

Task 1

A. Describe *each* of the seven data analytics life cycle phases, adding a reflection of your expertise to *each* phase described.

1. Business understanding is the first phase in the data analytics life cycle. At this point, the analyst meets with stakeholders to define the significant goals and questions to answer, set a budget, design a timeline, and clear any misunderstandings (Western Governors University, 2020). This step is primarily used to form a foundation on the wanted outcomes for the analysis. I have no experience with data analytics, but this step reminds me of some moments in my previous internship. In the internship, I would meet with my director frequently to chat about tasks, goals, and work completed previously and for the future.
2. Data acquisition involves the active gathering of data from any source that has been defined during business understanding. SQL may be used during this phase if no previously made datasets exist. I have had the opportunity to practice this phase through personal projects by finding databases from Kaggle and Real Python and importing them into Jupyter Notebook.
3. Data cleaning is the process by which the collected data is cleansed. During this process, the data is checked for inconsistencies, duplicates, missing values, and extreme outliers (Western Governors University, 2020). My only experience with this phase is what I have done during a follow-along tutorial. During the tutorial, I learned how to clean the data using Python in Jupyter Notebook.
4. Data exploration is the first phase in which the analyst begins to understand the numbers and relationships beyond the appeared values ((Western Governors University, 2020). Basic correlations and patterns are identified. Python or R is used to complete this phase. In data analytics, I have little to no expertise in using Python to discover patterns and correlations. I have experience understanding numbers and their relationships during my undergraduate chemistry and biology labs.
5. Data modeling takes the patterns and correlations found during the primary statistical analyses to form graphs and other statistical models during this phase. Graphs used might include histograms, boxplots, and scatter plots ((Western Governors University, 2020). Different techniques like regression, classification, and clustering may be used. Python, matplotlib, and R are used to create these visuals. I only practiced this step with the follow-along tutorial and Python self-studying.
6. Data mining and machine learning go beyond describing the data and finding relationships and patterns using models. Other techniques that might be used during this phase are statistical techniques and machine learning algorithms. Algorithms might be descriptive, diagnostic, predictive, or prescriptive (Western Governors University, 2020). Python or R or used in this step. I have yet to gain expertise with this step beyond my self-studying to become more familiar with Python.

7. Reporting and visualization use the newfound relationships and values to create an interactive dashboard, PowerPoint, or other presentation so stakeholders can make sound decisions to solve the original business problems (Western Governors University, 2020). This phase uses programs like Tableau or PowerPI. I have yet to gain expertise with programs like these. I do have experience with PowerPoint and presenting to people for school.

1. Propose a way, with at least one example of *each*, to gain expertise in *each* of the seven phases.

1. To gain experience in business understanding, it is ideal for learners to actively work with data and discuss primary goals and questions with stakeholders. However, these skills will only be obtained indirectly for many students, including myself. Therefore, to better practice the skills needed, I can contact professors in each class for the MSDA to clarify the project goals and expectations. Other students can talk with managers in their fields to set clear expectations for the projects and responsibilities they are expected to complete.
2. Data acquisition can be practiced through personal explorations of daily activities. Of course, data can be acquired through accessible sources like Kaggle and online blogs. However, to be actively involved in developing the data, personal information like spending, chores, fitness, caloric intake, and car maintenance can be tracked and stored in programs like Excel or SQL.
3. To gain expertise in data cleansing, specific Python coding and other processes can be learned in class or during side projects. For example, after finding a dataset through tracking a personal habit, Python and SQL can see misspelled words, inconsistent titles and labels, extreme outliers, duplicates, and any other inconsistent measurement. Different practices can be implemented using OpenRefine. This source is designed purely for working with messy data and practicing cleansing.
4. For data exploration, it is necessary to practice the code and application of different visualizations. Kaggle datasets or the personally cleansed tracking dataset can be used in Jupyter Notebook, terminal, or VSCode to determine relationships. Pandas, matplotlib, and seaborn can be used to create visualizations. Beyond that, find projects that are looking for volunteers.
5. Data modeling requires practice, so learners know how to use modeling techniques properly. Kaggle and other personal datasets can be used in Python, sci-kit-learn, matplotlib, and Seaborn coding programs. Modeling techniques like regression, classification, and clustering should be used during practice. Learners must ask questions at this stage to know exactly how and when to use each specific modeling technique.
6. For better expertise in data mining, it is essential to familiarize yourself with more advanced techniques and machine learning. Learners can participate in individual or team Kaggle competitions and should find passion projects to help join in.
7. Reporting and visualization are arguably the most accessible phases to familiarize yourself with. To do so, learners can take their gathered datasets and practice with them in Tableau and PowerPI. Additionally, there are resources like Tableau Public and websites like workout-wednesday.com.

2. Explain how the organization's goal and mission help the analyst identify the business requirement.

The organization's goal specifically targets the ideal outcome for the data analysis. This is where business problems are questioned and identified. The analyst should be able to demonstrate to the organization and stakeholders how the issue will be answered and solved. The organization's mission identifies the long-term goals of the organization. The mission might help stakeholders, and the data analyst determine the type of business problems that may be encountered. Further, the mission tells the data analyst the business requirements they should focus on during the analysis.

B. Apply your knowledge of the data analytics life cycle by selecting one data analytics tool or technique and describing how the tool or method might be used in one phase of the data analytics life cycle in an organization about which you have some knowledge.

Fuse Lenses is an online retail sunglasses and eyeglasses company based in Florida. The organization is slowly gaining popularity and traction. To better understand their consumer demographic, the most popular types of sunglasses, and consumer buying habits, they might benefit from using machine learning techniques during the data mining phase of the lifecycle. Fuse will benefit most from using a prescriptive algorithm to determine inventory purchases, advertising, marketing, and production (Johnson, 2022).

1. Include three risks of using the selected tool or technique for data analytics.

- By using a machine learning algorithm, there may be a risk of over-reliance on automation.
- The algorithm may compromise data quality if accurate data points lead to false predictions.
- To gather data for the prescriptive algorithm, it is necessary to store information about the consumers' habits, purchasing history, and other demographics. However, privacy concerns might arise if the consumer is unaware of this data collection.

2. Describe an organizational or technical problem using the selected tool or technique.

When the business grows or becomes smaller, the prescriptive algorithm must be updated as it is scaled up. Constant updates must occur if the algorithm was not initially designed to be flexible and scalable.

C. Describe the decision-making process of selecting the appropriate data analytics tool or technique from part B.

Before any tools are implemented in the business, defining the problem or key goals that must be solved and met is essential. Once the problem has been determined, details like cost, data requirements, and timelines should be determined. After that, available tools and techniques should be taken into consideration. Machine learning techniques are simply not an option for some businesses due to cost constraints. Once it has been determined that the business can and

will use machine learning as a data analytics tool to solve the problems, the type of algorithm should be chosen. All algorithms must be formed for some projects to ensure that data is truly accurate. For this hypothetical, we will assume that the business decided that prescriptive algorithms are the only tools used for the goals.

Once the technique has been developed, it is essential to ensure that the algorithm is flexible and scalable to be adaptable to business growth. Pilot testing will occur once everything has been checked, and stakeholders will be involved. The likelihood of machine learning implementation will be determined. Once the technique has been adopted, continuous monitoring should take place.

1. Justify the organizational or technical need for the selected tool or technique.

For businesses to determine that prescriptive algorithms are a desired tool for relevant problems or goals, there must be some sort of organizational or technical need. Such needs may occur when the business grows, and inventory levels, decision-making, scalability, flexibility, and larger demographics must improve.

For fuse lenses, if they are not already in place, prescriptive algorithms may improve the inventory order process, shipping process, flexibility, and other business needs that simply aren't there as they grow.

2. Summarize the results of using the selected tool or technique in the life cycle phase you selected in B.

Prescriptive algorithms will tell the business how they should act to achieve desired results and outcomes based on previous data points. The results may tell stakeholders how much inventory should be ordered for which products, who to market the inventory to, how many employees to hire, and make better decisions for the business. Stakeholders also have the option to become more involved and use the algorithm for scenario simulations.

3. Evaluate the three potential ethical problems of using the selected data analytics tool or technique identified in part B1 for this problem.

Unintentional misuse of data may occur during the algorithm's learning period; it may not do what it should be doing and may harm the consumers of the business. Data privacy may be compromised through either selling of the data or breached security measures. Bias through racial profiling may take place when determining the demographic that typically purchases the inventory at the business (Hillier, 2022).

D. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

Hillier, W. (2022, December 15). *Is big data dangerous?* . CareerFoundry.

<https://careerfoundry.com/en/blog/data-analytics/is-big-data-dangerous/>

Johnson, J. (2022, June 28). Business leaders and data analytics - Data analytics for business professionals [Video]. LinkedIn. <https://www.linkedin.com/learning/data-analytics-for->

business-professionals-14936642/business-leaders-and-data-analytics?resume=false&u=2045532

Western Governors University. (2020). *Understanding the Data Analytics Life Cycle*.

<https://apps.cgp-oex.wgu.edu/wgulearning/course/course->

v1:WGUx+OEX0342+v01/block-

v1:WGUx+OEX0342+v01+type@sequential+block@b066e149473141e7b4a042d693cbdc30/block-

v1:WGUx+OEX0342+v01+type@vertical+block@48d880c04a524a3aacecf41a853851ff