PROJECT

Module

Programming Essentials

Module Code

IT43002FP

Duration

20 hours

Title: Create an application through programming.

Tools, Equipment and Materials:

1 Personal Computer with Internet Access
2 Python Software Tools

Instructions:

1. This project is done in a group comprising of not more than 3 members in a team. The group needs to complete the scenario below based on the instruction from the Lecturer.

2. The project groups are required to do the following:
   i. Analyze and understand the project requirements;
   ii. Identify the appropriate library package(s) needed for the application based on the project requirements.
   iii. Display appropriate outcomes.
   iv. Present findings or insights for the data gathered or analyze.

Deliverables

The group has to zip the application folder and presentation slides into one compressed zip file for submission. There should only be 1 submission per group.

Ensure that the following are included in the zip file:
- Application source code with comments to explain the logic and functionalities;
- Presentation slides

The group has to present the slides and demonstrate the application to the Lecturer upon completion of the project.

PROJECT

Module

Programming Essentials

Module Code

IT43002FP

Duration

20 hours

Project Scenario

DataMax Pte Ltd has just clinched a deal from an important client *ExpenseTrack Pte Ltd*. The project is to create an application to analyse travellers' trend. The application should identity the top 3 countries of visitors to Singapore from a specific region over a span of *10* years.

ExpenseTrack believes that Python is one of the most appropriate programming languages for data analyses. *ExpenseTrack* is expecting to use Python to create script(s) to process the below data and present an important insight.

Being an experienced Data Analyst for *DataMax*, you are expected to create Python script(s) to analyse any one of the following period for any one region.

S/No	Years	Regions			Data
1	1978 - 1987	Asia	Europe	Others (Non-Asia & Non-Europe)	Int Monthly Visitor.xlsx
2	1988 - 1997				
3	1998 - 2007				
4	2008 - 2017				

Table 1

Countries in Asia - ' Brunei Darussalam ', ' Indonesia ', ' Malaysia ', ' Philippines ', ' Thailand ', ' Viet Nam ', ' Myanmar ', ' Japan ', ' Hong Kong ', ' China ', ' Taiwan ', ' Korea, Republic Of ', ' India ', ' Pakistan ', ' Sri Lanka ', ' Saudi Arabia ', ' Kuwait ', ' UAE '

Countries in Europe - ' United Kingdom ', ' Germany ', ' France ', ' Italy ', ' Netherlands ', ' Greece ', ' Belgium & Luxembourg ', ' Switzerland ', ' Austria ', ' Scandinavia ', ' CIS & Eastern Europe '

Countries in Others – ' USA ', ' Canada ', ' Australia ', ' New Zealand ', ' Africa '

PROJECT

Module

Programming Essentials

Module Code

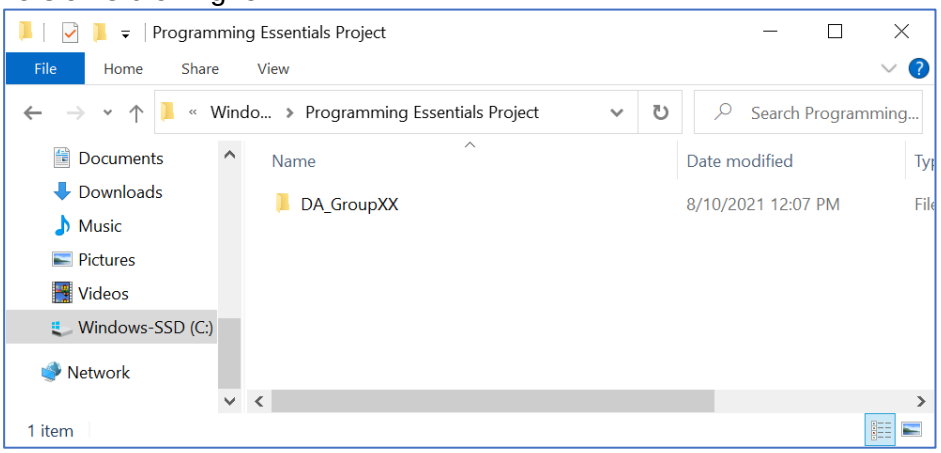
IT43002FP

Duration

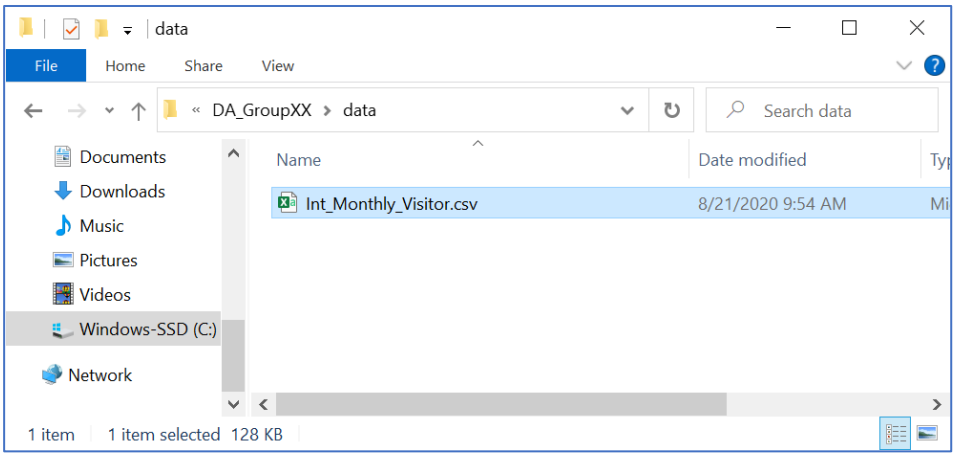
20 hours

Tasks

1. Create a project folder named DA_<Group Name> with script(s) to use Python version 3.0 or higher.



2. The path to the data file should be correctly defined for the imported library class to retrieve the relevant data.



3. The script(s) should import appropriate library for data analysis.

4. The script(s) should prompt user to enter the time period and region for data analysis and call the functions defined below in **VisitorsAnalyticsUtils** to perform the data analysis.
   i. The program should exit if user enter an invalid value for time period or region.
   ii. Refer to screenshot below as a reference for the prompt message and the expected user input.

```
Enter year period (1: 1978-1987, 2: 1988-1997, 3: 1998-2007, 4: 2008-2017): 1
Enter region (1: Asia, 2: Europe, 3: Others): 2
```

5. The script(s) should define a **class** named **VisitorsAnalyticsUtils**.

PROJECT

Module
Programming Essentials

Module Code
IT43002FP

Duration
20 hours

6. In **VisitorsAnalyticsUtils** class, create a **function** named **loadDataFile** to retrieve data from the given data file and load the data into an appropriate data structure for ease of computation and graphic display of results:
   i. Create appropriate class from the imported library / libraries
   ii. Use the appropriate function to load data from data file
   iii. Load the data into an appropriate data structure for ease of computation and graphic display of results
   iv. Display the loaded information (refer to sample output below)

```
Enter region (1: Asia, 2: Europe, 3: Others): 1
*** first 5 rows of data loaded ***
            Brunei Darussalam    Indonesia   ...   New Zealand    Africa
                                             ...
 1978 Jan                  0.0          0.0  ...          3612       587
 1978 Feb                  0.0          0.0  ...          2521       354
 1978 Mar                  0.0          0.0  ...          2727       405
 1978 Apr                  0.0          0.0  ...          3197       736
 1978 May                  0.0          0.0  ...          5130       514

[5 rows x 34 columns]
```

7. In **VisitorsAnalyticsUtils** class, create a **function** named **parseData** to prepare the data in Table 1 based on the time period and region entered by user:
   i. Identify the specified data for computation
   ii. Parse the data to retain the necessary regions only
   iii. Parse the data to retain the necessary 10 years period only (hint: split the critical information to different columns as numeric for easy processing)
   iv. Display the relevant information (refer to sample output below)

```
*** Parsed Data (Regions) ***
<class 'pandas.core.frame.DataFrame'>
Index: 120 entries, 1998 Jan to 2007 Dec
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0    USA            120 non-null    int64
 1    Canada         120 non-null    int64
 2    Australia      120 non-null    int64
 3    New Zealand    120 non-null    int64
 4    Africa         120 non-null    int64
 5   year            120 non-null    int64
dtypes: int64(6)
memory usage: 6.6+ KB
None
```

PROJECT

Module
Programming Essentials

Module Code
IT43002FP

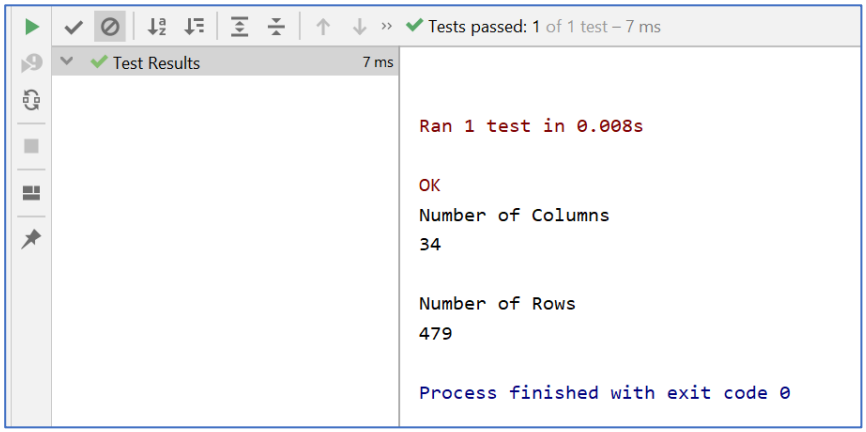
Duration
20 hours

```
*** Parsed Data (Years) ***
count     120.000000
mean     2002.500000
std         2.884324
min      1998.000000
25%      2000.000000
50%      2002.500000
75%      2005.000000
max      2007.000000
Name: year, dtype: float64
```

8. In **VisitorsAnalyticsUtils** class, create a **function** named **getTop3Countries** to display the top 3 countries in the assigned region and 10 years period (refer to sample output below).

```
*** Top 3 countries ***
 Australia       5527129
 USA             3514969
 New Zealand      989725
 Africa           833109
 Canada           712490
dtype: int64
```

9. The project should also include a Unit Test Case to test loadDataFile function. The test case should ensure the following (refer to sample output below):
   i. The data structure contains the correct number of rows.
   ii. The data structure contains the correct number of columns.



10. Finally, a presentation is required to share your insights.

- END -