```
In [1]:    import numpy as np
           import pandas as pd
           import matplotlib.pyplot as plt
```

```
In [2]:    def plot_time_distribution(df, dt_name):
               grouped_data = df.groupby('year').size()
               # Plotting
               plt.figure(figsize=(10, 6))
               bars = plt.bar(grouped_data.index, grouped_data.values, color='skyblue')

               # Adding labels to each bar
               for bar in bars:
                   yval = bar.get_height()
                   plt.text(bar.get_x() + bar.get_width()/2.0, yval, int(yval), va='bottom', ha='center', fontsize=12)

               plt.title(f'{dt_name} Number Of Videos vs Year')
               plt.xlabel('Year')
               plt.ylabel('Number of Videos')
               plt.xticks(grouped_data.index)
               plt.grid(axis='y', linestyle='--', alpha=0.7)
               plt.tight_layout()
               plt.show()
```

Statistics for YouCook2

Train-Val Data

```
In [3]:    df_tv = pd.read_csv('DatasetsWithTime/youcook2trainval_withtime.csv')

           df_tv.head()
```

Out[3]:

	Unnamed: 0	database	year	month	day	time
0	--bv0V6ZjWI	{'duration': 332.46, 'subset': 'training', 're...	2014.0	6.0	28.0	22:38:49
1	-AfxeTnCbVQ	{'duration': 440.58, 'subset': 'training', 're...	2014.0	2.0	20.0	17:35:36
2	-AwyG1JcMp8	{'duration': 307.5, 'subset': 'validation', 'r...	2014.0	4.0	10.0	16:06:07
3	-CP9gKR9GR4	{'duration': 430.84, 'subset': 'training', 're...	NaN	NaN	NaN	NaN
4	-ErPSunMfcs	{'duration': 154.28, 'subset': 'validation', '...	2015.0	4.0	11.0	15:50:25

```
In [4]:    plot_time_distribution(df_tv, "YouCook2 TrainVal")
```

YouCook2 TrainVal Number Of Videos vs Year
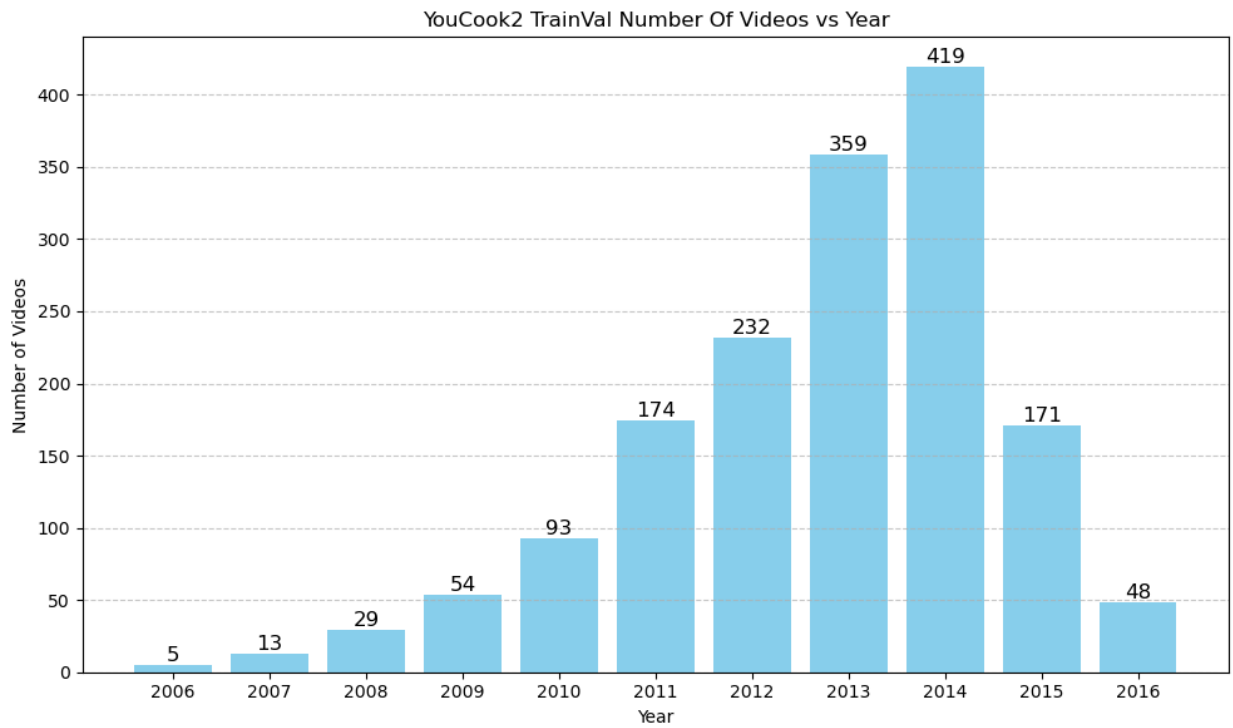
Test Data

```
In [5]:  df_t = pd.read_csv('DatasetsWithTime/youcook2test_withtime.csv')

         df_t.head()
```

Out[5]:

	vid	database	year	month	day	time
0	-rJ5CI0hXHA	{'subset': 'testing', 'duration': 138.7, 'anno...	2013.0	10.0	1.0	12:00:03
1	-w0WRDmDoG8	{'subset': 'testing', 'duration': 148.4, 'anno...	2011.0	2.0	7.0	12:30:10
2	06Nl8pAF2mo	{'subset': 'testing', 'duration': 235.6, 'anno...	2012.0	5.0	2.0	00:39:37
3	0LNyj8V_R1k	{'subset': 'testing', 'duration': 564.71, 'ann...	2014.0	8.0	5.0	11:05:37
4	0Q2k-EJCriQ	{'subset': 'testing', 'duration': 302.32, 'ann...	2013.0	3.0	16.0	10:00:26

```
In [6]:  plot_time_distribution(df_t, "YouCook2 Test")
```

YouCook2 Test Number Of Videos vs Year
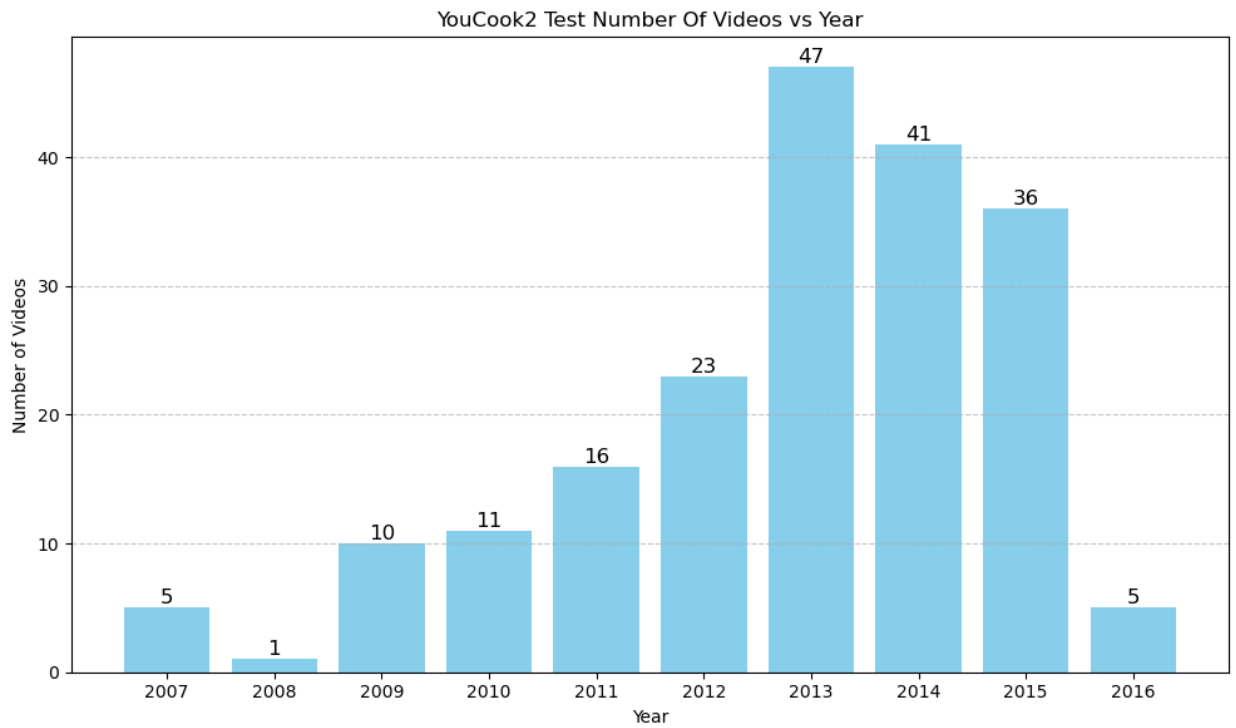
Total Data

```
In [7]: df_yc2total = pd.concat([df_tv, df_t], axis=0, ignore_index=True)

        plot_time_distribution(df_yc2total, "YouCook2 Total")
```


YouCook2 Total Number Of Videos vs Year
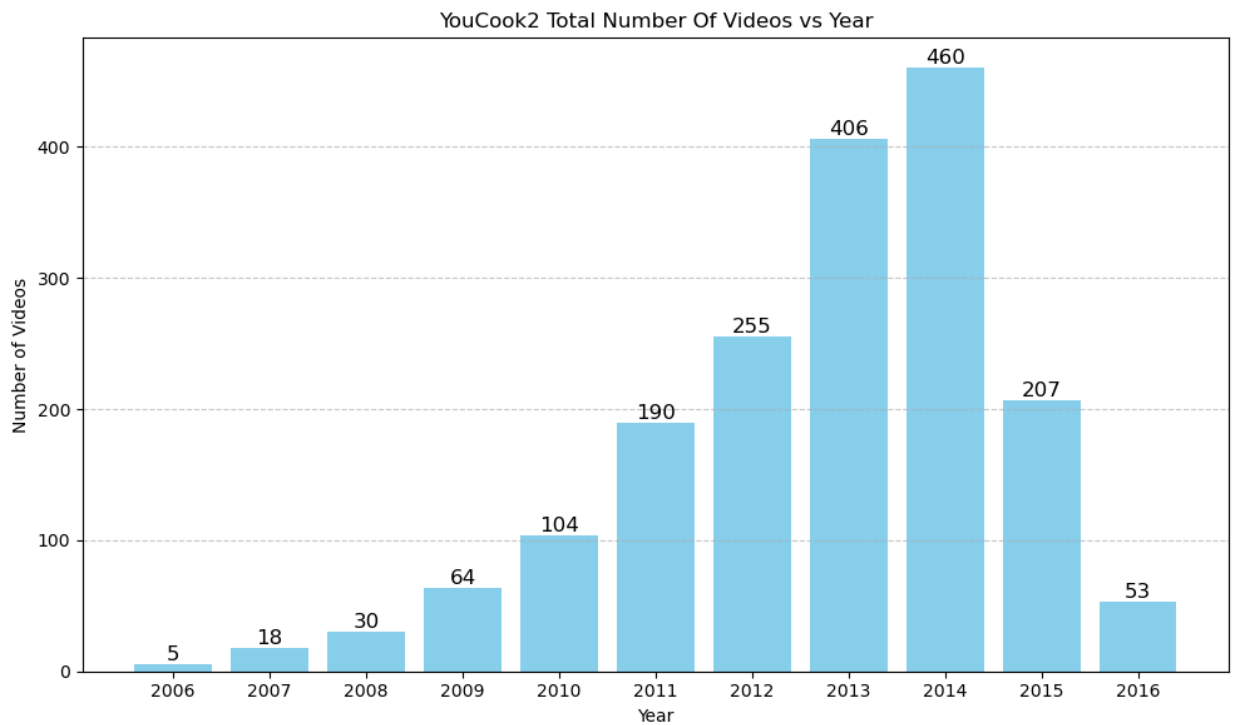
Statistics for HowTo100M

```
In [8]: df_htm = pd.read_csv('DatasetsWithTime/howto100m_withtime.csv')
```

```
df_htm.head()
```

Out[8]:

	Unnamed: 0	video_id	category_1	category_2	rank	task_id	year	month	day	ti
0	0	nVbIUDjzWY4	Cars & Other Vehicles	Motorcycles	27	52907	2015.0	6.0	26.0	00:53
1	1	CTPAZ2euJ2Q	Cars & Other Vehicles	Motorcycles	35	109057	NaN	NaN	NaN	N
2	2	rwmt7Cbuvfs	Cars & Other Vehicles	Motorcycles	99	52907	2017.0	5.0	6.0	01:50
3	3	HnTLh99gcxY	Cars & Other Vehicles	Motorcycles	35	52907	2014.0	9.0	7.0	23:32
4	4	EyP3HVhg1u0	Cars & Other Vehicles	Motorcycles	95	52906	2012.0	6.0	3.0	01:05

In [9]:
```
plot_time_distribution(df_htm, "HowTo100M")
```
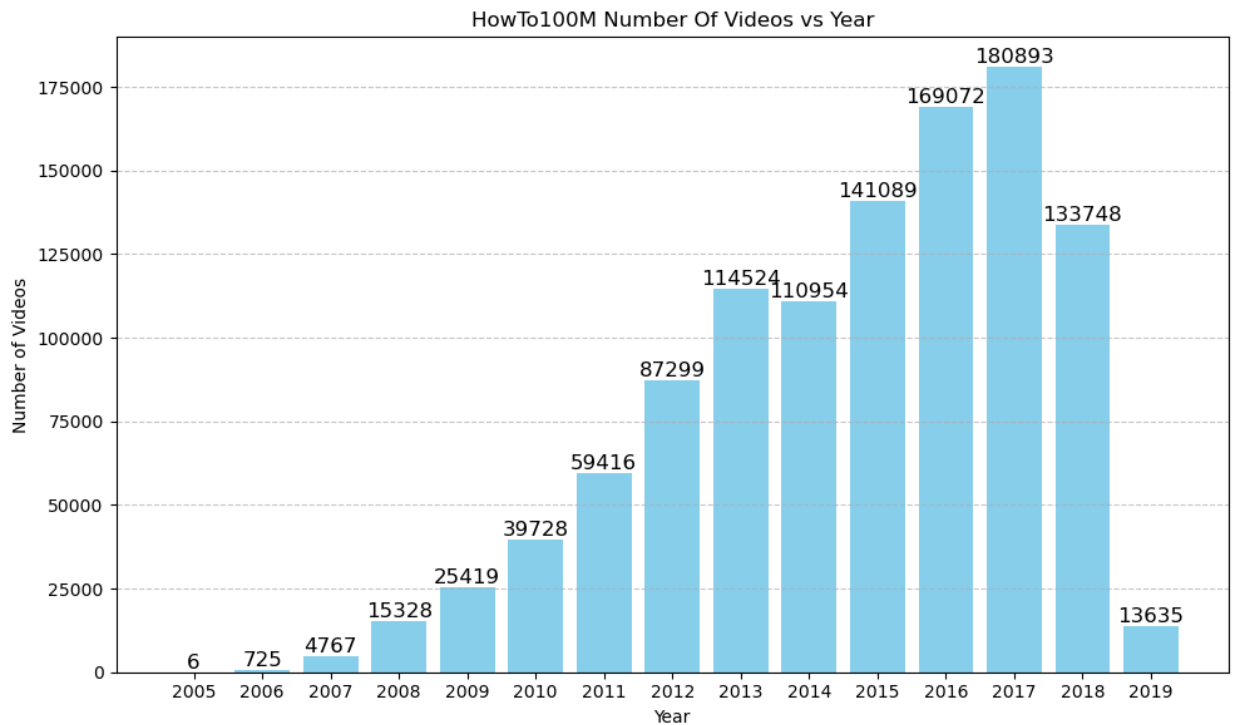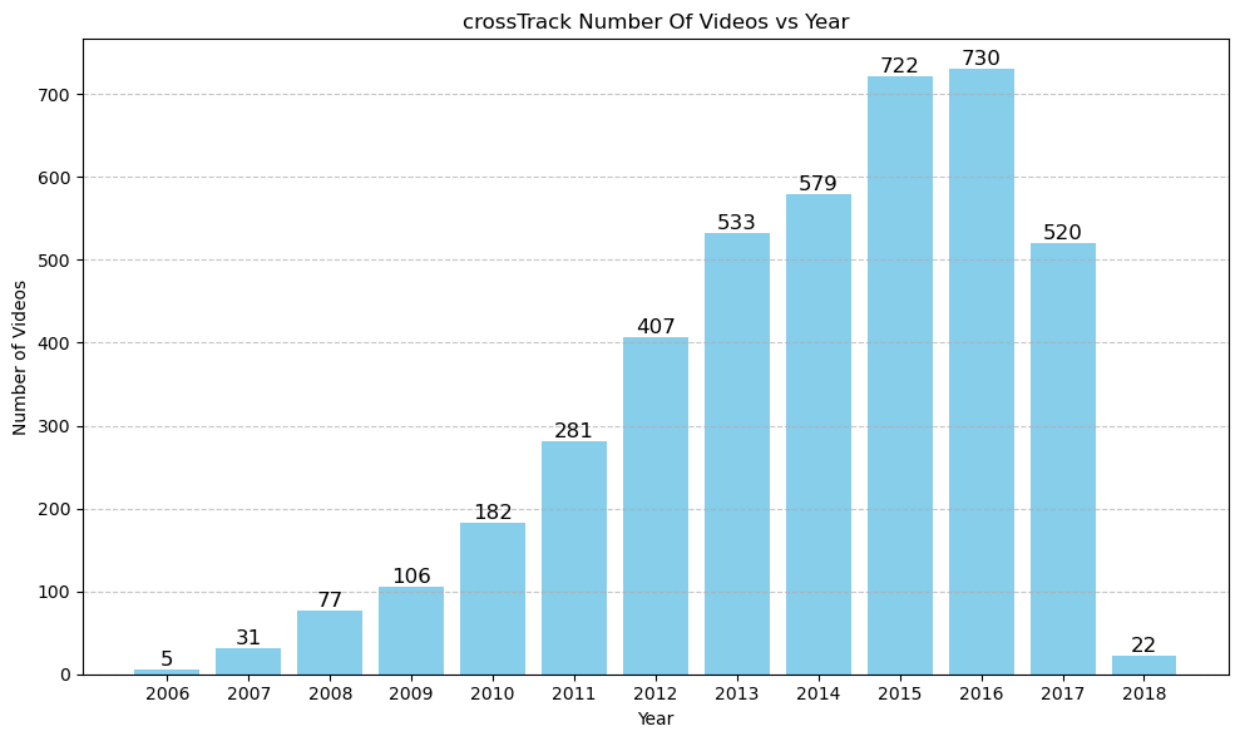

Statistics for crossTrack

CrossTracK

In [10]:
```
df_ct = pd.read_csv('DatasetsWithTime/crossTrack_withtime.csv')

df_ct.head()
```

Out[10]:

	Unnamed: 0.1	Unnamed: 0	id	vid	url	year	month
0	0	0	41950	UYTGe1ixPcc	https://www.youtube.com/watch?v=UYTGe1ixPcc	2016.0	2.0
1	1	1	41718	pd3bErSrLio	https://www.youtube.com/watch?v=pd3bErSrLio	2012.0	5.0
2	2	2	41950	4_qZdxTeqvU	https://www.youtube.com/watch?v=4_qZdxTeqvU	NaN	NaN
3	3	3	40610	17DCH80pzRk	https://www.youtube.com/watch?v=17DCH80pzRk	2015.0	4.0
4	4	4	91515	BanqwWgLqt8	https://www.youtube.com/watch?v=BanqwWgLqt8	2015.0	2.0

In [11]: `plot_time_distribution(df_ct, "crossTrack")`

Val Data

In [12]:
```
df_ctval = pd.read_csv('DatasetsWithTime/crossTrackVal_withtime.csv')

df_ctval.head()
```

`Out[12]:`

	Unnamed: 0.1	Unnamed: 0	id	vid	url	year	month
0	0	0	105222	rzxVluau83Q	https://www.youtube.com/watch?v=rzxVluau83Q	2017.0	3.0
1	1	1	105222	2uDelqlNWaE	https://www.youtube.com/watch?v=2uDelqlNWaE	NaN	NaN
2	2	2	105222	xDMGRE3YZFY	https://www.youtube.com/watch?v=xDMGRE3YZFY	NaN	NaN
3	3	3	105222	0mnuAqJCg5c	https://www.youtube.com/watch?v=0mnuAqJCg5c	2013.0	7.0
4	4	4	105222	W3-Nf-QgfaI	https://www.youtube.com/watch?v=W3-Nf-QgfaI	2015.0	2.0

`In [13]:` `plot_time_distribution(df_ctval, "crossTrackVal")`
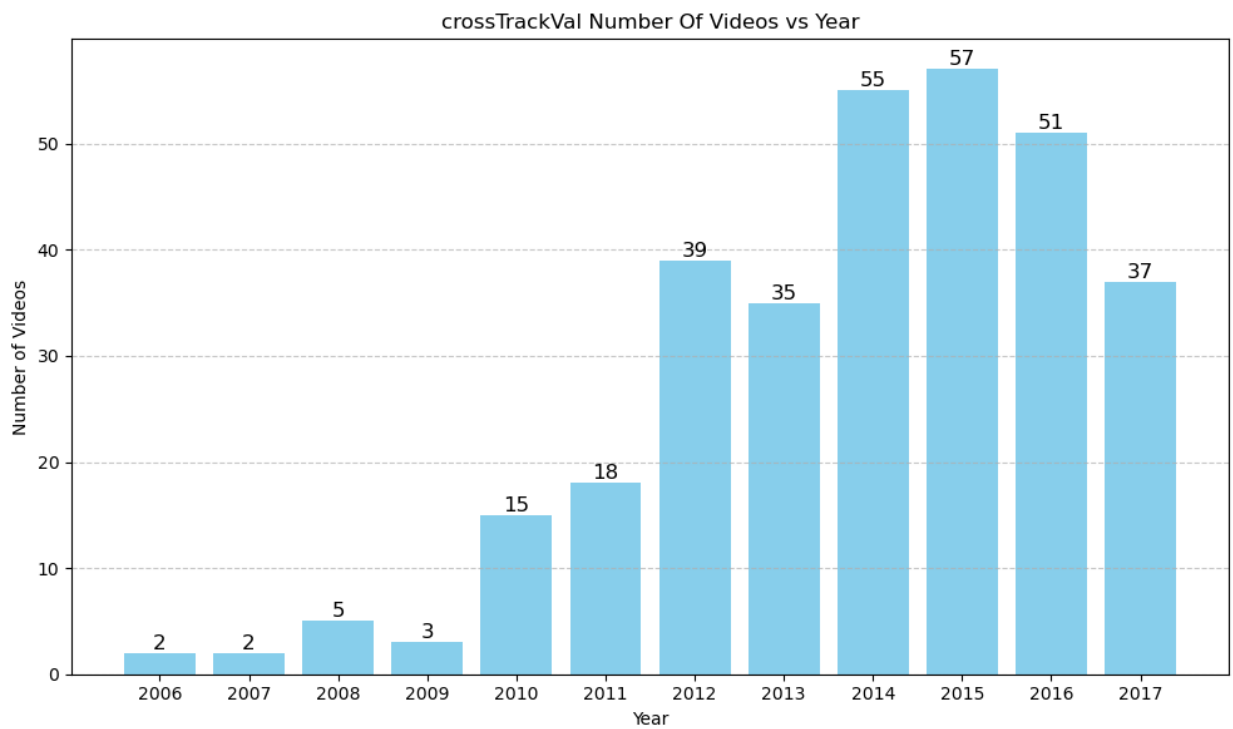
Statictics for COIN

`In [14]:`
```
df_c = pd.read_csv('DatasetsWithTime/coin_withtime.csv')

df_c.head()
```

Out[14]:

	Unnamed: 0	database	year	month	day	time
0	-0X2mXPy3Mc	{'recipe_type': 32, 'annotation': [{'id': '603...	2016.0	9.0	6.0	19:34:19
1	-0bbAfOuh-M	{'recipe_type': 78, 'annotation': [{'id': '352...	2016.0	6.0	5.0	13:54:06
2	-13Rr4vdydc	{'recipe_type': 62, 'annotation': [{'id': '426...	2013.0	10.0	2.0	02:28:10
3	-1_Gpwr9Npg	{'recipe_type': 124, 'annotation': [{'id': '48...	2016.0	5.0	12.0	14:56:14
4	-1cUDe-mecw	{'recipe_type': 9, 'annotation': [{'id': '155'...	2017.0	7.0	4.0	03:00:37

In [15]: plot_time_distribution(df_c, "COINI")