

MASTER - Notebook 1

Matteo Grazioso 884055

```
In [ ]: # Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import json
import warnings
warnings.filterwarnings('ignore')
```

```
In [ ]: # Display all columns and all rows
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
```

```
In [ ]: # The files contain the data of the validation of tickets in the city of public transport of Venice.

# Import the data into a dataframe of a txt file
# path = 'data/raw/validazioni.txt' # Period: 2022-05-13 to 2022-07-15
path = 'data/raw/esportazioneCompleta.txt' # Period: 2023-01-23 to 2023-03-14

df = pd.read_csv(path, header=0, sep='\t')
# Save the name of the file in a variable for future use extracting the name of the file from the path
file_name = path.split('/')[-1].split('.')[0]
```

```
In [ ]: # Check the first 5 rows of the data
df.head()
```

```
Out [ ]:
```

	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	DESCRIZIONE_TITOLO
0	13/01/2023 00:00	40834866809772548	162	Stazione MES	12101	Bigl.Aut.75'Mestre/Lido-tsc
1	13/01/2023 00:00	42242241686217732	3625	Aeroporto MA	12106	Bigl Aer-Venezia TSC
2	13/01/2023 00:00	42242241686217476	3625	Aeroporto MA	12106	Bigl Aer-Venezia TSC
3	13/01/2023 00:00	-3604990320	5049	Zattere "B"	23301	Mens.Studente Rete Unica
4	13/01/2023 00:00	-2824230951	5043	S. Toma' "B"	23303	Abb stud. ReteUnica 12 mesi

```
In [ ]: # Check the last 5 rows of the data
df.tail()
```

```
Out [ ]:
```

	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	DESCRIZIONE_TITOLO
5537461	14/03/2023 23:58	-2864643315	162	Stazione MES	11209	Bigl RETE UNICA 75'
5537462	14/03/2023 23:58	-2854956628	5026	Tronchetto F	11209	Bigl RETE UNICA 75'
5537463	14/03/2023 23:59	-2850025054	384	Mestre Centr	23101	Mensile ordinario Rete Unica
5537464	14/03/2023 23:59	-2824225710	5024	Tronchetto "	23101	Mensile ordinario Rete Unica
5537465	14/03/2023 23:59	-3604916033	5039	Rialto "C"	23101	Mensile ordinario Rete Unica

```
In [ ]: # Create a subset of the data with the first 10% of the rows and the last 10% of the rows
# df = df.iloc[:int(len(df)*0.1),:]
# df = df.append(df.iloc[-int(len(df)*0.1):,:])
```

Explorative Data Analysis

```
In [ ]: # Dates and hour of the validation of the ticket are in the same column 'DATA_VALIDAZIONE'
# Split the column 'DATA_VALIDAZIONE' into two columns 'DATA' and 'ORA' and convert them to datetime format
df.insert(0, 'DATA', pd.to_datetime(df['DATA_VALIDAZIONE'].str.split(' ').str[0], format='%d/%m/%Y'))
df.insert(1, 'ORA', pd.to_datetime(df['DATA_VALIDAZIONE'].str.split(' ').str[1], format='%H:%M').dt.time)

# Drop the column 'DATA_VALIDAZIONE'
# df.drop('DATA_VALIDAZIONE', axis=1, inplace=True)
```

```
# Display the first 5 rows of the dataframe
df.head()
```

```
Out [ ]:
```

	DATA	ORA	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	DESCRIZIONE_TITOLO
0	2023-01-13	00:00:00	13/01/2023 00:00	40834866809772548	162	Stazione MES	12101	Bigl.Aut.75'Mestre/Lido-tsc
1	2023-01-13	00:00:00	13/01/2023 00:00	42242241686217732	3625	Aeroporto MA	12106	Bigl Aer-Venezia TSC
2	2023-01-13	00:00:00	13/01/2023 00:00	42242241686217476	3625	Aeroporto MA	12106	Bigl Aer-Venezia TSC
3	2023-01-13	00:00:00	13/01/2023 00:00	-3604990320	5049	Zattere "B"	23301	Mens.Studente Rete Unica
4	2023-01-13	00:00:00	13/01/2023 00:00	-2824230951	5043	S. Toma' "B"	23303	Abb stud. ReteUnica 12 mesi

```
In [ ]: # Set the format of the timestamp
df['DATA_VALIDAZIONE'] = pd.to_datetime(df['DATA_VALIDAZIONE'], format='%d/%m/%Y %H:%M')
```

```
In [ ]: # Print the date of the first and last validation using both data and hour
print('First validation: ', df['DATA'].min(), df['ORA'].min())
print('Last validation: ', df['DATA'].max(), df['ORA'].max())

# Print the number of Serial numbers
print('Number of Serial numbers: ', df['SERIALE'].nunique())

# Print the number of validation (rows)
print('Number of validation: ', df.shape[0])

# Print the number of tickets
print('Number of tickets: ', df['DESCRIZIONE_TITOLO'].nunique())
# Print the number of titolo
print('Number of titolo: ', df['TITOLO'].nunique())
# TODO: why the number of unique TITOLO is different from the number of DESCRIZIONE_TITOLO?

# Print the number of FERMATA
print('Number of FERMATA: ', df['FERMATA'].nunique())
# Print the number of DESCRIZIONE
print('Number of DESCRIZIONE: ', df['DESCRIZIONE'].nunique())
# TODO: why the number of unique DESCRIZIONE is different from the number of FERMATA?
```

First validation: 2023-01-13 00:00:00 00:00:00
Last validation: 2023-03-14 00:00:00 23:59:00
Number of Serial numbers: 1349509
Number of validation: 5537466
Number of tickets: 315
Number of titolo: 316
Number of FERMATA: 1826
Number of DESCRIZIONE: 989

```
In [ ]: # Which is the most used ticket?  
df['DESCRIZIONE_TITOLO'].value_counts().head(10)
```

```
Out[ ]: Bigl RETE UNICA 75'          1167916  
Mensile ordinario Rete Unica      753855  
DailyP-Tpl19,90-C.Ve5,10        522245  
75'-Tpl 8,64-ComVe0,86          449274  
Bigl.Aut.75'Mestre/Lido-tsc      327816  
Annuale ordinario Rete Unica     311494  
48h-Tpl 29,90-ComVe5,10         278703  
72h-Tpl 38,40-ComVe6,60         199789  
Mens.Studente Rete Unica        123083  
7gg-Tpl 48,60-ComVe16,40        108925  
Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Which is the most frequent validation in date and hour?  
# Date and hour are in two different columns; DATA_VALIDAZIONE does not exist anymore  
df.groupby(['DATA', 'ORA'])['SERIALE'].count().sort_values(ascending=False).head(10)  
# TODO: #4 Re-aswer the question of the most frequent validation after cleaning operations
```

```
Out[ ]: DATA      ORA
2023-02-20 16:17:00    330
           17:44:00    301
           10:52:00    290
           15:23:00    288
2023-02-19 17:02:00    287
2023-02-18 16:56:00    286
2023-02-20 15:32:00    284
2023-02-18 16:11:00    283
           16:55:00    277
2023-02-20 16:16:00    276
Name: SERIALE, dtype: int64
```

```
In [ ]: # Which is the most frequent FERMATA?
df['DESCRIZIONE'].value_counts().head(10)
# TODO: #4 Re-aswer the question of the most frequent FERMATA after cleaning operations
```

```
Out[ ]: Lido S.M.E.      386315
P.le Roma "G"      361079
Rialto "C"      344344
San Marco-Sa      276783
VENEZIA      259102
Burano "C"      193367
P.le Roma "E"      176712
Ferrovia "B"      175434
S. Marco-San      169498
Rialto "B"      112853
Name: DESCRIZIONE, dtype: int64
```

Categories

```
In [ ]: # Add a new column with the code profile of the ticket
df.insert(7, "TICKET_CODE", 'TBD')
```

This column will be filled with the code of the ticket profile according to the ticket type and the ticket validity as follows:

1. One-day ticket

2. Two-day ticket
3. Three-day ticket
4. Weekly ticket (Seven-day ticket)
5. Monthly ticket
 - 5-STUD. Monthly ticket for students
 - 5-RET. Monthly ticket for retirees
 - 5-WKRS. Monthly ticket for workers
6. Annual ticket
 - 6-STUD. Annual ticket for students
 - 6-RET. Annual ticket for retirees
 - 6-WKRS. Annual ticket for workers
7. 75 minutes ticket
8. Other ticket (if it is necessary to add other types of tickets)

```
In [ ]: df.head()
```

Out []:	DATA	ORA	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	TICKET_CODE	DESCRIZIONE_TITOLO
0	2023-01-13	00:00:00	2023-01-13	40834866809772548	162	Stazione MES	12101	TBD	Bigl.Aut.75'Mestre/Lido-tsc
1	2023-01-13	00:00:00	2023-01-13	42242241686217732	3625	Aeroporto MA	12106	TBD	Bigl Aer-Venezia TSC
2	2023-01-13	00:00:00	2023-01-13	42242241686217476	3625	Aeroporto MA	12106	TBD	Bigl Aer-Venezia TSC
3	2023-01-13	00:00:00	2023-01-13	-3604990320	5049	Zattere "B"	23301	TBD	Mens.Studente Rete Unica
4	2023-01-13	00:00:00	2023-01-13	-2824230951	5043	S. Toma' "B"	23303	TBD	Abb stud. ReteUnica 12 mesi

```
In [ ]: # Create a dictionary with the ticket code and the ticket profile
dict_tickets = {'1': 'One-day ticket', '2': 'Two-day ticket', '3': 'Three-day ticket',
               '4': 'Seven-day ticket',
               '5': 'Monthly ticket', '5-STUD': 'Monthly ticket for students',
               '5-RET': 'Monthly ticket for retired', '5-WKRS': 'Monthly ticket for workers',
               '6': 'Annual ticket', '6-STUD': 'Annual ticket for students', '6-RET': 'Annual ticket for retired',
               '6-WKRS': 'Annual ticket for workers',
               '7': '75 minutes ticket', '8': 'Other ticket'}

# Export the dictionary to a json file
with open('data/dictionaries/dict_ticket_codes.json', 'w') as fp:
    json.dump(dict_tickets, fp)
```

```
In [ ]: # How many unique values are there in the column 'DESCRIZIONE_TITOLO'?
df['DESCRIZIONE_TITOLO'].nunique()
```

Out []: 315

```
In [ ]: # Which are the unique values of the column 'DESCRIZIONE_TITOLO'?
df['DESCRIZIONE_TITOLO'].unique()
```

```
Out[ ]: array(["Bigl.Aut.75'Mestre/Lido-tsc", 'Bigl Aer-Venezia TSC',  
  'Mens.Studente Rete Unica', 'Abb stud. ReteUnica 12 mesi ',  
  "75'-Tpl 8,64-ComVe0,86", "Bigl RETE UNICA 75'",  
  'Mensile ordinario Rete Unica', 'Linea 17-categoria B',  
  'Aeroporto-Venezia AR', 'Annuale ORDINARIO ISOLE',  
  'Mensile ORDINARIO ISOLE', 'Linea 17-categoria D',  
  'DailyP-Tpl19,90-C.Ve5,10', '48h-Tpl 29,90-ComVe5,10',  
  'Annuale ordinario Rete Unica', 'Atvo+Actv ann.Stud.F1',  
  'Annuale STUDENTE ISOLE', 'Mensile Ordinario extra',  
  'Linea 17-categoria C', 'Prenotazione Veicolo ABBONATO',  
  'Libera circ. RETE intera', 'Tessera di servizio ACTV',  
  'Supp Mens.navigazione', "Bigl.Mestre/Lido 75' a bordo",  
  'Annuale ord.res.PELLESTRINA', '72hAerAR-Tpl51,40-CVe6,60',  
  'Biglietto 72 ore Roll. Venice', 'Traghetto Carta Venezia',  
  'Traghetto residente LIDO', 'Mensile STUDENTE ISOLE',  
  '48ore online no aerobus', 'MOBILITY ordinario Rete Unica',  
  'Libera circ. DUE RETI', '7gg-Tpl 48,60-ComVe16,40',  
  'Abbonamento pensionati Actv', 'Linea 11-categoria B',  
  'Linea 11-categoria C', 'Linea 11+17 categoria D',  
  "75'-Tpl 6,64-ComVe0,86", '72h-Tpl 38,40-ComVe6,60',  
  'Ferry17-autocarri+35q.', 'Abb. Chioggia A20 ',  
  'Ferry17-carri+35q.rim.', '72 ore R.Venice online',  
  'Atvo+Actv mens.Lav.F2', 'Titolo CMVenezia',  
  'Abb. Rete Intera A20', 'L.17-auto "D" oltre metri 4,50',  
  'MOBILITY Ordinario extra', 'Annuale cat. D 17(un semestre)',  
  'Extra tratta 1', 'MOBILITY Supp.NAVIGAZIONE',  
  'Linea 17-categoria A', 'Extra tratta 4 ',  
  'Prenotaz OCCASIONALE si barra', 'Traghetto resid. PELLESTRINA',  
  'L.17-auto "C"da 4,01 a 4,50 mt', 'Semestrale Ceod A20',  
  'Supp.Annuale NAVIGAZIONE', "Bigl RETE UNICA 100'",  
  'Daily Pass Venezia Online', 'Abb. extra A20', 'Extra tratta 2 ',  
  'Annuale Ordinario extra', 'Atvo+Actv mens.Lav.F1',  
  'Atvo+Actv ann.Lav.F2', 'Extra tratta 6 ', 'Atvo+Actv ann.Lav.F1',  
  'Carnet CHIOGGIA 10c. TICKET', 'Mensile ord. res. PELLESTRINA',  
  'ARRIVA VENETO tratta 8-9-10', 'Apertura tornelli DUSSMAN',  
  'Traghetto residente BURANO', 'Atvo+Actv mens.Ord.F1',  
  'Extra tratta 7 ', 'Extra tratta 3 ', 'Aer+boat-Tpl16,50-C.Ve1,50',  
  'Stud. Rete Intera FAMILIARE', 'Abb. Rete Intera A5',  
  'abb. Mensile CHIOGGIA', "NA-Carnet nav. 10 corse da 75'",  
  'Annuale ORDINARIO bus Lido', 'S.Terr+Actv ORD. tr.6',
```


'Semestrale Ceod A20+Acc.', 'Mensile Studente extra',
'Extra tratta 5 ', 'S.Terr+Actv ORD. tr.9',
'Ord. Rete Intera FAMILIARE', 'Atvo+Actv ann.Stud.F2',
'Abb. over75 GRATUITO', 'Atvo+Actv mens.Ord.F3',
'ARRIVAExtra tr.2-3-4 BORDO', 'Mensile stud. PELLESTRINA',
'Traghetto residente S.ERASMO', 'Atvo+Actv mens.Stud.F2',
'Atvo+Actv mens.Stud.F1', 'Atvo+Actv mens.20%.F1',
'S.Terrr+Actv STUDENTE tr.6', 'Bigl.urbano CHIOGGIA',
'72ore online no aerobus', 'Abb.stud.Ann.PELLESTRINA',
'Annuale Studente extra', 'Atvo+Actv mens.Stud.F3',
'Linea 11+17 categoria B', 'Bicicletta "Palmare"',
'7 days online aerobus AR', 'Supp Mens.automobilistico',
'abb. Studente Mens. CHIOGGIA', 'Aeroporto-Venezia CS ONLINE',
'Bicicletta "concessionari"', 'Bicicletta "biglietteria"',
'Abb Annuale PeopleMover', "PeopleMover+Bus+Tram 75'",
'ARRIVA VENETO tratta 4', 'Atvo+Actv mens.20%.F2',
'Bordo 75min CartaVenezia', 'Abb Mensile PeopleMover',
'Supp. 12 mesi studente laguna', 'Atvo+Actv mens.Ord.F2',
"Biglietto di bordo CV 75'", 'Traghetto residente MURANO',
'Abb. Rete Intera A20 +Acc.', 'Atvo+Actv mens.20%.F3',
'ARRIVA VENETO tratta 1', 'Supp Annuale PeopleMover',
'Extra tratte 2-3-4 BORDO', '72 ore R.Venice+aeroporto AR',
'Abb. over75 Rete Unica 50% ', 'Bagaglio CartaVenezia',
'Tessera di servizio ARRIVA', '7 days online no aerobus',
"PeopleMover+Bus+Tram 75'carnet", 'Abb. over 75 A20',
'abbonamento 30 gg.PeopleMover', 'Daily Pass Ve. Online 1mese',
'Mens. cose animali RETE UNICA', 'Ciclomotore fino 50cc',
'48hAerAR-Tpl42,90-CVe5,10', '24hAerCS-Tpl26,90-CVe5,10',
'7 days online no aerobus 1mese', '72h online no aerobus 1mese',
'Mens. cose animali RETE INTERA', '7 days online aerobus CS',
'Linea 11+17 categoria C', 'ARRIVA VENETO tratta 7',
'Abb Stud. 12 mesi CHIOGGIA', 'S.Terr+Actv ORD. tr.7',
'Ev8-Tpl 52,00-C.Ve3,00', '72 ore R.Venice+aeroporto CS',
'72H R.Venice+aerop.AR online', 'Abb. ordinario bus. Lido',
'L.17-auto "AeB" fino a 4 metri', '7ggAerAR-Tpl61,60-CVe16,40',
'72ore online aerobus AR', 'Extra tratta 1 BORDO',
'Extra tratta 8-9-10 ', 'ARRIVA VENETO tratta 6',
'72ore online aerobus CS', '24h-24 ore ',
'48h online aerobus CS 1 mese', '72h R.Venice online 1 mese',
'7 days online aerobus CS 1mese', 'AtvoCanova+Actv 72Hroll.online',

'48hAerCS-Tpl36,90-CVe5,10', 'Atvo Canova+Actv 72H online',
'Tariffa carrozzina', "ord. navigazione 75' online",
"NA-Big.Aut.75' Mestre/Lido-csc", '72hAerCS-Tpl45,40-CVe6,60',
'Extra tratta 2 TVM', '7ggAerCS-Tpl55,60-CVe16,40',
'Aeroporto-Venezia AR ONLINE', 'abb.CHIOGGIA annuale',
'72H RVenice+aerop.CS online', 'T.Fusina Ve+ACTV 72 ore',
'Ferry11-autocarri+35q.', 'L.11-auto "D" oltre metri 4,50',
'Linea 11-categoria D', 'S.Terr+Actv ANN stud tr.6',
'Linea 11-categoria A', '48h online no aerobus 1mese',
'7 days online aerobus AR 1mese', 'abb. impersonale rete INTERA',
'Atvo Canova+Navig AR online', 'L.11-auto "C"da 4,01 a 4,50 mt',
'Atvo+Actv mens.5%.F2', '72h RVe+aerop.CS online 1 mese',
'Abb. extra A5', 'L.11-auto "AeB" fino a 4 metri',
'Supp navigazione FAMILIARE', 'S.Terr+Actv STUDENTE tr.7',
'Studente extra FAMILIARE', 'S.Terr+Actv ANN ord tr.9',
'ARRIVA Extra tr.1 BORDO', 'Supp. Annuale AUTOMOB.',
'72h online aerobus CS 1 mese', '48ore online aerobus AR',
'biglietto merci C.Semplice', 'ARRIVA VENETO tratta 5',
'Aerobus+boat online', '72h online aerobus AR 1 mese',
'Extra tratta 3 TVM', 'Ordinario Chioggia FAMILIARE',
'Ciclomotore oltre 50cc', 'NA-Traghetto ordinario',
'Atvo Canova+Navig AR', 'Extra tratta 1 TVM', 'Extra tratta 4 TVM',
'NA-Bicicletta e conducente CV', 'ARRIVA VENETO AEROPORTO',
'ARRIVA Extra tr. 5-6-7 BORDO', '72h R.Ve.+aer.AR online 1mese',
'Abb. Chioggia A20 + acc.', 'Extra tratta 5 TVM',
'Linea 11+17 categoria A', 'Studente Chioggia FAMILIARE',
'ARRIVA Aeroporto 0.Mensi', 'Ferry17-Trasporti pericolosi',
'Libera circ. FERRY LINEA 17', 'ARRIVA VENETO tratta 2',
'48ore online aerobus CS', "NA-C Aut. 10 corse 75' CARD ",
'Annuale cat. D linea 11', 'Atvo Canova+Navig 1 corsa',
'S.Terr+Actv ANN ord tr.6', 'T.Fusina Ve+ACTV 24 ore',
'Extra tratta 8-9-10 TVM', 'Carnet CHIOGGIA 10 c. CARD',
'S.Terr+Actv STUDENTE tr.5', 'S.Terr+Actv ANN stud tr.7',
'ARRIVA Extra tr.8-9-10 BORDO', 'DDGR1201-1297/2022 Extra',
'Biglietto Soc. Sportive', 'Ordinario extra FAMILIARE',
"NA75'-Tpl 13,28-ComVe1,72", 'Extra tratte 5-6-7 BORDO',
'Atvo Canova+Actv 72H', 'MENSILE park+Rete intera',
'24ore online no aerobus', 'S.Terr+Actv STUDENTE tr.2',
'24hAerAR-Tpl32,90-CVe5,10', 'ARRIVA VENETO tratta 3',
'SpiaggeAR-Tpl 14,75-ComVe1,25', 'Atvo+Actv mens.Lav.F3',

```
'Ev5-Tpl 38,50-C.Ve1,50', 'S.Terr+Actv STUDENTE tr.3',
'Ferry11-carri+35q.rim.', 'Libera circ. FERRY LINEA 17+11',
'Abb. Chioggia A5 ', 'Mens. studente bus LIDO',
'S.Terr+Actv ANN stud tr.5', 'S.Terr+Actv ORD. tr.8',
'ARRIVA Aeroporto BORDO', 'Biglietto MOTO FINO 50 cc',
'Libera circ. FERRY LINEA 11', 'Jesolo + Actv 24H',
'Supp Mensile PeopleMover', 'AtvoCanova+Navig 1corsa online',
'Cav-Trep - S.Marco AR', "ord. navig. 75' online 1 mese",
'S.Terr+Actv ORD. tr.5', 'S.Terr+Actv ANN stud tr.2',
'S.Terr+Actv ANN ord tr.2', 'S.Terr+Actv STUDENTE tr.8',
'Extra tratta 7 TVM', 'Abb. over 75 A5', 'S.Terr+Actv ORD. tr.3',
'Extra tratta 6 TVM', 'Cav -Trep + Actv 24H',
'Aerobus+boat online 1mese', "75'-Tpl 12,60-CVe2,40 online",
'Traghetto residente GIUDECCA', 'Extra tratte 8-9-10 BORDO',
'24ore online aerobus CS', '24ore online aerobus AR',
'NATragh-Tpl 4,41-C.Ve0,59', 'Abb studente bus LIDO 12 mesi ',
'MOBILITY studente ReteUnica', 'S.Terr+Actv ANN ord tr.8',
'48h online aerobus AR 1 mese', 'NATragh-Tpl 8,82-C.Ve1,18',
'Bus+People mover online', 'Integrazione rete MESTRE',
'Supp Mens.urbano CHIOGGIA', 'S.Terr+Actv ORD. tr.2',
'Jesolo - S.Marco AR', 'Navetta Arsenale CA',
'S.Terr+Actv ANN ord tr.7', 'S.Terr+Actv ORD. tr.4',
'Supp. 12 mesi studente automob', "VENDITA A BORDO 75' CV",
'DDGR1201-1297/2022 R. Unica', 'ARRIVA Integ.Aerop. BORDO',
'Suppl. Rete Mestre ATV0', 'Acc.L.R.A20 rete intera',
'NA-Gruppi e Scuole', 'Gruppi e scuole online 2viaggi',
'Ferry11-Trasporti pericolosi', 'Traghetto Gratuito',
'Biglietto scuole', 'S.Terr+Actv STUDENTE tr.9',
'NA-Bigl. CHIOGGIA CARD', 'Biglietto scuole online',
'ANNUALE park+Rete intera', 'Apertura tornelli P.SABBIONI',
'S.Terr+Actv ANN stud tr.8', 'S.Terr+Actv ANN ord tr.3',
'NA-C CHIOGGIA 10 c. CARD fs', 'Ferry17-AUTOBUS',
'S.Terr+Actv STUDENTE tr.4', 'Supp Mestre FAMILIARE',
'ARRIVA Misto Actv 24h', '24h online aerobus CS 1 mese',
'Ev3-Tpl 30,50-C.Ve1,50', 'S.Terr+Actv ANN stud tr.3'],
dtype=object)
```

```
In [ ]: # Get the number of unique values of the column 'DESCRIZIONE_TITOLO'
num_unique_DESCRIZIONE_TITOLO = len(df['DESCRIZIONE_TITOLO'].unique())
print('The number of unique values of the column DESCRIZIONE_TITOLO is: ', num_unique_DESCRIZIONE_TITOLO)
```

The number of unique values of the column DESCRIZIONE_TITOLO is: 315

```
In [ ]: # Convert the column 'DESCRIZIONE_TITOLO' into upper case
df['DESCRIZIONE_TITOLO'] = df['DESCRIZIONE_TITOLO'].str.upper()
# Count the number of unique values of the column 'DESCRIZIONE_TITOLO'
df['DESCRIZIONE_TITOLO'].value_counts()
```

Out[]:	BIGL RETE UNICA 75'	1167916
	MENSILE ORDINARIO RETE UNICA	753855
	DAILYP-TPL19,90-C.VE5,10	522245
	75'-TPL 8,64-COMVE0,86	449274
	BIGL.AUT.75'MESTRE/LIDO-TSC	327816
	ANNUALE ORDINARIO RETE UNICA	311494
	48H-TPL 29,90-COMVE5,10	278703
	72H-TPL 38,40-COMVE6,60	199789
	MENS.STUDENTE RETE UNICA	123083
	7GG-TPL 48,60-COMVE16,40	108925
	MENSILE ORDINARIO ISOLE	85577
	ANNUALE ORDINARIO ISOLE	74351
	ABB STUD. RETEUNICA 12 MESI	65034
	BIGLIETTO 72 ORE ROLL. VENICE	62143
	720RE ONLINE NO AEROBUS	56474
	ABB. RETE INTERA A20	50474
	BIGL AER-VENEZIA TSC	50419
	PEOPLEMOVER+BUS+TRAM 75'	45649
	MOBILITY ORDINARIO RETE UNICA	45048
	EXTRA TRATTA 2	38147
	TESSERA DI SERVIZIO ACTV	37209
	7 DAYS ONLINE NO AEROBUS	34462
	72 ORE R.VENICE ONLINE	31590
	480RE ONLINE NO AEROBUS	31472
	MENSILE ORDINARIO EXTRA	27239
	LIBERA CIRC. RETE INTERA	26236
	SUPP MENS.NAVIGAZIONE	24998
	EXTRA TRATTA 3	22071
	DAILY PASS VENEZIA ONLINE	17424
	ABB. OVER75 GRATUITO	15349
	LINEA 17-CATEGORIA C	14772
	TRAGHETTO CARTA VENEZIA	13195
	TITOLO CMVENEZIA	13116
	EXTRA TRATTA 4	11671
	75'-TPL 6,64-COMVE0,86	11049
	7GGAERAR-TPL61,60-CVE16,40	10113
	LINEA 17-CATEGORIA B	9969
	ANNUALE STUDENTE ISOLE	9965
	MENSILE ORD. RES. PELLESTRINA	9645
	72HAERAR-TPL51,40-CVE6,60	9459

TRAGHETTO RESIDENTE BURANO	9281
EXTRA TRATTA 1	8733
BIGL.MESTRE/LIDO 75' A BORDO	8607
L.17-AUTO "D" OLTRE METRI 4,50	8438
ABBONAMENTO PENSIONATI ACTV	8187
LINEA 17-CATEGORIA D	8152
LIBERA CIRC. DUE RETI	7645
ARRIVA VENETO TRATTA 8-9-10	7499
7 DAYS ONLINE AEROBUS AR	6906
MENSILE STUDENTE ISOLE	6614
72 ORE R.VENICE+AEROPORTO AR	6604
BIGL RETE UNICA 100'	6367
AEROPORTO-VENEZIA AR	6051
72HAERCS-TPL45,40-CVE6,60	6011
72 ORE R.VENICE+AEROPORTO CS	5965
LINEA 11-CATEGORIA C	5904
ANNUALE ORD.RES.PELLESTRINA	5830
ABB. OVER75 RETE UNICA 50%	5294
72H ONLINE NO AEROBUS 1MESE	5218
LINEA 11-CATEGORIA B	5044
720RE ONLINE AEROBUS AR	5011
72H R.VENICE ONLINE 1 MESE	4948
7 DAYS ONLINE NO AEROBUS 1MESE	4821
MENSILE STUDENTE EXTRA	4810
48HAERCS-TPL36,90-CVE5,10	4735
ATVO+ACTV MENS.LAV.F1	4562
EXTRA TRATTA 5	4502
TRAGHETTO RESID. PELLESTRINA	4394
72H R.VENICE+AEROP.AR ONLINE	4242
24HAERCS-TPL26,90-CVE5,10	4181
AER+BOAT-TPL16,50-C.VE1,50	4059
SUPP.ANNUALE NAVIGAZIONE	3583
ORD. NAVIGAZIONE 75' ONLINE	3509
PRENOTAZ OCCASIONALE SI BARRA	3458
ORD. RETE INTERA FAMILIARE	3412
ABB. EXTRA A20	3304
LINEA 11-CATEGORIA D	3198
TRAGHETTO RESIDENTE LIDO	3123
ABB. RETE INTERA A5	3023
ATVO+ACTV MENS.STUD.F1	2967

BIGLIETTO DI BORDO CV 75'	2936
MOBILITY SUPP.NAVIGAZIONE	2917
ANNUALE ORDINARIO EXTRA	2900
48HAERAR-TPL42,90-CVE5,10	2881
L.17-AUTO "C"DA 4,01 A 4,50 MT	2749
L.17-AUTO "AEB" FINO A 4 METRI	2733
SEMESTRALE CEOD A20+ACC.	2686
ATVO+ACTV MENS.LAV.F2	2679
7 DAYS ONLINE AEROBUS AR 1MESE	2668
72H RVENICE+AEROP.CS ONLINE	2601
48H ONLINE NO AEROBUS 1MESE	2486
CARNET CHIOGGIA 10C. TICKET	2263
NAVETTA ARSENALE CA	2251
FERRY17-AUTOCARRI+35Q.	2245
LINEA 17-CATEGORIA A	2212
ABB. MENSILE CHIOGGIA	2118
ANNUALE STUDENTE EXTRA	2056
ATVO+ACTV MENS.STUD.F2	2041
7GGAERCS-TPL55,60-CVE16,40	1952
STUD. RETE INTERA FAMILIARE	1938
ATVO+ACTV MENS.ORD.F1	1902
ATVO CANOVA+ACTV 72H ONLINE	1866
TRAGHETTO RESIDENTE MURANO	1720
TRAGHETTO RESIDENTE S.ERASMO	1714
ABB.STUD.ANN.PELLESTRINA	1678
720RE ONLINE AEROBUS CS	1667
L.11-AUTO "D" OLTRE METRI 4,50	1576
BICICLETTA "BIGLIETTERIA"	1565
LINEA 11+17 CATEGORIA C	1522
BORDO 75MIN CARTAVENEZIA	1466
ABB. OVER 75 A20	1455
ATVOCANOVA+ACTV 72HROLL.ONLINE	1427
AEROPORTO-VENEZIA AR ONLINE	1310
ABB. CHIOGGIA A20	1300
MOBILITY ORDINARIO EXTRA	1294
BIGL.URBANO CHIOGGIA	1287
EXTRA TRATTA 6	1250
DAILY PASS VE. ONLINE 1MESE	1236
SEMESTRALE CEOD A20	1226
480RE ONLINE AEROBUS AR	1223

MENSILE STUD. PELLESTRINA	1221
EXTRA TRATTA 2 TVM	1118
480RE ONLINE AEROBUS CS	1095
TARIFFA CARROZZINA	1069
7 DAYS ONLINE AEROBUS CS	992
LINEA 11-CATEGORIA A	956
TESSERA DI SERVIZIO ARRIVA	945
ABBONAMENTO 30 GG.PEOPLEMOVER	934
72H RVE+AEROP.CS ONLINE 1 MESE	927
ATVO+ACTV ANN.LAV.F1	923
BIGLIETTO SOC. SPORTIVE	918
72H ONLINE AEROBUS AR 1 MESE	903
ARRIVA VENETO TRATTA 1	888
ATVO+ACTV ANN.STUD.F1	817
EXTRA TRATTE 2-3-4 BORDO	796
EXTRA TRATTA 1 TVM	778
ATVO CANOVA+ACTV 72H	777
FERRY11-AUTOCARRI+35Q.	733
72H R.VE.+AER.AR ONLINE 1MESE	648
EXTRA TRATTA 8-9-10	640
L.11-AUTO "AEB" FINO A 4 METRI	630
ABB MENSILE PEOPLEMOVER	612
FERRY17-CARRI+35Q.RIM.	609
SUPP MENS.AUTOMOBILISTICO	607
BICICLETTA "PALMARE"	585
EV8-TPL 52,00-C.VE3,00	573
72H ONLINE AEROBUS CS 1 MESE	561
NA-BIG.AUT.75' MESTRE/LIDO-CSC	528
ATVO+ACTV MENS.ORD.F2	522
LINEA 11+17 CATEGORIA B	509
ABB. ORDINARIO BUS. LIDO	508
EXTRA TRATTA 3 TVM	500
ARRIVA VENETO TRATTA 7	497
EXTRA TRATTA 4 TVM	488
24HAERAR-TPL32,90-CVE5,10	486
T.FUSINA VE+ACTV 24 ORE	479
ABB. STUDENTE MENS. CHIOGGIA	477
LINEA 11+17 CATEGORIA D	474
PEOPLEMOVER+BUS+TRAM 75'CARNET	460
ARRIVA VENETO TRATTA 2	440

240RE ONLINE AEROBUS CS	439
7 DAYS ONLINE AEROBUS CS 1MESE	420
APERTURA TORNELLI DUSSMAN	413
T.FUSINA VE+ACTV 72 ORE	411
L.11-AUTO "C"DA 4,01 A 4,50 MT	408
BIGLIETTO SCUOLE	404
EXTRA TRATTA 7	383
BUS+PEOPLE MOVER ONLINE	370
ARRIVA VENETO TRATTA 4	362
ATVO+ACTV ANN.LAV.F2	354
24H-24 ORE	342
ATVO+ACTV ANN.STUD.F2	327
48H ONLINE AEROBUS CS 1 MESE	322
JESOLO - S.MARCO AR	313
ARRIVA VENETO TRATTA 3	311
NA-CARNET NAV. 10 CORSE DA 75'	301
ABB ANNUALE PEOPLEMOVER	287
AEROPORTO-VENEZIA CS ONLINE	284
ARRIVA VENETO TRATTA 6	283
S.TERR+ACTV ORD. TR.6	267
ATVO+ACTV MENS.20%.F1	250
ABB. IMPERSONALE RETE INTERA	249
ARRIVA VENETO AEROPORTO	243
BICICLETTA "CONCESSIONARI"	229
SUPP. 12 MESI STUDENTE LAGUNA	225
BAGAGLIO CARTAVENEZIA	220
ATVO+ACTV MENS.20%.F2	202
ORD. NAVIG. 75' ONLINE 1 MESE	195
CICLOMOTORE FINO 50CC	194
EXTRA TRATTA 1 BORDO	193
240RE ONLINE AEROBUS AR	181
ATVO CANOVA+NAVIG AR	180
ATVO CANOVA+NAVIG 1 CORSA	177
S.TERR+ACTV ORD. TR.7	146
BIGLIETTO MOTO FINO 50 CC	133
SUPP ANNUALE PEOPLEMOVER	132
ANNUALE CAT. D 17(UN SEMESTRE)	122
AEROBUS+BOAT ONLINE	121
BIGLIETTO MERCI C.SEMPLICE	120
SUPP NAVIGAZIONE FAMILIARE	116

NA75'-TPL 13,28-COMVE1,72	113
MENS. COSE ANIMALI RETE INTERA	111
ARRIVA VENETO TRATTA 5	110
ATVO+ACTV MENS.ORD.F3	109
EXTRA TRATTA 5 TVM	109
MENSILE PARK+RETE INTERA	106
48H ONLINE AEROBUS AR 1 MESE	104
NA-C AUT. 10 CORSE 75' CARD	100
ABB STUD. 12 MESI CHIOGGIA	98
ATVO+ACTV MENS.STUD.F3	96
ARRIVA EXTRA TR.8-9-10 BORDO	93
SUPP. ANNUALE AUTOMOB.	91
S.TERR+ACTV ANN STUD TR.6	88
ABB.CHIOGGIA ANNUALE	85
ARRIVA EXTRA TR.1 BORDO	85
CAV -TREP + ACTV 24H	84
S.TERRR+ACTV STUDENTE TR.6	82
ARRIVA EXTRA TR. 5-6-7 BORDO	80
ABB. EXTRA A5	80
ATVO CANOVA+NAVIG AR ONLINE	79
JESOLO + ACTV 24H	75
ARRIVAEXTRA TR.2-3-4 BORDO	72
ABB. RETE INTERA A20 +ACC.	72
ATVO+ACTV MENS.20%.F3	67
ANNUALE CAT. D LINEA 11	67
AEROBUS+BOAT ONLINE 1MESE	62
MENS. COSE ANIMALI RETE UNICA	61
EXTRA TRATTE 5-6-7 BORDO	55
ATVOCANOVA+NAVIG 1CORSIA ONLINE	49
ARRIVA AEROPORTO 0.MENS	46
LINEA 11+17 CATEGORIA A	45
S.TERR+ACTV ORD. TR.5	44
TRAGHETTO RESIDENTE GIUDECCA	41
S.TERR+ACTV ANN ORD TR.9	40
240RE ONLINE NO AEROBUS	37
S.TERR+ACTV STUDENTE TR.2	36
NA-TRAGHETTO ORDINARIO	35
FERRY17-TRASPORTI PERICOLOSI	33
S.TERR+ACTV ANN STUD TR.7	32
ATVO+ACTV MENS.LAV.F3	32

S.TERR+ACTV ORD. TR.9	29
ATVO+ACTV MENS.5%.F2	27
INTEGRAZIONE RETE MESTRE	26
EXTRA TRATTE 8-9-10 BORDO	25
STUDENTE EXTRA FAMILIARE	25
S.TERR+ACTV STUDENTE TR.7	25
EV5-TPL 38,50-C.VE1,50	23
MENS. STUDENTE BUS LIDO	23
EXTRA TRATTA 7 TVM	22
CAV-TREP - S.MARCO AR	22
FERRY11-CARRI+35Q.RIM.	22
ABB. CHIOGGIA A5	18
ORDINARIO EXTRA FAMILIARE	18
NATRAGH-TPL 8,82-C.VE1,18	17
ORDINARIO CHIOGGIA FAMILIARE	16
CICLOMOTORE OLTRE 50CC	15
FERRY17-AUTOBUS	14
EXTRA TRATTA 6 TVM	14
CARNET CHIOGGIA 10 C. CARD	14
S.TERR+ACTV STUDENTE TR.5	13
ANNUALE ORDINARIO BUS LIDO	13
S.TERR+ACTV ANN STUD TR.5	12
LIBERA CIRC. FERRY LINEA 17	11
S.TERR+ACTV ANN ORD TR.2	11
NA-BICICLETTA E CONDUCENTE CV	10
LIBERA CIRC. FERRY LINEA 17+11	10
VENDITA A BORDO 75' CV	9
SUPP MENS.URBANO CHIOGGIA	9
ARRIVA AEROPORTO BORDO	9
EXTRA TRATTA 8-9-10 TVM	9
PRENOTAZIONE VEICOLO ABBONATO	9
S.TERR+ACTV ANN STUD TR.8	9
SUPP MENSILE PEOPLEMOVER	8
LIBERA CIRC. FERRY LINEA 11	8
S.TERR+ACTV ANN ORD TR.6	8
NATRAGH-TPL 4,41-C.VE0,59	8
S.TERR+ACTV ANN ORD TR.8	7
STUDENTE CHIOGGIA FAMILIARE	7
NA-GRUPPI E SCUOLE	7
S.TERR+ACTV ORD. TR.8	7

FERRY11-TRASPORTI PERICOLOSI	7
BIGLIETTO SCUOLE ONLINE	6
24H ONLINE AEROBUS CS 1 MESE	6
ACC.L.R.A20 RETE INTERA	6
SUPP. 12 MESI STUDENTE AUTOMOB	5
MOBILITY STUDENTE RETEUNICA	5
ABB STUDENTE BUS LIDO 12 MESI	5
S.TERR+ACTV STUDENTE TR.3	5
DDGR1201-1297/2022 R. UNICA	5
GRUPPI E SCUOLE ONLINE 2VIAGGI	4
ARRIVA INTEG.AEROP. BORDO	4
TRAGHETTO GRATUITO	4
S.TERR+ACTV ORD. TR.3	4
S.TERR+ACTV ANN ORD TR.7	4
S.TERR+ACTV ORD. TR.2	4
ABB. CHIOGGIA A20 + ACC.	4
75'-TPL 12,60-CVE2,40 ONLINE	3
S.TERR+ACTV ANN STUD TR.2	3
S.TERR+ACTV ORD. TR.4	3
SPIAGGEAR-TPL 14,75-COMVE1,25	3
ARRIVA MISTO ACTV 24H	3
S.TERR+ACTV STUDENTE TR.4	2
ABB. OVER 75 A5	2
S.TERR+ACTV STUDENTE TR.8	2
S.TERR+ACTV STUDENTE TR.9	2
EV3-TPL 30,50-C.VE1,50	2
DDGR1201-1297/2022 EXTRA	2
SUPPL. RETE MESTRE ATVO	1
SUPP MESTRE FAMILIARE	1
APERTURA TORNELLI P.SABBIONI	1
NA-C CHIOGGIA 10 C. CARD FS	1
S.TERR+ACTV ANN ORD TR.3	1
ANNUALE PARK+RETE INTERA	1
NA-BIGL. CHIOGGIA CARD	1
S.TERR+ACTV ANN STUD TR.3	1
Name: DESCRIZIONE_TITOLO, dtype: int64	

One-day tickets

```
In [ ]: # Which type of ticket are one-day tickets and how many are there?
df[df['DESCRIZIONE_TITOLO'].str.contains('GIORNALIERO|24H|24ORE|24 ORE|DAILY')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out [ ]: DAILYP-TPL19,90-C.VE5,10      522245
DAILY PASS VENEZIA ONLINE          17424
24HAERCS-TPL26,90-CVE5,10          4181
DAILY PASS VE. ONLINE 1MESE        1236
24HAERAR-TPL32,90-CVE5,10           486
T.FUSINA VE+ACTV 24 ORE             479
24ORE ONLINE AEROBUS CS             439
24H-24 ORE                          342
24ORE ONLINE AEROBUS AR             181
CAV -TREP + ACTV 24H                84
JESOLO + ACTV 24H                   75
24ORE ONLINE NO AEROBUS             37
24H ONLINE AEROBUS CS 1 MESE         6
ARRIVA MISTO ACTV 24H               3
Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('GIORNALIERO|24H|24ORE|24 ORE|DAILY'), 'TICKET_CODE'] = '1'
```

```
In [ ]: # TICKET_CODE = 1: Information about one-day tickets
print("The number of one-day tickets is: ", df[df['TICKET_CODE'] == '1'].shape[0])
print("The number of tickets for each type of one-day ticket is: ")
df[df['TICKET_CODE'] == '1']['DESCRIZIONE_TITOLO'].value_counts()
```

```
The number of one-day tickets is: 547218
The number of tickets for each type of one-day ticket is:
```

```
Out[ ]: DAILYP-TPL19,90-C.VE5,10      522245
        DAILY PASS VENEZIA ONLINE      17424
        24HAERCS-TPL26,90-CVE5,10      4181
        DAILY PASS VE. ONLINE 1MESE     1236
        24HAERAR-TPL32,90-CVE5,10      486
        T.FUSINA VE+ACTV 24 ORE         479
        240RE ONLINE AEROBUS CS         439
        24H-24 ORE                      342
        240RE ONLINE AEROBUS AR         181
        CAV -TREP + ACTV 24H            84
        JESOLO + ACTV 24H               75
        240RE ONLINE NO AEROBUS         37
        24H ONLINE AEROBUS CS 1 MESE     6
        ARRIVA MISTO ACTV 24H           3
        Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: print("Information about the tickets with code 1 related to the serial number: ")
        df[df['TICKET_CODE'] == '1'].groupby('DESCRIZIONE_TITOLO')['SERIALE'].value_counts().groupby('DESCRIZIONE_TITOLO').
```

Information about the tickets with code 1 related to the serial number:

Out []:

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
24H ONLINE AEROBUS CS 1 MESE	3.0	2.000000	1.732051	1.0	1.00	1.0	2.50	4.0
24H-24 ORE	141.0	2.425532	1.631100	1.0	1.00	2.0	3.00	10.0
24HAERAR-TPL32,90-CVE5,10	113.0	4.300885	2.942560	1.0	2.00	4.0	7.00	13.0
24HAERCS-TPL26,90-CVE5,10	939.0	4.452609	2.570673	1.0	3.00	4.0	6.00	21.0
24ORE ONLINE AEROBUS AR	41.0	4.414634	3.154169	1.0	3.00	4.0	5.00	19.0
24ORE ONLINE AEROBUS CS	86.0	5.104651	1.958565	1.0	4.00	5.0	6.00	11.0
24ORE ONLINE NO AEROBUS	13.0	2.846154	1.281025	1.0	2.00	3.0	3.00	6.0
ARRIVA MISTO ACTV 24H	2.0	1.500000	0.707107	1.0	1.25	1.5	1.75	2.0
CAV -TREP + ACTV 24H	22.0	3.818182	1.622355	2.0	2.00	4.0	4.75	7.0
DAILY PASS VE. ONLINE 1MESE	297.0	4.161616	2.313476	1.0	2.00	4.0	6.00	12.0
DAILY PASS VENEZIA ONLINE	4432.0	3.931408	2.187197	1.0	2.00	4.0	5.00	16.0
DAILYP-TPL19,90-C.VE5,10	130117.0	4.013657	2.216815	1.0	2.00	4.0	5.00	74.0
JESOLO + ACTV 24H	24.0	3.125000	1.226962	2.0	2.00	3.0	4.00	6.0
T.FUSINA VE+ACTV 24 ORE	135.0	3.548148	2.028596	1.0	2.00	3.0	4.00	10.0

Two days tickets

```
In [ ]: # Which type of ticket are two-day tickets and how many are there?
df[df['DESCRIZIONE_TITOLO'].str.contains('48H|48ORE|48 ORE')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out[ ]: 48H-TPL 29,90-COMVE5,10      278703
        480RE ONLINE NO AEROBUS      31472
        48HAERCS-TPL36,90-CVE5,10    4735
        48HAERAR-TPL42,90-CVE5,10    2881
        48H ONLINE NO AEROBUS 1MESE   2486
        480RE ONLINE AEROBUS AR       1223
        480RE ONLINE AEROBUS CS       1095
        48H ONLINE AEROBUS CS 1 MESE   322
        48H ONLINE AEROBUS AR 1 MESE   104
        Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('48H|480RE|48 ORE'), 'TICKET_CODE'] = '2'
```

```
In [ ]: # TICKET_CODE = 2: Information about two-day tickets
print("The number of two-day tickets is: ", df[df['TICKET_CODE'] == '2'].shape[0])
print("The number of tickets for each type of two-day ticket is: ")
df[df['TICKET_CODE'] == '2']['DESCRIZIONE_TITOLO'].value_counts()
```

```
The number of two-day tickets is: 323021
The number of tickets for each type of two-day ticket is:
```

```
Out[ ]: 48H-TPL 29,90-COMVE5,10      278703
        480RE ONLINE NO AEROBUS      31472
        48HAERCS-TPL36,90-CVE5,10    4735
        48HAERAR-TPL42,90-CVE5,10    2881
        48H ONLINE NO AEROBUS 1MESE   2486
        480RE ONLINE AEROBUS AR       1223
        480RE ONLINE AEROBUS CS       1095
        48H ONLINE AEROBUS CS 1 MESE   322
        48H ONLINE AEROBUS AR 1 MESE   104
        Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: print("Information about the tickets with code 2 related to the serial number: ")
df[df['TICKET_CODE'] == '2'].groupby('DESCRIZIONE_TITOLO')['SERIALE'].value_counts().groupby('DESCRIZIONE_TITOLO').
Information about the tickets with code 2 related to the serial number:
```


Out []:

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
48H ONLINE AEROBUS AR 1 MESE	15.0	6.933333	3.575046	3.0	4.50	6.0	9.0	17.0
48H ONLINE AEROBUS CS 1 MESE	38.0	8.473684	3.523948	2.0	6.25	8.0	12.0	15.0
48H ONLINE NO AEROBUS 1MESE	359.0	6.924791	3.373810	1.0	5.00	7.0	9.0	19.0
48H-TPL 29,90-COMVE5,10	41726.0	6.679361	3.574409	1.0	4.00	6.0	9.0	41.0
48HAERAR-TPL42,90-CVE5,10	363.0	7.936639	3.602310	1.0	6.00	8.0	10.0	20.0
48HAERCS-TPL36,90-CVE5,10	590.0	8.025424	3.705779	1.0	5.00	8.0	11.0	21.0
48ORE ONLINE AEROBUS AR	150.0	8.153333	3.333251	3.0	6.00	8.0	10.0	18.0
48ORE ONLINE AEROBUS CS	136.0	8.051471	3.754148	1.0	6.00	8.0	10.0	19.0
48ORE ONLINE NO AEROBUS	4885.0	6.442579	3.523849	1.0	4.00	6.0	9.0	32.0

Three days tickets

```
In [ ]: # Which type of ticket are three-day tickets and how many are there?
# Do not consider the ticket that contains also 75
df[df['DESCRIZIONE_TITOLO'].str.contains('72H|72ORE|72 ORE')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out[ ]: 72H-TPL 38,40-COMVE6,60      199789
        BIGLIETTO 72 ORE ROLL. VENICE 62143
        72ORE ONLINE NO AEROBUS      56474
        72 ORE R.VENICE ONLINE      31590
        72HAERAR-TPL51,40-CVE6,60    9459
        72 ORE R.VENICE+AEROPORTO AR  6604
        72HAERCS-TPL45,40-CVE6,60    6011
        72 ORE R.VENICE+AEROPORTO CS  5965
        72H ONLINE NO AEROBUS 1MESE  5218
        72ORE ONLINE AEROBUS AR      5011
        72H R.VENICE ONLINE 1 MESE    4948
        72H R.VENICE+AEROP.AR ONLINE 4242
        72H RVENICE+AEROP.CS ONLINE   2601
        ATVO CANOVA+ACTV 72H ONLINE   1866
        72ORE ONLINE AEROBUS CS      1667
        ATVOCANOVA+ACTV 72HROLL.ONLINE 1427
        72H RVE+AEROP.CS ONLINE 1 MESE 927
        72H ONLINE AEROBUS AR 1 MESE  903
        ATVO CANOVA+ACTV 72H          777
        72H R.VE.+AER.AR ONLINE 1MESE 648
        72H ONLINE AEROBUS CS 1 MESE  561
        T.FUSINA VE+ACTV 72 ORE      411
        Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('72H|72ORE|72 ORE'), 'TICKET_CODE'] = '3'
```

```
In [ ]: # TICKET_CODE = 3: Information about three-day tickets
print("The number of three-day tickets is: ", df[df['TICKET_CODE'] == '3'].shape[0])
print("The number of tickets for each type of three-day ticket is: ")
df[df['TICKET_CODE'] == '3']['DESCRIZIONE_TITOLO'].value_counts()
```

The number of three-day tickets is: 409242

The number of tickets for each type of three-day ticket is:

```
Out[ ]: 72H-TPL 38,40-COMVE6,60      199789
        BIGLIETTO 72 ORE ROLL. VENICE 62143
        720RE ONLINE NO AEROBUS      56474
        72 ORE R.VENICE ONLINE      31590
        72HAERAR-TPL51,40-CVE6,60    9459
        72 ORE R.VENICE+AEROPORTO AR 6604
        72HAERCS-TPL45,40-CVE6,60    6011
        72 ORE R.VENICE+AEROPORTO CS 5965
        72H ONLINE NO AEROBUS 1MESE 5218
        720RE ONLINE AEROBUS AR      5011
        72H R.VENICE ONLINE 1 MESE   4948
        72H R.VENICE+AEROP.AR ONLINE 4242
        72H RVENICE+AEROP.CS ONLINE 2601
        ATV0 CANOVA+ACTV 72H ONLINE 1866
        720RE ONLINE AEROBUS CS      1667
        ATV0CANOVA+ACTV 72HROLL.ONLINE 1427
        72H RVE+AEROP.CS ONLINE 1 MESE 927
        72H ONLINE AEROBUS AR 1 MESE 903
        ATV0 CANOVA+ACTV 72H         777
        72H R.VE.+AER.AR ONLINE 1MESE 648
        72H ONLINE AEROBUS CS 1 MESE 561
        T.FUSINA VE+ACTV 72 ORE      411
        Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: print("Information about the tickets with code 3 related to the serial number: ")
        df[df['TICKET_CODE'] == '3'].groupby('DESCRIZIONE_TITOLO')['SERIALE'].value_counts().groupby('DESCRIZIONE_TITOLO').
```

Information about the tickets with code 3 related to the serial number:

Out[]:

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
72 ORE R.VENICE ONLINE	3663.0	8.624079	4.572296	1.0	6.00	8.0	11.0	47.0
72 ORE R.VENICE+AEROPORTO AR	600.0	11.006667	5.128723	1.0	7.00	11.0	14.0	39.0
72 ORE R.VENICE+AEROPORTO CS	600.0	9.941667	4.176602	1.0	7.00	10.0	13.0	26.0
72H ONLINE AEROBUS AR 1 MESE	78.0	11.576923	5.041437	2.0	7.25	10.0	15.0	28.0
72H ONLINE AEROBUS CS 1 MESE	61.0	9.196721	3.771912	3.0	6.00	9.0	12.0	19.0
72H ONLINE NO AEROBUS 1MESE	531.0	9.826742	5.009436	1.0	6.00	10.0	13.0	27.0
72H R.VE.+AER.AR ONLINE 1MESE	62.0	10.451613	5.621030	1.0	6.00	11.0	14.0	24.0
72H R.VENICE ONLINE 1 MESE	597.0	8.288107	5.081953	1.0	5.00	8.0	10.0	29.0
72H R.VENICE+AEROP.AR ONLINE	393.0	10.793893	4.927615	1.0	8.00	10.0	14.0	32.0
72H RVE+AEROP.CS ONLINE 1 MESE	100.0	9.270000	4.496587	1.0	6.00	9.0	12.0	23.0
72H RVENICE+AEROP.CS ONLINE	283.0	9.190813	4.577038	1.0	6.00	9.0	12.0	28.0
72H-TPL 38,40-COMVE6,60	23186.0	8.616795	4.641303	1.0	5.00	8.0	11.0	60.0
72HAERAR-TPL51,40-CVE6,60	882.0	10.724490	4.461215	1.0	8.00	10.0	13.0	36.0
72HAERCS-TPL45,40-CVE6,60	592.0	10.153716	4.953081	1.0	7.00	10.0	13.0	32.0
72ORE ONLINE AEROBUS AR	509.0	9.844794	4.467895	1.0	7.00	10.0	13.0	31.0
72ORE ONLINE AEROBUS CS	195.0	8.548718	4.295939	1.0	5.00	8.0	11.5	21.0
72ORE ONLINE NO AEROBUS	6594.0	8.564453	4.691662	1.0	5.00	8.0	11.0	76.0
ATVO CANOVA+ACTV 72H	83.0	9.361446	4.192264	1.0	6.00	9.0	12.0	21.0
ATVO CANOVA+ACTV 72H ONLINE	213.0	8.760563	4.849036	1.0	5.00	8.0	11.0	26.0
ATVOCANOVA+ACTV 72HROLL.ONLINE	159.0	8.974843	4.173378	1.0	6.00	9.0	11.0	21.0
BIGLIETTO 72 ORE ROLL. VENICE	7452.0	8.339104	4.740000	1.0	5.00	8.0	11.0	72.0
T.FUSINA VE+ACTV 72 ORE	58.0	7.086207	4.882063	1.0	3.00	7.0	10.0	22.0

Seven days tickets

```
In [ ]: # Which type of ticket are weekly tickets and how many are there?
# Exclude the tickets that contains also 72, 75 that are three-day tickets, 17, 48h, 57 that are other types of tic
# 'tratt*' and 'tr' that are reserved to specific routes
df[df['DESCRIZIONE_TITOLO'].str.contains('7GG|7DAYS|7 DAYS')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out [ ]: 7GG-TPL 48,60-COMVE16,40      108925
7 DAYS ONLINE NO AEROBUS             34462
7GGAERAR-TPL61,60-CVE16,40          10113
7 DAYS ONLINE AEROBUS AR              6906
7 DAYS ONLINE NO AEROBUS 1MESE        4821
7 DAYS ONLINE AEROBUS AR 1MESE        2668
7GGAERCS-TPL55,60-CVE16,40           1952
7 DAYS ONLINE AEROBUS CS               992
7 DAYS ONLINE AEROBUS CS 1MESE         420
Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('7GG|7DAYS|7 DAYS'), 'TICKET_CODE'] = '4'
```

```
In [ ]: # TICKET_CODE = 4: Information about weekly tickets
print("The number of weekly tickets is: ", df[df['TICKET_CODE'] == '4'].shape[0])
print("The number of tickets for each type of weekly ticket is: ")
df[df['TICKET_CODE'] == '4']['DESCRIZIONE_TITOLO'].value_counts()
```

The number of weekly tickets is: 171259
The number of tickets for each type of weekly ticket is:

```
Out [ ]: 7GG-TPL 48,60-COMVE16,40      108925
7 DAYS ONLINE NO AEROBUS             34462
7GGAERAR-TPL61,60-CVE16,40          10113
7 DAYS ONLINE AEROBUS AR              6906
7 DAYS ONLINE NO AEROBUS 1MESE        4821
7 DAYS ONLINE AEROBUS AR 1MESE        2668
7GGAERCS-TPL55,60-CVE16,40           1952
7 DAYS ONLINE AEROBUS CS               992
7 DAYS ONLINE AEROBUS CS 1MESE         420
Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: print("Information about the tickets with code 4 related to the serial number: ")
df[df['TICKET_CODE'] == '4'].groupby('DESCRIZIONE_TITOLO')['SERIALE'].value_counts().groupby('DESCRIZIONE_TITOLO').
```

Information about the tickets with code 4 related to the serial number:

```
Out [ ]:
```

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
7 DAYS ONLINE AEROBUS AR	476.0	14.508403	7.619223	1.0	10.00	14.0	19.0	65.0
7 DAYS ONLINE AEROBUS AR 1MESE	175.0	15.245714	6.096180	1.0	11.00	15.0	19.0	32.0
7 DAYS ONLINE AEROBUS CS	77.0	12.883117	7.998312	1.0	8.00	12.0	17.0	47.0
7 DAYS ONLINE AEROBUS CS 1MESE	27.0	15.555556	9.336996	3.0	9.00	11.0	26.0	31.0
7 DAYS ONLINE NO AEROBUS	2684.0	12.839791	8.234205	1.0	7.00	12.0	17.0	146.0
7 DAYS ONLINE NO AEROBUS 1MESE	327.0	14.743119	7.606589	1.0	9.00	14.0	19.0	43.0
7GG-TPL 48,60-COMVE16,40	7782.0	13.997044	8.506889	1.0	8.00	13.0	19.0	107.0
7GGAERAR-TPL61,60-CVE16,40	624.0	16.206731	7.786098	1.0	11.00	16.0	21.0	45.0
7GGAERCS-TPL55,60-CVE16,40	122.0	16.000000	8.888659	1.0	9.25	16.0	21.0	53.0

Monthly tickets

```
In [ ]: # Whick type of ticket are monthly tickets and how many are there?
df[df['DESCRIZIONE_TITOLO'].str.contains('MENSILE|30GG|30 GG|MENS')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out[ ]: MENSILE ORDINARIO RETE UNICA      753855
        MENS.STUDENTE RETE UNICA        123083
        MENSILE ORDINARIO ISOLE          85577
        MENSILE ORDINARIO EXTRA          27239
        SUPP MENS.NAVIGAZIONE            24998
        MENSILE ORD. RES. PELLESTRINA     9645
        MENSILE STUDENTE ISOLE           6614
        MENSILE STUDENTE EXTRA           4810
        ATVO+ACTV MENS.LAV.F1             4562
        ATVO+ACTV MENS.STUD.F1           2967
        ATVO+ACTV MENS.LAV.F2            2679
        ABB. MENSILE CHIOGGIA             2118
        ATVO+ACTV MENS.STUD.F2            2041
        ATVO+ACTV MENS.ORD.F1             1902
        MENSILE STUD. PELLESTRINA         1221
        ABBONAMENTO 30 GG.PEOPLEMOVER    934
        ABB MENSILE PEOPLEMOVER           612
        SUPP MENS.AUTOMOBILISTICO        607
        ATVO+ACTV MENS.ORD.F2             522
        ABB. STUDENTE MENS. CHIOGGIA      477
        ATVO+ACTV MENS.20%.F1             250
        ATVO+ACTV MENS.20%.F2            202
        MENS. COSE ANIMALI RETE INTERA    111
        ATVO+ACTV MENS.ORD.F3             109
        MENSILE PARK+RETE INTERA          106
        ATVO+ACTV MENS.STUD.F3             96
        ATVO+ACTV MENS.20%.F3             67
        MENS. COSE ANIMALI RETE UNICA     61
        ARRIVA AEROPORTO 0.MENS           46
        ATVO+ACTV MENS.LAV.F3             32
        ATVO+ACTV MENS.5%.F2              27
        MENS. STUDENTE BUS LIDO           23
        SUPP MENS.URBANO CHIOGGIA          9
        SUPP MENSILE PEOPLEMOVER           8
        Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('MENSILE|30GG|30 GG|MENS'), 'TICKET_CODE'] = '5'
```

```
In [ ]: # If DESCRIZIONE_TITOLO contains 'STUDENTE' or 'STUD' update the column 'TICKET_CODE' with '5-STUD' only for the ti
df.loc[(df['TICKET_CODE'] == '5') & (df['DESCRIZIONE_TITOLO'].str.contains('STUDENTE|STUD|STUD')), 'TICKET_CODE'] =
# If DESCRIZIONE_TITOLO contains 'LAVORATORE' or 'LAV' update the column 'TICKET_CODE' with '5-WKRS' only for the t
df.loc[(df['TICKET_CODE'] == '5') & (df['DESCRIZIONE_TITOLO'].str.contains('LAVORATORE|LAV|LAV')), 'TICKET_CODE'] =
# If DESCRIZIONE_TITOLO contains 'OVER 65' or '65+' or 'PENSIONATI' update the column 'TICKET_CODE' with '5-RET' on
df.loc[(df['TICKET_CODE'] == '5') & (df['DESCRIZIONE_TITOLO'].str.contains('OVER 65|65+|PENSIONATI')), 'TICKET_CODE']
```

```
In [ ]: # According to the page of agevolation of specific categories of people available at the site web of ACTV
# (https://actv.avmspa.it/it/content/categorie-agevolate-0), the DDGR 1201-1297/2022 tickets are monthly tickets fo

# Which type of ticket are yearly tickets for blind people and how many are there?
df[df['DESCRIZIONE_TITOLO'].str.contains('DDGR1201-1297/2022')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out [ ]: DDGR1201-1297/2022 R. UNICA      5
          DDGR1201-1297/2022 EXTRA      2
          Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('DDGR1201-1297/2022'), 'TICKET_CODE'] = '5'
```

```
In [ ]: # TICKET_CODE = 5: Information about monthly tickets
print("The number of monthly tickets is: ", df[df['TICKET_CODE'] == '5'].shape[0])
print("The number of monthly tickets for students is: ", df[df['TICKET_CODE'] == '5-STUD'].shape[0])
print("The number of monthly tickets for workers is: ", df[df['TICKET_CODE'] == '5-WKRS'].shape[0])
print("The number of monthly tickets for retired people is: ", df[df['TICKET_CODE'] == '5-RET'].shape[0])

print("The number of tickets for each type of monthly ticket (including the subtypes) is: ")
df[df['TICKET_CODE'].isin(['5', '5-STUD', '5-WKRS', '5-RET'])].groupby('TICKET_CODE')['DESCRIZIONE_TITOLO'].value_c
```

```
The number of monthly tickets is: 909012
The number of monthly tickets for students is: 141332
The number of monthly tickets for workers is: 7273
The number of monthly tickets for retired people is: 0
The number of tickets for each type of monthly ticket (including the subtypes) is:
```



```

Out[ ]: TICKET_CODE DESCRIZIONE_TITOLO
5          MENSILE ORDINARIO RETE UNICA      753855
          MENSILE ORDINARIO ISOLE            85577
          MENSILE ORDINARIO EXTRA            27239
          SUPP MENS.NAVIGAZIONE              24998
          MENSILE ORD. RES. PELLESTRINA      9645
          ABB. MENSILE CHIOGGIA              2118
          ATVO+ACTV MENS.ORD.F1              1902
          ABBONAMENTO 30 GG.PEOPLEMOVER      934
          ABB MENSILE PEOPLEMOVER            612
          SUPP MENS.AUTOMOBILISTICO          607
          ATVO+ACTV MENS.ORD.F2              522
          ATVO+ACTV MENS.20%.F1              250
          ATVO+ACTV MENS.20%.F2              202
          MENS. COSE ANIMALI RETE INTERA     111
          ATVO+ACTV MENS.ORD.F3              109
          MENSILE PARK+RETE INTERA           106
          ATVO+ACTV MENS.20%.F3              67
          MENS. COSE ANIMALI RETE UNICA      61
          ARRIVA AEROPORTO 0.MENS            46
          ATVO+ACTV MENS.5%.F2               27
          SUPP MENS.URBANO CHIOGGIA           9
          SUPP MENSILE PEOPLEMOVER            8
          DDGR1201-1297/2022 R. UNICA        5
          DDGR1201-1297/2022 EXTRA           2
5-STUD    MENS.STUDENTE RETE UNICA          123083
          MENSILE STUDENTE ISOLE             6614
          MENSILE STUDENTE EXTRA             4810
          ATVO+ACTV MENS.STUD.F1             2967
          ATVO+ACTV MENS.STUD.F2             2041
          MENSILE STUD. PELLESTRINA          1221
          ABB. STUDENTE MENS. CHIOGGIA       477
          ATVO+ACTV MENS.STUD.F3              96
          MENS. STUDENTE BUS LIDO             23
5-WKRS    ATVO+ACTV MENS.LAV.F1             4562
          ATVO+ACTV MENS.LAV.F2             2679
          ATVO+ACTV MENS.LAV.F3              32

```

Name: DESCRIZIONE_TITOLO, dtype: int64

```
In [ ]: print("Information about the tickets with code 5 (including the subtypes) related to the serial number: ")
df[df['TICKET_CODE'].isin(['5', '5-STUD', '5-WKRS', '5-RET'])].groupby(['TICKET_CODE', 'DESCRIZIONE_TITOLO'])['SERI
```

Information about the tickets with code 5 (including the subtypes) related to the serial number:

Out[]:

		count	mean	std	min	25%	50%	75%	max
TICKET_CODE	DESCRIZIONE_TITOLO								
5	ABB MENSILE PEOPLEMOVER	12.0	51.000000	23.916521	2.0	38.50	59.5	65.25	81.0
	ABB. MENSILE CHIOGGIA	163.0	12.993865	15.593009	1.0	3.00	8.0	18.00	104.0
	ABBONAMENTO 30 GG.PEOPLEMOVER	81.0	11.530864	13.244892	1.0	1.00	3.0	20.00	45.0
	ARRIVA AEROPORTO O.MENS	3.0	15.333333	11.590226	3.0	10.00	17.0	21.50	26.0
	ATVO+ACTV MENS.20%.F1	8.0	31.250000	42.005952	1.0	1.00	7.0	53.75	104.0
	ATVO+ACTV MENS.20%.F2	7.0	28.857143	30.975413	5.0	11.00	19.0	30.50	95.0
	ATVO+ACTV MENS.20%.F3	3.0	22.333333	29.365513	2.0	5.50	9.0	32.50	56.0
	ATVO+ACTV MENS.5%.F2	1.0	27.000000	NaN	27.0	27.00	27.0	27.00	27.0
	ATVO+ACTV MENS.ORD.F1	46.0	41.347826	34.092826	1.0	10.25	39.5	69.50	110.0
	ATVO+ACTV MENS.ORD.F2	11.0	47.454545	35.103742	5.0	26.00	39.0	66.00	109.0
	ATVO+ACTV MENS.ORD.F3	2.0	54.500000	13.435029	45.0	49.75	54.5	59.25	64.0
	DDGR1201-1297/2022 EXTRA	1.0	2.000000	NaN	2.0	2.00	2.0	2.00	2.0
	DDGR1201-1297/2022 R. UNICA	5.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
	MENS. COSE ANIMALI RETE INTERA	17.0	6.529412	11.801047	1.0	1.00	1.0	7.00	49.0
	MENS. COSE ANIMALI RETE UNICA	5.0	12.200000	10.986355	1.0	1.00	15.0	18.00	26.0
	MENSILE ORD. RES. PELLESTRINA	905.0	10.657459	13.825629	1.0	3.00	6.0	12.00	125.0
	MENSILE ORDINARIO EXTRA	2278.0	11.957419	20.620575	1.0	1.00	3.0	13.00	348.0
	MENSILE ORDINARIO ISOLE	4288.0	19.957323	23.656477	1.0	4.00	11.0	28.00	542.0
	MENSILE ORDINARIO RETE UNICA	31806.0	23.701660	27.568725	1.0	4.00	13.0	35.00	318.0
	MENSILE PARK+RETE INTERA	6.0	17.666667	21.341665	1.0	3.50	7.0	28.50	53.0
	SUPP MENS.AUTOMOBILISTICO	67.0	9.059701	15.227448	1.0	1.00	2.0	7.00	63.0
	SUPP MENS.NAVIGAZIONE	872.0	28.667431	25.505128	1.0	7.00	22.0	45.00	150.0
	SUPP MENS.URBANO CHIOGGIA	6.0	1.500000	0.836660	1.0	1.00	1.0	1.75	3.0

		count	mean	std	min	25%	50%	75%	max
TICKET_CODE	DESCRIZIONE_TITOLO								
5-STUD	SUPP MENSILE PEOPLEMOVER	2.0	4.000000	4.242641	1.0	2.50	4.0	5.50	7.0
	ABB. STUDENTE MENS. CHIOGGIA	73.0	6.534247	12.886343	1.0	1.00	2.0	6.00	83.0
	ATVO+ACTV MENS.STUD.F1	100.0	29.670000	28.253561	1.0	5.75	21.5	48.25	125.0
	ATVO+ACTV MENS.STUD.F2	63.0	32.396825	24.172458	1.0	14.00	28.0	53.50	93.0
	ATVO+ACTV MENS.STUD.F3	5.0	19.200000	18.322118	1.0	3.00	22.0	24.00	46.0
	MENS. STUDENTE BUS LIDO	11.0	2.090909	1.758098	1.0	1.00	2.0	2.00	7.0
	MENS.STUDENTE RETE UNICA	8949.0	13.753827	19.026200	1.0	2.00	6.0	17.00	164.0
	MENSILE STUD. PELLESTRINA	144.0	8.479167	12.034896	1.0	2.00	4.0	9.00	73.0
	MENSILE STUDENTE EXTRA	1179.0	4.079729	7.479599	1.0	1.00	2.0	3.00	80.0
	MENSILE STUDENTE ISOLE	494.0	13.388664	18.326395	1.0	3.00	6.0	16.00	156.0
5-WKRS	ATVO+ACTV MENS.LAV.F1	130.0	35.092308	31.207842	1.0	8.00	24.5	57.75	136.0
	ATVO+ACTV MENS.LAV.F2	53.0	50.547170	37.013805	1.0	21.00	41.0	73.00	156.0
	ATVO+ACTV MENS.LAV.F3	1.0	32.000000	NaN	32.0	32.00	32.0	32.00	32.0

Yearly tickets

```
In [ ]: # Which type of ticket are yearly tickets and how many are there?
df[df['DESCRIZIONE_TITOLO'].str.contains('ANNUALE|ANN|12MESI|12 MESI')]['DESCRIZIONE_TITOLO'].value_counts()
```

```

Out[ ]: ANNUALE ORDINARIO RETE UNICA      311494
        ANNUALE ORDINARIO ISOLE          74351
        ABB STUD. RETEUNICA 12 MESI       65034
        ANNUALE STUDENTE ISOLE           9965
        ANNUALE ORD.RES.PELLESTRINA      5830
        SUPP.ANNUALE NAVIGAZIONE         3583
        ANNUALE ORDINARIO EXTRA          2900
        ANNUALE STUDENTE EXTRA           2056
        ABB.STUD.ANN.PELLESTRINA         1678
        ATVO+ACTV ANN.LAV.F1             923
        ATVO+ACTV ANN.STUD.F1            817
        ATVO+ACTV ANN.LAV.F2             354
        ATVO+ACTV ANN.STUD.F2            327
        ABB ANNUALE PEOPLEMOVER          287
        SUPP. 12 MESI STUDENTE LAGUNA     225
        SUPP ANNUALE PEOPLEMOVER         132
        ANNUALE CAT. D 17(UN SEMESTRE)   122
        ABB STUD. 12 MESI CHIOGGIA       98
        SUPP. ANNUALE AUTOMOB.           91
        S.TERR+ACTV ANN STUD TR.6         88
        ABB.CHIOGGIA ANNUALE             85
        ANNUALE CAT. D LINEA 11           67
        S.TERR+ACTV ANN ORD TR.9          40
        S.TERR+ACTV ANN STUD TR.7         32
        ANNUALE ORDINARIO BUS LIDO        13
        S.TERR+ACTV ANN STUD TR.5         12
        S.TERR+ACTV ANN ORD TR.2          11
        S.TERR+ACTV ANN STUD TR.8          9
        S.TERR+ACTV ANN ORD TR.6           8
        S.TERR+ACTV ANN ORD TR.8           7
        ABB STUDENTE BUS LIDO 12 MESI     5
        SUPP. 12 MESI STUDENTE AUTOMOB    5
        S.TERR+ACTV ANN ORD TR.7          4
        S.TERR+ACTV ANN STUD TR.2         3
        ANNUALE PARK+RETE INTERA         1
        S.TERR+ACTV ANN ORD TR.3          1
        S.TERR+ACTV ANN STUD TR.3         1
        Name: DESCRIZIONE_TITOLO, dtype: int64

```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('ANNUALE|ANN|12MESI|12 MESI'), 'TICKET_CODE'] = '6'
```

```
In [ ]: # If DESCRIZIONE_TITOLO contains 'STUDENTE' or 'STUD' update the column 'TICKET_CODE' with '6-STUD' only for the ti
df.loc[(df['TICKET_CODE'] == '6') & (df['DESCRIZIONE_TITOLO'].str.contains('STUDENTE|STUD|STUD')), 'TICKET_CODE'] =
# If DESCRIZIONE_TITOLO contains 'LAVORATORE' or 'LAV' update the column 'TICKET_CODE' with '6-WKRS' only for the t
df.loc[(df['TICKET_CODE'] == '6') & (df['DESCRIZIONE_TITOLO'].str.contains('LAVORATORE|LAV|LAV')), 'TICKET_CODE'] =
# If DESCRIZIONE_TITOLO contains 'OVER 65' or '65+' or 'PENSIONATI' update the column 'TICKET_CODE' with '6-RET' on
df.loc[(df['TICKET_CODE'] == '6') & (df['DESCRIZIONE_TITOLO'].str.contains('OVER 65|65+|PENSIONATI')), 'TICKET_CODE']
```

```
In [ ]: # According to the page of agevolation of specific categories of people available at the site web of ACTV
# (https://actv.avmspa.it/it/content/categorie-agevolate-0), the for OVER 75 are yearly tickets for free

# Which type of ticket are yearly tickets for OVER 75 and how many are there?
df[df['DESCRIZIONE_TITOLO'].str.contains('OVER 75|OVER75|PENSIONATI')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out[ ]: ABB. OVER75 GRATUITO          15349
ABBONAMENTO PENSIONATI ACTV         8187
ABB. OVER75 RETE UNICA 50%          5294
ABB. OVER 75 A20                    1455
ABB. OVER 75 A5                      2
Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('OVER 75|OVER75|PENSIONATI'), 'TICKET_CODE'] = '6-RET'
```

```
In [ ]: # According to the page of yearly tickets available at the site web of ACTV
# (https://actv.avmspa.it/it/content/abbonamento-annuale-0), the bus pass for students has a validity of 12 months

# Which type of ticket are yearly tickets for students and how many are there?
# Exclude the tickets that have already the field TICKET_CODE populated with 5-STUD or 6-STUD
df[(df['DESCRIZIONE_TITOLO'].str.contains('STUDENTE|STUD|STUD')) & ~ (df['TICKET_CODE'].isin(['5-STUD', '6-STUD']))]
```

```
Out[ ]: STUD. RETE INTERA FAMILIARE      1938
        S.TERRR+ACTV STUDENTE TR.6      82
        S.TERR+ACTV STUDENTE TR.2       36
        S.TERR+ACTV STUDENTE TR.7       25
        STUDENTE EXTRA FAMILIARE       25
        S.TERR+ACTV STUDENTE TR.5       13
        STUDENTE CHIOGGIA FAMILIARE     7
        S.TERR+ACTV STUDENTE TR.3        5
        MOBILITY STUDENTE RETEUNICA     5
        S.TERR+ACTV STUDENTE TR.8        2
        S.TERR+ACTV STUDENTE TR.9        2
        S.TERR+ACTV STUDENTE TR.4        2
        Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[(df['DESCRIZIONE_TITOLO'].str.contains('STUDENTE|STUD|STUD')) & ~ (df['TICKET_CODE'].isin(['5-STUD', '6-STUD
```

```
In [ ]: # TICKET_CODE = 6: Information about annual tickets
print("The number of annual tickets is: ", df[df['TICKET_CODE'] == '6'].shape[0])
print("The number of annual tickets for students is: ", df[df['TICKET_CODE'] == '6-STUD'].shape[0])
print("The number of annual tickets for workers is: ", df[df['TICKET_CODE'] == '6-WKRS'].shape[0])
print("The number of annual tickets for retired people is: ", df[df['TICKET_CODE'] == '6-RET'].shape[0])

print("The number of tickets for each type of annual ticket (including the subtypes) is: ")
df[df['TICKET_CODE'].isin(['6', '6-STUD', '6-WKRS', '6-RET'])].groupby('TICKET_CODE')['DESCRIZIONE_TITOLO'].value_c

The number of annual tickets is: 399027
The number of annual tickets for students is: 82497
The number of annual tickets for workers is: 1277
The number of annual tickets for retired people is: 30287
The number of tickets for each type of annual ticket (including the subtypes) is:
```

Out[]:	TICKET_CODE	DESCRIZIONE_TITOLO	
	6	ANNUALE ORDINARIO RETE UNICA	311494
		ANNUALE ORDINARIO ISOLE	74351
		ANNUALE ORD.RES.PELLESTRINA	5830
		SUPP.ANNUALE NAVIGAZIONE	3583
		ANNUALE ORDINARIO EXTRA	2900
		ABB ANNUALE PEOPLEMOVER	287
		SUPP ANNUALE PEOPLEMOVER	132
		ANNUALE CAT. D 17(UN SEMESTRE)	122
		SUPP. ANNUALE AUTOMOB.	91
		ABB.CHIOGGIA ANNUALE	85
		ANNUALE CAT. D LINEA 11	67
		S.TERR+ACTV ANN ORD TR.9	40
		ANNUALE ORDINARIO BUS LIDO	13
		S.TERR+ACTV ANN ORD TR.2	11
		S.TERR+ACTV ANN ORD TR.6	8
		S.TERR+ACTV ANN ORD TR.8	7
		S.TERR+ACTV ANN ORD TR.7	4
		ANNUALE PARK+RETE INTERA	1
		S.TERR+ACTV ANN ORD TR.3	1
	6-RET	ABB. OVER75 GRATUITO	15349
		ABBONAMENTO PENSIONATI ACTV	8187
		ABB. OVER75 RETE UNICA 50%	5294
		ABB. OVER 75 A20	1455
		ABB. OVER 75 A5	2
	6-STUD	ABB STUD. RETEUNICA 12 MESI	65034
		ANNUALE STUDENTE ISOLE	9965
		ANNUALE STUDENTE EXTRA	2056
		STUD. RETE INTERA FAMILIARE	1938
		ABB.STUD.ANN.PELLESTRINA	1678
		ATVO+ACTV ANN.STUD.F1	817
		ATVO+ACTV ANN.STUD.F2	327
		SUPP. 12 MESI STUDENTE LAGUNA	225
		ABB STUD. 12 MESI CHIOGGIA	98
		S.TERR+ACTV ANN STUD TR.6	88
		S.TERRR+ACTV STUDENTE TR.6	82
		S.TERR+ACTV STUDENTE TR.2	36
		S.TERR+ACTV ANN STUD TR.7	32
		S.TERR+ACTV STUDENTE TR.7	25
		STUDENTE EXTRA FAMILIARE	25

	S.TERR+ACTV STUDENTE TR.5	13
	S.TERR+ACTV ANN STUD TR.5	12
	S.TERR+ACTV ANN STUD TR.8	9
	STUDENTE CHIOGGIA FAMILIARE	7
	ABB STUDENTE BUS LIDO 12 MESI	5
	MOBILITY STUDENTE RETEUNICA	5
	S.TERR+ACTV STUDENTE TR.3	5
	SUPP. 12 MESI STUDENTE AUTOMOB	5
	S.TERR+ACTV ANN STUD TR.2	3
	S.TERR+ACTV STUDENTE TR.4	2
	S.TERR+ACTV STUDENTE TR.8	2
	S.TERR+ACTV STUDENTE TR.9	2
	S.TERR+ACTV ANN STUD TR.3	1
6-WKRS	ATVO+ACTV ANN.LAV.F1	923
	ATVO+ACTV ANN.LAV.F2	354

Name: DESCRIZIONE_TITOLO, dtype: int64

```
In [ ]: print("Information about the tickets with code 6 (including the subtypes) related to the serial number: ")
df[df['TICKET_CODE'].isin(['6', '6-STUD', '6-WKRS', '6-RET'])].groupby(['TICKET_CODE', 'DESCRIZIONE_TITOLO'])['SERI
```

Information about the tickets with code 6 (including the subtypes) related to the serial number:

Out[]:

		count	mean	std	min	25%	50%	75%	max
TICKET_CODE	DESCRIZIONE_TITOLO								
6	ABB ANNUALE PEOPLEMOVER	6.0	47.833333	36.956280	18.0	21.00	32.5	65.75	110.0
	ABB.CHIOGGIA ANNUALE	7.0	12.142857	15.410263	1.0	2.00	4.0	20.50	35.0
	ANNUALE CAT. D 17(UN SEMESTRE)	1.0	122.000000	NaN	122.0	122.00	122.0	122.00	122.0
	ANNUALE CAT. D LINEA 11	1.0	67.000000	NaN	67.0	67.00	67.0	67.00	67.0
	ANNUALE ORD.RES.PELLESTRINA	429.0	13.589744	16.760252	1.0	3.00	6.0	18.00	98.0
	ANNUALE ORDINARIO BUS LIDO	12.0	1.083333	0.288675	1.0	1.00	1.0	1.00	2.0
	ANNUALE ORDINARIO EXTRA	230.0	12.608696	17.655458	1.0	1.00	4.0	17.75	75.0
	ANNUALE ORDINARIO ISOLE	3109.0	23.914764	24.020827	1.0	6.00	16.0	35.00	290.0
	ANNUALE ORDINARIO RETE UNICA	11762.0	26.483081	26.526202	1.0	6.00	18.0	39.00	223.0
	ANNUALE PARK+RETE INTERA	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	S.TERR+ACTV ANN ORD TR.2	1.0	11.000000	NaN	11.0	11.00	11.0	11.00	11.0
	S.TERR+ACTV ANN ORD TR.3	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	S.TERR+ACTV ANN ORD TR.6	4.0	2.000000	0.816497	1.0	1.75	2.0	2.25	3.0
	S.TERR+ACTV ANN ORD TR.7	1.0	4.000000	NaN	4.0	4.00	4.0	4.00	4.0
	S.TERR+ACTV ANN ORD TR.8	2.0	3.500000	2.121320	2.0	2.75	3.5	4.25	5.0
	S.TERR+ACTV ANN ORD TR.9	2.0	20.000000	5.656854	16.0	18.00	20.0	22.00	24.0
	SUPP ANNUALE PEOPLEMOVER	7.0	18.857143	32.017852	1.0	2.50	4.0	17.00	88.0
	SUPP. ANNUALE AUTOMOB.	5.0	18.200000	16.724234	1.0	3.00	18.0	29.00	40.0
	SUPP.ANNUALE NAVIGAZIONE	117.0	30.623932	37.780383	1.0	9.00	24.0	42.00	356.0
6-RET	ABB. OVER 75 A20	169.0	8.609467	15.270619	1.0	2.00	4.0	10.00	159.0
	ABB. OVER 75 A5	1.0	2.000000	NaN	2.0	2.00	2.0	2.00	2.0
	ABB. OVER75 GRATUITO	1564.0	9.813939	16.048568	1.0	2.00	4.0	10.00	200.0
	ABB. OVER75 RETE UNICA 50%	439.0	12.059226	15.086846	1.0	3.00	6.0	16.00	116.0

TICKET_CODE	DESCRIZIONE_TITOLO	count	mean	std	min	25%	50%	75%	max
6-STUD	ABBONAMENTO PENSIONATI ACTV	834.0	9.816547	15.638508	1.0	2.00	4.0	11.00	238.0
	ABB STUD. 12 MESI CHIOGGIA	19.0	5.157895	5.785645	1.0	2.00	3.0	6.50	23.0
	ABB STUD. RETEUNICA 12 MESI	5025.0	12.942090	17.474675	1.0	2.00	5.0	16.00	134.0
	ABB STUDENTE BUS LIDO 12 MESI	5.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
	ABB.STUD.ANN.PELLESTRINA	174.0	9.643678	12.063873	1.0	2.00	5.0	12.75	70.0
	ANNUALE STUDENTE EXTRA	643.0	3.197512	5.519424	1.0	1.00	1.0	3.00	67.0
	ANNUALE STUDENTE ISOLE	676.0	14.741124	17.600156	1.0	3.00	7.0	20.00	103.0
	ATVO+ACTV ANN.STUD.F1	25.0	32.680000	32.158099	1.0	10.00	18.0	37.00	103.0
	ATVO+ACTV ANN.STUD.F2	11.0	29.727273	27.295021	1.0	12.50	23.0	39.50	88.0
	MOBILITY STUDENTE RETEUNICA	2.0	2.500000	2.121320	1.0	1.75	2.5	3.25	4.0
	S.TERR+ACTV ANN STUD TR.2	2.0	1.500000	0.707107	1.0	1.25	1.5	1.75	2.0
	S.TERR+ACTV ANN STUD TR.3	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	S.TERR+ACTV ANN STUD TR.5	10.0	1.200000	0.421637	1.0	1.00	1.0	1.00	2.0
	S.TERR+ACTV ANN STUD TR.6	12.0	7.333333	20.375267	1.0	1.00	1.0	2.00	72.0
	S.TERR+ACTV ANN STUD TR.7	11.0	2.909091	2.427120	1.0	1.50	2.0	3.00	8.0
	S.TERR+ACTV ANN STUD TR.8	1.0	9.000000	NaN	9.0	9.00	9.0	9.00	9.0
	S.TERR+ACTV STUDENTE TR.2	22.0	1.636364	2.555022	1.0	1.00	1.0	1.00	13.0
	S.TERR+ACTV STUDENTE TR.3	4.0	1.250000	0.500000	1.0	1.00	1.0	1.25	2.0
	S.TERR+ACTV STUDENTE TR.4	2.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
	S.TERR+ACTV STUDENTE TR.5	8.0	1.625000	1.060660	1.0	1.00	1.0	2.00	4.0
	S.TERR+ACTV STUDENTE TR.7	9.0	2.777778	2.386304	1.0	1.00	1.0	4.00	7.0
	S.TERR+ACTV STUDENTE TR.8	1.0	2.000000	NaN	2.0	2.00	2.0	2.00	2.0
	S.TERR+ACTV STUDENTE TR.9	2.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0

TICKET_CODE	DESCRIZIONE_TITOLO	count	mean	std	min	25%	50%	75%	max
6-WKRS	S.TERRR+ACTV STUDENTE TR.6	18.0	4.555556	6.279961	1.0	1.00	2.0	4.00	25.0
	STUD. RETE INTERA FAMILIARE	118.0	16.423729	26.215545	1.0	2.00	4.0	20.00	150.0
	STUDENTE CHIOGGIA FAMILIARE	2.0	3.500000	3.535534	1.0	2.25	3.5	4.75	6.0
	STUDENTE EXTRA FAMILIARE	11.0	2.272727	1.848833	1.0	1.00	2.0	2.50	7.0
	SUPP. 12 MESI STUDENTE AUTOMOB	3.0	1.666667	0.577350	1.0	1.50	2.0	2.00	2.0
	SUPP. 12 MESI STUDENTE LAGUNA	22.0	10.227273	14.767828	1.0	1.00	3.0	12.50	56.0
	ATVO+ACTV ANN.LAV.F1	22.0	41.954545	32.322007	1.0	11.75	39.0	73.00	98.0
	ATVO+ACTV ANN.LAV.F2	5.0	70.800000	59.440727	4.0	38.00	64.0	86.00	162.0

75 minutes tickets

```
In [ ]: # Which type of ticket are 75' (75 minutes) tickets and how many are there?
df[df['DESCRIZIONE_TITOLO'].str.contains('75\'|75MIN|75 MIN')]['DESCRIZIONE_TITOLO'].value_counts()
```

```
Out[ ]: BIGL RETE UNICA 75' 1167916
75'-TPL 8,64-COMVE0,86 449274
BIGL.AUT.75'MESTRE/LIDO-TSC 327816
PEOPLEMOVER+BUS+TRAM 75' 45649
75'-TPL 6,64-COMVE0,86 11049
BIGL.MESTRE/LIDO 75' A BORDO 8607
ORD. NAVIGAZIONE 75' ONLINE 3509
BIGLIETTO DI BORDO CV 75' 2936
BORDO 75MIN CARTAVENEZIA 1466
NA-BIG.AUT.75' MESTRE/LIDO-CSC 528
PEOPLEMOVER+BUS+TRAM 75'CARNET 460
NA-CARNET NAV. 10 CORSE DA 75' 301
ORD. NAVIG. 75' ONLINE 1 MESE 195
NA75'-TPL 13,28-COMVE1,72 113
NA-C AUT. 10 CORSE 75' CARD 100
VENDITA A BORDO 75' CV 9
75'-TPL 12,60-CVE2,40 ONLINE 3
Name: DESCRIZIONE_TITOLO, dtype: int64
```

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[df['DESCRIZIONE_TITOLO'].str.contains('75\'|75MIN|75 MIN'), 'TICKET_CODE'] = '7'
```

```
In [ ]: # TICKET_CODE = 7: Information about 75' (75 minutes) tickets
print("The number of 75' (75 minutes) tickets is: ", df[df['TICKET_CODE'] == '7'].shape[0])
print("The number of tickets for each type of 75' (75 minutes) ticket is: ")
df[df['TICKET_CODE'] == '7'].groupby('TICKET_CODE')['DESCRIZIONE_TITOLO'].value_counts()
```

The number of 75' (75 minutes) tickets is: 2019931

The number of tickets for each type of 75' (75 minutes) ticket is:

```
Out[ ]: TICKET_CODE DESCRIZIONE_TITOLO
7      BIGL RETE UNICA 75' 1167916
      75'-TPL 8,64-COMVE0,86 449274
      BIGL.AUT.75'MESTRE/LIDO-TSC 327816
      PEOPLEMOVER+BUS+TRAM 75' 45649
      75'-TPL 6,64-COMVE0,86 11049
      BIGL.MESTRE/LIDO 75' A BORDO 8607
      ORD. NAVIGAZIONE 75' ONLINE 3509
      BIGLIETTO DI BORDO CV 75' 2936
      BORDO 75MIN CARTAVENEZIA 1466
      NA-BIG.AUT.75' MESTRE/LIDO-CSC 528
      PEOPLEMOVER+BUS+TRAM 75'CARNET 460
      NA-CARNET NAV. 10 CORSE DA 75' 301
      ORD. NAVIG. 75' ONLINE 1 MESE 195
      NA75'-TPL 13,28-COMVE1,72 113
      NA-C AUT. 10 CORSE 75' CARD 100
      VENDITA A BORDO 75' CV 9
      75'-TPL 12,60-CVE2,40 ONLINE 3
```

Name: DESCRIZIONE_TITOLO, dtype: int64

```
In [ ]: print("Information about the tickets with code 7 related to the serial number: ")
df[df['TICKET_CODE'] == '7'].groupby('DESCRIZIONE_TITOLO')['SERIALE'].value_counts().groupby('DESCRIZIONE_TITOLO').
```

Information about the tickets with code 7 related to the serial number:

Out []:

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
75'-TPL 12,60-CVE2,40 ONLINE	3.0	1.000000	0.000000	1.0	1.0	1.0	1.0	1.0
75'-TPL 6,64-COMVE0,86	10354.0	1.067124	0.348078	1.0	1.0	1.0	1.0	8.0
75'-TPL 8,64-COMVE0,86	414505.0	1.083881	0.371727	1.0	1.0	1.0	1.0	126.0
BIGL RETE UNICA 75'	167114.0	6.988738	9.324228	1.0	2.0	3.0	8.0	179.0
BIGL.AUT.75'MESTRE/LIDO-TSC	219856.0	1.491049	1.097226	1.0	1.0	1.0	2.0	25.0
BIGL.MESTRE/LIDO 75' A BORDO	8438.0	1.020028	0.143450	1.0	1.0	1.0	1.0	3.0
BIGLIETTO DI BORDO CV 75'	2114.0	1.388836	1.016750	1.0	1.0	1.0	1.0	13.0
BORDO 75MIN CARTAVENEZIA	1411.0	1.038979	0.214470	1.0	1.0	1.0	1.0	3.0
NA-BIG.AUT.75' MESTRE/LIDO-CSC	365.0	1.446575	0.943759	1.0	1.0	1.0	2.0	10.0
NA-C AUT. 10 CORSE 75' CARD	71.0	1.408451	0.854859	1.0	1.0	1.0	2.0	7.0
NA-CARNET NAV. 10 CORSE DA 75'	166.0	1.813253	1.157874	1.0	1.0	2.0	2.0	8.0
NA75'-TPL 13,28-COMVE1,72	91.0	1.241758	0.564519	1.0	1.0	1.0	1.0	4.0
ORD. NAVIG. 75' ONLINE 1 MESE	169.0	1.153846	0.408248	1.0	1.0	1.0	1.0	3.0
ORD. NAVIGAZIONE 75' ONLINE	3173.0	1.105893	0.377663	1.0	1.0	1.0	1.0	8.0
PEOPLEMOVER+BUS+TRAM 75'	44983.0	1.014806	0.128965	1.0	1.0	1.0	1.0	4.0
PEOPLEMOVER+BUS+TRAM 75'CARNET	170.0	2.705882	2.161106	1.0	1.0	2.0	3.0	10.0
VENDITA A BORDO 75' CV	7.0	1.285714	0.487950	1.0	1.0	1.0	1.5	2.0

Other types of tickets

```
In [ ]: # Which type of ticket are other tickets and how many are there?
# The other tickets are the tickets that are not already classified in the previous categories
df[~df['TICKET_CODE'].isin(['1', '2', '3', '4', '5', '5-STUD', '5-WKRS', '5-RET', '6', '6-STUD', '6-WKRS', '6-RET',
```

Out[]: ABB. RETE INTERA A20	50474
BIGL AER-VENEZIA TSC	50419
MOBILITY ORDINARIO RETE UNICA	45048
EXTRA TRATTA 2	38147
TESSERA DI SERVIZIO ACTV	37209
LIBERA CIRC. RETE INTERA	26236
EXTRA TRATTA 3	22071
LINEA 17-CATEGORIA C	14772
TRAGHETTO CARTA VENEZIA	13195
TITOLO CMVENEZIA	13116
EXTRA TRATTA 4	11671
LINEA 17-CATEGORIA B	9969
TRAGHETTO RESIDENTE BURANO	9281
EXTRA TRATTA 1	8733
L.17-AUTO "D" OLTRE METRI 4,50	8438
LINEA 17-CATEGORIA D	8152
LIBERA CIRC. DUE RETI	7645
ARRIVA VENETO TRATTA 8-9-10	7499
BIGL RETE UNICA 100'	6367
AEROPORTO-VENEZIA AR	6051
LINEA 11-CATEGORIA C	5904
LINEA 11-CATEGORIA B	5044
EXTRA TRATTA 5	4502
TRAGHETTO RESID. PELLESTRINA	4394
AER+BOAT-TPL16,50-C.VE1,50	4059
PRENOTAZ OCCASIONALE SI BARRA	3458
ORD. RETE INTERA FAMILIARE	3412
ABB. EXTRA A20	3304
LINEA 11-CATEGORIA D	3198
TRAGHETTO RESIDENTE LIDO	3123
ABB. RETE INTERA A5	3023
MOBILITY SUPP.NAVIGAZIONE	2917
L.17-AUTO "C"DA 4,01 A 4,50 MT	2749
L.17-AUTO "AEB" FINO A 4 METRI	2733
SEMESTRALE CEOD A20+ACC.	2686
CARNET CHIOGGIA 10C. TICKET	2263
NAVETTA ARSENALE CA	2251
FERRY17-AUTOCARRI+35Q.	2245
LINEA 17-CATEGORIA A	2212
TRAGHETTO RESIDENTE MURANO	1720

TRAGHETTO RESIDENTE S.ERASMO	1714
L.11-AUTO "D" OLTRE METRI 4,50	1576
BICICLETTA "BIGLIETTERIA"	1565
LINEA 11+17 CATEGORIA C	1522
AEROPORTO-VENEZIA AR ONLINE	1310
ABB. CHIOGGIA A20	1300
MOBILITY ORDINARIO EXTRA	1294
BIGL.URBANO CHIOGGIA	1287
EXTRA TRATTA 6	1250
SEMESTRALE CEOD A20	1226
EXTRA TRATTA 2 TVM	1118
TARIFFA CARROZZINA	1069
LINEA 11-CATEGORIA A	956
TESSERA DI SERVIZIO ARRIVA	945
BIGLIETTO SOC. SPORTIVE	918
ARRIVA VENETO TRATTA 1	888
EXTRA TRATTE 2-3-4 BORDO	796
EXTRA TRATTA 1 TVM	778
FERRY11-AUTOCARRI+35Q.	733
EXTRA TRATTA 8-9-10	640
L.11-AUTO "AEB" FINO A 4 METRI	630
FERRY17-CARRI+35Q.RIM.	609
BICICLETTA "PALMARE"	585
EV8-TPL 52,00-C.VE3,00	573
LINEA 11+17 CATEGORIA B	509
ABB. ORDINARIO BUS. LIDO	508
EXTRA TRATTA 3 TVM	500
ARRIVA VENETO TRATTA 7	497
EXTRA TRATTA 4 TVM	488
LINEA 11+17 CATEGORIA D	474
ARRIVA VENETO TRATTA 2	440
APERTURA TORNELLI DUSSMAN	413
L.11-AUTO "C"DA 4,01 A 4,50 MT	408
BIGLIETTO SCUOLE	404
EXTRA TRATTA 7	383
BUS+PEOPLE MOVER ONLINE	370
ARRIVA VENETO TRATTA 4	362
JESOLO - S.MARCO AR	313
ARRIVA VENETO TRATTA 3	311
AEROPORTO-VENEZIA CS ONLINE	284

ARRIVA VENETO TRATTA 6	283
S.TERR+ACTV ORD. TR.6	267
ABB. IMPERSONALE RETE INTERA	249
ARRIVA VENETO AEROPORTO	243
BICICLETTA "CONCESSIONARI"	229
BAGAGLIO CARTAVENEZIA	220
CICLOMOTORE FINO 50CC	194
EXTRA TRATTA 1 BORDO	193
ATVO CANOVA+NAVIG AR	180
ATVO CANOVA+NAVIG 1 CORSA	177
S.TERR+ACTV ORD. TR.7	146
BIGLIETTO MOTO FINO 50 CC	133
AEROBUS+BOAT ONLINE	121
BIGLIETTO MERCI C.SEMPLICE	120
SUPP NAVIGAZIONE FAMILIARE	116
ARRIVA VENETO TRATTA 5	110
EXTRA TRATTA 5 TVM	109
ARRIVA EXTRA TR.8-9-10 BORDO	93
ARRIVA EXTRA TR.1 BORDO	85
ARRIVA EXTRA TR. 5-6-7 BORDO	80
ABB. EXTRA A5	80
ATVO CANOVA+NAVIG AR ONLINE	79
ABB. RETE INTERA A20 +ACC.	72
ARRIVAEXTRA TR.2-3-4 BORDO	72
AEROBUS+BOAT ONLINE 1MESE	62
EXTRA TRATTE 5-6-7 BORDO	55
ATVOCANOVA+NAVIG 1CORSIA ONLINE	49
LINEA 11+17 CATEGORIA A	45
S.TERR+ACTV ORD. TR.5	44
TRAGHETTO RESIDENTE GIUDECCA	41
NA-TRAGHETTO ORDINARIO	35
FERRY17-TRASPORTI PERICOLOSI	33
S.TERR+ACTV ORD. TR.9	29
INTEGRAZIONE RETE MESTRE	26
EXTRA TRATTE 8-9-10 BORDO	25
EV5-TPL 38,50-C.VE1,50	23
CAV-TREP - S.MARCO AR	22
FERRY11-CARRI+35Q.RIM.	22
EXTRA TRATTA 7 TVM	22
ORDINARIO EXTRA FAMILIARE	18

ABB. CHIOGGIA A5	18
NATRAGH-TPL 8,82-C.VE1,18	17
ORDINARIO CHIOGGIA FAMILIARE	16
CICLOMOTORE OLTRE 50CC	15
FERRY17-AUTOBUS	14
EXTRA TRATTA 6 TVM	14
CARNET CHIOGGIA 10 C. CARD	14
LIBERA CIRC. FERRY LINEA 17	11
LIBERA CIRC. FERRY LINEA 17+11	10
NA-BICICLETTA E CONDUCENTE CV	10
EXTRA TRATTA 8-9-10 TVM	9
ARRIVA AEROPORTO BORDO	9
PRENOTAZIONE VEICOLO ABBONATO	9
LIBERA CIRC. FERRY LINEA 11	8
NATRAGH-TPL 4,41-C.VE0,59	8
NA-GRUPPI E SCUOLE	7
FERRY11-TRASPORTI PERICOLOSI	7
S.TERR+ACTV ORD. TR.8	7
BIGLIETTO SCUOLE ONLINE	6
ACC.L.R.A20 RETE INTERA	6
TRAGHETTO GRATUITO	4
GRUPPI E SCUOLE ONLINE 2VIAGGI	4
ARRIVA INTEG.AEROP. BORDO	4
ABB. CHIOGGIA A20 + ACC.	4
S.TERR+ACTV ORD. TR.2	4
S.TERR+ACTV ORD. TR.3	4
S.TERR+ACTV ORD. TR.4	3
SPIAGGEAR-TPL 14,75-COMVE1,25	3
EV3-TPL 30,50-C.VE1,50	2
SUPPL. RETE MESTRE ATVO	1
NA-BIGL. CHIOGGIA CARD	1
APERTURA TORNELLI P.SABBIONI	1
NA-C CHIOGGIA 10 C. CARD FS	1
SUPP MESTRE FAMILIARE	1

Name: DESCRIZIONE_TITOLO, dtype: int64

```
In [ ]: # Populate the column 'TICKET_CODE' with the code of the ticket profile according to the ticket type and the ticket
df.loc[~df['TICKET_CODE'].isin(['1','2','3','4','5','5-STUD','5-WKRS','5-RET','6','6-STUD','6-WKRS','6-RET','7']),
```

```
In [ ]: # TICKET_CODE = 7b: Information about other tickets
print("The number of other tickets is: ", df[df['TICKET_CODE'] == '8'].shape[0])
```

The number of other tickets is: 496090

```
In [ ]: print("Information about the tickets with code 8 related to the serial number: ")
df[df['TICKET_CODE'] == '8'].groupby('DESCRIZIONE_TITOLO')['SERIALE'].value_counts().groupby('DESCRIZIONE_TITOLO').
```

Information about the tickets with code 8 related to the serial number:

Out[]:

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
ABB. CHIOGGIA A20	197.0	6.598985	7.931891	1.0	1.00	3.0	8.00	49.0
ABB. CHIOGGIA A20 + ACC.	1.0	4.000000	NaN	4.0	4.00	4.0	4.00	4.0
ABB. CHIOGGIA A5	3.0	6.000000	3.605551	3.0	4.00	5.0	7.50	10.0
ABB. EXTRA A20	223.0	14.816143	21.417920	1.0	2.00	5.0	18.00	120.0
ABB. EXTRA A5	13.0	6.153846	4.862204	1.0	1.00	5.0	10.00	15.0
ABB. IMPERSONALE RETE INTERA	10.0	24.900000	15.652121	9.0	13.25	22.5	32.75	60.0
ABB. ORDINARIO BUS. LIDO	117.0	4.341880	10.550504	1.0	1.00	2.0	3.00	89.0
ABB. RETE INTERA A20 +ACC.	7.0	10.285714	14.636875	1.0	4.00	5.0	7.50	43.0
ABB. RETE INTERA A20	3045.0	16.576026	25.076975	1.0	3.00	7.0	20.00	365.0
ABB. RETE INTERA A5	267.0	11.322097	15.453824	1.0	2.00	5.0	15.00	93.0
ACC.L.R.A20 RETE INTERA	1.0	6.000000	NaN	6.0	6.00	6.0	6.00	6.0
AER+BOAT-TPL16,50-C.VE1,50	2296.0	1.767857	0.554352	1.0	1.00	2.0	2.00	5.0
AEROBUS+BOAT ONLINE	67.0	1.805970	0.679545	1.0	1.00	2.0	2.00	4.0
AEROBUS+BOAT ONLINE 1MESE	38.0	1.631579	0.488852	1.0	1.00	2.0	2.00	2.0
AEROPORTO-VENEZIA AR	3479.0	1.739293	0.463919	1.0	1.00	2.0	2.00	4.0
AEROPORTO-VENEZIA AR ONLINE	876.0	1.495434	0.516009	1.0	1.00	1.0	2.00	3.0
AEROPORTO-VENEZIA CS ONLINE	282.0	1.007092	0.084065	1.0	1.00	1.0	1.00	2.0
APERTURA TORNELLI DUSSMAN	19.0	21.736842	39.306265	1.0	3.00	7.0	17.00	148.0
APERTURA TORNELLI P.SABBIONI	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
ARRIVA AEROPORTO BORDO	9.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
ARRIVA EXTRA TR. 5-6-7 BORDO	76.0	1.052632	0.277836	1.0	1.00	1.0	1.00	3.0
ARRIVA EXTRA TR.1 BORDO	74.0	1.148649	0.838826	1.0	1.00	1.0	1.00	8.0
ARRIVA EXTRA TR.8-9-10 BORDO	89.0	1.044944	0.208355	1.0	1.00	1.0	1.00	2.0

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
ARRIVA INTEG.AEROP. BORDO	4.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
ARRIVA VENETO AEROPORTO	233.0	1.042918	0.203110	1.0	1.00	1.0	1.00	2.0
ARRIVA VENETO TRATTA 1	584.0	1.520548	1.312137	1.0	1.00	1.0	2.00	16.0
ARRIVA VENETO TRATTA 2	294.0	1.496599	1.196549	1.0	1.00	1.0	2.00	9.0
ARRIVA VENETO TRATTA 3	229.0	1.358079	1.073275	1.0	1.00	1.0	1.00	9.0
ARRIVA VENETO TRATTA 4	280.0	1.292857	1.015976	1.0	1.00	1.0	1.00	8.0
ARRIVA VENETO TRATTA 5	73.0	1.506849	1.119564	1.0	1.00	1.0	2.00	7.0
ARRIVA VENETO TRATTA 6	208.0	1.360577	0.921972	1.0	1.00	1.0	1.00	8.0
ARRIVA VENETO TRATTA 7	349.0	1.424069	0.975619	1.0	1.00	1.0	2.00	10.0
ARRIVA VENETO TRATTA 8-9-10	5124.0	1.463505	1.126593	1.0	1.00	1.0	2.00	38.0
ARRIVAEXTRA TR.2-3-4 BORDO	70.0	1.028571	0.167802	1.0	1.00	1.0	1.00	2.0
ATVO CANOVA+NAVIG 1 CORSA	152.0	1.164474	0.520368	1.0	1.00	1.0	1.00	5.0
ATVO CANOVA+NAVIG AR	92.0	1.956522	1.036828	1.0	1.00	2.0	2.00	8.0
ATVO CANOVA+NAVIG AR ONLINE	44.0	1.795455	0.667503	1.0	1.00	2.0	2.00	4.0
ATVOCANOVA+NAVIG 1CORSIA ONLINE	43.0	1.139535	0.350605	1.0	1.00	1.0	1.00	2.0
BAGAGLIO CARTAVENEZIA	199.0	1.105528	0.308007	1.0	1.00	1.0	1.00	2.0
BICICLETTA "BIGLIETTERIA"	1432.0	1.092877	0.470401	1.0	1.00	1.0	1.00	8.0
BICICLETTA "CONCESSIONARI"	221.0	1.036199	0.210091	1.0	1.00	1.0	1.00	3.0
BICICLETTA "PALMARE"	564.0	1.037234	0.189503	1.0	1.00	1.0	1.00	2.0
BIGL AER-VENEZIA TSC	49832.0	1.011780	0.111011	1.0	1.00	1.0	1.00	4.0
BIGL RETE UNICA 100'	1787.0	3.562955	5.987186	1.0	1.00	2.0	3.00	98.0
BIGL.URBANO CHIOGGIA	1142.0	1.126970	0.355980	1.0	1.00	1.0	1.00	4.0
BIGLIETTO MERCI C.SEMPLICE	118.0	1.016949	0.129631	1.0	1.00	1.0	1.00	2.0

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
BIGLIETTO MOTO FINO 50 CC	130.0	1.023077	0.150729	1.0	1.00	1.0	1.00	2.0
BIGLIETTO SCUOLE	115.0	3.513043	6.649581	1.0	1.00	1.0	1.00	44.0
BIGLIETTO SCUOLE ONLINE	3.0	2.000000	1.732051	1.0	1.00	1.0	2.50	4.0
BIGLIETTO SOC. SPORTIVE	234.0	3.923077	6.101648	1.0	1.00	1.0	2.75	32.0
BUS+PEOPLE MOVER ONLINE	352.0	1.051136	0.245063	1.0	1.00	1.0	1.00	3.0
CARNET CHIOGGIA 10 C. CARD	9.0	1.555556	1.013794	1.0	1.00	1.0	2.00	4.0
CARNET CHIOGGIA 10C. TICKET	1120.0	2.020536	1.430714	1.0	1.00	2.0	2.00	10.0
CAV-TREP - S.MARCO AR	10.0	2.200000	0.632456	2.0	2.00	2.0	2.00	4.0
CICLOMOTORE FINO 50CC	194.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
CICLOMOTORE OLTRE 50CC	15.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EV3-TPL 30,50-C.VE1,50	1.0	2.000000	NaN	2.0	2.00	2.0	2.00	2.0
EV5-TPL 38,50-C.VE1,50	11.0	2.090909	1.513575	1.0	1.00	2.0	2.50	6.0
EV8-TPL 52,00-C.VE3,00	67.0	8.552239	6.108324	1.0	2.50	8.0	11.00	24.0
EXTRA TRATTA 1	5258.0	1.660898	1.507346	1.0	1.00	1.0	2.00	15.0
EXTRA TRATTA 1 BORDO	164.0	1.176829	0.654723	1.0	1.00	1.0	1.00	6.0
EXTRA TRATTA 1 TVM	710.0	1.095775	0.410523	1.0	1.00	1.0	1.00	5.0
EXTRA TRATTA 2	21988.0	1.734901	1.596514	1.0	1.00	1.0	2.00	40.0
EXTRA TRATTA 2 TVM	1026.0	1.089669	0.389810	1.0	1.00	1.0	1.00	6.0
EXTRA TRATTA 3	13811.0	1.598074	1.383515	1.0	1.00	1.0	2.00	36.0
EXTRA TRATTA 3 TVM	463.0	1.079914	0.360505	1.0	1.00	1.0	1.00	5.0
EXTRA TRATTA 4	7494.0	1.557379	1.214202	1.0	1.00	1.0	2.00	28.0
EXTRA TRATTA 4 TVM	458.0	1.065502	0.418487	1.0	1.00	1.0	1.00	6.0
EXTRA TRATTA 5	2878.0	1.564281	3.580040	1.0	1.00	1.0	2.00	184.0

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
EXTRA TRATTA 5 TVM	95.0	1.147368	0.385051	1.0	1.00	1.0	1.00	3.0
EXTRA TRATTA 6	845.0	1.479290	1.114790	1.0	1.00	1.0	2.00	10.0
EXTRA TRATTA 6 TVM	13.0	1.076923	0.277350	1.0	1.00	1.0	1.00	2.0
EXTRA TRATTA 7	280.0	1.367857	0.674977	1.0	1.00	1.0	2.00	5.0
EXTRA TRATTA 7 TVM	22.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EXTRA TRATTA 8-9-10	474.0	1.350211	0.832563	1.0	1.00	1.0	2.00	12.0
EXTRA TRATTA 8-9-10 TVM	9.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EXTRA TRATTE 2-3-4 BORDO	703.0	1.132290	0.431484	1.0	1.00	1.0	1.00	4.0
EXTRA TRATTE 5-6-7 BORDO	53.0	1.037736	0.192380	1.0	1.00	1.0	1.00	2.0
EXTRA TRATTE 8-9-10 BORDO	23.0	1.086957	0.288104	1.0	1.00	1.0	1.00	2.0
FERRY11-AUTOCARRI+35Q.	733.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
FERRY11-CARRI+35Q.RIM.	22.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
FERRY11-TRASPORTI PERICOLOSI	7.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
FERRY17-AUTOBUS	14.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
FERRY17-AUTOCARRI+35Q.	2191.0	1.024646	0.338321	1.0	1.00	1.0	1.00	13.0
FERRY17-CARRI+35Q.RIM.	588.0	1.035714	0.365109	1.0	1.00	1.0	1.00	9.0
FERRY17-TRASPORTI PERICOLOSI	33.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
GRUPPI E SCUOLE ONLINE 2VIAGGI	1.0	4.000000	NaN	4.0	4.00	4.0	4.00	4.0
INTEGRAZIONE RETE MESTRE	6.0	4.333333	8.164966	1.0	1.00	1.0	1.00	21.0
JESOLO - S.MARCO AR	151.0	2.072848	0.589908	1.0	2.00	2.0	2.00	4.0
L.11-AUTO "AEB" FINO A 4 METRI	629.0	1.001590	0.039873	1.0	1.00	1.0	1.00	2.0
L.11-AUTO "C" DA 4,01 A 4,50 MT	408.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
L.11-AUTO "D" OLTRE METRI 4,50	1569.0	1.004461	0.075630	1.0	1.00	1.0	1.00	3.0

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
L.17-AUTO "AEB" FINO A 4 METRI	2680.0	1.019776	0.264873	1.0	1.00	1.0	1.00	9.0
L.17-AUTO "C" DA 4,01 A 4,50 MT	2670.0	1.029588	0.353043	1.0	1.00	1.0	1.00	11.0
L.17-AUTO "D" OLTRE METRI 4,50	8219.0	1.026646	0.452295	1.0	1.00	1.0	1.00	25.0
LIBERA CIRC. DUE RETI	378.0	20.224868	26.440960	1.0	3.00	7.0	28.00	134.0
LIBERA CIRC. FERRY LINEA 11	1.0	8.000000	NaN	8.0	8.00	8.0	8.00	8.0
LIBERA CIRC. FERRY LINEA 17	5.0	2.200000	2.167948	1.0	1.00	1.0	2.00	6.0
LIBERA CIRC. FERRY LINEA 17+11	3.0	3.333333	3.214550	1.0	1.50	2.0	4.50	7.0
LIBERA CIRC. RETE INTERA	2403.0	10.918019	16.336272	1.0	2.00	4.0	12.00	122.0
LINEA 11+17 CATEGORIA A	4.0	11.250000	12.175796	2.0	4.25	7.0	14.00	29.0
LINEA 11+17 CATEGORIA B	72.0	7.069444	5.918047	1.0	4.00	4.0	9.00	38.0
LINEA 11+17 CATEGORIA C	179.0	8.502793	7.796120	1.0	4.00	6.0	10.00	56.0
LINEA 11+17 CATEGORIA D	50.0	9.480000	10.254397	1.0	4.00	6.0	10.75	55.0
LINEA 11-CATEGORIA A	88.0	10.863636	15.521718	1.0	2.00	4.0	12.25	78.0
LINEA 11-CATEGORIA B	439.0	11.489749	16.106361	1.0	2.00	6.0	12.00	99.0
LINEA 11-CATEGORIA C	512.0	11.531250	16.690782	1.0	2.00	6.0	12.25	110.0
LINEA 11-CATEGORIA D	215.0	14.874419	20.478346	1.0	2.00	6.0	18.00	110.0
LINEA 17-CATEGORIA A	272.0	8.132353	13.081244	1.0	2.00	3.0	8.00	84.0
LINEA 17-CATEGORIA B	1426.0	6.990884	7.976231	1.0	2.00	4.0	8.00	84.0
LINEA 17-CATEGORIA C	2033.0	7.266109	6.838010	1.0	2.00	6.0	10.00	74.0
LINEA 17-CATEGORIA D	984.0	8.284553	9.644565	1.0	2.00	6.0	10.00	104.0
MOBILITY ORDINARIO EXTRA	128.0	10.109375	14.514850	1.0	1.00	2.0	13.50	66.0
MOBILITY ORDINARIO RETE UNICA	1972.0	22.843813	24.051757	1.0	4.00	14.0	35.00	143.0
MOBILITY SUPP.NAVIGAZIONE	116.0	25.146552	16.156439	1.0	11.75	23.0	36.00	86.0

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
NA-BICICLETTA E CONDUCENTE CV	9.0	1.111111	0.333333	1.0	1.00	1.0	1.00	2.0
NA-BIGL. CHIOGGIA CARD	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
NA-C CHIOGGIA 10 C. CARD FS	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
NA-GRUPPI E SCUOLE	2.0	3.500000	3.535534	1.0	2.25	3.5	4.75	6.0
NA-TRAGHETTO ORDINARIO	2.0	17.500000	20.506097	3.0	10.25	17.5	24.75	32.0
NATRAGH-TPL 4,41-C.VE0,59	8.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
NATRAGH-TPL 8,82-C.VE1,18	13.0	1.307692	0.480384	1.0	1.00	1.0	2.00	2.0
NAVETTA ARSENALE CA	1843.0	1.221378	0.472753	1.0	1.00	1.0	1.00	6.0
ORD. RETE INTERA FAMILIARE	153.0	22.300654	26.037949	1.0	4.00	10.0	35.00	151.0
ORDINARIO CHIOGGIA FAMILIARE	1.0	16.000000	NaN	16.0	16.00	16.0	16.00	16.0
ORDINARIO EXTRA FAMILIARE	3.0	6.000000	6.244998	1.0	2.50	4.0	8.50	13.0
PRENOTAZ OCCASIONALE SI BARRA	3200.0	1.080625	0.453943	1.0	1.00	1.0	1.00	13.0
PRENOTAZIONE VEICOLO ABBONATO	9.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
S.TERR+ACTV ORD. TR.2	4.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
S.TERR+ACTV ORD. TR.3	4.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
S.TERR+ACTV ORD. TR.4	1.0	3.000000	NaN	3.0	3.00	3.0	3.00	3.0
S.TERR+ACTV ORD. TR.5	8.0	5.500000	8.400680	1.0	1.00	1.0	6.00	25.0
S.TERR+ACTV ORD. TR.6	38.0	7.026316	12.177709	1.0	1.00	2.0	7.00	55.0
S.TERR+ACTV ORD. TR.7	15.0	9.733333	9.772166	1.0	1.50	4.0	15.50	30.0
S.TERR+ACTV ORD. TR.8	3.0	2.333333	2.309401	1.0	1.00	1.0	3.00	5.0
S.TERR+ACTV ORD. TR.9	12.0	2.416667	2.466441	1.0	1.00	1.0	2.25	8.0
SEMESTRALE CEOD A20	102.0	12.019608	20.030166	1.0	1.00	4.0	12.00	104.0
SEMESTRALE CEOD A20+ACC.	186.0	14.440860	24.672542	1.0	2.00	4.0	12.75	173.0

	count	mean	std	min	25%	50%	75%	max
DESCRIZIONE_TITOLO								
SPIAGGEAR-TPL 14,75-COMVE1,25	3.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
SUPP MESTRE FAMILIARE	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
SUPP NAVIGAZIONE FAMILIARE	3.0	38.666667	26.839026	18.0	23.50	29.0	49.00	69.0
SUPPL. RETE MESTRE ATVO	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
TARIFFA CARROZZINA	995.0	1.074372	0.284574	1.0	1.00	1.0	1.00	3.0
TESSERA DI SERVIZIO ACTV	1931.0	19.269291	36.260457	1.0	2.00	7.0	20.00	475.0
TESSERA DI SERVIZIO ARRIVA	119.0	7.941176	10.850521	1.0	1.00	2.0	10.00	54.0
TITOLO CMVENEZIA	786.0	16.687023	25.532470	1.0	2.00	6.0	19.00	240.0
TRAGHETTO CARTA VENEZIA	2409.0	5.477377	6.712011	1.0	1.00	3.0	7.00	57.0
TRAGHETTO GRATUITO	4.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
TRAGHETTO RESID. PELLESTRINA	638.0	6.887147	6.789025	1.0	2.00	5.0	9.75	48.0
TRAGHETTO RESIDENTE BURANO	850.0	10.918824	9.055476	1.0	4.00	8.0	16.00	50.0
TRAGHETTO RESIDENTE GIUDECCA	7.0	5.857143	4.879500	1.0	2.00	6.0	7.50	15.0
TRAGHETTO RESIDENTE LIDO	489.0	6.386503	7.066132	1.0	2.00	4.0	8.00	51.0
TRAGHETTO RESIDENTE MURANO	251.0	6.852590	6.021809	1.0	2.00	5.0	10.00	38.0
TRAGHETTO RESIDENTE S.ERASMO	149.0	11.503356	11.129231	1.0	3.00	9.0	15.00	65.0

Summary of the ticket profiles

```
In [ ]: # Print the number of tickets for each ticket profile code ordered by the code of the ticket profile; print the nam
df['TICKET_CODE'].value_counts().sort_index().rename(dict_tickets).reindex(dict_tickets.values(), fill_value=0)
```

```
Out[ ]: One-day ticket          547218
        Two-day ticket         323021
        Three-day ticket       409242
        Seven-day ticket       171259
        Monthly ticket         909012
        Monthly ticket for students 141332
        Monthly ticket for retired      0
        Monthly ticket for workers    7273
        Annual ticket           399027
        Annual ticket for students    82497
        Annual ticket for retired    30287
        Annual ticket for workers    1277
        75 minutes ticket         2019931
        Other ticket             496090
        Name: TICKET_CODE, dtype: int64
```

```
In [ ]: print("The number of tickets for each ticket profile code ordered by the code of the ticket profile, with a reference to the name of the serial number, is:")
        df.groupby('TICKET_CODE')['SERIALE'].value_counts().groupby('TICKET_CODE').describe()
```

The number of tickets for each ticket profile code ordered by the code of the ticket profile, with a reference to the name of the serial number, is:

Out []:

	count	mean	std	min	25%	50%	75%	max
TICKET_CODE								
1	136365.0	4.012892	2.219777	1.0	2.0	4.0	5.0	74.0
2	48262.0	6.693071	3.576822	1.0	4.0	6.0	9.0	41.0
3	46891.0	8.727517	4.701423	1.0	5.0	8.0	11.0	76.0
4	12294.0	13.930291	8.364992	1.0	8.0	13.0	18.0	146.0
5	40038.0	22.703731	26.971151	1.0	3.0	12.0	33.0	542.0
5-STUD	10967.0	12.887025	18.626881	1.0	2.0	5.0	15.0	164.0
5-WKRS	183.0	39.743169	33.530610	1.0	9.0	32.0	65.0	156.0
6	15644.0	25.506712	26.009371	1.0	6.0	17.0	38.0	356.0
6-RET	2970.0	10.197643	16.036240	1.0	2.0	5.0	11.0	238.0
6-STUD	6827.0	12.083931	17.111862	1.0	2.0	4.0	15.0	150.0
6-WKRS	27.0	47.296296	38.958375	1.0	12.5	40.0	75.0	162.0
7	870919.0	2.319310	4.730368	1.0	1.0	1.0	2.0	179.0
8	177925.0	2.788197	8.279868	1.0	1.0	1.0	2.0	475.0

```
In [ ]: # Countplot of the column 'TICKET_CODE'
fig, ax = plt.subplots(figsize=(15,8))
# Countplot of the column 'TICKET_CODE'
sns.countplot(x='TICKET_CODE', data=df, order=df['TICKET_CODE'].value_counts().sort_index().index)
plt.title('Countplot of the column TICKET_CODE', fontsize=20)
plt.xlabel('Ticket code', fontsize=15)
plt.ylabel('Count (in millions)', fontsize=15)

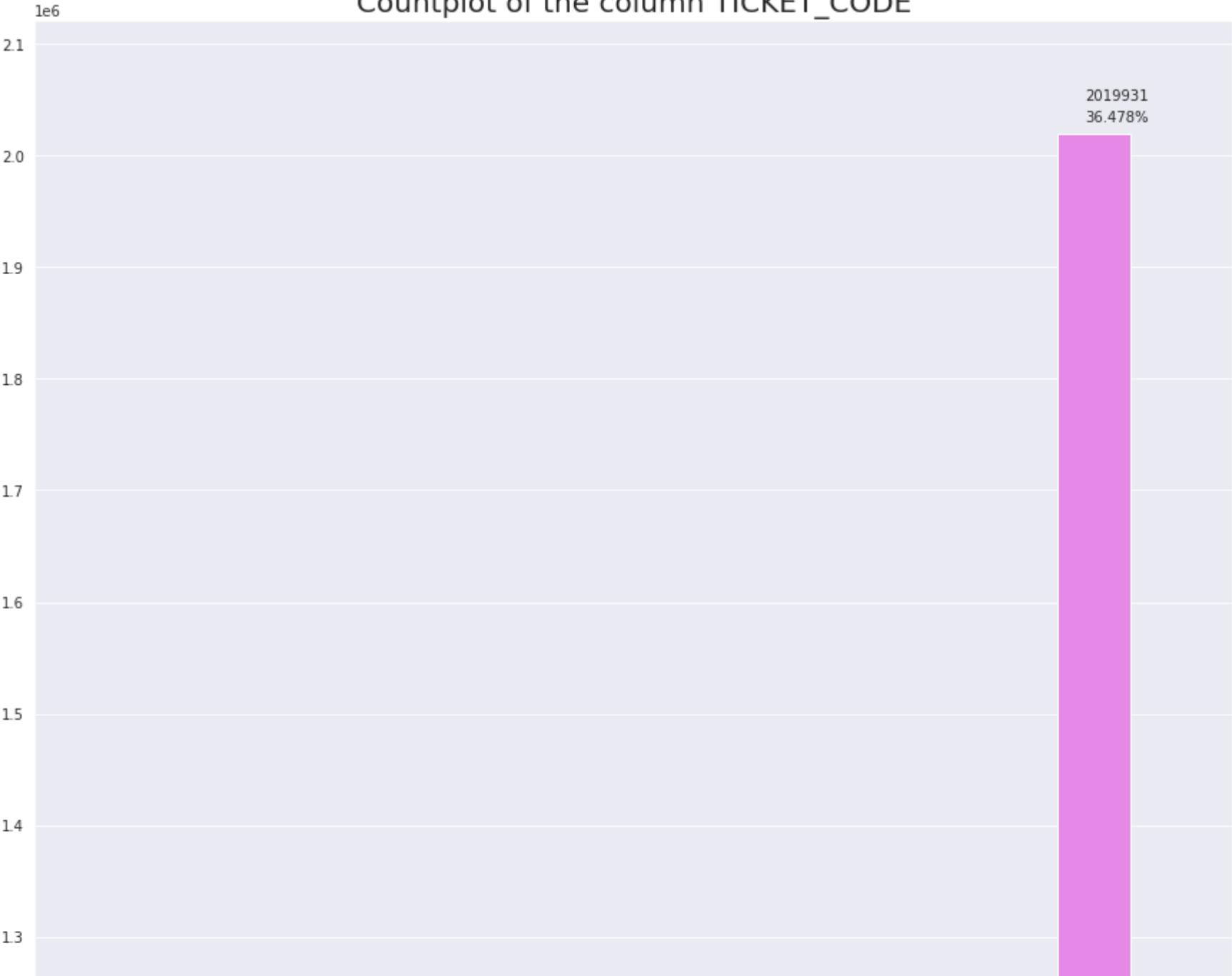
# Change yticks to have a better visualization
scale = np.arange(0, max(df['TICKET_CODE'].value_counts())+100000, 100000)
plt.yticks(scale)

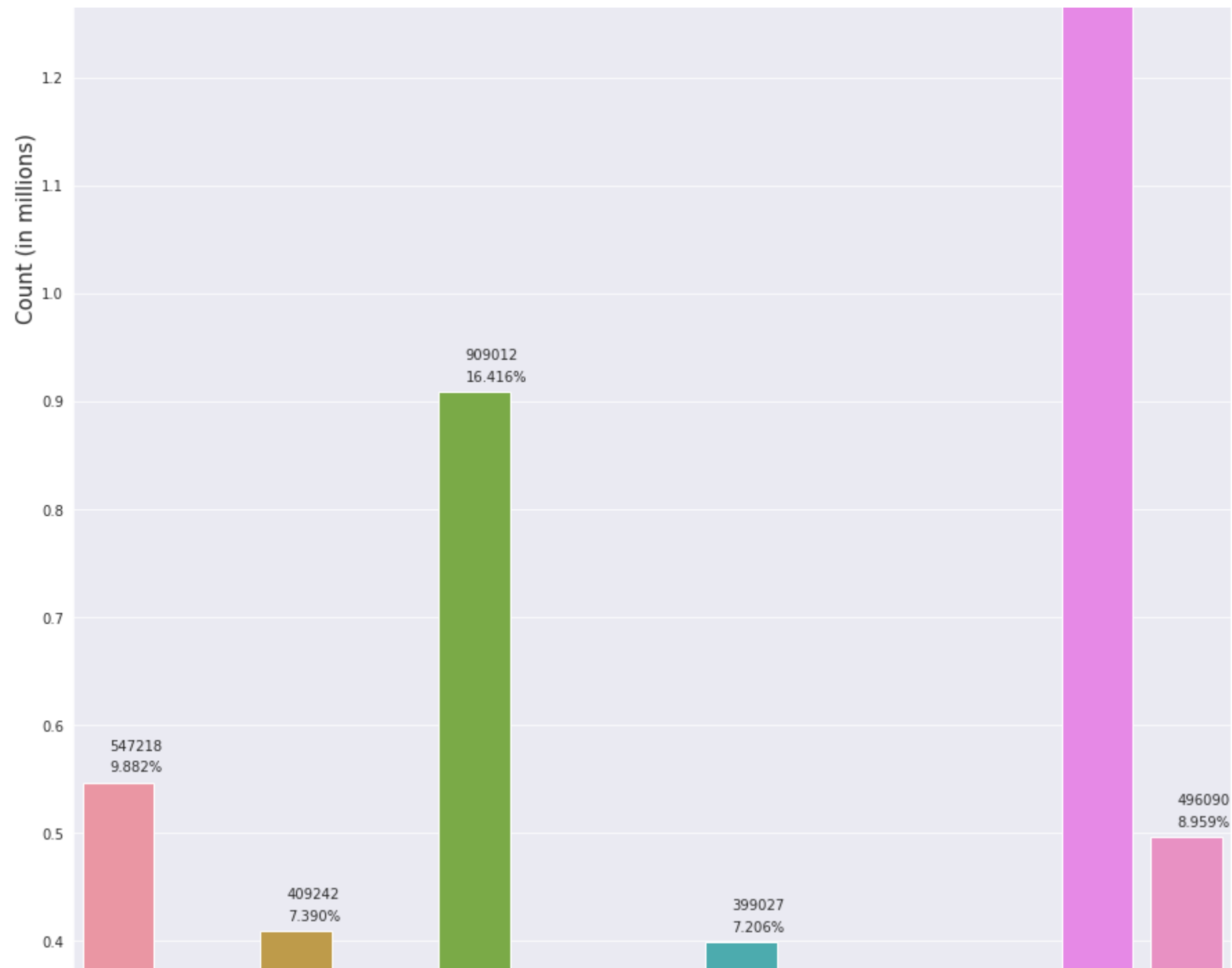
# Add the percentage of each category on top of the bars
for p in ax.patches:
    ax.annotate('{:.3f}%'.format(100*p.get_height()/len(df)), (p.get_x()+0.3, p.get_height()+10000))
```

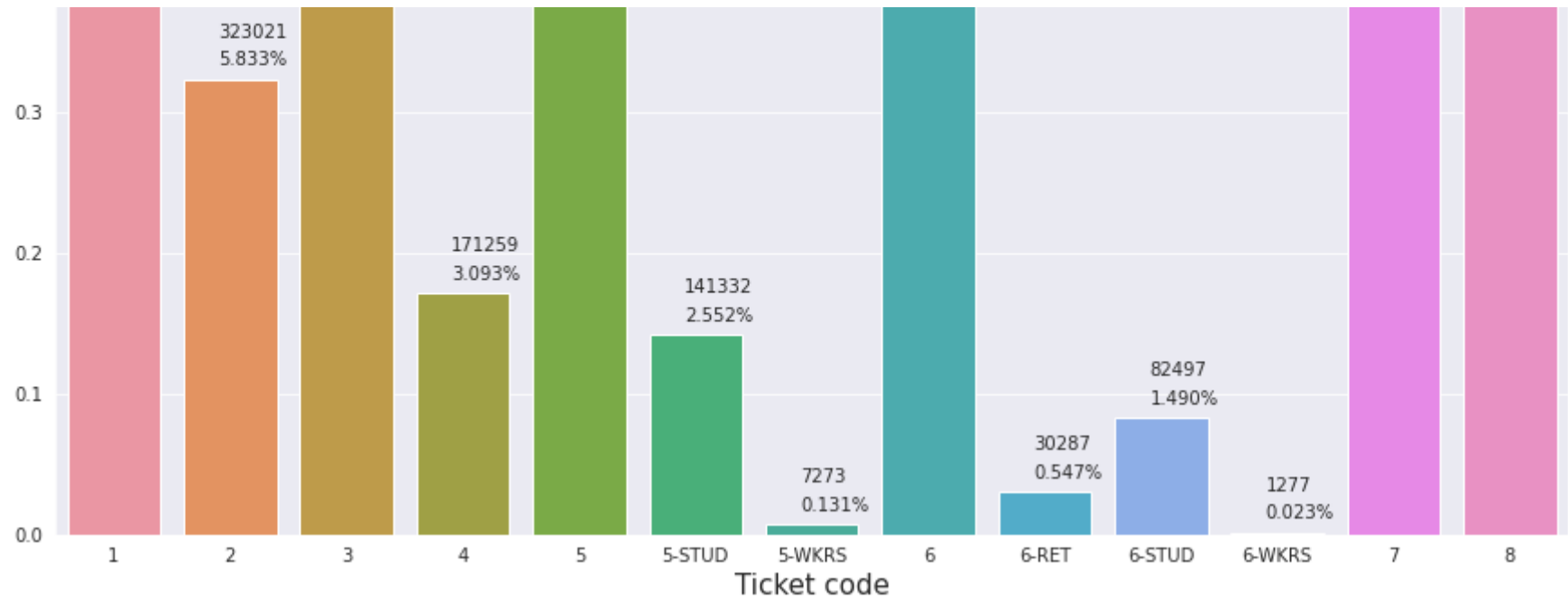
```
# Add the count of each category on top of the bars
for p in ax.patches:
    ax.annotate('{:.0f}'.format(p.get_height()), (p.get_x()+0.3, p.get_height()+30000))

# Add a padding on the top of the plot
plt.subplots_adjust(top=3)
```

Countplot of the column TICKET_CODE





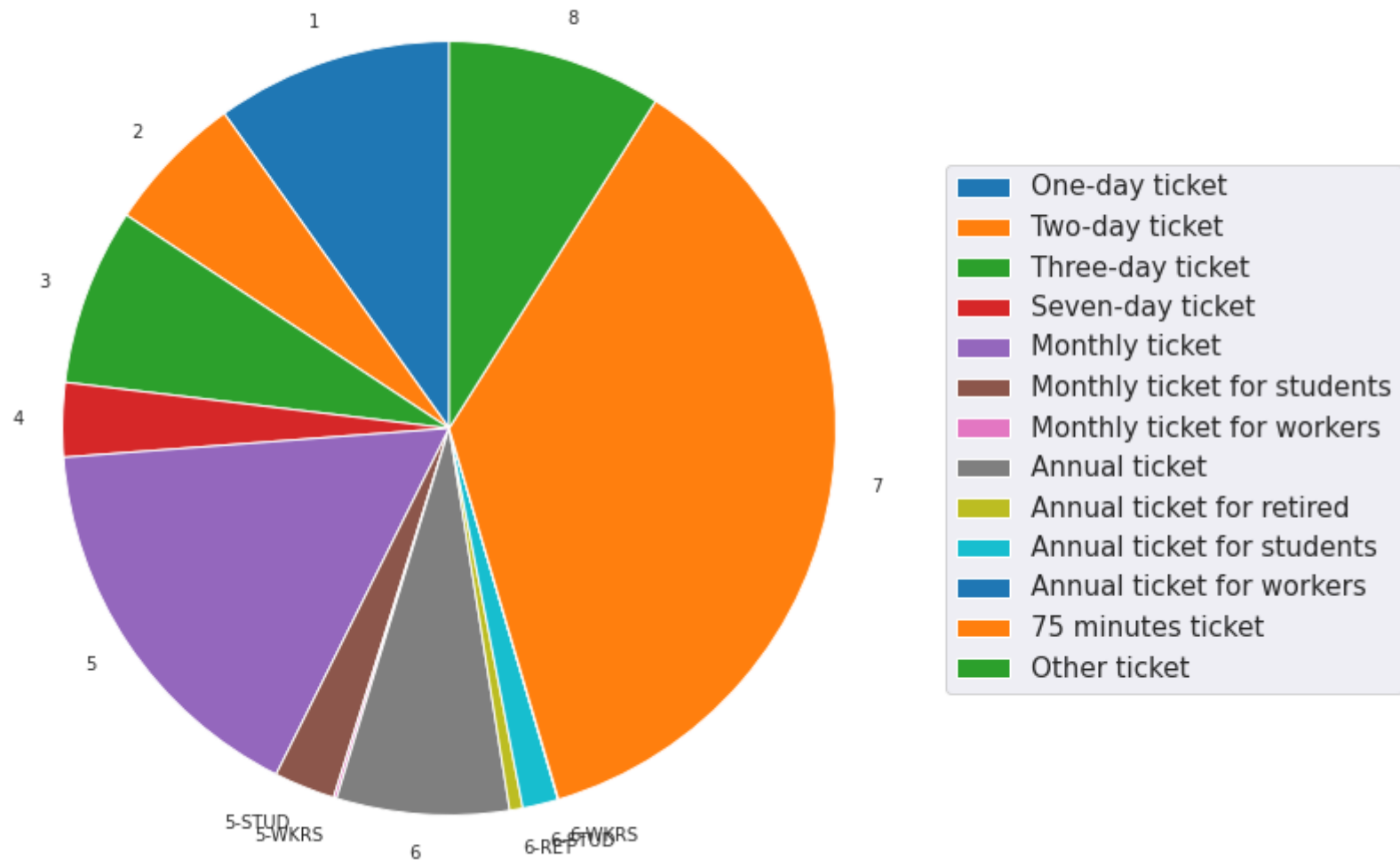


```
In [ ]: # Plot a pie chart of the column 'TICKET_CODE'
fig, ax = plt.subplots(figsize=(20,10))
df['TICKET_CODE'].value_counts().sort_index().plot.pie(startangle=90)

# Add the name of the ticket profile on the pie chart
plt.legend(labels=df['TICKET_CODE'].value_counts().sort_index().rename(dict_tickets).index, loc='center left', bbox

plt.title('Pie chart of the column TICKET_CODE', fontsize=20)
plt.ylabel('')
fig.patch.set_facecolor('white')
plt.show()
```

Pie chart of the column TICKET_CODE



Delete the validation that are with TICKET_CODE = 8 (other tickets)

```
In [ ]: shape_before = df.shape[0]

# Delete 8 tickets because they are not useful for the analysis
```

```
df = df[df['TICKET_CODE'] != '8']

# Print the number of rows before and after the deletion of the 8 tickets and the difference
print('The number of rows before the deletion of the 8 tickets is: {}'.format(shape_before))
print('The number of rows after the deletion of the 8 tickets is: {}'.format(df.shape[0]))
print('The difference is: {}'.format(shape_before - df.shape[0]))

# TODO: to de-comment later
```

The number of rows before the deletion of the 8 tickets is: 5537466
The number of rows after the deletion of the 8 tickets is: 5041376
The difference is: 496090

Data Cleaning

Stops similar

```
In [ ]: # Print the number of unique values in the column 'DESCRIZIONE' that are the names of the stops
print('The number of unique values in the column DESCRIZIONE is: {}'.format(df['DESCRIZIONE'].nunique()))
```

The number of unique values in the column DESCRIZIONE is: 885

```
In [ ]: def get_common_prefix(string_list):
        """
        This function returns the common prefix of a list of strings.
        If there is no common prefix, it returns an empty string.
        :param string_list: list of strings
        :return: string that is the common prefix of the list of strings
        """
        first_prefix = string_list[0].split(" ")[0]
        # Create and empty dictionary
        prefix_dict = {}

        # Iterate over the list of strings
        for string in string_list[1:]:
            # Check if the string starts with the first prefix
            if not string.startswith(first_prefix):
                # If the string does not start with the first prefix, split the string and take the first word
                first_prefix = string.split(" ")[0]
```

```

        if string.startswith(first_prefix):
            # In the dictionary add the new prefix as key and the list of strings that have this prefix as value
            prefix_dict[first_prefix] = [string for string in string_list if string.startswith(first_prefix)]
        else:
            # In the dictionary add the new prefix as key and the list of strings that have this prefix as value
            prefix_dict[first_prefix] = [string for string in string_list if string.startswith(first_prefix)]
    return prefix_dict

```

```

In [ ]: # To avoid problem regarding the letters in uppercase and lowercase, convert all the letters in uppercase
df['DESCRIZIONE'] = df['DESCRIZIONE'].str.upper()

```

```

In [ ]: # Use the function get_common_prefix to find the common prefix of the strings in the column 'DESCRIZIONE' and print

# Create a string list with the unique values of the column 'DESCRIZIONE'
string_list = df['DESCRIZIONE'].unique().tolist()

dict_prefix = get_common_prefix(string_list)
for key, value in dict_prefix.items():
    print('{}: {}'.format(key, value))

# Print the number of keys in the dictionary
print('The number of keys in the dictionary is: {}'.format(len(dict_prefix.keys())))

```

ZATTERE: ['ZATTERE "B"', 'ZATTERE "A"', 'ZATTERE']

S.: ['S. TOMA\' "B"', 'S. MARCO-SAN', 'S. MARCUOLA-', 'S. STAE', 'S. MARCO VAL', 'S.ELENA-STAD', 'S. MARIA DEL', 'S. SILVESTRO', 'S. ANGELO', 'S. ALVISE', 'S. GIORGIO', 'S. PIETRO DI', 'S. BASILIO', 'S. MARTA', 'S. SERVULO', 'S. ERASMO PU', 'S. ERASMO CA', 'S. ERASMO CH', 'S. SAMUELE', 'S. MARCO (GI', 'S. LAZZARO', 'S. TOMA\' "A"', 'S. M ARCO GIA']

VENEZIA: ['VENEZIA CORS', 'VENEZIA', 'VENEZIA PIAZ', 'VENEZIA RAMP']

P.LE: ['P.LE ROMA "G', 'P.LE ROMA "E', 'P.LE ROMA "A', 'P.LE ROMA "C', 'P.LE ROMA "D', 'P.LE ROMA "B', 'P.LE ROMA "F']

FERROVIA: ['FERROVIA "B"', 'FERROVIA "D"', 'FERROVIA "E"', 'FERROVIA PIO', 'FERROVIA "A"', 'FERROVIA "C"']

SAN: ['SAN MARCO CA', 'SAN MARCO SA', 'SAN MARCO-SA', "SANT'ANTONIO", 'SAN MARCO MO', "SAN DONA' MA", "SAN DONA' C E", "SAN DONA' RI", "SAN DONA' PI", 'SAN MARCO BO', "SAN DONA' PA", 'SAN MARCO MA', 'SAN LIBERALE', "SAN DONA' V A", "SAN DONA' FA", 'SAN MARCO FO', 'SANTA MARGHE', "SAN DONA' PE", 'SANTA MARIA ', 'SAN MARCO MU', "SAN NICOLO'", "SAN NICOLO'", 'SAN ROCCO BR', 'SANSOVINO VE', 'SAN PIETRO B', 'SANPIETRO CA', 'SAN PIETRO P', 'SANT'ANNA S', 'SAN PIETRO C', 'SANT'ANNA V', 'SAN TROVASO', 'SANBRUSON MA', "SANT'ANGELO ", 'SANBRUSON CA', 'SANDON', 'SAN GIUS EPP', 'SANT ANGELO ', "SANT' ALBERT"]

LIBERTA': ["LIBERTA' SAN", "LIBERTA' FIN"]

TRONCHETTO: ['TRONCHETTO F', 'TRONCHETTO "', 'TRONCHETTO M', 'TRONCHETTO T', 'TRONCHETTO V']

RIALTO: ['RIALTO "C"', 'RIALTO MERCA', 'RIALTO "D"', 'RIALTO "B"', 'RIALTO "A"']

AEROPORTO: ['AEROPORTO MA']

GIARDINI: ['GIARDINI BIE', 'GIARDINI "B"', 'GIARDINI "A"']

SOTTOMARINA: ['SOTTOMARINA ']

SANT'ANTONIO: ["SANT'ANTONIO"]

CA': ["CA' D'ORO", "CA' ROSSA VO", "CA' ROSSA SE", "CA' ROSSA BI", "CA' REZZONIC", "CA' SABBIONI", "CA' ROSSA OB", "CA' SOLARO C", "CA' MARCELLO", "CA' BRENTCELL", "CA' LIN ERAC", "CA' LIN CAST", "CA' BIANCA L", "CA' SOLARO P", "C A' LIN PITA", "CA' LIN GATT", "CA' BIANCA C", "CA' SOLARO O", "CA' BIANCA P"]

ALTINIA: ["ALTINIA CA' ", 'ALTINIA MUNI', 'ALTINIA INDR', 'ALTINIA FAVA', 'ALTINIA SAN ', 'ALTINIA BERI', 'ALTINIA 181', 'ALTINIA FORT']

MESTRE: ['MESTRE RAMP', 'MESTRE CENTR', 'MESTRE VIA M', 'MESTRE VIA H', 'MESTRE VIA T', 'MESTRE STAZI', 'MESTRE V IA P', 'MESTRE VIA A', 'MESTRE PIAZZ']

CHIOGGIA: ['CHIOGGIA', 'CHIOGGIA VIA', 'CHIOGGIA SAN', 'CHIOGGIA CAM', 'CHIOGGIA STA', 'CHIOGGIA BER', 'CHIOGGIA I SO', 'CHIOGGIA RID', 'CHIOGGIA PAR', 'CHIOGGIA VAL', 'CHIOGGIA NEN', 'CHIOGGIA CA', 'CHIOGGIA OSP']

SACCA: ['SACCA FISOLA']

LIDO: ['LIDO S.M.E. ', 'LIDO S. NICO']

GIUDECCA: ['GIUDECCA PAL']

ARSENALE: ['ARSENALE "B"', 'ARSENALE "A"']

OLIVI: ['OLIVI']

GALLO: ['GALLO BIBLIO', 'GALLO TORTA', 'GALLO MOROSI', 'GALLO GIOLIT', "GALLO SANT'A", 'GALLO BRAGAD', 'GALLO LORE DA', 'GALLO DEI KI', 'GALLO CONTAR', 'GALLO BARBAR', 'GALLO GALOPP', 'GALLO DEI GI']

MARTIRI: ['MARTIRI DELL']

S.ELENA-STAD: ['S.ELENA-STAD']

ZITELLE: ['ZITELLE "B"', 'ZITELLE "A"']
PORTO: ['PORTO MARGHE', 'PORTOSECCO', 'PORTOSECCO C', 'PORTOSECCO L']
REDENTORE: ['REDENTORE']
PUNTA: ['PUNTA SABBIO']
BORGO: ['BORGO SAN GI', 'BORGORICCO M', 'BORGORICCO', 'BORGORRICO S']
F.TE: ['F.TE NOVE "C"', 'F.TE NOVE "A"', 'F.TE NOVE "D"', 'F.TE NOVE "B"']
PADOVA: ['PADOVA SAN L', 'PADOVA FIERA', 'PADOVA TURAZ', 'PADOVA GOZZI', 'PADOVA EST', 'PADOVA OSPED']
BURANO: ['BURANO "C"', 'BURANO "A"', 'BURANO "B"']
TRE: ['TRE ARCHI', 'TREPORTI', 'TREVISO', 'TREVISO SELV', 'TREZZO BATTU', 'TREVISO FS', 'TRENTO FAGAR', 'TREZZO TE
RRA', 'TREVISO PIND', 'TRENTO GAZZE', 'TREVISO SAN ', 'TREVISO GHIR', 'TRENTO PODGO', 'TREVISO LORE']
MARGHERA: ['MARGHERA VIA', 'MARGHERA NAV', 'MARGHERA CIM', 'MARGHERA SAP', 'MARGHERA ROM', 'MARGHERA SAL']
MIRANESE: ['MIRANESE SAN', 'MIRANESE GIU', 'MIRANESE PER', 'MIRANESE PIE', 'MIRANESE SEL', 'MIRANESE IVA', 'MIRANE
SE VIV', 'MIRANESE CAL', 'MIRANESE PIA', 'MIRANESE LAZ', 'MIRANESE MON', 'MIRANESE AVA']
TRIESTE: ['TRIESTE MIRA', 'TRIESTE CATE', 'TRIESTE ERAC', 'TRIESTE ROBI', 'TRIESTE BOSC', 'TRIESTE MAZZ', 'TRIESTE
PARC', 'TRIESTE STAZ']
OLMO: ['OLMO CALVI', 'OLMO GIOVANN', 'OLMO MINZONI', 'OLMO PAPA LU', 'OLMO VITTORI', 'OLMO PELLICO']
ZELARINO: ['ZELARINO MUN', 'ZELARINO PAR', 'ZELARINO CHI', 'ZELARINO CAR']
CAMPALTO: ['CAMPALTO VIA', 'CAMPALTO PIA']
VALLI: ['VALLI PIAZZA', 'VALLI VIA PO', 'VALLI VALFIO', 'VALLI PONTE ']
LOVA: ['LOVA STRADA ']
BELFREDO: ['BELFREDO TER']
PIAVE: ['PIAVE FIUME', 'PIAVE PODGOR', 'PIAVE PUCCIN']
PERTINI: ['PERTINI CHIE', 'PERTINI FOSC', 'PERTINI CARR', 'PERTINI GAVA', 'PERTINI VESP']
TREPORTI: ['TREPORTI']
MURANO: ['MURANO NAVAG', 'MURANO COLON', 'MURANO VENIE', 'MURANO MUSEO', 'MURANO FARO ', 'MURANO SEREN', 'MURANO D
A MU']
STAZIONE: ['STAZIONE MES', 'STAZIONE PAD', 'STAZIONE MAR', 'STAZIONE DI ']
TESSERA: ['TESSERA VIA ', 'TESSERA SCUO']
CAMPAGNA: ['CAMPAGNA LUP']
ROSARA: ['ROSARA STRAD']
PASQUALIGO: ['PASQUALIGO M']
PARK: ['PARK PETROLI']
DESE: ['DESE CENTRO', 'DESE CICOGNE', 'DESE LITOMAR', 'DESE FS']
VALLENARI: ['VALLENARI ST']
MONTE: ['MONTE CELO F', 'MONTE CERVIN', 'MONTE GRAPPA']
CORSO: ['CORSO DEL PO']
BISSUOLA: ['BISSUOLA ESI', 'BISSUOLA CAD', 'BISSUOLA PIS', 'BISSUOLA COL', 'BISSUOLA TEV', 'BISSUOLA VAR']
PALEOCAPA: ['PALEOCAPA PA']
DOLO: ['DOLO CENTRO', 'DOLO MANZONI', 'DOLO CA' TRO', 'DOLO OSPEDAL', 'DOLO SCUOLE', 'DOLO MUNICIP', 'DOLO BANDIE
R', 'DOLO MONACHE', 'DOLO CIVICO ', 'DOLO DORIA', 'DOLO PRESICC', 'DOLO GUARDIA']

CASTELLANA: ['CASTELLANA P', 'CASTELLANA C', 'CASTELLANA S', 'CASTELLANA M', 'CASTELLANA B']
FIESSO: ['FIESSO GEMIT', 'FIESSO D'ART', 'FIESSO BARBA', 'FIESSO PIOVE', 'FIESSO PIOGH']
SALZANO: ['SALZANO MATT', 'SALZANO LORE', 'SALZANO DE G', 'SALZANO MUNI', 'SALZANO TOSC', 'SALZANO CHIE', 'SALZANO CONF', 'SALZANO ODDO', 'SALZANO PONT', 'SALZANO CIMI', 'SALZANO SOGA']
ASSEGGIANO: ['ASSEGGIANO M', 'ASSEGGIANO C', 'ASSEGGIANO E', 'ASSEGGIANO V', 'ASSEGGIANO P', 'ASSEGGIANO D']
ORLANDA: ['ORLANDA PINE', 'ORLANDA CASI', 'ORLANDA SAN ', 'ORLANDA PIOV', 'ORLANDA BAGA', 'ORLANDA DON ', 'ORLANDA CENT', 'ORLANDA 200']
TERRAGLIO: ['TERRAGLIO PE', 'TERRAGLIO NI', 'TERRAGLIO BO', 'TERRAGLIO TE', 'TERRAGLIO VI', 'TERRAGLIO GA', 'TERRAGLIO CA', 'TERRAGLIO FA']
PIOVE: ['PIOVE DI SAC']
MALCONTENTA: ['MALCONTENTA ']
27: ['27 OTTOBRE D']
TRIVIGNANO: ['TRIVIGNANO B', 'TRIVIGNANO P', 'TRIVIGNANO L', 'TRIVIGNANO G', 'TRIVIGNANO C']
PASSO: ['PASSO CAMPAL']
GARIBALDI: ['GARIBALDI MU', 'GARIBALDI C', 'GARIBALDI OG', 'GARIBALDI GI']
MIRA: ['MIRANESE SAN', 'MIRANESE GIU', 'MIRANESE PER', 'MIRANESE PIE', 'MIRANESE SEL', 'MIRA PORTE', 'MIRANESE IV A', 'MIRANESE VIV', 'MIRANESE CAL', 'MIRA RISCOSS', 'MIRANESE PIA', 'MIRA EGO', 'MIRANESE LAZ', 'MIRANESE MON', 'MIRANO FOSSA', 'MIRANO CENTR', 'MIRANESE AVA', 'MIRA ALIGHIE', 'MIRANO MATTE', 'MIRANO GRIMA', 'MIRANO SPORT', 'MIRANO SCUOL', 'MIRA BELLINI', 'MIRA CENTRO', 'MIRA MONTESS', 'MIRANO GRAMS', 'MIRANO BATTI', 'MIRA PIAZZA ', 'MIRANO PESTR', 'MIRANO TREV', 'MIRA ALBRIZZ', 'MIRANO DANTE', 'MIRANO CARDU', 'MIRA RIVIERA', 'MIRA CAMPI', 'MIRANO B OLLA', 'MIRANO CARAV', 'MIRA BERNINI', 'MIRANOTAGLIO', 'MIRANO LOCAL', 'MIRANO MARIU', 'MIRANO PERUG', 'MIRA PERTI NI', 'MIRA 25 APRI', 'MIRANO GALIL', 'MIRANO BOSCH', 'MIRA PONTE D', 'MIRANO OLMO', 'MIRA SAN MON', 'MIRA MUNICI P', 'MIRANO MINZ', 'MIRANO MODIG', 'MIRANO VAROT', 'MIRA MARE ME', 'MIRA GINESTR', 'MIRANO BARBA', 'MIRANO BOS C', 'MIRANO LUNEO']
FORTE: ['FORTE MARGHE']
SABBADINO: ['SABBADINO BA', 'SABBADINO L', 'SABBADINO PA', 'SABBADINO LA']
GOBBI: ['GOBBI CA' DO', 'GOBBI MANDAR', 'GOBBI DON BO', 'GOBBI SAN DO', 'GOBBI ORLAND', 'GOBBI MIRTIL', 'GOBBI VAL LEN']
DON: ['DON STURZO V', 'DON STURZO P']
CASONA: ['CASONA VALLE', 'CASONA BISSU', 'CASONA MARZI']
CARDUCCI: ['CARDUCCI FEL', 'CARDUCCI PAS']
CAPPUCCINA: ['CAPPUCCINA B', 'CAPPUCCINA V', 'CAPPUCCINA S']
CIRCONVALLAZ: ['CIRCONVALLAZ']
MALAMOCCO: ['MALAMOCCO CE', 'MALAMOCCO BA', 'MALAMOCCO AL', 'MALAMOCCO ST', 'MALAMOCCO BE', 'MALAMOCCO OC', 'MALAMOCCO GA', 'MALAMOCCO PA']
BACINI: ['BACINI - ARS']
CAMPONOGARA: ['CAMPONOGARA ', 'CAMPONOGARA']
PADANA: ['PADANA CIVIC', 'PADANA DELLE', 'PADANA STAZI', 'PADANA DEL L']
PELLESTRINA: ['PELLESTRINA ']
BECCARIA: ['BECCARIA COR', 'BECCARIA CHI', 'BECCARIA CAN', 'BECCARIA ORO', 'BECCARIA PAR']

CREA: ['CREA']
PESEGGIA: ['PESEGGIA CA', 'PESEGGIA', 'PESEGGIA PER']
BANDIERA: ['BANDIERA GHE']
TEVERE: ['TEVERE BAGLI', 'TEVERE PARCO']
TRIESTINA: ['TRIESTINA PR', 'TRIESTINA MO', 'TRIESTINA TO', 'TRIESTINA AL', 'TRIESTINA LA', 'TRIESTINA UL', 'TRIESTINA ZO', 'TRIESTINA PI', 'TRIESTINA PA', 'TRIESTINA AE', 'TRIESTINA FO', 'TRIESTINA ZU', 'TRIESTINA SC']
MARTELLAGO: ['MARTELLAGO', 'MARTELLAGO V', 'MARTELLAGO S', 'MARTELLAGO D', 'MARTELLAGO P', 'MARTELLAGOTR', 'MARTELLAGO F', 'MARTELLAGO C', 'MARTELLAGO A', 'MARTELLAGO G']
GUGLIE: ['GUGLIE "A"', 'GUGLIE "B"']
SCORZE: ['SCORZE' ROMA", "SCORZE' MUNI", "SCORZE' CAPO", "SCORZE' MOGL", "SCORZE' FERM", "SCORZE' VENE", "SCORZE' ORTI", "SCORZE' BOSC"]
MARCON: ['MARCON GENOV', 'MARCON OBERD', 'MARCON CULT', 'MARCON DELLA', 'MARCON MUNIC', 'MARCON MARMO', 'MARCONI CORA', 'MARCONI MARC', 'MARCON LOMBA', 'MARCON MATTE', 'MARCON COOPE', 'MARCON STADI', 'MARCONI LORE', 'MARCON VITTO', 'MARCON REPUB', 'MARCON CULTU', 'MARCON ALTIN', 'MARCON PEROS', 'MARCONI FIUM', 'MARCON MILAN', 'MARCONI DES', 'MARCON ANCON']
TREVISO: ['TREVISO', 'TREVISO SELV', 'TREVISO FS', 'TREVISO PIND', 'TREVISO SAN ', 'TREVISO GHIR', 'TREVISO LORE']
MAERNE: ['MAERNE ISONZ', 'MAERNE CHIES', 'MAERNE MULIN', 'MAERNE CENTR', 'MAERNE FS', 'MAERNE TASSO', 'MAERNE CIVIC', 'MAERNE CIMIT', 'MAERNE GUARD', 'MAERNE CIRCO']
CELESTIA: ['CELESTIA']
MADONNA: ['MADONNA DELL', 'MADONNA MARI']
CAPOLINEA: ['CAPOLINEA CA', 'CAPOLINEA FU']
RIVA: ['RIVA DE BIAS', 'RIVALE CHIES']
ILARIA: ['ILARIA ALPI ']
ACCADEMIA: ['ACCADEMIA "B', 'ACCADEMIA "A']
ROBEGANO: ['ROBEGANO CEN', 'ROBEGANO PUC', 'ROBEGANO MON', 'ROBEGANO 25 ']
MAZZORBO: ['MAZZORBO']
OSPEDALE: ['OSPEDALE DEL', 'OSPEDALE MIR', 'OSPEDALE']
SALAMONIO: ['SALAMONIO MA']
CALABRIA: ['CALABRIA CAM']
PEOPLE: ['PEOPLE MOVER']
TITO: ['TITO CASTELL', 'TITO SELVANE']
NOALE: ['NOALE', 'NOALE OSPEDA', 'NOALE DEI NO', 'NOALE BACCHI', 'NOALE ONGARI', 'NOALE MORO', 'NOALE ZONA I', 'NOALE LIVENZ', 'NOALE LANCER']
VILLABONA: ['VILLABONA PI', 'VILLABONA BO', 'VILLABONA MO', 'VILLABONA 87', 'VILLABONA 8', 'VILLABONA ON']
SANTA: ['SANTA MARGHE', 'SANTA MARIA ']
PERON: ['PERON BASEGG', 'PERON SARAGA']
SALUTE: ['SALUTE']
ROMEA: ['ROMEA CIMITE', 'ROMEA CIVICO', 'ROMEA FOSSET', 'ROMEA MARINE', 'ROMEI CIVIC', 'ROMEA DEL BO', 'ROMEA FISOLA', 'ROMEA PRIMAV']
ORIAGO: ['ORIAGO STAZI', 'ORIAGO CENTR', 'ORIAGO SOMMO', 'ORIAGO ROMAG', 'ORIAGO VENEZ', 'ORIAGO FERRO']

PIAZZALE: ['PIAZZALE GIO', 'PIAZZALE RAV']
RIVALE: ['RIVALE CHIES']
CONCHE: ['CONCHE STRAD']
ALBERONI: ['ALBERONI DEL', 'ALBERONI FAR', 'ALBERONI OTT', 'ALBERONI SAN', 'ALBERONI STE', 'ALBERONI GOL', 'ALBERONI CA', 'ALBERONI COL']
PIAZZA: ['PIAZZALE GIO', 'PIAZZA MERCA', 'PIAZZALE RAV']
GAMBARARE: ['GAMBARARE VI', 'GAMBARARE CI', 'GAMBARARE PO', 'GAMBARARE']
MOGLIANO: ['MOGLIANO RON', 'MOGLIANO CEN', 'MOGLIANO RIM', 'MOGLIANO MAR', 'MOGLIANO RAG', 'MOGLIANO TOM', 'MOGLIANO MUN', 'MOGLIANO CAM', 'MOGLIANO LIC', 'MOGLIANO FS', 'MOGLIANO CIM', 'MOGLIANO TER', 'MOGLIANO DEI', 'MOGLIANO BUR', 'MOGLIANO BEL', 'MOGLIANO ZER', 'MOGLIANO COR', 'MOGLIANO MEU', 'MOGLIANO GHE']
PAOLUCCI: ['PAOLUCCI LON']
BOTTENIGO: ['BOTTENIGO CA', 'BOTTENIGO PI', 'BOTTENIGO BO', 'BOTTENIGO MA']
SPINEA: ['SPINEA PIAZZ', 'SPINEA MARTI', 'SPINEA ORGNA', 'SPINEA POZZU', 'SPINEA ALFIE', 'SPINEA GIORG', 'SPINEA R EPUB', 'SPINEA SANRE', 'SPINEA ROSSI', 'SPINEA CENTR', 'SPINEA DESEN', 'SPINEA VILLA', 'SPINEA SAN R', 'SPINEA LUN EO']
CANAL: ['CANAL LEONE']
PREGANZIOL: ['PREGANZIOL M', 'PREGANZIOL ', 'PREGANZIOL F', 'PREGANZIOL G', 'PREGANZIOL B']
RISORGIMENTO: ['RISORGIMENTO']
FARO: ['FARO ROCCHET']
FAVRETTI: ['FAVRETTI MES']
PALIAGA: ['PALIAGA CA']
BORBIAGO: ['BORBIAGO MIL', 'BORBIAGO CEN']
CAPPELLA: ['CAPPELLA']
RINASCITA: ['RINASCITA EM', 'RINASCITA BE']
MATTUGLIE: ['MATTUGLIE PE', 'MATTUGLIE DI']
BRENDOLE: ['BRENDOLE ARS', 'BRENDOLE STI', 'BRENDOLE FAV', 'BRENDOLE PRO', 'BRENDOLE']
CORRENTI: ['CORRENTI CAP']
VIGONZA: ['VIGONZA PERA', 'VIGONZA BACH']
GIOVANNACCI: ['GIOVANNACCI']
CALUCCI: ['CALUCCI CIME', 'CALUCCI QUAR']
RIELTA: ['RIELTA PARCO', 'RIELTA CA R']
MIRANO: ['MIRANO FOSSA', 'MIRANO CENTR', 'MIRANO MATTE', 'MIRANO GRIMA', 'MIRANO SPORT', 'MIRANO SCUOL', 'MIRANO G RAMS', 'MIRANO BATTI', 'MIRANO PESTR', 'MIRANO TREVI', 'MIRANO DANTE', 'MIRANO CARDU', 'MIRANO BOLLA', 'MIRANO CAR AV', 'MIRANOTAGLIO', 'MIRANO LOCAL', 'MIRANO MARIU', 'MIRANO PERUG', 'MIRANO GALIL', 'MIRANO BOSCH', 'MIRANO OLM O', 'MIRANO MINZ', 'MIRANO MODIG', 'MIRANO VAROT', 'MIRANO BARBA', 'MIRANO BOSC', 'MIRANO LUNEO']
LAVELLI: ['LAVELLI PAOL']
DE: ['DESE CENTRO', 'DESE CICOGNE', 'DESE LITOMAR', 'DE NICOLA CH', 'DEI MURAZZI', 'DESE FS', 'DELLE MESSI']
EINAUDI: ['EINAUDI CAST']
VIGNOLE: ['VIGNOLE']
ZENDRINI: ['ZENDRINI VIL']

GIARE: ['GIARE STRADA']
ISOLA: ['ISOLA UNIONE', 'ISOLA VERDE', 'ISOLA VERDE ']
TORINO: ['TORINO ROSSE', 'TORINO', 'TORINO UNIVE']
ANCONA: ['ANCONA CARBO', 'ANCONA TORIN']
STRA: ['STRA DANTE', 'STRA PIAZZA ', 'STRA LOREDAN', 'STRA FOSSOLO']
FITTIZIA: ['FITTIZIA']
GOZZI: ['GOZZI CAPPUC']
SAMBUGHE: ['SAMBUGHE']
CAVARZERE: ['CAVARZERE VI', 'CAVARZERE AU', 'CAVARZERE C']
MORANZANI: ['MORANZANI 32', 'MORANZANI EL', 'MORANZANI CO']
LUGHETTO: ['LUGHETTO STR', 'LUGHETTO MAR', 'LUGHETTO DI ']
CAZZAGHETTO: ['CAZZAGHETTO ']
CAVERGNAGO: ['CAVERGNAGO T', 'CAVERGNAGO M']
CIRCONVALAZI: ['CIRCONVALAZI']
TREZZO: ['TREZZO BATTU', 'TREZZO TERRA']
AGENZIA: ['AGENZIA ENT']
TORCELLO: ['TORCELLO']
CIMITERO: ['CIMITERO SAN']
GATTA: ['GATTA SCARAN', 'GATTA PISACA', 'GATTA SANTA ', 'GATTA 90', 'GATTA SCARAM', 'GATTA PROTAG', 'GATTA IMMAC
O', 'GATTA VERCI']
SALICI: ['SALICI VILLA']
GAZZERA: ['GAZZERA ALTA']
QUARNARO: ['QUARNARO CAL']
CAFASSO: ['CAFASSO BOTT']
TOSATTO: ['TOSATTO IMPA', 'TOSATTO PACC']
PASINI: ['PASINI FRATE', 'PASINI LAVOR']
CAVALCAVIA: ['CAVALCAVIA V']
LUGO: ['LUGO STRADA ', 'LUGO ZONA IN']
SCUOLA: ['SCUOLA MEDIA', 'SCUOLA ZENDR']
CAVANIS: ['CAVANIS CAPO']
BUSA: ['BUSA DI VIGO']
D'ANNUNZIO: ['D'ANNUNZIO P", "D'ANNUNZIO S", "D'ANNUNZIO D"]
CAPRICCIO: ['CAPRICCIO CE']
CALVI: ['CALVI PARMES']
MARSALA: ['MARSALA CENT']
CERTOSA: ['CERTOSA', 'CERTOSA A RI']
SPIRITO: ['SPIRITO SANT']
MARCONI: ['MARCONI CORA', 'MARCONI MARC', 'MARCONI LORE', 'MARCONI FIUM', 'MARCONI DES ']
CAVANELLA: ['CAVANELLA PI', 'CAVANELLA', 'CAVANELLA BO']
CALCROCI: ['CALCROCI', 'CALCROCI CAV', 'CALCROCI CHI']

DEI: ['DEI MURAZZI ']
VISINONI: ['VISINONI COM', 'VISINONI POL', 'VISINONI SEL']
PALAZZO: ['PALAZZO DEL ']
PORTOSECCH: ['PORTOSECCH', 'PORTOSECCH C', 'PORTOSECCH L']
SANSOVINO: ['SANSOVINO VE']
LUNGOMARE: ['LUNGOMARE AD']
GAGGIO: ['GAGGIO CIMIT', 'GAGGIO', 'GAGGIO FERMI', 'GAGGIO VIVAL']
GALILEI: ['GALILEI DARS']
COLOMBO: ['COLOMBO']
MAZZOCCO: ['MAZZOCCO']
NAZIONI: ['NAZIONI UNIT']
TRENTO: ['TRENTO FAGAR', 'TRENTO GAZZE', 'TRENTO PODGO']
BRONDOLO: ['BRONDOLO', 'BRONDOLO DOL']
VIA: ['VIA DEI CANT', 'VIA VILLABON']
VESPUCCI: ['VESPUCCI GAR', 'VESPUCCI SAN', "VESPUCCI CA'", 'VESPUCCI GRI', 'VESPUCCI PIG', 'VESPUCCI CAT', 'VESPUCCI BOE']
AREOPORTO: ['AREOPORTO MA']
FOSSO: ["FOSSO' CENTR", "FOSSO' ARZAR", "FOSSO' RONCA", "FOSSO' FAVAL", "FOSSO' PROVI", "FOSSO' ZONA ", "FOSSO' B OSEL", "FOSSO'"]
BOJON: ['BOJON', 'BOJON FS', 'BOJON RIVELL']
GALTA: ['GALTA ARGINE']
LIETTOLI: ['LIETTOLI TRE']
FRESCADA: ['FRESCADA']
PROZZOLO: ['PROZZOLO', 'PROZZOLOTOGL']
SELVANESE: ['SELVANESE PL']
GRAN: ['GRAN VIALE']
CAROMAN: ['CAROMAN']
PETTORAZZA: ['PETTORAZZA S']
SAMBRUSON: ['SAMBRUSON', 'SAMBRUSON MA']
PAGANELLO: ['PAGANELLO TI']
FISICA: ['FISICA DEPOS']
VALLON: ['VALLON FORTE', 'VALLON VALDE', 'VALLON BORGO', 'VALLON DE NI']
COLMELLO: ['COLMELLO']
PERAROLO: ['PERAROLO QUA']
CAMPORESE: ['CAMPORESE GR']
SANT'ANNA: ['SANT'ANNA S', 'SANT'ANNA V']
RONZINELLA: ['RONZINELLA F']
AZOTO: ['AZOTO SOTTAN']
CASALE: ['CASALE SUL S', 'CASALE BIVIO']
ADRIA: ['ADRIA SCUOLE', 'ADRIA MARCON', 'ADRIA FS', 'ADRIA OSPEDA']

DELLE: ['DELLE MESSI ']
PONTE: ['PONTE DI BRE']
FUSINA: ['FUSINA CENTR']
DOSSON: ['DOSSON']
MARTELLAGOTR: ['MARTELLAGOTR']
GARDIGIANO: ['GARDIGIANO']
OSPIZIO: ['OSPIZIO MARI']
CALTANA: ['CALTANA MARI', 'CALTANA', 'CALTANA PIOG']
VIGONOVO: ['VIGONOVO GAL', 'VIGONOVO', 'VIGONOVO 1 M', 'VIGONOVO DE ', 'VIGONOVO ALF']
CA: ['CA' D'ORO", 'CAMPALTO VIA', 'CAMPAGNA LUP', 'CASTELLANA P', 'CASTELLANA C', "CA' ROSSA VO", 'CASONA VALLE',
'CARDUCCI FEL', 'CASONA BISSU', 'CAPPUCCINA B', 'CAMPONOGARA ', 'CARDUCCI PAS', "CA' ROSSA SE", "CA' ROSSA BI", 'C
APOLINEA CA', 'CASTELLANA S', 'CALABRIA CAM', 'CAPPUCCINA V', "CA' REZZONIC", 'CAPPUCCINA S', "CA' SABBIONI", "CA'
ROSSA OB", "CA' SOLARO C", 'CANAL LEONE', 'CASONA MARZI', 'CAPPELLA', 'CALUCCI CIME', 'CASTELLANA M', 'CAMPALTO PI
A', 'CAVARZERE VI', 'CAZZAGHETTO ', 'CAVERGNAGO T', "CA' MARCELLO", 'CAVERGNAGO M', 'CAFASSO BOTT', 'CAVALCAVIA
V', 'CASTELLANA B', 'CAVANIS CAPO', 'CAPRICCIO CE', 'CALVI PARMES', 'CAVANELLA PI', 'CALCROCI', "CA' BRENTTELL", 'C
AROMAN', 'CALCROCI CAV', 'CAMPONOGARA', "CA' LIN ERAC", 'CAMPORESE GR', "CA' LIN CAST", 'CASALE SUL S', 'CAVARZERE
AU', "CA' BIANCA L", 'CALTANA MARI', 'CA SOLARO PA', 'CALTANA', 'CALUCCI QUAR', 'CAPOLINEA FU', "CA' SOLARO P", "C
A' LIN PITA", "CA' LIN GATT", "CA' BIANCA C", 'CALCROCI CHI', 'CAMPOCROCE', "CA' SOLARO O", 'CAVANELLA', "CA' BIAN
CA P", 'CAZZAGO PASC', 'CA' BIANCA ', 'CAZZAGO', 'CAMPOVERARDO', 'CAVARZERE C', 'CASALE BIVIO', 'CALTANA PIOG',
'CAVANELLA BO', 'CAMPOLONGO L', 'CAMPOCROCE C', 'CAMPOLONGO 8']
LE: ['LE GRAZIE', 'LE CRETE']
MONIEGO: ['MONIEGO TREV']
FAVIGNANA: ['FAVIGNANA']
MORANDI: ['MORANDI NICE']
TERMINAL: ['TERMINAL RO-']
BERNINI: ['BERNINI DI V']
CORTIVO: ['CORTIVO TOMB']
RIO: ['RIO SAN MART']
ZIANIGO: ['ZIANIGO']
FORNASE: ['FORNASE PERU']
MELLAREDO: ['MELLAREDO CA']
ZERO: ['ZERO BRANCO', 'ZERO BRANCO ']
SFMR: ['SFMR SPINEA']
RUSTEGHELLO: ['RUSTEGHELLO']
QUARTO: ["QUARTO D'ALT"]
MIRANOTAGLIO: ['MIRANOTAGLIO']
SANBRUSON: ['SANBRUSON MA', 'SANBRUSON CA']
MAIANO: ['MAIANO ALBER']
CAMPOCROCE: ['CAMPOCROCE', 'CAMPOCROCE C']
ROMEAI: ['ROMEAI CIVIC']

ALTA: ['ALTA MILANO']
SOPPRESSA: ['SOPPRESSA - ']
SCALTENIGO: ['SCALTENIGO F', 'SCALTENIGO P', 'SCALTENIGO']
P.ZZA: ['P.ZZA DELLA ']
BORGORICCO: ['BORGORICCO M', 'BORGORICCO']
STIGLIANO: ['STIGLIANO']
CAZZAGO: ['CAZZAGO PASC', 'CAZZAGO']
MARANO: ['MARANO']
SANT'ANGELO: ["SANT'ANGELO "]
CA': ['CA' BIANCA ']
CORTE: ['CORTE']
ULSS: ['ULSS']
CAMPOVERARDO: ['CAMPOVERARDO']
PIANIGA: ['PIANIGA ROMA', 'PIANIGA BOSC']
MATTEOTTI: ['MATTEOTTI CE']
BORGORRICO: ['BORGORRICO S']
MESTRINA: ['MESTRINA SPA']
ROTTANOVA: ['ROTTANOVA CO']
TORRE: ['TORRE ENAC']
VETERNIGO: ['VETERNIGO', 'VETERNIGO PI']
BORROMINI: ['BORROMINI BE']
PROZZOLOTOGL: ['PROZZOLOTOGL']
BADOERE: ['BADOERE']
FERM.SERV.: ['FERM.SERV. D']
CAMPOLONGO: ['CAMPOLONGO L', 'CAMPOLONGO 8']
SANDON: ['SANDON']
BOSCHETTA: ['BOSCHETTA']
SCANDOLARA: ['SCANDOLARA']
CHIMICA: ['CHIMICA INGR']
SANT': ["SANT'ANTONIO", "SANT'ANGELO ", "SANT' ALBERT"]
VETREGO: ["VETREGO PRA'", 'VETREGO']
GRAMSCI: ['GRAMSCI TOGL']
PALUELLO: ['PALUELLO VEN']
LAZZARETTO: ['LAZZARETTO N']
VILLA: ['VILLABONA PI', 'VILLABONA BO', 'VILLABONA MO', 'VILLABONA 87', 'VILLABONA 8', 'VILLABONA ON', 'VILLA DEL BO']
The number of keys in the dictionary is: 292

Update some keys in the dictionary

```
In [ ]: # Rename the key 'P.le' with 'P.le Roma'
dict_prefix['P.LE ROMA'] = dict_prefix.pop('P.LE')
# Rename the key 'F.TE' with 'F.TE NOVE'
dict_prefix['F.TE NOVE'] = dict_prefix.pop('F.TE')
```

```
In [ ]: # Print the values of the dictionary with the keys 'S.' and 'San'
print('The values of the dictionary with the key S. are: {}'.format(dict_prefix['S.']))
print('The values of the dictionary with the key San are: {}'.format(dict_prefix['SAN']))
```

The values of the dictionary with the key S. are: ['S. TOMA\ 'B"', 'S. MARCO-SAN', 'S. MARCUOLA-', 'S. STAE', 'S. MARCO VAL', 'S.ELENA-STAD', 'S. MARIA DEL', 'S. SILVESTRO', 'S. ANGELO', 'S. ALVISE', 'S. GIORGIO', 'S. PIETRO D I', 'S. BASILIO', 'S. MARTA', 'S. SERVULO', 'S. ERASMO PU', 'S. ERASMO CA', 'S. ERASMO CH', 'S. SAMUELE', 'S. MARC O (GI', 'S. LAZZARO', 'S. TOMA\ 'A"', 'S. MARCO GIA']

The values of the dictionary with the key San are: ['SAN MARCO CA', 'SAN MARCO SA', 'SAN MARCO-SA', "SANT'ANTONI O", 'SAN MARCO MO', "SAN DONA' MA", "SAN DONA' CE", "SAN DONA' RI", "SAN DONA' PI", 'SAN MARCO BO', "SAN DONA' P A", 'SAN MARCO MA', 'SAN LIBERALE', "SAN DONA' VA", "SAN DONA' FA", 'SAN MARCO FO', 'SANTA MARGHE', "SAN DONA' P E", 'SANTA MARIA ', 'SAN MARCO MU', "SAN NICOLO' ", "SAN NICOLO'", 'SAN ROCCO BR', 'SANSOVINO VE', 'SAN PIETRO B', 'SANPIETRO CA', 'SAN PIETRO P', 'SANT'ANNA S', 'SAN PIETRO C', 'SANT'ANNA V', 'SAN TROVASO', 'SANBRUSON MA', "SAN T'ANGELO ", 'SANBRUSON CA', 'SANDON', 'SAN GIUSEPP', 'SANT ANGELO ', "SANT' ALBERT"]

S.Erasmo

```
In [ ]: # Create a new key in the dictionary with the key S.ERASMO; insert as value the list of strings that have the prefix
dict_prefix['S.ERASMO'] = [string for string in dict_prefix['S.'] if string.startswith('S.ERASMO')]

# Add the value 'S. Erasmo Pu' originally in the key 'San' to the key 'S.ERASMO'
dict_prefix['S.ERASMO'].append('S. ERASMO PU')

# Remove the strings that have the prefix 'S.ERASMO' from the keys 'S.' and 'San'
dict_prefix['S.'] = [string for string in dict_prefix['S.'] if not string.startswith('S.ERASMO')]
dict_prefix['S.'] = [string for string in dict_prefix['S.'] if not string.startswith('S. ERASMO PU')]

# Print the values of the dictionary with the key 'S.ERASMO'
print('The values of the dictionary with the key S.ERASMO are: {}'.format(dict_prefix['S.ERASMO']))
```

The values of the dictionary with the key S.ERASMO are: ['S. ERASMO PU']

San Marco

```
In [ ]: # Create a new key in the dictionary with the key 'San Marco'; insert as value the list of strings that have the pr
dict_prefix['SAN MARCO'] = [string for string in dict_prefix['SAN'] if string.startswith('SAN MARCO')]

# Add the value S. MARCO (Gi', 'S. Pietro in Gu') originally in the key 'S.' to the key 'San Marco'
dict_prefix['SAN MARCO'].append('S. MARCO (GI')

# Remove the strings that have the prefix 'San Marco' from the keys 'S.' and 'San'
dict_prefix['SAN'] = [string for string in dict_prefix['SAN'] if not string.startswith('SAN MARCO')]
dict_prefix['S.'] = [string for string in dict_prefix['S.'] if not string.startswith('S. MARCO (GI')]

# Print the values of the dictionary with the key 'San Marco'
print('The values of the dictionary with the key San Marco are: {}'.format(dict_prefix['SAN MARCO']))
```

The values of the dictionary with the key San Marco are: ['SAN MARCO CA', 'SAN MARCO SA', 'SAN MARCO-SA', 'SAN MARCO MO', 'SAN MARCO BO', 'SAN MARCO MA', 'SAN MARCO FO', 'SAN MARCO MU', 'S. MARCO (GI']

San Dona'

```
In [ ]: # Create a new key in the dictionary with the key 'San Dona'; insert as value the list of strings that have the pre
dict_prefix['SAN DONA'] = [string for string in dict_prefix['SAN'] if string.startswith('SAN DONA')]

# Remove the strings that have the prefix 'San Dona' from the keys 'S.' and 'San'
dict_prefix['SAN'] = [string for string in dict_prefix['SAN'] if not string.startswith('SAN DONA')]

# Print the values of the dictionary with the key 'San Dona'
print('The values of the dictionary with the key San Dona are: {}'.format(dict_prefix['SAN DONA']))
```

The values of the dictionary with the key San Dona are: ['SAN DONA' MA", "SAN DONA' CE", "SAN DONA' RI", "SAN DONA' PI", "SAN DONA' PA", "SAN DONA' VA", "SAN DONA' FA", "SAN DONA' PE"]

San Pietro

```
In [ ]: # Create a new key in the dictionary with the key 'San Pietro'; insert as value the list of strings that have the w
dict_prefix['SAN PIETRO'] = [string for string in dict_prefix['SAN'] if 'PIETRO' in string] + [string for string in

# Remove the strings that have the word 'Pietro' from the keys 'S.' and 'San'
dict_prefix['SAN'] = [string for string in dict_prefix['SAN'] if 'PIETRO' not in string]
dict_prefix['S.'] = [string for string in dict_prefix['S.'] if 'PIETRO' not in string]
```

```
# Print the values of the dictionary with the key 'San Pietro'
```

```
print('The values of the dictionary with the key San Pietro are: {}'.format(dict_prefix['SAN PIETRO']))
```

The values of the dictionary with the key San Pietro are: ['SAN PIETRO B', 'SANPIETRO CA', 'SAN PIETRO P', 'SAN PIETRO C', 'S. PIETRO DI']

Ca' Rossa

```
In [ ]: # Create a new key in the dictionary with the key 'Ca' Rossa'; insert as value the list of strings that have the wo  
dict_prefix['CA\ ' ROSSA'] = [string for string in dict_prefix['CA\ '] if 'CA' in string and 'ROSSA' in string]
```

```
# Remove the strings that have the word 'Ca' Rossa' from the keys 'Ca'
```

```
dict_prefix['CA\ '] = [string for string in dict_prefix['CA\ '] if 'CA' not in string or 'ROSSA' not in string]
```

```
# Print the values of the dictionary with the key 'Ca Rossa'
```

```
print('The values of the dictionary with the key Ca\ ' Rossa are: {}'.format(dict_prefix['CA\ ' ROSSA']))
```

The values of the dictionary with the key Ca' Rossa are: ["CA' ROSSA V0", "CA' ROSSA SE", "CA' ROSSA BI", "CA' ROS SA OB"]

Manage the remaining values in the keys 'S.' and 'San' and others

```
In [ ]: # Manage the remaining values in the keys 'S.', 'San', 'Santa', 'Sant', 'Ca', 'Piazza', 'Piazzale', 'Stazione', '  
# Create a new key for each value in the keys as above and assign the value as value of the new key  
# Remove the values from the keys as above
```

```
if 'S.' in dict_prefix:  
    for value in dict_prefix['S.']:   
        dict_prefix[value] = [value]  
    dict_prefix.pop('S.')
```

```
if 'SAN' in dict_prefix:  
    for value in dict_prefix['SAN']:   
        dict_prefix[value] = [value]  
    dict_prefix.pop('SAN')
```

```
if 'SANTA' in dict_prefix:  
    for value in dict_prefix['SANTA']:   
        dict_prefix[value] = [value]  
    dict_prefix.pop('SANTA')
```



```
if 'SANT\' ' in dict_prefix:
    for value in dict_prefix['SANT\'']:
        dict_prefix[value] = [value]
    dict_prefix.pop('SANT\' ')

if 'CA\' ' in dict_prefix:
    for value in dict_prefix['CA\'']:
        dict_prefix[value] = [value]
    dict_prefix.pop('CA\' ')

if 'PIAZZA' in dict_prefix:
    for value in dict_prefix['PIAZZA']:
        dict_prefix[value] = [value]
    dict_prefix.pop('PIAZZA')

if 'PIAZZALE' in dict_prefix:
    for value in dict_prefix['PIAZZALE']:
        dict_prefix[value] = [value]
    dict_prefix.pop('PIAZZALE')

if 'VIA' in dict_prefix:
    for value in dict_prefix['VIA']:
        dict_prefix[value] = [value]
    dict_prefix.pop('VIA')

if 'STAZIONE' in dict_prefix:
    for value in dict_prefix['STAZIONE']:
        dict_prefix[value] = [value]
    dict_prefix.pop('STAZIONE')

if 'TREVISO' in dict_prefix:
    for value in dict_prefix['TREVISO']:
        dict_prefix[value] = [value]
    dict_prefix.pop('TREVISO')

if 'TRENTO' in dict_prefix:
    for value in dict_prefix['TRENTO']:
        dict_prefix[value] = [value]
    dict_prefix.pop('TRENTO')
```

```

if 'INCR.' in dict_prefix:
    for value in dict_prefix['INCR.']:
        dict_prefix[value] = [value]
    dict_prefix.pop('INCR.')

if 'DE' in dict_prefix:
    for value in dict_prefix['DE']:
        dict_prefix[value] = [value]
    dict_prefix.pop('DE')

```

Treviso and Trento

```

In [ ]: # Remove the values Treviso, Trento, Trezzo and Treporti from the key 'Tre'
dict_prefix['TRE'] = [string for string in dict_prefix['TRE'] if 'TREVISO' not in string and 'TRENTO' not in string]

# Print the values of the dictionary with the key 'Tre'
print('The values of the dictionary with the key Tre are: {}'.format(dict_prefix['TRE']))

# TODO: Correct the values of the keys 'Treviso' and 'Trento' with the correct values

```

The values of the dictionary with the key Tre are: ['TRE ARCHI']

Keys with only an item

```

In [ ]: # If a key has only one value, then rename the key with the value
# Use copy() to avoid RuntimeError: dictionary changed size during iteration
for key, value in dict_prefix.copy().items():
    if len(value) == 1:
        dict_prefix[value[0]] = dict_prefix.pop(key)

```

Finally, the update dictionary is

```

In [ ]: # Print the dictionary in the new format
for key, value in dict_prefix.items():
    print('{}: {}'.format(key, value))

```

ZATTERE: ['ZATTERE "B"', 'ZATTERE "A"', 'ZATTERE']
VENEZIA: ['VENEZIA CORS', 'VENEZIA', 'VENEZIA PIAZ', 'VENEZIA RAMP']
FERROVIA: ['FERROVIA "B"', 'FERROVIA "D"', 'FERROVIA "E"', 'FERROVIA PIO', 'FERROVIA "A"', 'FERROVIA "C"']
LIBERTA': ['LIBERTA' SAN", 'LIBERTA' FIN"]
TRONCHETTO: ['TRONCHETTO F', 'TRONCHETTO "', 'TRONCHETTO M', 'TRONCHETTO T', 'TRONCHETTO V']
RIALTO: ['RIALTO "C"', 'RIALTO MERCA', 'RIALTO "D"', 'RIALTO "B"', 'RIALTO "A"']
GIARDINI: ['GIARDINI BIE', 'GIARDINI "B"', 'GIARDINI "A"']
ALTINIA: ['ALTINIA CA' "', 'ALTINIA MUNI', 'ALTINIA INDR', 'ALTINIA FAVA', 'ALTINIA SAN ', 'ALTINIA BERI', 'ALTINIA 181', 'ALTINIA FORT']
MESTRE: ['MESTRE RAMP', 'MESTRE CENTR', 'MESTRE VIA M', 'MESTRE VIA H', 'MESTRE VIA T', 'MESTRE STAZI', 'MESTRE V
IA P', 'MESTRE VIA A', 'MESTRE PIAZZ']
CHIOGGIA: ['CHIOGGIA', 'CHIOGGIA VIA', 'CHIOGGIA SAN', 'CHIOGGIA CAM', 'CHIOGGIA STA', 'CHIOGGIA BER', 'CHIOGGIA I
SO', 'CHIOGGIA RID', 'CHIOGGIA PAR', 'CHIOGGIA VAL', 'CHIOGGIA NEN', 'CHIOGGIA CA', 'CHIOGGIA OSP']
LIDO: ['LIDO S.M.E. ', 'LIDO S. NICO']
ARSENALE: ['ARSENALE "B"', 'ARSENALE "A"']
GALLO: ['GALLO BIBLIO', 'GALLO TORTA', 'GALLO MOROSI', 'GALLO GIOLIT', 'GALLO SANT'A", 'GALLO BRAGAD', 'GALLO LORE
DA', 'GALLO DEI KI', 'GALLO CONTAR', 'GALLO BARBAR', 'GALLO GALOPP', 'GALLO DEI GI']
ZITELLE: ['ZITELLE "B"', 'ZITELLE "A"']
PORTO: ['PORTO MARGHE', 'PORTOSECCO', 'PORTOSECCO C', 'PORTOSECCO L']
BORGO: ['BORGO SAN GI', 'BORGORICCO M', 'BORGORICCO', 'BORGORRICO S']
PADOVA: ['PADOVA SAN L', 'PADOVA FIERA', 'PADOVA TURAZ', 'PADOVA GOZZI', 'PADOVA EST', 'PADOVA OSPED']
BURANO: ['BURANO "C"', 'BURANO "A"', 'BURANO "B"']
MARGHERA: ['MARGHERA VIA', 'MARGHERA NAV', 'MARGHERA CIM', 'MARGHERA SAP', 'MARGHERA ROM', 'MARGHERA SAL']
MIRANESE: ['MIRANESE SAN', 'MIRANESE GIU', 'MIRANESE PER', 'MIRANESE PIE', 'MIRANESE SEL', 'MIRANESE IVA', 'MIRANE
SE VIV', 'MIRANESE CAL', 'MIRANESE PIA', 'MIRANESE LAZ', 'MIRANESE MON', 'MIRANESE AVA']
TRIESTE: ['TRIESTE MIRA', 'TRIESTE CATE', 'TRIESTE ERAC', 'TRIESTE ROBI', 'TRIESTE BOSC', 'TRIESTE MAZZ', 'TRIESTE
PARC', 'TRIESTE STAZ']
OLMO: ['OLMO CALVI', 'OLMO GIOVANN', 'OLMO MINZONI', 'OLMO PAPA LU', 'OLMO VITTORI', 'OLMO PELLICO']
ZELARINO: ['ZELARINO MUN', 'ZELARINO PAR', 'ZELARINO CHI', 'ZELARINO CAR']
CAMPALTO: ['CAMPALTO VIA', 'CAMPALTO PIA']
VALLI: ['VALLI PIAZZA', 'VALLI VIA PO', 'VALLI VALFIO', 'VALLI PONTE ']
PIAVE: ['PIAVE FIUME', 'PIAVE PODGOR', 'PIAVE PUCCIN']
PERTINI: ['PERTINI CHIE', 'PERTINI FOSC', 'PERTINI CARR', 'PERTINI GAVA', 'PERTINI VESP']
MURANO: ['MURANO NAVAG', 'MURANO COLON', 'MURANO VENIE', 'MURANO MUSEO', 'MURANO FARO ', 'MURANO SEREN', 'MURANO D
A MU']
TESSERA: ['TESSERA VIA ', 'TESSERA SCUO']
DESE: ['DESE CENTRO', 'DESE CICOGNE', 'DESE LITOMAR', 'DESE FS']
MONTE: ['MONTE CELO F', 'MONTE CERVIN', 'MONTE GRAPPA']
BISSUOLA: ['BISSUOLA ESI', 'BISSUOLA CAD', 'BISSUOLA PIS', 'BISSUOLA COL', 'BISSUOLA TEV', 'BISSUOLA VAR']
DOLO: ['DOLO CENTRO', 'DOLO MANZONI', 'DOLO CA' TRO", 'DOLO OSPEDAL', 'DOLO SCUOLE', 'DOLO MUNICIP', 'DOLO BANDIE

R', 'DOLO MONACHE', 'DOLO CIVICO ', 'DOLO DORIA', 'DOLO PRESICC', 'DOLO GUARDIA']
CASTELLANA: ['CASTELLANA P', 'CASTELLANA C', 'CASTELLANA S', 'CASTELLANA M', 'CASTELLANA B']
FIESSO: ['FIESSO GEMIT', 'FIESSO D'ART', 'FIESSO BARBA', 'FIESSO PIOVE', 'FIESSO PIOGH']
SALZANO: ['SALZANO MATT', 'SALZANO LORE', 'SALZANO DE G', 'SALZANO MUNI', 'SALZANO TOSC', 'SALZANO CHIE', 'SALZANO CONF', 'SALZANO ODDO', 'SALZANO PONT', 'SALZANO CIMI', 'SALZANO SOGA']
ASSEGGIANO: ['ASSEGGIANO M', 'ASSEGGIANO C', 'ASSEGGIANO E', 'ASSEGGIANO V', 'ASSEGGIANO P', 'ASSEGGIANO D']
ORLANDA: ['ORLANDA PINE', 'ORLANDA CASI', 'ORLANDA SAN ', 'ORLANDA PIOV', 'ORLANDA BAGA', 'ORLANDA DON ', 'ORLANDA CENT', 'ORLANDA 200']
TERRAGLIO: ['TERRAGLIO PE', 'TERRAGLIO NI', 'TERRAGLIO BO', 'TERRAGLIO TE', 'TERRAGLIO VI', 'TERRAGLIO GA', 'TERRAGLIO CA', 'TERRAGLIO FA']
TRIVIGNANO: ['TRIVIGNANO B', 'TRIVIGNANO P', 'TRIVIGNANO L', 'TRIVIGNANO G', 'TRIVIGNANO C']
GARIBALDI: ['GARIBALDI MU', 'GARIBALDI C', 'GARIBALDI OG', 'GARIBALDI GI']
MIRA: ['MIRANESE SAN', 'MIRANESE GIU', 'MIRANESE PER', 'MIRANESE PIE', 'MIRANESE SEL', 'MIRA PORTE', 'MIRANESE IV A', 'MIRANESE VIV', 'MIRANESE CAL', 'MIRA RISCOSS', 'MIRANESE PIA', 'MIRA EGEO', 'MIRANESE LAZ', 'MIRANESE MON', 'MIRANO FOSSA', 'MIRANO CENTR', 'MIRANESE AVA', 'MIRA ALIGHIE', 'MIRANO MATTE', 'MIRANO GRIMA', 'MIRANO SPORT', 'MIRANO SCUOL', 'MIRA BELLINI', 'MIRA CENTRO', 'MIRA MONTESS', 'MIRANO GRAMS', 'MIRANO BATTI', 'MIRA PIAZZA ', 'MIRANO PESTR', 'MIRANO TREV', 'MIRA ALBRIZZ', 'MIRANO DANTE', 'MIRANO CARDU', 'MIRA RIVIERA', 'MIRA CAMPI', 'MIRANO B OLLA', 'MIRANO CARAV', 'MIRA BERNINI', 'MIRANOTAGLIO', 'MIRANO LOCAL', 'MIRANO MARIU', 'MIRANO PERUG', 'MIRA PERTI NI', 'MIRA 25 APRI', 'MIRANO GALIL', 'MIRANO BOSCH', 'MIRA PONTE D', 'MIRANO OLMO', 'MIRA SAN MON', 'MIRA MUNICI P', 'MIRANO MINZ', 'MIRANO MODIG', 'MIRANO VAROT', 'MIRA MARE ME', 'MIRA GINESTR', 'MIRANO BARBA', 'MIRANO BOS C', 'MIRANO LUNEO']
SABBADINO: ['SABBADINO BA', 'SABBADINO L', 'SABBADINO PA', 'SABBADINO LA']
GOBBI: ['GOBBI CA' DO', 'GOBBI MANDAR', 'GOBBI DON BO', 'GOBBI SAN DO', 'GOBBI ORLAND', 'GOBBI MIRTIL', 'GOBBI VAL LEN']
DON: ['DON STURZO V', 'DON STURZO P']
CASONA: ['CASONA VALLE', 'CASONA BISSU', 'CASONA MARZI']
CARDUCCI: ['CARDUCCI FEL', 'CARDUCCI PAS']
CAPPUCCINA: ['CAPPUCCINA B', 'CAPPUCCINA V', 'CAPPUCCINA S']
MALAMOCCO: ['MALAMOCCO CE', 'MALAMOCCO BA', 'MALAMOCCO AL', 'MALAMOCCO ST', 'MALAMOCCO BE', 'MALAMOCCO OC', 'MALAMOCCO GA', 'MALAMOCCO PA']
CAMPONOGARA: ['CAMPONOGARA ', 'CAMPONOGARA']
PADANA: ['PADANA CIVIC', 'PADANA DELLE', 'PADANA STAZI', 'PADANA DEL L']
BECCARIA: ['BECCARIA COR', 'BECCARIA CHI', 'BECCARIA CAN', 'BECCARIA ORO', 'BECCARIA PAR']
PESEGGIA: ['PESEGGIA CA', 'PESEGGIA', 'PESEGGIA PER']
TEVERE: ['TEVERE BAGLI', 'TEVERE PARCO']
TRIESTINA: ['TRIESTINA PR', 'TRIESTINA MO', 'TRIESTINA TO', 'TRIESTINA AL', 'TRIESTINA LA', 'TRIESTINA UL', 'TRIESTINA ZO', 'TRIESTINA PI', 'TRIESTINA PA', 'TRIESTINA AE', 'TRIESTINA FO', 'TRIESTINA ZU', 'TRIESTINA SC']
MARTELLAGO: ['MARTELLAGO', 'MARTELLAGO V', 'MARTELLAGO S', 'MARTELLAGO D', 'MARTELLAGO P', 'MARTELLAGOTR', 'MARTELLAGO F', 'MARTELLAGO C', 'MARTELLAGO A', 'MARTELLAGO G']
GUGLIE: ['GUGLIE "A"', 'GUGLIE "B"']

SCORZE': ['SCORZE' ROMA", "SCORZE' MUNI", "SCORZE' CAPO", "SCORZE' MOGL", "SCORZE' FERM", "SCORZE' VENE", "SCORZE' ORTI", "SCORZE' BOSC"]
MARCON: ['MARCON GENOV', 'MARCON OBERD', 'MARCON CULT', 'MARCON DELLA', 'MARCON MUNIC', 'MARCON MARMO', 'MARCONI CORA', 'MARCONI MARC', 'MARCON LOMBA', 'MARCON MATTE', 'MARCON COOPE', 'MARCON STADI', 'MARCONI LORE', 'MARCON VIT TO', 'MARCON REPUB', 'MARCON CULTU', 'MARCON ALTIN', 'MARCON PEROS', 'MARCONI FIUM', 'MARCON MILAN', 'MARCONI DES ', 'MARCON ANCON']
MAERNE: ['MAERNE ISONZ', 'MAERNE CHIES', 'MAERNE MULIN', 'MAERNE CENTR', 'MAERNE FS', 'MAERNE TASSO', 'MAERNE CIVI C', 'MAERNE CIMIT', 'MAERNE GUARD', 'MAERNE CIRCO']
MADONNA: ['MADONNA DELL', 'MADONNA MARI']
CAPOLINEA: ['CAPOLINEA CA', 'CAPOLINEA FU']
RIVA: ['RIVA DE BIAS', 'RIVALE CHIES']
ACCADEMIA: ['ACCADEMIA "B', 'ACCADEMIA "A']
ROBEGANO: ['ROBEGANO CEN', 'ROBEGANO PUC', 'ROBEGANO MON', 'ROBEGANO 25 ']
OSPEDALE: ['OSPEDALE DEL', 'OSPEDALE MIR', 'OSPEDALE']
TITO: ['TITO CASTELL', 'TITO SELVANE']
NOALE: ['NOALE', 'NOALE OSPEDA', 'NOALE DEI NO', 'NOALE BACCHI', 'NOALE ONGARI', 'NOALE MORO', 'NOALE ZONA I', 'NO ALE LIVENZ', 'NOALE LANCER']
VILLABONA: ['VILLABONA PI', 'VILLABONA BO', 'VILLABONA MO', 'VILLABONA 87', 'VILLABONA 8', 'VILLABONA ON']
PERON: ['PERON BASEGG', 'PERON SARAGA']
ROMEA: ['ROMEA CIMITE', 'ROMEA CIVICO', 'ROMEA FOSSET', 'ROMEA MARINE', 'ROMEI CIVIC', 'ROMEA DEL BO', 'ROMEA FIS OLA', 'ROMEA PRIMAV']
ORIAGO: ['ORIAGO STAZI', 'ORIAGO CENTR', 'ORIAGO SOMMO', 'ORIAGO ROMAG', 'ORIAGO VENEZ', 'ORIAGO FERRO']
ALBERONI: ['ALBERONI DEL', 'ALBERONI FAR', 'ALBERONI OTT', 'ALBERONI SAN', 'ALBERONI STE', 'ALBERONI GOL', "ALBERO NI CA'", 'ALBERONI COL']
GAMBARARE: ['GAMBARARE VI', 'GAMBARARE CI', 'GAMBARARE PO', 'GAMBARARE']
MOGLIANO: ['MOGLIANO RON', 'MOGLIANO CEN', 'MOGLIANO RIM', 'MOGLIANO MAR', 'MOGLIANO RAG', 'MOGLIANO TOM', 'MOGLIA NO MUN', 'MOGLIANO CAM', 'MOGLIANO LIC', 'MOGLIANO FS', 'MOGLIANO CIM', 'MOGLIANO TER', 'MOGLIANO DEI', 'MOGLIANO BUR', 'MOGLIANO BEL', 'MOGLIANO ZER', 'MOGLIANO COR', 'MOGLIANO MEU', 'MOGLIANO GHE']
BOTTENIGO: ['BOTTENIGO CA', 'BOTTENIGO PI', 'BOTTENIGO BO', 'BOTTENIGO MA']
SPINEA: ['SPINEA PIAZZ', 'SPINEA MARTI', 'SPINEA ORGNA', 'SPINEA POZZU', 'SPINEA ALFIE', 'SPINEA GIORG', 'SPINEA R EPUB', 'SPINEA SANRE', 'SPINEA ROSSI', 'SPINEA CENTR', 'SPINEA DESEN', 'SPINEA VILLA', 'SPINEA SAN R', 'SPINEA LUN EO']
PREGANZIOL: ['PREGANZIOL M', 'PREGANZIOL ', 'PREGANZIOL F', 'PREGANZIOL G', 'PREGANZIOL B']
BORBIAGO: ['BORBIAGO MIL', 'BORBIAGO CEN']
RINASCITA: ['RINASCITA EM', 'RINASCITA BE']
MATTUGLIE: ['MATTUGLIE PE', 'MATTUGLIE DI']
BRENDOLE: ['BRENDOLE ARS', 'BRENDOLE STI', 'BRENDOLE FAV', 'BRENDOLE PRO', 'BRENDOLE']
VIGONZA: ['VIGONZA PERA', 'VIGONZA BACH']
CALUCCI: ['CALUCCI CIME', 'CALUCCI QUAR']
RIELTA: ['RIELTA PARCO', "RIELTA CA' R"]

MIRANO: ['MIRANO FOSSA', 'MIRANO CENTR', 'MIRANO MATTE', 'MIRANO GRIMA', 'MIRANO SPORT', 'MIRANO SCUOL', 'MIRANO G
RAMS', 'MIRANO BATTI', 'MIRANO PESTR', 'MIRANO TREVI', 'MIRANO DANTE', 'MIRANO CARDU', 'MIRANO BOLLA', 'MIRANO CAR
AV', 'MIRANOTAGLIO', 'MIRANO LOCAL', 'MIRANO MARIU', 'MIRANO PERUG', 'MIRANO GALIL', 'MIRANO BOSCH', 'MIRANO OLM
O', 'MIRANO MINZ', 'MIRANO MODIG', 'MIRANO VAROT', 'MIRANO BARBA', 'MIRANO BOSC', 'MIRANO LUNEO']
ISOLA: ['ISOLA UNIONE', 'ISOLA VERDE', 'ISOLA VERDE ']
TORINO: ['TORINO ROSSE', 'TORINO', 'TORINO UNIVE']
ANCONA: ['ANCONA CARBO', 'ANCONA TORIN']
STRA: ['STRA DANTE', 'STRA PIAZZA ', 'STRA LOREDAN', 'STRA FOSSOLO']
CAVARZERE: ['CAVARZERE VI', 'CAVARZERE AU', 'CAVARZERE C']
MORANZANI: ['MORANZANI 32', 'MORANZANI EL', 'MORANZANI CO']
LUGHETTO: ['LUGHETTO STR', 'LUGHETTO MAR', 'LUGHETTO DI ']
CAVERGNAGO: ['CAVERGNAGO T', 'CAVERGNAGO M']
TREZZO: ['TREZZO BATTU', 'TREZZO TERRA']
GATTA: ['GATTA SCARAN', 'GATTA PISACA', 'GATTA SANTA ', 'GATTA 90', 'GATTA SCARAM', 'GATTA PROTAG', 'GATTA IMMAC
O', 'GATTA VERCI']
TOSATTO: ['TOSATTO IMPA', 'TOSATTO PACC']
PASINI: ['PASINI FRATE', 'PASINI LAVOR']
LUGO: ['LUGO STRADA ', 'LUGO ZONA IN']
SCUOLA: ['SCUOLA MEDIA', 'SCUOLA ZENDR']
D'ANNUNZIO: ["D'ANNUNZIO P", "D'ANNUNZIO S", "D'ANNUNZIO D"]
CERTOSA: ['CERTOSA', 'CERTOSA A RI']
MARCONI: ['MARCONI CORA', 'MARCONI MARC', 'MARCONI LORE', 'MARCONI FIUM', 'MARCONI DES ']
CAVANELLA: ['CAVANELLA PI', 'CAVANELLA', 'CAVANELLA BO']
CALCROCI: ['CALCROCI', 'CALCROCI CAV', 'CALCROCI CHI']
VISINONI: ['VISINONI COM', 'VISINONI POL', 'VISINONI SEL']
PORTOSECCE: ['PORTOSECCE', 'PORTOSECCE C', 'PORTOSECCE L']
GAGGIO: ['GAGGIO CIMIT', 'GAGGIO', 'GAGGIO FERMI', 'GAGGIO VIVAL']
BRONDOLO: ['BRONDOLO', 'BRONDOLO DOL']
VESPUCCI: ['VESPUCCI GAR', 'VESPUCCI SAN', "VESPUCCI CA'", 'VESPUCCI GRI', 'VESPUCCI PIG', 'VESPUCCI CAT', 'VESPUCC
CI BOE']
FOSSO: ["FOSSO' CENTR", "FOSSO' ARZAR", "FOSSO' RONCA", "FOSSO' FAVAL", "FOSSO' PROVI", "FOSSO' ZONA ", "FOSSO' B
OSEL", "FOSSO'"]
BOJON: ['BOJON', 'BOJON FS', 'BOJON RIVELL']
PROZZOLO: ['PROZZOLO', 'PROZZOLOTOGL']
SAMBRUSON: ['SAMBRUSON', 'SAMBRUSON MA']
VALLON: ['VALLON FORTE', 'VALLON VALDE', 'VALLON BORGO', 'VALLON DE NI']
SANT'ANNA: ['SANT'ANNA S', 'SANT'ANNA V']
CASALE: ['CASALE SUL S', 'CASALE BIVIO']
ADRIA: ['ADRIA SCUOLE', 'ADRIA MARCON', 'ADRIA FS', 'ADRIA OSPEDA']
CALTANA: ['CALTANA MARI', 'CALTANA', 'CALTANA PIOG']

VIGONOVO: ['VIGONOVO GAL', 'VIGONOVO', 'VIGONOVO 1 M', 'VIGONOVO DE ', 'VIGONOVO ALF']
 CA: ["CA' D'ORO", 'CAMPALTO VIA', 'CAMPAGNA LUP', 'CASTELLANA P', 'CASTELLANA C', "CA' ROSSA VO", 'CASONA VALLE',
 'CARDUCCI FEL', 'CASONA BISSU', 'CAPPUCCINA B', 'CAMPONOGARA ', 'CARDUCCI PAS', "CA' ROSSA SE", "CA' ROSSA BI", 'C
 APOLINEA CA', 'CASTELLANA S', 'CALABRIA CAM', 'CAPPUCCINA V', "CA' REZZONIC", 'CAPPUCCINA S', "CA' SABBIONI", "CA'
 ROSSA OB", "CA' SOLARO C", 'CANAL LEONE', 'CASONA MARZI', 'CAPPELLA', 'CALUCCI CIME', 'CASTELLANA M', 'CAMPALTO PI
 A', 'CAVARZERE VI', 'CAZZAGHETTO ', 'CAVERGNAGO T', "CA' MARCELLO", 'CAVERGNAGO M', 'CAFASSO BOTT', 'CAVALCAVIA
 V', 'CASTELLANA B', 'CAVANIS CAPO', 'CAPRICCIO CE', 'CALVI PARMES', 'CAVANELLA PI', 'CALCROCI', "CA' BRENTTELL", 'C
 AROMAN', 'CALCROCI CAV', 'CAMPONOGARA', "CA' LIN ERAC", 'CAMPORESE GR', "CA' LIN CAST", 'CASALE SUL S', 'CAVARZERE
 AU', "CA' BIANCA L", 'CALTANA MARI', 'CA SOLARO PA', 'CALTANA', 'CALUCCI QUAR', 'CAPOLINEA FU', "CA' SOLARO P", "C
 A' LIN PITA", "CA' LIN GATT", "CA' BIANCA C", 'CALCROCI CHI', 'CAMPOCROCE', "CA' SOLARO O", 'CAVANELLA', "CA' BIAN
 CA P", 'CAZZAGO PASC', 'CA' BIANCA ', 'CAZZAGO', 'CAMPOVERARDO', 'CAVARZERE C', 'CASALE BIVIO', 'CALTANA PIOG',
 'CAVANELLA BO', 'CAMPOLONGO L', 'CAMPOCROCE C', 'CAMPOLONGO 8']
 LE: ['LE GRAZIE', 'LE CRETE']
 ZERO: ['ZERO BRANCO', 'ZERO BRANCO ']
 SANBRUSON: ['SANBRUSON MA', 'SANBRUSON CA']
 CAMPOCROCE: ['CAMPOCROCE', 'CAMPOCROCE C']
 SCALTENIGO: ['SCALTENIGO F', 'SCALTENIGO P', 'SCALTENIGO']
 BORGORICCO: ['BORGORICCO M', 'BORGORICCO']
 CAZZAGO: ['CAZZAGO PASC', 'CAZZAGO']
 PIANIGA: ['PIANIGA ROMA', 'PIANIGA BOSC']
 VETERNIGO: ['VETERNIGO', 'VETERNIGO PI']
 CAMPOLONGO: ['CAMPOLONGO L', 'CAMPOLONGO 8']
 VETREGO: ["VETREGO PRA'", 'VETREGO']
 VILLA: ['VILLABONA PI', 'VILLABONA BO', 'VILLABONA MO', 'VILLABONA 87', 'VILLABONA 8', 'VILLABONA ON', 'VILLA DEL
 BO']
 P.LE ROMA: ['P.LE ROMA "G', 'P.LE ROMA "E', 'P.LE ROMA "A', 'P.LE ROMA "C', 'P.LE ROMA "D', 'P.LE ROMA "B', 'P.LE
 ROMA "F']
 F.TE NOVE: ['F.TE NOVE "C', 'F.TE NOVE "A', 'F.TE NOVE "D', 'F.TE NOVE "B']
 SAN MARCO: ['SAN MARCO CA', 'SAN MARCO SA', 'SAN MARCO-SA', 'SAN MARCO MO', 'SAN MARCO BO', 'SAN MARCO MA', 'SAN M
 ARCO FO', 'SAN MARCO MU', 'S. MARCO (GI)']
 SAN DONA: ["SAN DONA' MA", "SAN DONA' CE", "SAN DONA' RI", "SAN DONA' PI", "SAN DONA' PA", "SAN DONA' VA", "SAN DO
 NA' FA", "SAN DONA' PE"]
 SAN PIETRO: ['SAN PIETRO B', 'SANPIETRO CA', 'SAN PIETRO P', 'SAN PIETRO C', 'S. PIETRO DI']
 CA' ROSSA: ["CA' ROSSA VO", "CA' ROSSA SE", "CA' ROSSA BI", "CA' ROSSA OB"]
 AEROPORTO MA: ['AEROPORTO MA']
 SOTTOMARINA : ['SOTTOMARINA ']
 SANT'ANTONIO: ["SANT'ANTONIO"]
 SACCA FISOLA: ['SACCA FISOLA']
 GIUDECCA PAL: ['GIUDECCA PAL']
 OLIVI: ['OLIVI']

MARTIRI DELL: ['MARTIRI DELL']
S.ELENA-STAD: ['S.ELENA-STAD']
REDENTORE: ['REDENTORE']
PUNTA SABBIO: ['PUNTA SABBIO']
TRE ARCHI: ['TRE ARCHI']
LOVA STRADA : ['LOVA STRADA ']
BELFREDO TER: ['BELFREDO TER']
TREPORTI: ['TREPORTI']
CAMPAGNA LUP: ['CAMPAGNA LUP']
ROSARA STRAD: ['ROSARA STRAD']
PASQUALIGO M: ['PASQUALIGO M']
PARK PETROLI: ['PARK PETROLI']
VALLENARI ST: ['VALLENARI ST']
CORSO DEL PO: ['CORSO DEL PO']
PALEOCAPA PA: ['PALEOCAPA PA']
PIOVE DI SAC: ['PIOVE DI SAC']
MALCONTENTA : ['MALCONTENTA ']
27 OTTOBRE D: ['27 OTTOBRE D']
PASSO CAMPAL: ['PASSO CAMPAL']
FORTE MARGHE: ['FORTE MARGHE']
CIRCONVALLAZ: ['CIRCONVALLAZ']
BACINI - ARS: ['BACINI - ARS']
PELESTRINA : ['PELESTRINA ']
CREA: ['CREA']
BANDIERA GHE: ['BANDIERA GHE']
CELESTIA: ['CELESTIA']
ILARIA ALPI : ['ILARIA ALPI ']
MAZZORBO: ['MAZZORBO']
SALAMONIO MA: ['SALAMONIO MA']
CALABRIA CAM: ['CALABRIA CAM']
PEOPLE MOVER: ['PEOPLE MOVER']
SALUTE: ['SALUTE']
RIVALE CHIES: ['RIVALE CHIES']
CONCHE STRAD: ['CONCHE STRAD']
PAOLUCCI LON: ['PAOLUCCI LON']
CANAL LEONE: ['CANAL LEONE']
RISORGIMENTO: ['RISORGIMENTO']
FARO ROCCHET: ['FARO ROCCHET']
FAVRETTI MES: ['FAVRETTI MES']
PALIAGA CA' : ["PALIAGA CA' "]

CAPPELLA: ['CAPPELLA']
CORRENTI CAP: ['CORRENTI CAP']
GIOVANNACCI : ['GIOVANNACCI ']
LAVELLI PAOL: ['LAVELLI PAOL']
EINAUDI CAST: ['EINAUDI CAST']
VIGNOLE: ['VIGNOLE']
ZENDRINI VIL: ['ZENDRINI VIL']
GIARE STRADA: ['GIARE STRADA']
FITTIZIA: ['FITTIZIA']
GOZZI CAPPUC: ['GOZZI CAPPUC']
SAMBUGHE': ["SAMBUGHE'"]
CAZZAGHETTO : ['CAZZAGHETTO ']
CIRCONVALAZI: ['CIRCONVALAZI']
AGENZIA ENT: ['AGENZIA ENT']
TORCELLO: ['TORCELLO']
CIMITERO SAN: ['CIMITERO SAN']
SALICI VILLA: ['SALICI VILLA']
GAZZERA ALTA: ['GAZZERA ALTA']
QUARNARO CAL: ['QUARNARO CAL']
CAFASSO BOTT: ['CAFASSO BOTT']
CAVALCAVIA V: ['CAVALCAVIA V']
CAVANIS CAPO: ['CAVANIS CAPO']
BUSA DI VIGO: ['BUSA DI VIGO']
CAPRICCIO CE: ['CAPRICCIO CE']
CALVI PARMES: ['CALVI PARMES']
MARSALA CENT: ['MARSALA CENT']
SPIRITO SANT: ['SPIRITO SANT']
PALAZZO DEL : ['PALAZZO DEL ']
LUNGOMARE AD: ['LUNGOMARE AD']
GALILEI DARS: ['GALILEI DARS']
COLOMBO: ['COLOMBO']
MAZZOCCO: ['MAZZOCCO']
NAZIONI UNIT: ['NAZIONI UNIT']
AREOPORTO MA: ['AREOPORTO MA']
GALTA ARGINE: ['GALTA ARGINE']
LIETTOLI TRE: ['LIETTOLI TRE']
FRESCADA: ['FRESCADA']
SELVANESE PL: ['SELVANESE PL']
GRAN VIALE: ['GRAN VIALE']
CAROMAN: ['CAROMAN']

PETTORAZZA S: ['PETTORAZZA S']
PAGANELLO TI: ['PAGANELLO TI']
FISICA DEPOS: ['FISICA DEPOS']
COLMELLO: ['COLMELLO']
PERAROLO QUA: ['PERAROLO QUA']
CAMPORESE GR: ['CAMPORESE GR']
RONZINELLA F: ['RONZINELLA F']
AZOTO SOTTAN: ['AZOTO SOTTAN']
PONTE DI BRE: ['PONTE DI BRE']
FUSINA CENTR: ['FUSINA CENTR']
DOSSON: ['DOSSON']
MARTELLAGOTR: ['MARTELLAGOTR']
GARDIGIANO: ['GARDIGIANO']
OSPIZIO MARI: ['OSPIZIO MARI']
MONIEGO TREV: ['MONIEGO TREV']
FAVIGNANA: ['FAVIGNANA']
MORANDI NICE: ['MORANDI NICE']
TERMINAL RO-: ['TERMINAL RO-']
BERNINI DI V: ['BERNINI DI V']
CORTIVO TOMB: ['CORTIVO TOMB']
RIO SAN MART: ['RIO SAN MART']
ZIANIGO: ['ZIANIGO']
FORNASE PERU: ['FORNASE PERU']
MELLAREDO CA: ['MELLAREDO CA']
SFMR SPINEA: ['SFMR SPINEA']
RUSTEGHELLO: ['RUSTEGHELLO']
QUARTO D'ALT: ["QUARTO D'ALT"]
MIRANOTAGLIO: ['MIRANOTAGLIO']
MAIANO ALBER: ['MAIANO ALBER']
ROMEAI CIVIC: ['ROMEAI CIVIC']
ALTA MILANO: ['ALTA MILANO']
SOPPRESSA - : ['SOPPRESSA - ']
P.ZZA DELLA : ['P.ZZA DELLA ']
STIGLIANO: ['STIGLIANO']
MARANO: ['MARANO']
CA' BIANCA : ['CA' BIANCA ']
CORTE: ['CORTE']
ULSS: ['ULSS']
CAMPOVERARDO: ['CAMPOVERARDO']
MATTEOTTI CE: ['MATTEOTTI CE']

BORGORRICO S: ['BORGORRICO S']
MESTRINA SPA: ['MESTRINA SPA']
ROTTANOVA CO: ['ROTTANOVA CO']
TORRE ENAC: ['TORRE ENAC']
BORROMINI BE: ['BORROMINI BE']
PROZZOLOTOGL: ['PROZZOLOTOGL']
BADOERE: ['BADOERE']
FERM.SERV. D: ['FERM.SERV. D']
SANDON: ['SANDON']
BOSCHETTA: ['BOSCHETTA']
SCANDOLARA: ['SCANDOLARA']
CHIMICA INGR: ['CHIMICA INGR']
GRAMSCI TOGL: ['GRAMSCI TOGL']
PALUELLO VEN: ['PALUELLO VEN']
LAZZARETTO N: ['LAZZARETTO N']
S. ERASMO PU: ['S. ERASMO PU']
S. TOMA 'B': ['S. TOMA\ 'B']
S. MARCO-SAN: ['S. MARCO-SAN']
S. MARCUOLA-: ['S. MARCUOLA-']
S. STAE: ['S. STAE']
S. MARCO VAL: ['S. MARCO VAL']
S. MARIA DEL: ['S. MARIA DEL']
S. SILVESTRO: ['S. SILVESTRO']
S. ANGELO: ['S. ANGELO']
S. ALVISE: ['S. ALVISE']
S. GIORGIO: ['S. GIORGIO']
S. BASILIO: ['S. BASILIO']
S. MARTA: ['S. MARTA']
S. SERVULO: ['S. SERVULO']
S. ERASMO CA: ['S. ERASMO CA']
S. ERASMO CH: ['S. ERASMO CH']
S. SAMUELE: ['S. SAMUELE']
S. LAZZARO: ['S. LAZZARO']
S. TOMA 'A': ['S. TOMA\ 'A']
S. MARCO GIA: ['S. MARCO GIA']
SAN LIBERALE: ['SAN LIBERALE']
SANTA MARGHE: ['SANTA MARGHE']
SANTA MARIA : ['SANTA MARIA '']
SAN NICOLO' : ["SAN NICOLO' "]
SAN NICOLO': ["SAN NICOLO'"]

SAN ROCCO BR: ['SAN ROCCO BR']
SANSOVINO VE: ['SANSOVINO VE']
SANT'ANNA S: ['SANT'ANNA S']
SANT'ANNA V: ['SANT'ANNA V']
SAN TROVASO: ['SAN TROVASO']
SANBRUSON MA: ['SANBRUSON MA']
SANT'ANGELO : ["SANT'ANGELO "]
SANBRUSON CA: ['SANBRUSON CA']
SAN GIUSEPP: ['SAN GIUSEPP']
SANT ANGELO : ['SANT ANGELO ']
SANT' ALBERT: ["SANT' ALBERT"]
CA' D'ORO: ["CA' D'ORO"]
CA' REZZONIC: ["CA' REZZONIC"]
CA' SABBIONI: ["CA' SABBIONI"]
CA' SOLARO C: ["CA' SOLARO C"]
CA' MARCELLO: ["CA' MARCELLO"]
CA' BRENTTELL: ["CA' BRENTTELL"]
CA' LIN ERAC: ["CA' LIN ERAC"]
CA' LIN CAST: ["CA' LIN CAST"]
CA' BIANCA L: ["CA' BIANCA L"]
CA' SOLARO P: ["CA' SOLARO P"]
CA' LIN PITA: ["CA' LIN PITA"]
CA' LIN GATT: ["CA' LIN GATT"]
CA' BIANCA C: ["CA' BIANCA C"]
CA' SOLARO O: ["CA' SOLARO O"]
CA' BIANCA P: ["CA' BIANCA P"]
PIAZZALE GIO: ['PIAZZALE GIO']
PIAZZA MERCA: ['PIAZZA MERCA']
PIAZZALE RAV: ['PIAZZALE RAV']
VIA DEI CANT: ['VIA DEI CANT']
VIA VILLABON: ['VIA VILLABON']
STAZIONE MES: ['STAZIONE MES']
STAZIONE PAD: ['STAZIONE PAD']
STAZIONE MAR: ['STAZIONE MAR']
STAZIONE DI : ['STAZIONE DI ']
TREVISO SELV: ['TREVISO SELV']
TREVISO FS: ['TREVISO FS']
TREVISO PIND: ['TREVISO PIND']
TREVISO SAN : ['TREVISO SAN ']
TREVISO GHIR: ['TREVISO GHIR']

```
TREVIS0 LORE: ['TREVIS0 LORE']
TRENTO FAGAR: ['TRENTO FAGAR']
TRENTO GAZZE: ['TRENTO GAZZE']
TRENTO PODGO: ['TRENTO PODGO']
DESE CENTRO: ['DESE CENTRO']
DESE CICOGNE: ['DESE CICOGNE']
DESE LITOMAR: ['DESE LITOMAR']
DE NICOLA CH: ['DE NICOLA CH']
DEI MURAZZI : ['DEI MURAZZI ']
DESE FS: ['DESE FS']
DELLE MESSI : ['DELLE MESSI ']
```

```
In [ ]: # Export the dictionary in a json file
name_file = 'dict_prefix_' + file_name.split('.')[0] + '.json'
with open('data/dictionaries/' + name_file, 'w') as fp:
    json.dump(dict_prefix, fp)
```

Useless stamps

```
In [ ]: # TODO: #1 Remove useless rows that have a minimum temporal gap for the same serial and fermata
# DE-COMMENT THE FOLLOWING LINES OF CODE
```

```
In [ ]: # Find the serial with the highest number of validations, and the same for each ticket profile, save the results i
dict_serial = {}
for ticket in df['TICKET_CODE'].unique():
    dict_serial[ticket] = df[df['TICKET_CODE'] == ticket]['SERIALE'].value_counts().index[0]

# Print the serial with the highest number of validations, and the same for each ticket profile
for ticket in df['TICKET_CODE'].unique():
    print('The serial with the highest number of validations for the ticket profile {} is: {}'.format(ticket, dict_s
```

The serial with the highest number of validations for the ticket profile 7 is: -2821794789
The serial with the highest number of validations for the ticket profile 5-STUD is: -2854964619
The serial with the highest number of validations for the ticket profile 6-STUD is: -3604953805
The serial with the highest number of validations for the ticket profile 5 is: -2821768344
The serial with the highest number of validations for the ticket profile 6 is: -2855011795
The serial with the highest number of validations for the ticket profile 1 is: 65694522718453509
The serial with the highest number of validations for the ticket profile 2 is: 40551647001685764
The serial with the highest number of validations for the ticket profile 3 is: 41960768052577796
The serial with the highest number of validations for the ticket profile 4 is: 41679293073258756
The serial with the highest number of validations for the ticket profile 6-RET is: -4089257922
The serial with the highest number of validations for the ticket profile 5-WKRS is: -2824296001
The serial with the highest number of validations for the ticket profile 6-WKRS is: -3613105285

```
In [ ]: # Group by the serial and the fermata
# df = df.groupby(['SERIALE', 'FERMATA']).apply(lambda x: x.sort_values(by='DATA_VALIDAZIONE', ascending=True))

# Print the first 5 rows of the df
# df.head()
# DO NOT DE-COMMENT THIS CELL
```

```
In [ ]: # Reset the index of the df and drop the old index in order to have a new index starting from 0 to the number of rows
# It is necessary to have a new index because the groupby function has created a multi-index
df.reset_index(drop=True, inplace=True)
```

```
In [ ]: # Create a new column 'MIN_TEMPORAL_GAP' that contains the minimum temporal gap between two validations for the same serial
df = df.groupby(['SERIALE', 'DATA', 'DESCRIZIONE']).apply(lambda x: x.assign(MIN_TEMPORAL_GAP = x['DATA_VALIDAZIONE']
```

```
In [ ]: df.head(20)
```

Out[]:

	DATA	ORA	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	TICKET_CODE	DESCRIZIONE_TITOLO
0	2023-01-13	00:00:00	2023-01-13 00:00:00	40834866809772548	162	STAZIONE MES	12101	7	BIGL.AUT.75'MESTRE/LIDO-TSC
1	2023-01-13	00:00:00	2023-01-13 00:00:00	-3604990320	5049	ZATTERE "B"	23301	5-STUD	MENS.STUDENTE RETE UNICA
2	2023-01-13	00:00:00	2023-01-13 00:00:00	-2824230951	5043	S. TOMA' "B"	23303	6-STUD	ABB STUD. RETEUNICA 12 MESI
3	2023-01-13	00:00:00	2023-01-13 00:00:00	40552750134805252	5013	S. MARCO-SAN	11101	7	75'-TPL 8,64-COMVE0,86
4	2023-01-13	00:01:00	2023-01-13 00:01:00	-3604964420	6084	VENEZIA CORS	11209	7	BIGL RETE UNICA 75'
5	2023-01-13	00:01:00	2023-01-13 00:01:00	-2855032233	6084	VENEZIA CORS	11209	7	BIGL RETE UNICA 75'
6	2023-01-13	00:01:00	2023-01-13 00:01:00	-3604965107	5031	P.LE ROMA "G"	23301	5-STUD	MENS.STUDENTE RETE UNICA
7	2023-01-13	00:01:00	2023-01-13 00:01:00	-3613059169	5032	FERROVIA "B"	23101	5	MENSILE ORDINARIO RETE UNICA
8	2023-01-13	00:01:00	2023-01-13 00:01:00	-3604913072	6057	SAN MARCO CA	11209	7	BIGL RETE UNICA 75'
9	2023-01-13	00:01:00	2023-01-13 00:01:00	65694113023229189	1392	LIBERTA' SAN	12101	7	BIGL.AUT.75'MESTRE/LIDO-TSC
10	2023-01-13	00:01:00	2023-01-13 00:01:00	65694113021064965	1392	LIBERTA' SAN	12101	7	BIGL.AUT.75'MESTRE/LIDO-TSC
11	2023-01-13	00:01:00	2023-01-13 00:01:00	-3604966090	5031	P.LE ROMA "G"	11209	7	BIGL RETE UNICA 75'
12	2023-01-13	00:01:00	2023-01-13 00:01:00	-2821741139	5026	TRONCHETTO F	11209	7	BIGL RETE UNICA 75'
13	2023-01-13	00:02:00	2023-01-13 00:02:00	-2821834621	6061	SAN MARCO SA	11209	7	BIGL RETE UNICA 75'
14	2023-01-13	00:02:00	2023-01-13 00:02:00	-3613076684	5013	S. MARCO-SAN	23101	5	MENSILE ORDINARIO RETE UNICA
15	2023-	00:02:00	2023-01-13	-2821793317	5013	S. MARCO-	11209	7	BIGL RETE UNICA 75'

	DATA	ORA	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	TICKET_CODE	DESCRIZIONE_TITOLO
	01-13		00:02:00			SAN			
16	2023-01-13	00:02:00	2023-01-13 00:02:00	-2854788116	5013	S. MARCO-SAN	23101	5	MENSILE ORDINARIO RETE UNICA
17	2023-01-13	00:02:00	2023-01-13 00:02:00	-3613106338	5039	RIALTO "C"	23101	5	MENSILE ORDINARIO RETE UNICA
18	2023-01-13	00:04:00	2023-01-13 00:04:00	40552750539688708	506	VENEZIA	12101	7	BIGL.AUT.75'MESTRE/LIDO-TSC
19	2023-01-13	00:04:00	2023-01-13 00:04:00	40552750539692036	506	VENEZIA	12101	7	BIGL.AUT.75'MESTRE/LIDO-TSC

In []: `df.tail(20)`

Out[]:

	DATA	ORA	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	TICKET_CODE	DESCRIZIONE_TI
5041356	2023-03-14	23:57:00	2023-03-14 23:57:00	40834866531627780	5138	RIALTO MERCA	11107	2	48H-TPL 2 COMV
5041357	2023-03-14	23:57:00	2023-03-14 23:57:00	-3604965688	5043	S. TOMA' "B"	23102	6	ANNUALE ORDIN RETE U
5041358	2023-03-14	23:57:00	2023-03-14 23:57:00	40553391555332612	5031	P.LE ROMA "G	11261	1	DAILYP-TPL19,90-C.V
5041359	2023-03-14	23:57:00	2023-03-14 23:57:00	-3604943928	5031	P.LE ROMA "G	11209	7	BIGL RETE UNIC
5041360	2023-03-14	23:57:00	2023-03-14 23:57:00	40553391555332868	5031	P.LE ROMA "G	11261	1	DAILYP-TPL19,90-C.V
5041361	2023-03-14	23:57:00	2023-03-14 23:57:00	-2855030333	5031	P.LE ROMA "G	23101	5	MENSILE ORDINARIO L
5041362	2023-03-14	23:57:00	2023-03-14 23:57:00	65694182440336389	509	VENEZIA	12101	7	BIGL.AUT.75'MESTRE/
5041363	2023-03-14	23:57:00	2023-03-14 23:57:00	-2854816188	5024	TRONCHETTO "	23101	5	MENSILE ORDINARIO L
5041364	2023-03-14	23:58:00	2023-03-14 23:58:00	40834866535624196	5045	CA' REZZONIC	11107	2	48H-TPL 2 COMV
5041365	2023-03-14	23:58:00	2023-03-14 23:58:00	-2818096743	5024	TRONCHETTO "	11209	7	BIGL RETE UNIC
5041366	2023-03-14	23:58:00	2023-03-14 23:58:00	-2818101150	5024	TRONCHETTO "	23101	5	MENSILE ORDINARIO L
5041367	2023-03-14	23:58:00	2023-03-14 23:58:00	40834866535623940	5045	CA' REZZONIC	11107	2	48H-TPL 2 COMV
5041368	2023-03-14	23:58:00	2023-03-14 23:58:00	40835096720245252	5031	P.LE ROMA "G	11261	1	DAILYP-TPL19,90-C.V
5041369	2023-03-14	23:58:00	2023-03-14 23:58:00	40553621743596292	5031	P.LE ROMA "G	11261	1	DAILYP-TPL19,90-C.V
5041370	2023-03-14	23:58:00	2023-03-14 23:58:00	-2824223843	5003	LIDO S.M.E.	23102	6	ANNUALE ORDIN RETE U
5041371	2023-	23:58:00	2023-03-14	-2864643315	162	STAZIONE	11209	7	BIGL RETE UNIC

	DATA	ORA	DATA_VALIDAZIONE	SERIALE	FERMATA	DESCRIZIONE	TITOLO	TICKET_CODE	DESCRIZIONE_TI
	03-14		23:58:00			MES			
5041372	2023-03-14	23:58:00	2023-03-14 23:58:00	-2854956628	5026	TRONCHETTO F	11209	7	BIGL RETE UNIC
5041373	2023-03-14	23:59:00	2023-03-14 23:59:00	-2850025054	384	MESTRE CENTR	23101	5	MENSILE ORDINARIO L
5041374	2023-03-14	23:59:00	2023-03-14 23:59:00	-2824225710	5024	TRONCHETTO "	23101	5	MENSILE ORDINARIO L
5041375	2023-03-14	23:59:00	2023-03-14 23:59:00	-3604916033	5039	RIALTO "C"	23101	5	MENSILE ORDINARIO L

```
In [ ]: df['MIN_TEMPORAL_GAP'].value_counts()
```

```
Out[ ]: 0.0      68001
        1.0      25914
        2.0      17556
        3.0      10004
        4.0       6943
        5.0       5300
        6.0       3982
        7.0       3140
        8.0       2553
        9.0       1992
       10.0       1813
       11.0       1405
       12.0       1146
       13.0       1026
       14.0        862
       15.0        742
       16.0        703
       17.0        651
       18.0        568
       19.0        474
       20.0        467
       21.0        453
       22.0        420
      419.0        397
       23.0        392
      420.0        389
       25.0        379
       24.0        369
      360.0        359
      421.0        337
       27.0        333
      359.0        330
      409.0        330
       26.0        329
      361.0        322
      418.0        321
      480.0        313
      417.0        312
      300.0        311
      326.0        309
```

28.0	308
301.0	307
482.0	306
408.0	304
219.0	303
382.0	303
358.0	300
30.0	297
339.0	296
363.0	293
303.0	293
362.0	292
299.0	292
479.0	291
387.0	288
277.0	288
341.0	287
396.0	287
342.0	286
311.0	285
381.0	284
357.0	283
333.0	283
313.0	282
31.0	281
29.0	281
241.0	281
240.0	280
395.0	279
386.0	277
298.0	277
289.0	276
349.0	276
455.0	276
334.0	275
325.0	275
158.0	275
238.0	274
297.0	274
250.0	273

390.0	273
302.0	273
328.0	272
355.0	270
321.0	270
331.0	269
374.0	269
239.0	269
304.0	268
422.0	268
410.0	268
385.0	267
383.0	267
375.0	267
371.0	265
372.0	265
329.0	264
279.0	264
308.0	264
186.0	264
481.0	263
268.0	263
348.0	263
337.0	262
369.0	262
356.0	262
270.0	260
478.0	260
150.0	260
416.0	259
284.0	259
340.0	259
350.0	259
351.0	259
280.0	258
468.0	258
243.0	258
338.0	258
285.0	257
296.0	257

332.0	257
373.0	257
344.0	257
346.0	257
441.0	257
159.0	257
189.0	257
353.0	256
245.0	255
430.0	255
265.0	255
248.0	254
424.0	254
310.0	254
399.0	254
431.0	254
397.0	253
444.0	253
391.0	253
343.0	253
306.0	252
376.0	252
312.0	252
324.0	252
275.0	251
244.0	251
218.0	251
256.0	250
318.0	250
288.0	250
204.0	250
380.0	249
315.0	249
405.0	249
439.0	248
392.0	248
389.0	248
323.0	248
274.0	248
195.0	247

456.0	247
251.0	247
411.0	246
367.0	246
393.0	246
379.0	246
384.0	246
314.0	246
221.0	246
290.0	246
184.0	245
307.0	245
295.0	245
388.0	244
206.0	244
278.0	244
427.0	244
467.0	243
406.0	243
220.0	242
440.0	242
309.0	242
222.0	242
208.0	241
252.0	241
203.0	241
272.0	241
407.0	241
264.0	240
330.0	240
320.0	240
273.0	240
477.0	240
32.0	239
364.0	239
214.0	238
258.0	238
254.0	238
335.0	238
156.0	238

232.0	238
435.0	238
423.0	237
192.0	237
157.0	236
249.0	236
271.0	236
217.0	236
370.0	236
260.0	235
263.0	235
283.0	235
215.0	235
434.0	235
469.0	235
33.0	234
128.0	234
211.0	233
305.0	233
433.0	233
210.0	233
216.0	233
347.0	233
483.0	233
404.0	233
451.0	232
445.0	232
449.0	232
425.0	232
196.0	232
401.0	232
197.0	231
336.0	231
400.0	231
155.0	231
282.0	231
267.0	231
247.0	230
352.0	230
200.0	229

443.0	229
458.0	229
120.0	229
354.0	228
266.0	228
327.0	228
540.0	228
432.0	227
317.0	227
182.0	227
144.0	227
269.0	226
287.0	226
322.0	226
436.0	226
193.0	226
34.0	226
213.0	226
484.0	225
366.0	225
36.0	225
191.0	225
188.0	225
492.0	224
292.0	224
394.0	224
127.0	224
259.0	224
476.0	224
205.0	223
414.0	223
450.0	223
491.0	223
378.0	222
179.0	222
291.0	222
365.0	222
402.0	222
181.0	222
448.0	221

465.0	221
209.0	221
473.0	221
281.0	221
502.0	221
170.0	221
261.0	220
415.0	220
235.0	220
246.0	220
234.0	220
201.0	220
180.0	219
164.0	219
345.0	219
413.0	219
194.0	219
276.0	218
257.0	218
185.0	218
161.0	218
442.0	218
130.0	217
510.0	216
475.0	216
471.0	216
253.0	216
237.0	215
454.0	215
398.0	215
493.0	215
403.0	215
255.0	215
133.0	215
168.0	215
460.0	215
262.0	215
152.0	214
229.0	214
145.0	214

134.0	214
294.0	214
516.0	214
286.0	214
446.0	214
35.0	213
160.0	213
485.0	213
138.0	213
178.0	212
231.0	212
447.0	212
174.0	212
453.0	212
538.0	211
190.0	211
207.0	211
118.0	211
501.0	211
242.0	210
466.0	210
541.0	209
464.0	209
457.0	209
165.0	209
230.0	209
226.0	208
511.0	208
470.0	208
429.0	208
437.0	207
452.0	207
539.0	207
474.0	207
368.0	206
426.0	206
319.0	206
172.0	206
462.0	205
135.0	205

227.0	205
202.0	205
167.0	204
459.0	204
173.0	204
154.0	204
236.0	204
377.0	204
518.0	204
166.0	203
212.0	203
412.0	203
140.0	202
496.0	202
153.0	202
104.0	201
177.0	201
293.0	201
124.0	201
125.0	201
171.0	200
513.0	200
126.0	200
488.0	198
534.0	198
519.0	198
37.0	197
183.0	197
225.0	196
500.0	196
428.0	196
119.0	196
122.0	195
148.0	194
228.0	194
487.0	194
115.0	194
499.0	194
162.0	193
461.0	193

512.0	193
508.0	193
472.0	193
505.0	192
490.0	192
121.0	192
533.0	191
151.0	191
198.0	191
504.0	190
532.0	190
223.0	190
438.0	190
169.0	190
316.0	190
146.0	190
132.0	190
149.0	189
187.0	188
52.0	188
101.0	188
107.0	188
116.0	188
489.0	188
529.0	187
137.0	187
38.0	187
175.0	187
497.0	185
199.0	185
503.0	185
79.0	184
536.0	184
507.0	184
163.0	184
100.0	184
89.0	184
84.0	184
48.0	183
97.0	183

49.0	183
143.0	182
136.0	182
112.0	182
86.0	180
554.0	180
542.0	180
142.0	180
517.0	179
59.0	179
486.0	179
61.0	178
63.0	178
514.0	178
80.0	177
521.0	177
54.0	177
498.0	177
506.0	177
66.0	177
69.0	176
96.0	176
109.0	175
527.0	175
94.0	174
463.0	174
50.0	174
40.0	174
129.0	173
139.0	173
515.0	173
39.0	173
108.0	173
509.0	172
141.0	172
524.0	172
224.0	172
537.0	171
70.0	171
131.0	171

72.0	171
105.0	171
522.0	171
528.0	171
95.0	170
520.0	170
494.0	169
176.0	169
523.0	169
110.0	168
91.0	167
551.0	167
43.0	167
51.0	166
550.0	166
53.0	166
147.0	166
106.0	165
99.0	165
102.0	165
85.0	165
90.0	165
552.0	164
93.0	164
495.0	164
77.0	163
233.0	163
114.0	162
601.0	162
111.0	162
47.0	161
560.0	161
123.0	160
562.0	160
62.0	159
88.0	158
71.0	157
598.0	157
530.0	157
45.0	157

561.0	156
117.0	156
572.0	156
98.0	156
544.0	155
599.0	155
525.0	155
535.0	155
42.0	154
56.0	154
65.0	153
78.0	153
55.0	153
75.0	152
563.0	152
81.0	151
566.0	151
60.0	150
546.0	150
595.0	150
103.0	149
58.0	149
46.0	149
68.0	148
569.0	147
543.0	147
113.0	147
555.0	146
549.0	146
41.0	146
553.0	146
73.0	146
57.0	145
545.0	145
92.0	145
64.0	144
567.0	144
559.0	141
526.0	141
600.0	141

76.0	141
571.0	140
547.0	140
74.0	139
82.0	139
565.0	139
580.0	139
581.0	138
579.0	137
602.0	137
582.0	136
564.0	135
87.0	135
570.0	135
83.0	135
531.0	133
585.0	133
568.0	133
556.0	132
44.0	131
575.0	130
574.0	130
594.0	129
548.0	128
576.0	126
577.0	124
558.0	123
67.0	123
603.0	123
573.0	122
588.0	121
604.0	118
597.0	117
578.0	117
607.0	113
593.0	113
557.0	112
596.0	111
606.0	109
587.0	108

586.0	108
589.0	106
625.0	105
583.0	104
592.0	102
584.0	100
609.0	98
610.0	96
591.0	95
640.0	95
605.0	92
632.0	92
611.0	89
626.0	88
613.0	87
636.0	87
612.0	86
622.0	86
628.0	85
618.0	85
623.0	84
608.0	84
590.0	83
630.0	81
660.0	81
637.0	80
620.0	80
633.0	79
653.0	78
617.0	78
621.0	78
639.0	76
635.0	75
634.0	75
629.0	75
631.0	74
624.0	74
615.0	73
627.0	72
641.0	72

673.0	69
616.0	69
638.0	69
671.0	68
645.0	67
659.0	66
663.0	66
707.0	66
649.0	64
670.0	64
619.0	64
648.0	63
651.0	63
652.0	63
666.0	62
647.0	62
662.0	62
650.0	61
658.0	61
646.0	61
614.0	60
681.0	59
679.0	58
694.0	58
667.0	58
644.0	58
643.0	57
682.0	56
706.0	54
685.0	54
661.0	54
711.0	53
687.0	53
674.0	52
684.0	52
654.0	51
683.0	51
686.0	51
655.0	50
656.0	50

657.0	50
721.0	49
701.0	49
698.0	49
675.0	49
665.0	49
672.0	49
689.0	49
691.0	48
695.0	48
642.0	47
717.0	46
719.0	46
688.0	46
704.0	45
708.0	45
705.0	45
664.0	45
668.0	45
690.0	44
722.0	44
680.0	44
720.0	43
702.0	42
677.0	42
750.0	40
693.0	40
676.0	40
669.0	39
692.0	38
723.0	37
703.0	36
718.0	35
732.0	34
749.0	33
696.0	33
678.0	33
697.0	32
709.0	32
728.0	31

730.0	31
713.0	30
731.0	30
725.0	30
710.0	30
734.0	30
724.0	30
751.0	30
759.0	29
740.0	29
700.0	29
715.0	29
714.0	29
744.0	28
742.0	27
738.0	27
712.0	27
716.0	27
741.0	27
743.0	26
699.0	26
726.0	25
754.0	25
736.0	25
786.0	24
782.0	24
748.0	24
727.0	23
755.0	22
761.0	22
746.0	22
739.0	21
733.0	21
735.0	21
762.0	20
781.0	19
774.0	19
729.0	18
791.0	18
769.0	18

777.0	18
770.0	18
753.0	18
779.0	18
756.0	18
812.0	17
757.0	17
758.0	17
778.0	17
803.0	17
780.0	17
737.0	16
745.0	16
800.0	16
752.0	16
814.0	16
837.0	16
793.0	16
747.0	15
787.0	15
767.0	15
818.0	15
811.0	15
801.0	15
765.0	15
771.0	14
766.0	14
763.0	14
820.0	14
790.0	13
775.0	13
822.0	13
792.0	13
773.0	13
789.0	13
810.0	13
843.0	12
788.0	12
819.0	12
764.0	12

816.0	12
760.0	12
776.0	11
914.0	11
892.0	11
863.0	11
857.0	11
870.0	11
804.0	11
802.0	11
828.0	11
842.0	11
1012.0	10
963.0	10
796.0	10
807.0	10
798.0	10
848.0	10
905.0	10
927.0	10
815.0	10
840.0	10
824.0	10
846.0	10
821.0	10
813.0	9
772.0	9
860.0	9
831.0	9
784.0	9
823.0	9
929.0	9
908.0	9
851.0	9
829.0	9
768.0	9
868.0	9
849.0	9
951.0	9
808.0	9

838.0	9
986.0	8
785.0	8
835.0	8
925.0	8
850.0	8
1009.0	8
856.0	8
805.0	8
909.0	8
956.0	8
806.0	8
794.0	8
847.0	8
844.0	8
795.0	8
890.0	8
903.0	8
855.0	8
902.0	8
875.0	7
926.0	7
817.0	7
836.0	7
882.0	7
833.0	7
1114.0	7
897.0	7
906.0	7
878.0	7
830.0	7
940.0	7
809.0	7
858.0	7
928.0	7
898.0	7
907.0	7
886.0	7
845.0	7
799.0	7

975.0	6
924.0	6
827.0	6
950.0	6
911.0	6
853.0	6
877.0	6
917.0	6
874.0	6
1132.0	6
1030.0	6
852.0	6
945.0	6
896.0	6
832.0	6
1438.0	6
880.0	6
932.0	6
919.0	6
861.0	6
839.0	6
834.0	6
901.0	6
783.0	6
881.0	6
893.0	6
797.0	6
923.0	6
920.0	5
899.0	5
891.0	5
952.0	5
885.0	5
1165.0	5
879.0	5
1167.0	5
1038.0	5
826.0	5
841.0	5
1037.0	5

859.0	5
1061.0	5
1059.0	5
904.0	5
865.0	5
964.0	5
1075.0	5
1049.0	5
1013.0	5
1035.0	5
984.0	5
930.0	5
887.0	5
900.0	5
1146.0	4
1129.0	4
1066.0	4
876.0	4
998.0	4
864.0	4
994.0	4
989.0	4
934.0	4
1028.0	4
949.0	4
1003.0	4
869.0	4
1001.0	4
910.0	4
1032.0	4
1187.0	4
1043.0	4
1149.0	4
938.0	4
991.0	4
825.0	4
1204.0	4
921.0	4
1046.0	4
1048.0	4

1431.0	4
862.0	4
895.0	4
1005.0	4
942.0	4
1427.0	4
955.0	4
889.0	4
1214.0	4
968.0	4
1091.0	3
931.0	3
918.0	3
1384.0	3
1047.0	3
1380.0	3
888.0	3
1029.0	3
1131.0	3
948.0	3
1116.0	3
1228.0	3
1071.0	3
1182.0	3
961.0	3
912.0	3
1095.0	3
1301.0	3
1426.0	3
1432.0	3
988.0	3
1105.0	3
1259.0	3
1145.0	3
1002.0	3
995.0	3
933.0	3
1213.0	3
922.0	3
1415.0	3

871.0	3
1433.0	3
1430.0	3
1195.0	3
1170.0	3
936.0	3
1141.0	3
979.0	3
944.0	3
867.0	3
958.0	3
1033.0	3
1173.0	3
1021.0	3
1025.0	3
1234.0	3
1060.0	3
1370.0	3
1096.0	3
1175.0	3
1413.0	3
976.0	3
980.0	3
1052.0	3
866.0	3
1093.0	3
943.0	3
1076.0	3
1023.0	3
1010.0	3
1036.0	3
935.0	3
884.0	2
1237.0	2
996.0	2
1101.0	2
1374.0	2
1110.0	2
959.0	2
1045.0	2

1011.0	2
992.0	2
987.0	2
1157.0	2
982.0	2
1015.0	2
1156.0	2
1421.0	2
1055.0	2
1000.0	2
1331.0	2
873.0	2
1057.0	2
1414.0	2
1058.0	2
1166.0	2
894.0	2
1168.0	2
1302.0	2
1439.0	2
1128.0	2
1139.0	2
1127.0	2
1016.0	2
1177.0	2
1191.0	2
1008.0	2
1118.0	2
1291.0	2
1274.0	2
1103.0	2
990.0	2
1250.0	2
916.0	2
957.0	2
1026.0	2
1418.0	2
1099.0	2
985.0	2
1178.0	2

1044.0	2
1241.0	2
1086.0	2
1152.0	2
1112.0	2
1295.0	2
1183.0	2
1201.0	2
1209.0	2
1366.0	2
1087.0	2
872.0	2
1289.0	2
977.0	2
1434.0	2
1367.0	2
1085.0	2
1135.0	2
1081.0	2
1188.0	2
1386.0	2
1223.0	2
1031.0	2
1160.0	2
1113.0	1
1416.0	1
1300.0	1
1192.0	1
1098.0	1
1349.0	1
854.0	1
1106.0	1
1282.0	1
1189.0	1
1429.0	1
1409.0	1
1410.0	1
1389.0	1
941.0	1
1163.0	1

1411.0	1
1292.0	1
1428.0	1
1253.0	1
1275.0	1
1309.0	1
1424.0	1
1437.0	1
1276.0	1
1084.0	1
1130.0	1
1225.0	1
1226.0	1
1233.0	1
1208.0	1
1362.0	1
1423.0	1
1041.0	1
1251.0	1
1396.0	1
1397.0	1
915.0	1
1215.0	1
999.0	1
1124.0	1
1123.0	1
1293.0	1
1100.0	1
1403.0	1
1227.0	1
1375.0	1
972.0	1
1398.0	1
1217.0	1
1341.0	1
1174.0	1
1102.0	1
1181.0	1
1007.0	1
1027.0	1

1245.0	1
1120.0	1
1020.0	1
1097.0	1
1412.0	1
1180.0	1
1344.0	1
1318.0	1
1014.0	1
1121.0	1
1176.0	1
1311.0	1
1197.0	1
1017.0	1
1395.0	1
1371.0	1
1372.0	1
1067.0	1
960.0	1
1133.0	1
997.0	1
954.0	1
1203.0	1
978.0	1
1136.0	1
1284.0	1
1376.0	1
1051.0	1
1162.0	1
1218.0	1
1080.0	1
1261.0	1
1354.0	1
1339.0	1
1321.0	1
1326.0	1
1327.0	1
1019.0	1
1229.0	1
1319.0	1

1077.0	1
1238.0	1
1090.0	1
1202.0	1
1392.0	1
974.0	1
1312.0	1
1324.0	1
1224.0	1
1155.0	1
1220.0	1
1232.0	1
1221.0	1
1297.0	1
1381.0	1
1407.0	1
1039.0	1
1179.0	1
971.0	1
1420.0	1
1050.0	1
1159.0	1
993.0	1
1401.0	1
1144.0	1
1290.0	1
1363.0	1
1158.0	1
1246.0	1
1314.0	1
1018.0	1
1305.0	1
953.0	1
913.0	1
1056.0	1
1148.0	1
1231.0	1
973.0	1
983.0	1
967.0	1

```
1379.0      1
1268.0      1
962.0       1
1053.0      1
1254.0      1
1330.0      1
1332.0      1
1348.0      1
1359.0      1
1369.0      1
1387.0      1
947.0       1
946.0       1
1184.0      1
1342.0      1
1352.0      1
Name: MIN_TEMPORAL_GAP, dtype: int64
```

```
In [ ]: # How many rows have a minimum temporal gap equal to NaN?
df[df['MIN_TEMPORAL_GAP'].isna()].shape[0]
```

```
Out[ ]: 4751190
```

```
In [ ]: # Cleaning operation: remove the rows using the minimum temporal gap

# Find a reasonable delta of MIN_TEMPORAL_GAP to remove the rows that have a minimum temporal gap for the same series
# Print the minimum value of the column MIN_TEMPORAL_GAP
print('The minimum value of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].min()))

# Print the maximum value of the column MIN_TEMPORAL_GAP
print('The maximum value of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].max()))

# Print the mean value of the column MIN_TEMPORAL_GAP
print('The mean value of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].mean()))

# Print the median value of the column MIN_TEMPORAL_GAP
print('The median value of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].median()))

# Print the standard deviation of the column MIN_TEMPORAL_GAP
print('The standard deviation of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].std()))
```

```
# Print the 0.05th percentile of the column MIN_TEMPORAL_GAP
print('The 0.05th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.05)))

# Print the 0.10th percentile of the column MIN_TEMPORAL_GAP
print('The 0.10th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.10)))

# Print the 25th percentile of the column MIN_TEMPORAL_GAP
print('The 25th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.25)))

# Print the 75th percentile of the column MIN_TEMPORAL_GAP
print('The 75th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.75)))

# Print the 90th percentile of the column MIN_TEMPORAL_GAP
print('The 90th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.90)))

# Print the 95th percentile of the column MIN_TEMPORAL_GAP
print('The 95th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.95)))

# Print the 99th percentile of the column MIN_TEMPORAL_GAP
print('The 99th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.99)))

# Print the 99.9th percentile of the column MIN_TEMPORAL_GAP
print('The 99.9th percentile of the column MIN_TEMPORAL_GAP is: {}'.format(df['MIN_TEMPORAL_GAP'].quantile(0.999)))

# Decide the delta of MIN_TEMPORAL_GAP using the 25th percentile of the column MIN_TEMPORAL_G
delta = df['MIN_TEMPORAL_GAP'].quantile(0.1)
if delta == 0:
    delta = df['MIN_TEMPORAL_GAP'].quantile(0.25)
if delta == 0:
    delta = df['MIN_TEMPORAL_GAP'].median()
print('The delta of MIN_TEMPORAL_GAP is: {}'.format(delta))
```

The minimum value of the column MIN_TEMPORAL_GAP is: 0.0
The maximum value of the column MIN_TEMPORAL_GAP is: 1439.0
The mean value of the column MIN_TEMPORAL_GAP is: 158.56392107131288
The median value of the column MIN_TEMPORAL_GAP is: 9.0
The standard deviation of the column MIN_TEMPORAL_GAP is: 212.0631750424889
The 0.05th percentile of the column MIN_TEMPORAL_GAP is: 0.0
The 0.10th percentile of the column MIN_TEMPORAL_GAP is: 0.0
The 25th percentile of the column MIN_TEMPORAL_GAP is: 1.0
The 75th percentile of the column MIN_TEMPORAL_GAP is: 312.0
The 90th percentile of the column MIN_TEMPORAL_GAP is: 487.0
The 95th percentile of the column MIN_TEMPORAL_GAP is: 571.0
The 99th percentile of the column MIN_TEMPORAL_GAP is: 732.0
The 99.9th percentile of the column MIN_TEMPORAL_GAP is: 1129.0
The delta of MIN_TEMPORAL_GAP is: 1.0

```
In [ ]: # Cleaning operation: remove the rows using the minimum temporal gap

# Save the number of rows before the cleaning operation
shape_before = df.shape[0]

# Delete the rows that have a minimum temporal gap for the same serial and fermata more than the delta calculated b
# Do not remove the rows with NaN values because they are the first validations of the day of a specific serial and
df = df[(df['MIN_TEMPORAL_GAP'] > delta) | (df['MIN_TEMPORAL_GAP'].isna())]

# Print the number of rows before and after the cleaning operation and the difference
print('The number of rows before the cleaning operation is: {}'.format(shape_before))
print('The number of rows after the cleaning operation is: {}'.format(df.shape[0]))
print('The difference is: {}'.format(shape_before - df.shape[0]))
# Calculate the percentage of rows that has just been deleted
print('The percentage of rows that has just been deleted is: {}'.format(round((shape_before - df.shape[0])/shape_b

The number of rows before the cleaning operation is: 5041376
The number of rows after the cleaning operation is: 4947461
The difference is: 93915
The percentage of rows that has just been deleted is: 1.86%
```

```
In [ ]: # Delete the column MIN_TEMPORAL_GAP because it is not useful anymore
df.drop('MIN_TEMPORAL_GAP', axis=1, inplace=True)
```

```
In [ ]: # Create a new dataframe, copied from the original one
df_new = df.copy()
```

```

# Update the column 'DESCRIZIONE' of the new df with the new values of the dictionary:
# the value that are present in the dataframe are the values of the dictionary; you have to substitute with the key
for key, value in dict_prefix.items():
    df_new['DESCRIZIONE'] = df_new['DESCRIZIONE'].replace(value, key)

# Print the head of the new dataframe
print(df_new.head())

# Export the new dataframe in a txt file
# The name of the file is dataset_cleaned followed by the name (file_name variable) of the file that has been cleaned
name_file = 'dataset_cleaned_' + file_name.split('.')[0] + '.txt'
df_new.to_csv('data/processed/' + name_file, sep='\t', index=False)

print('The script has finished')

```

	DATA	ORA	DATA_VALIDAZIONE	SERIALE	FERMATA	\
0	2023-01-13	00:00:00	2023-01-13 00:00:00	40834866809772548	162	
1	2023-01-13	00:00:00	2023-01-13 00:00:00	-3604990320	5049	
2	2023-01-13	00:00:00	2023-01-13 00:00:00	-2824230951	5043	
3	2023-01-13	00:00:00	2023-01-13 00:00:00	40552750134805252	5013	
4	2023-01-13	00:01:00	2023-01-13 00:01:00	-3604964420	6084	

	DESCRIZIONE	TITOLO	TICKET_CODE	DESCRIZIONE_TITOLO
0	STAZIONE MES	12101	7	BIGL.AUT.75'MESTRE/LIDO-TSC
1	ZATTERE	23301	5-STUD	MENS.STUDENTE RETE UNICA
2	S. TOMA' "B"	23303	6-STUD	ABB STUD. RETEUNICA 12 MESI
3	S. MARCO-SAN	11101	7	75'-TPL 8,64-COMVE0,86
4	VENEZIA	11209	7	BIGL RETE UNICA 75'

The script has finished