

3 CAPSTONE PROPOSALS FOR FOUNDATIONS IN DATA SCIENCE (SPRINGBOARD)

ARUN BHARADWAJ, 24 JULY COHORT

Goal of the capstone project is to present findings of a research project using the concepts learned from the course. The concepts include descriptive and inferential statistics, R programming, problem statement identification, data extraction, data cleaning, story-telling and final analysis.

Objectives for selecting the capstone project dataset are as follows:

- 1) Interesting topic that may be of use to potential employer
- 2) Rich and informative dataset
- 3) Allows me to use API (application programming interface) and pull data from web
- 4) Includes elements of data mining and text analytics

It is my goal to achieve at least 3 of the 4 above stated objectives.

CAPSTONE PROPOSAL 1 (YELP REVIEWS – COMPETITOR COMPARISON AND TEXT ANALYTICS)

It will be interesting to compare the structure and semantics of customer reviews for a fast-growing upstart company compared to established incumbents. For the capstone, a one year old smart real estate brokerage can be compared to other brokerages in Seattle area.

CAPSTONE PROPOSAL 2 (ZILLOW - ZESTIMATE ERROR FOR DIFFERENT TYPES OF HOMES)

It will be interesting to compare the magnitude of Zillow's Z estimate error for different types of homes. The homes can be divided based on their physical and geographical attributes. Zillow has already uploaded datasets of houses, their attributes and Z estimate errors on Kaggle.

CAPSTONE PROPOSAL 3 (FINANCE – PREDICTIVE MODELING USING TEXT ANALYTICS)

It will be interesting to understand the effect of text information on finance. Annual reports contain valuable information describing the company's past and expected future. Assuming that markets are rational, it can be expected that investors and analysts incorporate all publicly available information when determining stock prices. Similarly, text mining of federal reserve meeting minutes, which are the biggest determinant of interest rates, can be used to predict future prices of bank stocks.