Module 4 Challenge Quiz – Part 1

1. You are a data analyst at a retail company. You want to analyze customer data to identify trends in customer behavior. Which of the following is a benefit of using cloud data analytics for this task?

- You can reduce costs for hardware and software.
- You are able to streamline the supply chain.
- You can custom tailor communications to customers.
- **You are able to access large amounts of data from multiple sources.**

Feedback: Cloud data analytics enables access to vast datasets from various sources without the need for local infrastructure. The other options may be benefits of analytics in general, but this one is specific to cloud capabilities.

2. You are a data engineer at a healthcare company. You want to build a data pipeline to process and analyze patient data from multiple sources. Which of the following Google Cloud Platform services is best suited for this task?

- **Cloud Data Fusion**
- Looker
- Dataproc
- BigQuery

Feedback: Cloud Data Fusion is designed for building and managing ETL pipelines across multiple data sources. Looker is for visualization, Dataproc is for big data processing, and BigQuery is for querying large datasets.

3. You are a data analyst on a team that is working on a new project. You need to access a data source that is currently only accessible to the engineering team. What should you do?

- Change the permissions on the data source so that everyone has access to it.
- Delete the data source and create a new one that you can access.
- Create a new role for yourself that has access to the data source.
- **Request access to the data source from the engineering team.**

Feedback: Proper protocol is to request access through the appropriate channels. Changing permissions or deleting data without authorization is not secure or professional.

4. A data analyst works for a company that runs clinical trials for a health care product. Stakeholders want to know how long participants spend filling out surveys during the clinical trials. The analyst receives a business data request for a report about survey timing. What type of request is this?

- Internal request for a data dashboard
- External request for a data report
- **Internal request for a data report**
- External request for a data dashboard

Feedback: The request comes from stakeholders within the company and asks for a report, not a dashboard. Hence, it's an internal request for a data report.

5. Which of the following is an example of an outlier within data?

- A product that sells for the average price of any other product of the same type.
- The mode of social media platform used by most people across the world.
- The median income for a family of four accessed from a nation's census data.
- **A student who scores a 100% on a test while everyone else scores under 40%.**

Feedback: This is a clear example of an outlier—an observation point that is distant from other observations in the dataset.

6. A data analyst is working for an app-based gaming company. The advertising team wants to know what percent of users are clicking on ads during gameplay. What data management tool should the analyst use to host all data once it has been compiled?

- Cloud Data Fusion
- BigQuery
- Dataflow
- **Google Cloud Storage**

Feedback: Google Cloud Storage is used to host and store compiled data. BigQuery is for querying, Dataflow for processing, and Cloud Data Fusion for integration.

7. A cloud data analyst works for a major online retailer. They are tasked with managing the company's unstructured data. The analyst wants to use a fully managed service that maximizes open-source data tools for batch processing, querying, streaming, and machine learning. Which of the following services should the data analyst use?

- Google Cloud Storage

- **Dataproc**
- Cloud Data Fusion
- Dataflow

Feedback: Dataproc is a fully managed service that supports open-source tools like Apache Spark and Hadoop for processing large-scale data.

8. A data analyst working on a new project needs to find out why a specific query was written a year ago. Where is the best place to find this information about the query?

- In the data team's documentation
- In the code version control system
- **In the data request central system**
- In the task tracking system

Feedback: The data request central system stores the context and communication around data requests, including the rationale behind queries.

9. A data analyst reviewing a data request notices that the requestor has not specified how often the data needs to be refreshed. What should the analyst do?

- Assume the data does not need refreshing.
- **Ask the requestor for clarification.**
- Refresh the data daily.
- Refresh the data weekly.

Feedback: Clarifying with the requestor ensures the data is refreshed at the appropriate frequency and avoids unnecessary processing.

10. A data professional has been asked to work on a data project that uses a data playbook. How often should the data playbook be updated?

- **The data playbook should be updated as needed.**
- The data playbook should be updated at least once a month.
- The data playbook should be updated at least once a quarter.
- The data playbook should be updated at least once a year.

Feedback: A data playbook is a living document and should be updated whenever processes or tools change to remain accurate and useful.