A blue and black logo

Description automatically generated

**Information and Data Collation using Python with Excel**

**Functionality & Code Documentation**

**Dated: November 2023**

**For: Catering & Logs Team**

Table of Contents

[Agenda & Purpose 3](#_Toc152167315)

[Background 3](#_Toc152167316)

[Solution 3](#_Toc152167317)

[Constraints & Assumptions 3](#_Toc152167318)

# Agenda & Purpose

## Agenda

This documentation serves as a purpose to document down the functionalities of the code and explanation. Some knowledge of Python programming is beneficial as there will be code snippets in this documentation to demonstrate the functionalities.

## Background

Catering and Logs (C&L) team needs an efficient and reliable way to compile an overview of the courses that is being provided by SMU Academy. The C&L team is responsible to order catering services whenever there are courses that is being offered. Navigating between multiple windows of different Excel sheets is a challenge to effectively carry out the task.

## Solution

The solution created to tackle the problem is a Python program script that can generate the desired output of overview consistently. Within clicks of the Python script, it should be able to compile Excel sheets, manipulate data, and generate an output Excel sheet with overview.

## Constraints & Assumptions

Before coming up with the Python program, there are certain assumptions and constraints that were made to scoping the requirement:

* There will be 3 exact Excel (not CSV) files with certain naming standardised (“gvSession.xlsx”, “Manage Schedule.xlsx”, “Enrolment Summary.xlsx”) present in the same directory with the Python script. Otherwise, an error will be thrown.
* The headers’ name (a.k.a. the first row of Excel) does not change often.
* All the “manageSchedule.xlsx” *(“Schedule”)* records will have more than or equals to 0 records of “gvSession.xlsx” *(“Session”)* records. If there is a Session and its Schedule is not reflected, this means that the Session does not have a Schedule.
* Schedule will be the primary file that collates the data and information. There will be a key “Schedule #” *(“Schedule Number”)* that maps everything together.

# Functionalities & Code

## Main Code

This is the entry/starting point of the code. This will make sure that the code runs sequentially and executes. This part of the code is usually located the most bottom part of the code because Python executes from top-down, *i.e. executing A then B allow B to ‘use’ A, but not the other way round.*



A screen shot of a computer program

Description automatically generatedUnder the main function, it consists of various functions that executes step-by-step, which it is readable by its naming convention. In this part, I will try to summarise the code’s functionality, if need be, more elaboration can be found in the code functions explanation.

* Line 278-281: This is to check if there exist 3 exact files together with the same directory as the source code. The name of the files is stated in the code.
* Line 284: Read data from the relevant Excel files.
* Line 285: Convert the data that is being read into Dictionary­­­­[[1]](#footnote-1) so that it is easier to use
* Line 286: Compile all the different Sessions to a Schedule.
* Line 287: Get course audience and map it to relevant Schedule Number.
* Line 288: Get course pillar and map it to relevant Schedule Number.
* Line 290: Structure the data so that it is ready to export it out to an Excel sheet.
* Line 292-293: Get the current datetime in a certain format and use it as part of the filename when exporting the data.
* Line 295: Converts the Dictionary data into a DataFrame[[2]](#footnote-2) data so that it is easier to export as it has built-in functions to do so.
* Line 296: Sort the data by “Start Date” column first, then “Course No.” column. The order will be increasing.
* Line 298-300: Prepare the file type to be exported, and exports it to an Excel sheet. “index=False” means that there is no numbering in the very first column.

1. Dictionary is a type of data in Python. For more information about Python Dictionary, click [here](https://www.programiz.com/python-programming/dictionary). [↑](#footnote-ref-1)
2. DataFrame is type of data by a popular library by Pandas that is good for analyzing data, but it is not built-in Python data types. For more information about DataFrame, click [here](https://www.w3schools.com/python/pandas/pandas_dataframes.asp). [↑](#footnote-ref-2)