Capstone 2 Project Proposal

By: Sanana Alizada

1. **What is the business problem?**

   Customer Lifetime Value Prediction for E-commerce platforms.

It is essential to be able to predict Customer Lifetime Value, so that customers could be grouped in the different levels of CLV(high, medium, low). Therefore, it will create an opportunity for businesses to set a certain group of customers as a main target and build strategies around their behaviors to increase customer retention as well as profit margin.

Constraints within solution space:

Potential economic recession which can cause customers with high or medium CLTV to lose their buying power.

2. **Who are the intended stakeholders, and why is this problem relevant to them?**
   Stakeholders are mainly the E-commerce store or online store of any given physical store owners/Sales & Marketing team/Website design and development team. The problem is relevant to them, because of the goals such as, increasing sales, keeping most profitable customers, customer retention, and establishing an effective platform.

3. **How is the anticipation that the intended clients will use the results to address the original business problem?**
   Store owners or Sales & Marketing Team can use the results to build new Marketing Strategies, promotional campaigns or loyalty programs after knowing the interactions of the existing and potential most profitable customers. Operations Team can improve the efficiency of customer service. The Technical Team can use the results to focus on building a more user-friendly online platform.

4. **Where are the datasets available from?**
   IBM Watson Marketing Customer Value Data

   Relevant customer data will be analyzed. Most profitable customers and their interactions will be analyzed by use of predictive analytics.

   Dataset consists of 24 columns:

   Customer ID
   State
   Customer Lifetime Value
   Response
   Coverage
   Education
   Effective To Date
   EmploymentStatus
   Gender
   Income
   Location Code
   Marital Status
   Monthly Premium Auto
   Months Since Last Claim
   Months Since Policy Inception

Number of Open Complaints
Number of Policies
Policy Type
Policy
Renew Offer Type
Sales Channel
Total Claim Amount
Vehicle Class
Vehicle Size

Data types:
String- 14
Integer- 6
Decimal- 6
Other- 2

5. **How will the problem be addressed?**

Data Science method will be applied, which entails the following phases: wrangling, exploratory analysis, and modeling.

I anticipate that this business problem can be modeled in two different ways:

1. As a classification problem, where the target would be classes of customers according to categories defined by the CLTV.

2. As a regression problem, where the target is the actual CLTV.