Zacary Harmon Performance Assessment 1 D596

Questions are in **bold**. My responses are styled regularly.

1. **Describe*each* of the seven phases of the data analytics life cycle.**
   1. Business understanding
      1. This involves direct collaboration with the business to understand the business’s objectives and to ensure that our work coincides with these objectives.
   2. Data acquisition
      1. This step involves the collection of data, whether that be from an API, a database, or web scraping.
   3. Data cleaning
      1. This step involves editing the acquired data to fix inconsistencies, null values, and errors, etc. This is to ensure the data is reliable enough to undergo the level of detail involved in professional use.
   4. Data exploration
      1. This step is where the analyst works with the cleaned data set to find patterns and insights that align with the objectives of the data project. The analyst can do so by physically exploring the data set or creating visualizations, etc.
   5. Predictive modeling
      1. This step utilizes algorithms and machine learning to predict future trends based on historical data.
   6. Data mining
      1. Data mining is the process of exploring an extremely large dataset using algorithms and machine learning models to gather insights that would be too strenuous or time-consuming for a human to undergo.
   7. Reporting and visualization
      1. This is arguably the most important step as it is the culmination of all the other steps where our findings and insights are presented and visualized to the stakeholders of this data project.
2. **Propose a way that you might gain expertise in *each* of the seven phases. Provide an example of *each* phase.**
   1. Business understanding
      1. I can gain expertise by drafting some data projects with colleagues. When drafting these projects, we can collaborate and discuss the needs of each project and the data we’d like to focus on in each one. This will simulate real business meetings where a data project would be created and priorities would be managed.
   2. Data acquisition
      1. I can gain expertise by working on home projects that source data sets on my own. By writing code to pull data from APIs/ the web and load them into a data set, I am sourcing data and doing the same work I would be doing in a professional setting.
   3. Data cleaning
      1. I can gain expertise by working with Python to clean and organize the data I am acquiring in the homework projects I mentioned in the previous step. By directly doing this work, I am gaining expertise in skills needed in the daily role.
   4. Data exploration
      1. I like to gain expertise in this area by utilizing Jupytr notebook to explore the data sets I’ve collected and visualize the data in different formats. By exploring in Jupytr notebook in Python I can quickly and linearly see what each step is doing to the data as well as visualizing the dataset at the same time.
   5. Predictive modeling
      1. I can gain expertise in this by working on my own machine learning algorithms to explore past data sets and practice utilizing this tool to gather insights. I can specifically do project where I use a machine learning tool to gather weather data from previous years in my city and compare them to see how temperatures have changed.
   6. Data mining
      1. In the same vein, I can utilize the machine learning project just mentioned to undergo data mining and find insights and create visualizations using this weather data.
   7. Reporting and visualization
      1. To gain expertise in this I like to take every data project I do and complete it by fully visualizing and preparing to prepare every project. By forming conclusions and writing papers/powerpoints I can further my professional communication skills.
3. **Explain how the goal and mission of the organization help the analyst to identify the business requirement.**
   1. By understanding the essential business goals and missions of the organization, you can make informed decisions about your data project. If the company is trying to sell buttons, then the data analyst should have in mind to focus on insights that affect the company's sales of buttons. Likewise, if the company is focused on preserving natural spaces, the analyst might be focused on creating visualizations to show the percentage of land in the city that is natural spaces.

**B.  Select one data analytics tool or technique in one phase of the data analytics life cycle in an organization.**

**1.  Describe the decision-making process of selecting the appropriate data analytics tool or technique for the identified phase of the data analytics life cycle.**

My tool of choice would be Python. I would use this for data acquisition, data cleaning, or data exploration but I could also use it for modeling, visualization, etc. I would use python when I need to manipulate a dataset or explore a dataset for insights. My decision making process for using tools would be, if I need to acquire data, I would use python to extract that into a dataset. If I need to manipulate data, I would use SQL to query and alter the dataset or I would use Python if there was a lot of complicated manipulation. Lastly, if I wanted to visualize my data I would use tableau or power BI. While I could also use Python, the two previously mentioned tools are far more in depth for visualizations.

1. **Describe an organizational or technical problem that will be addressed using the selected tool or technique.**
   1. An organizational problem that can be fixed could be that we need to extract, transform, and load data from an API into our PostgreSQL database. Python speeds up many processes by being able to perform all of these tasks on it’s own, it can call the API and gather the required data, clean the data for duplicates and errors, and then load it into our SQL database all in one script.
2. **Justify the organizational or technical need for the selected tool or technique with evidence to support claims.**
   1. Python can be used for nearly every step of the data analytics process, and in addition, makes collaboration between analysts far easier due to the increased readability of python compared to other languages. (UCD Professional Academy) Aside from easier collaboration between analysts, there just simply isn’t another tool as flexible as Python. SQL is the peak of querying databases, but we can write SQL queries using python, we can visualize, extract, transform, and load data all using Python. In an article from New Horizons.com they explain how “Python boasts a rich ecosystem of libraries and tools specifically designed for data analysis… these libraries provide pre-built functions and modules that streamline various aspects of data analysis, from data manipulation and visualization to statistical analysis and machine learning.” (Karl)

**C.  Evaluate potential legal or ethical risks or impact on essential business functions by doing the following:**

**1.  Discuss three risks of using the selected tool or technique for data analytics from part B.**

1. Lack of experience. Using Python could become a crutch for data analysts as it’s flexibility could inhibit analyst from gaining experience in other tools due to the wide scope of python. This can greatly impact business functions by making mediocre analysts. An analyst who solely uses python will struggle to keep pace with the analyst who is flexible and knows which tool to use for which task. While python can do many things well, there are tools who excel in one or two aspects of the data analytics lifecycle. For example, while python can successfully do data visualization, Tableau or PowerBI will allow for more interactivity and depth to be inserted into your visualization than python will allow.

2. Lack of complex thought. Due to pythons many libraries and wide support, data analysts could lose professional grasp of advanced coding concepts and algorithms that are bypassed using python libraries. While these libraries save time in the short run they do lower the knowledge requirement to utilize these tools. This again impacts a daily business function by lowering the skills of data analysts. For example, if there is a need to run a particularly complex query in SQL but the analyst is only used to using a python library for that task, the analyst will be unable to quickly and efficiently complete this query due to a loss of skills thanks to their reliance on the ease of python libraries.

3. Lastly, Python can struggle with large datasets. Due to being an interpreted language, python can be slower than compiled languages like C++. This can become troublesome for large datasets that would be no issue for a language like SQL. For any large corporation they are going to use very large databases to store the information for their employees, sales metrics, business data, etc. While python can excel for personal projects and small files, it is not optimized to run millions of rows in a database like SQL is. An analyst that uses python as a crutch can dramatically slow down the process of working with these essential large databases that companies rely on.

**2.  Based on the risks discussed in part C1, discuss three potential legal or ethical problems.**

1. With a lack of experience, you could run into an ethical issue of privacy concern if analysts do not have the experience needed to notice when something needs to be fixed or omitted for privacy concerns. Knowing which tool to use and what data should look like is important to maintain consistency and privacy. A lack of experience in using the tools can lead to an analyst not being familiar enough with a software that they do not notice a privacy error, or inconsistencies in their work, which can lead to legal and ethical repercussions.

2. With a lack of complex thought, mistakes can easily be made regarding privacy. Again, getting too comfortable using a tool can be a benefit, but it can also allow for sloppy results and inaccurate data sets. It’s vital to remain diligent and double-check your work for accuracy and consistency as a lack of detail in analytic work can lead to legal or ethical repercussions.

3. Struggling with large data sets can be an issue. If not optimized properly, you, once again, can have inaccurate and inconsistent results from your data. This can completely break a Python model, and if the model makes decisions using inaccurate data, that can lead to illegal and unethical situations. Using tools how they were meant to be used can avoid inaccurate and inconsistent results, which will improve your quality of work and help the analyst get better results.

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

**References**

“Why Do Data Analysts Use Python?” *UCD Professional Academy*, www.ucd.ie/professionalacademy/resources/why-do-data-analysts-use-python/. Accessed 11 Aug. 2025.

Karl, Taylor. “Benefits of Python for Data Analytics Explained.” *New Horizons*, New Horizons, 30 July 2024, www.newhorizons.com/resources/blog/benefits-of-python-for-data-analytics.

All other information came from the WGU lesson materials.

**E.  Demonstrate professional communication in the content and presentation of your submission.**

Thank you for taking the time to review and grade my submission.