# Project Proposal

## Faith, Hope and Love in TED Talks

Sean Davern
Seattle, Fall Cohort, weeks 6.5-9

The words faith, hope and love show up prominently in a collection[1] of 2,467 TED talk transcripts[2]:

```
Transcripts that contain the words:
  faith 168
  hope 1022
  love 1175
  faith & hope 88
  hope & love 520
  faith & love 101
  faith, hope & love 56
  faith or hope or love 1712
```

I propose using NLP to process the languange content (transcripts) of the collection of TED talks to assess how the authors talk about each of faith, hope and love. This analysis would use unsupervised learning techniques, e.g. clustering, to hopefully differentiate and group clusters involving faith, hope and love the talks were discussing. For example, are the authors describing their love for something, encouraging the expression love, or talking about love in some other way? My hope is that this differentiation is correlated with what makes talks rate highly or affect popularity.

As a secondary goal, I'd then use supervised learning with the clustering output, talk ratings and views as a way to assess popularity and/or audience reaction to the groups.

The overall goal of the analysis is to see if viewers are resonating with or enjoying any particular type or collections of what I believe are three core human values. If so, the analysis can be used to adapt or assess recommendations.

## Data

The data I propose[1] to use is available from kaggle. The data comes in two files containing the indicated fields:

ted_main.csv: comments, **description**, duration, event, film_date, languages, name, num_speaker, publication_date, **ratings**, related_talks, speaker_occupation, tags, **title**, url, **views**

transcripts.csv: **transcript**, url