [2]: #df = pd.read_excel("Z:\Godswill\Project\Python\Casualty Project.xlsx")
df = pd.read_excel("C:\Python\Casualty Project.xlsx")

In [ ]: #print(df)

df.head() #first 5 rows

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

In [ ]: df.dtypes #datatypes

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

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

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

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

## Graphs

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

## Casualty Settlement Amount within Company States

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

## Relationship between Casualty settlement amounts and Soft Tissue Indicator

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

## Casualty Settlement Amount by Loss Year

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

## General Injury Type Vs Casualty Settlement Amount

```
In [ ]: 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)
```

## 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 [ ]: 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))
```

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

## Categorical and Quantitative Variables

```
In [ ]: #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 [ ]: MD.plot.barh(x='Loss Cause(Formula)', y='Casualty_Settlement Amount', rot=0)
#plt.xlabel('Casualty Settlement Amount')
#plt.ylabel('natural_Log_Claim Cost')
```

```
In [ ]:
```

## Exposure Limit vs Policy Deductible

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

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

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

## Demographics

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

```
In [ ]: df["State"].unique
```

## Year

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

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

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

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