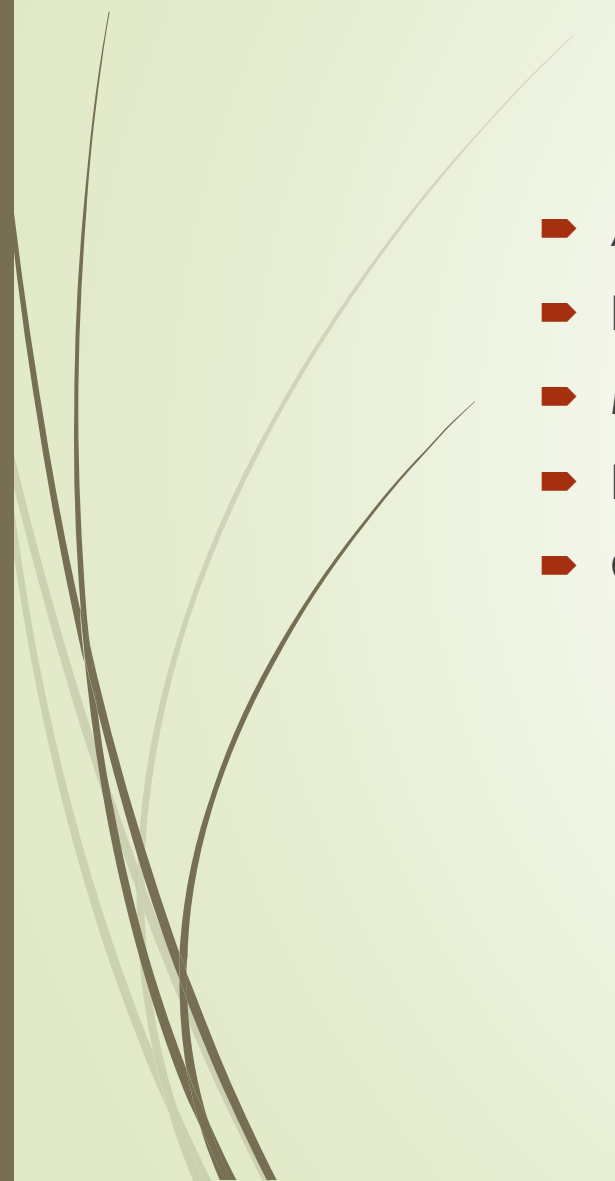CAS KAGGLE

AÑO DE SALIDA DE CANCIONES (2010-2019)

Adrián García Tapia
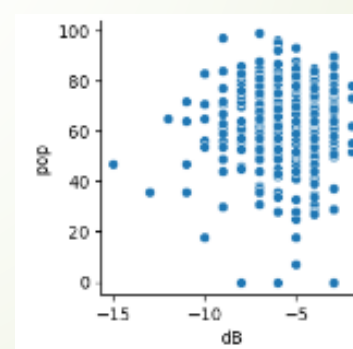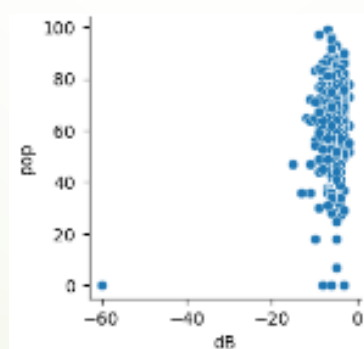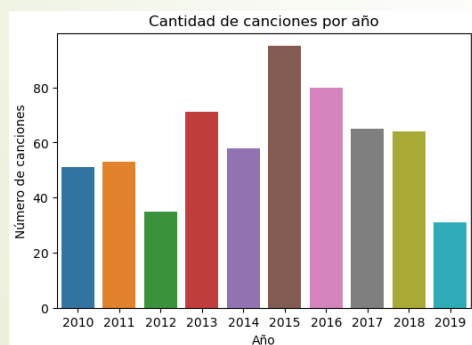
ÍNDICE

- Análisis de datos
- Preprocessing
- Métricas de evaluación / Selección de modelo
- Entrenamiento
- Conclusiones

ANÁLISIS DE DATOS

	Unnamed: 0	title	artist	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop
0	1	Hey, Soul Sister	Train	neo mellow	2010	97	89	67	-4	8	80	217	19	4	83
1	2	Love The Way You Lie	Eminem	detroit hip hop	2010	87	93	75	-5	52	64	263	24	23	82
2	3	TiK ToK	Kesha	dance pop	2010	120	84	76	-3	29	71	200	10	14	80
3	4	Bad Romance	Lady Gaga	dance pop	2010	119	92	70	-4	8	71	295	0	4	79
4	5	Just the Way You Are	Bruno Mars	pop	2010	109	84	64	-5	9	43	221	2	4	78
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
598	599	Find U Again (feat. Camila Cabello)	Mark Ronson	dance pop	2019	104	66	61	-7	20	16	176	1	3	75
599	600	Cross Me (feat. Chance the Rapper & PnB Rock)	Ed Sheeran	pop	2019	95	79	75	-6	7	61	206	21	12	75
600	601	No Brainer (feat. Justin Bieber, Chance the Ra...	DJ Khaled	dance pop	2019	136	76	53	-5	9	65	260	7	34	70
601	602	Nothing Breaks Like a Heart (feat. Miley Cyrus)	Mark Ronson	dance pop	2019	114	79	60	-6	42	24	217	1	7	69
602	603	Kills You Slowly	The Chainsmokers	electropop	2019	150	44	70	-9	13	23	213	6	6	67

PREPROCESSING

	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop
0	-0.289803	2010	-0.894045	1.144728	0.191584	0.873847	-0.749399	1.235396	-0.224762	0.223987	-0.584865	1.147726
1	-0.105885	2010	-1.305241	1.393869	0.801413	0.286731	2.613881	0.521493	1.123029	0.464858	1.956808	1.077609
2	-0.560973	2010	0.051707	0.833300	0.877641	1.460964	0.855803	0.833826	-0.722858	-0.209582	0.752858	0.937376
3	-0.560973	2010	0.010587	1.331584	0.420270	0.873847	-0.749399	0.833826	2.060623	-0.691326	-0.584865	0.867259
4	1.442618	2010	-0.400609	0.833300	-0.037101	0.286731	-0.672961	-0.415505	-0.107562	-0.594977	-0.584865	0.797142
5	1.769425	2010	-2.209874	0.957871	0.648956	0.286731	-0.520084	0.075304	-0.312661	-0.498628	0.752858	0.727025
6	-0.560973	2010	0.051707	0.459588	0.801413	0.873847	-1.055152	1.324634	-0.634959	-0.691326	0.083997	0.727025
7	-0.560973	2010	1.203057	0.335017	-0.951844	-0.300385	-0.443646	-0.638599	0.009637	-0.354105	-0.584865	0.727025
8	-1.344974	2010	-1.058524	-2.094115	-1.256758	-1.474617	-0.443646	-1.709454	-0.254061	2.873575	-0.718637	0.656908
9	-1.744569	2010	0.298425	0.085875	1.106327	0.873847	-0.825837	0.387636	0.302635	-0.065059	-0.584865	0.446558

MÉTRICAS DE EVALUACIÓN / SELECCIÓN DE MODELO

```
DummyClassifier mean f1-score: 0.027274550011392118
```

```
SVC mean f1-score: 0.154779612348218
KNeighborsClassifier mean f1-score: 0.12530078744784628
RandomForestClassifier mean f1-score: 0.20216567282047157
GradientBoostingClassifier mean f1-score: 0.16780158956148322
```

CONCLUSIONES

F1-Score (Train): 0.17302
F1-Score (Test): 0.20572

precision vs. recall curve

class 2010. AUC = 0.31
class 2011. AUC = 0.26
class 2012. AUC = 0.29
class 2013. AUC = 0.16
class 2014. AUC = 0.12
class 2015. AUC = 0.23
class 2016. AUC = 0.24
class 2017. AUC = 0.17
class 2018. AUC = 0.22
class 2019. AUC = 0.83