

CURSO UNIVERSITARIO AVANZADO EN DATA ANALYST: Análisis e Interpretación de Datos

**Actividad grupal: Definición de un
problema estadístico: modelización
y propuesta de soluciones**

Estrategia 1: David Toscano Recalde

Estrategia 2: Nicolás Felipe Trujillo Montero e Iris Aguado Fernández

Conclusiones y redacción de artículo: Todos los miembros



Análisis del bienestar de un país en función de las emociones encontradas en las canciones más escuchadas por la población

David Toscano Recalde, Nicolás Felipe Trujillo Montero, Iris Aguado Fernández, UNIR

Resumen

En este artículo se va a realizar un análisis para estudiar cómo se relacionan diferentes variables económicas como el desempleo con las emociones que se pueden asociar a las canciones de Spotify más escuchadas por la población de Estados Unidos.

1. Introducción y estado del arte

La relación entre la música y el ser humano es un tema ampliamente estudiado, se han generado diversos estudios para encontrar relaciones entre los gustos musicales de ciertas poblaciones con sus creencias, estado anímico o características intrínsecas de las mismas (DeWall et al., 2011) (McAuslan & Waung, 2018). A continuación, se van a presentar algunos estudios que relacionan el estilo de música escuchado en una determinada época, con patrones económicos y sociales presentes durante ese periodo de tiempo. En primer lugar, los estudios de Dolan et al. (2008), Bianchi (2016) y Hill et. al (2012) relacionan el desempleo con la medida subjetiva del bienestar en una sociedad. Otros estudios han sugerido que en tiempos económicos y sociales retadores es más probable que las canciones tengan un matiz más serio y reflexivo (Pettijohn et al., 2012). Por otro lado, el estudio de McAuslan and Waung (2018) concluyó que los términos asociados a individualismo en la música top de la década de 2010 aumentaron significativamente con respecto a los de otras década pasadas. Además, se ha estudiado que dentro de una misma sociedad los gustos musicales sugieren comportamientos diferentes para cada estrato socioeconómico pero manteniendo una coherencia entre los valores de la música con los valores que describen a tal estrato (Snibbe & Markus, 2005).

2. Metodología y resultados

En este estudio el objetivo va a ser analizar la relación entre las emociones asociadas a las canciones más escuchadas frente a características económicas de Estados Unidos, como puede ser la tasa de desempleo. A continuación, se plasmará el objetivo realizando una comparativa con los resultados obtenidos en los artículos mencionados anteriormente. Se van a utilizar datos de Spotify, concretamente

un listado del top 200 diario, donde se define el título de la canción, la fecha y el número de streams. Por otro lado, se utilizará un listado de 90.000 canciones en las que se define el título de la canción y las emociones asociadas a ella, como pueden ser felicidad, optimismo o agresividad (Akiki, C., & Burghardt, M. 2020). Las emociones se medirán como la cantidad de streams de las canciones relacionadas con cada una de ellas, frente al total de streams para cada año, teniendo en cuenta que un stream es una reproducción de más de 30 segundos en la plataforma Spotify. Por último, se emplearán datos de variables económicas de Estados Unidos obtenidos de 'The World Bank'.

Para analizar la relación entre todos estos datos se plantean dos estrategias diferentes, que serán definidas a continuación. En ambas estrategias, las variables independientes serán las emociones asociadas a las canciones más escuchadas y la variable dependiente será la tasa de desempleo anual.

2.1. Primera estrategia

En primer lugar, se ha realizado la combinación de los diferentes dataset para unificarlos en uno solo. En el dataset final los datos estarán agrupados por años en el intervalo de 2017 a 2021 y tendrá variables asociadas a las emociones, que representarán el número de veces que ha aparecido dicha emoción en el conjunto de streams totales que han sido registrados cada año. Además, contendrá cinco variables económicas que representarán la inflación, el GNI (Gross National Income) per capita, el índice de precios al consumidor y la tasa de desempleo tanto en jóvenes (15 - 24 años) como en adultos.

A partir de estos datos se realizará un estudio de correlación entre las diferentes variables mediante una matriz de correlación de Pearson en la que los datos se muestran como un mapa de calor. Los tonos azules representan correlación positiva mientras los tonos rojos representan correlación negativa.

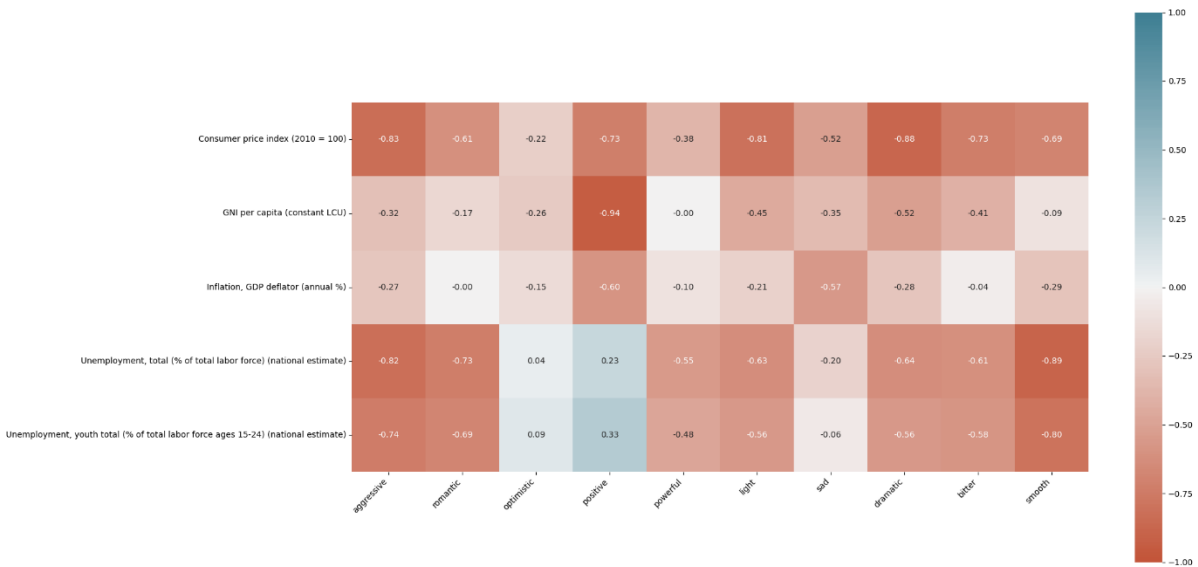


Figura 1. Matriz de correlación

Analizando la Figura 1 se puede observar que las variables con una correlación mayor con el desempleo son 'aggressive' y 'smooth', dicha correlación es negativa. Sin embargo, este dato contradice lo observado en estudios anteriores ya que la correlación encontrada para estos estudios en las variables que representan emociones negativas es directamente proporcional al desempleo. La variable con mayor correlación positiva frente al desempleo es 'positive' pero dicho porcentaje es bajo por lo que no se puede establecer una relación concluyente entre esta variable y el desempleo. Además, se puede observar que el desempleo en adultos jóvenes no tuvo correlaciones significativamente diferentes frente a las asociadas al desempleo en la población adulta.

Por otro lado, la correlación más alta encontrada se observa entre la variable 'positive' y el 'GNI per capita' que se define como la producción nacional y extranjera total reclamada por un país dividido entre el número de residentes de dicho país. Esta correlación negativa nos indica que cuanto mayor es la cantidad facturada por cada habitante, la positividad es más baja en ese país.

2.2. Segunda estrategia

Al igual que en la estrategia anterior, se ha realizado la combinación de los diferentes dataset para unificarlos en uno solo, la diferencia es que se utiliza un dataset económico diferente que contiene los datos asociados al desempleo de Estados Unidos pero en lugar de estar agrupado por años se encuentran distribuidos por meses. Para poder hacer una comparativa con los resultados obtenidos anteriormente, se procede a elegir la mediana de la variable 'Value', que representa el porcentaje de

desempleo, de entre los doce meses que contiene cada año, obteniendo de este modo un valor de dicha variable por año. A continuación, para el resto de variables se elige el valor asociado a la fila de la mediana elegida anteriormente. Después, se procede a realizar una regresión lineal para ver la posible relación existente entre el desempleo y las variables asociadas a las diferentes emociones. Después de realizar una primera regresión se observan datos poco concluyentes por lo que se estudia la posibilidad de la existencia de outliers, que una vez detectados se eliminan. Tras la eliminación de dichos valores atípicos, se vuelve a realizar la regresión lineal obteniendo resultados más aceptables que se analizan a continuación.

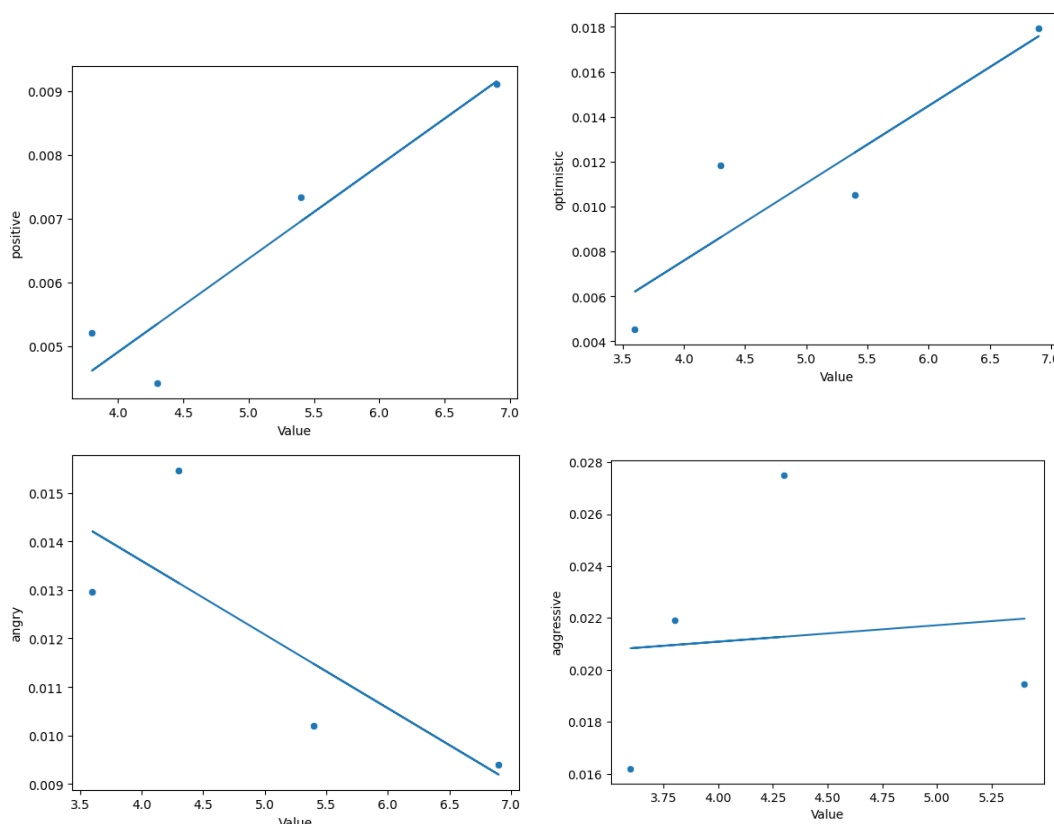


Figura 2. Resultados de la regresión lineal

Analizando los resultados mostrados en la Figura 2, observamos que las emociones positive, optimistic y aggressive son directamente proporcionales al desempleo (con $R^2 = 0.90, 0.81, 0.01$, respectivamente) mientras que angry es inversamente proporcional (con $R^2 = 0.63$).

3. Conclusiones y líneas futuras

Los resultados obtenidos en este trabajo muestran que contraintuitivamente, y opuesto a lo que se tiene en la bibliografía, la agresividad tiene una relación negativa mayor con el desempleo en la

primera estrategia, mientras que en la segunda estrategia la relación es positiva pero con un porcentaje muy bajo. De forma similar se puede apreciar lo que ocurre con las variables 'positive' y 'optimistic'.

Además en la primera estrategia se evidencia como, en general, el desempleo muestra correlaciones más altas que los otros indicadores económicos generales para las diferentes emociones seleccionadas, lo que podría señalar que el desempleo es un buen indicador del estado general de un país o al menos es más reactivo a las emociones de la música popular.

Comparando ambas estrategias, vemos como tanto para la variable 'positive' como para 'optimistic', la relación que se obtiene es directamente proporcional con el desempleo. Sin embargo, en la segunda estrategia dicha relación es bastante mayor que la observada en la primera. Por el contrario, para la variable 'aggressive' en la primera estrategia la relación obtenida es inversa, mientras que en la segunda es directamente proporcional aunque bastante baja.

4. Discusiones

Los años 2020 y 2021 estuvieron marcados por la pandemia de Covid-19, y esto influyó en la economía mundial y en la sociedad. Además, que el propio comportamiento de los usuarios de Spotify durante la pandemia pudo haber cambiado. Se necesita ampliar los datos a más años, principalmente porque los años de Covid-19 tuvieron comportamientos que podrían ser tratados como anomalías. Al usar la plataforma de Spotify nos estamos restringiendo a analizar datos de personas que utilizan esta plataforma, posiblemente adultos jóvenes, sesgando el resultado.

Referencias

- DeWall, C. N., Pond, R. S. J., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, 5, 200–207. <http://dx.doi.org/10.1037/a0023195>
- McAuslan, P., & Waung, M. (2018). Billboard hot 100 songs: Selfpromoting over the past 20 years. *Psychology of Popular Media Culture*, 7, 171–184. <http://dx.doi.org/10.1037/ppm0000118>
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29, 94–122. <http://dx.doi.org/10.1016/j.joep.2007.09.001>

Bianchi, E. C. (2016). American individualism rises and falls with the economy: Cross-temporal evidence that individualism declines when the economy falters. *Journal of Personality and Social Psychology*, 111, 567–584. <http://dx.doi.org/10.1037/pspp0000114>

Hill, S. E., Rodeheffer, C. D., Griskevicius, V., Durante, K., & White, A. E. (2012). Boosting beauty in an economic decline: Mating, spending, and the lipstick effect. *Journal of Personality and Social Psychology*, 103, 275–291. <http://dx.doi.org/10.1037/a0028657>

Pettijohn, T. F., II, Eastman, J. T., & Richard, K. G. (2012). And the beat goes on: Popular billboard song beats per minute and key signatures vary with social and economic conditions. *Current Psychology*, 31, 313–317. <http://dx.doi.org/10.1007/s12144-012-9149-y>

Snibbe, A. C., & Markus, H. R. (2005). You can't always get what you want: Educational attainment, agency, and choice. *Journal of Personality and Social Psychology*, 88, 703–720. <http://dx.doi.org/10.1037/0022-3514.88.4.703>

Akiki, C., & Burghardt, M. (2020, November). Toward a Musical Sentiment (MuSe) Dataset for Affective Distant Hearing. In *CHR* (pp. 225-235).

Top 1 de Spotify (charts.csv): <https://www.kaggle.com/datasets/dhruvildave/spotify-charts?resource=download>

Lista de Canciones junto con Emocion asociadas (muse_v3.csv):
<https://www.kaggle.com/code/cakiki/muse-dataset-getting-started/data>

Datos Macroeconómicos a nivel global (US_SECONOMICS.csv):
<https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS>

Datos Macroeconómicos a nivel Estados Unidos (DP_LIVE.csv):
<https://databank.worldbank.org/reports.aspx?source=world-development-indicators#>