Course Project Logistics

Students,

I have posted a list of candidate projects and resources on Piazza. You can either select from them or come up with your own projects that interest you. (Will send a detailed note about what I'm looking for)

Course Project Proposal is due on Mar 30 (will send the instructions shortly)

Course Project Report is due on April 30 (will send the instructions later)

Course Project Presentations on Final Exam day/time. (will send the instructions later)

-- Jana

0. Data Mining for Social Good Projects

https://dssg.uchicago.edu/projects/

Datasets: https://data.cityofchicago.org/browse?category=Education

1. Kaggle Projects

https://www.kaggle.com/competitions

https://www.kaggle.com/datasets

Some selected projects:

1.1 Instacart Market Basket Analysis (What will I buy next? 3 Million Instacart Orders, Open Sourced)

https://www.kaggle.com/c/instacart-market-basket-analysis/data

1.2 Amazon Reviews for Sentiment Analysis (Let’s get sentimental. Few million rows of Amazon customer review text and star ratings)

https://www.kaggle.com/bittlingmayer/amazonreviews

1.3 Indian Premier League | Kaggle (Love Cricket? This is the dataset for you. This dataset has IPL data from all seasons and all matches. Can you predict the winner for the next season?)

https://www.kaggle.com/manasgarg/ipl

1.4 Walmart Recruiting - Store Sales Forecasting (Hmm… you think you can forecast? Data from 45 stores in the US, also bakes in the seasonality and key events so be prepared for ups and downs.)

https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data

1.5 Trending YouTube Video Statistics and Comments (How about good old Exploratory Data Analysis ( EDA) and insights generation? Can you identify the attributes that make a video popular? 200 trending videos from US and UK)

https://www.kaggle.com/datasnaek/youtube

1.6 Credit Card Fraud Detection (Let’s play fraudster. Data from a European credit card. Can you cope up with the low incidence rate of 0.17% ?)

https://www.kaggle.com/dalpozz/creditcardfraud

1.7 Climate Change: Earth Surface Temperature Data (Is Global Warming for real? Global temperature data from year 1750 onwards. What will be a good way to statistically segment this data?)

https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

1.8 Health care analytics. (Healthcare analytics is booming. Data from US Department of Health on individual and small businesses. What drives the plan rate? Who makes the most money?)

https://www.kaggle.com/hhs/health-insurance-marketplace

1.9 Used cars database (370,000 used cars data scraped from Germany Ebay.)

https://www.kaggle.com/orgesleka/used-cars-database/kernels

1.10 Human Resources Analytics (Why do employees leave? Note that this is simulated data and not very large in size)

https://www.kaggle.com/ludobenistant/hr-analytics

1.11 SMS Spam Classification (Classify SMS to spam or non-spam)

https://www.kaggle.com/uciml/sms-spam-collection-dataset

1.12 Twitter User Gender Classification (Predict user gender based on Twitter profile information)

https://www.kaggle.com/crowdflower/twitter-user-gender-classification

1.13 Fake News Challenge (Multi-class classification)

https://github.com/FakeNewsChallenge/fnc-1

1.14 Toxic Comment Classification (Sentiment Analysis)

https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data

2. Some more projects

2.1 NYT Taxi Cab data analysis

http://www.nyc.gov/html/tlc/html/about/trip\_record\_data.shtml

Exploratory data analysis

- Trips per day/month/hr

- Average tips (in dollars) by location

- Visualization of trips on map

- Trends and try to predict trends (seasonal/etc.)

- Find anomalies in trends which might indicate some events

2.2 Five Thirty Eight

Can You Beat FiveThirtyEight’s NFL Predictions? (https://projects.fivethirtyeight.com/nfl-predictions-game/)

How does it work: https://fivethirtyeight.com/features/how-our-2017-nfl-predictions-work/

More datasets are available for ideas: https://github.com/fivethirtyeight/data