

```
import pandas as pd
```

```
data=pd.read_csv('/content/Most popular 1000 Youtube videos - Copy(prac).csv')
```

```
data.head()
```

	rank	Video	Video views	Likes	Dislikes	Category	published
0	1	Lil Nas X - Old Town Road (Official Movie) ft....	5,40,71,677	34,97,955	78,799	Music	2019
1	2	20 Tennis shots if they were not filmed, NOBOD...	34,71,237	19,023	859	NaN	2017
2	3	JoJo Siwa - Karma (Official Video)	3,42,06,747	2,93,563	NaN	Music	2024
3	4	David Kushner - Daylight (Official Music Video)	1,85,58,390	6,80,732	NaN	Music	2023
4	5	Wiz Khalifa - See You Again ft. Charlie Puth [...	6,54,79,81,039	4,44,28,537	NaN	Music	2015

Next steps:

[Generate code with data](#)[View recommended plots](#)[New interactive sheet](#)

```
data.isnull().sum() #checking for null values
```

```
rank    0
Video    0
Video views  0
Likes    0
Dislikes  473
Category  18
published  0
```

dtype: int64

```
data['Dislikes']=pd.to_numeric(data['Dislikes'],errors='coerce') # converting data type
```

```
data['Dislikes'].fillna(data['Dislikes'].mean(),inplace=True) # replacing all null values with the mean data of that column
```

<ipython-input-5-bddb5517eb71>:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment. The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or 'df[col] = df[col].method(value, inplace=True)'

```
data['Dislikes'].fillna(data['Dislikes'].mean(),inplace=True) # replacing all null values with the mean data of that column
```

```
data.isnull().sum() #checking null values
```

```
rank    0
Video    0
Video views  0
Likes    0
Dislikes    0
Category  18
published  0
```

dtype: int64

```
data['rank'].dtypes
```

```
dtype('int64')
```

```
data['Video'].dtypes
```

```
dtype('O')
```

```
data['Video'].dtypes
```

```
dtype('O')
```

```
data['Category'].head(10)
```

```
dtype: object
```

	Category
0	Music
1	NaN
2	Music
3	Music
4	Music
5	Music
6	Music
7	Music
8	Music
9	Music

```
data['Category'].dtypes
```

```
dtype('O')
```

```
data=data.dropna(subset=['Category']) #remove rows where category column has missing values
```

```
data['Category'].head(10)
```

```
dtype: object
```

	Category
0	Music
2	Music
3	Music
4	Music
5	Music
6	Music
7	Music
8	Music
9	Music
10	Entertainment

```
len(data)
```

```
982
```

```
data.isnull().sum()
```



	0
rank	0
Video	0
Video views	0
Likes	0
Dislikes	0
Category	0
published	0

dtype: int64

```
data.duplicated().sum()
```



```
np.int64(0)
```

```
data.dtypes
```



	0
rank	int64
Video	object
Video views	object
Likes	object
Dislikes	float64
Category	object
published	int64

dtype: object

```
data['Dislikes']=data['Dislikes'].astype(int)      #astype(dataType): used to convert data type
```



<ipython-input-18-09b2128b8d53>:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-data](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-data)  
data['Dislikes']=data['Dislikes'].astype(int) #astype(dataType): used to convert data type

```
data.dtypes
```



	0
rank	int64
Video	object
Video views	object
Likes	object
Dislikes	int64
Category	object
published	int64

dtype: object

```
data.describe()
```

	rank	Dislikes	published
count	982.000000	982.000000	982.000000
mean	500.858452	329.997963	2019.155804
std	289.200196	146.759479	5.358804
min	1.000000	0.000000	2005.000000
25%	248.250000	331.000000	2017.000000
50%	502.500000	331.000000	2021.000000
75%	750.750000	331.000000	2024.000000
max	1000.000000	990.000000	2025.000000

data.head()

	rank	Video	Video views	Likes	Dislikes	Category	published
0	1	Lil Nas X - Old Town Road (Official Movie) ft....	5,40,71,677	34,97,955	331	Music	2019
2	3	JoJo Siwa - Karma (Official Video)	3,42,06,747	2,93,563	331	Music	2024
3	4	David Kushner - Daylight (Official Music Video)	1,85,58,390	6,80,732	331	Music	2023
4	5	Wiz Khalifa - See You Again ft. Charlie Puth [...	6,54,79,81,039	4,44,28,537	331	Music	2015
5	6	JP Saxe - If the World Was Ending (Official Vi...	7,68,34,495	8,04,353	331	Music	2019

Next steps: [Generate code with data](#) [View recommended plots](#) [New interactive sheet](#)

```
years=data['published'].unique()
print(years)
```

[2019 2024 2023 2015 2025 2020 2018 2021 2017 2022 2011 2007 2009 2016  
2006 2012 2010 2008 2014 2013 2005]

```
data_backup=data.copy() #creating backup of original data
print(len(data))
```

982

```
df = pd.DataFrame(published_songs_by_year)
```

```
print(df)
```

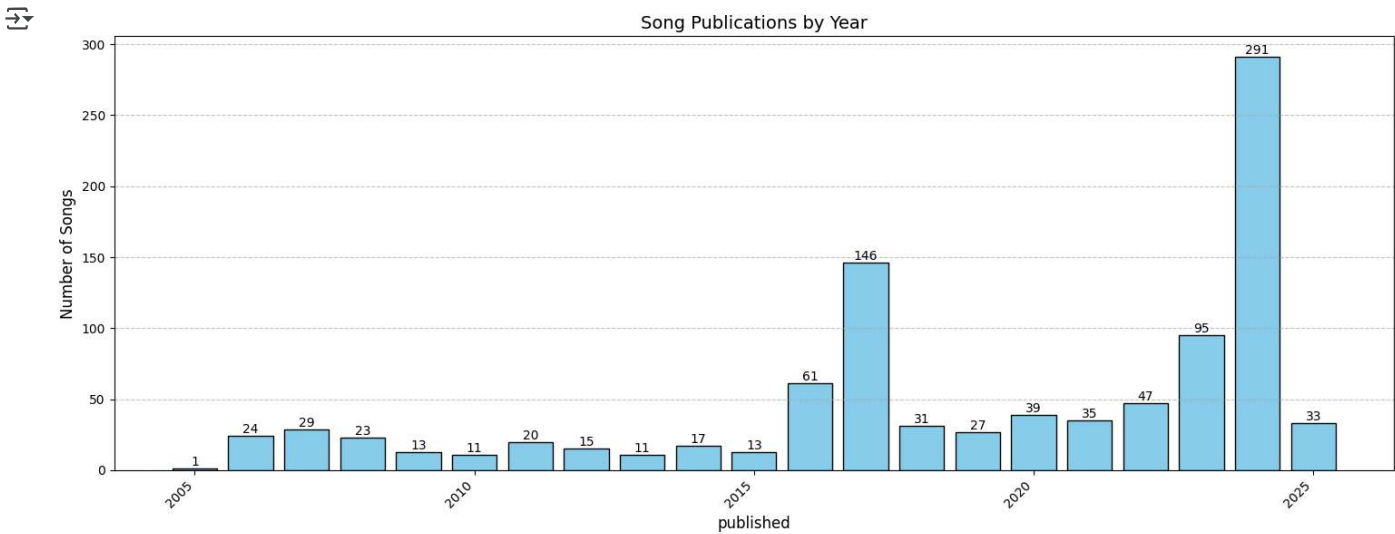
	published	Video
0	2005	1
1	2006	24
2	2007	29
3	2008	23
4	2009	13
5	2010	11
6	2011	20
7	2012	15
8	2013	11
9	2014	17
10	2015	13
11	2016	61
12	2017	146
13	2018	31
14	2019	27
15	2020	39
16	2021	35
17	2022	47
18	2023	95
19	2024	291
20	2025	33

```
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(15, 6))
bars = plt.bar(df['published'], df['Video'], color='skyblue', edgecolor='black')
plt.title('Song Publications by Year', fontsize=14)
plt.xlabel('published', fontsize=12)
plt.ylabel('Number of Songs', fontsize=12)
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

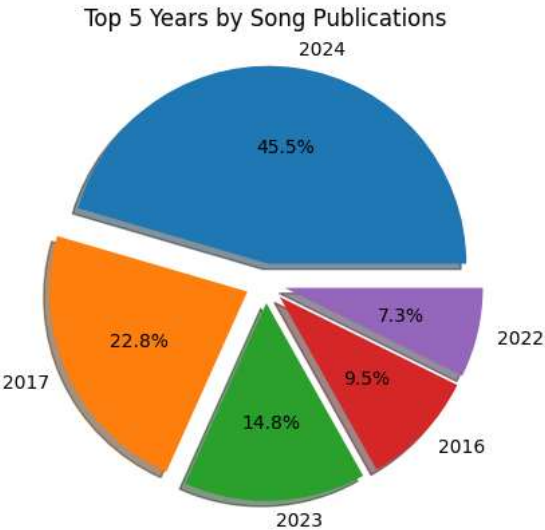
```
# Add value labels
for bar in bars:
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2., height,
             f'{int(height)}', ha='center', va='bottom')

plt.tight_layout()
plt.show()
```



```
top5 = df.nlargest(5, 'Video')
plt.pie(top5['Video'], labels=top5['published'], autopct='%1.1f%%',
        explode=[0.1]*5, shadow=True)
plt.title('Top 5 Years by Song Publications')
```

```
Text(0.5, 1.0, 'Top 5 Years by Song Publications')
```






```
data.head()
```

	rank	Video	Video views	Likes	Dislikes	Category	published
0	1	Lil Nas X - Old Town Road (Official Movie) ft....	5,40,71,677	34,97,955	331	Music	2019
2	3	JoJo Siwa - Karma (Official Video)	3,42,06,747	2,93,563	331	Music	2024
3	4	David Kushner - Daylight (Official Music Video)	1,85,58,390	6,80,732	331	Music	2023
4	5	Wiz Khalifa - See You Again ft. Charlie Puth [...]	6,54,79,81,039	4,44,28,537	331	Music	2015
5	6	JP Saxe - If the World Was Ending (Official Vi...	7,68,34,495	8,04,353	331	Music	2019

Next steps: [Generate code with data](#) [View recommended plots](#) [New interactive sheet](#)

```
data[['rank', 'Video']].tail()
```



	rank	Video	
995	996	New Champ Kayn/Rhaast Leak for LOL (Moobeat cr...	
996	997	Ford Mustang Launch (street)	
997	998	Eminem is gay - The Interview	
998	999	Yakuza OST - Baka Mitai (ばかみたい) Kiryu full ver...	
999	1000	What a Twist #memes #shorts #movie	

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.