Capstone Ideas

Sam Connelly June 6th, 2018

Final

Time Series stock prediction of consumer goods stocks; in specific retail/fashion/luxury goods. Further exploring the effect of the recent ‘musical chairs’ of creative directors and how it affects their houses stock price.

Round 2:

Enron email dataset, Social network analysis vs more modern social network communication like twitter.

- Comparing communication, email in the workplace vs the more modern forms of social communication
- How different are the social structures of a workplace vs social structures and hierarchy of those circles in the outside world
- How different are the messages and structure of the messages between a company, like that and social forms of communications which don't involve co-workers.

London police records vs london crime data

- Based on crime and location did someone call the police, and what can this tell us about the neighborhood dychotomies of london.
- Crime data and police intervention is a great place to find details on the city and communities is what attracts me to this project.
- Is there an extra likelihood that someone will get arrested given a certain crime.

Company acquisition data and the bidding process which helps entrepreneurs curtail their company to getting bought out faster.

Network analysis on the food and fast food industry

- Goal is to highlight the negative impact on marketing and fast food and that correlation

Ted talk NLP

Price of debt/borrowing cost, details on certain startups/small businesses and predicting their success with a heavy influence on the borrowing cost in the city which they are in

18,393 Pitchfork reviews, nlp, clustering and prediction. A whole bunch of helpful topics which show up in fashion and retail

Moma and Met Gala

- Network analysis of where the artwork originated vs who gifted it to the museum
  - NLP analysis of reputation of the artists vs reputation and businesses acumen of who donates the pieces
    - Where is the art originating vs whose hands is it ending in
    - How different are the artists vs the people who donate them
- Try to re-cluster trip-advisor reviews
  - Majority of reviews are excellent or very good, which is only helpful to a certain point

Round 1:

The apartment search process (Likely Stretch)

Problems:

- Too much repetitive data
- Communication process between roommates
- Reliance on spreadsheet application
- Not an effective way to organize comments/preferences/likes and dislikes on options of interest

Capstone solution focus

- Creating a machine learning model which is able to work off your requirements and comments on ones you've already considered
- Some sort of chatbot which pulls in recommendations, possibly slack, that allows for roommates to better share and communicate
- More appealing dashboard that details what your wishlist options have and don't have.

Predicting the guest on a late night TV show

https://github.com/fivethirtyeight/data/tree/master/daily-show-guests

Large fashion houses/conglom stock pricing

- Predicting the price and performance of a fashion group or retail brand based off social media and such
- Measuring the timely effect on seasonal collections vs the immediate impact of social media what makes a group stock price fluctuate more?
    - Is there a delayed effect from seasonal collections? Do they only see a bump in stock price when the collection is previewed? When it hits the runway? When the looks become ready to wear and they start selling in the stores.? Or is it during markdown season when people really want to buy

Large fashion houses/conglom stock pricing, Cont.

What data would you need to do this:

- Stock pricing for the list of companies below:
  - Kering, Moncler SpA, PVH, Michael Kors, LVMH, RL
- Would want to see sales data during different times of the year to visualize the traditional fashion cycle
- You would want not only the brands overall sales data but the department stores which carry those brands
- The dates of the big fashion cycle
- Possibly certain headlines from media sources allowing the model to pull from nytimes and others so it would know when an event would affect the price.

Fashion and apparel classification using CNN

http://cs231n.stanford.edu/reports/2015/pdfs/BLAO_KJAG_CS231N_FinalPaperFashionClassification.pdf

http://ceur-ws.org/Vol-2009/fmt-proceedings-2017-paper2.pdf

https://cdn.oreillystatic.com/en/assets/1/event/258/Deep%20learning%20in%20the%20fashion%20industry%20Presentation.pdf (involving case study on gilt)

https://conferences.oreilly.com/artificial-intelligence/ai-ny-2017/public/schedule/detail/59111 (involving case study on gilt)

CNN Classification Kaggle competition

- Data set:
  - https://github.com/zalandoresearch/fashion-mnist
- Competition
  - https://www.kaggle.com/c/ttic-31020-hw2-fashion-mnist-large
  - https://www.kaggle.com/c/fashion-mnist-itba-lab-ml/data

Old Kaggle competitions

https://www.kaggle.com/c/imaterialist-challenge-fashion-2018/data

https://www.kaggle.com/c/mercari-price-suggestion-challenge/data

https://www.kaggle.com/c/two-sigma-financial-modeling/data

Others

- One method being deployed by retailors to discover more about what customers might want is the use of cognitive computing - programs that simulate human thought process and mimi the functions of the brain.
    - If you can give a retailer a two-week jump on trend prediction, that parles into two extra weeks in selling time in stores!
- Trend and retail analysis - WGSN
- WGSN : Her company conducts catwalk analytics, with teams of experts tagging each outfit - noting its garment type, style colour fabric and other details - as it is presented on the runway. Analysing this data reveals whether skirts or trousers are dominant in a particular season and, if it is trousers for example, whether they are mostly wide leg, flared-leg or bootleg style

Others

- Zara case study and the new fashion of retail
  - https://www.amazon.com/New-Science-Retailing-Transforming-Performance/dp/1422110575
  - https://www.forbes.com/sites/walterloeb/2013/10/14/zaras-secret-to-success-the-new-science-of-retailing-a-must-read/#d4b5e99534f9
- Interesting Data Set digging into immigration and its impact on science, specifically the nobel prize winners
  - https://github.com/BuzzFeedNews/2017-01-immigration-and-science
- https://github.com/BuzzFeedNews/2015-02-texas-cpa-deficiencies
- RX and Pharma data
  - https://www.kaggle.com/nlm-nih/nlm-rxnorm

NYC party noise complaints and NYC Taxi Data

https://www.kaggle.com/somesnm/partynyc

Fashion Datasets

- http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html (CNN, set of images)
- https://github.com/zalandoresearch/fashion-mnist (CNN, set of images)
- https://data.world/datasets/fashion
- http://www.st.ewi.tudelft.nl/~bozzon/fashion10000dataset/ (CNN, set of images)
- https://github.com/xthan/polyvore-dataset (CNN, scape from SSENSE, formerly plyvore)
- https://fashionunited.com/global-fashion-industry-statistics
-

Datasets

- http://developer.nytimes.com/ (ny times API)