

Hotelbeds

Carlos A. García

April 14, 2019

Primer establim el directori de feina

```
setwd("D:/Users/cagarcia/uoc/M2.954 - Estadística avançada/hotelbeds")
```

Llegim el fitxer proporcionat a la pràctica

```
bookings <- read.csv2("dataTT00.csv", header = TRUE, sep = ",", dec = ".")
attach(bookings)
colnames(bookings)[colnames(bookings)=="i..TT00"] <- "TT00"

summary(bookings)
```

```
##          TT00          CREATED_SUNDAY    CREATED_SUNDAY_CANCEL_1_HOUR
## Min.      : 9      Min.      : 0.000    Min.      : 0.000
## 1st Qu.: 48011    1st Qu.: 0.000    1st Qu.: 0.000
## Median :141182    Median : 0.000    Median : 0.000
## Mean   :152476    Mean   : 2.096    Mean   : 0.125
## 3rd Qu.:259156    3rd Qu.: 0.000    3rd Qu.: 0.000
## Max.   :340037    Max.   :1507.000    Max.   :290.000
## CREATED_SUNDAY_CANCEL_24_HOUR    CREATED_MONDAY
## Min.      : 0.0000              Min.      : 0.000
## 1st Qu.: 0.0000              1st Qu.: 0.000
## Median : 0.0000              Median : 0.000
## Mean   : 0.1829              Mean   : 4.203
## 3rd Qu.: 0.0000              3rd Qu.: 1.000
## Max.   :313.0000              Max.   :1990.000
## CREATED_MONDAY_CANCEL_1_HOUR    CREATED_MONDAY_CANCEL_24_HOUR
## Min.      : 0.0000              Min.      : 0.0000
## 1st Qu.: 0.0000              1st Qu.: 0.0000
## Median : 0.0000              Median : 0.0000
## Mean   : 0.2686              Mean   : 0.4384
## 3rd Qu.: 0.0000              3rd Qu.: 0.0000
## Max.   :437.0000              Max.   :487.0000
## CREATED_TUESDAY    CREATED_TUESDAY_CANCEL_1_HOUR
## Min.      : 0.000    Min.      : 0.0000
## 1st Qu.: 0.000    1st Qu.: 0.0000
## Median : 0.000    Median : 0.0000
## Mean   : 4.385    Mean   : 0.2706
## 3rd Qu.: 2.000    3rd Qu.: 0.0000
## Max.   :1929.000    Max.   :426.0000
## CREATED_TUSDAY_CANCEL_24_HOUR    CREATED_WEDNESDAY
## Min.      : 0.0000              Min.      : 0.000
## 1st Qu.: 0.0000              1st Qu.: 0.000
## Median : 0.0000              Median : 0.000
## Mean   : 0.4404              Mean   : 4.345
## 3rd Qu.: 0.0000              3rd Qu.: 2.000
## Max.   :460.0000              Max.   :1735.000
## CREATED_WEDNES_CANCEL_1_HOUR    CREATED_WEDNES_CANCEL_24_HOUR
```

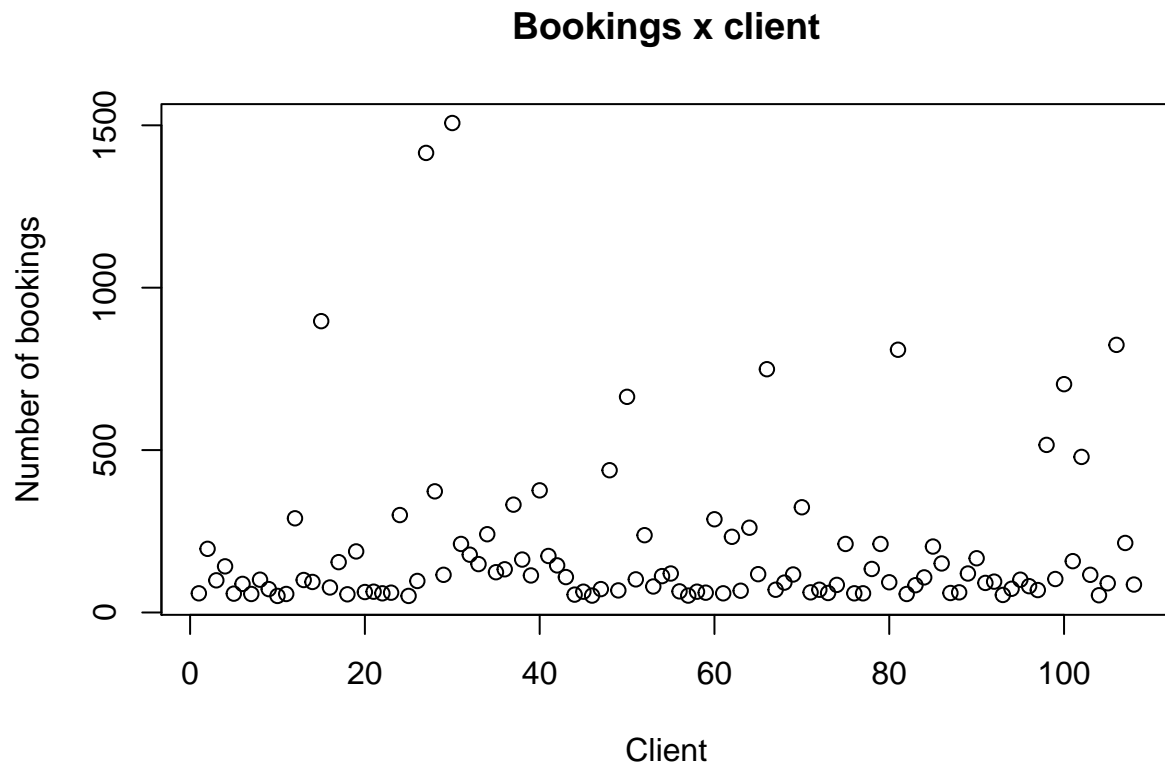
```
## Min. : 0.0000 Min. : 0.0000
## 1st Qu.: 0.0000 1st Qu.: 0.0000
## Median : 0.0000 Median : 0.0000
## Mean : 0.2791 Mean : 0.4498
## 3rd Qu.: 0.0000 3rd Qu.: 0.0000
## Max. :421.0000 Max. :457.0000
## CREATED_THURSDAY CREATED_THURSDAY_CANCEL_1_HOUR
## Min. : 0.000 Min. : 0.0000
## 1st Qu.: 0.000 1st Qu.: 0.0000
## Median : 0.000 Median : 0.0000
## Mean : 4.231 Mean : 0.2538
## 3rd Qu.: 2.000 3rd Qu.: 0.0000
## Max. :1697.000 Max. :408.0000
## CREATED_THURS_CANCEL_24_HOUR CREATED_FRIDAY
## Min. : 0.000 Min. : 0.000
## 1st Qu.: 0.000 1st Qu.: 0.000
## Median : 0.000 Median : 0.000
## Mean : 0.415 Mean : 4.128
## 3rd Qu.: 0.000 3rd Qu.: 1.000
## Max. :449.000 Max. :1700.000
## CREATED_FRIDAY_CANCEL_1_HOUR CREATED_FRIDAY_CANCEL_24_HOUR
## Min. : 0.0000 Min. : 0.0000
## 1st Qu.: 0.0000 1st Qu.: 0.0000
## Median : 0.0000 Median : 0.0000
## Mean : 0.2421 Mean : 0.3675
## 3rd Qu.: 0.0000 3rd Qu.: 0.0000
## Max. :354.0000 Max. :383.0000
## CREATED_SATURDAY CREATED_SATURDAY_CANCEL_1_HOUR
## Min. : 0.000 Min. : 0.0000
## 1st Qu.: 0.000 1st Qu.: 0.0000
## Median : 0.000 Median : 0.0000
## Mean : 2.619 Mean : 0.1659
## 3rd Qu.: 1.000 3rd Qu.: 0.0000
## Max. :1408.000 Max. :325.0000
## CREATED_SATUR_CANCEL_24_HOUR
## Min. : 0.0000
## 1st Qu.: 0.0000
## Median : 0.0000
## Mean : 0.2228
## 3rd Qu.: 0.0000
## Max. :356.0000
```

```
created_sunday_nn <- bookings [c("TT00", "CREATED_SUNDAY", "CREATED_SUNDAY_CANCEL_1_HOUR", "CREATED_SUNDAY_CANCEL_24_HOUR")]
created_sunday_nn <- created_sunday_nn[created_sunday_nn$CREATED_SUNDAY > 50,]
summary(created_sunday_nn )
```

```
## TT00 CREATED_SUNDAY CREATED_SUNDAY_CANCEL_1_HOUR
## Min. : 417 Min. : 51.00 Min. : 0.0
## 1st Qu.: 74085 1st Qu.: 64.75 1st Qu.: 1.0
## Median :217083 Median : 101.50 Median : 3.0
## Mean :181410 Mean : 193.03 Mean : 12.7
## 3rd Qu.:263787 3rd Qu.: 197.75 3rd Qu.: 7.0
## Max. :339416 Max. :1507.00 Max. :290.0
## CREATED_SUNDAY_CANCEL_24_HOUR
## Min. : 0.00
```

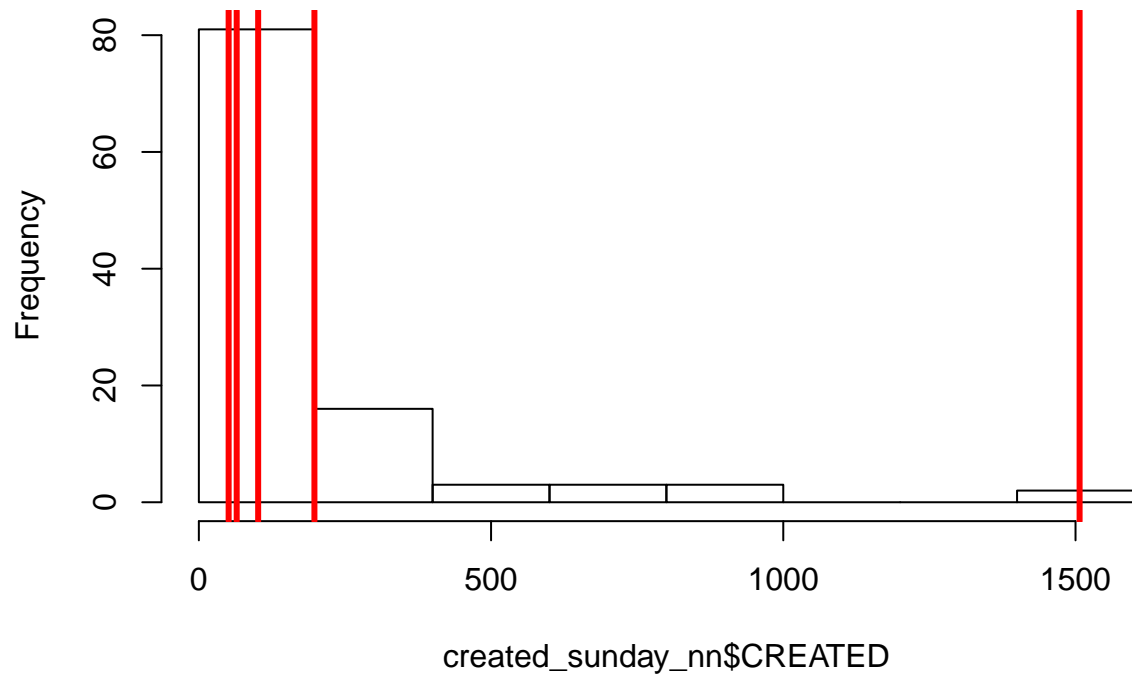
```
## 1st Qu.: 2.00
## Median : 6.50
## Mean   : 18.12
## 3rd Qu.: 13.50
## Max.   :313.00
```

```
colnames(created_sunday_nn) <- c("TT00", "CREATED", "C1", "C24")
plot(created_sunday_nn$CREATED, col="black", main="Bookings x client",
      xlab="Client", ylab="Number of bookings")
```

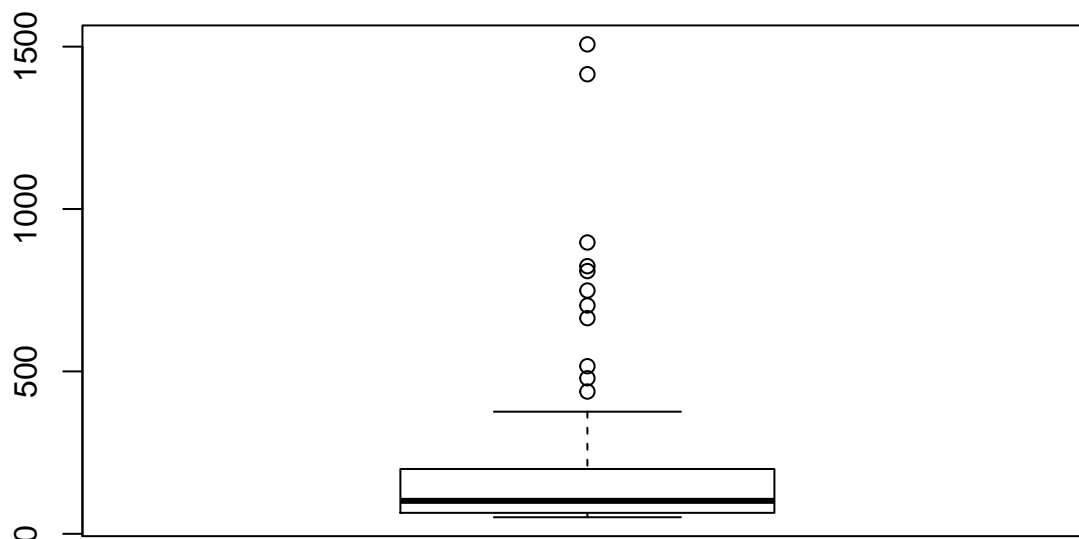


```
hist(created_sunday_nn$CREATED)
abline(v=quantile(created_sunday_nn$CREATED), col="red", lwd=3)
```

Histogram of created_sunday_nn\$CREATED



```
boxplot(created_sunday_nn$CREATED)
boxplot(created_sunday_nn$CREATED)$out
```



```
## [1] 897 1415 1507 438 664 749 809 516 703 479 824
```

```
shapiro.test(created_sunday_nn$CREATED)
```

```
##
## Shapiro-Wilk normality test
##
## data: created_sunday_nn$CREATED
## W = 0.56869, p-value = 2.698e-16
```

```
cor(created_sunday_nn[, c("CREATED", "C1", "C24")])
```

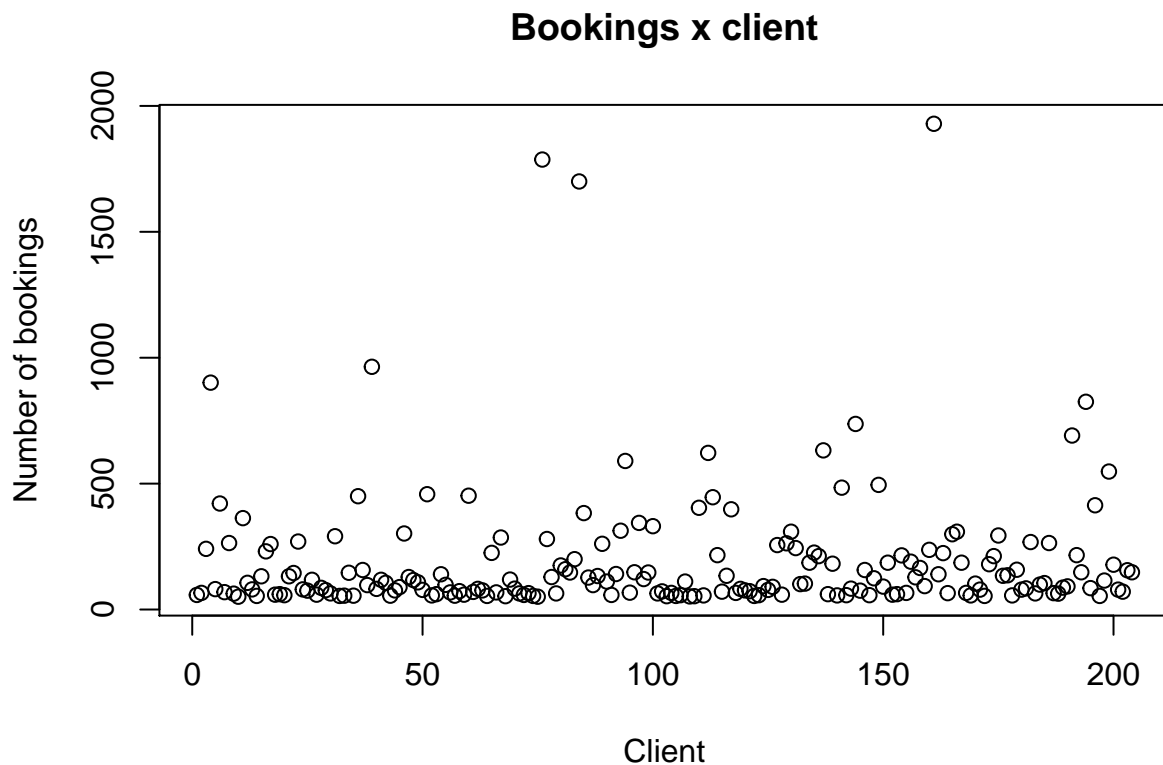
```
##          CREATED          C1          C24
## CREATED 1.0000000 0.5439297 0.6620170
## C1      0.5439297 1.0000000 0.9834777
## C24     0.6620170 0.9834777 1.0000000
```

```
created_tuesday_nn <- bookings [c("TTOO", "CREATED_TUESDAY", "CREATED_TUESDAY_CANCEL_1_HOUR", "CREATED_TUESDAY"),]
created_tuesday_nn <- created_tuesday_nn[created_tuesday_nn$CREATED_TUESDAY > 50,]
summary(created_tuesday_nn )
```

```
##      TTOO      CREATED_TUESDAY  CREATED_TUESDAY_CANCEL_1_HOUR
## Min.   : 322   Min.   : 51.0   Min.   : 0.00
## 1st Qu.: 45982 1st Qu.: 66.0   1st Qu.: 2.00
## Median :185601 Median : 105.5  Median : 4.00
## Mean   :156491 Mean   : 191.8   Mean   : 12.33
## 3rd Qu.:257582 3rd Qu.: 216.0   3rd Qu.: 11.00
## Max.   :323936 Max.   :1929.0   Max.   :426.00
```

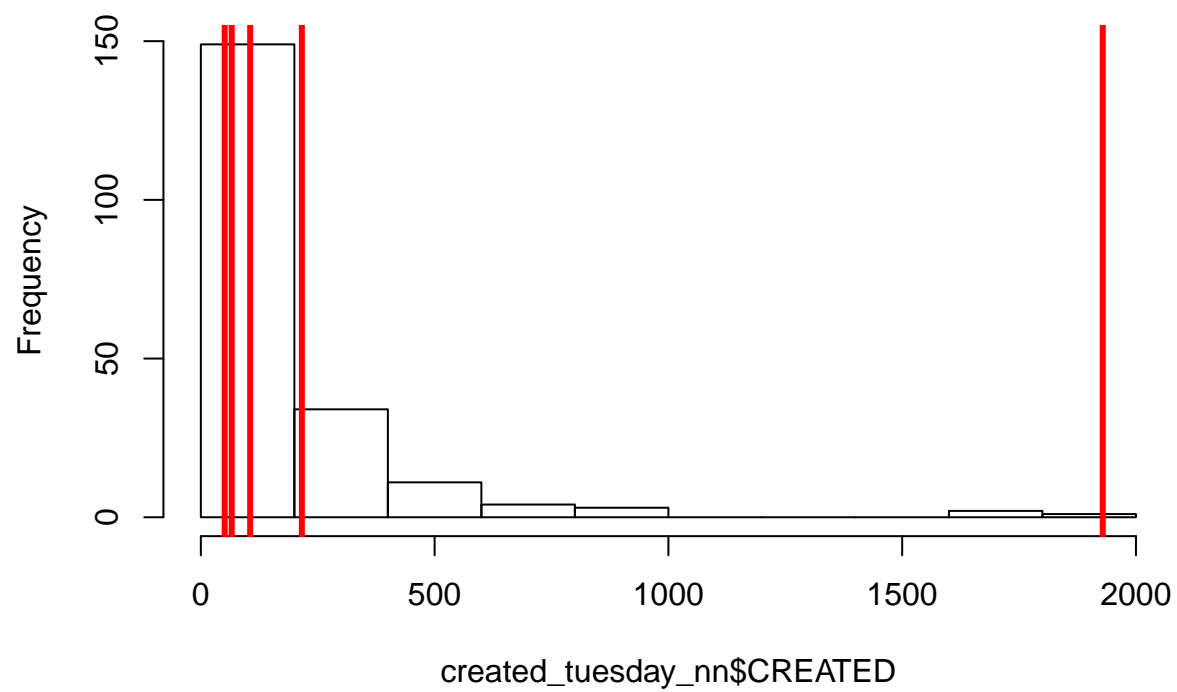
```
## CREATED_TUESDAY_CANCEL_24_HOUR
## Min.   : 0.00
## 1st Qu.: 4.00
## Median : 9.00
## Mean   : 19.69
## 3rd Qu.: 19.00
## Max.   : 460.00
```

```
colnames(created_tuesday_nn) <- c("T00", "CREATED", "C1", "C24")
plot(created_tuesday_nn$CREATED, col="black", main="Bookings x client",
      xlab="Client", ylab="Number of bookings")
```

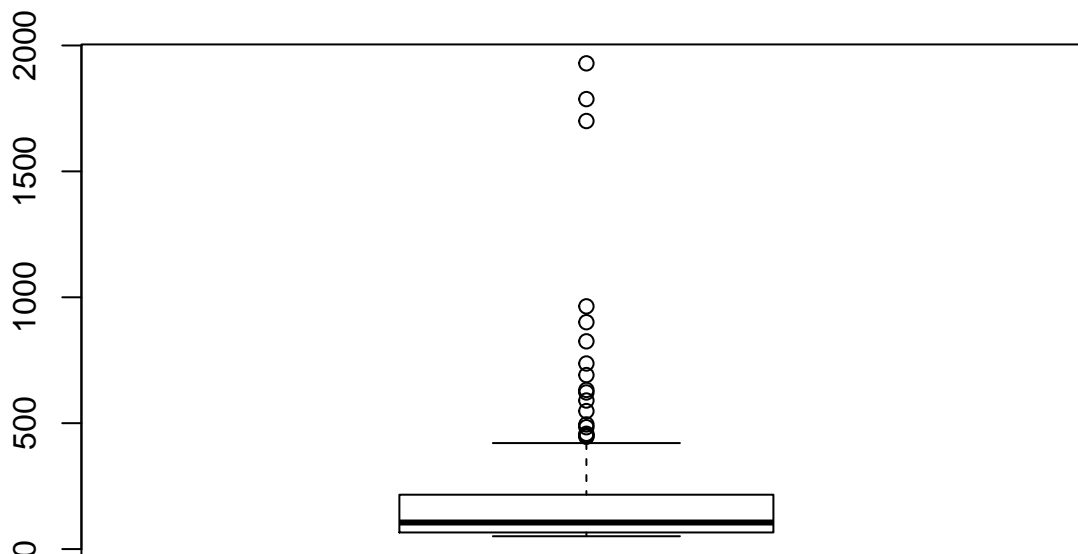


```
hist(created_tuesday_nn$CREATED)
abline(v=quantile(created_tuesday_nn$CREATED), col="red", lwd=3)
```

Histogram of created_tuesday_nn\$CREATED



```
boxplot(created_tuesday_nn$CREATED)
boxplot(created_tuesday_nn$CREATED)$out
```



```
## [1] 901 450 964 458 452 1787 1700 590 622 446 632 484 737 495
## [15] 1929 691 825 548
```

```
shapiro.test(created_tuesday_nn$CREATED)
```

```
##
## Shapiro-Wilk normality test
##
## data: created_tuesday_nn$CREATED
## W = 0.52752, p-value < 2.2e-16
```

```
cor(created_tuesday_nn[, c("CREATED", "C1", "C24")])
```

```
##          CREATED          C1          C24
## CREATED 1.0000000 0.6054281 0.7425328
## C1       0.6054281 1.0000000 0.9706181
## C24      0.7425328 0.9706181 1.0000000
```

```
created_tuesday_nn$C1C24[created_tuesday_nn$C1>0] <-
  created_tuesday_nn$C1[created_tuesday_nn$C1>0] /
  created_tuesday_nn$C24[created_tuesday_nn$C1>0]
summary(created_tuesday_nn$C1C24[created_tuesday_nn$C24>0])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.04762 0.43147 0.53846 0.58042 0.71429 1.00000     13
```

```
created_tuesday_nn[created_tuesday_nn$C1 == created_tuesday_nn$C24,]
```

```
##      TT00 CREATED  C1 C24 C1C24
```


##	31	322	58	0	0	NA
##	267	3847	51	0	0	NA
##	1678	25812	75	1	1	1
##	2427	30542	291	291	291	1
##	2450	31218	55	2	2	1
##	3206	41164	118	4	4	1
##	3295	42087	55	1	1	1
##	3317	42373	89	0	0	NA
##	3634	46025	56	3	3	1
##	3923	49927	59	0	0	NA
##	4194	53874	54	1	1	1
##	4340	56026	68	0	0	NA
##	5136	66776	58	2	2	1
##	5557	72899	51	3	3	1
##	5891	78935	64	2	2	1
##	6834	114539	58	0	0	NA
##	7471	137701	121	0	0	NA
##	9152	197160	53	0	0	NA
##	10114	224717	54	0	0	NA
##	10569	238408	244	2	2	1
##	10647	240234	103	2	2	1
##	10893	246153	61	7	7	1
##	11327	258662	128	5	5	1
##	11339	258918	93	1	1	1
##	11875	271852	56	5	5	1
##	12229	279027	54	0	0	NA
##	12632	287183	136	6	6	1
##	13297	296492	79	0	0	NA
##	13323	297125	84	1	1	1
##	14277	317028	71	1	1	1

```
created_tuesday_nn[created_tuesday_nn$CREATED == created_tuesday_nn$C24,]
```

##	TT00	CREATED	C1	C24	C1C24
##	2427	30542	291	291	291
					1