**Data Wrangling Steps Done as part of Capstone1**

1. Initial Setup

Imported all the required libraries and read the csv file as a Pandas dataframe.

import numpy as np

import pandas as pd

from matplotlib import pyplot as plt

import matplotlib

%matplotlib inline

import seaborn as sns

data = pd.read\_csv("Desktop/Python/startup\_funding.csv")

Understanding the shape and format of data

data.head

data.shape

2.Exploring the data in a Dataframe

We can access the data types of each column in a Dataframe as follows

data.dtype

We can display the index, columns, and the underlying numpy data separately:

data.index

data.columns

data.values

To get a quick statistical summary of your data, use the .describe() function

data.describe()

drop columns with all values as NAN

data.dropna(how='all')

data.fillna(method='ffill')

To remove duplicate rows:

data.drop\_duplicates()

To convert the names of columns in upper case:

data.rename(columns=str.upper)

To divide the investments into buckets/bins , created

investment = [100000, 250000, 500000, 1000000]

converted the string into float after replacing , with nothing

data["AmountInUSD"] = data["AmountInUSD"].apply(lambda x: float(str(x).replace(",","")))

Note that pd.value\_counts(cats) are the bin counts for the result of pandas.cut.

cats = pd.cut(data["AmountInUSD"], investment)

pd.value\_counts(cats)