K-Means

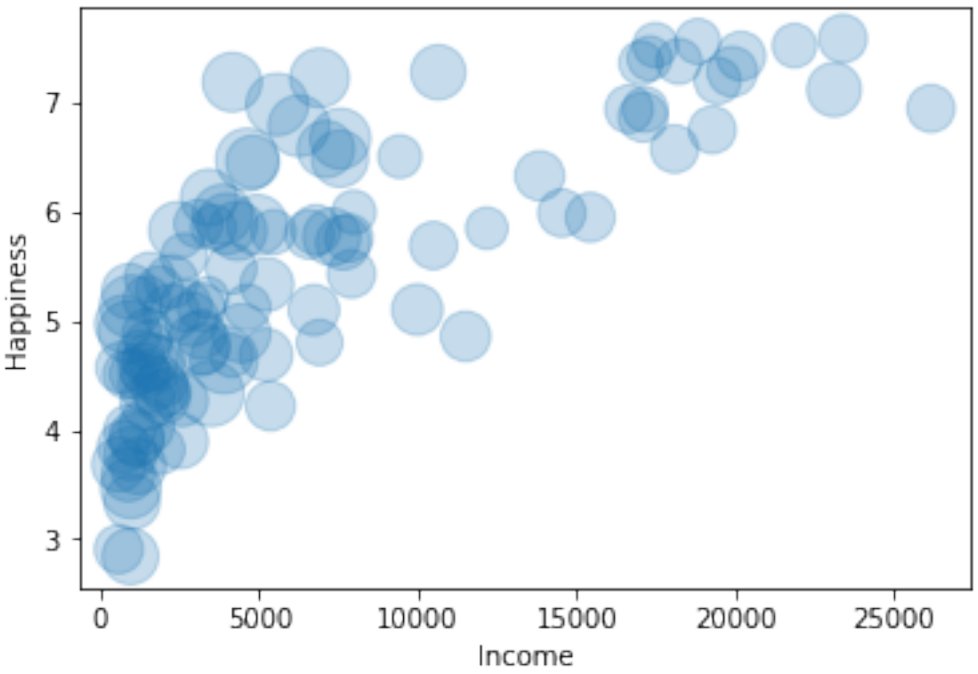
August 11, 2020

In [3]:
```
#by eyeballing the graph we can justify that, there're countries who're well off and a
#low income but are also as happy as the high income countries,
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('happyscore_income.csv')

happy = data['happyScore']
income = data['avg_income']
inequality = data['income_inequality']

plt.xlabel("Income")
plt.ylabel("Happiness")
plt.scatter(income, happy, s=inequality*10, alpha=0.25)
plt.show()
```



In [15]:
```
from sklearn.cluster import KMeans
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

data = pd.read_csv('happyscore_income.csv')

happy = data['happyScore']
income = data['avg_income']
inequality = data['income_inequality']

#for k means this format is needed
income_and_happiness = np.column_stack((income, happy))
#print(income_and_happiness)

km_result = KMeans(n_clusters=3).fit(income_and_happiness)

clusters = km_result.cluster_centers_

plt.scatter(income, happy)
plt.scatter(clusters[:,0], clusters[:,1], s=1000)
plt.show()
```