AI ML | MODULE PROJECT

greatlearning
Power Ahead

1 AIML module projects are designed to have a detailed hands on to integrate theoretical knowledge with actual practical implementations.

2 AIML module projects are designed to enable you as a learner to work on realtime industry scenarios, problems and datasets.

3 AIML module projects are designed to enable you simulating the designed solution using AIML techniques onto python technology platform.

4 AIML module projects are designed to be scored using a predefined rubric based system.

5 AIML module projects are designed to enhance your learning above and beyond. Hence, it might require you to experiment, research, self learn and implement.

5 Takeaways

AIML | MODULE PROJECT

greatlearning
Power Ahead

INTRODUCTION TO **PYTHON**

PART I

AIML module project part I consists of stepwise task based assessment which can be accomplished using python programming skills.

PART II

AIML module project part II consists of automating all the tasks from Part I which can be accomplished using python programming skills.

TOTAL SCORE 60

PART ONE | PROJECT **BASED** | TOTAL SCORE | 35

Part I consists of the following steps:

1. Create a pandas data frame consisting of 4 columns and 20 rows.

2. Column 1 should have values between 0 to 25.

3. Column 2 should have values between -10 to 10.

4. Column 3 should have values by multiplying column 1 and 2.

5. Column 4 should have value "Positive" if column 3 is a positive number.

6. Column 4 should have value "Negative" if column 3 is a negative number.

7. Create a new data frame of 1 column and 20 rows having random alphanumeric values of size 5. Name the column as 'Password'

8. Merge both the data frames to one single data frame.

9. Create a 2 new columns i.e. 'Word' and 'Number' by separating the alphabets and numbers from column "Password".

10. Create a new column 'Anomaly'. This column should have value 'Found' if the row entry in 'Number' is empty.

11. Insert a new column "Date" by inserting the current date. [format: mmm-dd-yyyy]

12. Insert a new column "Time" by inserting the current time. [format: hh-mm-ss]

13. Create a new column 'Comment'. This column should have value 'Non office hours of working' if the row entry in 'Time' is not between 9 am to 7 pm.

14. Export the data frames to excel, csv, html and json formats.

15. Measure and display how much time it takes to execute the code. Optimise your code to reduce the time.

PART TWO | PROJECT **BASED** | TOTAL SCORE | 25

- Automate all the tasks listed in part I.
- Your function should be able to create datasets with N rows. [Where N is defined by the user. N is any integer.]
- Your function should be able to export datasets onto local machine in excel, csv, json, html or all formats. User can choose this.
- Purpose of this function is to allow user to generate a dataset with N rows and store it on local machine.
- You might need to write a web of complex interrelated functions

Hint: You might need to write a web of complex interrelated functions to make a flawless automation. Write individual or interrelated functions to automate steps in part I. Anything that needs just a trigger to start and auto-execute all the steps. Please find the below screenshot for your reference: [Just a sample for you to visualise. You can be more innovative in your approach]

```
1  Main_Function(33,"html")

Starting the main function
1. Create a pandas data frame consisting of 4 columns and 20 rows
2. Column 1 should have values between 0 to 25
3. Column 2 should have values between -10 to 10
4. Column 3 should have values by multiplying column 1 and 2
5. Column 4 should have value "Positive" if column 3 is a positive number.
6. Column 4 should have value "Negative" if column 3 is a negative number
7. Create a new data frame of 1 column and 20 rows having random alphanumeric values of size 5
8. Merge both the data frames to one single data frame
9. Create a 2 new columns i.e. 'Word' and 'Number' by separating the alphabets and numbers from column "Password
10. Create a new column 'Anomaly'. This column should have value 'Found' if the row entry in 'Number' is empty.
11. Insert a new column "Date" by inserting the current date. [ format: mmm-dd-yyyy ]
12. Insert a new column "Time" by inserting the current time. [ format: hh-mm-ss ]
13. Create a new column 'Comment'. This column should have value 'Non office hours of working' if the row entry in '
Time' is not between 9 am to 7 pm.
Export the data frames to html
Time taken 0.078 seconds
Ending the main function
```

LEARNING OUTCOME

Deep understanding on creating, manipulating, imputing, deleting and merging data/data-frames into python data frames using numpy and pandas library.

Possibility for creating new attributes or features as per the business requirement.

Understanding on how to create, split & manipulate non numeric data.

Building optimised python codes using python's execution time cycles as a reference.

Hands on practise on understanding automation and code reusability using python programming.

" *Put yourself in the shoes of an actual* "

DATA SCIENTIST

THAT's **YOU**

Assume that you are working at the company which has received the above problem statement from internal/external client. Finding the best solution for the problem statement will enhance the business/operations for your organisation/project. You are responsible for the complete delivery. Put your best analytical thinking hat to squeeze the raw data into relevant insights and later into an AIML working model.

PLEASE **NOTE**

Designing a data driven decision product typically traces the following process:

1. Data and insights:

    Warehouse the relevant data. Clean and validate the data as per the the functional requirements of the problem statement. Capture and validate all possible insights from the data as per the the functional requirements of the problem statement. Please remember there will be numerous ways to achieve this. Sticking to relevance is of utmost importance. Pre-process the data which can be used for relevant AIML model.

2. AIML training:

    Use the data to train and test a relevant AIML model. Tune the model to achieve the best possible learnings out of the data. This is an iterative process where your knowledge on the above data can help to debug and improvise. Different AIML models react differently and perform depending on quality of the data. Baseline your best performing model and store the learnings for future usage.

3. AIML end product:

    Design a trigger or user interface for the business to use the designed AIML model for future usage. Maintain, support and keep the model/product updated by continuous improvement/training. These are generally triggered by time, business or change in data.

greatlearning
Power Ahead

IMPORTANT
POINTERS

Project should be submitted as a single ".**html**" and ".**ipynb**" file. Follow the below best practices where your submission should be:

- ".html" and ".ipynb" files should be an exact match.
- Pre-run codes with all outputs intact.
- Error free & machine independent i.e. run on any machine without adding any extra code.
- Well commented for clarity on code designed, assumptions made, approach taken, insights found and results obtained.

Project should be submitted on or before the deadline given by the program office.

SUBMISSION

Project submission should be an original work from you as a learner. If any percentage of plagiarism found in the submission, the project will not be evaluated and no score will be given.

greatlearning
Power Ahead

HAPPY
LEARNING