

1. City Block : 2

Euclidean: $\sqrt{2}$

Supremum: 1

Cosine: $\frac{1}{2}$

Simple: $\frac{8}{10}$

Jaccard: $\frac{1}{3}$

2. a) $SAC = \frac{9}{10}$

$$J = \frac{2}{3}$$

$$\cos \approx 0.76 \quad \text{by } P.\text{corr}(Q)$$

$$s_{\cos} = \frac{2}{\sqrt{6}}$$

b) Jaccard coefficient because there are a lot of 0s and we want to ignore them

c) No $S(P,Q)$ is too sensitive

- | Even for a single miss match, the measure is pegged at 0

3. a) Lay out the numeric data in memory

b) Pandas

c) dask

d) Scikit-learn

e) IPython, Jupyter Notebook