1. **Capstone Project Proposal**

**Project – Acquired Valued Shoppers Challenge**

**Problem-**

Consumer brands often offer discounts to attract new shoppers to buy their products. The most valuable customers are those who return after this initial incented purchase. So the task of this project is to predict the shoppers that are most likely to repeat their purchase.

In addition to the task of predicting the loyal shoppers, I would also like to answer the following questions using the data –

1. What factors determine a loyal shopper? Factors such as – Number of offers, price of commodity, Brand, Category etc.
2. What factors make a shopper disinterested and leave?
3. Determine shoppers that are likely to purchase out of necessity (e.g.: if there is only store available in their locality)
4. What brands and companies of the items are most likely to be purchased when given an offer?
5. What would be the overall return on investment based on the offers put and the consumer purchases?
6. What offers would interest whom?
7. Are the offers really required for a particular set of loyal customers? Like people who purchase based on necessity as mentioned above (point 3) or if they are just loyal to the brand and don’t really care about any offer?
8. How soon does a customer buy an item after he receives an offer?

**Client -**

The client here is a retail corporation that operates a chain of stores in multiple locations. Solving this challenge will help the retail corporation to identify the shoppers that become loyal to a product. It can also target the coupons to these loyal shoppers who are most likely to make a repeat purchase. Also, it would help them to effectively target specific offers to specific consumers.

**Data –**

The data I am going to use is from Kaggle. It contains the following files

* **transactions.csv** - contains transaction history for all customers for a period of at least 1 year prior to their offered incentive
* **trainHistory.csv** - contains the incentive offered to each customer and information about the behavioral response to the offer
* **testHistory.csv** - contains the incentive offered to each customer but does not include their response (you are predicting the repeater column for each id in this file)
* **offers.csv** - contains information about the offers

I will be continuously looking at combining this data with more relevant data that I come across.

**Brief Approach (doesn’t include implementation details) –**

I am planning to add more interesting data to this to make it more challenging in terms of cleaning it. I would then use the training data to build a machine learning model capable of predicting a loyal customer on the test data. The model would be further enhanced and capable of answering all the questions mentioned above in the problem.

**Final Deliverables -**

My final deliverables would include a code shared via my GitHub repository, a PowerPoint presentation and final report.

1. **Data wrangling work for Shark tank project**

The original data which I found related to Shark tank products had the following data –

1. Season , episode
2. Company name
3. Deal – yes/No
4. Industry
5. Gender of Entreprenuer
6. Ask Equity
7. Ask Valuation
8. Deal Equity
9. Deal Valuation
10. Royalty deal - yes/no and how much?
11. Investor
12. $ per share among investors

I was able to add more parameters to this data set –

1. Get the website of the company/product by using BING search API and some basic python code
2. Get the Facebook page of the company/product using BING search API and using that to call Facebook API to get the popularity of the product/service such as – likes, reviews, rating, business\_category, date\_founded
3. I came across a site called [shartankblog.com](http://shartankblog.com/) that has information about each and every shark tank product/service. I was able to get the web page by using BING API search - product+shartankblog. This website has really useful info about the products which I can utilize - like, I was thinking I can get the price of each item, number of units already sold, profit margin etc. I was also thinking, if it would be a good idea to some basic sentiment analysis on the webpage
4. I was able to use ALEXA web api to get the popularity of a website by getting information such as - #Of inward links, rank, rank by country, # page views, categories etc

Services used to add more parameters –

1. BING Search API
2. Facebook API
3. ALEXA Web API