Neeraj Chopra's Best Throws Analysis and Prediction using Machine Learning


```python
import pandas as pd

data = {
    'Rank': list(range(1, 35)),
    'Throw': [
        '89.94m', '89.45m', '89.34m', '89.30m', '89.08m', '88.88m',
'88.77m', '88.67m', '88.44m', '88.39m',
        '88.36m', '88.17m', '88.13m', '88.07m', '88.06m', '88.00m',
'87.86m', '87.80m', '87.73m', '87.66m',
        '87.58m', '87.46m', '87.43m', '87.03m', '87.00m', '86.92m',
'86.84m', '86.79m', '86.69m', '86.67m',
        '86.65m', '86.52m', '86.48m', '86.47m'
    ],
    'Competition': [
        'Stockholm Diamond League 2022', 'Paris 2024 Olympics - F',
'Paris 2024 Olympics - Q', 'Paavo Nurmi Games (Turku)',
        'Lausanne Diamond League 2022', 'Asian Games 2023 (Hangzhou)',
'World Athletics Championships 2023 (Budapest) - Q',
        'Doha Diamond League 2023', 'Zurich Diamond League Final
2022', 'World Athletics Championships 2022 (Oregon) - Q',
        'Doha Diamond League 2024', 'World Athletics Championships
2023 (Budapest) - F', 'World Athletics Championships 2022 (Oregon) -
F',
```

```python
        'Indian Grand Prix 3 (Patiala)', 'Asian Games 2018 (Jakarta)',
'Zurich Diamond League Final 2022', 'ACNW League Meeting 1
(Potchefstroom)',
        'Federation Cup (Patiala)', 'World Athletics Championships
2023 (Budapest) - F', 'Lausanne Diamond League 2023',
        'Tokyo 2020 Olympics - F', 'Stockholm Diamond League 2022',
'Doha Diamond League 2018', 'Tokyo 2020 Olympics - F',
        'Zurich Diamond League Final 2022', 'Paavo Nurmi Games
(Turku)', 'Stockholm Diamond League 2022', 'Kuortane Games 2021',
        'Kuortane Games 2022', 'Stockholm Diamond League 2022', 'Tokyo
2020 Olympics - Q', 'Doha Diamond League 2023',
        'World U20 Championships 2016 (Bydgoszcz)', 'Commonwealth
Games 2018 (Gold Coast)'
    ],
    'Date': [
        'June 30, 2022', 'August 8, 2024', 'August 6, 2024', 'June 14,
2022', 'August 26, 2022', 'October 4, 2023',
        'August 25, 2023', 'May 5, 2023', 'September 8, 2022', 'July
21, 2022', 'May 10, 2024', 'August 27, 2023',
        'July 23, 2022', 'March 5, 2021', 'August 27, 2018',
'September 8, 2022', 'January 28, 2020', 'March 17, 2021',
        'August 27, 2023', 'June 30, 2023', 'August 7, 2021', 'June
30, 2022', 'May 4, 2018', 'August 7, 2021',
        'September 8, 2022', 'June 14, 2022', 'June 30, 2022', 'June
26, 2021', 'June 18, 2022', 'June 30, 2022',
        'August 4, 2021', 'May 5, 2023', 'July 23, 2016', 'April 14,
2018'
    ]
}

df = pd.DataFrame(data)

df

    Rank  Throw                                        Competition  \
0      1  89.94m                    Stockholm Diamond League 2022
1      2  89.45m                            Paris 2024 Olympics - F
2      3  89.34m                            Paris 2024 Olympics - Q
3      4  89.30m                          Paavo Nurmi Games (Turku)
4      5  89.08m                      Lausanne Diamond League 2022
5      6  88.88m                         Asian Games 2023 (Hangzhou)
6      7  88.77m  World Athletics Championships 2023 (Budapest) - Q
7      8  88.67m                            Doha Diamond League 2023
8      9  88.44m                   Zurich Diamond League Final 2022
9     10  88.39m   World Athletics Championships 2022 (Oregon) - Q
10    11  88.36m                            Doha Diamond League 2024
11    12  88.17m  World Athletics Championships 2023 (Budapest) - F
12    13  88.13m    World Athletics Championships 2022 (Oregon) - F
13    14  88.07m                          Indian Grand Prix 3 (Patiala)
14    15  88.06m                         Asian Games 2018 (Jakarta)
```

```
15    16    88.00m                        Zurich Diamond League Final 2022
16    17    87.86m            ACNW League Meeting 1 (Potchefstroom)
17    18    87.80m                            Federation Cup (Patiala)
18    19    87.73m    World Athletics Championships 2023 (Budapest) - F
19    20    87.66m                       Lausanne Diamond League 2023
20    21    87.58m                            Tokyo 2020 Olympics - F
21    22    87.46m                    Stockholm Diamond League 2022
22    23    87.43m                         Doha Diamond League 2018
23    24    87.03m                            Tokyo 2020 Olympics - F
24    25    87.00m                    Zurich Diamond League Final 2022
25    26    86.92m                         Paavo Nurmi Games (Turku)
26    27    86.84m                    Stockholm Diamond League 2022
27    28    86.79m                               Kuortane Games 2021
28    29    86.69m                               Kuortane Games 2022
29    30    86.67m                    Stockholm Diamond League 2022
30    31    86.65m                            Tokyo 2020 Olympics - Q
31    32    86.52m                         Doha Diamond League 2023
32    33    86.48m        World U20 Championships 2016 (Bydgoszcz)
33    34    86.47m            Commonwealth Games 2018 (Gold Coast)

                        Date
0            June 30, 2022
1           August 8, 2024
2           August 6, 2024
3            June 14, 2022
4          August 26, 2022
5          October 4, 2023
6          August 25, 2023
7               May 5, 2023
8        September 8, 2022
9             July 21, 2022
10             May 10, 2024
11         August 27, 2023
12            July 23, 2022
13            March 5, 2021
14         August 27, 2018
15       September 8, 2022
16        January 28, 2020
17           March 17, 2021
18         August 27, 2023
19            June 30, 2023
20           August 7, 2021
21            June 30, 2022
22               May 4, 2018
23           August 7, 2021
24       September 8, 2022
25            June 14, 2022
26            June 30, 2022
27            June 26, 2021
```

```
28        June 18, 2022
29        June 30, 2022
30      August 4, 2021
31         May 5, 2023
32        July 23, 2016
33       April 14, 2018
```

```python
df = df.drop('Rank', axis = 1)
```

```python
df.shape
```

```
(34, 3)
```

```python
df.columns
```

```
Index(['Throw', 'Competition', 'Date'], dtype='object')
```

```python
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

import warnings
warnings.filterwarnings('ignore')
```

```python
df['Throw'] = df['Throw'].str.replace('m', '').astype(float)
```

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34 entries, 0 to 33
Data columns (total 3 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Throw        34 non-null     float64
 1   Competition  34 non-null     object
 2   Date         34 non-null     object
dtypes: float64(1), object(2)
memory usage: 944.0+ bytes
```
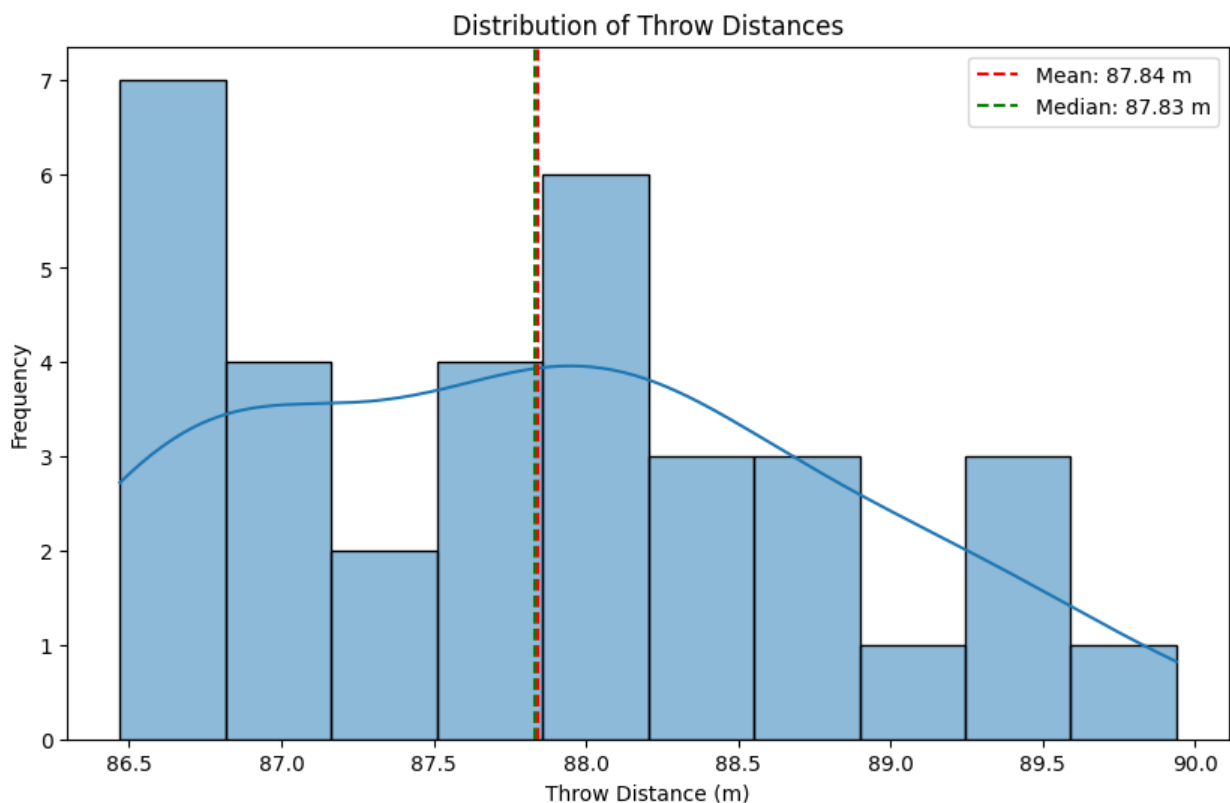
```python
df.describe()
```

```
           Throw
count  34.000000
mean   87.842059
std     0.971144
min    86.470000
25%    86.940000
50%    87.830000
75%    88.427500
max    89.940000
```

```python
plt.figure(figsize=(10, 6))
sns.histplot(df['Throw'], bins=10, kde=True)
plt.title('Distribution of Throw Distances')
plt.xlabel('Throw Distance (m)')
plt.ylabel('Frequency')

mean_throw = df['Throw'].mean()
median_throw = df['Throw'].median()
plt.axvline(mean_throw, color='r', linestyle='--', label=f'Mean:
{mean_throw:.2f} m')
plt.axvline(median_throw, color='g', linestyle='--', label=f'Median:
{median_throw:.2f} m')
plt.legend()

plt.show()
```



```python
plt.figure(figsize=(14, 8))
ax = sns.countplot(y='Competition', data=df,
order=df['Competition'].value_counts().index)
plt.title('Number of Top Throws by Competition')
plt.xlabel('Number of Throws')
plt.ylabel('Competition')

for p in ax.patches:
```
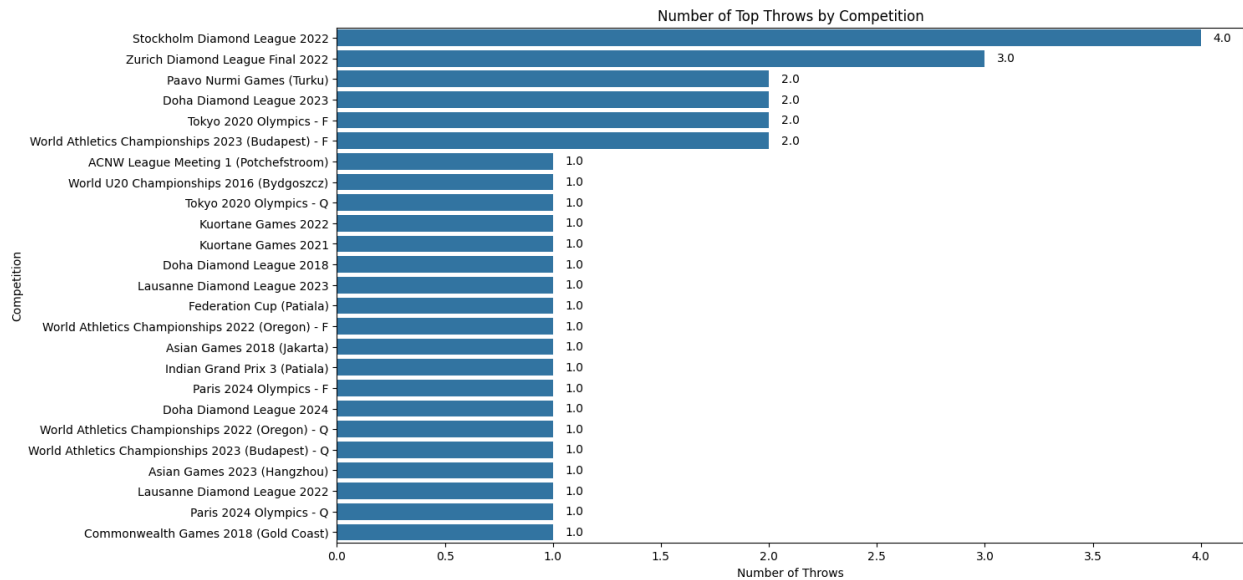
```
    ax.annotate(f'{p.get_width()}', (p.get_width() + 0.1, p.get_y() +
p.get_height() / 2),
                ha='center', va='center', fontsize=10, color='black')

plt.show()
```



```
df
```

```
    Throw                                    Competition  \
0   89.94                    Stockholm Diamond League 2022
1   89.45                         Paris 2024 Olympics - F
2   89.34                         Paris 2024 Olympics - Q
3   89.30                     Paavo Nurmi Games (Turku)
4   89.08                     Lausanne Diamond League 2022
5   88.88                     Asian Games 2023 (Hangzhou)
6   88.77   World Athletics Championships 2023 (Budapest) - Q
7   88.67                         Doha Diamond League 2023
8   88.44                 Zurich Diamond League Final 2022
9   88.39    World Athletics Championships 2022 (Oregon) - Q
10  88.36                         Doha Diamond League 2024
11  88.17   World Athletics Championships 2023 (Budapest) - F
12  88.13    World Athletics Championships 2022 (Oregon) - F
13  88.07                     Indian Grand Prix 3 (Patiala)
14  88.06                       Asian Games 2018 (Jakarta)
15  88.00                 Zurich Diamond League Final 2022
16  87.86             ACNW League Meeting 1 (Potchefstroom)
17  87.80                         Federation Cup (Patiala)
18  87.73   World Athletics Championships 2023 (Budapest) - F
19  87.66                     Lausanne Diamond League 2023
20  87.58                         Tokyo 2020 Olympics - F
21  87.46                    Stockholm Diamond League 2022
```

```
22  87.43                              Doha Diamond League 2018
23  87.03                              Tokyo 2020 Olympics - F
24  87.00                   Zurich Diamond League Final 2022
25  86.92                          Paavo Nurmi Games (Turku)
26  86.84                   Stockholm Diamond League 2022
27  86.79                                 Kuortane Games 2021
28  86.69                                 Kuortane Games 2022
29  86.67                   Stockholm Diamond League 2022
30  86.65                              Tokyo 2020 Olympics - Q
31  86.52                              Doha Diamond League 2023
32  86.48          World U20 Championships 2016 (Bydgoszcz)
33  86.47                  Commonwealth Games 2018 (Gold Coast)

                    Date
0          June 30, 2022
1         August 8, 2024
2         August 6, 2024
3          June 14, 2022
4        August 26, 2022
5        October 4, 2023
6        August 25, 2023
7             May 5, 2023
8      September 8, 2022
9           July 21, 2022
10          May 10, 2024
11       August 27, 2023
12          July 23, 2022
13         March 5, 2021
14       August 27, 2018
15     September 8, 2022
16      January 28, 2020
17        March 17, 2021
18       August 27, 2023
19         June 30, 2023
20        August 7, 2021
21         June 30, 2022
22           May 4, 2018
23        August 7, 2021
24     September 8, 2022
25         June 14, 2022
26         June 30, 2022
27         June 26, 2021
28         June 18, 2022
29         June 30, 2022
30        August 4, 2021
31            May 5, 2023
32         July 23, 2016
33        April 14, 2018
```
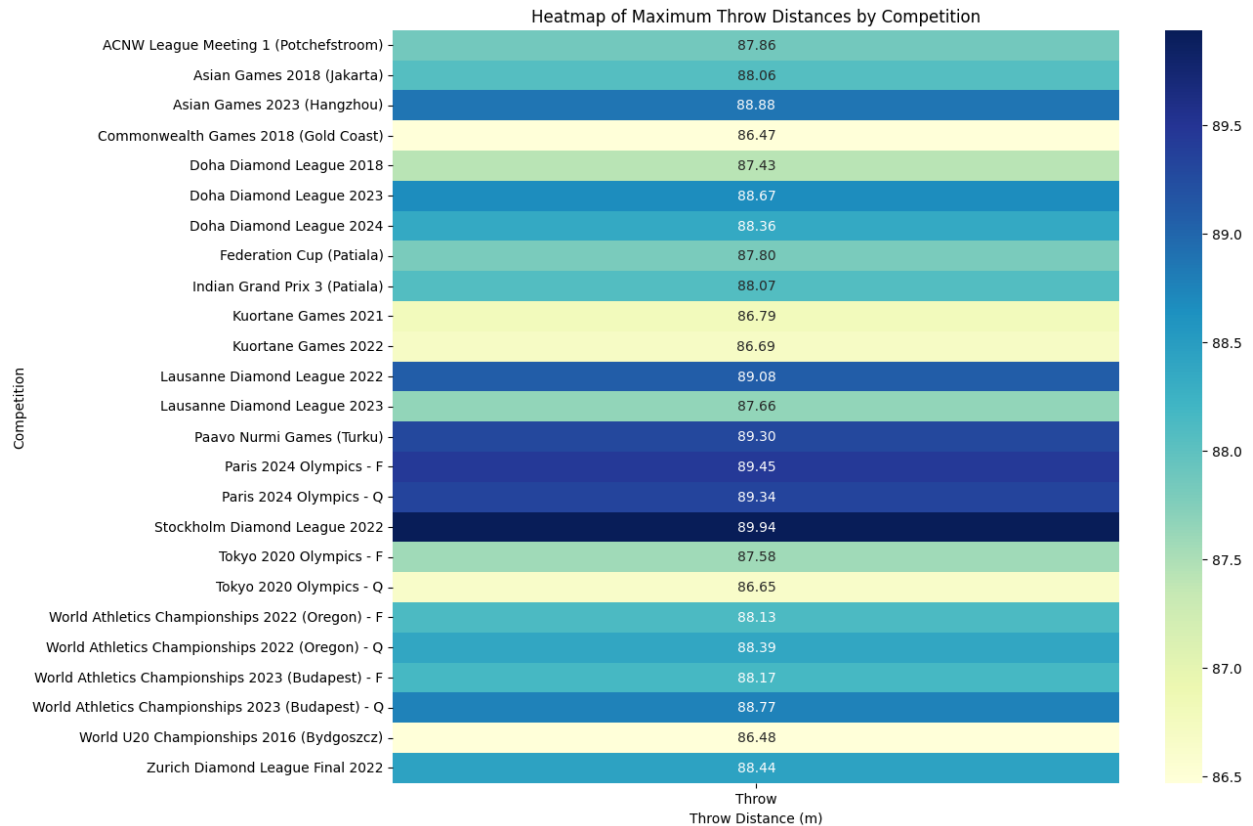
```
df['Date'] = pd.to_datetime(df['Date'], format='%B %d, %Y')
df['Date'] = df['Date'].dt.strftime('%Y-%m-%d')

df
```

	Throw	Competition	Date
0	89.94	Stockholm Diamond League 2022	2022-06-30
1	89.45	Paris 2024 Olympics - F	2024-08-08
2	89.34	Paris 2024 Olympics - Q	2024-08-06
3	89.30	Paavo Nurmi Games (Turku)	2022-06-14
4	89.08	Lausanne Diamond League 2022	2022-08-26
5	88.88	Asian Games 2023 (Hangzhou)	2023-10-04
6	88.77	World Athletics Championships 2023 (Budapest) - Q	2023-08-25
7	88.67	Doha Diamond League 2023	2023-05-05
8	88.44	Zurich Diamond League Final 2022	2022-09-08
9	88.39	World Athletics Championships 2022 (Oregon) - Q	2022-07-21
10	88.36	Doha Diamond League 2024	2024-05-10
11	88.17	World Athletics Championships 2023 (Budapest) - F	2023-08-27
12	88.13	World Athletics Championships 2022 (Oregon) - F	2022-07-23
13	88.07	Indian Grand Prix 3 (Patiala)	2021-03-05
14	88.06	Asian Games 2018 (Jakarta)	2018-08-27
15	88.00	Zurich Diamond League Final 2022	2022-09-08
16	87.86	ACNW League Meeting 1 (Potchefstroom)	2020-01-28
17	87.80	Federation Cup (Patiala)	2021-03-17
18	87.73	World Athletics Championships 2023 (Budapest) - F	2023-08-27
19	87.66	Lausanne Diamond League 2023	2023-06-30
20	87.58	Tokyo 2020 Olympics - F	2021-08-07
21	87.46	Stockholm Diamond League 2022	2022-06-

```
30
22  87.43                                    Doha Diamond League 2018    2018-05-
04
23  87.03                                    Tokyo 2020 Olympics - F    2021-08-
07
24  87.00                          Zurich Diamond League Final 2022    2022-09-
08
25  86.92                                 Paavo Nurmi Games (Turku)    2022-06-
14
26  86.84                               Stockholm Diamond League 2022    2022-06-
30
27  86.79                                        Kuortane Games 2021    2021-06-
26
28  86.69                                        Kuortane Games 2022    2022-06-
18
29  86.67                               Stockholm Diamond League 2022    2022-06-
30
30  86.65                                    Tokyo 2020 Olympics - Q    2021-08-
04
31  86.52                                    Doha Diamond League 2023    2023-05-
05
32  86.48                     World U20 Championships 2016 (Bydgoszcz)    2016-07-
23
33  86.47                          Commonwealth Games 2018 (Gold Coast)    2018-04-
14

plt.figure(figsize=(12, 10))
pivot_table = df.pivot_table(values='Throw', index='Competition',
aggfunc='max')
sns.heatmap(pivot_table, annot=True, fmt='.2f', cmap='YlGnBu',
cbar=True)
plt.title('Heatmap of Maximum Throw Distances by Competition')
plt.xlabel('Throw Distance (m)')
plt.ylabel('Competition')
plt.show()
```

Heatmap of Maximum Throw Distances by Competition

Competition	Throw Distance (m)
ACNW League Meeting 1 (Potchefstroom)	87.86
Asian Games 2018 (Jakarta)	88.06
Asian Games 2023 (Hangzhou)	88.88
Commonwealth Games 2018 (Gold Coast)	86.47
Doha Diamond League 2018	87.43
Doha Diamond League 2023	88.67
Doha Diamond League 2024	88.36
Federation Cup (Patiala)	87.80
Indian Grand Prix 3 (Patiala)	88.07
Kuortane Games 2021	86.79
Kuortane Games 2022	86.69
Lausanne Diamond League 2022	89.08
Lausanne Diamond League 2023	87.66
Paavo Nurmi Games (Turku)	89.30
Paris 2024 Olympics - F	89.45
Paris 2024 Olympics - Q	89.34
Stockholm Diamond League 2022	89.94
Tokyo 2020 Olympics - F	87.58
Tokyo 2020 Olympics - Q	86.65
World Athletics Championships 2022 (Oregon) - F	88.13
World Athletics Championships 2022 (Oregon) - Q	88.39
World Athletics Championships 2023 (Budapest) - F	88.17
World Athletics Championships 2023 (Budapest) - Q	88.77
World U20 Championships 2016 (Bydgoszcz)	86.48
Zurich Diamond League Final 2022	88.44

Throw

```python
highest_throw = df.loc[df['Throw'].idxmax()]
lowest_throw = df.loc[df['Throw'].idxmin()]

print("Highest Throw:")
print(f"Throw: {highest_throw['Throw']} m")
print(f"Competition: {highest_throw['Competition']}")
print(f"Date: {highest_throw['Date']}\n")

print("Lowest Throw:")
print(f"Throw: {lowest_throw['Throw']} m")
print(f"Competition: {lowest_throw['Competition']}")
print(f"Date: {lowest_throw['Date']}")
```

```
Highest Throw:
Throw: 89.94 m
Competition: Stockholm Diamond League 2022
Date: 2022-06-30

Lowest Throw:
Throw: 86.47 m
Competition: Commonwealth Games 2018 (Gold Coast)
Date: 2018-04-14
```

```python
from statsmodels.tsa.arima.model import ARIMA
```

```
df_sorted = df.sort_values(by='Date')

df_sorted.set_index('Date', inplace=True)

df_sorted
```

	Throw	Competition
Date		
2016-07-23	86.48	World U20 Championships 2016 (Bydgoszcz)
2018-04-14	86.47	Commonwealth Games 2018 (Gold Coast)
2018-05-04	87.43	Doha Diamond League 2018
2018-08-27	88.06	Asian Games 2018 (Jakarta)
2020-01-28	87.86	ACNW League Meeting 1 (Potchefstroom)
2021-03-05	88.07	Indian Grand Prix 3 (Patiala)
2021-03-17	87.80	Federation Cup (Patiala)
2021-06-26	86.79	Kuortane Games 2021
2021-08-04	86.65	Tokyo 2020 Olympics - Q
2021-08-07	87.03	Tokyo 2020 Olympics - F
2021-08-07	87.58	Tokyo 2020 Olympics - F
2022-06-14	86.92	Paavo Nurmi Games (Turku)
2022-06-14	89.30	Paavo Nurmi Games (Turku)
2022-06-18	86.69	Kuortane Games 2022
2022-06-30	86.84	Stockholm Diamond League 2022
2022-06-30	87.46	Stockholm Diamond League 2022
2022-06-30	89.94	Stockholm Diamond League 2022
2022-06-30	86.67	Stockholm Diamond League 2022
2022-07-21	88.39	World Athletics Championships 2022 (Oregon) - Q
2022-07-23	88.13	World Athletics Championships 2022 (Oregon) - F
2022-08-26	89.08	Lausanne Diamond League 2022
2022-09-08	88.44	Zurich Diamond League Final 2022
2022-09-08	87.00	Zurich Diamond League Final 2022
2022-09-08	88.00	Zurich Diamond League Final 2022
2023-05-05	88.67	Doha Diamond League 2023
2023-05-05	86.52	Doha Diamond League 2023
2023-06-30	87.66	Lausanne Diamond League 2023
2023-08-25	88.77	World Athletics Championships 2023 (Budapest) - Q
2023-08-27	87.73	World Athletics Championships 2023 (Budapest) - F
2023-08-27	88.17	World Athletics Championships 2023 (Budapest) - F
2023-10-04	88.88	Asian Games 2023 (Hangzhou)
2024-05-10	88.36	Doha Diamond League 2024
2024-08-06	89.34	Paris 2024 Olympics - Q
2024-08-08	89.45	Paris 2024 Olympics - F

```
df_sorted = df_sorted.drop('Competition', axis = 1)

df_sorted
```

	Throw
Date	
2016-07-23	86.48
2018-04-14	86.47

```
2018-05-04  87.43
2018-08-27  88.06
2020-01-28  87.86
2021-03-05  88.07
2021-03-17  87.80
2021-06-26  86.79
2021-08-04  86.65
2021-08-07  87.03
2021-08-07  87.58
2022-06-14  86.92
2022-06-14  89.30
2022-06-18  86.69
2022-06-30  86.84
2022-06-30  87.46
2022-06-30  89.94
2022-06-30  86.67
2022-07-21  88.39
2022-07-23  88.13
2022-08-26  89.08
2022-09-08  88.44
2022-09-08  87.00
2022-09-08  88.00
2023-05-05  88.67
2023-05-05  86.52
2023-06-30  87.66
2023-08-25  88.77
2023-08-27  87.73
2023-08-27  88.17
2023-10-04  88.88
2024-05-10  88.36
2024-08-06  89.34
2024-08-08  89.45
```

```python
plt.figure(figsize=(20, 10))

plt.plot(df_sorted.index, df_sorted['Throw'], marker='o',
linestyle='-', label='Thows')
plt.xticks(rotation=45)

for x, y in zip(df_sorted.index, df_sorted['Throw']):
    plt.annotate(f'{y}', (x, y), textcoords="offset points",
xytext=(0, 5), ha='center')

plt.title('Line Plot with Annotations')
plt.xlabel('X-axis Label')
plt.ylabel('Y-axis Label')

plt.legend()

plt.show()
```
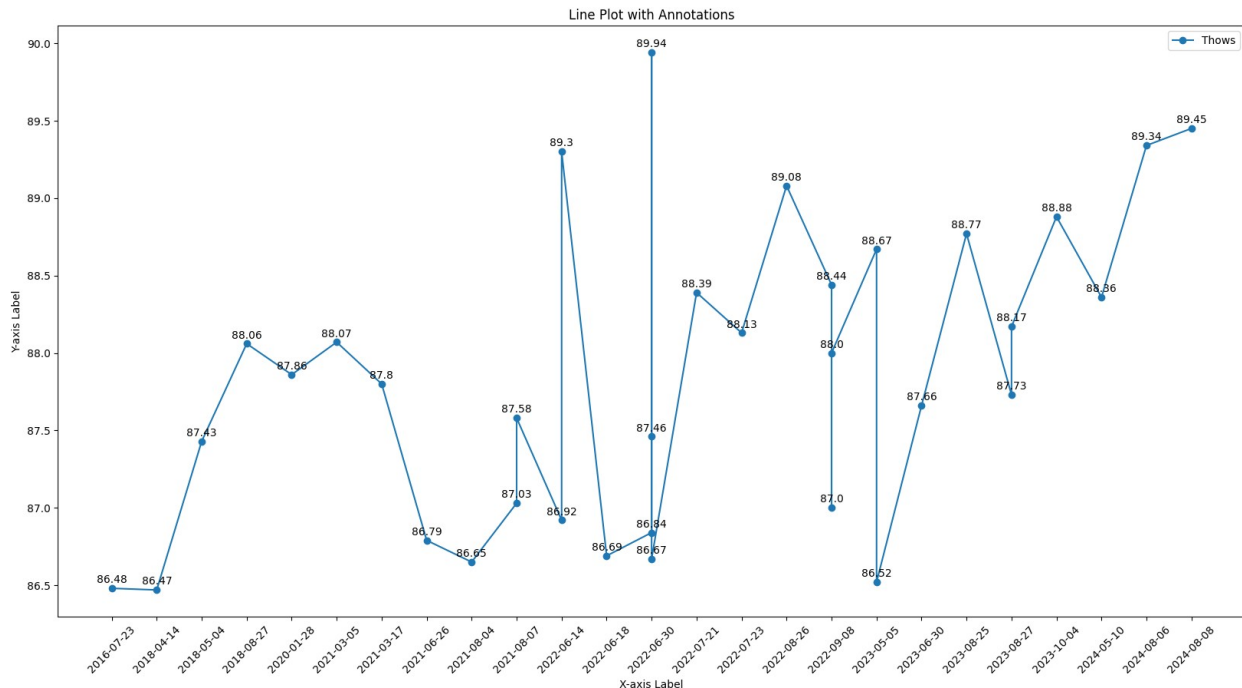
Line Plot with Annotations

```python
from sklearn.linear_model import LinearRegression
from datetime import datetime

df_sorted.index = pd.to_datetime(df_sorted.index)

df_sorted['Date_ordinal'] = df_sorted.index.map(datetime.toordinal)

X = df_sorted['Date_ordinal'].values.reshape(-1, 1)
y = df_sorted['Throw']

model = LinearRegression()
model.fit(X, y)

LinearRegression()

df_sorted['Predicted_Throw'] = model.predict(X)

plt.figure(figsize=(10, 6))
plt.plot(df_sorted.index, df_sorted['Throw'], label='Actual Throw',
marker='o')
plt.plot(df_sorted.index, df_sorted['Predicted_Throw'],
label='Predicted Throw', linestyle='--')
plt.xlabel('Date')
plt.ylabel('Throw')
plt.title('Linear Regression on Throw Data Over Time')
plt.legend()
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Linear Regression on Throw Data Over Time

```
print(df_sorted[['Throw', 'Predicted_Throw']])

             Throw   Predicted_Throw
Date
2016-07-23   86.48         86.502623
2018-04-14   86.47         86.917877
2018-05-04   87.43         86.931060
2018-08-27   88.06         87.006860
2020-01-28   87.86         87.348950
2021-03-05   88.07         87.613921
2021-03-17   87.80         87.621831
2021-06-26   86.79         87.688403
2021-08-04   86.65         87.714110
2021-08-07   87.03         87.716087
2021-08-07   87.58         87.716087
2022-06-14   86.92         87.921077
2022-06-14   89.30         87.921077
2022-06-18   86.69         87.923714
2022-06-30   86.84         87.931623
2022-06-30   87.46         87.931623
2022-06-30   89.94         87.931623
2022-06-30   86.67         87.931623
2022-07-21   88.39         87.945465
2022-07-23   88.13         87.946783
2022-08-26   89.08         87.969194
2022-09-08   88.44         87.977763
2022-09-08   87.00         87.977763
```

```
2022-09-08   88.00        87.977763
2023-05-05   88.67        88.135295
2023-05-05   86.52        88.135295
2023-06-30   87.66        88.172207
2023-08-25   88.77        88.209118
2023-08-27   87.73        88.210437
2023-08-27   88.17        88.210437
2023-10-04   88.88        88.235484
2024-05-10   88.36        88.379834
2024-08-06   89.34        88.437837
2024-08-08   89.45        88.439156
```

Thanks !!!