**Lessons Learned Report**

I enjoyed working on this course. I believe it served as a great foundational introduction the work, questions and concerns that data analysts and data scientists deal with. The instructions and questions asked were mostly clear and the scope of the project was both believable and informative as to just how important data and data analytics can be.

While there is still much that I have to learn, especially as it relates to feature engineering, feature selection, model creation and eventual model deployment --- all of which I will refer to as ML --- I think I have cemented my understanding of the necessity of preprocessing (data cleaning/data wrangling) and having some way to “view” the data in a preliminary or holistic sort of way (EDA). What I most enjoyed about this project was the EDA and having to refine and clarify my EDA for the sake of the presentation.

This experience has made more mindful about how I will go about taking the “initial” first look at the data. That is, initially my EDA was very formulaic and simple. I simply want to see distributions of the different variables, both in absolute terms (value counts) and relative terms (percentages). This was the bulk of my EDA. However, throughout this process, I learned about how to aggregate the data using the groupby() function. I experimented with creating sub-data frames (“zoomed in look at the overall data”). I learned about how to easily change the data-types of variables with the astype() function and how to check for null and duplicate values. I also learned how to create stacked bar charts and a more reliable, albeit not as efficient, way of creating graphs.

Throughout this process, I’ve been exposed to new vocabulary and their importance in relation to the “full process” of data analytics (data science pipeline or data science cycle), such as imputation, discretization, normalization, encoding and others. I am excited, and hopeful that I will be vigilant, in understanding when and how to utilize these terms. A fundamental take for me is that, before any visualization or calculation or ML process is applied, the data must be properly cleaned, formatted and designed. Any missing values, duplicate values, variables with improper data-types, highly skewed data, presence of outliers --- among many other issues that I have not yet encountered, considered or been exposed to --- this all must be dealt with before anything else can take place.

Something I would have loved to deal with is simulated talks, or listening to real conversations or lectures, of the various stakeholders that I would be working alongside of in performing my analytics work. In the case of the consumer electronics industry, I would have loved to learn about privacy, right-to-repair, customizability (such as “jail-breaking”), planned obsolescence, electricity usage, REM (rare earth minerals) and other important terms and processes related to this industry.

The reason why I bring up these more social and economic concerns is that all of these concerns are vital for me to properly understand the “context” of the data and what the data actually is. More than that, this context absolutely is necessary for every stakeholder to be aware of. This knowledge and insight would help in making recommendations, deciding what additional data needs to be collected, how data can be manipulated, and how aggregations and other applications can be done on the data so as to generate insights. Of course, I recognize that the subject of this paragraph is outside of the scope of this program and that, especially as an entry level data analyst, the day-to-day of my work will not remotely approach or capture the depth --- for the field of customer electronics or ANY field --- which I am crudely trying to refer to here.

That being said, it is exactly these concerns, and the questions that can be asked from them, that excite and motivate me. To summarize, I am more excited with coming up with questions and more opportunities for data collection, than I am in providing absolute or “good to hear” answers. This is because, I believe, solutions to questions can only come from research and understanding. Solutions and answers are inevitable and have only to be optimized and carried out --- which really are just applied questions in their own right.

To return to this particular course, I would like to be introduced to and explore more the limitations and implications of ML. I will need to spend more time learning about classification problems and main algorithms (DC, RF, GB) used for them. This also includes ML-related applications, such as recommenders, predictive messaging (“automated customer reply bots”) and so on. I am not still yet confident, convinced or particularly certain as to how I can relate myself with ML and its possible scope.

I want to work on collective challenges and areas of lacking. I want to optimize existing and emerging systems and processes that are responsible for products and services related to human well-being. Given how broad and severe this is, I would like to start within the fields of education and human services. I would love to use my developing data analysis skills to help companies and government agencies in providing greater material value to their customers and constituents.

My key takeaways, thus far, has been the following:

1. Analysis is only as good as the data itself. Clean data is king. Bad analysis comes from bad data and bad communication (storytelling).
2. ML isn’t some magic thing. It can be properly used or abused or under-utilized. Understanding the problem scope and the data are the only ways to ensure that ML can be used effectively. Beyond that… tune models and collect more data during deployment and repeat this.
3. Data visualization is an important way to communicate, but it must be done sensibly (ethically) and within a “greater context” (story-telling).
4. The work of a data analyst/data scientist is collaborative and requires connecting with many points within an organization. I can’t ask every question, answer every question, collect every data point or develop every end-system. So, it is vital that I understand exactly what my role is and properly **apply the fundamentals**.

With these preliminary thoughts out of the way, to answer the question of future data analytics efforts for Blackwell:

1. Collect more data in the hopes that clear trends can be parsed out, visualized and utilized by ML-related applications.

2. Develop recommendation systems that can recommend products to customers given particular demographic data.

3. Develop Market Basket Analysis procedures for the sake of cross-selling.

4. Develop robust automated customer service infrastructure that allows customers to explore their concerns and then speak with a human representative should their concerns not be addressed.

5. Optimize the physical layout of physical stores, via product placement, signage, and other physical elements that may impact customer purchasing choices.

6. Develop infrastructure to collect and store data and centralized them in secure locations. Ideally, have these data collection schemes be standardized so as to help ease the work of preprocessing.

7. Use ML to predict and detects changes in customer churn and customer gain.

8. Develop a system that compares the prices, products, in-store locations between Blackwell and its key competitors – evaluating these discrepancies can help Blackwell “trim the fat” and become more “lean” --- offering better deals, reducing operational costs, opening stores in my geographically strategic locations

9. Use ML to ascertain purchasing patterns more robustly, such as purchase time (time of day, day, month, year), purchase reason (gift, hobby, work, education, etc.), spending limits, amount of time spent looking at products (physically or online), how customers travel in a store or what webpages they navigate through --- that is, simply just expand demographic data (within legal and ethical reason) so as to achieve recommendation 1.

10. Use Sentiment Analysis and other ML-applicated related to language, so as to help marketing and product departments determine the “correct” language with which to promote customer satisfaction, brand loyalty, and customer anticipation for Blackwell products.

11. Use ML so that Blackwell can better scrutinize its supply-chain process, such as how it interacts with manufacturers, distributors, worker productivity, and so on.