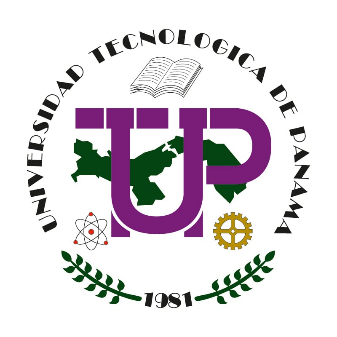
**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**

**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**

**LICENCIATURA EN INGENIERÍA EN SISTEMAS Y COMPUTACIÓN**

**PROYECTO FINAL DE PROBABILIDAD APLICADA A TIC**

**PROFESOR JUAN MARCOS CASTILLO, PhD**

**Nombre del proyecto: ¿Qué factores hacen popular a una canción?**

**Del ritmo al Dato.**

**Logotipo, nombre de la empresa

El contenido generado por IA puede ser incorrecto.Estudiantes: Emily Kennion, 8-1028-238**

**Daniel Nieto, 8-1019-552**

**Loyi Zou, 8-1015-602**

**Kenner Martelo, 20-14-8503**

**Keoni Castillo, 8-1032-1858**

**Wilson Wu, 2-756-299**

*(Empresa de nosotros ficticia)*

**Fecha: 30 de Julio del 2025**

**I SEMESTRE**

**Indice**

[**Introducción** 3](#_Toc204714164)

[**Justificación del Proyecto** 4](#_Toc204714165)

[**Antecedentes** 4](#_Toc204714166)

[**Definición del Problema** 5](#_Toc204714167)

[**Análisis descriptivos** 5](#_Toc204714168)

[**Gráfico de dispersión por canción: relación entre bailabilidad, popularidad y género.** 5](#_Toc204714169)

[**Gráfico de dispersión por canción: relación entre popularidad, Volumen promedio de la canción en decibelios y género.** 6](#_Toc204714170)

[**Grafica de promedio de popularidad por año de lanzamiento** 7](#_Toc204714171)

[**Gráfico de promedio de popularidad por Artistas con más de 3 canciones** 8](#_Toc204714172)

[**Análisis estocásticos** 10](#_Toc204714173)

[**Análisis con diferentes modelos estocásticos** 11](#_Toc204714174)

[**Determinación de la base de datos** 11](#_Toc204714175)

[**Preprocesamiento y limpieza** 11](#_Toc204714176)

[**Selección de variables** 11](#_Toc204714177)

[**Selección de Modelos** 13](#_Toc204714178)

[**Conclusiones** 15](#_Toc204714179)

[**Recomendaciones y futuros estudios** 16](#_Toc204714180)

[**Bibliografía** 17](#_Toc204714181)

# **Introducción**

**Del ritmo al dato: descubriendo qué hace popular a una canción**

Se desarrollará un análisis con herramientas de probabilidad y estadísticas ya facilitadas por nuestro profesor. Se responderá nuestra incógnita principal, **¿Qué factores influyen en la popularidad de una canción?** Y le agregamos un títuloatrayente, **“Del ritmo al dato”.**

Nuestro equipo, **KDWL Music** —nombre ficticio formado por nuestras siglas—, utilizó diversas aplicaciones como Excel, Power BI, Word, Canva, entre otras que se mostrarán a lo largo del desarrollo del contenido.

La base del análisis fue una base de datos combinada, resultado de la integración de dos datasets provenientes de Kaggle, modificados para responder a nuestros intereses específicos.

En particular, nuestra base de datos contiene registros desde 1987 hasta 2024. Aunque abarca un rango temporal amplio, no es una base de datos muy extensa, lo que permite trabajar con una muestra manejable pero representativa de esos años. Esto facilita un análisis más enfocado, donde se pueden observar tendencias relevantes sin perder precisión. Al mismo tiempo, contiene campos que permiten comparaciones factibles y significativas, como lo son: el año en que salió la canción, su nivel de energía, duración, bailabilidad, popularidad en Spotify, vistas en YouTube, entre otras características técnicas y de impacto.

Aplicamos distintos modelos gráficos para visualizar los datos y extraer conclusiones, incluyendo histogramas, diagramas de dispersión, gráficos de líneas, entre otros. Estas visualizaciones nos permitieron desarrollar un análisis profundo que nos acerca a una respuesta sólida a nuestra incógnita.

*"Es mejor tener una respuesta aproximada a la pregunta correcta que una respuesta exacta a la pregunta equivocada."*  
— **John W. Tukey**

# **Justificación del Proyecto**

El presente proyecto nace de nuestro genuino interés por la música. Al revisar las opciones disponibles para trabajar con bases de datos, no dudamos en seleccionar una relacionada con canciones y su popularidad. Aunque muchas veces creemos que nuestras preferencias musicales son completamente personales, surge la curiosidad: ¿y si en realidad seguimos ciertos patrones compartidos con millones de personas?

Esta investigación surge del entusiasmo del grupo **KDWL Music** por combinar una temática creativa con herramientas analíticas. Representa una oportunidad de aplicar conceptos técnicos y probabilísticos en un contexto real, cercano y dinámico como lo es la industria musical.

¿Son realmente nuestros gustos los que determinan lo que escuchamos, o simplemente seguimos tendencias? Este tipo de cuestionamientos fueron los que nos impulsaron a elegir y desarrollar este proyecto.

# **Antecedentes**

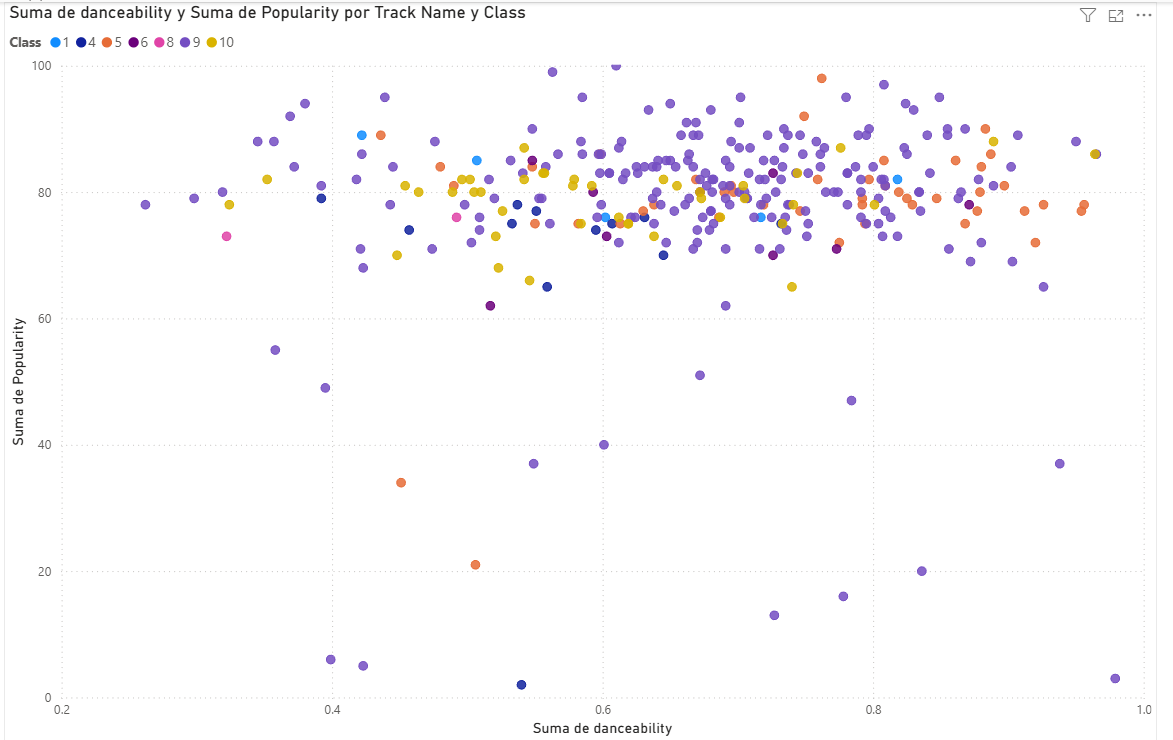
En nuestro caso, el grupo **KDWL Music** decidió analizar los patrones detrás de la popularidad de una canción desde un enfoque propio. A partir de conversaciones grupales, llegamos a la hipótesis de que sí podrían existir factores comunes que, más allá de los gustos individuales, influyen en que una canción se vuelva viral o ampliamente escuchada. Este proyecto nace precisamente de ese interés por comprender, desde una perspectiva probabilística y estadística, qué características hacen que una canción destaque.

Este análisis toma aún más relevancia considerando que vivimos la era de masificación de las plataformas digitales, como **YouTube** y **Spotify**, que no solo distribuyen música, sino que también generan métricas detalladas sobre el consumo musical. Estas plataformas son claves tanto en la difusión como en la recolección de datos que utilizamos para nuestro estudio.

# **Definición del Problema**

## **Análisis descriptivos**

### **Gráfico de dispersión por canción: relación entre bailabilidad, popularidad y género.**

****

Con una correlación de R= 0.07239928

*(cálculos realizados en el Excel)*

**Descriptivamente:**Se puede observar que no existe una relación clara entre la bailabilidad (*danceability*) y la popularidad (*popularity*) de las canciones. Existen valores con baja bailabilidad que se encuentran entre los más populares, lo cual permite descartar la hipótesis de que *a mayor bailabilidad, mayor popularidad*. Por tanto, la bailabilidad no es un factor determinante para la popularidad de una canción en este conjunto de datos.

**Estadísticamente:**  
El análisis de correlación arrojó un valor de R = 0.0723, lo que indica una relación prácticamente nula entre las variables. Esto confirma que no hay una asociación significativa entre qué tan bailable es una canción y qué tan popular resulta ser.

Se utilizaron colores para representar diferentes géneros musicales según la variable *Class*. Se observan colores predominantes como el morado (Pop), el naranja (Hip Hop) y el amarillo (Rock), lo que también permite identificar la distribución de géneros dentro del gráfico.

Se hicieron las mismas graficas con Valence y Energy y arrojaron resultados similares, dejandonos claro que estos factores de “emociones” (por las variables bailabilidad, valence y energy) no afectan la popularidad de una canción.

### **Gráfico de dispersión por canción: relación entre popularidad, Volumen promedio de la canción en decibelios y género.**

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Con una correlación de R= 0.13956348

**Descriptivamente**

Existe una tendencia donde las canciones con un nivel de loudness entre -10 y -5 dB tienden a ser más populares. Las canciones demasiado silenciosas o excesivamente ruidosas suelen tener menos aceptación.

**Estadísticamente**

El coeficiente de correlación de Pearson entre el volumen de una canción y su popularidad es de 0.139, lo cual indica una relación positiva débil. Es decir, las canciones con mayor volumen tienden ligeramente a ser más populares, aunque este factor por sí solo no determina la popularidad de manera significativa.

Se observa que ciertos géneros musicales tienden a concentrarse en rangos específicos de volumen (loudness) y popularidad. En particular, los géneros más comerciales se ubican en niveles de loudness moderados (entre -10 y -5 dB) y muestran mayor popularidad. Por otro lado, géneros con menos representación tienden a tener una distribución más dispersa en términos de popularidad, posiblemente reflejando un alcance más de nicho o alternativo.

### **Grafica de promedio de popularidad por año de lanzamiento**

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Además de las características sonoras, la variable ‘año de lanzamiento’ mostró cierta influencia en la popularidad. Entre los años 2000 y 2005 se observa un pico significativo en el promedio de popularidad, lo cual puede deberse a la permanencia de canciones icónicas, el efecto nostalgia en las audiencias actuales y la transición a plataformas digitales. Esto sugiere que factores contextuales también pueden incidir en la popularidad.

Se observa que las canciones lanzadas antes de 2005 presentan una popularidad promedio alta y constante, probablemente debido a la permanencia de grandes clásicos en las plataformas actuales. Sin embargo, entre 2005 y 2009 se registra una caída significativa en la popularidad promedio, seguida de una recuperación inestable desde 2010 en adelante. Esta variabilidad reciente puede atribuirse a cambios en los hábitos de consumo musical, el impacto de las redes sociales y los algoritmos de recomendación.

### **Gráfico de promedio de popularidad por Artistas con más de 3 canciones**

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

El análisis muestra que el **nombre del artista tiene una influencia significativa en la popularidad**, dado que artistas con múltiples canciones en el dataset como Doja Cat, The Weeknd o Ed Sheeran mantienen promedios de popularidad consistentemente altos.

**Popularidad vs. Año de lanzamiento:**

* + La popularidad no muestra una tendencia clara con el tiempo.
  + Hubo un pico en los 2000, posiblemente por algún artista o fenómeno musical particular, pero no hay una relación directa fuerte.
  + Correlación muy baja (R = 0.139), lo que indica que el año de lanzamiento no influye significativamente.

1. **Popularidad por Género (colores en un gráfico anterior):**
   * Algunos géneros pueden tener una ligera tendencia a ser más populares, pero no hay un género dominante absoluto.
   * Esto sugiere que el género por sí solo no determina la popularidad, aunque puede influir dependiendo del artista o del año.
2. **Popularidad por Artista:**
   * Claramente hay artistas que tienen promedios de popularidad sistemáticamente más altos (Doja Cat, The Weeknd, Ed Sheeran).
   * Esto indica que el artista sí tiene un impacto fuerte en la popularidad, probablemente por su base de fans, visibilidad en redes, o campañas de marketing.

**Conclusión tentativa**

Con base en los análisis realizados hasta ahora:

El artista es el factor que más influye en la popularidad de una canción, seguido posiblemente por características musicales específicas. El año de lanzamiento y el género parecen tener un efecto limitado por sí solos. Esto sugiere que **el reconocimiento del artista y el diseño sonoro** (como lo que transmite una canción: si es energética, alegre, etc.) pueden ser claves para alcanzar popularidad.

## **Análisis estocásticos**

*(cálculos realizados en el Excel)*

1. Probabilidad de que una canción seleccionada al azar tenga una popularidad mayor a 85

se observó que la probabilidad de que una canción seleccionada al azar tenga una popularidad mayor a 85 es de aproximadamente un **21.4%.**

1. Probabilidad de que una canción sea popular dado que su valence (grado de positividad) es > 0.7

al calcular la probabilidad condicional de que una canción sea popular dado que su valence (grado de positividad) es mayor a 0.7, se obtuvo el mismo valor (21.4%). Esto indica que tener una valencia alta no aumenta ni disminuye la probabilidad de popularidad, por lo que este atributo no parece estar relacionado de manera significativa con el éxito de una canción en este conjunto.

1. Comparar la probabilidad de que una canción sea popular para un artista específico versus la probabilidad general (21.4%).

Artista 1-Doja Cat: Al analizar el impacto del artista en la popularidad de las canciones, se observó que el 80% de las canciones de Doja Cat (que en la gráfica de artistas fue la primera) superan el umbral de popularidad de 85, frente a un promedio general del 21.4%. analizado.

Artista 2- Olivia Rodrigo: se observó que el 71.42% de las canciones de Olivia Rodrigo superan el umbral de popularidad de 85.

Artista 3-BTS: se observó que el 66.66% de las canciones de BTS superan el umbral de popularidad de 85.

Este patrón revela que los artistas con gran reconocimiento tienden a tener una proporción significativamente mayor de canciones exitosas en comparación con el promedio general. Por lo tanto, se puede concluir que:

***El artista es un factor relevante que influye de manera significativa en la popularidad de una canción.***

# **Análisis con diferentes modelos estocásticos**

## **Determinación de la base de datos**

Nuestro interés surgió al encontrar la base de datos llamada *“Million Dataset”*, que inicialmente fue nuestra primera opción. Sin embargo, notamos que estaba actualizada solo hasta el año 2011. Buscamos en Kaggle una alternativa más reciente y encontramos otra base de datos, que, si bien incluía canciones más actuales, carecía de suficientes métricas de popularidad como vistas en YouTube, likes y, sobre todo, el año de lanzamiento, una variable que considerábamos fundamental para nuestro análisis.

Posteriormente, encontramos una segunda base que sí contenía esas métricas faltantes, pero no incluía características musicales como bailabilidad, energía, key (llave), duración, entre otras. Por lo tanto, decidimos combinar ambas bases de datos para lograr un conjunto más completo y representativo.

Cabe destacar que ambas bases están en inglés y, aunque no son muy extensas, contienen canciones que se considerarían populares, abarcando un rango de años desde 1987 hasta 2024.

## **Preprocesamiento y limpieza**

Durante el proceso de combinación, surgieron algunos inconvenientes. Ambas bases incluían una columna llamada "popularity", pero cada una la calculaba de forma distinta, lo que generaba incoherencias al momento de graficar los datos. Además, al unirlas, algunas canciones se repitieron.

Como solución, eliminamos una de las columnas de popularidad y realizamos un proceso de limpieza, eliminando todos los registros duplicados para asegurar la coherencia del análisis.

## **Selección de variables**

Para abordar el objetivo del proyecto, se seleccionaron las variables más relevantes del conjunto de datos. Se identificó la popularidad como la variable dependiente principal, representando el éxito de la canción. Como variables independientes, se eligieron atributos musicales como bailabilidad, energía, valence, duración, entre otros, además de métricas complementarias como vistas y likes en YouTube. Estas variables fueron seleccionadas por su potencial influencia en el desempeño de una canción y su utilidad para responder la pregunta de investigación.

Las variables de la base de datos son:

|  |  |
| --- | --- |
| Variable | Significado |
| Artist Name | Nombre del artista |
| Track Name | Nombre de la canción |
| popularity | Un valor de 0 a 100 que indica qué tan popular es la canción en Spotify (basado en reproducciones recientes, tendencias, etc.). |
| danceability | De 0.0 a 1.0. Indica qué tan adecuada es la canción para bailar. Combina ritmo, tempo, regularidad de beats, etc. |
| energy | De 0.0 a 1.0. Mide la intensidad y actividad de la canción. Más energía = más fuerte y rápida (como el rock o EDM). |
| key | Representa la **clave musical** (de 0 a 11). Ej: 0 = C (Do), 1 = C#/Db, … 11 = B (Si). |
| loudness | Volumen promedio de la canción en decibelios (dB). Cuanto más negativo, más baja. Va de aprox. -60 a 0. |
| mode | Si la canción está en **modo mayor (1)** o **modo menor (0)**. Mayor = más alegre, menor = más melancólica. |
| speechiness | De 0.0 a 1.0. Mide la presencia de voz hablada. Rap tiene valores altos, música instrumental valores bajos. |
| acousticness | De 0.0 a 1.0. Qué tan acústica es la canción. Canta autor, guitarra acústica, etc. tienen valores altos. |
| instrumentalness | De 0.0 a 1.0. Probabilidad de que la canción **no tenga voces**. Música instrumental = alto valor. |
| liveness | De 0.0 a 1.0. Estima si fue grabada en vivo. Conciertos en vivo = valores más altos (>0.8). |
| valence | De 0.0 a 1.0. Mide cuán **positiva o feliz** se siente la canción. Alto = alegre, bajo = triste o tensa. |
| tempo | La velocidad de la canción, medida en **BPM (beats por minuto)**. Más alto = más rápida. |
| duration\_ms | Duración total de la canción en milisegundos. Puedes dividir entre 1000 para obtener segundos. |
| time\_signature | El compás musical. 4 = compás 4/4 (el más común), 3 = 3/4 (como un vals), etc. |
| Class | Genero de las canciones. Van desde el 1 al 10, cada numero un genero diferente.   1. Folk 2. Alt music 3. Blues 4. Bollywood 5. Country 6. Hip Hop 7. Indie 8. Instrumental 9. Metal 10. Pop 11. Rock |
| Release year | Año en que salió la canción. |
| YT\_Likes | Likes de la canción en youtube |
| YT\_Views | Las vistas de la canción en youtube. |

Para análisis más concluyentes dividimos las variables en dos campos.

Las variables independientes que serían las características musicales:

* danceability
* valence
* energy
* duration\_ms
* tempo
* acousticness
* instrumentalness
* liveness
* speechiness
* key, mode, loudness

Con las variables dependientes que serían las métricas de éxito:

* popularity
* views
* likes

## **Selección de Modelos**

**Gráficos de pilas**

Se utilizó un gráfico de barras verticales para visualizar el promedio de popularidad de los artistas que tenían más de tres canciones en el conjunto de datos.  
Este modelo permitió filtrar artistas con representación suficiente (para evitar sesgos por pocos datos) y observar qué artistas tienden a producir canciones con mayor éxito enpromedio. La visualización sirvió para identificar artistas con alta consistencia en popularidad, como Doja Cat u Olivia Rodrigo, lo que apoyó el análisis posterior sobre la influencia del artista en la popularidad.

**Gráficos de dispersión (scatter plots)**

Se utilizaron gráficos de dispersión para analizar la relación entre variables numéricas, como loudness, danceability, energy y la popularidad. Estos modelos permitieron visualizar si existía o no una correlación entre los atributos musicales de una canción y su nivel de popularidad. Aunque en algunos casos la correlación fue débil, el análisis fue clave para descartar o confirmar posibles influencias.

**Gráfico de líneas**

Se utilizó un gráfico de líneas para analizar la evolución temporal del promedio de popularidad de las canciones por año de lanzamiento. Este modelo permite observar tendencias a lo largo del tiempo, identificar picos o caídas en la popularidad promedio y explorar posibles cambios en la industria musical o en las preferencias del público. Es especialmente útil para detectar patrones de comportamiento estacional o históricos, como el descenso abrupto en cierto año o la estabilidad previa a los 2000.

# **Conclusiones**

Podemos concluir que no existe un conjunto de características específicas que garantice que una canción será popular. Sin embargo, se ha identificado que el artista tiene un impacto considerable en la popularidad de una canción. Esto puede explicarse por la presencia de una base de fans leales que siguen activamente la trayectoria del artista, así como por la fuerza de su marca personal, lo cual genera una mayor exposición y receptividad ante cada nuevo lanzamiento. En muchos casos, el nombre del artista impulsa la curiosidad de los oyentes que desean “estar a la moda” o al tanto de las tendencias.

Por otro lado, características acústicas como la bailabilidad, la energía o ruido no son determinantes por sí solas. Una canción lenta puede volverse icónica, mientras que una canción energética no necesariamente tendrá éxito. Estos atributos pueden acompañar a una canción popular, pero no la definen como tal.

Un hallazgo relevante de este estudio es que, si bien muchos factores intervienen en el éxito de una canción, el **artista** se posiciona como uno de los elementos con mayor peso. Aunque su impacto no es absoluto ni exclusivo, sí representa una influencia significativa que no debe pasarse por alto al analizar la popularidad musical.

# **Recomendaciones y futuros estudios**

A partir de los hallazgos de este estudio, se sugieren las siguientes recomendaciones prácticas:

* Contar con una base de datos individual y completa que no requiera modificaciones posteriores, lo cual facilitaría el análisis y la replicabilidad del estudio.
* Incorporar opiniones reales de los usuarios, ya que, en muchas ocasiones, las canciones se repiten varias veces, lo que puede hacer que el número de reproducciones no refleje necesariamente una valoración positiva o un gusto genuino.

Durante el desarrollo del proyecto surgieron diversas ideas que, debido al nivel de conocimiento y al tiempo disponible, no fue posible implementar. Por ello, para futuros estudios, ya sea individuales o en grupo, se podrían explorar temas como:

* ¿De qué manera influyen las canciones en la salud mental? ¿Pueden mejorar el estado de ánimo o, por el contrario, afectarlo negativamente?

En este sentido, encontramos una base de datos relacionada que nos habría servido para profundizar en esta línea, pero no se logró alcanzar un consenso claro para su inclusión en el presente estudio

# **Bibliografía**

*Music Genre Classification*. (2021, August 7). <https://www.kaggle.com/datasets/purumalgi/music-genre-classification/data>

*Most streamed Spotify songs 2024*. (2024, June 15). <https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-songs-2024?resource=download>

JulCsc. (n.d.). *Documentación de guía de Power BI - Power BI*. Microsoft Learn. <https://learn.microsoft.com/es-es/power-bi/guidance/>

Li, H., Fei, X., Yang, M., & Chao, K. (2020). E3MSD – a new music information retrieval architecture for an original music identifier. *ResearchGate*. <https://www.researchgate.net/publication/344658704_E3MSD_-_A_New_Music_Information_Retrieval_Architecture_for_an_Original_Music_Identifier>

Islam, M. (2024, August 2). *How to Use COUNTIFS Function with Multiple Criteria in Excel?* ExcelDemy. <https://www.exceldemy.com/countifs-with-multiple-criteria/?utm_source=>