

NGallo_Task 1 The Data Analytics Life Cycle_A

- 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. Propose a way, with at least **one** example of *each*, that you might gain expertise in *each* of the **seven** phases.
 2. Explain how the goal and mission of the organization help the analyst to identify the business requirement.

The data analytics lifecycle is a way to obtain insights from data. By using the seven phases of this life cycle, analysts are able to take a problem and break it down into how to solve the data-driven problem.

Business understanding is where stakeholders collaborate to define the project objective, scope of project, and what success metrics to include. An example of business understanding is attending a sprint planning meeting with the stakeholders on the project.

When planning and regularly attending sprint meetings, there is limited room for misaligned stakeholder expectations or lack of clarity from teammates.

As someone with 7 years of industry experience, I have experience collaborating with stakeholders to define project goals as well as what is expected of each teammate on the project.

Data acquisition is the process of collecting relevant data from databases, APIs, etc. An example of data acquisition is building a script using Python programming language to pull data from online platforms like X (formerly Twitter).

Collecting data is important to better track and contribute to documentation. This is a good phase to build a relationship with the IT team to ensure access and secure data sources. I have basic knowledge of Python and minimal experience using SQL in my previous role. I plan to learn Python and SQL more in-depth to gain data collection skills.

Data cleaning is where the collected raw data from the data acquisition phase is cleaned, transformed, and integrated into analysis-ready formats. An example of data cleaning is transforming a public dataset with thousands of rows and items into a readable format that is ready for further analysis.

The process of data cleaning will likely consume the majority of the analysts' time. It's important to incorporate automation rules to reduce manual errors in large datasets. I have experience transforming medium-sized datasets when working as an SEO strategist. I know how to download data from Google Search Console and manipulate the data with filters and categorization of keywords.

Data exploration is where analysts explore the data to identify patterns, trends, and relationships using statistical and visualization tools. An example of data exploration is utilizing player stats, such as passes and goals, to evaluate any patterns in their performance across matches.

Data exploration has the potential to create skewed distributions that can invalidate models. Therefore, it's important to use visualization tools like Tableau to evaluate the data before further analysis. In my past role, I consistently used visualization tools like LookerStudio (formerly DataStudio by Google) to identify any keyword trends in the e-commerce industry.

Data modeling is where predictive models like regression and neural networks are built. This phase happens when the development of mathematical or computational models is ready to be used to derive insights from the data. An example of data modeling can be shown in creating a shopping list. Data modeling can be used to take items, quantities, categories, and priorities and create a model for an efficient grocery shopping trip. Not only can this be used for the consumer as they shop, but it can also give insights to evaluate how consumers shop and predict buying patterns.

I have not built any predictive models in my previous role; however, I worked with data scientists and engineers on my team who were building language learning models for predicting how customers are searching for desired products.

Data mining is where the results of the analysis are interpreted and communicated to stakeholders. An example of data mining is explaining the results and any concepts, such as clustering, to others on the project. This is an important phase for practicing communication and for the analyst to present their findings in a digestible way for non-technical teammates. In my previous SEO strategist role, I've interpreted smaller datasets around keyword data and search engine data (clicks, impressions, click-thru rate) to find valuable information for our marketing and retention departments.

Reporting & visualization is the last phase of the data analytics life cycle. This is where the results are communicated to stakeholders via dashboards and reports. An example of reporting & visualization is using a Tableau dashboard to show the trends of movie ticket purchases for the newest blockbuster. A variety of charts can be created to show these data results. To explain a few:

- Histogram, a way to graph numerical data in groups to represent frequencies
- Scatterplot, a graph for visualizing correlations or relationships
- Heat map, a visual representation to show frequency or interaction using colors

I created a weekly dashboard for my SEO team as well as company-wide standups to discuss the latest trends in our keyword data, how our customers are searching on search engines, and valuable news articles in the SEO & search world other non-technical stakeholders should be aware of.

The seven phases of the data analytics life cycle are important to know in order to obtain insights from the data. By using the life cycle, data analysts can find problems, break them down, and offer solutions.

NGallo_Task 1 The Data Analytics Life Cycle_B

- B. Apply your knowledge of the data analytics life cycle by selecting one data analytics tool or technique and describing how the tool or technique might be used in one phase of the data analytics life cycle in an organization about which you have some knowledge.
1. Include **three** risks of using the selected tool or technique for data analytics.
 2. Describe an organizational or technical problem using the selected tool or technique.

SQL (Structured Query Language) is one of the most important data analytics tools. SQL is a programming language that allows a data analyst to interact and communicate with relational databases. This tool is commonly used in the data cleaning phase. In this phase, an analyst will use SQL to clean, filter, and structure the raw data to prepare it for the data exploration phase.

While there are always risks with data analysis and the tools that accompany it, here are the three risks of using SQL. The first risk is the complex SQL queries that come with managing massive databases slowing down systems. The second risk is human error opportunities with over-filtering or incorrectly combining data when joining tables. Lastly, the final risk is recognizing the limitations SQL has with unstructured data like social media content, customer reviews and other diverse data sources.

An organizational problem when using SQL for a retail or e-commerce company is slow query execution. This is likely to happen during peak buying seasons like holidays (Black Friday). The database is being overloaded with customer purchases and inventory restocking data. The execution of queries may take a bit longer to run through the millions of lines of data to pull the right information for the analyst.

NGallo_Task 1 The Data Analytics Life Cycle_C

- C. Describe the decision-making process of selecting the appropriate data analytics tool or technique from part B.
1. Justify the organizational or technical need for the selected tool or technique.
 2. Summarize the results of using the selected tool or technique in the life cycle phase you selected in B.
 3. Evaluate the 3 potential ethical problems of using the selected data analytics tool or technique identified in part B1 for this particular problem.

The justification for using SQL for a retail or e-commerce company comes with SQL being the most efficient way to access and transform data, especially in larger datasets. When using SQL for data cleaning, this tool will standardize the dataset after manipulating and filtering the data for customer purchases, inventory restocks, and other necessary items to track for a retail company.

There are risks and problems with managing company and consumer data, but here are the three potential ethical problems of using SQL for retail or e-commerce datasets. First, slow query execution could add pressure on the analyst to use incomplete datasets which then creates a biased dataset to work with. Second, over-filtering any purchases or consumer data can indirectly exclude marginalized customer groups from the dataset which then can exclude them from future company strategies like marketing assets. Lastly, the limitation on structured data could affect the company's relationship with customer feedback and user needs if shown via reviews or social media.