**Introduction**

Key questions: what is the effect of news on an asset (say ticker: APPL) price.

Start with whitelist of “reputable” resources.

Model is built on bitcoin (then expanded to other asset classes)

Q1 filtering (more traditional programming project)

Q2 gives a relevance score (learn what matters)

Q3 gives value score (rank)

**Tentative Steps:**

Build a model for selecting what constitutes relevant information

1. Flag a rally (positive or negative) by statistical analysis
2. Establish an initial data set (event\_i) within said rally
   1. Flag dates/times that deviated from mean (to be calculated)
   2. Use those dates and times to filter through repuable resources (previously white listed) consider an X-hr window of times
   3. Create a model for predicting price change at event i
3. Establish a data set (event\_i+j) within rally for j = 1, …, n
   1. Repeat process from (2) -- so we flag based on ‘sufficien’ deviation from the mean, and continue recording so long as the data compels us to (the rally remains)

Once we have established a model, we can indentify what it is on the articles that matters, and what the effect is on the price.

Quetions:

How do we structure the data?

Model (for Q2):

1. will need an ensemble model
2. Will need to consider time series aspect (daily?)