

**Topic 1 - Single Topic**

## Question #1

Topic 1

You are building an ML model to detect anomalies in real-time sensor data. You will use Pub/Sub to handle incoming requests. You want to store the results for analytics and visualization. How should you configure the pipeline?

- A. 1 = Dataflow, 2 = AI Platform, 3 = BigQuery
- B. 1 = DataProc, 2 = AutoML, 3 = Cloud Bigtable
- C. 1 = BigQuery, 2 = AutoML, 3 = Cloud Functions
- D. 1 = BigQuery, 2 = AI Platform, 3 = Cloud Storage

## Question #2

Topic 1

Your organization wants to make its internal shuttle service route more efficient. The shuttles currently stop at all pick-up points across the city every 30 minutes between 7 am and 10 am. The development team has already built an application on Google Kubernetes Engine that requires users to confirm their presence and shuttle station one day in advance. What approach should you take?

- A. 1. Build a tree-based regression model that predicts how many passengers will be picked up at each shuttle station. 2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the prediction.
- B. 1. Build a tree-based classification model that predicts whether the shuttle should pick up passengers at each shuttle station. 2. Dispatch an available shuttle and provide the map with the required stops based on the prediction.
- C. 1. Define the optimal route as the shortest route that passes by all shuttle stations with confirmed attendance at the given time under capacity constraints. 2. Dispatch an appropriately sized shuttle and indicate the required stops on the map.
- D. 1. Build a reinforcement learning model with tree-based classification models that predict the presence of passengers at shuttle stops as agents and a reward function around a distance-based metric. 2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the simulated outcome.

You were asked to investigate failures of a production line component based on sensor readings. After receiving the dataset, you discover that less than 1% of the readings are positive examples representing failure incidents. You have tried to train several classification models, but none of them converge. How should you resolve the class imbalance problem?

- A. Use the class distribution to generate 10% positive examples.
- B. Use a convolutional neural network with max pooling and softmax activation.
- C. Downsample the data with upweighting to create a sample with 10% positive examples.
- D. Remove negative examples until the numbers of positive and negative examples are equal.

You want to rebuild your ML pipeline for structured data on Google Cloud. You are using PySpark to conduct data transformations at scale, but your pipelines are taking over 12 hours to run. To speed up development and pipeline run time, you want to use a serverless tool and SQL syntax. You have already moved your raw data into Cloud Storage. How should you build the pipeline on Google Cloud while meeting the speed and processing requirements?

- A. Use Data Fusion's GUI to build the transformation pipelines, and then write the data into BigQuery.
- B. Convert your PySpark into SparkSQL queries to transform the data, and then run your pipeline on Dataproc to write the data into BigQuery.
- C. Ingest your data into Cloud SQL, convert your PySpark commands into SQL queries to transform the data, and then use federated queries from BigQuery for machine learning.
- D. Ingest your data into BigQuery using BigQuery Load, convert your PySpark commands into BigQuery SQL queries to transform the data, and then write the transformations to a new table.

[Next Questions →](#)