

處理 excel 檔案 (公司的出勤表)

目前人工的做法

标准报表.xlsx - LibreOffice Calc ← 打卡系統產生的檔案

檔案(F) 編輯(E) 檢視(V) 插入(I) 格式(O) 樣式(Y) 工作表(S) 資料

SimSun 9 B I U A

J3 fx Σ = 陳01

3	部门	峰騰		姓名	陳01				
4	日期	2020-08-01 2020-08-31		工号	11				
5	旷工(天)	请假(天)	出差(天)	上班(天)	加班(小时)	迟到(次)	早退(分)		
6				正常	特殊				
7	1	0	0	20	0:00 0:00	20	576	2	586

8

9 考 勤 表

10	日期	第一时段	第二时段	加班时段
11	星期	上班	下班	签到 签退
12	01 六			
13	02 日			
14	03 一	08:28	17:35	
15	04 二	08:29	17:10	
16	05 三	08:30	17:34	
17	06 四	08:27	17:32	
18	07 五	08:29	17:31	

出勤.xlsx - LibreOffice Calc ← 最終希望看到的出勤檔案

檔案(F) 編輯(E) 檢視(V) 插入(I) 格式(O) 樣式(Y) 工作表

標楷體 12 B I U A

D6 fx Σ =

1	9月份員工出勤資料表								
2	上班時間	上下班時間	8:30	08:30	08:30	08:30	8:30		
3	單位	姓名	卡號	張01	張02	張03	張04	張05	張06
4				32752	51438	42485	46759	38765	45745
5	卡號								
6	3(-)	上	8:28						
7		下	17:35						
8	4(-)	上							
9		下							
10	5(三)	上							
11		下							
12	6(四)	上							
13		下							
14	7(五)	上							
15		下							
16	10(-)	上							
17		下							
18	11(-)	上							
19		下							

人工填入

處理 Excel 檔案 (不討論作法 B)

- 作法 A：讀取 XLS(X) → 產生 XLSX
 - 使用 Python 的 **openpyxl** 第三方模組
 - 無法處理較舊的 XLS 格式，需要先轉成 XLSX 才能用
- 作法 B：讀取 XLS(X) → 產生 txt(純文字格式)
 - 公司有三份打卡資料，98-1 和 88 類似，純文字格式
 - 98-5 就是需要處理的 excel 格式原始檔案
 - 98-1 和 88 有公式自動將 txt 轉成 excel 表格

Standard toolbar with icons for file operations, editing, and formatting. Below the icons, the font is set to '宋体', size '9', and bold 'B' is active. The formula bar shows 'J7' and '=fx Σ = | 576'.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR
7	1	0	0	0	20	0:00	0:00	20	576	2	586	21	0	0	0	0:00	0:00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0:00	9:14	21	121	1	300				

考 勤 表								考 勤 表								考 勤 表							
日期	星期	第一时段		第二时段		加班时段		日期	星期	第一时段		第二时段		加班时段		日期	星期	第一时段		第二时段		加班时段	
		上班	下班	上班	下班	签到	签退			上班	下班	上班	下班	签到	签退			上班	下班	上班	下班	签到	签退
01	六							01	六							01	六						
02	日							02	日							02	日						
03	一	08:28	17:35					03	一				旷工			03	一	08:05	17:30				
04	二	08:29	17:10					04	二				旷工			04	二	08:10	17:30				
05	三	08:30	17:34					05	三				旷工			05	三	08:06	17:30				
06	四	08:27	17:32					06	四				旷工			06	四	08:04	17:30				
07	五	08:29	17:31					07	五				旷工			07	五	08:04	17:30				
08	六							08	六							08	六						
09	日							09	日							09	日						
10	一	08:31	17:35					10	一				旷工			10	一	08:04	17:30				
11	二	08:30	17:32					11	二				旷工			11	二	08:05	17:30				
12	三	08:29	17:31					12	三				旷工			12	三	08:08	17:30				
13	四	08:29	17:31					13	四				旷工			13	四	08:07	17:30				
14	五	08:28	12:05					14	五				旷工			14	五	08:02	17:30				
15	六							15	六							15	六						
16	日							16	日							16	日						
17	一	08:26	17:08					17	一				旷工			17	一	08:06	17:30				
18	二	08:30	17:30					18	二				旷工			18	二	08:06	17:30				
19	三	08:28	17:09					19	三				旷工			19	三	08:07	17:30				
20	四	08:28	17:31					20	四				旷工			20	四	08:01	17:30				
21	五	08:30	17:32					21	五				旷工			21	五	08:15	12:00				
22	六							22	六							22	六						
23	日							23	日							23	日						
24	一					旷工		24	一				旷工			24	一	08:03	17:30				
25	二	08:32	17:31					25	二				旷工			25	二	08:06	17:30				
26	三	08:32	17:32					26	三				旷工			26	三	08:04	17:30				
27	四	08:26	12:09					27	四				旷工			27	四	08:08	17:30				
28	五	08:26	17:37					28	五				旷工			28	五	08:04	17:30				
29	六							29	六							29	六						
30	日							30	日							30	日						
31	一	08:28	17:37					31	一				旷工			31	一	08:06	17:30				

(1)
一次先抓完一整
個月的出勤表
(不含加班)

L

M

R

(2)
之後再更新先前
空白的加班欄位

07:47 17:01


```

1 import openpyxl
2
3 from openpyxl.utils import get_column_letter, column_index_from_string
4
5 import os
6
7 filename_in = '1 标准报表.xlsx'
8 #filename_in = '6月.xlsx'
9
10 wb = openpyxl.load_workbook(filename_in)
11
12 #print(wb.worksheets[4].cell(row=12, column=1).value)
13 #print(wb.worksheets[4]['A12'].value)
14 header_1 = [wb.worksheets[4].cell(row=12+i, column=1).value for i in range(31)]
15 header_2 = header_1.copy()
16 header = []
17 for i in range(31):
18     header.append(header_1[i])
19     header.append(header_2[i])
20 header.insert(0, '姓名')
21 header.append('遲到(次)')
22 header.append('遲到(分)')
23
24 def get_block_times(person, row, col, days):
25     for i in range(days):
26         pos_row = row + i
27         pos_col = col
28         person.append(sheet.cell(row=pos_row, column=pos_col).value)
29         pos_col = col + 2
30         person.append(sheet.cell(row=pos_row, column=pos_col).value)
31
32 #employee_name = sheet['J3'].value # J3(L) ; Y3(M) ; AN3(R)
33
34 # weekday & dummy weekend
35 def get_person_times(row_begin, col_begin):
36     person = []
37     #
38     row = row_begin
39     col = col_begin
40     get_block_times(person, row, col, 31)
41     return person
42
43 people_all = []
44 for sheet in wb.worksheets[4:]:
45
46     row_begin = 12 #14 # 12-> whole month ; 14-> weekday only
47     col_begin = 2 # 2(L) ; 'Q' -> 17(M) ; 'AF' -> 32(R)
48     person_left month = get_person_times(row_begin, col_begin)
49
50     person_left_name = sheet['J3'].value
51     person_left = person_left month.copy()
52     person_left.insert(0, person_left_name)
53
54     # 遲到(次)
55     person_left.append(sheet.cell(row=row_begin-5, column=col_begin+7).value)

```

(這個程式原始碼尚未填入周末
的加班資料也尚未更正遲到統計)

處理最左邊 column
要顯示的字樣

定義要抓的 block

上班時間
下班時間

定義：處理一人一月出勤

指定左上角第一筆位置

從左邊算來第五張 sheet 開始處理個人出勤表，每張有 3 人

最左邊人的出勤資料

插入最左邊人的名字

```

57 # 遲到(分)
58 person_left.append(sheet.cell(row=row_begin-5, column=col_begin+8).value)
59
60 people_all.append(person_left)
61 print('got', person_left_name)
62
63 col_begin = 17 # 'Q' -> 17
64 person_middle month = get_person_times(row_begin, col_begin)
65
66 person_middle_name = sheet['Y3'].value
67 person_middle = person_middle month.copy()
68 person_middle.insert(0, person_middle_name)
69
70 # 遲到(次)
71 person_middle.append(sheet.cell(row=row_begin-5, column=col_begin+7).value)
72 # 遲到(分)
73 person_middle.append(sheet.cell(row=row_begin-5, column=col_begin+8).value)
74
75 people_all.append(person_middle)
76 print('got', person_middle_name)
77
78 col_begin = 32 # 'AF' -> 32
79 person_right month = get_person_times(row_begin, col_begin)
80
81 person_right_name = sheet['AN3'].value
82 person_right = person_right month.copy()
83 person_right.insert(0, person_right_name)
84
85 # 遲到(次)
86 person_right.append(sheet.cell(row=row_begin-5, column=col_begin+7).value)
87 # 遲到(分)
88 person_right.append(sheet.cell(row=row_begin-5, column=col_begin+8).value)
89
90 people_all.append(person_right)
91 print('got', person_right_name)
92
93 people_all.insert(0, header)
94
95 # generate new workbook
96 wb_out = openpyxl.Workbook() # Create a blank workbook.
97 sheet_out = wb_out.active
98 sheet_out.title = 'wt'
99
100 col_index = 1
101 for people in people_all:
102     row_index = 1
103     for record in people:
104         sheet_out.cell(row=row_index, column=col_index).value = record
105         row_index = row_index + 1
106         col_index = col_index + 1
107
108 basename tmp = os.path.basename(filename_in)
109 filename tmp = 'tmp3_' + basename tmp # 遲到(次), 遲到(分)
110
111 wb_out.save(filename tmp) # Save the workbook.
112

```

擷取系統自動產生的遲到紀錄

換處理中間的人

換處理最右邊的人

將總表擷取的資料寫入新 excel file

外圈：名稱和人的數量

內圈：名稱和每人內的內容

```

1 import openpyxl # 在command line視窗下打pip install openpyxl來安裝
2 from openpyxl.utils import get_column_letter, column_index_from_string # 轉換用
3 import os # 只是用來處理檔案名稱
4
5 filename_in = r'1-標準報表.xlsx' # 選擇要處理的檔案,記得先將xls存成xlsx格式
6 wb = openpyxl.load_workbook(filename_in) # 將使用者指定的xlsx檔載入
7
8 # --- 先準備好用來標示全體員工出勤總表的資訊名稱 ---
9 # 擷取A12~A42的日期名稱,並依次將一週日期擴充為兩項,分成上班和下班
10 header_0 = [wb.worksheets[4].cell(row=i+1, column=1).value for i in range(31)]
11 header = []
12 for date in header_0:
13     header.append(date + '(上)')
14     header.append(date + '(下)')
15 header.insert(0, '姓名') # 在每月的日期最前面加上"姓名"欄位
16 header.append('達到(次)') # 在每月的日期最後面加上"達到(次)"欄位
17 header.append('達到(分)') # 最後面再加上"達到(分)"欄位
18
19 def get_block_times(person, row, col, days):
20     """
21     在excel中,擷取上班和下班的時間,並將結果存入person.
22     Input
23     person: 單一員工的出勤日期list(執行後會被更新)
24     row: 擷取區塊左上角欄位的row座標
25     col: 擷取區塊左上角欄位的column座標
26     ...
27     days: 天數(擷取區塊的高度)
28
29     for i in range(days):
30         pos_row = row + i # row座標依序增加(0,1,2,3...)
31         pos_col = col
32         person.append(sheet.cell(row=pos_row, column=pos_col).value)
33         pos_col = col + 2 # 下班欄位和上班欄位間隔一個excel欄位
34         person.append(sheet.cell(row=pos_row, column=pos_col).value)
35
36 def get_person_times(row_begin, col_begin):
37     """
38     擷取單一員工31天(一個月)的上下班出勤或加班資料.
39     Input
40     row_begin: 單一員工左上角欄位的row座標
41     col_begin: 單一員工左上角欄位的column座標
42
43     Output
44     ... person: 單一員工一個月的上下班出勤或加班資料(list)
45
46     person = []
47     row = row_begin
48     col = col_begin
49     get_block_times(person, row, col, 31)
50     return person
51
52 def work_lag(person_raw):
53     """
54     重新計算達到次數和達到分鐘,而不是擷取系統算好的.
55
56     work_on = 0 # 每個月的總上班天數
57     work_on_lag_count = 0 # 達到次數
58     work_on_lag_time = 0 # 達到分鐘
59     for time in person_raw[1:2]: # 只處理上班時間,跳過下班時間
60         if time: # 若欄位有值才繼續,排除空白資料(None)
61             if '.' in time: # 排除"0"資料,有意義的時間會包含"."
62                 work_on += 1 # 只有是有意義的時間才會累積上班天數
63                 if time > '08:30': # 當上班時間超過08:30,才回計算次數和時間
64                     work_on_lag_count += 1
65                     hour, minute = time.split(':')
66                     minutes = int(hour) * 60 + int(minute)
67                     minutes = minutes - (9*60+30)
68                     work_on_lag_time += minutes
69
70     return (work_on_lag_count, work_on_lag_time)
71
72 people_all = [] # 準備好全體公司員工的出勤總表
73
74 for sheet in wb.worksheets[4:]: # 從第5個spreadsheet開始統計員工上下班時間
75     # --- 左邊的員工 ---
76     # 處理單一spreadsheet內的個人出勤總表
77     row_begin = 12 # x月一日出勤時間的起始座標 (row)
78     col_begin = 2 # x月一日加班時間的起始座標 (column)
79     person_left_month = get_person_times(row_begin, col_begin) # 抓取整月出勤表
80     # 重新計算上班的達到次數和達到分鐘
81     count, time = work_lag(person_left_month)
82     # 抓取整月的加班總表並加入出勤總表內

```

作法A 的程式原始碼 (已包含處理周末 加班和修正遲到統計)

```

80 row_begin = 12 # x月一日加班時間的起始座標 (row)
81 col_begin = 2 + 9 # x月一日加班時間的起始座標 (column)
82 person_left_add = get_person_times(row_begin, col_begin)
83 for i, time_add in enumerate(person_left_add):
84     if time_add:
85         person_left_month[i] = time_add
86 # 抓取人名並放到個人出勤總表開頭
87 person_left_name = sheet['J3'].value
88 person_left = person_left_month.copy()
89 person_left.insert(0, person_left_name)
90 # 捨棄系統自動產生的達到次數和達到分鐘
91 person_left.append(sheet.cell(row=row_begin-5, column=col_begin+7).value)
92 #person_left.append(sheet.cell(row=row_begin-5, column=col_begin+8).value)
93 #將重新計算好的達到次數和達到分鐘加到個人出勤總表
94 person_left.append(count)
95 person_left.append(time)
96 # 將個人出勤總表加到全體員工出勤總表內
97 people_all.append(person_left)
98 print('got', person_left_name)
99
100 # --- 中間的員工 ---
101 row_begin = 12
102 col_begin = 17 # 'Q' -> 17
103 person_middle_month = get_person_times(row_begin, col_begin)
104
105 count, time = work_lag(person_middle_month)
106
107 row_begin = 12
108 col_begin = 17 + 9
109 person_middle_add = get_person_times(row_begin, col_begin)
110 for i, time_add in enumerate(person_middle_add):
111     if time_add:
112         person_middle_month[i] = time_add
113 person_middle_name = sheet['Y3'].value
114 person_middle = person_middle_month.copy()
115 person_middle.insert(0, person_middle_name)
116 person_middle.append(count)
117 person_middle.append(time)
118 people_all.append(person_middle)
119 print('got', person_middle_name)
120
121 # --- 右邊的員工 ---
122 row_begin = 12
123 col_begin = 32 # 'AE' -> 32
124 person_right_month = get_person_times(row_begin, col_begin)
125
126 count, time = work_lag(person_right_month)
127
128 row_begin = 12
129 col_begin = 32 + 9
130 person_right_add = get_person_times(row_begin, col_begin)
131 for i, time_add in enumerate(person_right_add):
132     if time_add:
133         person_right_month[i] = time_add
134 person_right_name = sheet['AN3'].value
135 person_right = person_right_month.copy()
136 person_right.insert(0, person_right_name)
137 person_right.append(count)
138 person_right.append(time)
139 people_all.append(person_right)
140 print('got', person_right_name)
141
142 people_all.insert(0, header) # 將用來標示的資訊加入全體員工出勤總表最開頭
143
144 # --- 將上面整理好的公司全體員工出勤表存成excel檔案 ---
145 # generate new workbook
146 wb_out = openpyxl.Workbook() # Create a blank workbook.
147 sheet_out = wb_out.active
148 sheet_out.title = 'wt'
149 col_index = 1
150 for people in people_all:
151     row_index = 1
152     for record in people:
153         sheet_out.cell(row=row_index, column=col_index).value = record
154         row_index = row_index + 1
155         col_index = col_index + 1
156 basename_tmp = os.path.basename(filename_in)
157 filename_tmp = 'tmp985A_' + basename_tmp
158 wb_out.save(filename_tmp) # Save the workbook.

```