Taylor Flagg
D596 – The Data Analytics Journey
April 6, 2024

Task 1

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

1. Business understanding – Also known as discovery, business understanding is the first phase of the data analytics life cycle and is centered on developing a strong business sense. During this stage, the analyst collaborates with stakeholders to clarify key objectives and questions, establish a budget, outline a timeline, and resolve potential misunderstandings (Western Governors University, 2025). From my experience as a BI analyst, I believe this step is key to providing an actionable set of objectives in the future. Establishing a solid understanding of expectations ensures all parties are on the same page should obstacles arise.
2. Data acquisition – During the data acquisition phase, analysts will start collecting data relevant to the stakeholder's request. Common data sources include using SQL to pull the data from a database, web scraping, or other sources like flat files (Western Governors University, 2025). This step tends to be the lion's share of my effort as a BI Analyst. Often, I find myself presented with a stakeholder looking to answer some specific question but unsure where the data lives or the specific logic needed to generate an actionable dataset.
3. Data cleaning – The data cleaning phase utilizes tools such as (but not limited to) SQL, Python, or Excel to massage and transform the data to ensure quality, reliability, and relevance (Western Governors University, 2025). I've suffered multiple times from situations that could've been prevented with thorough attention upstream, ultimately rendering any data analysis created irrelevant.
4. Data exploration - During the data exploration phase, the analyst explores the data to understand its structure and relationships. This step often involves using data visualization tools and looking at the features of the data, including measures of central tendency, variability, and distributions (Western Governors University, 2025). I've learned through experience that this step helps provide a well-rounded view of the data you're working with. By identifying center, variability, and distribution measures, I can understand what values I should expect and possible missed outliers.
5. Predictive modeling – The predictive modeling phase is where the analyst steps from descriptive or diagnostic analysis into predictive analysis to ask what will happen. This is done using the present data to build models via tools like Python and R to leverage old data to forecast future results (Western Governors University, 2025). While my professional experience is limited mainly to descriptive analysis,

Taylor Flagg
D596 – The Data Analytics Journey
April 6, 2024

I've done a handful of reports that utilize current data and previous trends to forecast inventory or utilization metrics in the future.

6. Data mining – The data mining phase uses tools such as Python and R to identify patterns in large datasets through statistical techniques and machine learning algorithms (Western Governors University, 2025). While I have little experience with data mining in my professional career, I've pursued online courses where I've used Python and leveraged machine learning algorithms to mine large datasets to build models for linear regressions and other predictive analytics.
7. Reporting and visualization - In this phase, the analyst communicates the story behind the data, using visualizations and interactive dashboards to share key findings. Tools like Tableau or PowerBI empower non-technical users to explore the data and identify trends or patterns. The main goal is to deliver actionable insights tailored to the needs of different stakeholders (Western Governors University, 2025). This step is one of my personal favorites. Tools like Power BI allow me to turn the raw data into visualizations where I can tell a story to end-users and non-technical audiences. I've also found it incredibly rewarding to see others interact with the visuals and discover insights on their own.

A1. Propose a way, with at least one example of each, that you might gain expertise in each of the seven phases.

1. Like most soft skills, a key way to develop expertise in the business understanding phase would be through doing. For example, sitting in on requirement-gathering sessions and actively listening to stakeholders can help better understand the organization and its needs.
2. Expertise in data acquisition can be developed by refining your SQL skills and understanding the data source you are working with. Doing so forms an understanding of what data lives where in the EDW and how different tables connect.
3. Much like data acquisition, data cleaning can be developed by taking dirty data sets and massaging them to a point where you can identify outliers, remove duplicates, and account for errors.
4. Data exploration can be reinforced by iterating on clean data sets from the previous phase and working with tools like Python to identify and visualize the dataset's descriptive statistics.
5. Building off the last step, an analyst can work on predictive modeling skills by using functions via the Pandas Python library to build linear regression models to identify future predictions or correlations between separate variables.

Taylor Flagg
D596 – The Data Analytics Journey
April 6, 2024

6. The data mining phase would be the most difficult to develop skills in. I would look online to find a large dataset and utilize a Python library like TensorFlow to create a model capable of iterating over the large dataset and performing functions such as identifying patterns, classifying data points, and identifying data clusters.
7. Finally, when it comes to the reporting and visualization stage, I've found it best to iterate through the previous steps using data practical to your life and build dashboards and visuals for things helpful to you in tools like PowerBI. For instance, by creating a report visualizing monthly spending habits and income and debt predictions, my wife and I were able to identify valuable insights into our spending habits and adjust accordingly.

A2. Explain how the goal and mission of the organization help the analyst to identify the business requirement.

- An organization's mission outlines its core purpose and values, helping the analyst understand its aims. This ensures that business requirements align with the organization's long-term purpose. Goals, on the other handhelp the analyst pinpoint immediate needs and prioritize requirements that directly support measurable objectives. Together, the mission and goals guide the analyst in identifying business requirements that are both strategically relevant and practically actionable.

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.

- Previously, I worked on a project where the stakeholders requested the ability to view trends in the daily utilization rates of equipment at several of their facilities. During the data acquisition phase of the project, we encountered an obstacle – the data source we were provided only showed current utilization statistics. Considering the requirements they set forth and the data I had on hand, I utilized Python to write a script to extract the utilization data on a daily cadence and stored it in an SQL table my team was managing.

B1. Include three risks of using the selected tool or technique for data analytics.

1. If the infrastructure for the data capture failed, we would miss out on data points for days at a time.
2. If the data capture process were poorly implemented or captured data incorrectly, it would impact further analytics downstream.

Taylor Flagg
D596 – The Data Analytics Journey
April 6, 2024

3. Any change to the type, scope, or granularity of the data I was extracting wouldn't be able to change or capture previous data retroactively.

B2. Describe an organizational or technical problem using the selected tool or technique.

- As the scope of data capture expanded—such as by incorporating more facilities or additional data types—the resulting data bloat increased the strain on storage and processing resources. This posed organizational challenges, including an overall slower data pipeline, higher infrastructure costs in terms of cloud resources, and delays in decision-making due to longer processing times. While Python is highly effective for the data acquisition phase due to its flexible nature and capability to handle diverse data formats, its performance can degrade with large datasets unless integrated with more scalable solutions like cloud-based storage. Overall, this highlighted the need for the organization to reassess both tooling strategies and infrastructure planning to ensure scalability.

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

- After we received the requirements from the stakeholders and realized that the data we had available wouldn't be able to fulfill it directly, we took a step back to review which tools we'd have available to capture the data. Ultimately, we looked at Python and SSIS and chose Python partly due to the ease of use but predominantly because of the freedom to transform the data quickly as the project evolved.

C1. Justify the organizational or technical need for the selected tool or technique.

- Given the initial request and the type of data we had access to, some secondary acquisition was required. Without using Python or some other tool to facilitate the data acquisition phase, we would've been unable to track historical changes in the utilization data, rendering the stakeholder's requirements impossible.

C2. Summarize the results of using the selected tool or technique in the life cycle phase you selected in B.

- In general, the effort to utilize Python during the data acquisition phase proved to be a big win by allowing the project to get off the ground. Once we could finalize the logic and solidify the data pipeline, we moved into other stages of the analytics life cycle. This led to more informed business decisions and an overall increase in the initiative's adoption across different facilities.

C3. Evaluate the 3 potential ethical problems of using the selected data analytics tool or technique identified in part B1 for this particular problem.

Taylor Flagg
D596 – The Data Analytics Journey
April 6, 2024

1. Inaccurately gathered data via the Python script could introduce bias and lead to inaccurate business decisions. Due to the nature of the equipment (healthcare), poor business decisions could significantly impact facilities and their patients.
2. By effectively obfuscating the logic for the data capture process using a Python script, it could introduce a lack of transparency when dealing with facility-level stakeholders by removing the element of equity in what they were able to see versus what we saw on our end.
3. Extracting the data from the corporately managed EDW to the one my team manages could introduce a security risk by having an additional avenue of attack. Additionally, since the team-managed server is independent of the corporate EDW, there is not the same level of support should issues arise.

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

Western Governors University. 2025. "Understanding the Data Analytics Life Cycle." WGU.edu. Western Governors University. 2025. https://apps.cgp-oex.wgu.edu/learning/course/course-v1:WGUx+OEX0342+v01/block-v1:WGUx+OEX0342+v01+type@sequential+block@b066e149473141e7b4a042d693cbdc30/block-v1:WGUx+OEX0342+v01+type@vertical+block@48d880c04a524a3aacecf41a853851ff.