**Functional Specification: Whylogs and WhyLabs Integration**

**Overview**

This Python script is designed to monitor and analyze datasets using the ***Whylogs*** library and integrate these logs with ***WhyLabs*** for further analysis and visualization. The script performs the following key functions:

* Trains a K-Nearest Neighbors (KNN) model on the Iris dataset.
* Loads and processes multiple batches of data.
* Generates predictions using the trained KNN model.
* Logs the dataset profiles and classification metrics to ***WhyLabs***, both with and without segmentation.

**Functional Components**

1. **Environment Setup (set\_*WhyLabs*\_env)**
   * **Purpose:** Sets the necessary environment variables to connect the ***Whylogs*** library with the ***WhyLabs*** platform.
   * **Inputs:**
     + dataset\_id: The ID of the ***WhyLabs*** dataset where the logs will be sent.
   * **Outputs:** Sets the environment variables for ***WhyLabs*** organization ID, API key, and default dataset ID.
2. **Model Training (train\_knn\_model)**
   * **Purpose:** Loads the Iris dataset and trains a KNN model to predict the species of iris flowers based on their features.
   * **Inputs:** None (internally loads the Iris dataset).
   * **Outputs:**
     + knn: Trained KNN model.
     + df\_iris: DataFrame containing the Iris dataset.
3. **Load Datasets (load\_datasets)**
   * **Purpose:** Loads datasets from provided URLs, extracts the specified features and target values.
   * **Inputs:**
     + urls: List of URLs pointing to the CSV files to be loaded.
     + feature\_names: List of feature names to be extracted from the datasets.
   * **Outputs:**
     + datasets: List of DataFrames, each containing the features for one batch.
     + targets: List of Series, each containing the target values for one batch.
4. **Predict and Append (predict\_and\_append)**
   * **Purpose:** Uses the trained KNN model to generate predictions and appends the results to the corresponding DataFrames.
   * **Inputs:**
     + dfs: List of DataFrames containing the dataset batches.
     + knn: Trained KNN model.
     + class\_names: List of class names corresponding to the target labels.
   * **Outputs:** Updated DataFrames with cls\_output (predicted class) and prob\_output (prediction probability) columns.
5. **Log Profiles and Metrics (log\_profiles\_and\_metrics)**
   * **Purpose:** Logs dataset profiles and classification metrics to ***WhyLabs***.
   * **Inputs:**
     + dfs: List of DataFrames containing the dataset batches.
     + targets: List of Series containing the target values.
     + model\_id: The ***WhyLabs*** dataset ID to log the data.
     + segment\_column: (Optional) The column name used for segmenting the data in ***WhyLabs***.
   * **Outputs:** Logs profiles and metrics to ***WhyLabs***, with segmentation if segment\_column is provided.
6. **Main Execution Flow (main)**
   * **Purpose:** Orchestrates the entire process from model training to logging the profiles and metrics.
   * **Steps:**
7. **Train the KNN Model:** Trains a KNN model using the Iris dataset.
8. **Load Datasets:** Loads datasets both with and without the state feature.
9. **Generate Predictions:** Predicts outcomes for both sets of datasets using the trained model.
10. **Log to WhyLabs (Model-6):** Logs the profiles and metrics for datasets without state to ***WhyLabs*** under model-6.
11. **Log to WhyLabs (Model-7):** Logs the profiles and metrics for datasets with state to ***WhyLabs*** under model-7, using state as the segmentation column.
12. **Create Reference Profile:** Logs a reference profile for the Iris dataset to ***WhyLabs***.

**Error Handling and Warnings**

* **Future Warnings Suppression:** The script suppresses specific future warnings related to ***Whylogs*** to ensure smooth operation without deprecation warnings cluttering the output.
* **Dimensionality and Format Checking:** The script includes checks to ensure that columns like state are correctly formatted and 1-dimensional before logging to ***WhyLabs***, raising errors if these conditions are not met.

**Use Cases**

* **Model Monitoring:** This script is used for monitoring a KNN model's predictions across different datasets over time.
* **Data Drift Detection:** By logging profiles and classification metrics, this script helps detect data drift and model performance degradation.
* **Segmentation Analysis:** The use of state as a segmentation column in ***WhyLabs*** allows for granular analysis of model performance across different states.

This specification provides a detailed functional overview of the script, ensuring it meets the needs of users who want to monitor, log, and analyze machine learning model performance using ***Whylogs*** and ***WhyLabs***.