

Contents

1	Data Exploration	1
1.1	Fix path stuff	1
1.2	Imports	1
1.3	Read csv	1
1.4	Look at data	2
1.5	Competitor information	3
1.6	Picking top 75%	3

1 Data Exploration

1.1 Fix path stuff

```
import sys
import os
src_dir = os.path.join(os.getcwd(), '..', 'src')
sys.path.append(src_dir)
```

None

1.2 Imports

```
import pandas as pd
from matplotlib import use
import matplotlib.pyplot as plt
from d02_intermediate.transform_int_data import clean_csv

matplotlib.use('GTK3Agg')
```

1.3 Read csv

```
df = clean_csv()
df
```

	Location	Year	Competitor	Round	Success
0	Cairo	2023	Xenos Dionysios	Semi	Win
1	Cairo	2023	Amirali Didar	Semi	Loss
2	Cairo	2023	Nakano Hiroki	Semi	Loss
3	Cairo	2023	Da Costa Steven	Semi	Win
4	Cairo	2023	Xenos Dionysios	Final	Loss

..
103	Matosinhos	2022	Da Costa Steven	Final	Win
104	Matosinhos	2022	Pisino Noah	Bronze	Win
105	Matosinhos	2022	Oubaya Said	Bronze	Loss
106	Matosinhos	2022	Aghalarzade Tural	Bronze	Loss
107	Matosinhos	2022	Nakano Soichiro	Bronze	Win

[108 rows x 5 columns]

1.4 Look at data

```
comp_counts = df.groupby("Competitor")["Competitor"].count()
comp_counts
```

Competitor	
Aghalarzade Tural	5
Almasatfa Abdel Rahman	10
Alzahrani Sultan	1
Amirali Didar	8
Da Costa Steven	6
De Vivo Gianluca	5
Deniz Muratcan	1
Elsawy Ali	8
Figueira Vinicius	2
Friere Fuentes Tomas	1
Ghaith Afeef	2
Kozaki Yugo	5
Lotfy Ahmed	5
Maresca Luca	4
Milkhailichenko Danila	1
Muratov Assylbek	2
Nakano Hiroki	2
Nakano Soichiro	4
Nhuyen Devid	2
Oubaya Said	4
Oulad Haj Amar Youness	3
Pisino Noah	1
Pokorny Stefan	1
Rodrigues Jesus Leonardo	1
Rosiello Jess	2

```

Sharafutdinov Ernest      1
Tadissi Yves Martial      8
Uygur Burak              4
Xenos Dionysios           8
Yanovskyi Davyd          1
Name: Competitor, dtype: int64

```

1.5 Competitor information

```

comp_counts.describe()

count    30.000000
mean      3.600000
std       2.685851
min       1.000000
25%       1.000000
50%       2.500000
75%       5.000000
max      10.000000
Name: Competitor, dtype: float64

```

1.6 Picking top 75%

```

names = comp_counts[comp_counts >= 5].index

df[df["Competitor"].isin(names)].reset_index(drop=True)

   Location  Year Competitor Round Success
0      Cairo  2023   Xenos Dionysios   Semi    Win
1      Cairo  2023   Amirali Didar   Semi   Loss
2      Cairo  2023   Da Costa Steven   Semi    Win
3      Cairo  2023   Xenos Dionysios  Final   Loss
4      Cairo  2023   Da Costa Steven  Final    Win
..      ...   ...         ...     ...    ...
63  Matosinhos  2022  Almasatfa Abdel Rahman   Semi    Win
64  Matosinhos  2022   Da Costa Steven   Semi    Win
65  Matosinhos  2022  Almasatfa Abdel Rahman  Final   Loss
66  Matosinhos  2022   Da Costa Steven  Final    Win
67  Matosinhos  2022  Aghalarzade Tural  Bronze   Loss

[68 rows x 5 columns]

```