Contents

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

use('GTK3Agg')
```

```
None
```

1.3 Read csv

```
#df = clean_csv()     # No longer in use

medal_data_path = os.path.join("data", "01_raw", "medal_data.csv")
# Fix path if called from ntoebook
```

```
if not os.path.exists(medal_data_path):
medal_data_path = os.path.join("../", "data", "01_raw", "medal_data.csv")

df = pd.read_csv(medal_data_path)
df
```

1.4   Look at data

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

1.5   Competitor information

```
round_counts.describe()
```

1.6   Picking top 75%

```
names = round_counts[round_counts >= 5].index

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

1.7   Change success to True/False

```
df["Success"] = (df["Success"] == "Win")
```

1.8   Record total rounds

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

1.9   Record total wins

```
total_wins = df.groupby("Competitor")["Success"].sum()
total_wins
```