

Project Title: Data Digger

Team Name: Mercury Squad

Team Members: Sean Spearman (L), Aimee Galang (S), Ray Smets, Tim Shen, Cody Brown

Problem

HG Data currently has more data than they know how to portray. They still have a team that must compile reports for customers manually through searching through nearly 500,000,000 documents, making inferences and correlations a costly time consuming step in their product delivery. We hope to automate this process through a smart correlation engine resulting in eye catching visualizations to suit customers' desires. With this process automated, HG Data has the potential to take on more clients and provide them with not only with desired statistics but also with other relevant data packaged in a consumable format.

Why Important

Of course, correlation between two data sets does not necessarily mean causation. However, observing such correlations can still be very beneficial. With more information at hand, one can make better decisions. For example, if there were a correlation between the sales of a product and the months of the year, then one could plan ahead for such occurrences and market their product more effectively. By using such correlations, they could thus increase their profit. Although comparing the amount of correlation between two data sets can be done relatively simply, the problem becomes greatly more difficult if one tries to compare large sets of data. Our project would make this process much easier, and would thus save the user time spent on gathering so much data. Also, it would include more potentially relevant data than what one could collect on their own, or even show correlations between data that the user hadn't even thought of comparing.

Outcome

Our product will have the ability to process a wide variety of data, infer previously unseen

correlations and potential causality, and package that information in highly readable data visualizations made available to users through a website interface.

Milestones

- Solidify and set up tech stack
- Rudimentary data acquisition system
- Functionality to automatically find correlations in data
- Basic web interface
- Automatically generate the right type of graphics to best understand any given data

Possible Technologies:

Data Acquisition: Facebook Graph API, Untappd API, Weather Underground API, etc.

Data Storing: Amazon RDS for Sql

Accessing and Rendering Data: Express, Angular, Node

Correlating Data: Matlab

Data Output Visualization: D3.js