

碩管一甲 M09218001 周彥廷
數據分析與應用-作業 1：dplyr 特徵擷取實作

安裝 dplyr 套件與載入 dplyr

```
install.packages("dplyr")
```

```
library(dplyr)
```

```
# load dataset1.csv
```

載入 data.csv 與清洗資料將空格刪除

```
data<-read.csv("C:/Users/MCUT/Desktop/dataset1.csv",header=TRUE,  
sep = ",",na.strings="")
```

```
data1<-tibble::as_tibble(data)
```

or

```
data1<-tbl_df(data)
```

```
> class(data1)  
[1] "tbl_df"      "tbl"         "data.frame"
```

```
# start coding to 空白轉換 'NA' and use the complete.cases remove the 'NA'
```

```
newdata<-data1[complete.cases(data1),]
```

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2016/9/2	2016/10/19	N70707	F8157711	125040	191360	0	2
2	2016/5/12	2016/5/31	N21223	F8218210	179997	165000	0	2
3	2016/11/25	2017/1/23	N99991	F2103140	157199	157000	0	1
4	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1
5	2017/2/14	2017/6/27	N97971	F9102122	92336	119000	0	2
6	2015/8/12	2015/12/18	N82828	F4971164	175500	175500	0	1
7	2015/12/4	2015/12/30	N70707	F3135120	143000	143000	0	2
8	2016/2/15	2016/3/3	N83831	F6479685	142825	134400	0	3
9	2016/2/17	2016/5/24	N79801	F8131110	127151	127150	0	1
10	2016/7/15	2016/8/8	N71717	F3135120	136973	136946	0	3

篩選

```
1."辦事員"=N70707
```

```
f_1<-dplyr::filter(newdata,辦事員=='N70707')
```

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2016/9/2	2016/10/19	N70707	F8157711	125040	191360	0	2
2	2015/12/4	2015/12/30	N70707	F3135120	143000	143000	0	2
3	2015/12/14	2016/1/19	N70707	F1915213	117515	168677	0	3
4	2015/12/16	2016/1/19	N70707	F2933162	51490	110000	8612	5
5	2016/3/7	2016/4/1	N70707	F3657512	180500	180500	0	1

2."決包金額"介於 100000~130000

f_2<-dplyr::filter(newdata,決包金額>100000,決包金額<130000)

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2017/2/14	2017/6/27	N97971	F9102122	92336	119000	0	2
2	2016/2/17	2016/5/24	N79801	F8131110	127151	127150	0	1
3	2017/3/31	2017/5/2	N83831	F2117120	103500	101000	0	2
4	2016/2/15	2016/4/26	N13132	F1454010	116571	116000	0	1
5	2016/11/25	2016/12/23	N21223	F1464976	115500	111500	0	1
6	2017/4/5	2017/5/11	N65656	F9171150	116572	116572	0	1

3."決包金額">180000 且 "報價廠商家數" = 1

f_3<-dplyr::filter(newdata,決包金額>180000&報價廠商家數==1)

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2016/12/13	2017/1/13	N99991	F3130116	187999	187000	0	1
2	2016/5/30	2016/7/15	N13132	F6118571	185452	185000	0	1
3	2016/7/19	2016/8/22	N22222	F1115100	202140	200000	0	1
4	2017/4/25	2017/5/24	N22238	F1975172	190842	190000	0	1
5	2016/2/24	2016/5/25	N99991	F7751367	191472	191472	0	1
6	2016/7/12	2016/8/5	N88938	F2981011	190686	188000	227000	1

4."決包金額">180000 或 "報價廠商家數" = 1

f_4<-dplyr::filter(newdata,決包金額>180000|報價廠商家數==1)

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2016/9/2	2016/10/19	N70707	F8157711	125040	191360	0	2
2	2016/11/25	2017/1/23	N99991	F2103140	157199	157000	0	1
3	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1
4	2015/8/12	2015/12/18	N82828	F4971164	175500	175500	0	1
5	2016/2/17	2016/5/24	N79801	F8131110	127151	127150	0	1

5."辦事員"開頭為 N1 者

f_6<-dplyr::filter(newdata,grepl(pattern = 'N1',x=辦事員,fixed=TRUE))

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1
2	2016/2/15	2016/4/26	N13132	F1454010	116571	116000	0	1
3	2016/5/30	2016/7/15	N13132	F6118571	185452	185000	0	1
4	2017/5/25	2017/6/29	N14157	F3947911	148175	136800	0	2
5	2017/6/23	2017/7/28	N14157	F8998995	129161	120000	0	1
6	2016/12/28	2017/2/13	N14157	F2100545	147500	147500	0	1
7	2017/1/24	2017/3/22	N14157	F4637947	151321	151000	0	1

排序

1.按照"呈核日"排序(由小而大排)

a_1<-dplyr::arrange(newdata,呈核日)

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2015/9/30	2015/10/16	N39393	F2159102	136417	136417	-27585	1
2	2015/10/14	2015/11/10	N65656	F1831429	115000	115000	0	3
3	2015/10/14	2015/11/18	N64647	F7161110	160493	160000	0	1
4	2015/10/13	2015/11/23	N66666	F5606431	157116	126000	-101926	1
5	2015/10/23	2015/11/25	N99991	F7611449	172938	172938	0	1
6	2015/10/29	2015/11/26	N97979	F8175130	116675	116500	0	1
7	2015/9/21	2015/12/16	N18197	F4577144	139955	170000	0	4

2.按照"辦事員"、"報價廠商家數"排序(由小而大排)

a_2<-dplyr::arrange(newdata,辦事員,報價廠商家數)

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2015/7/31	2015/12/23	N11183	F3107529	89079	198000	-47198	4
2	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1
3	2016/2/15	2016/4/26	N13132	F1454010	116571	116000	0	1
4	2016/5/30	2016/7/15	N13132	F6118571	185452	185000	0	1
5	2017/6/15	2017/6/21	N13132	F2162110	179489	169431	0	1
6	2016/6/22	2016/7/15	N13132	F1143130	140000	110000	0	1
7	2016/2/17	2016/3/29	N13132	F2136100	180111	180000	0	1
8	2016/6/23	2016/7/5	N13132	F1115666	153007	153000	210317	1

篩選後排序

1.選出"辦事員"開頭為 N1 者再依"報價廠商家數"排序

a_3<-dplyr::arrange(f_6,報價廠商家數)

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數
1	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1
2	2016/2/15	2016/4/26	N13132	F1454010	116571	116000	0	1
3	2016/5/30	2016/7/15	N13132	F6118571	185452	185000	0	1
4	2017/6/23	2017/7/28	N14157	F8998995	129161	120000	0	1
5	2016/12/28	2017/2/13	N14157	F2100545	147500	147500	0	1

增加特徵

1.作業天數<-dplyr::mutate(newdata,作業天數=as.Date(呈核日,"%Y/%m/%d")-as.Date(收件日,"%Y/%m/%d"))

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數	作業天數
1	2016/9/2	2016/10/19	N70707	F8157711	125040	191360	0	2	47
2	2016/5/12	2016/5/31	N21223	F8218210	179997	165000	0	2	19
3	2016/11/25	2017/1/23	N99991	F2103140	157199	157000	0	1	59
4	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1	64
5	2017/2/14	2017/6/27	N97971	F9102122	92336	119000	0	2	133
6	2015/8/12	2015/12/18	N82828	F4971164	175500	175500	0	1	128
7	2015/12/4	2015/12/30	N70707	F3135120	143000	143000	0	2	26

2. 追加比例 = (決包金額+追加金額)/決包金額

追加比例<-dplyr::mutate(作業天數,追加比例=(決包金額+追加金額)/決包金額)

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數	作業天數	追加比例
1	2016/9/2	2016/10/19	N70707	F8157711	125040	191360	0	2	47	1.00000000
2	2016/5/12	2016/5/31	N21223	F8218210	179997	165000	0	2	19	1.00000000
3	2016/11/25	2017/1/23	N99991	F2103140	157199	157000	0	1	59	1.00000000
4	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1	64	1.00000000
5	2017/2/14	2017/6/27	N97971	F9102122	92336	119000	0	2	133	1.00000000
6	2015/8/12	2015/12/18	N82828	F4971164	175500	175500	0	1	128	1.00000000

3. 超預算：若決包金額>預算金額，則為超預算，將其標示為 1、反之為 0

超預算<-dplyr::mutate(追加比例,超預算=ifelse(決包金額>預算金額,1,0))

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數	作業天數	追加比例	超預算
1	2016/9/2	2016/10/19	N70707	F8157711	125040	191360	0	2	47	1.00000000	1
2	2016/5/12	2016/5/31	N21223	F8218210	179997	165000	0	2	19	1.00000000	0
3	2016/11/25	2017/1/23	N99991	F2103140	157199	157000	0	1	59	1.00000000	0
4	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1	64	1.00000000	0
5	2017/2/14	2017/6/27	N97971	F9102122	92336	119000	0	2	133	1.00000000	1
6	2015/8/12	2015/12/18	N82828	F4971164	175500	175500	0	1	128	1.00000000	0

4. 獨家報價：若報價廠商數 = 1，則為獨家，將其標示為 1、反之為 0

獨家報價<-dplyr::mutate(超預算,獨家報價=ifelse(報價廠商家數==1,1,0))

	收件日	呈核日	辦事員	施工廠商	預算金額	決包金額	追加金額	報價廠商家數	作業天數	追加比例	超預算	獨家報價
1	2016/9/2	2016/10/19	N70707	F8157711	125040	191360	0	2	47	1.00000000	1	0
2	2016/5/12	2016/5/31	N21223	F8218210	179997	165000	0	2	19	1.00000000	0	0
3	2016/11/25	2017/1/23	N99991	F2103140	157199	157000	0	1	59	1.00000000	0	1
4	2015/10/22	2015/12/25	N13132	F8131110	179000	179000	0	1	64	1.00000000	0	1
5	2017/2/14	2017/6/27	N97971	F9102122	92336	119000	0	2	133	1.00000000	1	0
6	2015/8/12	2015/12/18	N82828	F4971164	175500	175500	0	1	128	1.00000000	0	1

●計算每位"辦事員"的

決包金額最小值

決包金額最大值

平均作業天數

平均追加比例

超預算件數

獨家報價件數

先剔除呈核日、收件日的遺漏值列，將滿足"平均作業天數">15、"超預算件數">3、"獨家報價件數">3"辦事員"的明細列出來

#增加特徵<-將作業天數、追加比例、超預算與獨家報價

增加特徵<-dplyr::mutate(newdata,作業天數=as.Date(呈核日,"%Y/%m/%d")-
as.Date(收件日,"%Y/%m/%d"),追加比例=(決包金額+追加金額)/決包金額,超預
算=ifelse(決包金額>預算金額,1,0),獨家報價=ifelse(報價廠商家數==1,1,0))

計算每位"辦事員"的

明細表 1<-dplyr::group_by(增加特徵 ,辦事員) %>%

dplyr::summarise(決包金額最小值=min(決包金額),決包金額最大值=max(決包
金額),平均作業天數=mean(作業天數,na.rm=TRUE), 平均追加比例=mean(追加比
例,na.rm=TRUE),超預算件數=n(),獨家報價件數=n())%>% dplyr::filter(平均作業
天數>15,超預算件數>3,獨家報價件數>3)

	辦事員	決包金額最小值	決包金額最大值	平均作業天數	平均追加比例	超預算件數	獨家報價件數
1	N13132	102000	200000	39.9050279329609	1.0328174	179	179
2	N13145	101000	200000	37.2753623188406	1.0193436	69	69
3	N14157	102000	200000	38.4033613445378	0.9962371	119	119
4	N18197	102000	200000	42.8440366972477	1.0026354	109	109
5	N21223	100500	200000	25.4255319148936	1.0103008	141	141
6	N22222	105000	200000	30.2337662337662	1.0095463	77	77
7	N22238	100950	200000	30.043956043956	1.1009263	91	91
8	N25251	100350	200000	44.3260869565217	1.0158316	138	138
9	N33128	102864	199000	25.1578947368421	1.0000000	38	38
10	N39393	100485	199992	40.763358778626	1.0003862	131	131


```
# 計算每位"辦事員"的 使用 sum()
明細表 2<-dplyr::group_by(增加特徵,辦事員) %>%
dplyr::summarise(決包金額最小值=min(決包金額),決包金額最大值=max(決包金
額),平均作業天數=mean(作業天數,na.rm=TRUE), 平均追加比例=mean(追加比
例,na.rm=TRUE),超預算件數=sum(超預算),獨家報價件數=sum(獨家報價)) %>%
dplyr::filter(平均作業天數>15,超預算件數>3,獨家報價件數>3)
```

	辦事員	決包金額最小值	決包金額最大值	平均作業天數	平均追加比例	超預算件數	獨家報價件數
1	N13132	102000	200000	39.9050279329609	1.0328174	35	75
2	N13145	101000	200000	37.2753623188406	1.0193436	22	23
3	N14157	102000	200000	38.4033613445378	0.9962371	24	49
4	N18197	102000	200000	42.8440366972477	1.0026354	42	50
5	N21223	100500	200000	25.4255319148936	1.0103008	31	70
6	N22222	105000	200000	30.2337662337662	1.0095463	19	30
7	N22238	100950	200000	30.043956043956	1.1009263	14	46
8	N25251	100350	200000	44.3260869565217	1.0158316	26	63
9	N33128	102864	199000	25.1578947368421	1.0000000	4	27
10	N39393	100485	199992	40.763358778626	1.0003862	19	58

●計算每位"辦事員"與"施工廠商"的

決包金額最小值

決包金額最大值

平均作業天數

平均追加比例

超預算件數

獨家報價件數

先剔除呈核日、收件日的遺漏值列，將滿足"平均作業天數">30、"平均追加比例">0.2、"獨家報價件數">3"辦事員"與"施工廠商"的明細列出來

計算每位"辦事員"與"施工廠商"的

```
明細表 2<-dplyr::group_by(增加特徵,辦事員,施工廠商) %>%
```

```
dplyr::summarise(決包金額最小值=min(決包金額),決包金額最大值=max(決包金
額),平均作業天數=mean(作業天數,na.rm=TRUE), 平均追加比例=mean(追加比
例,na.rm=TRUE),超預算件數=n(),獨家報價件數=n())%>% dplyr::filter(平均作業
天數>30,平均追加比例>0.2,獨家報價件數>3)
```

	辦事員	施工廠商	決包金額最小值	決包金額最大值	平均作業天數	平均追加比例	超預算件數	獨家報價件數
1	N13145	F1173150	105000	200000	56.6	0.8210259	5	5
2	N18197	F1109120	112000	179250	50.75	1.0000000	4	4
3	N18197	F4577144	106706	187790	61.8	0.9921803	5	5
4	N18197	F5568110	137000	194619	32.6666666666667	1.0000000	6	6
5	N18197	F8100140	108760	185000	48	1.0000000	4	4
6	N21223	F2911151	119000	189000	31	1.0000000	4	4
7	N22238	F3107529	108000	198000	35.1666666666667	1.0598746	6	6
8	N39393	F2653762	130000	190000	48.25	0.9173389	4	4
9	N39393	F8139581	101620	161950	36.75	1.0336237	4	4

計算每位"辦事員"與"施工廠商"的 使用 sum()

對應明細表 4<-dplyr::group_by(增加特徵,辦事員,施工廠商) %>%

dplyr::summarise(決包金額最小值=min(決包金額),決包金額最大值=max(決包金額),平均作業天數=mean(作業天數,na.rm=TRUE), 平均追加比例=mean(追加比例,na.rm=TRUE),超預算件數=sum(超預算),獨家報價件數=sum(獨家報價))%>%

dplyr::filter(平均作業天數>30,平均追加比例>0.2,獨家報價件數>3)

	辦事員	施工廠商	決包金額最小值	決包金額最大值	平均作業天數	平均追加比例	超預算件數	獨家報價件數
1	N18197	F8100140	108760	185000	48	1.0000000	1	4
2	N39393	F8139581	101620	161950	36.75	1.0336237	1	4
3	N69701	F1101120	102000	174000	55.3333333333333	0.9863848	5	4
4	N88938	F1103961	101000	160700	50	1.0000000	2	6
5	N88938	F1123100	108000	190000	42.1428571428571	1.0000000	1	6
6	N96971	F1103961	115617	189000	50.75	1.0000000	1	4
7	N96971	F1123100	122000	170000	38	1.0000000	0	4
8	N96971	F2127120	124000	155000	71.75	1.0000000	0	4