**Introduction**

Let say we have been consulted by a major footwear manufacturer, say, adidas or Nike, to provide advanced analytics to help advance their business. As with most businesses, the goal is to increase sales (profits) and decrease costs, and there are of course several ways to achieve the goal. After certain sessions with our client, we understand that as a manufacturer, they do not know a lot about their consumers, but they might be missing important opportunities for not knowing enough about their customers. Therefore, they are hoping, through data analytics, to learn more about their customer profiles, to understand where they stand with other competitors or potential competitors, to tap on opportunities, to be able to predict future trends. Successful business analytics start with questions that are relevant and could have impact. Considering the following as data that is currently available from the client company.

* Any transactions from our client’s ecommerce website.
* Through partnership, we also have data from one specialty retail chain (e.g., Dicks’ sporting goods).
* Through subscription, we also have access to aggregate weekly sales (in dollars and in units) at product subcategory - brand level (e.g. running shoes - Nike) at different types of retailers, including major department stores, mid-size department stores, shoe chains and specialty stores. You can also break the data into different geographical regions (northeast, west, central, south).

Of course, there is a lot of data out there that has not been collected by our client company but is available to grab, for example, social media data, consumer reviews data, as well as data from public sources (e.g. census bureau) etc.

Our client is seeking help from you to help them advance/protect their business and to help them design an appropriate data infrastructure that will help address important questions that have impact. Given this background, we must:

1. Identifying the problem and relevant data:
2. List one specific question (or a few related questions) we would like to focus on.
3. Discuss how addressing this question may benefit and/or provide useful insights to our client.
4. List the high-level data source (e.g. consumer’s data, review data, etc.) we would need to answer the question we identify. We must be realistic on what we could get and what we cannot have, and how we will get it.
5. Provide the database schema(s) (based on either relational schema or star schema) that will help us address the questions we identified above. I should have specific data in terms of attributes/columns we need to track and collect and how they help address the question(s) we identify?
6. Based on schema(s), provide sample queries, and discuss how we will use the query results to address our question(s).