

Fait par : EL HANAFI Maha

Compte Rendu

Lab 7 : Introduction ro Data Analysis -Assignment 2

Dataset « ontime_february»

X.1	AIRLINE_ID	FL_NUM	ORIGIN_AIRPORT_ID	DEST_AIRPORT_ID	DEP_DELAY_NEW	ARR_DELAY_NEW	CANCELLED	CANCELLATION_CODE	CARRIER_DELAY	WEATHER_DELAY	NAS_DELAY	SECURITY_DELAY	LATE_AIRCRAFT_DELAY
1	1	19393	2938	10257	11697	0	0	0		N/A	N/A	N/A	N/A
2	2	19393	26	10257	13204	0	0	0		N/A	N/A	N/A	N/A
3	3	19393	30	10257	13204	2	0	0		N/A	N/A	N/A	N/A
4	4	19393	2218	10257	15304	6	0	0		N/A	N/A	N/A	N/A
5	5	19393	495	10397	11697	N/A	N/A	1	B	N/A	N/A	N/A	N/A
6	6	19393	582	10397	11697	33	13	0		N/A	N/A	N/A	N/A
7	7	19393	651	10397	11697	22	6	0		N/A	N/A	N/A	N/A
8	8	19393	899	10397	11697	0	0	0		N/A	N/A	N/A	N/A
9	9	19393	2243	10397	11697	6	0	0		N/A	N/A	N/A	N/A
10	10	19393	203	10397	12451	0	0	0		N/A	N/A	N/A	N/A
11	11	19393	1229	10397	12451	0	0	0		N/A	N/A	N/A	N/A
12	12	19393	1579	10397	12451	8	0	0		N/A	N/A	N/A	N/A
13	13	19393	24	10397	13204	49	36	0		36	0	0	0
14	14	19393	102	10397	13204	0	0	0		N/A	N/A	N/A	N/A
15	15	19393	494	10397	13204	44	33	0		26	0	0	0
16	16	19393	2045	10397	13204	35	29	0		29	0	0	0
17	17	19393	2591	10397	13204	0	0	0		N/A	N/A	N/A	N/A
18	18	19393	4225	10397	13204	0	0	0		N/A	N/A	N/A	N/A
19	19	19393	4747	10397	13204	8	0	0		N/A	N/A	N/A	N/A
20	20	19393	2216	10397	14027	48	27	0		18	0	0	0
21	21	19393	4233	10397	14027	0	0	0		N/A	N/A	N/A	N/A
22	22	19393	458	10397	14635	28	12	0		N/A	N/A	N/A	N/A
23	23	19393	4403	10397	14635	17	2	0		N/A	N/A	N/A	N/A
24	24	19393	4471	10397	14635	5	0	0		N/A	N/A	N/A	N/A
25	25	19393	362	10397	15304	3	0	0		N/A	N/A	N/A	N/A

```

> air1=unique(ontime_february$AIRLINE_ID)
> nbr=c()
> nbr_c=c()
> nbr_d=c()
>
> for(co in air1)
+ {
+   s=subset(ontime_february, ontime_february$AIRLINE_ID==co)
+   #nombre d'annulation
+   nbr_c=c(nbr_c, sum(s$CANCELLED==1,na.rm=TRUE))
+   nbr_d=c(nbr_d, sum(s$ARR_DELAY_NEW>0,na.rm=TRUE))
+   nbr=c(nbr,nrow(s))
+ }

> airline=data.frame(Airline=air1,nbr=nbr,nbr_cancelled=nbr_c, nbr_deployed=nbr_d)
> airline
  Airline  nbr nbr_cancelled nbr_deployed
1  19393 16246           506          5833
2  19790 12601           307          4708
3  19805 11176           397          4531
4  19930   224             0             71
5  19977 5845           116          2393
6  20304   14             0             7
7  20355 5844           180          2412
8  20366 1239           29           534
9  20398 1972           145           805
10 20409 9705           514          5307
11 20416 3792           97          1919
12 20436 2019           73          1283

>
> air <- table(ontime_february$AIRLINE_ID)
> air
19393 19790 19805 19930 19977 20304 20355 20366 20398 20409 20416 20436
16246 12601 11176   224  5845   14  5844  1239  1972  9705  3792  2019

```

```
> #Q2
> airline$prc_c=airline$nbr_cancelled/airline$nbr
> airline$prc_d=airline$nbr_deployed/airline$nbr
> |
```

Data

airline 12 obs. of 6 variables

```
> #Q3
> index=which.min(airline$prc_d)
> x=airline$Airline[index]
>
> view(airline)
> >
```

	Airline	nbr	nbr_cancelled	nbr_deployed	prc_c	prc_d
1	19393	16246	506	5833	0.03114613	0.3590422
2	19790	12601	307	4708	0.02436315	0.3736211
3	19805	11176	397	4531	0.03552255	0.4054223
4	19930	224	0	71	0.00000000	0.3169643
5	19977	5845	116	2393	0.01984602	0.4094098
6	20304	14	0	7	0.00000000	0.5000000
7	20355	5844	180	2412	0.03080082	0.4127310
8	20366	1239	29	534	0.02340597	0.4309927
9	20398	1972	145	805	0.07352941	0.4082150
10	20409	9705	514	5307	0.05296239	0.5468315
11	20416	3792	97	1919	0.02558017	0.5060654
12	20436	2019	73	1283	0.03615651	0.6354631

```
> #Q4
> min(airline$prc_d)*(airline$nbr)
[1] 5149.4018 3994.0670 3542.3929 71.0000 1852.6562 4.4375 1852.3393 392.7188 625.0536 3076.1384 1201.9286 639.9509
> (1-min(airline$prc_d))*(airline$nbr)
[1] 11096.5982 8606.9330 7633.6071 153.0000 3992.3438 9.5625 3991.6607 846.2812 1346.9464 6628.8616 2590.0714 1379.0491
> |
```