# Proposed Capstone Projects

## Customized Portfolio Analysis

Background

Client has investments in around 200 holdings, or positions, in an individualized and dynamic portfolio of stocks and ETFs of stock and bonds, primarily characterized by a diversified growth strategy, with a smattering of other positions characterized by other strategies, such as hedging and tax-advantageous, as well as positions allocated to income strategies that will be opportunistically grown over time. Client categorizes the positions by their own strategies, including each of the diversified growth positions in categories of secular (long-term – 5-10 year) growth themes, such as Big Data, or Internet of Things.

Stakeholder Requirements

Client requires better insight into their holdings and self-defined categories, such as exposure to each of their own categories. The investment community categorizes stocks in standardized fashion, by Sector (such as Technology, or Healthcare) and Industry (such as “Computers, Phones & Household Electronics”, or “Biotechnology & Medical Research”), but don’t drill down further. Furthermore, these Sector and Industry classifications do not represent my clients “angle” or the secular theme that the client needs to re-evaluate over time.

1. What is the problem you want to solve?
   1. Client requires daily data acquisition of their portfolio positions
   2. Pair the portfolio holdings with categorizations from another list
   3. Display percentage allocations according to these customized categories

will your client DO or DECIDE based on your analysis that they wouldn’t have

otherwise?

Client is an individual investor who will use daily updates to monitor performance of each of the positions to continually re-evaluate positions, risks, and buy/sell decisions.

1. What data are you going to use for this? How will you acquire this data?
   1. Daily downloads of client’s portfolio from their brokerage account - (private)
   2. Current updates of client’s custom categorization file - (private)
   3. Historical daily end-of-day price data from Quandl:

<https://www.quandl.com/data/EOD-End-of-Day-US-Stock-Prices>

1. In brief, outline your approach to solving this problem (knowing that this might change later).
   1. Daily data acquisition of their portfolio positions and historical price data
   2. Pair the portfolio holdings with categorizations from customer’s customization list
   3. Display percentage allocations according to these customized categories
   4. Display individual positions and their performance against time, relative to each other
2. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.
   1. Code
   2. Slide deck containing the dashboard displaying:
      * 1. Pie chart of percentage allocation, by the customized categories
        2. Pie chart of percentage allocation, by traditional metrics of Sector and Industry
        3. Time-based x-y plot of performance of each position from a baseline of 0% gain at time of purchase, color coded by category, hover-labeled with stock ticker symbol

## Online Retail Efficiency Study

1. What is the problem you want to solve?

Find outliers or other poorly performing sales inventory items (SKUs) by revenue on the low side for possible replacement with those on the high revenue side. This will be accomplished through the study of the retailer’s transactional database.

1. Who is your client and why do they care about this problem? In other words, what

will your client DO or DECIDE based on your analysis that they wouldn’t have

otherwise?

Client is an online retailer and wholesale distributor with a current limitation on warehouse capacity, looking to make decisions on dropping inventory items (SKUs) to make room for other higher performing inventory items.

1. What data are you going to use for this? How will you acquire this data?
   1. Online Retail Data Set from UCI ML Repository

<http://archive.ics.uci.edu/ml/datasets/online+retail>

1. In brief, outline your approach to solving this problem (knowing that this might change later).

Calculate revenue by stock code, by summing quantities sold over the period and multiplying the total units by the unit price(s) to determine and plot revenue per SKU.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.
   1. Code
   2. Slide deck containing the plot displaying:

Revenue by stock code, sorted and plotted by descending revenue

## Online Offer/Click Study

1. What is the problem you want to solve?

Determine absolute and relative effectiveness and discover other patterns in German online offer/response (click) data to determine effectiveness by category, merchant, user, time-of-day, or day-of-week.

1. Who is your client and why do they care about this problem? In other words, what

will your client DO or DECIDE based on your analysis that they wouldn’t have

otherwise?

Online retailers in Germany need to understand the absolute and relative effectiveness of their online advertising and promotional campaigns. They may make decisions on targeting advertising and promotions, and may make pricing decisions based on action taken (clicks) based on their campaigns

1. What data are you going to use for this? How will you acquire this data?

KASANDR dataset from UCI ML Repository

<http://archive.ics.uci.edu/ml/datasets/KASANDR>

1. In brief, outline your approach to solving this problem (knowing that this might change later).

Import the data.

Plot the implicit feedback of click/no-click binary data versus category, merchant, user, time-of-day, or day-of-week.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.
   1. Code
   2. Slide deck containing the plot displaying:

Click/no-click binary data versus category, merchant, user, time-of-day, or day-of-week.