

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

In [2]: df_tracks=pd.read_csv("C:\\Users\\91778\\Downloads\\tracks.csv")
df_tracks.head()
```

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechin
0	35wqR4Xm31BWEWsa1Q	Carve	6	126903	0	[Uir]	[45Hl06XoId0i4LBEVpIs]	22-02-1922	0.645	0.4450	0	-13.338	1	0.45
1	021HsdsPcrjgSK7JTKV	Capitulo 2.16 - Banquero Anarquista	0	98200	0	[Fernando Pessoa]	[14fPCoQnZwqk5wt9DxYr]	01-06-1922	0.695	0.2630	0	-22.136	1	0.95
2	07A5yehShnoedVJAZNnc	Vivo para Queerente - Remasterizado	0	181640	0	[Ignacio Corsini]	[5LlOoJbXvVSAMkBS2Um3X2]	21-03-1922	0.434	0.1770	1	-21.180	1	0.05
3	08FmJdXhJyLrnf6pH8k45	El Prisionero - Remasterizado	0	176907	0	[Ignacio Corsini]	[5LlOoJbXvVSAMkBS2Um3X2]	21-03-1922	0.321	0.0946	7	-27.961	1	0.05
4	08y9GtqcCwIOGskdwj5e	Lady of the Evening	0	163080	0	[Dick Haymes]	[3B8JGZsyX9sJchTqcSA7Su]	1922	0.402	0.1580	3	-16.900	0	0.05

```
In [3]: #null values
pd.isnull(df_tracks).sum()
```

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechin
0	id	0	71	0	0	0	0	0	0	0	0	0	0	0
1	name	0	0	0	0	0	0	0	0	0	0	0	0	0
2	popularity	0	0	0	0	0	0	0	0	0	0	0	0	0
3	duration_ms	0	0	0	0	0	0	0	0	0	0	0	0	0
4	explicit	0	0	0	0	0	0	0	0	0	0	0	0	0
5	artists	0	0	0	0	0	0	0	0	0	0	0	0	0
6	id_artists	0	0	0	0	0	0	0	0	0	0	0	0	0
7	release_date	0	0	0	0	0	0	0	0	0	0	0	0	0
8	danceability	0	0	0	0	0	0	0	0	0	0	0	0	0
9	energy	0	0	0	0	0	0	0	0	0	0	0	0	0
10	key	0	0	0	0	0	0	0	0	0	0	0	0	0
11	loudness	0	0	0	0	0	0	0	0	0	0	0	0	0
12	mode	0	0	0	0	0	0	0	0	0	0	0	0	0
13	speechiness	0	0	0	0	0	0	0	0	0	0	0	0	0
14	acousticness	0	0	0	0	0	0	0	0	0	0	0	0	0
15	instrumentalness	0	0	0	0	0	0	0	0	0	0	0	0	0
16	liveness	0	0	0	0	0	0	0	0	0	0	0	0	0
17	valence	0	0	0	0	0	0	0	0	0	0	0	0	0
18	tempo	0	0	0	0	0	0	0	0	0	0	0	0	0
19	time_signature	0	0	0	0	0	0	0	0	0	0	0	0	0
dtype:	int64	object	int64	int64	int64	object	object	object	float64	float64	int64	float64	int64	float64

```
In [4]: df_tracks.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 586672 entries, 0 to 586671
Data columns (total 20 columns):
#   Column              Non-Null Count  Dtype
---  --
0    id                  586672 non-null object
1    name                586661 non-null object
2    popularity          586672 non-null int64
3    duration_ms        586672 non-null int64
4    explicit            586672 non-null int64
5    artists             586672 non-null object
6    id_artists          586672 non-null object
7    release_date        586672 non-null object
8    danceability         586672 non-null float64
9    energy              586672 non-null float64
10   key                 586672 non-null int64
11   loudness            586672 non-null float64
12   mode                586672 non-null int64
13   speechiness         586672 non-null float64
14   acousticness        586672 non-null float64
15   instrumentalness     586672 non-null float64
16   liveness             586672 non-null float64
17   valence              586672 non-null float64
18   tempo               586672 non-null float64
19   time_signature       586672 non-null int64
dtypes: float64(9), int64(6), object(5)
memory usage: 89.5+ MB

In [5]: sorted_df=df_tracks.sort_values("popularity",ascending=True).head(10)
sorted_df
```

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechin
546130	181rTRRCqgZPwPzTUCvQM	Newspaper Reports On Albin, February 1935	0	896575	0	[Norris Goff, Chester Lusk, Carlton Brice...	[3WCQwCpDMPpGzntQz8qumy- 7V48uJ4Bg8Sga7G5G5...	20-02-1935	0.595	0.262	8	-17.746	1	
546222	0yC3Zv5KMm8lT8EFc6Q	恋は水の上 で	0	188440	0	[Hibari Misora]	[1m5pMY5blqJwdxJ7vqQuNq]	1949	0.418	0.388	0	-8.580	1	
546221	0y48Hhwe52099UqYjgRCO	私の誕生日	0	173467	0	[Hibari Misora]	[1m5pMY5blqJwdxJ7vqQuNq]	1949	0.642	0.178	5	-11.700	1	
546220	0xCmgR9ka07hKz3D6Pxy	エロいチョコ クロ(田 CHOCCLO)	0	205280	0	[Hibari Misora]	[1m5pMY5blqJwdxJ7vqQuNq]	1949	0.695	0.467	0	-12.236	0	
546219	0BXS3yGvK5X7KwFH2wros	恋は不思議 なもの	0	185733	0	[Hibari Misora]	[1m5pMY5blqJwdxJ7vqQuNq]	1949	0.389	0.388	2	-8.221	1	
546218	0qrKnQYDYJXhKFATHYVS9	ゆうはど うしたの (WHATSA MALLA U)	0	183427	0	[Hibari Misora]	[1m5pMY5blqJwdxJ7vqQuNq]	1949	0.631	0.249	5	-11.883	1	
546217	0nq5DxDeK5wEzp3AUQAQ45	Screen Director's Playhouse, Music For Million...	0	1767071	0	[Wilms Herbert, June Allyson, Joseph Kiar...]	[2bnIBQWvmmVvxFO84EVM3Y. yVW5adNgYrHFZGa5...	10-04-1949	0.533	0.317	7	-13.047	1	
546216	0KGEdsxLYfCdxM9bezo	ブルーマン ゴ	0	162147	0	[Hibari Misora]	[1m5pMY5blqJwdxJ7vqQuNq]	1949	0.529	0.546	0	-6.462	0	
546215	0bC3PUZvUUXyY7yqoCqY	Screen Director's Playhouse, Trade Winds direc...	0	1776652	0	[Talley Haller's Tay Gerritt, Lurette...]	[7NhkJJT7tUUGVWU8SjXJ. 3kYeelpRCpJz4iQYDv...	19-05-1949	0.599	0.321	0	-15.428	0	
546214	0VWm0u5y6MiWwG0yH41P	Screen Director's Playhouse, It's A Wonderful ...	0	1767576	0	[Joseph Granby, Jimmy Stewart, Tedre...	[6GKS98C4LJzqPOOpHAX253. 56Bz8aExmrx898bly...	08-05-1949	0.645	0.341	8	-12.177	1	

```
In [6]: df_tracks.describe().transpose()
```

	count	mean	std	min	25%	50%	75%	max
popularity	586672.0	27.570053	18.370642	0.0	13.0000	27.000000	41.00000	100.000
duration_ms	586672.0	230515.67286	126526.087418	3344.0	175093.00000	214893.000000	263867.00000	5621218.000
explicit	586672.0	0.044086	0.205286	0.0	0.0000	0.000000	0.00000	1.000
danceability	586672.0	0.563594	0.166103	0.0	0.4530	0.577000	0.68600	0.991
energy	586672.0	0.542036	0.251923	0.0	0.3430	0.549000	0.74800	1.000
key	586672.0	5.221603	3.519423	0.0	2.0000	5.000000	8.00000	11.000
loudness	586672.0	-10.206067	5.089328	-60.0	-12.8910	-9.243000	-6.48200	5.376
mode	586672.0	0.658797	0.474114	0.0	0.0000	1.000000	1.00000	1.000
speechiness	586672.0	0.104864	0.178893	0.0	0.0340	0.044300	0.07630	0.971
acousticness	586672.0	0.449863	0.348837	0.0	0.0969	0.422000	0.78500	0.996
instrumentalness	586672.0	0.113451	0.266868	0.0	0.0000	0.000024	0.00955	1.000
liveness	586672.0	0.213935	0.184326	0.0	0.0983	0.139000	0.27800	1.000
valence	586672.0	0.552292	0.257671	0.0	0.3460	0.564000	0.76900	1.000
tempo	586672.0	118.464857	29.754108	0.0	95.6000	117.384000	136.32100	246.381
time_signature	586672.0	3.873382	0.473162	0.0	4.0000	4.000000	4.00000	5.000

```
In [7]: most_popular=df_tracks.query("popularity>90",inplace=False).sort_values("popularity",ascending=False)
most_popular[:18]
```

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechin
93802	4I3y8OLHqaGvP12zqHqI	Peaches (feat. Daniel Caesars & Giveon)	100	198082	1	[Justin Bieber, Daniel Caesars, Giveon]	[1uNF0ZAHBGIllmzmpC3s/ 2DwKvLUtqUVOYrcdK5...	19-03-2021	0.677	0.696	0	-6.181	1	
93803	7PN2XDkMSvN7XUKlOWJ2CS	drivers license	99	242014	1	[Olivia Rodrigo]	[1McMsEnEETtX1kmY4oIG]	08-01-2021	0.585	0.436	10	-8.761	1	
93804	30fmpvYv8UaQ7OmEz8277	Astronaut In The Ocean	98	132780	0	[Masked Wolf]	[1uL7g3DN3SbsuOQJSEq2EdF]	06-01-2021	0.778	0.695	4	-6.865	0	
92810	SQO7980w1wacV478qGRGL3g	Save Your Tears	97	215627	1	[The Weeknd]	[1Xy04u8uXC1ZnmMpaIF05PJ]	20-03-2020	0.680	0.826	0	-5.487	1	
92811	6IDDoYtWwMLTdkpJkC1B	telepatia	97	160191	0	[Kali Uchis]	[1U1e1zk54VvEUz3yblPjIM]	04-12-2020	0.653	0.524	11	-9.016	0	
92813	0VjW4WGUZAMZmYdzvXM3b	Blinding Lights	96	200040	0	[The Weeknd]	[1Xy04u8uXC1ZnmMpaIF05PJ]	20-03-2020	0.514	0.730	1	-5.934	1	
93805	7MAhCtH4idCbHKGvGmh	Leave The Door Open	96	242096	0	[Bruno Mars, Anderson Paak, Silk Sonic]	[0d5uEVLvB5YfK9QJzeBzAOC/ 3JK9MfCA42L4AUMGU...	05-03-2021	0.586	0.616	5	-7.964	1	
92814	6ISlSGlGaz2bPZDdaFXN	The Business	95	164000	0	[Tiesto]	[2z5DJnhHvPH3dv3cQv9BZ]	16-09-2020	0.798	0.620	8	-7.079	0	
91866	60yMvPSSKw6O3sWmRNBRI	Streets	94	226987	1	[Drea Cat]	[5c9JLQr7Y7YOSnnhXOPd5]	07-11-2019	0.749	0.463	11	-8.433	1	
92816	3FAJ30ONH0QVbMCM5R6Knc	Heartbreak Anniversary	94	198371	0	[Giveon]	[4tndE5e7UuFO4CUXgwJ7IP]	27-03-2020	0.449	0.465	0	-8.964	1	

```
In [8]: df_tracks.set_index("release_date",inplace=True)
df_tracks.index=pd.to_datetime(df_tracks.index)
df_tracks.head()
```

C:\Users\91778\AppData\Local\Temp\ipykernel\_23644\3148201796.py:2: UserWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default) was specified. This may lead to inconsistently parsed dates! Specify a format to ensure consistent parsing.

```
df_tracks.index=pd.to_datetime(df_tracks.index)
```

	id	name	popularity	duration_ms	explicit	artists	id_artists	danceability	energy	key	loudness	mode	speechiness
1922-02-22	35wqR4Xm31BWEWsa1Q	Carve	6	126903	0	[Uir]	[45Hl06XoId0i4LBEVpIs]	0.645	0.4450	0	-13.338	1	0.4510
1922-01-06	021HsdsPcrjgSK7JTKV	Capitulo 2.16 - Banquero Anarquista	0	98200	0	[Fernando Pessoa]	[14fPCoQnZwqk5wt9DxYr]	0.695	0.2630	0	-22.136	1	0.9570
1922-03-21	07A5yehShnoedVJAZNnc	Vivo para Queerente - Remasterizado	0	181640	0	[Ignacio Corsini]	[5LlOoJbXvVSAMkBS2Um3X2]	0.434	0.1770	1	-21.180	1	0.0512
1922-03-21	08FmJdXhJyLrnf6pH8k45	El Prisionero - Remasterizado	0	176907	0	[Ignacio Corsini]	[5LlOoJbXvVSAMkBS2Um3X2]	0.321	0.0946	7	-27.961	1	0.0504
1922-01-01	08y9GtqcCwIOGskdwj5e	Lady of the Evening	0	163080	0	[Dick Haymes]	[3B8JGZsyX9sJchTqcSA7Su]	0.402	0.1580	3	-16.900	0	0.0390

```
In [9]: df_tracks[["artists"]].iloc[18]
```

artists
['Victor Boucher']

```
Name: 1922-01-01 00:00:00, dtype: object

In [10]: df_tracks["duration"]>df_tracks["duration_ms"].apply(lambda x: round(x/1000) )
df_tracks.drop("duration_ms",inplace=True,axis=1)
```

```
In [11]: df_tracks.duration.head()
```

release_date
1922-02-22 127
1922-01-06 98
1922-03-21 182
1922-03-21 177
1922-01-01 163

```
Name: duration, dtype: int64

In [12]: corr_df=df_tracks.drop(["key","mode","explicit"],axis=1).corr(method="pearson")
plt.figure(figsize=(14,6))
heatmap=sns.heatmap(corr_df,corr_df,annot=True,fmt=".1g",vmin=-1,vmax=1,center=0,cmap="inferno",linewidths=1,linecolor="Black")
heatmap.set_title("Correlation Heatmap Between Variables")
heatmap.set_xticklabels(heatmap.get_xticklabels(),rotation=90)

C:\Users\91778\AppData\Local\Temp\ipykernel_23644\1889893544.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
corr_df=df_tracks.drop(["key","mode","explicit"],axis=1).corr(method="pearson")

Out[12]: [Text(0.5, 0, 'popularity'),
Text(1.5, 0, 'danceability'),
Text(2.5, 0, 'energy'),
Text(3.5, 0, 'loudness'),
Text(4.5, 0, 'speechiness'),
Text(5.5, 0, 'acousticness'),
Text(6.5, 0, 'instrumentalness'),
Text(7.5, 0, 'liveness'),
Text(8.5, 0, 'valence'),
Text(9.5, 0, 'tempo'),
Text(10.5, 0, 'time_signature'),
Text(11.5, 0, 'duration')]

Correlation Heatmap Between Variables
```

	popularity	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature	duration
popularity	1	0.2	0.3	0.3	-0.05	-0.4	0.2	-0.05	0.005	0.07	0.09	0.03
danceability	0.2	1	0.2	0.3	0.2	-0.2	-0.1	0.5	0.04	0.1	-0.1	-0.1
energy	0.3	0.2	1	0.6	-0.05	-0.7	-0.2	0.1	0.4	0.2	0.2	0.02
loudness	0.3	0.3	0.6	1	-0.2	-0.5	-0.3	0.03	0.3	0.2	0.2	0.0003
speechiness	-0.05	0.2	-0.05	-0.2	1	0.07	-0.1	0.2	0.05	-0.09	-0.1	-0.1
acousticness	-0.4	-0.2	-0.7	-0.5	0.07	1	0.2	-0.005	-0.2	0.2	-0.2	-0.06
instrumentalness	-0.2	-0.2	-0.2	-0.3	-0.1	0.2	1	-0.04	-0.2	-0.06	-0.4	0.07
liveness	-0.05	-0.1	0.1	0.03	0.2	-0.005	-0.04	1	-3e-05	-0.01	-0.02	0.002
valence	0.005	0.5	0.4	0.3	0.05	-0.2	-0.2	-3e-05	1	0.1	0.1	-0.2
tempo	0.07	-0.04	0.2	0.2	-0.09	-0.2	-0.06	-0.01	0.1	1	0.03	-0.001
time_signature	0.09	0.1	0.2	0.2	-0.1	-0.2	-0.04	-0.02	0.1	0.03	1	0.04
duration	0.03	-0.1	0.02	0.0003	-0.1	-0.06	0.07	0.002	-0.2	0.001	0.04	1

```
In [13]: sample_df=df_tracks.sample(int(0.004*len(df_tracks)))
print(len(df_tracks))

586672

In [14]: plt.figure(figsize=(10,
```