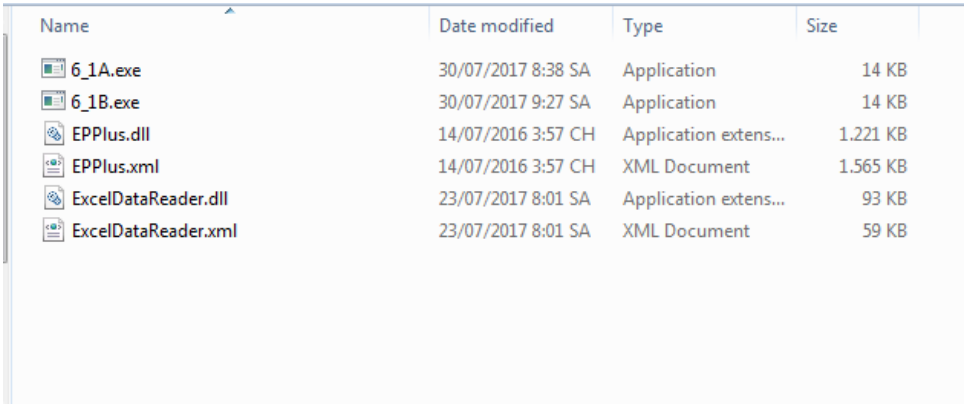
BÁO CÁO BÀI TẬP NHÓM THỰC HÀNH 1A

Nhóm 6

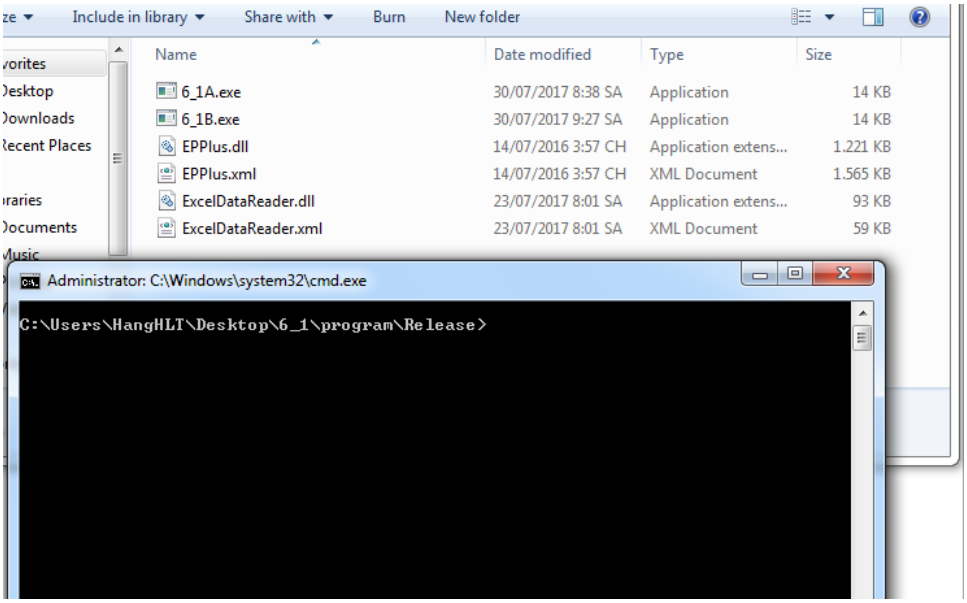
MSSV	Tên - Công việc	Đóng góp
1642004	Trần Chí Bảo - Câu 1a	25%
1642008	Trận Lệ Diễm Châu - Câu 1f	25%
1642035	Nguyễn Thành Lai - Câu 1d,e	25%
1642088	Bùi Thanh Vân - Câu 1b,c	25%

Cách sử dụng:

Muốn sử dụng cần phải mở thư mục chứa phần mềm :



Mở commandline vào thư mục hiện tại chứa file thực thi exe:



Cú pháp sử dụng các chức năng như sau:

Trước tiên để thống nhất để dễ sử dụng các loại chỉ thị

[đường dẫn tới file input]: Là đường dẫn tới file input:

Vd: D:\input.xlsx
-> Nên là được dẫn tuyệt đối như trên

-> Nếu file ngay cạnh thư mục thực hi thì chỉ cần tên file: vd: input.xlsx

[đường dẫn tới file output]: Là đường dẫn muốn xuất ra kèm tên của nó

Vd: D:\output.xlsx

-> Nên là đường dẫn tuyệt đối như trên

-> Nếu file ngay cạnh thư mục thực hi thì chỉ cần tên file: vd: input.xlsx

[danh_sách_các_field]: Là danh sách các field muốn áp dụng, vd như trong bảng có các trường tensv, diem, namsinh...

Thì kết quả khi sử dụng trong cmd là field1,field2,field3

Vd: tensv,diem,namsinh

[loại_rời_rạc]: Loại rời rạc hóa muốn áp dụng có 2 loại là theo chiều rộng và theo độ sâu lần lượt là các giá trị khi sử dụng cmd là: width, height

Vd: width

->Cách nhau bằng dấu phẩy không có khoảng trắng dư thừa

[loại_chuẩn hóa]: Loại chuẩn hóa muốn áp dụng có 2 loại là min-max và z-score ần lượt mang các giá trị khi sử dụng vào cmd là: min-max z-score

Vd: z-score

[bin]: bin của giỏ là số nguyên

Vd: 5

1. Để xóa các field trong file excel

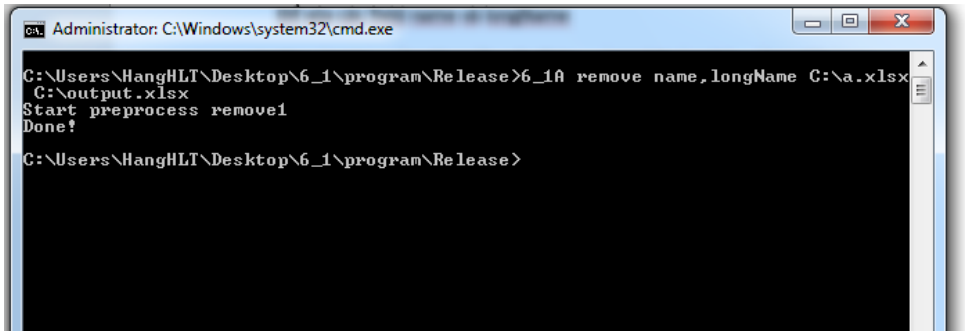
6_1A remove [danh_sách_các_field] [đường dẫn tới file input] [đường dẫn tới file output]

Vd:

Để xóa các field name và longName

country	name	longName	foundingDate	population	capital	largestCity	area
14	Abkhazia	Republic o	33044	242862	Sukhumi		5381,646
15	Abkhazia	Republic o	33110	242862	Sukhumi		5381,646
16	Abkhazia	Republic o	39686	242862	Sukhumi		5381,646
17	Abyei	Abyei Area	38361		Abyei (town)		6553,249
18	Abyei	Abyei Area	39691		Abyei (town)		6553,249
31	Adélie Lan	Adélie Land					432000
36	Adjara	Autonomous Republic of Ajar		393700	Batumi		1802,465
38	Aerica	Aerican En	31905		Montreal		9000000
41	Afghanista	Islamic Rep	7171	32564342	Kabul	Kabul	405276,3
42	Afghanista	font-size:8	7171	32564342	Kabul	Kabul	405276,3
44			33392	1,05E+09		Nigeria	29865860
45			33392	1,05E+09		Lagos	29865860
63	Akrotiri an	Sovereign Base Areas		7700	Episkopi Cantonment		157,7157
64	Akrotiri an	of		7700	Episkopi Cantonment		157,7157
65	Akrotiri an	Akrotiri and Dhekelia		7700	Episkopi Cantonment		157,7157

Chạy với câu lệnh sau:



Kết quả:

country	foundingD	population	capital	largestCity	area
14	33044	242862	Sukhumi		5381,646
15	33110	242862	Sukhumi		5381,646
16	39686	242862	Sukhumi		5381,646
17	38361		Abyei (town)		6553,249
18	39691		Abyei (town)		6553,249
31					432000
36		393700	Batumi		1802,465
38	31905		Montreal		9000000
41	7171	32564342	Kabul	Kabul	405276,3
42	7171	32564342	Kabul	Kabul	405276,3
44	33392	1,05E+09		Nigeria	29865860
45	33392	1,05E+09		Lagos	29865860
63		7700	Episkopi Cantonment		157,7157

1. Chuẩn hóa các field tronng file excel

6_1A standardized [loại_chuẩn hóa] [danh_sách_các_field] [đường dẫn tới file input] [đường dẫn tới file output]

Vd: Để chuẩn hóa cột population và area theo chuẩn hóa min-max

Gõ lệnh trên cmd:

6_1A standardized min-max population,area C:\a.xlsx C:\output.xlsx

population	capital	largestCity	area
0,000104	Sukhumi		0,000107
0,000104	Sukhumi		0,000107
0,000104	Sukhumi		0,000107
	Abyei (town)		0,000131
	Abyei (town)		0,000131
			0,008619
0,000169	Batumi		3,59E-05
	Montreal		0,17957
0,013988	Kabul	Kabul	0,008086
0,013988	Kabul	Kabul	0,008086
0,452378		Nigeria	0,595889
0,452378		Lagos	0,595889
3,31E-06	Episkopi Cantonment		3,13E-06
3,31E-06	Episkopi Cantonment		3,13E-06
3,31E-06	Episkopi Cantonment		3,13E-06
1,23E-05	Mariehamn		1,96E-05
0,001243	Tirana	Tirana	0,000356
0,001243	Tirana	Tirana	0,000356

2. Rời rạc hóa dữ liệu trong file excel

6_1A binning [loại_rời_rạc] [danh_sách_các_field] [bin] [đường dẫn tới file input] [đường dẫn tới file output]

VD: Để rời rạc hóa dữ liệu 2 trường population và area theo bin = 5000 theo rộng ta gõ lệnh:

6_1A binning width population 5000 C:\a.xlsx C:\output.xlsx

3. Xóa các record mà field được chọn thiếu dữ liệu thiếu dữ liệu

6_1A removeMissingInstance [danh_sách_các_field] [đường dẫn tới file input] [đường dẫn tới file output]

Vd: Để kiểm tra các record mà field population và area bị thiếu thì xóa record đó đi sử dụng lệnh:

6_1A removeMissingInstance population,area C:\a.xlsx C:\output.xlsx

Lưu ý chương trình chỉ sử dụng được khi dòng đầu tiên chứa các field name mà field name đó phải có giá trị là chuỗi không dấu và không cách nhau bằng khoảng trắng(hợp lệ: population, ngay_sinh), và các dữ liệu trong cùng 1 cột (ngoài field name tên cột) đó phải là cùng kiểu dữ liệu...