**Problem: Fetch product reviews and analyze them**

# HOW TO RUN

1. Download and install python virtual env
2. Activate the virtual env
3. Goto root of the folder
4. Take clone of the repo
5. Goto root of the project repo
6. Run command
    a. pip install -r requirements.txt
7. Quick Run
    a. Goto folder app
    b. run python main.py
8. Quick Run
    a. Goto folder app
    b. run python main.py

9. Modifications
    a. Open main.py
        i. Modify json path with new file
            1. *NOTE: INPUT SAMPLE PROVIDED IN THE ASSIGNMENT IS NOT A VALID FORMAT I HAVE UPDATED THE INPUT JSON FORMAT PLEASE LOOK INTO THE FILE **app/input.json***
        ii. Update the api_key with new key
            1. *NOTE: DUE TO SOME TECHNICAL ISSUE I WAS NOT ABLE TO CREATE ACCOUNT IN RAINFOREST I SHARED THIS INFORMATION EARLIER I DEVELOPED THIS USING **"demo"** KEY.*

# EXPLANATION & ASSUMPTIONS:

**EXPLANATION**

1. The solution is divided into four layers
    a. Runner Layer
    b. Pipeline Layer
    c. Data Layer
    d. API Layer
2. Runner Layer
    a. This layer is the entry point which consumes the input json file and credentials of api.
3. Pipeline Layer
    a. This layer has the core processing engine which takes input and provides the final result.
    b. This is taking a list of product ids from the runner layer.
    c. Using the product ids it is sending the data to the data layer which returns the data in merged dataframe datatype.
    d. Then Final calculations happens on the dataframe
4. Data Layer
    a. This layer takes product ids from pipelines makes api call to api layer and finally create a dataframe and send back to pipeline
5. API Layer
    a. This has the Rainforest api class which has the logic to contact the api server of rainforest.

**ASSUMPTIONS:**

1. The Json format provided in the assignment file has incorrect format, i have corrected the format and using it, you can find it in the input.json file.
2. The output should be a stdout(standard output) which means it will print the results and not save anywhere
3. Logging and exception handling are not excluded.

**BELOW DIAGRAM EXPLAINS HOW DATA IS TRAVELING IN THE 4 FOUR LAYERS.**

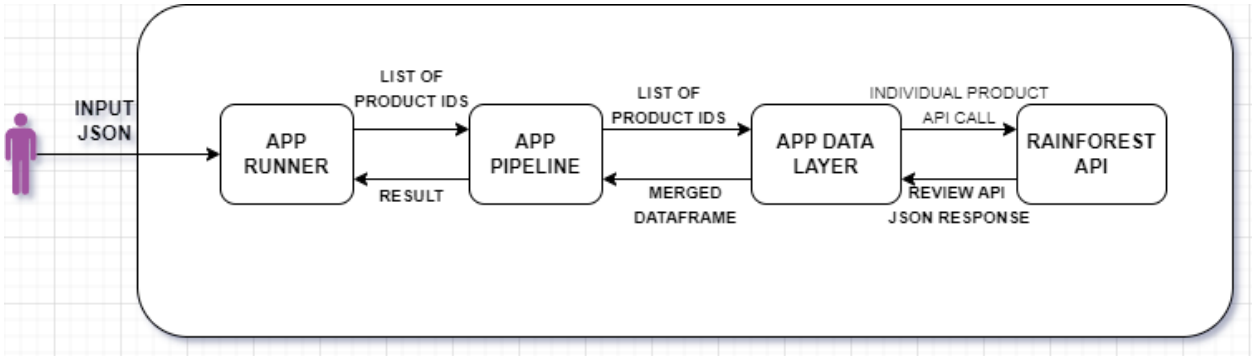


**FIGURE: DATA FLOW DIAGRAM**