Searching for Success

Using google searches to **predict** the stock market

Jahnavi Jasti, Khyati Parekh, Erin Orbits, Abhishek Varma, Angel Wang

Motivation

- Every fiscal quarter, every publicly traded company releases their quarterly earnings to the public
- The earnings announcements are frequently significant events for a company's share price
- Amateur investors lack the resources of institutional investors
- Predicting quarterly performance can help level the playing field
- Google Search volume for company keywords may indicate whether a company is enjoying a better quarter than usual and thus a good indicator on stock performance.

Use Case

- An amateur investor can know after selecting a company:
  - the expected date of next Earnings Report;
  - the predicted probability of Stock Price Change after Earnings Report is released; and
  - the visualization of the correlation between relevant Google keyword searches and the daily stock price
- No prediction of the stock price itself

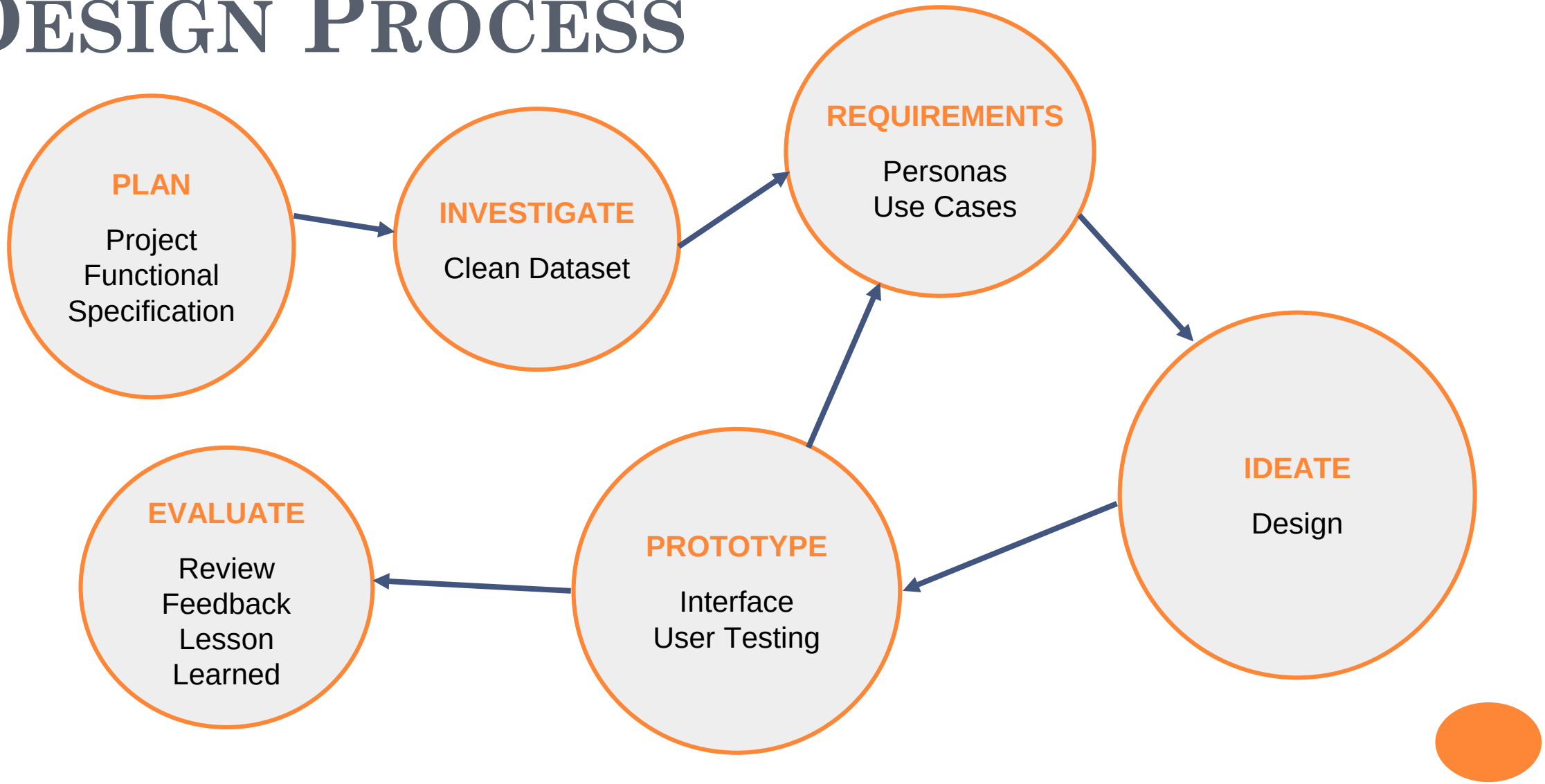
Data

- **Daily Google Search Index Data**
  - Unbiased sample of Google search data from Google Trends
  - Normalized data
  - Data is indexed at 100, with 100 as the maximum search interest within the time frame selected
  - 15 keywords search trends for Q-2 2007-2017
- **Daily Stock Price**
  - Daily adjusted close price within the Q-2 for 2007-2017 from Yahoo Finance

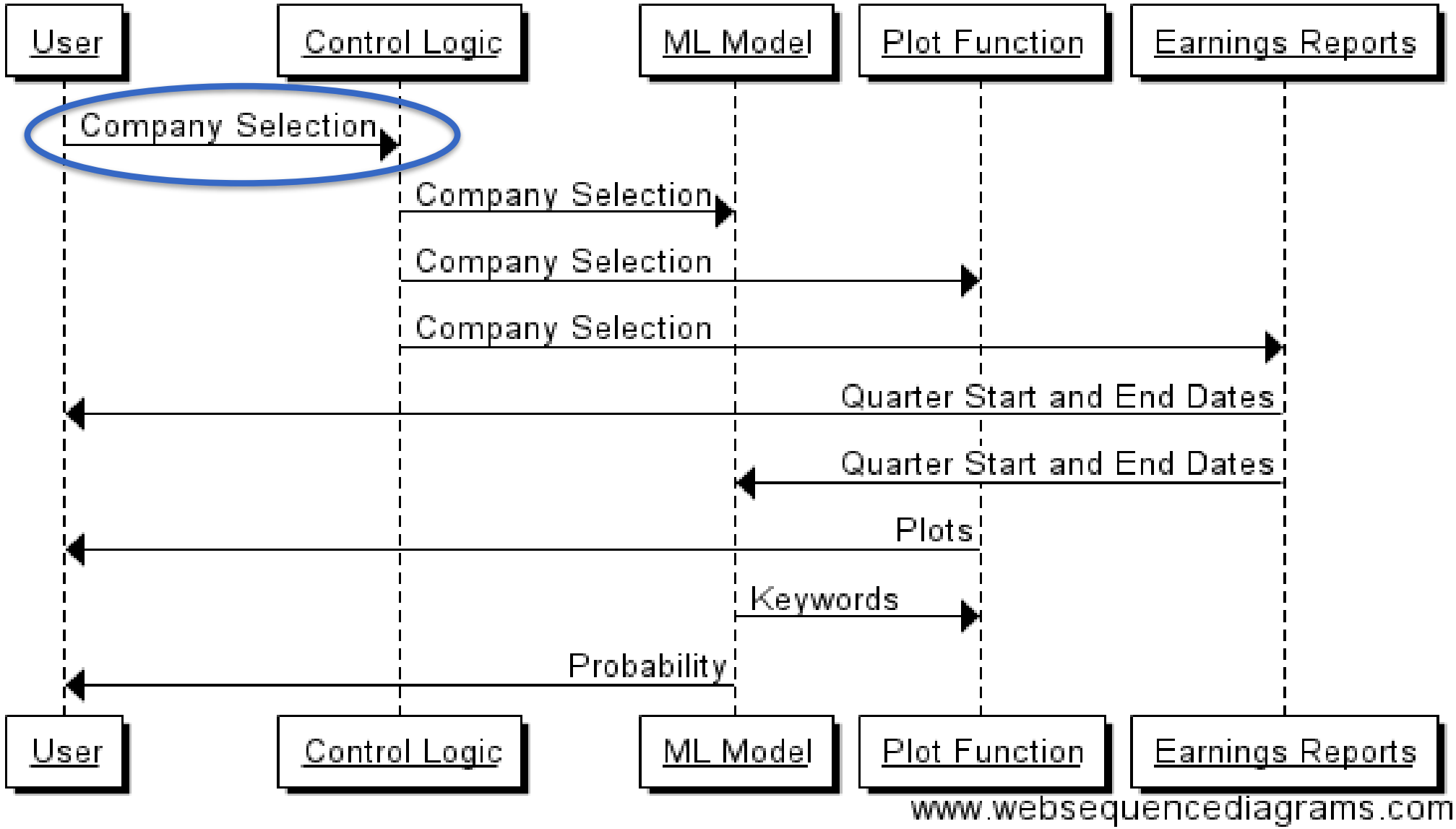
Packages used

- **Pytrends**: Download Google trend data automatically
- **Datetime**: Extract and format the dates and times for Time Series analysis

Design Process



Interaction Design



Control Logic – Earnings Report Module

get_quarter_begin()
get_quarter_end()

The start and end dates for each quarter define the range of Google trends data in the model

get_earnings_data()

1. Get the earnings release date for each company from Yahoo finance
2. This date is used in the ML model for prediction

Control Logic – Prediction Stock Price Module

feature_selection()
- Get trend data for all Google Trend keywords
- Get daily stock price data

1. Generate the keyword list with the 5 keywords with high impact on Stock Price
2. The keyword list is used by the ML model for prediction
3. The keyword list is used by the plot function

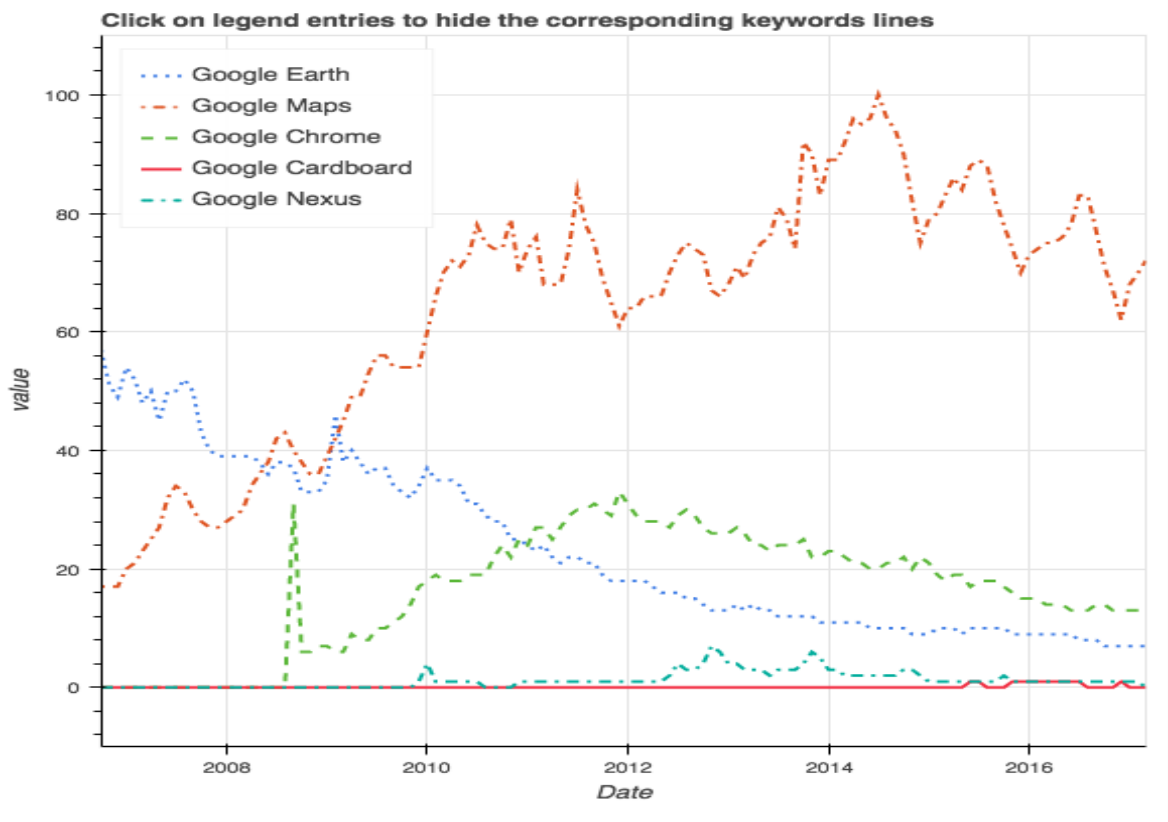
prediction_model:
- Keywords identified by feature_selection()
- Daily stock price data
- Earnings Report date from get_earnings_data()

Likelihood the Stock Price will increase after the quarterly Earnings Report is released

Control Logic – Historical Data Plot Module

historical_data_plot()

- Keywords identified by feature_selection()
- Daily stock price data



The Story

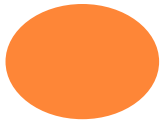
Josh graduated in 2013 from the University of Washington with Bachelor's degree in Computer Science. Since graduation, he's worked at Facebook as a software developer.

He saves around 1k every month and is interested in learning how and when to invest in the stock market.

He's interested in investing in technology companies. He heard about Searching For Success, and as a regular Google search user, he wants to give it a try.

DEMO

SearchingForSuccess/Submodule/Interface.ipynb



Repository Structure

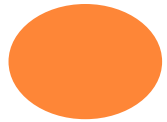
- Searching For Success/
  - |- README.md
  - |- LICENSE
  - |- SearchingForSuccess/
    - |- **init**.py
    - |- submodule/
      - |- **init**.py
      - |- EarningReport.py
      - |- PredictionStockPrice.py
      - |- HistoricalDataPlot.py
    - |- tests/
      - |- **init**.py
      - |- test_EarningReport.py
      - |- test_PredictionStockPrice.py
      - |- test_HistoricalDataPlot.py
    - |- setup.py
  - |- doc/
    - |- Design Specification and Project Plan
    - |- Functional Specification
    - |- Presentation
  - |- examples/
    - |- Example with Amazon
  - .gitignore

Lessons Learned

- Define Project scope as early as possible
- Working with new libraries can be difficult, *e.g.* manipulating dates with the datetime library and parsing JSON files

Problems Encountered

- Key libraries like pytrends didn't work on the Python 3 platform for everyone in the group
- Difficulty scraping web information, tried to use Beautiful Soup and other scraping tools, but wasn't able to get data beyond the Yahoo quote data
- Difficulty developing unit tests for the plotting functions



Thanks!