In [20]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [21]:
```
# Load the dataset
df = pd.read_csv("odb.csv")
df
```

Out[21]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21368	86.23	49
1	1	V Kohli (INDIA)	2008-2021	254	245	39	12169	183	59.07	13061	93.17	43
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	RG Sharma (INDIA)	2007-2021	227	220	32	9205	264	48.96	10354	88.90	29
4	4	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.20	28
...	...	...	...	...	...	...	...	...	...	...	...	...
114	114	SK Raina (INDIA)	2005-2018	226	194	35	5615	116*	35.31	6005	93.50	5
115	115	H Masakadza (ZIM)	2001-2019	209	208	4	5658	178*	27.73	7728	73.21	5
116	116	RB Richardson (WI)	1983-1996	224	217	30	6248	122	33.41	9801	63.74	5
117	117	BB McCullum (NZ)	2002-2016	260	228	28	6083	166	30.41	6312	96.37	5
118	118	Saleem Malik (PAK)	1982-1999	283	256	38	7170	102	32.88	9383	76.41	5

119 rows × 16 columns

In [22]:
```
df.shape
```

Out[22]: (119, 16)

In [23]:
```
df.isnull()
```

Out[23]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	!
0	False	False	False	False	False	False	False	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	False	False	False	Fal
...	...	...	...	...	...	...	...	...	...	...	...	...	
114	False	False	False	False	False	False	False	False	False	False	False	False	Fal
115	False	False	False	False	False	False	False	False	False	False	False	False	Fal
116	False	False	False	False	False	False	False	False	False	False	False	False	Fal
117	False	False	False	False	False	False	False	False	False	False	False	False	Fal
118	False	False	False	False	False	False	False	False	False	False	False	False	Fal

119 rows × 16 columns

In [24]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119 entries, 0 to 118
Data columns (total 16 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  119 non-null    int64
 1   Player      119 non-null    object
 2   Span        119 non-null    object
 3   Mat         119 non-null    int64
 4   Inns        119 non-null    int64
 5   NO          119 non-null    int64
 6   Runs        119 non-null    int64
 7   HS          119 non-null    object
 8   Ave         119 non-null    float64
 9   BF          119 non-null    int64
 10  SR          119 non-null    float64
 11  100         119 non-null    int64
 12  50          119 non-null    int64
 13  0           119 non-null    int64
 14  4s          119 non-null    object
 15  6s          119 non-null    object
dtypes: float64(2), int64(9), object(5)
memory usage: 15.0+ KB
```

In [25]:
```
df.describe()
```

Out[25]:

	Unnamed: 0	Mat	Inns	NO	Runs	Ave	
count	119.000000	119.000000	119.000000	119.00000	119.000000	119.000000	119.0000
mean	59.000000	194.672269	184.411765	21.10084	6306.890756	39.347983	7897.9411
std	34.496377	94.879684	89.115276	15.81803	3091.187863	7.013547	3894.8479
min	0.000000	33.000000	32.000000	1.00000	1447.000000	23.570000	1360.0000
25%	29.500000	124.500000	120.000000	8.00000	3928.500000	34.910000	5086.0000
50%	59.000000	190.000000	181.000000	16.00000	5964.000000	38.000000	7678.0000
75%	88.500000	247.500000	236.500000	33.00000	8079.000000	42.450000	9780.5000
max	118.000000	463.000000	452.000000	84.00000	18426.000000	67.000000	21368.0000

In [26]:
```
df_clean = df.dropna(how='all')
df_clean
df
```

Out[26]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21368	86.23	49
1	1	V Kohli (INDIA)	2008-2021	254	245	39	12169	183	59.07	13061	93.17	43
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	RG Sharma (INDIA)	2007-2021	227	220	32	9205	264	48.96	10354	88.90	29
4	4	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.20	28
...	...	...	...	...	...	...	...	...	...	...	...	...
114	114	SK Raina (INDIA)	2005-2018	226	194	35	5615	116*	35.31	6005	93.50	5
115	115	H Masakadza (ZIM)	2001-2019	209	208	4	5658	178*	27.73	7728	73.21	5
116	116	RB Richardson (WI)	1983-1996	224	217	30	6248	122	33.41	9801	63.74	5
117	117	BB McCullum (NZ)	2002-2016	260	228	28	6083	166	30.41	6312	96.37	5
118	118	Saleem Malik (PAK)	1982-1999	283	256	38	7170	102	32.88	9383	76.41	5

119 rows × 16 columns

In [29]:
```
df_clean = df.dropna(axis=1)
df
```

Out[29]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21368	86.23	49
1	1	V Kohli (INDIA)	2008-2021	254	245	39	12169	183	59.07	13061	93.17	43
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	RG Sharma (INDIA)	2007-2021	227	220	32	9205	264	48.96	10354	88.90	29
4	4	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.20	28
...	...	...	...	...	...	...	...	...	...	...	...	...
114	114	SK Raina (INDIA)	2005-2018	226	194	35	5615	116*	35.31	6005	93.50	5
115	115	H Masakadza (ZIM)	2001-2019	209	208	4	5658	178*	27.73	7728	73.21	5
116	116	RB Richardson (WI)	1983-1996	224	217	30	6248	122	33.41	9801	63.74	5
117	117	BB McCullum (NZ)	2002-2016	260	228	28	6083	166	30.41	6312	96.37	5
118	118	Saleem Malik (PAK)	1982-1999	283	256	38	7170	102	32.88	9383	76.41	5

119 rows × 16 columns

In [30]:
```
df.head(10)
```

Out[30]:

	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
0	0	SR Tendulkar (INDIA)	1989-2012	463	452	41	18426	200*	44.83	21368	86.23	49
1	1	V Kohli (INDIA)	2008-2021	254	245	39	12169	183	59.07	13061	93.17	43
2	2	RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30
3	3	RG Sharma (INDIA)	2007-2021	227	220	32	9205	264	48.96	10354	88.90	29
4	4	ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.20	28
5	5	HM Amla (SA)	2008-2019	181	178	14	8113	159	49.46	9178	88.39	27
6	6	AB de Villiers (Afr/SA)	2005-2018	228	218	39	9577	176	53.50	9473	101.09	25
7	7	CH Gayle (ICC/WI)	1999-2019	301	294	17	10480	215	37.83	12019	87.19	25
8	8	KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25
9	9	SC Ganguly (Asia/INDIA)	1992-2007	311	300	23	11363	183	41.02	15416	73.70	22

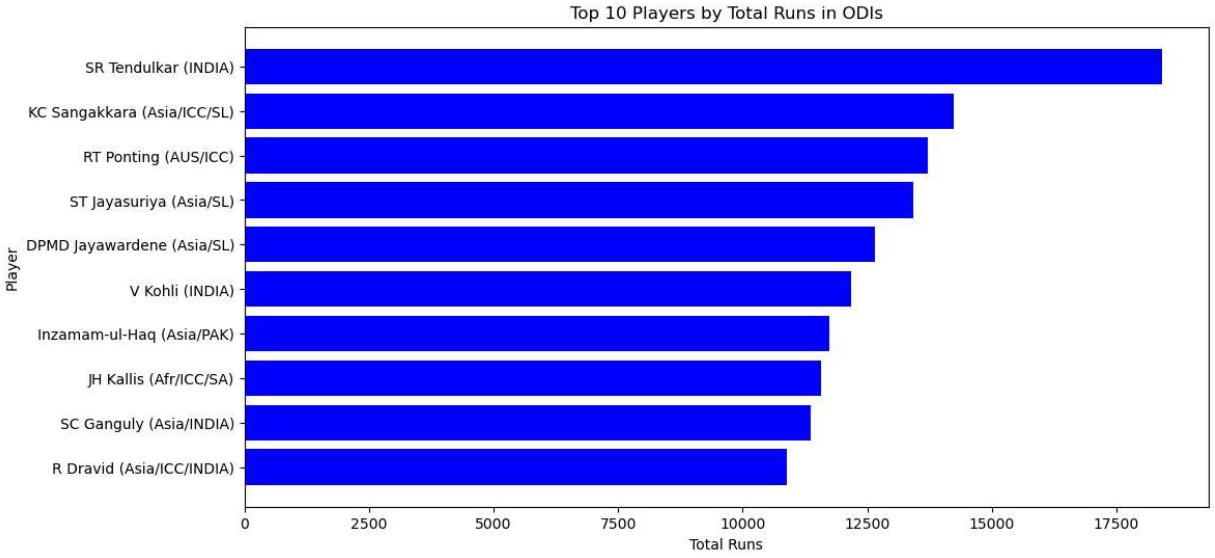
In [31]:
```
df.tail(10)
```

Out[31]:

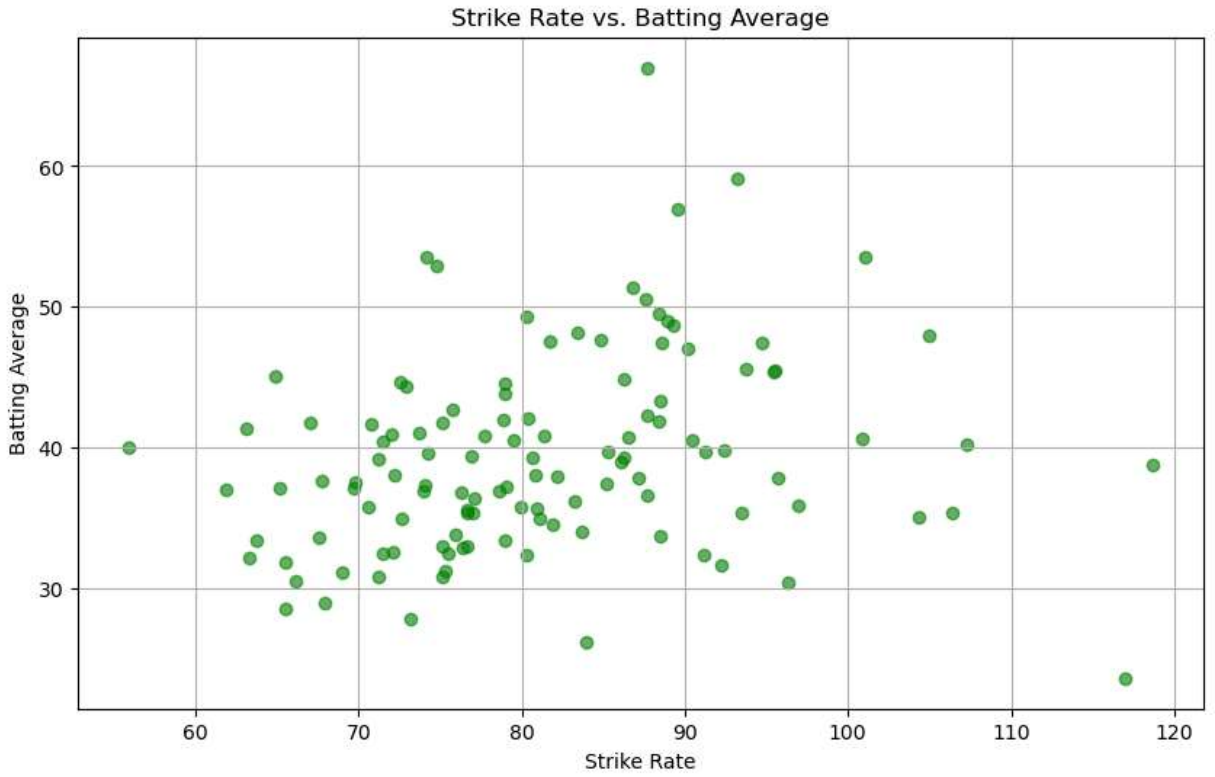
	Unnamed: 0	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100
109	109	Aamer Sohail (PAK)	1990-2000	156	155	5	4780	134	31.86	7297	65.50	5
110	110	RR Sarwan (WI)	2000-2013	181	169	33	5804	120*	42.67	7663	75.74	5
111	111	DC Boon (AUS)	1984-1995	181	177	16	5964	122	37.04	9157	65.13	5
112	112	PD Collingwood (ENG)	2001-2011	197	181	37	5092	120*	35.36	6614	76.98	5
113	113	DR Martyn (AUS)	1992-2006	208	182	51	5346	144*	40.80	6877	77.73	5
114	114	SK Raina (INDIA)	2005-2018	226	194	35	5615	116*	35.31	6005	93.50	5
115	115	H Masakadza (ZIM)	2001-2019	209	208	4	5658	178*	27.73	7728	73.21	5
116	116	RB Richardson (WI)	1983-1996	224	217	30	6248	122	33.41	9801	63.74	5
117	117	BB McCullum (NZ)	2002-2016	260	228	28	6083	166	30.41	6312	96.37	5
118	118	Saleem Malik (PAK)	1982-1999	283	256	38	7170	102	32.88	9383	76.41	5

In [32]:

```
import matplotlib.pyplot as plt

# Top 10 players by total runs
top_runs = df.nlargest(10, 'Runs')

plt.figure(figsize=(12, 6))
plt.barh(top_runs['Player'], top_runs['Runs'], color='blue')
plt.xlabel("Total Runs")
plt.ylabel("Player")
plt.title("Top 10 Players by Total Runs in ODIs")
plt.gca().invert_yaxis()
plt.show()
```

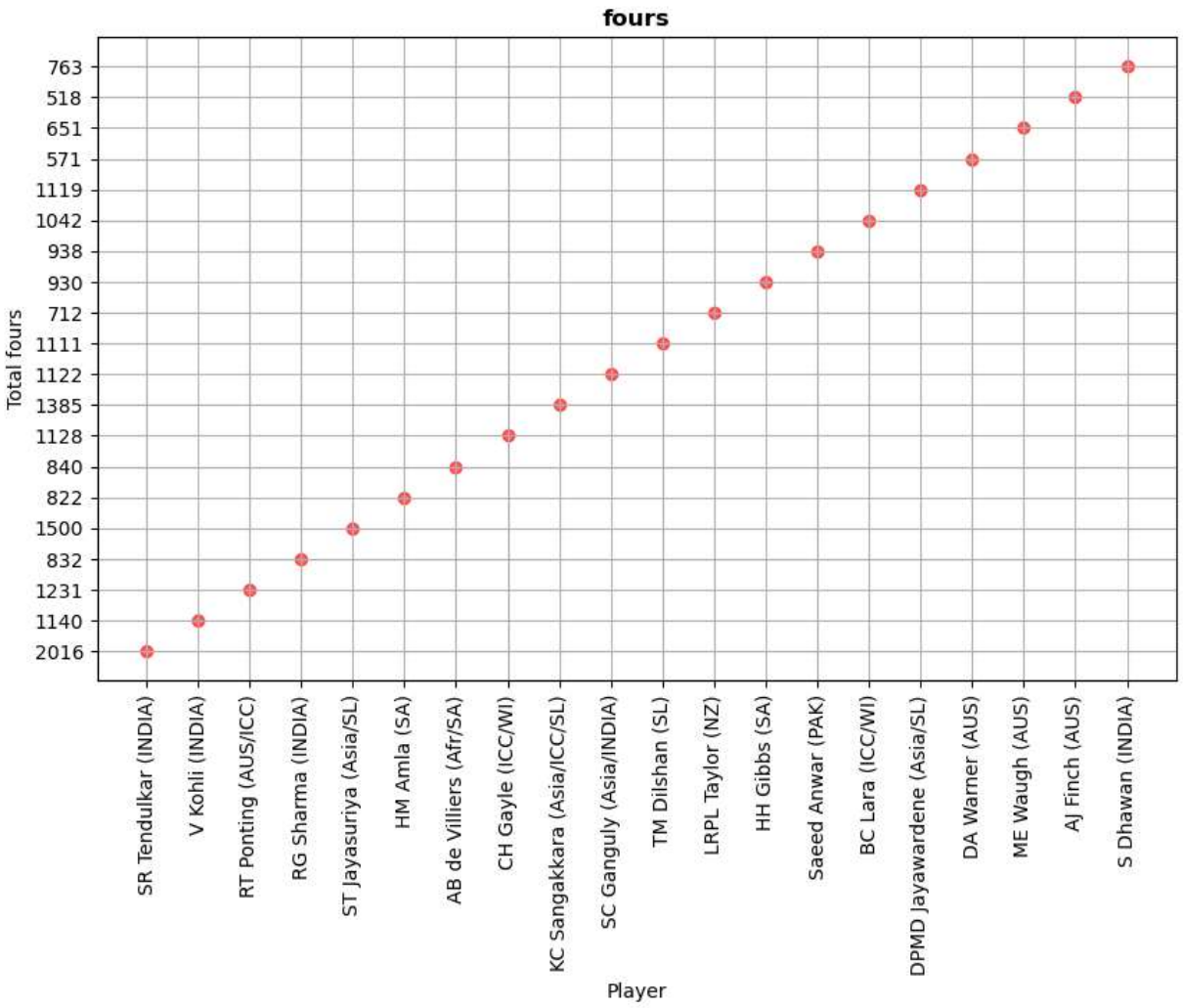


In [33]:
```
# Scatter plot: Strike Rate vs. Batting Average
plt.figure(figsize=(10, 6))
plt.scatter(df['SR'], df['Ave'], color='green', alpha=0.6)
plt.xlabel("Strike Rate")
plt.ylabel("Batting Average")
plt.title("Strike Rate vs. Batting Average")
plt.grid(True)
plt.show()
```

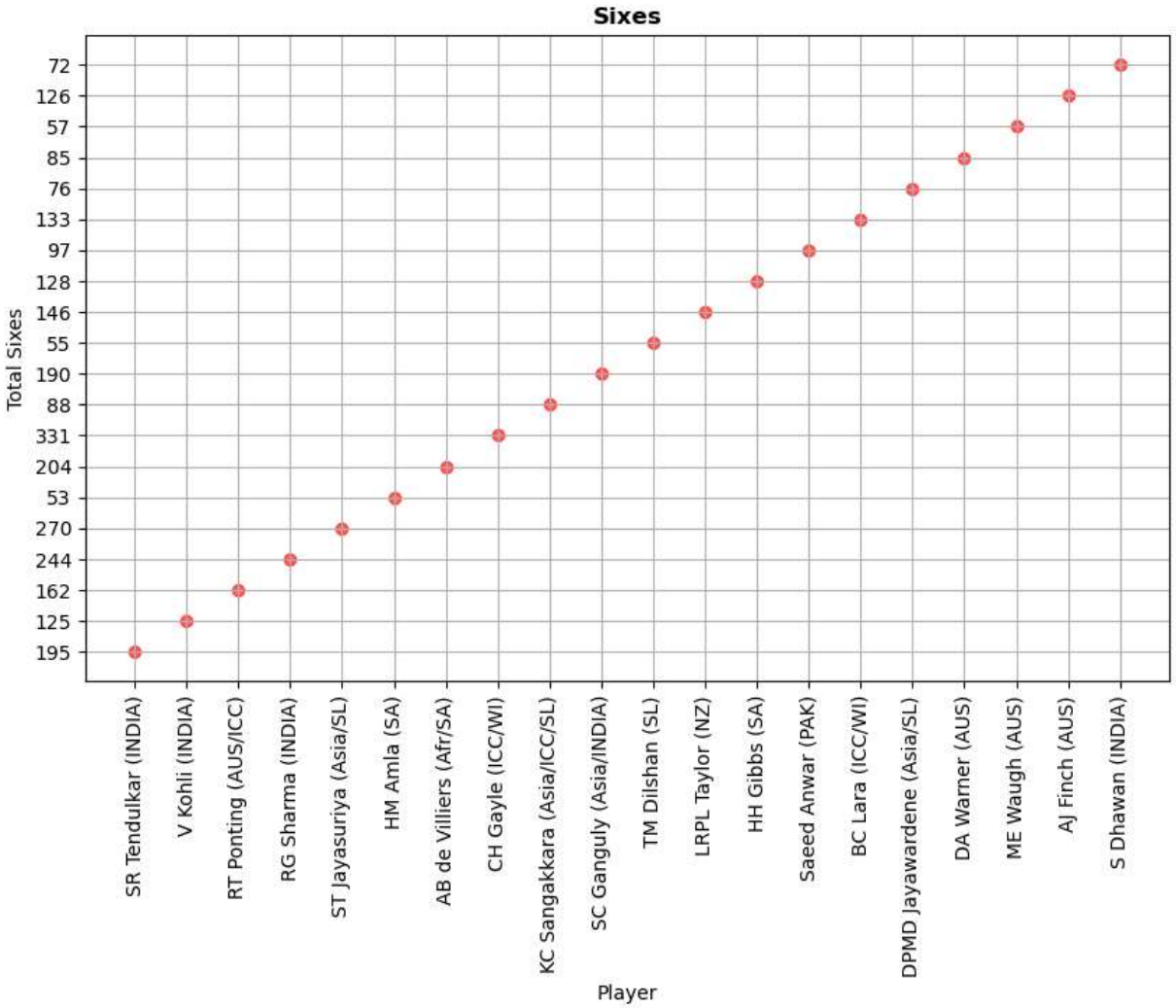


In [36]:
```
# Scatter plot: Sixes
plt.figure(figsize=(10, 6))
plt.scatter(df['Player'][:20], df['4s'][:20], color='red', alpha=0.6)
plt.xticks(rotation=90)
plt.xlabel("Player")
```

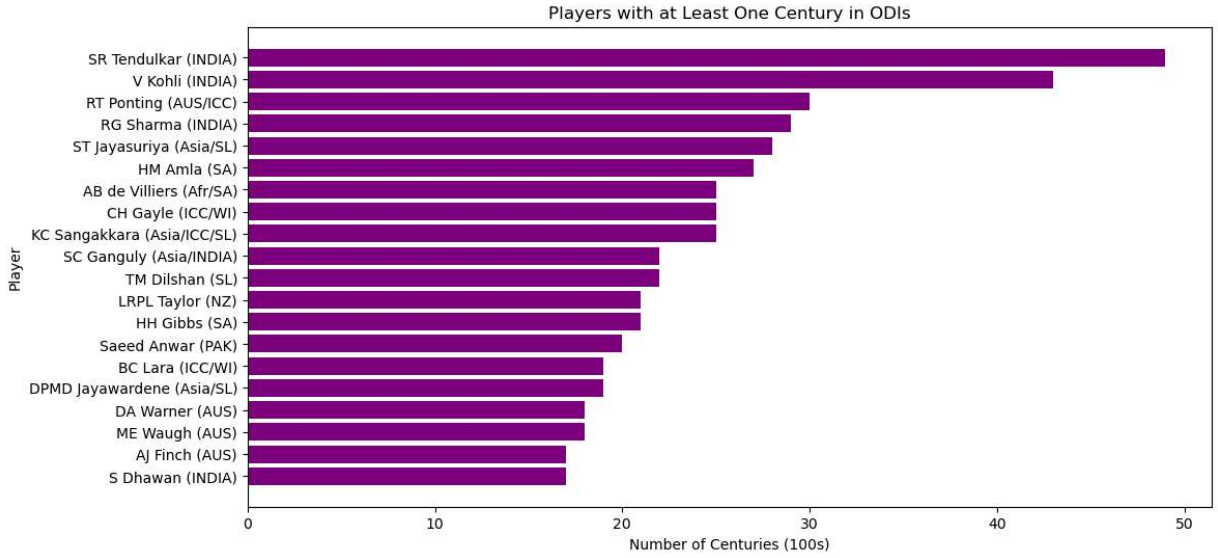
```
plt.ylabel("Total fours")
plt.title("fours", fontweight="bold")
plt.grid(True)
plt.show()
```



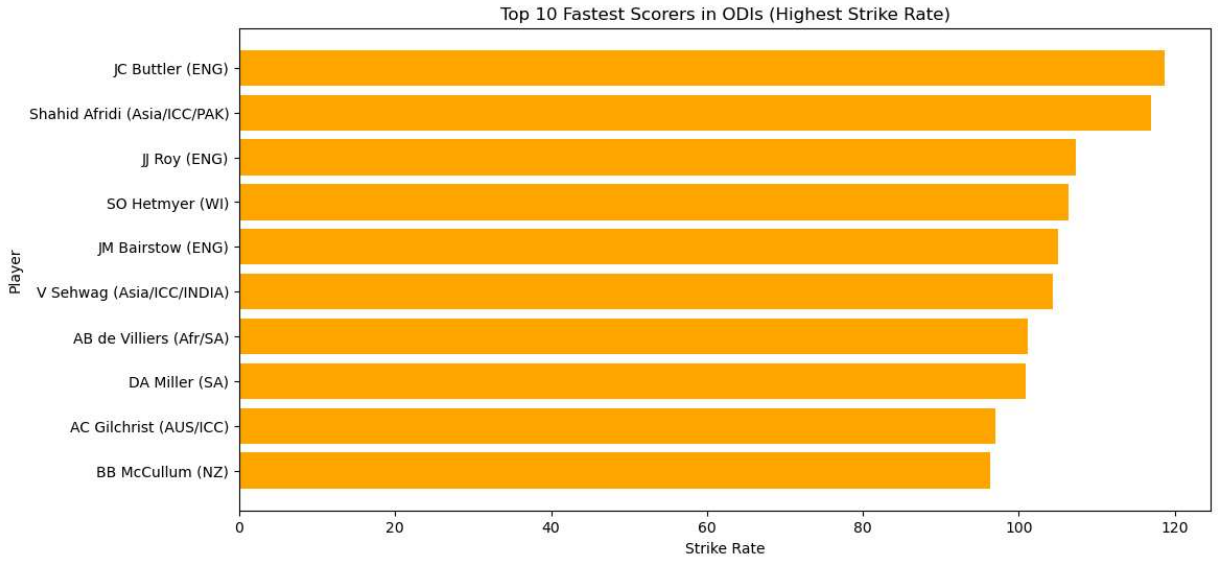
In [37]:
```
# Scatter plot: Sixes
plt.figure(figsize=(10, 6))
plt.scatter(df['Player'][:20], df['6s'][:20], color='red', alpha=0.6)
plt.xticks(rotation=90)
plt.xlabel("Player")
plt.ylabel("Total Sixes")
plt.title("Sixes", fontweight="bold")
plt.grid(True)
plt.show()
```



In [38]:
```
# Players with at least one century (100s >= 1)
century_players = df[df['100'] > 0]

# Bar chart: Players vs. Number of Centuries
plt.figure(figsize=(12, 6))
plt.barh(century_players['Player'][:20], century_players['100'][:20], color='purple
plt.xlabel("Number of Centuries (100s)")
plt.ylabel("Player")
plt.title("Players with at Least One Century in ODIs")
plt.gca().invert_yaxis()
plt.show()
```

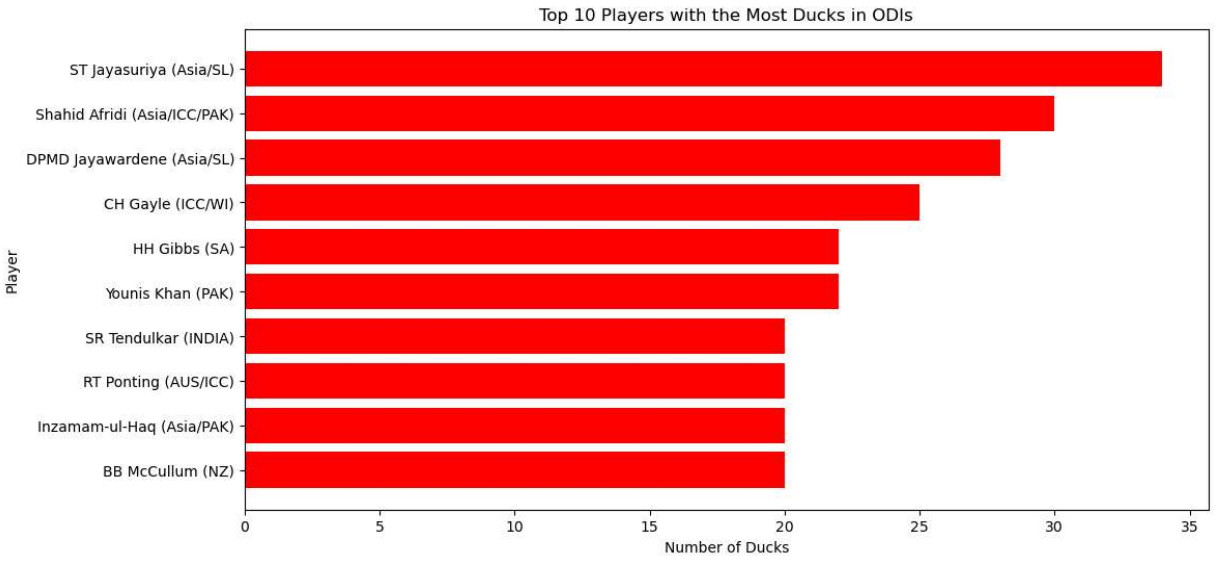


In [40]:
```python
# Top 10 players by Strike Rate
top_strike_rate = df.nlargest(10, 'SR')

# Bar chart for Strike Rate
plt.figure(figsize=(12, 6))
plt.barh(top_strike_rate['Player'], top_strike_rate['SR'], color='orange')
plt.xlabel("Strike Rate")
plt.ylabel("Player")
plt.title("Top 10 Fastest Scorers in ODIs (Highest Strike Rate)")
plt.gca().invert_yaxis()
plt.show()
```



In [41]:
```python
# Top 10 players with the most ducks
top_ducks = df.nlargest(10, '0')

# Bar chart for most ducks
plt.figure(figsize=(12, 6))
plt.barh(top_ducks['Player'], top_ducks['0'], color='red')
plt.xlabel("Number of Ducks")
plt.ylabel("Player")
plt.title("Top 10 Players with the Most Ducks in ODIs")
```

```
plt.gca().invert_yaxis()
plt.show()
```



In [ ]: