

Data Science Lab - Assignment 1

Lorenzo Mauri

6 maggio 2020

```
library(magrittr) #package for concatenating commands %>%
library(tibble)  #set data as tibble object
library(dplyr)   # data management (filter,select,summarize,ecc...)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(lubridate) #manipulating dates

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##   date

library(tidyr)

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:magrittr':
##
##   extract

library(pander)

file_path = "C:\\Users\\loren\\Desktop\\repos_github\\Covid19_prediction\\data\\comune_giorno.csv"
data1=read.csv(file_path)
data1=data1%>%head(20)
data1

##   REG PROV NOME_REGIONE NOME_PROVINCIA NOME_COMUNE COD_PROVCOM
## 1    1    1    Piemonte      Torino      Agliè      1001
## 2    1    1    Piemonte      Torino      Agliè      1001
## 3    1    1    Piemonte      Torino      Agliè      1001
## 4    1    1    Piemonte      Torino      Agliè      1001
## 5    1    1    Piemonte      Torino      Agliè      1001
## 6    1    1    Piemonte      Torino      Agliè      1001
## 7    1    1    Piemonte      Torino      Agliè      1001
## 8    1    1    Piemonte      Torino      Agliè      1001
## 9    1    1    Piemonte      Torino      Agliè      1001
## 10   1    1    Piemonte      Torino      Agliè      1001
## 11   1    1    Piemonte      Torino      Agliè      1001
```

##	12	1	1	Piemonte	Torino	Agliè	1001	
##	13	1	1	Piemonte	Torino	Agliè	1001	
##	14	1	1	Piemonte	Torino	Agliè	1001	
##	15	1	1	Piemonte	Torino	Agliè	1001	
##	16	1	1	Piemonte	Torino	Agliè	1001	
##	17	1	1	Piemonte	Torino	Agliè	1001	
##	18	1	1	Piemonte	Torino	Agliè	1001	
##	19	1	1	Piemonte	Torino	Agliè	1001	
##	20	1	1	Piemonte	Torino	Agliè	1001	
##		DATA_INIZIO_DIFF	CL_ETA	GE	MASCHI_15	MASCHI_16	MASCHI_17	MASCHI_18
##	1		1 aprile	17 102	0	0	0	0
##	2		1 aprile	18 104	0	0	0	0
##	3		1 aprile	18 105	0	0	0	0
##	4		1 aprile	17 106	1	0	0	0
##	5		1 aprile	18 106	0	0	0	1
##	6		1 aprile	20 106	0	0	0	0
##	7		1 aprile	16 107	0	0	0	0
##	8		1 aprile	17 107	0	0	0	0
##	9		1 aprile	21 107	0	0	0	0
##	10		1 aprile	19 108	0	0	0	0
##	11		1 aprile	19 109	0	0	0	0
##	12		1 aprile	20 109	0	0	0	0
##	13		1 aprile	14 110	0	0	0	0
##	14		1 aprile	10 113	0	0	0	0
##	15		1 aprile	17 117	0	0	0	0
##	16		1 aprile	19 117	0	0	0	0
##	17		1 aprile	19 118	0	0	0	0
##	18		1 aprile	20 118	0	0	0	0
##	19		1 aprile	19 119	0	1	0	0
##	20		1 aprile	13 120	0	0	0	0
##		MASCHI_19	MASCHI_20	FEMMINE_15	FEMMINE_16	FEMMINE_17	FEMMINE_18	
##	1		0	0	0	0	0	1
##	2		0	0	0	1	0	0
##	3		0	1	0	0	0	0
##	4		0	0	0	0	0	0
##	5		0	0	0	0	0	0
##	6		0	0	0	0	0	1
##	7		0	0	0	0	0	0
##	8		0	0	0	0	1	0
##	9		0	0	0	1	0	0
##	10		0	0	0	1	0	0
##	11		0	0	0	0	0	1
##	12		0	0	0	0	0	0
##	13		0	1	0	0	0	0
##	14		1	0	0	0	0	0
##	15		0	0	0	1	0	0
##	16		0	0	0	0	0	1
##	17		0	0	0	0	0	0
##	18		0	0	1	0	0	0
##	19		0	0	0	0	0	0
##	20		0	0	0	1	0	0
##		FEMMINE_19	FEMMINE_20	TOTALE_15	TOTALE_16	TOTALE_17	TOTALE_18	TOTALE_19
##	1		0	0	0	0	0	1
##	2		0	0	0	1	0	0

```

## 3      0      0      0      0      0      0      0
## 4      0      0      1      0      0      0      0
## 5      0      0      0      0      0      1      0
## 6      0      0      0      0      0      1      0
## 7      1      0      0      0      0      0      1
## 8      0      0      0      0      1      0      0
## 9      0      0      0      1      0      0      0
## 10     0      0      0      1      0      0      0
## 11     0      0      0      0      0      1      0
## 12     0      1      0      0      0      0      0
## 13     0      0      0      0      0      0      0
## 14     0      0      0      0      0      0      1
## 15     0      0      0      1      0      0      0
## 16     0      0      0      0      0      1      0
## 17     0      1      0      0      0      0      0
## 18     0      0      1      0      0      0      0
## 19     0      0      0      1      0      0      0
## 20     0      0      0      1      0      0      0
##      TOTALE_20
## 1      0
## 2      0
## 3      1
## 4      0
## 5      0
## 6      0
## 7      0
## 8      0
## 9      0
## 10     0
## 11     0
## 12     1
## 13     1
## 14     0
## 15     0
## 16     0
## 17     1
## 18     0
## 19     0
## 20     0

```

```
#reshaping dataset
```

```
data1=gather(data1,'MASCHI_15','MASCHI_16','MASCHI_17','MASCHI_18','MASCHI_19','MASCHI_20','FEMMINE_15')
```

```
#elimino le righe con 0 morti e le colonne dei totali
```

```
data1=data1%>%subset(MORTI>0)
```

```
data2=data1%>%separate(`GENERE/ANNO`,into=c('GENERE','ANNO'),sep='_')
```

```
data2$GE=ifelse(data2$GE%>%nchar()==3,paste('0',data2$GE,data2$ANNO,sep=''),paste(data2$GE,data2$ANNO,sep=''))
data2$DATA=data2$GE%>%mdy()
```

```
data3=data2%>%select('DATA','NOME_REGIONE','NOME_PROVINCIA','NOME_COMUNE','GENERE','MORTI','CL_ETÀ','DATA')
```

```
head(data3,20)
```

##	DATA	NOME_REGIONE	NOME_PROVINCIA	NOME_COMUNE	GENERE	MORTI
## 4	2015-01-06	Piemonte	Torino	Agliè	MASCHI	1
## 39	2016-01-19	Piemonte	Torino	Agliè	MASCHI	1
## 65	2018-01-06	Piemonte	Torino	Agliè	MASCHI	1
## 94	2019-01-13	Piemonte	Torino	Agliè	MASCHI	1
## 103	2020-01-05	Piemonte	Torino	Agliè	MASCHI	1
## 113	2020-01-10	Piemonte	Torino	Agliè	MASCHI	1
## 138	2015-01-18	Piemonte	Torino	Agliè	FEMMINE	1
## 142	2016-01-04	Piemonte	Torino	Agliè	FEMMINE	1
## 149	2016-01-07	Piemonte	Torino	Agliè	FEMMINE	1
## 150	2016-01-08	Piemonte	Torino	Agliè	FEMMINE	1
## 155	2016-01-17	Piemonte	Torino	Agliè	FEMMINE	1
## 160	2016-01-20	Piemonte	Torino	Agliè	FEMMINE	1
## 168	2017-01-07	Piemonte	Torino	Agliè	FEMMINE	1
## 181	2018-01-02	Piemonte	Torino	Agliè	FEMMINE	1
## 186	2018-01-06	Piemonte	Torino	Agliè	FEMMINE	1
## 191	2018-01-09	Piemonte	Torino	Agliè	FEMMINE	1
## 196	2018-01-17	Piemonte	Torino	Agliè	FEMMINE	1
## 207	2019-01-07	Piemonte	Torino	Agliè	FEMMINE	1
## 232	2020-01-09	Piemonte	Torino	Agliè	FEMMINE	1
## 237	2020-01-18	Piemonte	Torino	Agliè	FEMMINE	1

##	CL_ETA	DATA_INIZIO_DIFF	COD_PROVCOM	PROV	REG	GE
## 4	17	1 aprile	1001	1	1	010615
## 39	19	1 aprile	1001	1	1	011916
## 65	18	1 aprile	1001	1	1	010618
## 94	10	1 aprile	1001	1	1	011319
## 103	18	1 aprile	1001	1	1	010520
## 113	14	1 aprile	1001	1	1	011020
## 138	20	1 aprile	1001	1	1	011815
## 142	18	1 aprile	1001	1	1	010416
## 149	21	1 aprile	1001	1	1	010716
## 150	19	1 aprile	1001	1	1	010816
## 155	17	1 aprile	1001	1	1	011716
## 160	13	1 aprile	1001	1	1	012016
## 168	17	1 aprile	1001	1	1	010717
## 181	17	1 aprile	1001	1	1	010218
## 186	20	1 aprile	1001	1	1	010618
## 191	19	1 aprile	1001	1	1	010918
## 196	19	1 aprile	1001	1	1	011718
## 207	16	1 aprile	1001	1	1	010719
## 232	20	1 aprile	1001	1	1	010920
## 237	19	1 aprile	1001	1	1	011820

```
data3%>%arrange(DATA)%>%head(20)
```

##	DATA	NOME_REGIONE	NOME_PROVINCIA	NOME_COMUNE	GENERE	MORTI	CL_ETA
## 1	2015-01-06	Piemonte	Torino	Agliè	MASCHI	1	17
## 2	2015-01-18	Piemonte	Torino	Agliè	FEMMINE	1	20
## 3	2016-01-04	Piemonte	Torino	Agliè	FEMMINE	1	18
## 4	2016-01-07	Piemonte	Torino	Agliè	FEMMINE	1	21
## 5	2016-01-08	Piemonte	Torino	Agliè	FEMMINE	1	19
## 6	2016-01-17	Piemonte	Torino	Agliè	FEMMINE	1	17
## 7	2016-01-19	Piemonte	Torino	Agliè	MASCHI	1	19
## 8	2016-01-20	Piemonte	Torino	Agliè	FEMMINE	1	13
## 9	2017-01-07	Piemonte	Torino	Agliè	FEMMINE	1	17

## 10	2018-01-02	Piemonte	Torino	Agliè FEMMINE	1	17
## 11	2018-01-06	Piemonte	Torino	Agliè MASCHI	1	18
## 12	2018-01-06	Piemonte	Torino	Agliè FEMMINE	1	20
## 13	2018-01-09	Piemonte	Torino	Agliè FEMMINE	1	19
## 14	2018-01-17	Piemonte	Torino	Agliè FEMMINE	1	19
## 15	2019-01-07	Piemonte	Torino	Agliè FEMMINE	1	16
## 16	2019-01-13	Piemonte	Torino	Agliè MASCHI	1	10
## 17	2020-01-05	Piemonte	Torino	Agliè MASCHI	1	18
## 18	2020-01-09	Piemonte	Torino	Agliè FEMMINE	1	20
## 19	2020-01-10	Piemonte	Torino	Agliè MASCHI	1	14
## 20	2020-01-18	Piemonte	Torino	Agliè FEMMINE	1	19
##	DATA_INIZIO_DIFF	COD_PROVCOM	PROV REG	GE		
## 1	1 aprile	1001	1 1	010615		
## 2	1 aprile	1001	1 1	011815		
## 3	1 aprile	1001	1 1	010416		
## 4	1 aprile	1001	1 1	010716		
## 5	1 aprile	1001	1 1	010816		
## 6	1 aprile	1001	1 1	011716		
## 7	1 aprile	1001	1 1	011916		
## 8	1 aprile	1001	1 1	012016		
## 9	1 aprile	1001	1 1	010717		
## 10	1 aprile	1001	1 1	010218		
## 11	1 aprile	1001	1 1	010618		
## 12	1 aprile	1001	1 1	010618		
## 13	1 aprile	1001	1 1	010918		
## 14	1 aprile	1001	1 1	011718		
## 15	1 aprile	1001	1 1	010719		
## 16	1 aprile	1001	1 1	011319		
## 17	1 aprile	1001	1 1	010520		
## 18	1 aprile	1001	1 1	010920		
## 19	1 aprile	1001	1 1	011020		
## 20	1 aprile	1001	1 1	011820		