BIG DATA PROJECT

PALIO DI SIENA

“

Il Palio è il Palio. Nessuna interpretazione sociologica, storica, antropologica, potrebbe spiegarlo. Sublimazione e dannazione insieme del fato in ogni singolo senese e nella sua cittadinanza. Rogo furente della senesità, in ogni caso impareggiabile conferma di essa.

(Mario Luzi, italian poet)

1.

RULES OF THE RACE

HOW DOES IT WORK?

2

races each year, on July 2nd and

August 16th

HOW DOES IT WORK?

10 out of 17 Contrade take part in the race

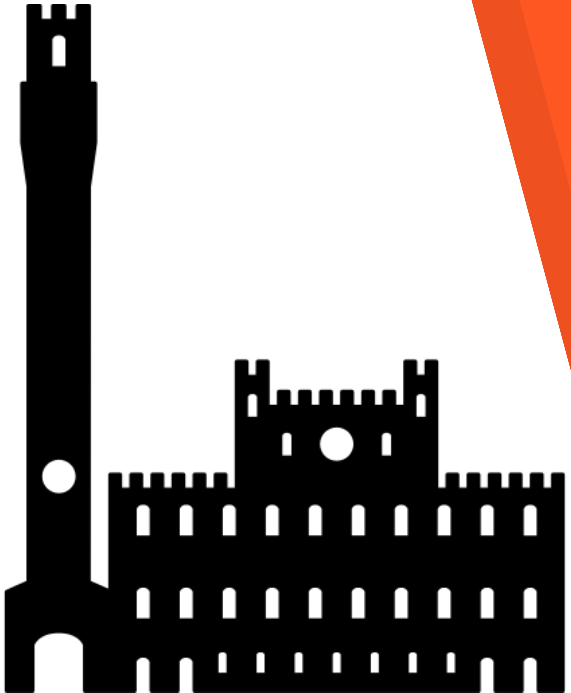
3 are chosen by draw

HOW DOES IT WORK?

3

laps around Piazza del Campo, the

first to finish gets the Drappellone



2.

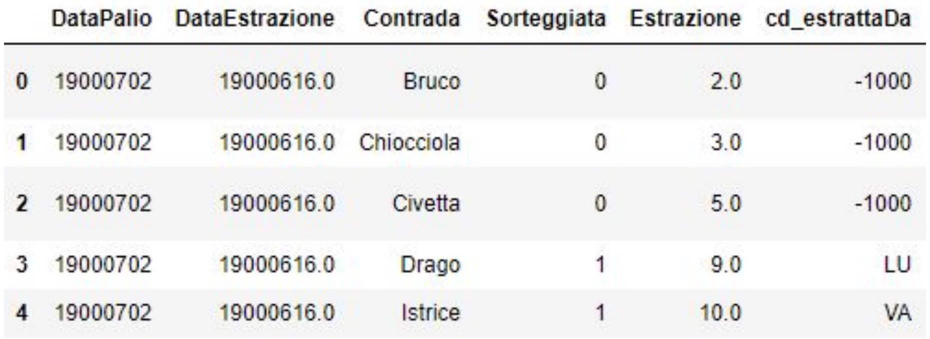
STATISTICS ON PALIO

LIBRARY FOR DATA ANALYSIS: PANDAS

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

DATASET ATTRIBUTES (1)

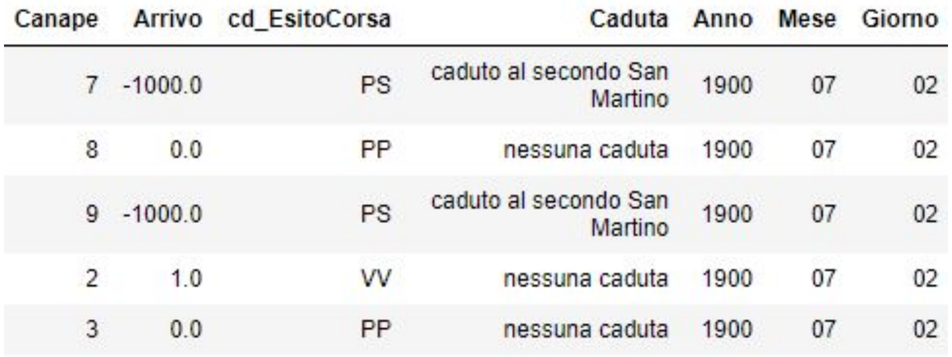
1. DataPalio: date of the race without spaces
2. DataEstrazione: date of the draw of Contrade
3. Contrada: name of the Contrada participating in the race
4. Sorteggiata: 1 if Contrada was chosen by draw, 0 if Contrada participates on its own right
5. Estrazione: number of extraction in which the Contrada was drawn
6. cd_estrattaDa: Contrada by which the corresponding Contrada was extracted



	DataPalio	DataEstrazione	Contrada	Sorteggiata	Estrazione	cd_estrattaDa
0	19000702	19000616.0	Bruco	0	2.0	-1000
1	19000702	19000616.0	Chiocciola	0	3.0	-1000
2	19000702	19000616.0	Civetta	0	5.0	-1000
3	19000702	19000616.0	Drago	1	9.0	LU
4	19000702	19000616.0	Istrice	1	10.0	VA

DATASET ATTRIBUTES (2)

1. Canape: position of Contrada near *canape*
2. Arrivo: final position of Contrada, 1 in case of win
3. cd_EsitoCorsa: race outcome
   a. PS: jockey fell from the horse (horse ran *“scosso”*)
   b. PP: Contrada didn’t run/lost the race
   c. VV: Contrada won the Palio
4. Caduta: if and where the jockey fell from the horse
5. Anno, Mese, Giorno: year, month, day of the race splitted in three attributes

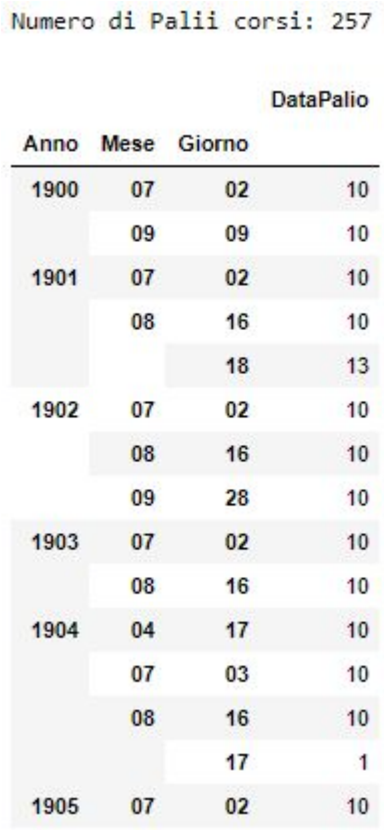


Canape	Arrivo	cd_EsitoCorsa	Caduta	Anno	Mese	Giorno
7	-1000.0	PS	caduto al secondo San Martino	1900	07	02
8	0.0	PP	nessuna caduta	1900	07	02
9	-1000.0	PS	caduto al secondo San Martino	1900	07	02
2	1.0	VV	nessuna caduta	1900	07	02
3	0.0	PP	nessuna caduta	1900	07	02

119 YEARS OF PALIO

The dataset contains data about the Palii from 1900 to 2019. How many were they? Let’s count them!

```
palii =
(pd.DataFrame(df.groupby(['Anno','Mese','Giorno']).count()['Da
taPalio']))
print('Numero di Palii corsi:',len(palii))
palii.head(15)
```



```
Numero di Palii corsi: 257
```

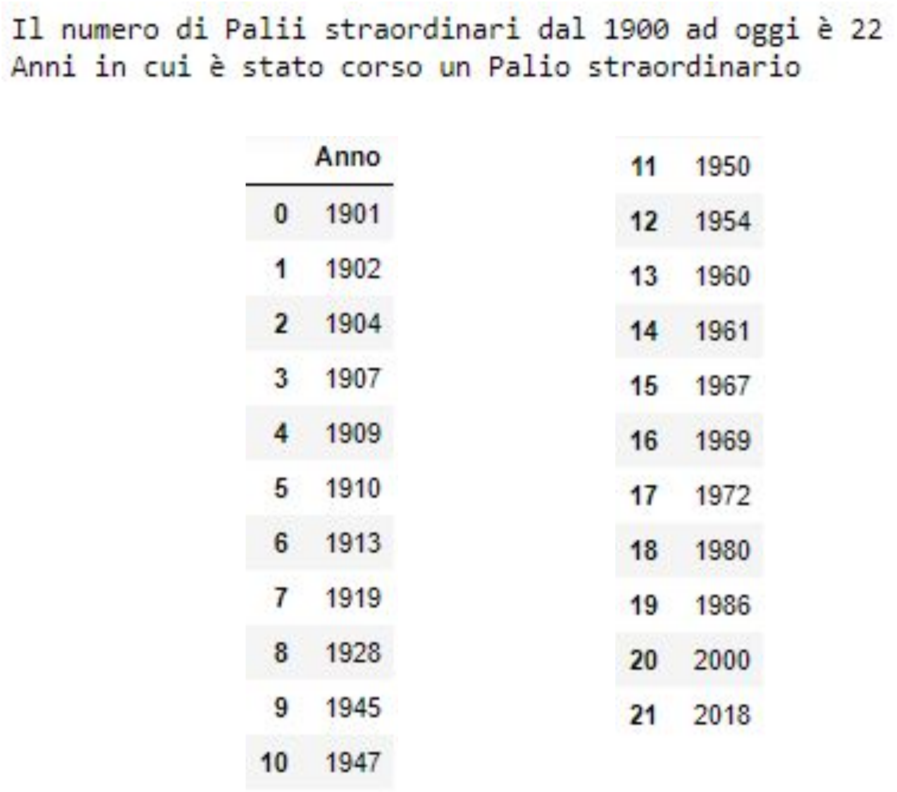
			DataPalio
Anno	Mese	Giorno	
1900	07	02	10
	09	09	10
1901	07	02	10
	08	16	10
		18	13
1902	07	02	10
	08	16	10
	09	28	10
1903	07	02	10
	08	16	10
1904	04	17	10
	07	03	10
	08	16	10
		17	1
1905	07	02	10

EXTRAORDINARY PALII

So, if the numbers don’t lie, there are many Palii in excess. What’s the trick? To celebrate anniversaries or important events, a Palio can be organized: the last time was in 2018, when Siena celebrated 100 years after the end of the 1st World War.

```
a = pd.DataFrame(df).groupby(['Anno']).count()
a = a[a >= 30].dropna()
print('Il numero di Palii straordinari dal 1900 ad oggi
è',len(a))
print('Anni in cui è stato corso un Palio straordinario')
pd.DataFrame(a.index)
```

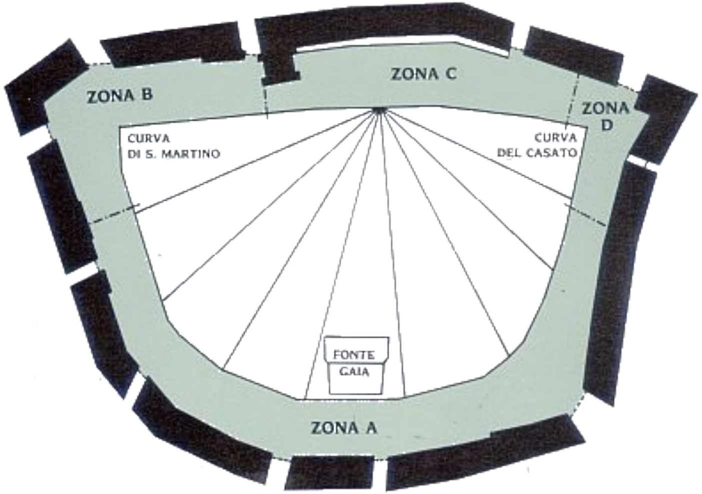
EXTRAORDINARY PALII



```
Il numero di Palii straordinari dal 1900 ad oggi è 22
Anni in cui è stato corso un Palio straordinario
```

	Anno
0	1901
1	1902
2	1904
3	1907
4	1909
5	1910
6	1913
7	1919
8	1928
9	1945
10	1947
11	1950
12	1954
13	1960
14	1961
15	1967
16	1969
17	1972
18	1980
19	1986
20	2000
21	2018

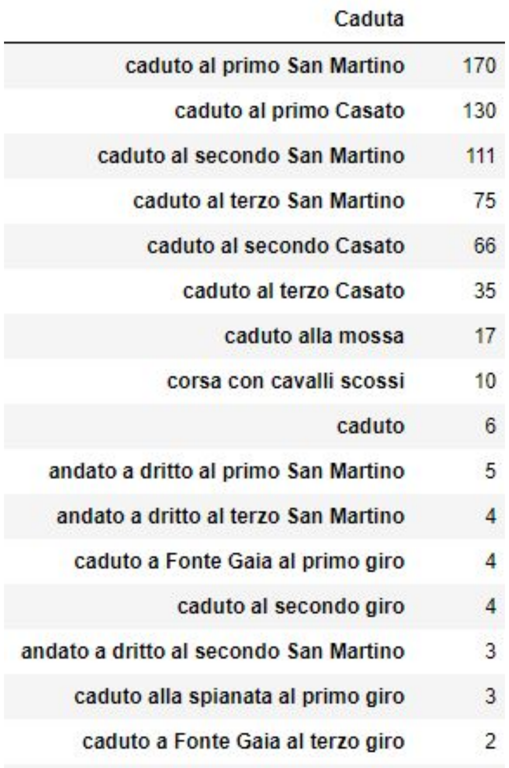
THE MOST DANGEROUS BEND



The San Martino bend has a bending angle of 95°, while the Casato bend is 92° bending. That's why they're considered more dangerous points of the track with respect to the others: is this true? And which is the most dangerous?

```
df2 = pd.DataFrame(columns = ['Cadute'])
df2.Cadute = (df[(df.Caduta != 'nessuna
caduta')].groupby('Caduta').count()['DataPalio']).sort_values
(axis = 0, ascending = False)
```

THE MOST DANGEROUS BEND



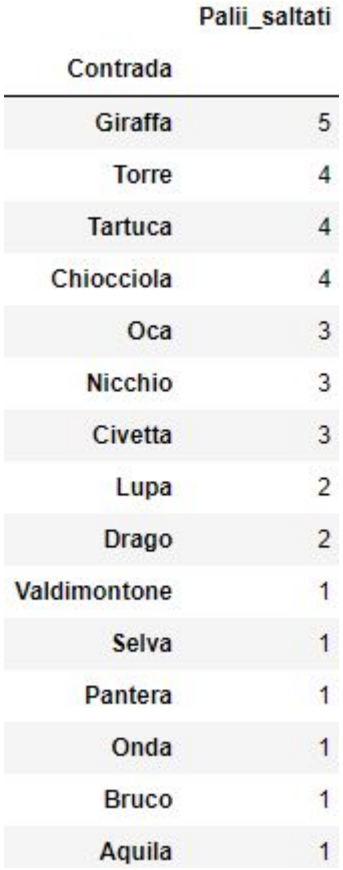
Caduta	
caduto al primo San Martino	170
caduto al primo Casato	130
caduto al secondo San Martino	111
caduto al terzo San Martino	75
caduto al secondo Casato	66
caduto al terzo Casato	35
caduto alla mossa	17
corsa con cavalli scossi	10
caduto	6
andato a dritto al primo San Martino	5
andato a dritto al terzo San Martino	4
caduto a Fonte Gaia al primo giro	4
caduto al secondo giro	4
andato a dritto al secondo San Martino	3
caduto alla spianata al primo giro	3
caduto a Fonte Gaia al terzo giro	2

NUMBER OF NON-PARTICIPATIONS

It may happen that one or more Contrade cannot or don’t want to take part in the race, due to protests, injury to the horse or to the jockey during a Prova of the Palio. However, show must go on and the race is run.

```
df3 = pd.DataFrame(columns = ['Palii_saltati'])
df3.Palii_saltati = pd.DataFrame(df[(df.cd_EsitoCorsa ==
'NP') | (df.cd_EsitoCorsa ==
'NC')]).groupby('Contrada').count()['DataPalio'].sort_va
lues(axis = 0, ascending = False)
```

NUMBER OF NON-PARTICIPATIONS



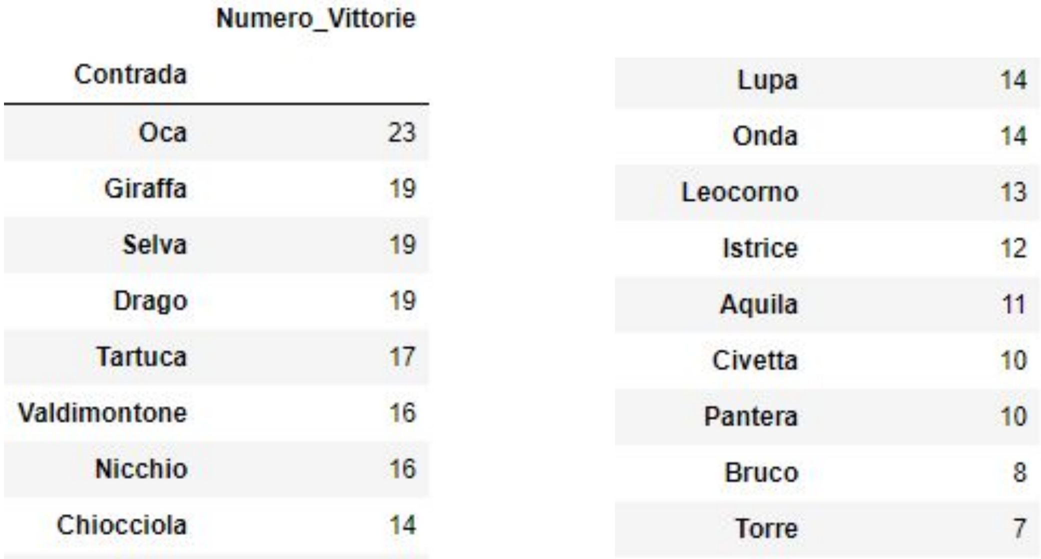
	Palii_saltati
Contrada	
Giraffa	5
Torre	4
Tartuca	4
Chiocciola	4
Oca	3
Nicchio	3
Civetta	3
Lupa	2
Drago	2
Valdimontone	1
Selva	1
Pantera	1
Onda	1
Bruco	1
Aquila	1

NUMBER OF WINS, CONTRADA BY CONTRADA

Who won the highest number of Palii?

```
vittorie_contrade = pd.DataFrame(columns = ['Numero_Vittorie'])
vittorie_contrade.Numero_Vittorie =
pd.DataFrame(df[(df.Arrivo.apply(int) ==
1)]).groupby('Contrada').count()['DataPalio'].sort_values(axis =
0, ascending = False)
```

NUMBER OF WINS, CONTRADA BY CONTRADA



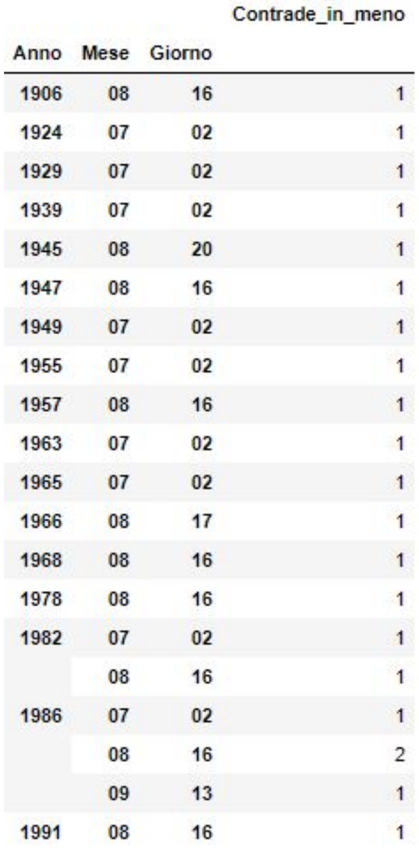
	Numero_Vittorie
Contrada	
Oca	23
Giraffa	19
Selva	19
Drago	19
Tartuca	17
Valdimontone	16
Nicchio	16
Chiocciola	14
Lupa	14
Onda	14
Leocorno	13
Istrice	12
Aquila	11
Civetta	10
Pantera	10
Bruco	8
Torre	7

PALII RUN BY LESS THAN 10 CONTRADE

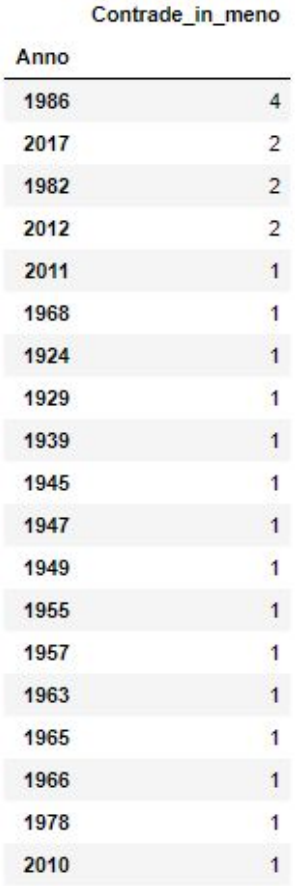
Almost the same as the previous case, but we are interested in knowing how many Contrade didn’t run a specific Palio…

```
df3 = pd.DataFrame(columns = ['Contrade_in_meno'])
df3.Contrade_in_meno = pd.DataFrame(df[(df.cd_EsitoCorsa == 'NP')
| (df.cd_EsitoCorsa ==
'NC')]).groupby(['Anno','Mese','Giorno']).count()['DataPalio']
df3
```

PALII RUN BY LESS THAN 10 CONTRADE



			Contrade_in_meno
Anno	Mese	Giorno	
1906	08	16	1
1924	07	02	1
1929	07	02	1
1939	07	02	1
1945	08	20	1
1947	08	16	1
1949	07	02	1
1955	07	02	1
1957	08	16	1
1963	07	02	1
1965	07	02	1
1966	08	17	1
1968	08	16	1
1978	08	16	1
1982	07	02	1
	08	16	1
1986	07	02	1
	08	16	2
	09	13	1
1991	08	16	1



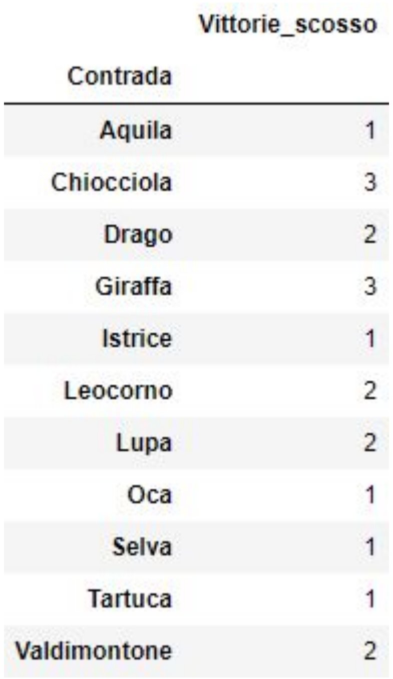
	Contrade_in_meno
Anno	
1986	4
2017	2
1982	2
2012	2
2011	1
1968	1
1924	1
1929	1
1939	1
1945	1
1947	1
1949	1
1955	1
1957	1
1963	1
1965	1
1966	1
1978	1
2010	1

CONTRADE WHICH WON THE PALIO WITH THE CAVALLO SCOSSO

An important rule of the Palio says: since the horse is assigned by fate and the jockey is chosen by men, the horse can win even without jockey (i.e., “scosso”). How many times did each Contrada win the Palio with the *“cavallo scosso”*?

```
df2 = pd.DataFrame(columns = ['Vittorie_scosso'])
df2.Vittorie_scosso = (df[(df.Arrivo.apply(int) == 1) &
(df.Caduta != 'nessuna
caduta')].groupby('Contrada').count()['DataPalio'])
```

CONTRADE WHICH WON THE PALIO WITH THE CAVALLO SCOSSO



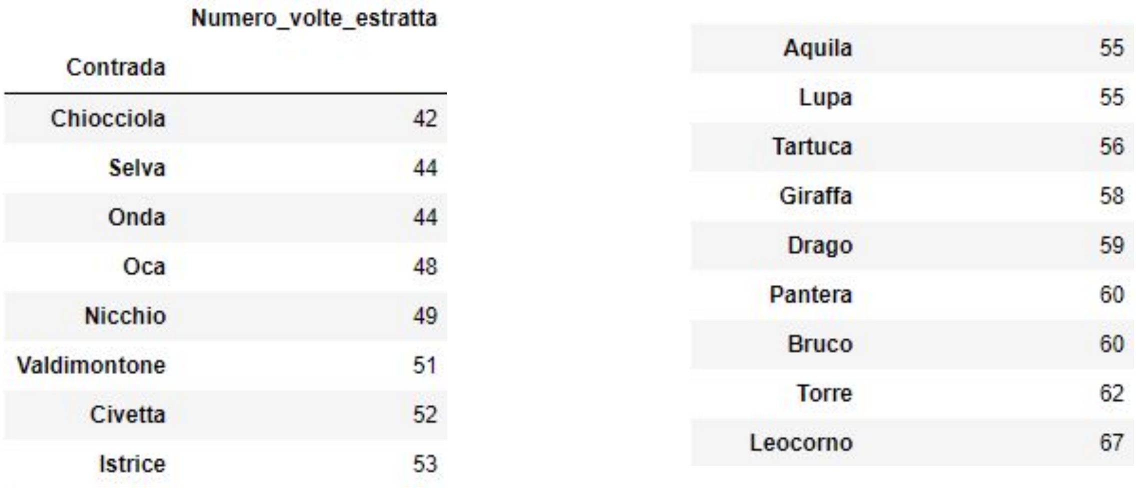
	Vittorie_scosso
Contrada	
Aquila	1
Chiocciola	3
Drago	2
Giraffa	3
Istrice	1
Leocorno	2
Lupa	2
Oca	1
Selva	1
Tartuca	1
Valdimontone	2

NUMBER OF TIMES EXTRACTED, CONTRADA BY CONTRADA

Which is the most extracted Contrada of the last 100 years?

```
sorteggi = pd.DataFrame(columns = ['Numero_volte_estratta'])
sorteggi.Numero_volte_estratta = pd.DataFrame(df[df.Sorteggiata
== 1]).groupby('Contrada').count()['DataPalio'].sort_values(axis
= 0, ascending = True)
```

NUMBER OF TIMES EXTRACTED, CONTRADA BY CONTRADA

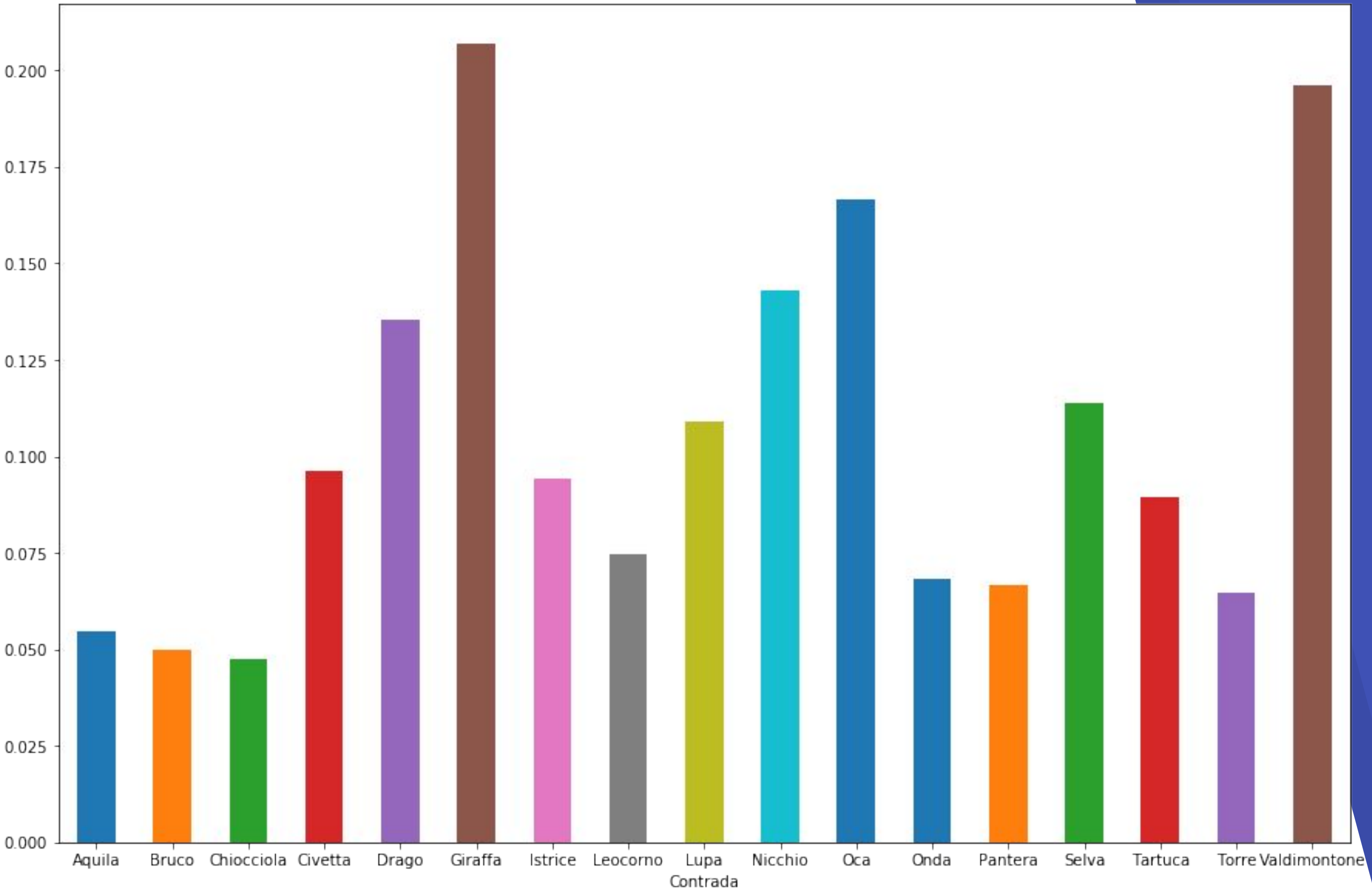


	Numero_volte_estratta
Contrada	
Chiocciola	42
Selva	44
Onda	44
Oca	48
Nicchio	49
Valdimontone	51
Civetta	52
Istrice	53
Aquila	55
Lupa	55
Tartuca	56
Giraffa	58
Drago	59
Pantera	60
Bruco	60
Torre	62
Leocorno	67

WHICH CONTRADA HAS THE BEST RATE **EXTRACTIONS VS. WINS**?

To be extracted is a matter of luck, but being extracted and winning the Palio is a matter of fate!

```
estratte_vittoriose = pd.DataFrame(df[(df.Sorteggiata.apply(int)
== 1) & (df.Arrivo.apply(int) ==
1)]).groupby('Contrada').count()['DataPalio'].sort_values(axis =
0, ascending = True)
vittorie_vs_estrazioni =
estratte_vittoriose/sorteggi.Numero_volte_estratta
contrade = pd.concat([estratte_vittoriose,
sorteggi.Numero_volte_estratta], axis=1, sort = True)
ax = vittorie_vs_estrazioni.plot.bar(rot = 0, figsize = (15,10))
```


0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Aquila
Bruco
Chiocciola
Civetta
Drago
Giraffa
Istrice
Leocorno
Lupa
Nicchio
Oca
Onda
Pantera
Selva
Tartuca
Torre
Valdimontone
Contrada

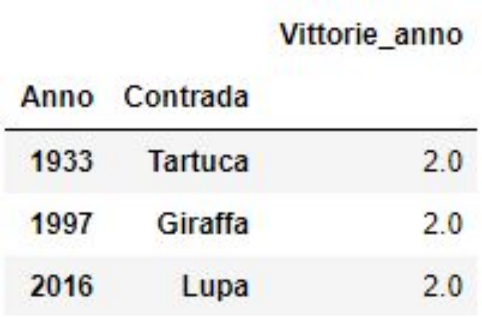
CONTRADE WHICH REALIZED A “CAPPOTTO”

In case of two wins in the same year, it is said the Contrada realized a “cappotto”. The probability of winning twice in one year is low, but it is not impossible!

How many Contrade did it more than once in the last century, and when?

CONTRADE WHICH REALIZED A “CAPPOTTO”

```
cappotti_1 = pd.DataFrame(df[(df.Arrivo == 1) & ((df.Mese ==
'07') | (df.Mese ==
'08'))]).groupby(['Anno','Contrada']).count()
cappotti = pd.DataFrame(columns = ['Vittorie_anno'])
cappotti['Vittorie_anno'] = pd.DataFrame(cappotti_1[cappotti_1
> 1]).dropna()['Arrivo']
pd.DataFrame(cappotti)
```



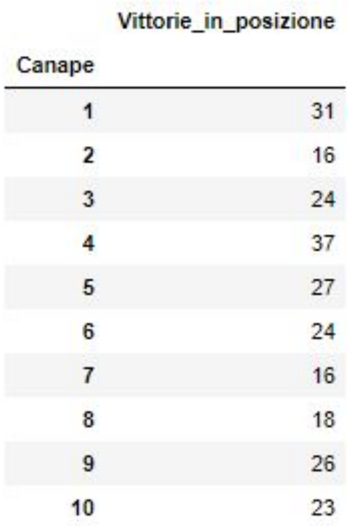
		Vittorie_anno
Anno	Contrada	
1933	Tartuca	2.0
1997	Giraffa	2.0
2016	Lupa	2.0

(DEMYSTIFICATION OF) LOW POSITION AT THE CANAPE

Since the track is not perfectly circular, we would expect that the lower the position, the closer to the *canape*, the more likely the win: can we prove it?

```
posizione = pd.DataFrame(columns = ['Vittorie_in_posizione'])
posizione.Vittorie_in_posizione = pd.DataFrame(df[(df.Canape
!= -1000) & (df.Canape != 0) & (df.Arrivo.apply(int) ==
1)]).groupby(['Canape']).count()['DataPalio']
pd.DataFrame(posizione['Vittorie_in_posizione'].astype('int'))
```

(DEMYSTIFICATION OF) LOW POSITION AT THE CANAPE



	Vittorie_in_posizione
Canape	
1	31
2	16
3	24
4	37
5	27
6	24
7	16
8	18
9	26
10	23

PALII RUN BY “LE QUATTRO VERDI”

“Le quattro verdi” are the four Contrade having green as one of their colours:

1. Drago
2. Selva
3. Oca
4. Bruco

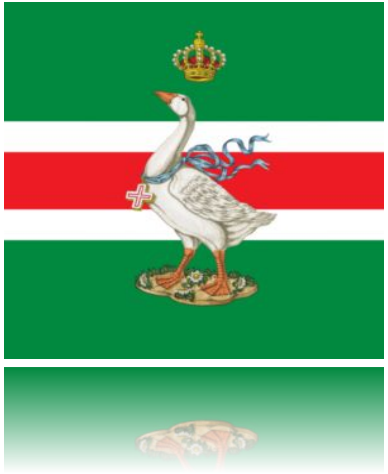
There’s a legend which tells that, when all these Contrade run the Palio, something memorable (bad or good) is going to happen: disqualifications due to bad behaviour on the track, death or injury of horses during the Prove, unexpected rain before the race are just few examples of what happened in the last century.

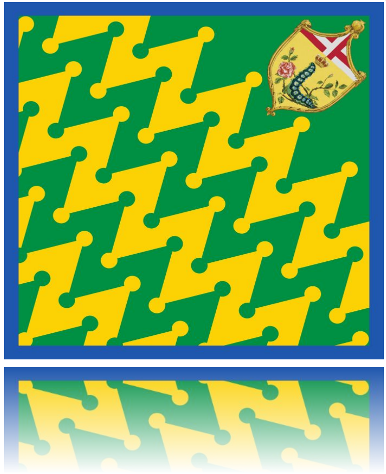
PALII RUN BY “LE QUATTRO VERDI”

```
import numpy as np
a = df[(df.Contrada == 'Selva') | (df.Contrada == 'Oca') | (df.Contrada ==
'Drago') | (df.Contrada == 'Bruco')].groupby('DataPalio').count()
a = a[a == 4].dropna()
quattro_verdi = pd.DataFrame(a).index
print('I Palii corsi con "le quattro verdi" in Campo sono',
len(quattro_verdi))
b = pd.DataFrame()
for e in quattro_verdi:
    b = b.append(df[(df.DataPalio == e)])['Anno']
pd.DataFrame(b).index
(pd.DataFrame(df.Anno[pd.DataFrame(b).index]).groupby('Anno').count())
```

PALII RUN BY “LE QUATTRO VERDI”

1910, 1912, 1919, 1925, 1934, 1936, 1937, 1938, 1939, 1945, 1949, 1964, 1966, 1970, 1972, 1980, 1987, 1989, 2002, 2004, 2006, 2008, 2010





THANK YOU!