

```
In [ ]: import pandas as pd

data = pd.read_csv("csv/new_dataset.csv")
data.head()
```

```
Out[ ]:
```

	track_id	artists	track_name	popularity	duration_ms	explicit	danceability	energy	key	loudness
0	5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	Comedy	73	230666	False	0.676	0.4610	1	-6.74
1	4qPNDBW1i3p13qLCt0Ki3A	Ben Woodward	Ghost - Acoustic	55	149610	False	0.420	0.1660	1	-17.23
2	1iJBSr7s7jYXzM8EGcbK5b	Ingrid Michaelson; ZAYN	To Begin Again	57	210826	False	0.438	0.3590	0	-9.73
3	6lfxq3CG4xtTiEg7opyCyx	Kina Grannis	Can't Help Falling In Love	71	201933	False	0.266	0.0596	0	-18.57
4	5vjLSffimiIP26QG5WcN2K	Chord Overstreet	Hold On	82	198853	False	0.618	0.4430	2	-9.68



```
In [ ]: data.describe()
```

Out[ ]:

	popularity	duration_ms	danceability	energy	key	loudness	mode	speechiness
count	114000.000000	1.140000e+05	114000.000000	114000.000000	114000.000000	114000.000000	114000.000000	114000.000000
mean	33.238535	2.280292e+05	0.566800	0.641383	5.309140	-8.258960	0.637553	0.084652
std	22.305078	1.072977e+05	0.173542	0.251529	3.559987	5.029337	0.480709	0.105732
min	0.000000	0.000000e+00	0.000000	0.000000	0.000000	-49.531000	0.000000	0.000000
25%	17.000000	1.740660e+05	0.456000	0.472000	2.000000	-10.013000	0.000000	0.035900
50%	35.000000	2.129060e+05	0.580000	0.685000	5.000000	-7.004000	1.000000	0.048900
75%	50.000000	2.615060e+05	0.695000	0.854000	8.000000	-5.003000	1.000000	0.084500
max	100.000000	5.237295e+06	0.985000	1.000000	11.000000	4.532000	1.000000	0.965000



Modify / Clean Dataset

In [ ]:

```
#data = data.drop('track_name',axis=1)
data
```

Out[ ]:

	track_id	artists	popularity	duration_ms	explicit	danceability	energy	key	loudness	mo
0	5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	73	230666	False	0.676	0.4610	1	-6.746	
1	4qPNDBW1i3p13qLCt0Ki3A	Ben Woodward	55	149610	False	0.420	0.1660	1	-17.235	
2	1iJBSr7s7jYXzM8EGcbK5b	Ingrid Michaelson;ZAYN	57	210826	False	0.438	0.3590	0	-9.734	
3	6lfxq3CG4xtTiEg7opyCyx	Kina Grannis	71	201933	False	0.266	0.0596	0	-18.515	
4	5vjLSffimilP26QG5WcN2K	Chord Overstreet	82	198853	False	0.618	0.4430	2	-9.681	
...	...	...	...	...	...	...	...	...	...	...
113995	2C3TZjDRiAzdyViavDJ217	Rainy Lullaby	21	384999	False	0.172	0.2350	5	-16.393	
113996	1hlz5L4IB9hN3WRYPOCGPw	Rainy Lullaby	22	385000	False	0.174	0.1170	0	-18.318	
113997	6x8ZfSoqDjuNa5SVP5QjvX	Cesária Evora	22	271466	False	0.629	0.3290	0	-10.895	
113998	2e6sXL2bYv4bSz6VTdnfLs	Michael W. Smith	41	283893	False	0.587	0.5060	7	-10.889	
113999	2hETkH7cOfqzm3LqZDHzf5	Cesária Evora	22	241826	False	0.526	0.4870	1	-10.204	

114000 rows × 18 columns



In [ ]: %pwd

```
In [ ]: path = '/root/AAI-540/Module2/csv/new_dataset.csv'
data.to_csv(path,index=False)
```

```
In [ ]: data2 = pd.read_csv("csv/new_dataset.csv")
data2.head()
```

Out [ ]:

	track_id	artists	popularity	duration_ms	explicit	danceability	energy	key	loudness	mode	sp
0	5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	73	230666	False	0.676	0.4610	1	-6.746	0	
1	4qPNDBW1i3p13qLCt0Ki3A	Ben Woodward	55	149610	False	0.420	0.1660	1	-17.235	1	
2	1iJBSr7s7jYXzM8EGcbK5b	Ingrid Michaelson;ZAYN	57	210826	False	0.438	0.3590	0	-9.734	1	
3	6lfxq3CG4xtTiEg7opyCyx	Kina Grannis	71	201933	False	0.266	0.0596	0	-18.515	1	
4	5vjLSffimiIP26QG5WcN2K	Chord Overstreet	82	198853	False	0.618	0.4430	2	-9.681	1	



In [ ]:

```
%%html

<p><b>Shutting down your kernel for this notebook to release resources.</b></p>
<button class="sm-command-button" data-commandlinker-command="kernelmenu:shutdown" style="display:none;">Shutdown Ker

<script>
try {
  els = document.getElementsByClassName("sm-command-button");
  els[0].click();
}
catch(err) {
  // NoOp
}
</script>
```

**Shutting down your kernel for this notebook to release resources.**

In [ ]: