```
%load_ext autoreload
%autoreload 2

import pandas as pd
import numpy as np

import dsba6211project.Utils.getData as gd

df = gd.GetData().getOpportunityData()

df.shape


from plotnine import *

df.describe()



tooManyMissing = list(df.isna().sum()[df.isna().sum() > 10000].index)
tooManyMissing

df = df.drop(tooManyMissing, axis=1)
df.shape


from sklearn.impute import SimpleImputer
df.loc[:,:] = SimpleImputer(strategy="most_frequent").fit_transform(df)

df.describe()

for c in df.columns[:5]:
    if c in ["Amount" , "AccountId" , "CloseDate"]:
        continue
    print(c)
    p = ggplot(df , aes(x=c,y="Amount")) + geom_point()
    print(p)

```