Task 1:

The Data Analytics Life Cycle

Keniyah Chestnut

The Data Analytics Journey — D596

ID: 012601305

**A. Phases of the Data Analytics Life Cycle**

**1. Business Understanding**

**Definition:**  
This phase focuses on defining business goals and identifying key problems that data analysis can help solve.

**My Expertise:**  
During my internship as a data analyst, I participated in meetings with stakeholders where I took notes on project requirements and business goals. I also helped turn business questions into measurable objectives. These meetings helped me learn how data is used to support business decisions.

**2. Data Acquisition**

**Definition:**  
Collecting data from sources like APIs, SQL databases, or flat files.

**My Expertise:**  
I’ve written SQL queries to extract relevant data from relational databases and imported it into Python for analysis. My familiarity with SQL has made it easier to transition to using Python libraries like pandas.read\_sql() easier.

**3. Data Cleaning**

**Definition:**  
Fixing or removing incorrect, corrupted, or incomplete data to prepare it for analysis.

**My Expertise:**  
I used Python and Pandas to clean datasets by removing duplicates, filling in missing values, and changing data types. One time, I cleaned a very messy user activity log. This made it easier for our team to analyze app engagement smoothly and effectively.

**4. Data Exploration**

**Definition:**  
Data Exploration is analyzing the data through visualization and statistical summaries to discover patterns and trends.

**My Expertise:**  
I’ve used Python libraries like Matplotlib and Pandas to create charts and run summary statistics. For example, I made a bar chart that showed product engagement across different user groups. That helped the team make smarter marketing decisions.

**5. Predictive Modeling**

**Definition:**  
Using statistical or machine learning models to forecast outcomes.

**My Expertise:**  
I built simple regression models in Python to predict sales trends based on seasonal data. I used Scikit-learn to train and test the models and evaluated them using RMSE and R-squared.

**6. Data Mining / Machine Learning**

**Definition:**  
Finding patterns in large datasets using algorithms like clustering, classification, or association rules.

**My Expertise:**  
In my coursework, I used K-means clustering to group customers by their transaction behavior. I plan to grow in this area by learning more complex techniques like decision trees and ensemble models..

**7. Reporting and Visualization**

**Definition:**  
Presenting data insights in a clear and meaningful way through dashboards, reports, or presentations.

**My Expertise:**  
I created interactive dashboards using Tableau and used Python to make visualizations. I shared these with my internship supervisor, and they helped guide decisions about how to improve the product based on user activity trends.

**A1. Gaining Expertise in Each Phase**

1. **Business Understanding:**  
   To get a deeper understanding of this concept, I will try to attend more meetings with stakeholders and listen to how the upper level of members on my team turn the proper into data questions. By being apart of these conversations it would help me understand what matters most to the organization and how data can support business decisions.
2. **Data Acquisition:**  
   I want to keep building my skills with SQL by working on large databases. I also plan to learn how to connect to APIs using Python so I can pull in outside data when needed.
3. **Data Cleaning:**  
   I plan to download open-source datasets and practice cleaning them using Python. I’ll focus on removing errors, fixing missing data, and making sure the data is ready for analysis.
4. **Data Exploration:**  
   I want to get better at using Python libraries like Matplotlib and Seaborn to make charts and graphs. I’ll also practice with tools like Power BI or Tableau to help me visually explore data and spot trends.
5. **Predictive Modeling:**  
   To gain experience here, I’ll take some courses/certifications in machine learning. I’d like to try out models like logistic regression and random forests so I can make predictions based on data.
6. **Data Mining/Machine Learning:**  
   To gain more exposure to this I will work on side projects using real-world datasets to uncover insights using clustering or classification.
7. **Reporting and Visualization:**  
   I plan to create an analytics portfolio with interactive dashboards and visual reports. This will help me get better at telling stories with data and sharing insights in a way others can understand.

**A2. How Organization Goals & Mission Guide Analysts**

**Organizational Goals:**  
An organization’s goal is usually something short-term and specific, like “improve customer satisfaction by 15% in the next 6 months” or “cut costs on shipping by 10% this year.” These goals help the data analyst know exactly what the company wants to achieve. Once the goal is clear, the analyst can figure out what kind of data to collect and what questions need to be answered. The goal basically acts like a guide, so the analyst doesn’t waste time analyzing things that aren’t helpful. A good data project should help the organization meet its goal in a smart and efficient way.

**Organizational Mission:**  
While a goal is short-term, the mission of a company is more long-term and big-picture. A mission tells you what the company stands for and what it hopes to do in the future. For example, if a company’s mission is “to make healthy food affordable for everyone,” then every data analysis project should support that mission. If it doesn’t, it’s probably not worth doing. Knowing the mission helps the analyst stay focused on what really matters to the organization and make sure their work fits with the company’s values and purpose.

**B. Applying a Tool: Python in Data Exploration**

**Tool/Technique:** Python  
**Phase:** Data Exploration

**Application:**  
As a data analyst intern, I used Python to explore customer usage logs. During the Data Exploration phase, I worked with Pandas to clean and organize the data. I also used Matplotlib to create bar charts and histograms. These visuals helped me find patterns in how users interacted with the app, including peak usage times and key demographics. Our team used these insights to adjust app features to better fit the needs of our users.

**B1. Three Risks of Using Python**

Three risks I might face while using Python to analyze consumer data include:

1. **Data Privacy**

It is important to protect any private or personal information in a dataset. If sensitive data like full names or home addresses are included, they should be removed or hidden. Only the data that is necessary for the analysis should be used.

**2. Security Breaches**

If the data is saved on a local computer or in an unprotected file, there is a risk that someone could steal it. To stay safe, data should be encrypted and stored in a secure place. Only people who need access to the data should be able to view it.

1. **Bias or Misleading Results**

If the dataset is not complete or does not include a fair sample of the population, it can lead to biased results. For example, if only one age group responds to a survey, the results will not represent all users. This can cause the data to tell the wrong story.

**B2. Technical Problem with Python**

**Problem:**  
Python is great for analyzing data and creating simple visualizations, but it is not always the best tool for making interactive dashboards. During my internship, we needed a dashboard that included a clickable map. While it is possible to do this in Python using Plotly or Dash, it would have taken a lot of time and coding. Instead, we used Tableau, which made it faster and easier to build the interactive features we needed.

**C. Decision-Making Process**

**C1. Tool Selection Justification**

Choosing the right tool depends on what the project needs. Some tools are better for certain tasks. For example, SQL is good for working with structured data, R is good for statistical models, and Tableau is helpful for dashboards. I chose Python because I already know SQL, so it felt like a natural next step. Python was also the best choice for cleaning and exploring the data before creating visual reports.

**C2. Results of Tool Use**

With Python, I was able to explore the data and find important trends. I shared my charts and summaries with the product team, and they used the results to make changes to the app. One example was improving app updates by scheduling them during low-traffic times, based on the data I found.

**C3. Ethical Considerations**

When using consumer data, analysts must be careful to act responsibly. Here are three ethical risks that should be avoided:

1. **Leaving Out Bias or Data Problems**

If there are problems with the data, such as missing values or uneven responses, these issues should be clearly explained. If not, the results might seem more reliable than they actually are.

1. **Not Being Clear or Honest**

People should know how their data is being used. Analysts should not hide how the data was collected or change the results to fit a certain goal. It is important to stay honest and follow scientific methods.

1. **Not Protecting Personal Data**

Sensitive data like names, birthdates, or home addresses should not be collected unless absolutely necessary. The data that is collected must be stored safely so that it is not stolen or misused.

**References**

Wing, J. M. (2019). The Data Life Cycle. Harvard Data Science Review, 1(1). <https://doi.org/10.1162/99608f92.e26845b4>