**Capstone Project Ideas**

[In order of preference]

**Synopsis**

**Price Prediction**

Derive optimal pricing for products based on past transaction history and price sensitivity analysis.

Data availability: Sales transactions data set available

* 1. Profit Optimization – [WA\_Retail-SalesMarketing\_-ProfitCost.csv]

**Market Basket Analysis**

Analysis of transaction data to uncover patterns in purchases – which items are bought together frequently.

Data availability: Grocery data set available

1. Belgian Retailer
2. Groceries data set

**Targeted Marketing**

Customer analytics to determine high-value customers to be targeted through direct marketing campaigns.

Data availability: Charity/ Donor data set available

1. Kaggle competition – Raising money to fund an organizational mission
   * 1. [https://www.kaggle.com/c/Raising-Money-to-Fund-an-Organizational-Mission]
2. US Superstore data
   * 1. Tableau trial version sample dataset
3. Kaggle competition - Determine whether to send a direct mail piece to a customer
   * 1. https://www.kaggle.com/c/springleaf-marketing-response

Data Sources:

1. BestBuy – <http://developer.bestbuy.com/apis>
2. Kaggle.com competitions
3. KDD Cup competitions
4. Tableau trial version – sample data sets
5. IBM Watson - <https://community.watsonanalytics.com/guide-to-sample-datasets/>

**Detail**

1. **Price Prediction**

In this case, we use a data set containing sales data for various products for a sporting goods retailer. Data are available across a number of years (2004 – 2007), for a three-level merchandising hierarchy. Data include list and sale prices and cost information for individual products.

The potential questions that we plan to answer at this time are as follows:

* Which products are the most & least profitable? Across the years?
* How does profit respond to changing prices?
* What is the optimal price (band) for a (Illustrative) product?

Approach:

1. Pricing data will be normalized across products. Normalization will enable us to analyze the profit-price relationship at a higher level in the merchandising hierarchy (like product type instead of individual products.)

For this we have chosen the products “**Personal Accessories ⇨ Watches ⇨ Mountain Man (Extreme, Deluxe, Combination, Analog, and Digital)**” as illustrative.

The accompanying file [WA - analysis \_suppl.xlsx] details the normalization process (‘scaling’ tab).

1. For each product (extreme, deluxe, etc.) the price range has been extracted from the data. This price range has been normalized to a range 1 – 50 as shown in the ‘scaling’ tab of the excel sheet.

Assumptions:

1. The data analysis is performed for United States; it is assumed that, due to practically constant inflation rates, the prices across the years from 2004 to 2007 do not require any inflation adjustment.
2. This project is an effort to demonstrate an adequate understanding of the subject matter, and therefore does not claim that the predicted prices are of any immediate use in real life situations given the particular data set in question (2004 – 2007). However, the methodology used is very relevant for contemporary real-time effort in any price optimization endeavor.
3. Seasonality in the data is not a concern – since the products are for outdoor sports; it is assumed that seasonality has been taken care of in the data consolidated per year.

* Data source: <https://community.watsonanalytics.com/wp-content/uploads/2015/03/WA_Retail-SalesMarketing_-ProfitCost.csv>

1. **Market Basket Analysis**

Analysis of transaction data to uncover patterns in purchases – which items are bought together frequently.

Data availability: Grocery data set available

* 1. Belgian Retailer
  2. Groceries data set

1. **Targeted Marketing**

Customer analytics to determine high-value customers to be targeted through direct marketing campaigns

Data availability: Charity/ Donor data set available

* 1. Kaggle competition – Raising money to fund an organizational mission
     1. [https://www.kaggle.com/c/Raising-Money-to-Fund-an-Organizational-Mission]
  2. US Superstore data
     1. Tableau trial version sample dataset
  3. Kaggle competition - Determine whether to send a direct mail piece to a customer
     1. https://www.kaggle.com/c/springleaf-marketing-response