



# Ranking de Música de Spotify

Predicción de popularidad de canciones utilizando modelos de Deep Learning

Integrantes:

1. Isidora Jara
2. Néstor Manríquez
3. Daniel Reyes

Profesor:

Francisco Plaza



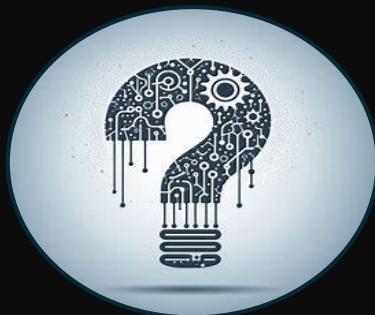
0:00

15:40

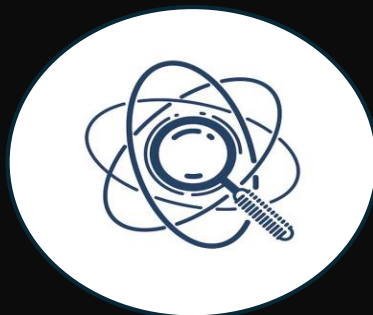
# Índice



Objetivos



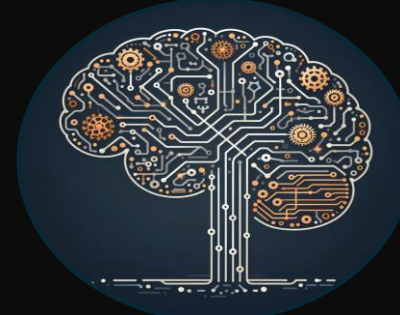
Descripción de  
Datos



Análisis  
exploratorio



Esquema  
metodológico



Descripción de  
modelo

# Objetivos



## General:

Desarrollar un modelo de Deep Learning para predecir qué canciones serán las más populares en la plataforma de streaming de Spotify.

## Específicos:

1. Recopilar y preparar los datos.
2. Realizar un análisis exploratorio de los datos.
3. Desarrollar y entrenar un modelo de Deep Learning.
4. Evaluar el modelo.
5. Optimizar y ajustar el modelo en caso de ser necesario.





# Descripción de datos

Variable	Descripción
País	Indica si la canción es global o de uno de los 34 países donde Spotify opera.
Popularidad	Puntuación basada en días en el Top 200 y su posición diaria.
Top_Max	Posición más alta alcanzada en el top en 1401 días de datos.
Título	Nombre de la canción.
Artista	Nombre del artista.
Formato	Publicada como sencillo o parte de un álbum.
Explícito	Indica si tiene advertencia de contenido explícito.
Bailable	Adecuación para bailar (0.0 a 1.0).
Energía	Intensidad y actividad de la canción (0.0 a 1.0).
Tono	Tono general de la pista (0 = Do, 1 = Do#, etc.).
Volumen	Sonoridad en decibelios (dB).
Modo	Modalidad (1 = mayor, 0 = menor).
Hablado	Presencia de palabras habladas (0.0 a 1.0).
Acústico	Confianza en si la pista es acústica (0.0 a 1.0).
Instrumental	Probabilidad de que la pista sea instrumental (0.0 a 1.0).
Valencia	Positividad musical de la pista (0.0 a 1.0).
BPM	Tempo en pulsos por minuto.
Duración_ms	Duración de la pista en milisegundos.
Género	Género predominante del artista.
Lanzamiento	Fecha de publicación de la canción.
Días_Lanzamiento	Días desde el lanzamiento hasta la fecha de referencia.



7:00

15:40



# Análisis

## Exploración

##Datos faltantes####

País	Popularidad	Top_Max	Título	Artista
0	0	0	0	0
Formato	Explícito	Bailable	Energía	Tono
0	0	0	0	0
Volumen	Modo	Hablado	Acústico	Instrumental
0	0	0	0	0
Valencia	BPM	Duración_ms	Género	Lanzamiento
0	0	0	0	0
Días_Lanzamiento				
0				

## ##Estadísticas descriptivas####

	n	mean	sd	median	min	max
Popularidad	167397	5475.11	13172.30	529.60	0.8	233766.90
Top_Max	167397	89.52	57.99	84.00	1.0	200.00
Bailable	167397	0.67	0.14	0.69	0.0	0.98
Energía	167397	0.64	0.17	0.66	0.0	1.00
Tono	167397	5.27	3.64	5.00	0.0	11.00
Volumen	167397	-6.52	2.61	-6.12	-60.0	1.63
Hablado	167397	0.12	0.12	0.07	0.0	0.97
Acústico	167397	0.24	0.25	0.15	0.0	1.00
Instrumental	167397	0.01	0.09	0.00	0.0	0.99
Valencia	167397	0.49	0.22	0.49	0.0	0.99
BPM	167397	121.03	28.76	119.95	0.0	230.27
Duración_ms	167397	206392.04	50141.49	202214.00	30133.0	3653957.00
Días_Lanzamiento	167397	1337.54	2453.66	834.00	9.0	44128.00

Variable type: character								
	skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
1	País	0	1	2	11	0	35	0
2	Título	0	1	1	195	0	46060	0
3	Artista	0	1	1	618	0	25076	0
4	Formato	0	1	5	11	0	3	0
5	Género	0	1	3	16	0	21	0
6	Link	0	1	53	53	0	60996	0



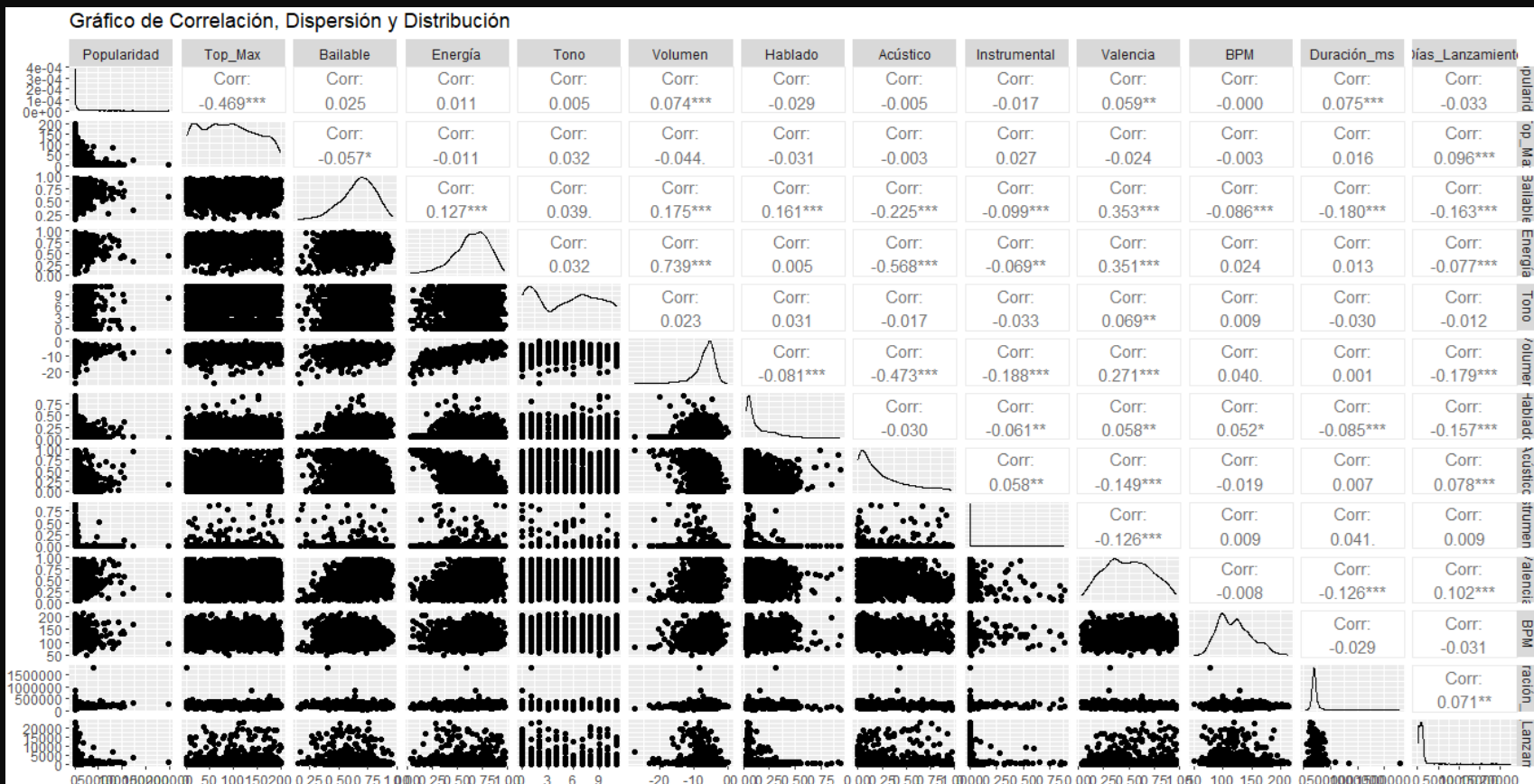
# Análisis Exploratorio

Matriz de Correlación

Popularidad	1.00	-0.48	0.08	0.02	0.01	0.09	-0.04	0.00	-0.04	0.06	-0.01	0.02	-0.05
Top_Max	-0.48	1.00	-0.09	-0.00	-0.00	-0.05	-0.04	0.01	0.04	-0.02	-0.01	0.04	0.08
Bailable	0.08	-0.09	1.00	0.13	0.01	0.18	0.20	-0.22	-0.11	0.35	-0.06	-0.18	-0.17
Energía	0.02	-0.00	0.13	1.00	0.01	0.73	0.03	-0.56	-0.07	0.37	0.09	0.01	-0.07
Tono	0.01	-0.00	0.01	0.01	1.00	-0.01	0.03	0.00	-0.01	0.04	-0.01	-0.00	-0.01
Volumen	0.09	-0.05	0.18	0.73	-0.01	1.00	-0.06	-0.46	-0.20	0.28	0.07	0.01	-0.19
Hablado	-0.04	-0.04	0.20	0.03	0.03	-0.06	1.00	-0.04	-0.07	0.05	0.10	-0.12	-0.15
Acústico	0.00	0.01	-0.22	-0.56	0.00	-0.46	-0.04	1.00	0.06	-0.12	-0.08	-0.01	0.09
Instrumental	-0.04	0.04	-0.11	-0.07	-0.01	-0.20	-0.07	0.06	1.00	-0.10	-0.01	0.05	0.02
Valencia	0.06	-0.02	0.35	0.37	0.04	0.28	0.05	-0.12	-0.10	1.00	0.03	-0.14	0.09
BPM	-0.01	-0.01	-0.06	0.09	-0.01	0.07	0.10	-0.08	-0.01	0.03	1.00	0.01	-0.01
Duración_ms	0.02	0.04	-0.18	0.01	-0.00	0.01	-0.12	-0.01	0.05	-0.14	0.01	1.00	0.09
Días_Lanzamiento	-0.05	0.08	-0.17	-0.07	-0.01	-0.19	-0.15	0.09	0.02	0.09	-0.01	0.09	1.00
	Popularidad	Bailable	Energía	Tono	Volumen	Acústico	Valencia	BPM	Días_Lanzamiento				



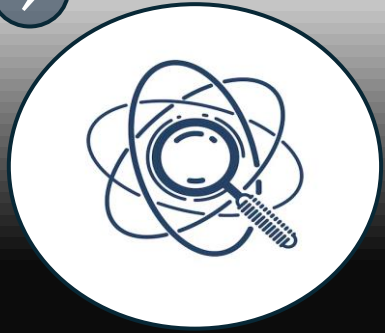
# Análisis Exploratorio



9:20

15:40





# Análisis Exploratorio

## ##Correlación Parcial###



	Popularidad	Top_Max	Bailable	Energía	Tono	Volumen	Hablado	Acústico	Instrumental	Valencia	BPM	Duración_ms	Días_Lanzamiento
Popularidad	1.000000000	-0.473472457	0.03481096	-0.032782481	0.014418366	0.065697432	-0.06331576	0.031353994	-0.003197294	0.04527903	-0.005633240	0.053486605	-0.007640047
Top_Max	-0.473472457	1.000000000	-0.03296865	0.011447606	0.007187116	-0.002497189	-0.04632902	0.006964644	0.016735810	0.01867298	-0.009832982	0.044243520	0.046992535
Bailable	0.034810962	-0.032968647	1.000000000	-0.156089093	-0.010448279	0.081763257	0.17607884	-0.224659440	-0.037850493	0.35603768	-0.106402146	-0.110220273	-0.146839458
Energía	-0.032782481	0.011447606	-0.15608909	1.000000000	0.016377193	0.623149022	0.12598136	-0.402217106	0.123128892	0.29673136	0.009896504	0.022898593	0.066593144
Tono	0.014418366	0.007187116	-0.01044828	0.016377193	1.000000000	-0.023775213	0.02865964	0.005473652	-0.010010633	0.03726885	-0.015734476	0.005387283	-0.014507446
Volumen	0.065697432	-0.002497189	0.08176326	0.623149022	-0.023775213	1.000000000	-0.19096257	-0.059756866	-0.227102198	0.01405607	0.028465046	0.028542830	-0.218367778
Hablado	-0.063315757	-0.046329020	0.17607884	0.125981357	0.028659644	-0.190962572	1.000000000	0.010271664	-0.090053298	-0.01249356	0.117606935	-0.063521813	-0.147026217
Acústico	0.031353994	0.006964644	-0.22465944	-0.402217106	0.005473652	-0.059756866	0.01027166	1.000000000	-0.001584526	0.17674976	-0.058433364	-0.021789608	-0.004268836
Instrumental	-0.003197294	0.016735810	-0.03785049	0.123128892	-0.010010633	-0.227102198	-0.09005330	-0.001584526	1.000000000	-0.03962022	0.005023107	0.033262581	-0.051437226
Valencia	0.045279033	0.018672983	0.35603768	0.296731360	0.037268848	0.014056072	-0.01249356	0.176749762	-0.039620216	1.000000000	0.037871780	-0.108978681	0.192059131
BPM	-0.005633240	-0.009832982	-0.10640215	0.009896504	-0.015734476	0.028465046	0.11760694	-0.058433364	0.005023107	0.03787178	1.000000000	0.003995418	0.004416233
Duración_ms	0.053486605	0.044243520	-0.11022027	0.022898593	0.005387283	0.028542830	-0.06352181	-0.021789608	0.033262581	-0.10897868	0.003995418	1.000000000	0.081236796
Días_Lanzamiento	-0.007640047	0.046992535	-0.14683946	0.066593144	-0.014507446	-0.218367778	-0.14702622	-0.004268836	-0.051437226	0.19205913	0.004416233	0.081236796	1.000000000

##Tabla Formato####

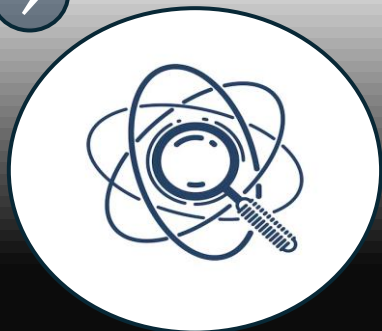
Album compilation	91080	2285	Sencillo	74032
-------------------	-------	------	----------	-------

##Tabla País####

Argentina	Australia	Austria	Belgium	Brazil	Canada	Chile	Colombia
2980	3923	6453	5919	3299	5532	3387	3096
Costa Rica	Denmark	Ecuador	Finland	France	Germany	Global	Indonesia
2909	5123	2608	6665	6357	6887	5395	3024
Ireland	Italy	Malaysia	Mexico	Netherlands	New Zealand	Norway	Peru
5125	5148	3840	2763	6602	3501	5994	2619
Philippines	Poland	Portugal	Singapore	Spain	Sweden	Switzerland	Taiwan
2732	5146	4514	3560	4242	6820	7619	7463
Turkey	UK	USA					
3305	6287	6560					





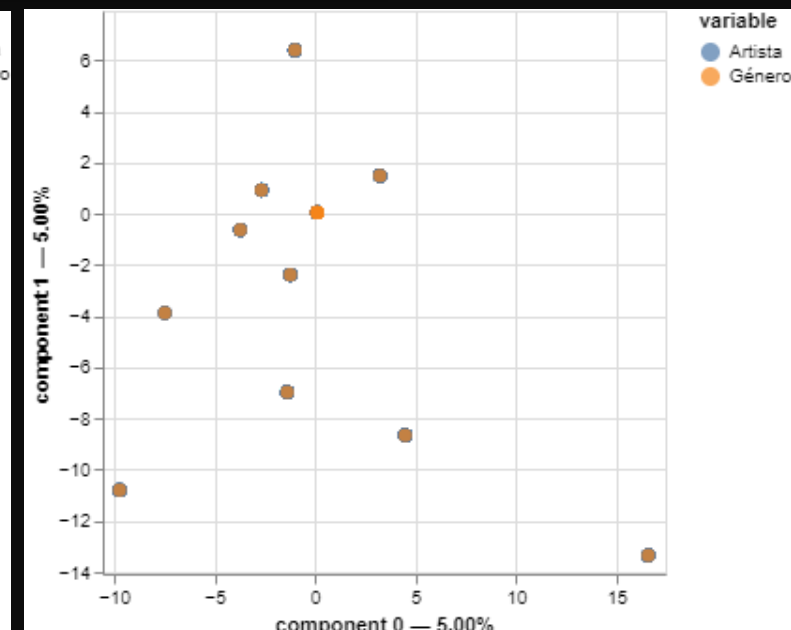
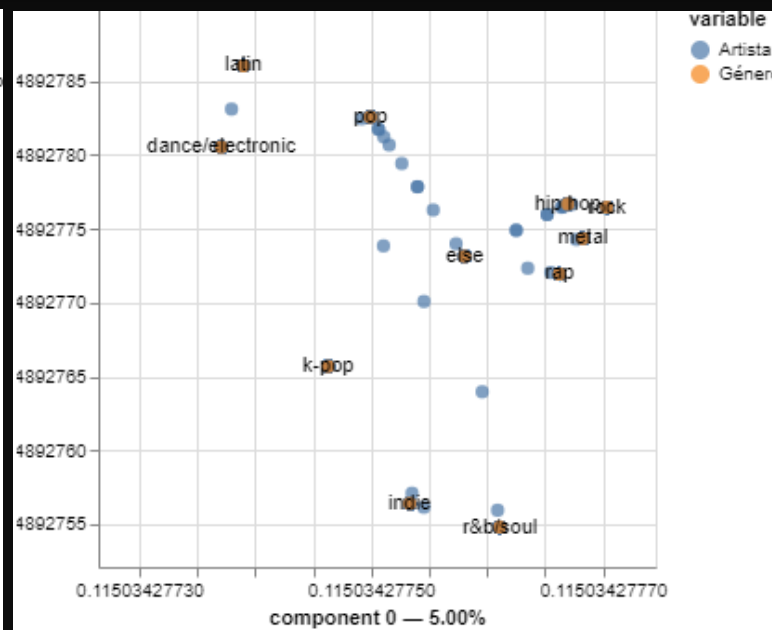
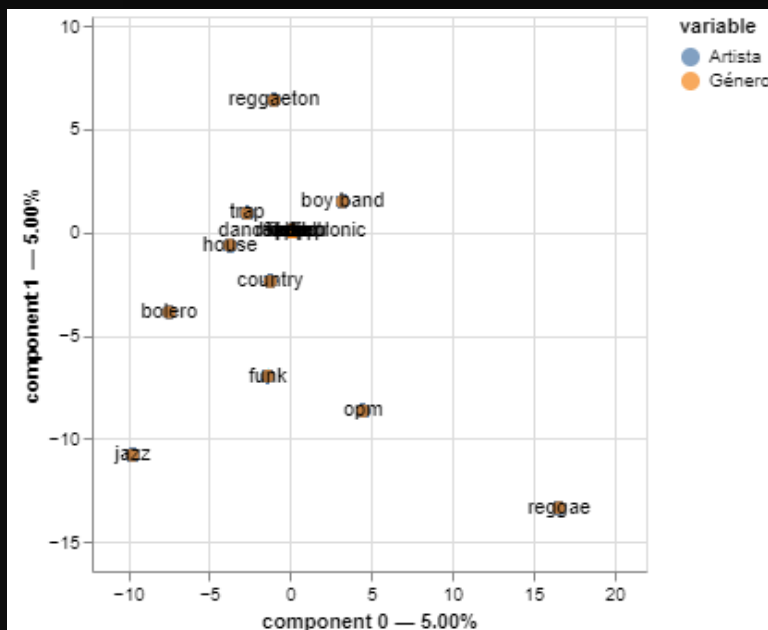


# Análisis Exploratorio

##Tabla Género####

bolero	boy band	country	dance/electronic	else
188	1934	689	5654	8306
funk	hip hop	house	indie	jazz
999	35665	4682	7271	130
k-pop	latin	metal	opm	pop
4495	9478	2286	340	58485
r&b/soul	rap	reggae	reggaeton	rock
4432	13553	127	1001	6001
trap				
1681				

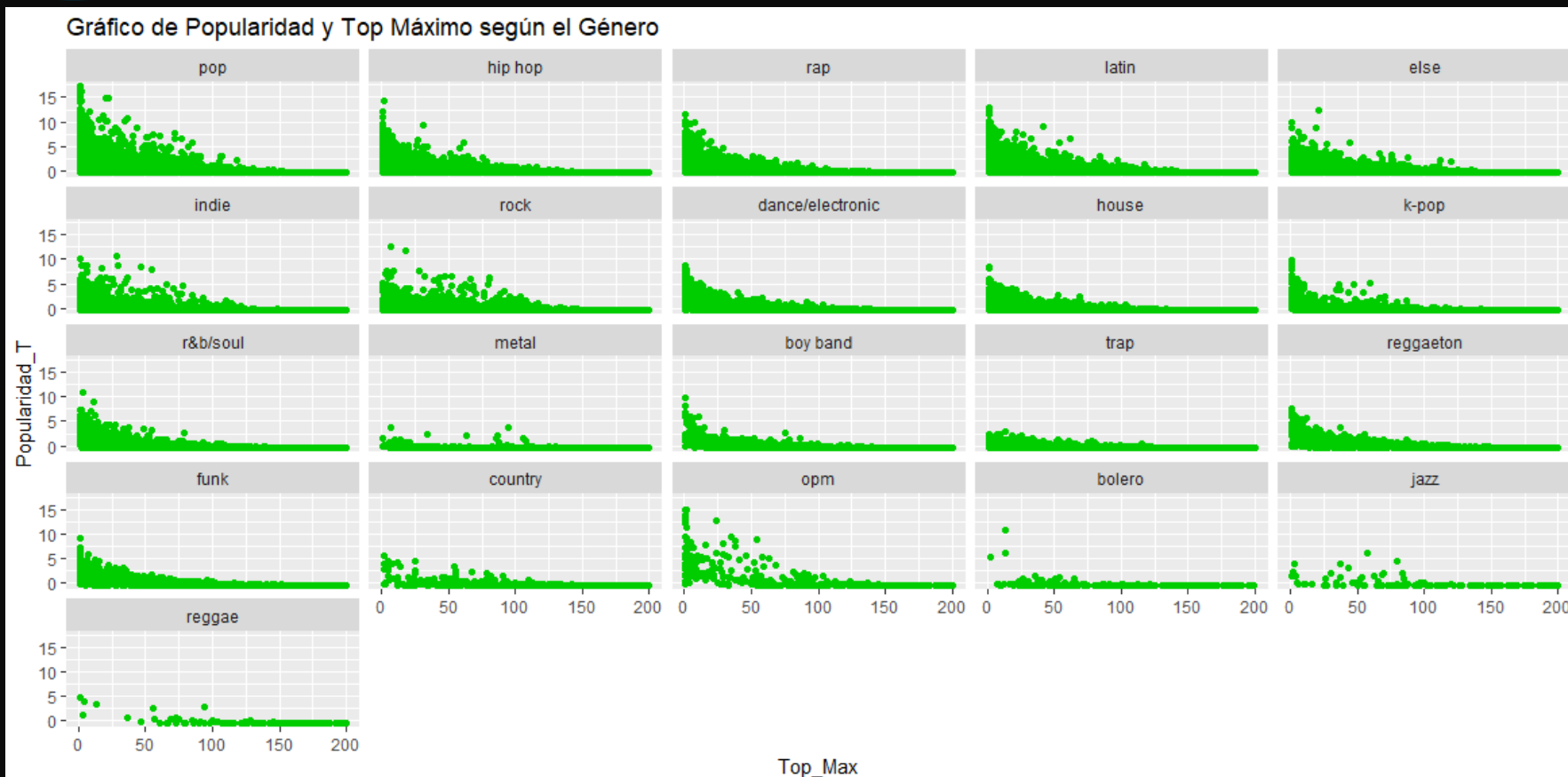
##Artista####



9:20

15:40

# Análisis Exploratorio





# Análisis Exploratorio

Bo1ero

	País	Popularidad	Top_Max	Título	Artista	Formato	Explícito	Bailable	Energía	Tono	Volumen
	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Mexico	149128.	13	ahora...	Luis M...	Álbum	0	0.78	0.877	6	-3.68
2	Mexico	86978.	13	la in...	Luis M...	Álbum	0	0.374	0.727	10	-4.35
3	Mexico	74695.	2	culpa...	Luis M...	Álbum	0	0.54	0.367	0	-14.2
4	Mexico	23872	36	cuand...	Luis M...	Álbum	0	0.686	0.868	0	-3.89
5	Ecuador	22843.	54	ahora...	Luis M...	Álbum	0	0.78	0.877	6	-3.68
6	Peru	19530.	51	ahora...	Luis M...	Álbum	0	0.78	0.877	6	-3.68
7	Costa Ri...	17642.	26	la in...	Luis M...	Álbum	0	0.374	0.727	10	-4.35
8	Ecuador	17050.	65	la in...	Luis M...	Álbum	0	0.374	0.727	10	-4.35
9	Peru	15717.	50	la in...	Luis M...	Álbum	0	0.374	0.727	10	-4.35
10	Argentina	15197.	41	ahora...	Luis M...	Álbum	0	0.78	0.877	6	-3.68



Canción

## Ahora Te Puedes Marchar



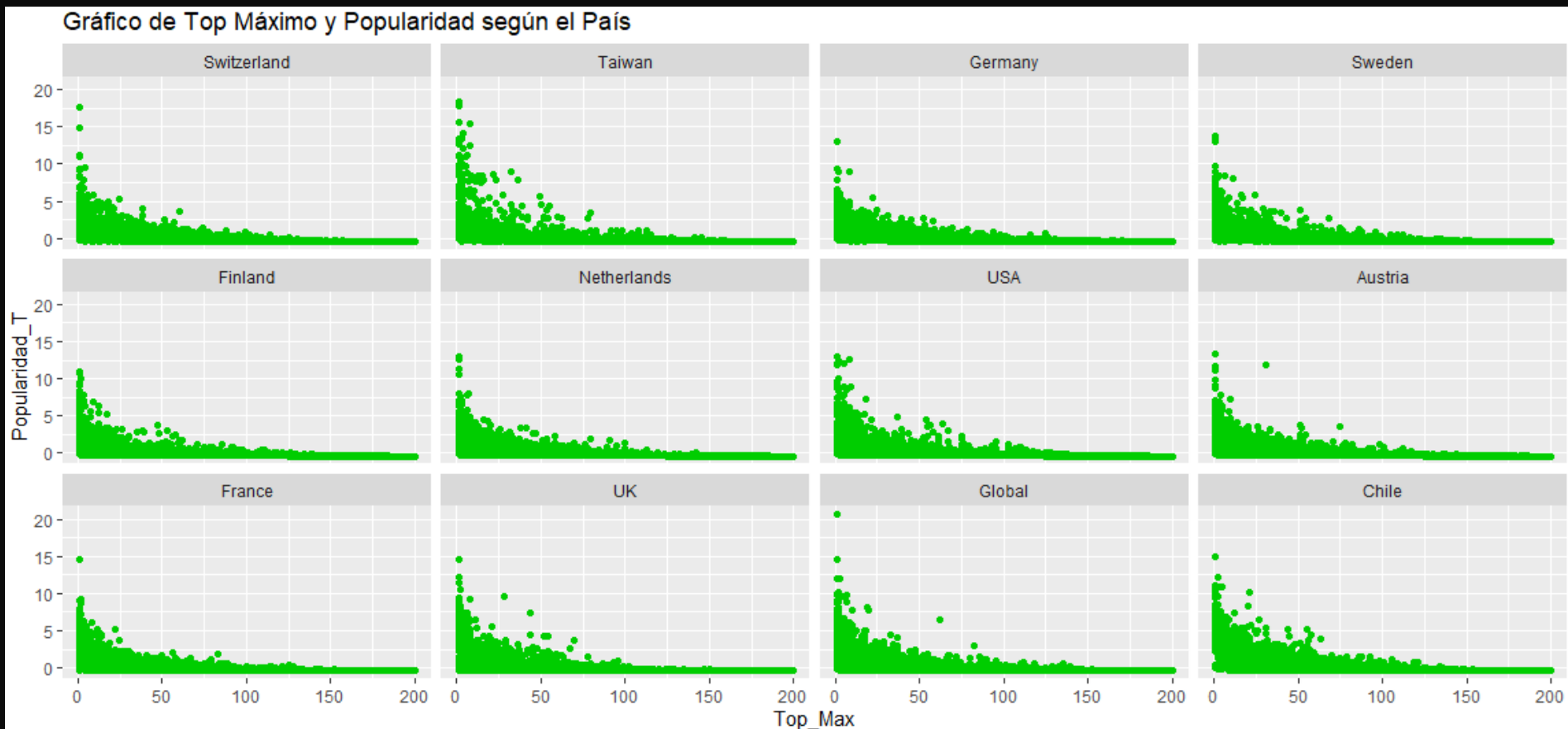
Luis Miguel • Soy Como Quiero Ser • 1987 • 3:11 • 728,701,018

Ahora Te Puedes Marchar  
Luis Miguel

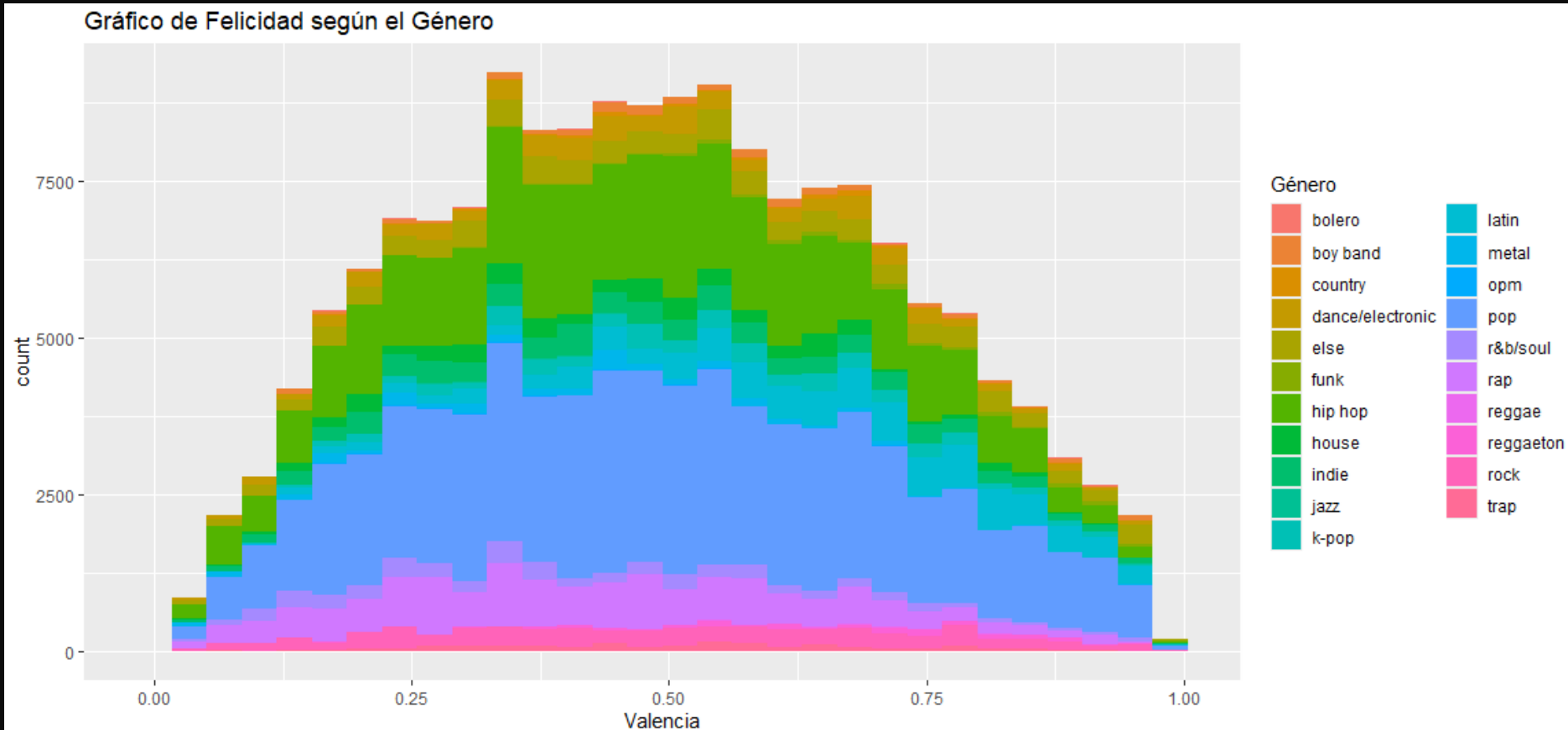
9:20

15:40

# Análisis Exploratorio



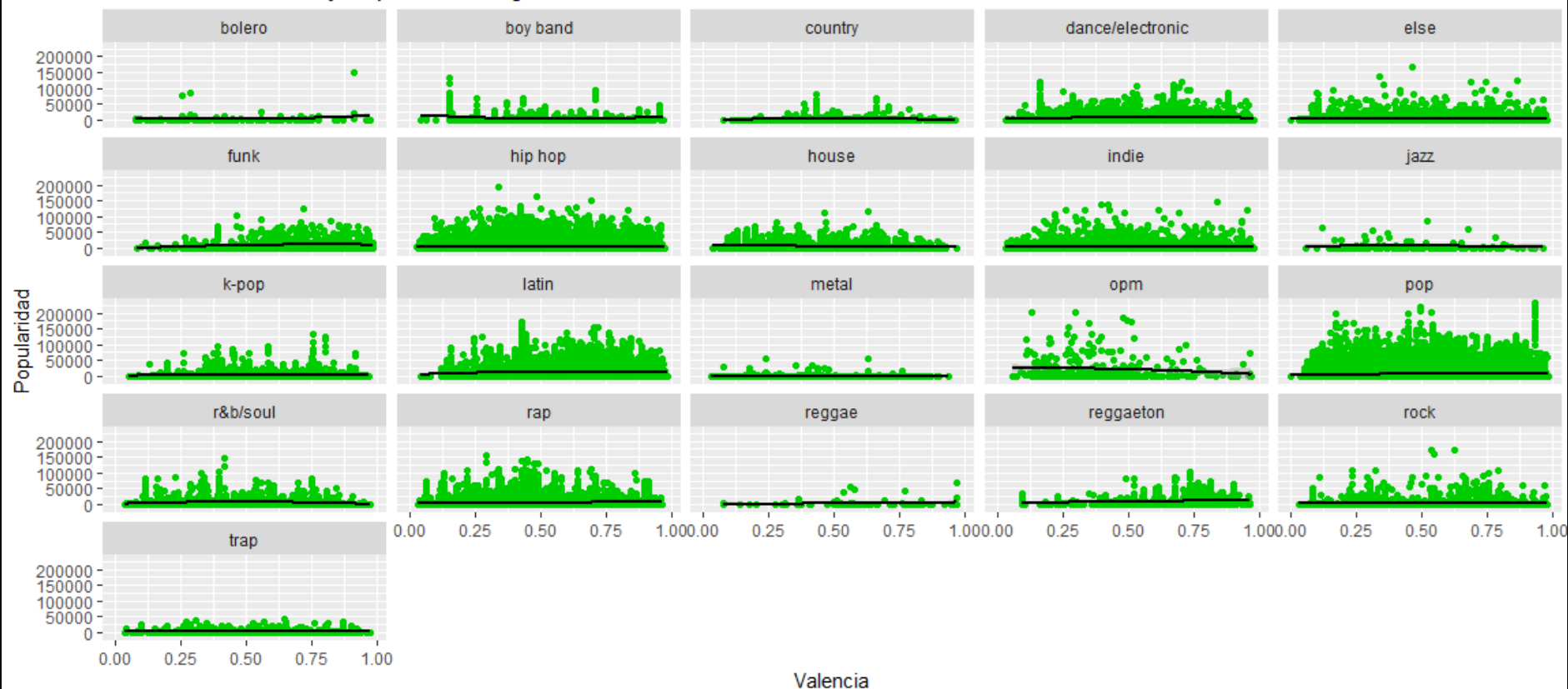
# Análisis Exploratorio





# Análisis Exploratorio

Gráfico de Felicidad y Popularidad según el Género



# Análisis Exploratorio

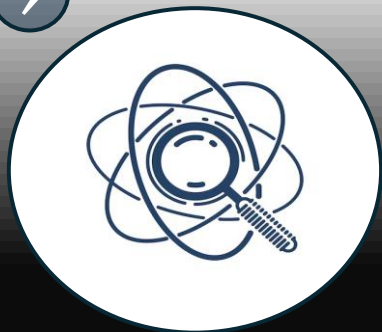


Gráfico de BPM y Popularidad según si es Explícito

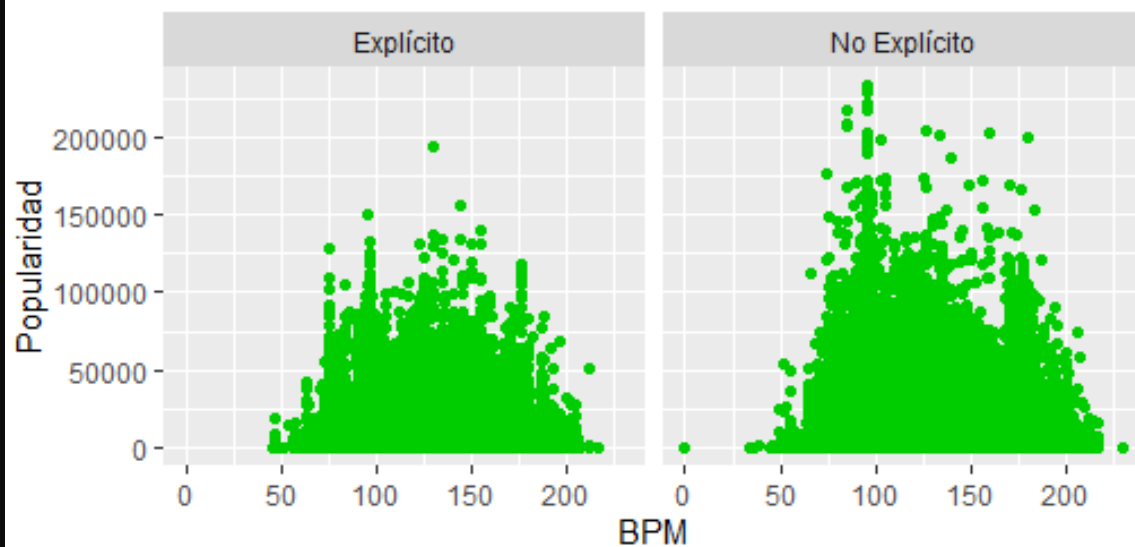
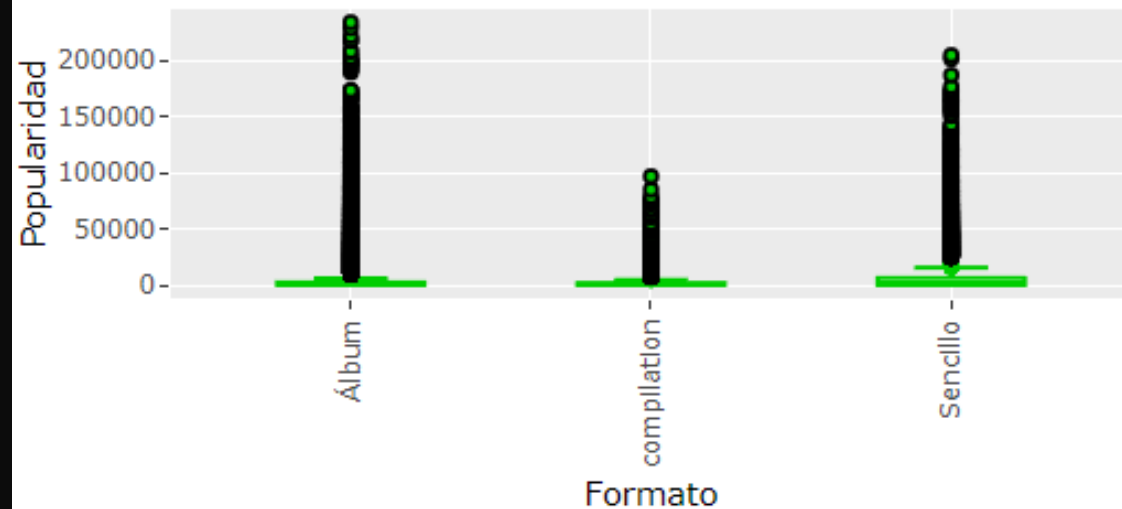
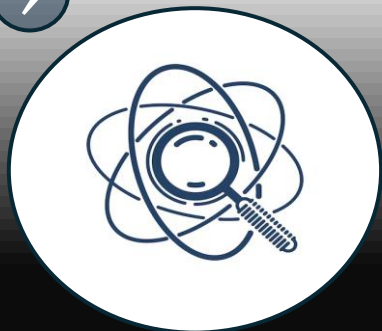


Gráfico Caja de la Popularidad y el Formato

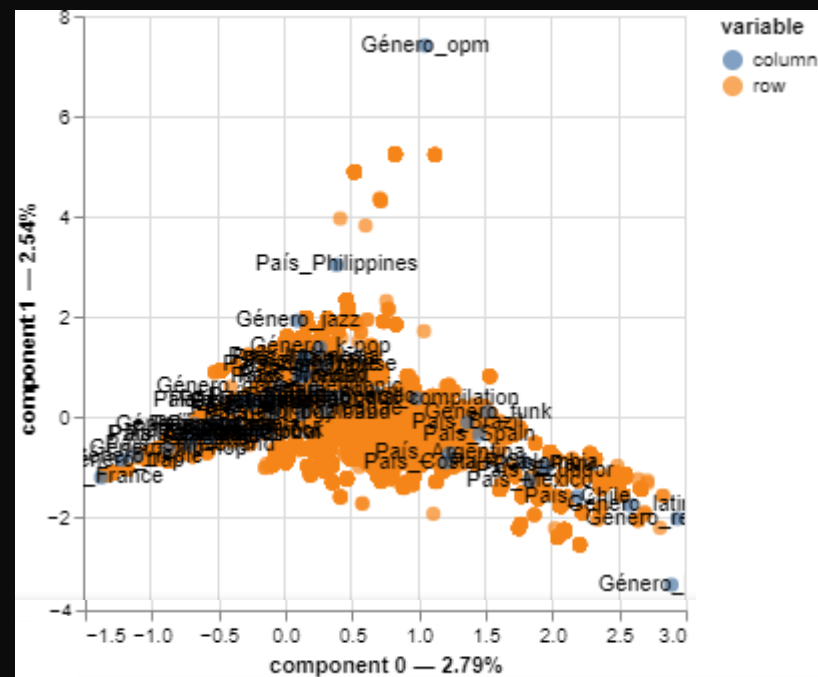
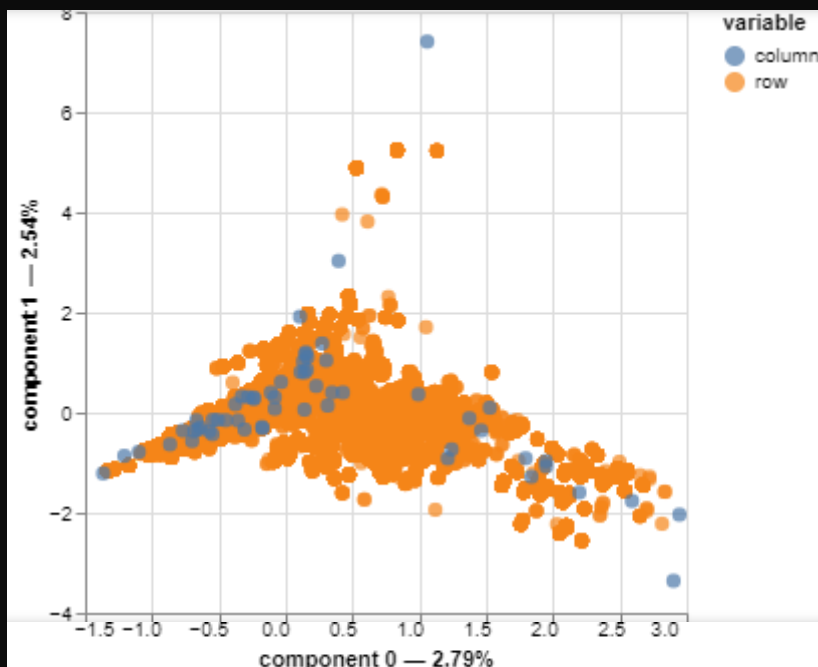




# Análisis Exploratorio

Importance of components:

	PC1	PC2
Standard deviation	5.014e+04	2.443e+03
Proportion of Variance	9.976e-01	2.370e-03
Cumulative Proportion	9.976e-01	1.000e+00



9:20

15:40



# Análisis Exploratorio de las Variables

- Numerical Encoding

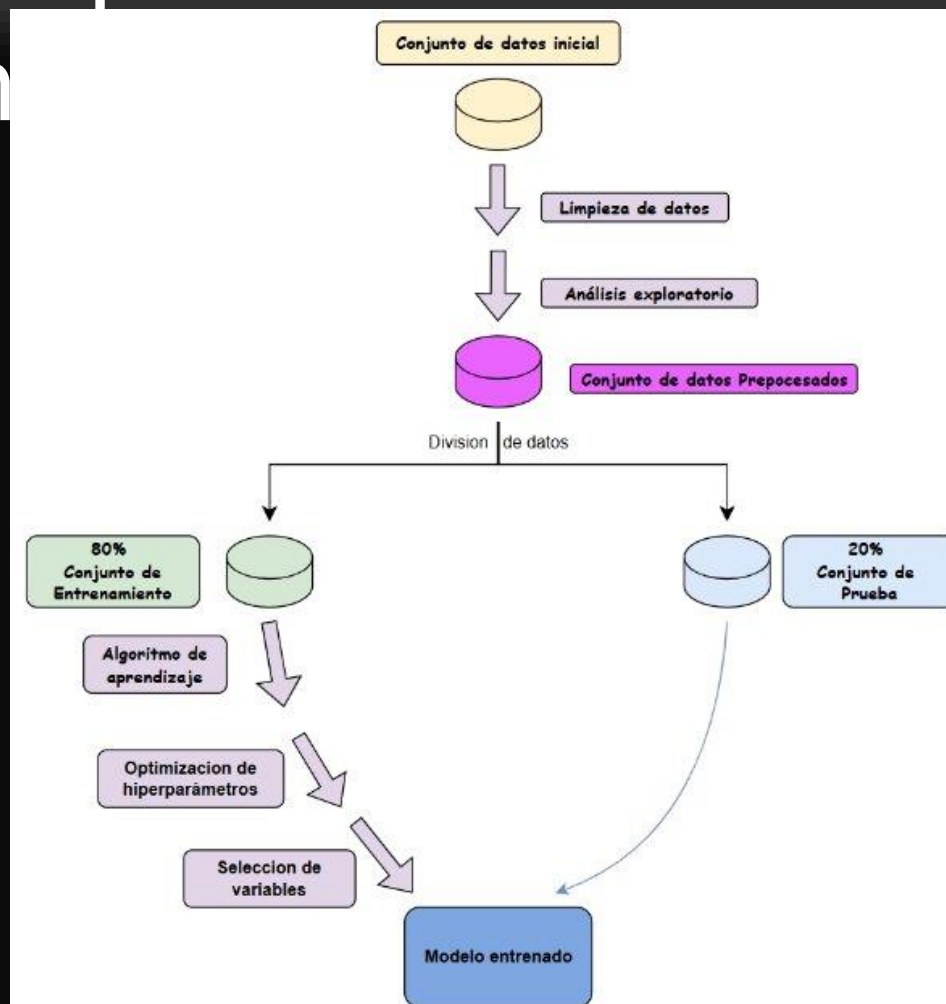
```
$ codGen      <dbl> 17, 17, 17, 17, 17, 17, 17, 16, 17, 17, 16, 17, 17, 17, 17, ...  
              'opm'= 16,  
              'pop'=17,  
$ codPaís     <dbl> 4, 9, 37, 26, 29, 20, 29, 26, 29, 27, 26, 29, 29, 33, 23, 2, 20...  
              'Belgium'= 4,  
              'Costa Rica'= 9,  
              'Global'=37
```

- One Hot Encoding

```
$ Explícito_Explícito <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
$ Explícito_no.explicito <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
$ Formato_Album      <dbl> 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0,  
$ Formato_compilation <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
$ Formato_Sencillo   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,
```

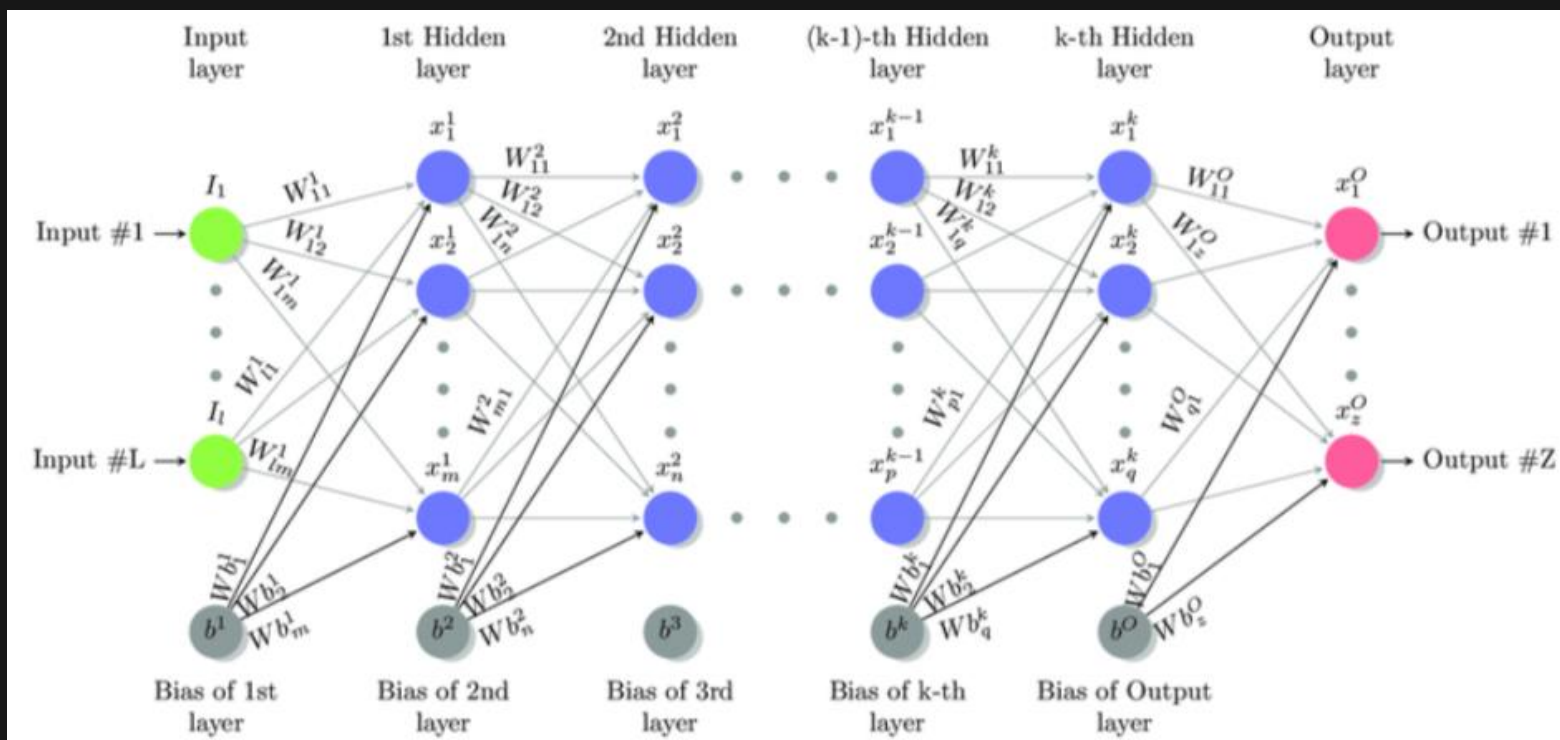
# Esquema

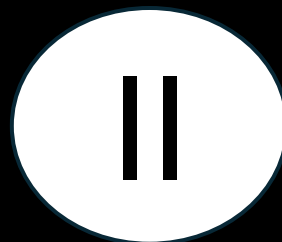
## m





# Descripción de Modelo





15:40

15:40

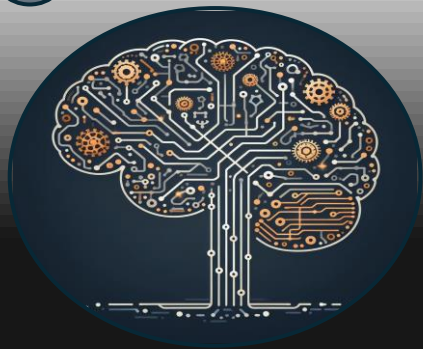
Letras

El Fin.

Muchas gracias por su  
Se aceptan preguntas.  
atencion.

Referencias

- Clases de Deep Learning.



# Descripción de Modelo de Entrada

Top\_Max

Explícito

Valencia

País

BPM

Artista

Formato

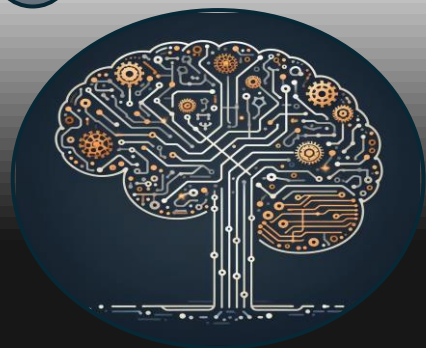
Género

- Capas Ocultas

La cantidad de capas ocultas se definirán más adelante de acuerdo a los requerimientos del modelo

- Capa de Salida

Popularidad



# Descripción de Modelo de Activación

De acuerdo al rendimiento del modelo,  
utilizaremos 'sigmoide', 'Tangente Hiperbólica' o 'ReLU'.

- Función de Coste

De acuerdo al rendimiento del modelo,  
utilizaremos el error cuadrático medio o  
el error absoluto medio.

- Optimizador

utilizaremos 'Adam'.

