Lab1-Isaac Stertzbach

Task 1: Table is in the code

import pandas as pd

import ipaddress

import seaborn as sns

data = pd.read\_csv("majestic\_million.csv")

list = ["IDN\_Domain", "IDN\_TLD", "PrevGlobalRank", "PrevRefSubNets", "PrevRefIPs", "PrevTldRank"]

data = data.drop(list, axis=1)

Task 2: Table is in the code

data.loc[51:61]

Task 3: Answer: 16149

print(len(pd.unique(data["RefIPs"])))

Task 4: Answer: 12052

print(len(pd.unique(data["RefSubNets"])))

Task 5: Answer: 0.7.134.169

maxRefSubNet = data["RefSubNets"].max() # Question 5

print(ipaddress.ip\_address(maxRefSubNet))

Task 6: Answer: 0.0.1.9

minRefIp = data["RefIPs"].min() # Question 6

print(ipaddress.ip\_address(minRefIp))

Task 7: Answer: 88363

print(len(data[data.TLD.isin(["edu","org"])])) # Question 7

count\_edu = len(data[data.TLD.isin(["edu"])])

count\_org = len(data[data.TLD.isin(["org"])])

new\_data = data[data.TLD.isin(["edu","org"])]

Task 8: Answer: org – 0.951507

Edu – 0.048493

new\_data['TLD'].value\_counts(normalize=True)

Task 9: Graph is in the code

new\_data['TLD'].value\_counts().plot(kind='bar')

Task 10: Graph is in the code

sns.boxplot(x=new\_data['GlobalRank'], y=new\_data['TLD'], showmeans=True);