

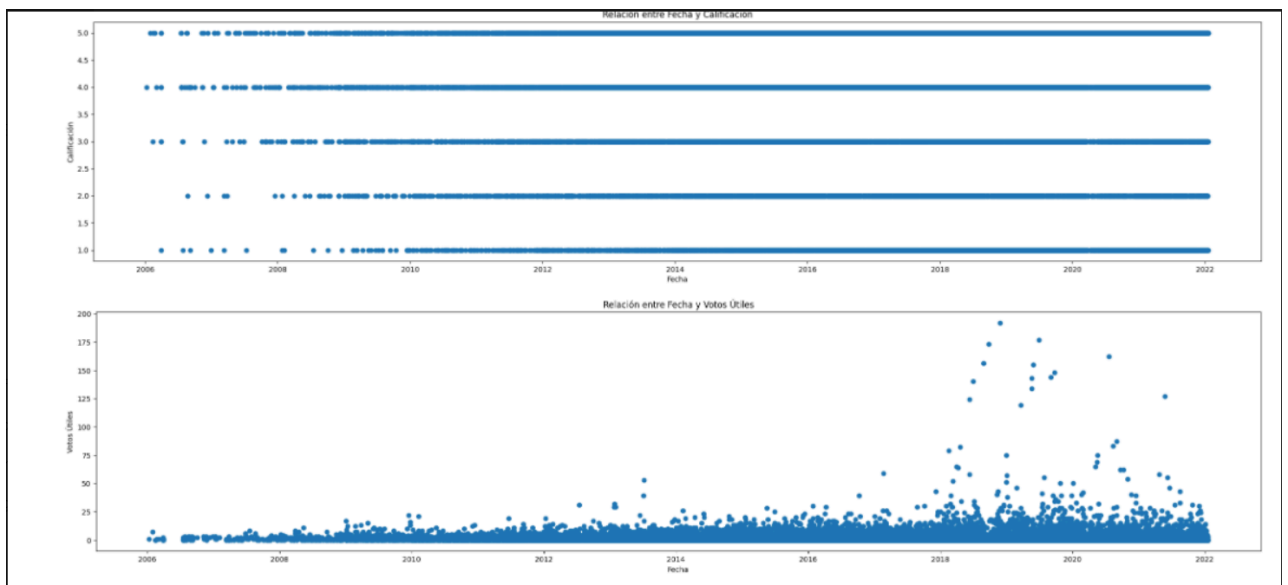
## Informe sobre la Calidad de los Datos - Análisis Exploratorio de Datos (EDA)

El análisis exploratorio de datos (EDA) es una fase crucial para comprender la calidad y fiabilidad de los datos antes de realizar análisis más profundos o desarrollar modelos predictivos. Durante esta fase, se evaluaron varios aspectos fundamentales que impactan en la calidad de los datos:

1. **Valores Faltantes:** Se detectaron columnas con datos faltantes en varias variables. Estos valores ausentes pueden distorsionar los resultados de los modelos de análisis si no se tratan correctamente. Dependiendo del contexto, se utilizarán técnicas de imputación como la media, la mediana o modelos predictivos, o bien, la eliminación de filas o columnas.
2. **Duplicados:** Se identificaron registros duplicados, lo cual es un problema común que puede generar sesgos en los análisis, especialmente si las filas repetidas corresponden a entradas similares. Se procederá a eliminar estos duplicados para garantizar la integridad de los datos.
3. **Outliers (Valores atípicos):** Se analizaron los valores extremos que podrían afectar los resultados. Estos outliers deben evaluarse cuidadosamente, ya que pueden indicar errores de captura o valores legítimos que, si no se tratan adecuadamente, sesgan las estadísticas descriptivas.
4. **Consistencia y Formato:** Se observó que varias columnas presentan inconsistencias en sus formatos (por ejemplo, diferentes unidades de medida o tipos de datos en una misma columna). Para mejorar la calidad de los datos, es necesario unificar los formatos y estandarizar las unidades de medida.
5. **Distribución de los Datos:** Se evaluó la distribución de las variables principales para entender su comportamiento y determinar si se requieren transformaciones adicionales para abordar posibles sesgos en la distribución de los datos.
6. **Correlación entre Variables:** Se analizaron las correlaciones entre variables clave para identificar posibles relaciones lineales o no lineales que podrían influir en los análisis o en el diseño de los modelos predictivos.

**Conclusión y Recomendaciones:** Para asegurar que el dataset sea adecuado para el análisis y modelado, es crucial aplicar técnicas de limpieza de datos adicionales. Esto incluye la imputación de valores faltantes, la eliminación de duplicados, la transformación de outliers y la normalización de los formatos de los datos. Solo así se

podrá proceder a las siguientes etapas del análisis con un conjunto de datos limpio y coherente.



#### 📊 Gráfica superior: Relación entre Fecha y Calificación

Esta gráfica muestra la evolución de las calificaciones a lo largo del tiempo. Se observa que las calificaciones se mantienen distribuidas en diferentes niveles (1 a 5) a través de los años, pero la densidad en ciertos valores parece consistente, indicando una estabilidad en la distribución de calificaciones.

#### En relación con los KPIs:

- *Satisfacción mensual del usuario*: Si se agrupa esta información por períodos (por ejemplo, trimestral o anual), podrías obtener un índice de satisfacción promedio en cada período y observar su evolución. Si la proporción de calificaciones altas (4 o 5) ha aumentado, esto podría indicar una mejora en la satisfacción.
- *Índice de sentimiento positivo*: Al observar la tendencia de las calificaciones altas (superiores a 3.5), podrías evaluar si el sentimiento positivo ha mejorado en comparación con años anteriores.

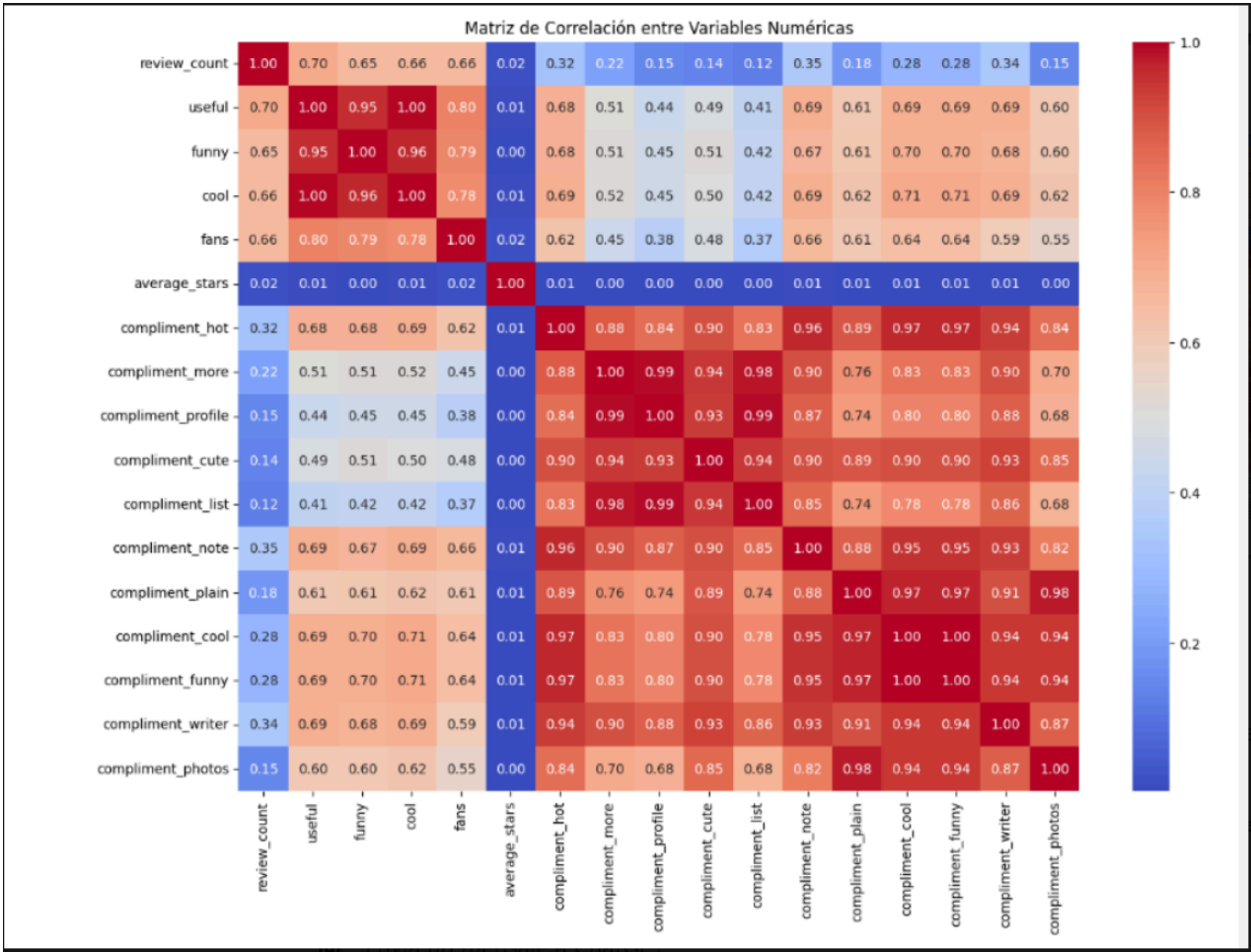
#### 📊 Gráfica inferior: Relación entre Fecha y Votos Útiles

La segunda gráfica muestra cómo los votos útiles (probablemente en reseñas) han

cambiado con el tiempo. Se observa un aumento en la cantidad de votos útiles en ciertos años, especialmente a partir de 2016 y alcanzando picos hacia 2019-2020.

**En relación con los KPIs:**

- *Tasa de crecimiento de reseñas positivas y negativas:* Si los votos útiles están correlacionados con reseñas positivas o de alta calidad, este aumento podría indicar una mayor participación y aceptación de los usuarios. Un análisis de crecimiento en votos útiles junto con el tipo de reseña (positiva o negativa) puede proporcionar insights sobre la percepción de los usuarios.
- *Índice de sentimiento positivo:* Un incremento en votos útiles para reseñas positivas podría sugerir que más usuarios encuentran útiles las experiencias positivas, lo cual impacta en el índice de sentimiento positivo.



**1. Satisfacción mensual del usuario:**

- o *Average Stars*: La correlación de `average_stars` con otras variables es muy baja, lo que indica que la satisfacción medida en términos de puntaje promedio de estrellas no tiene una relación fuerte con la cantidad de reseñas (`review_count`) u otros aspectos como `useful`, `funny` o `cool`. Esto podría sugerir que factores externos (como el contenido o calidad del servicio/producto) tienen un impacto más significativo en la satisfacción y no tanto el volumen de interacciones.

## 2. Tasa de crecimiento de reseñas positivas y negativas trimestral:

- o La variable `review_count` tiene una correlación positiva moderada a alta con variables como `useful`, `funny`, `cool` y `fans`, lo cual indica que los negocios con más reseñas suelen recibir también más reacciones y ser considerados útiles o entretenidos. Esta correlación sugiere que un crecimiento en reseñas podría venir acompañado de un aumento en las interacciones, lo cual puede ayudar a monitorear cambios en la satisfacción del cliente trimestralmente.
- o Si observas un incremento en `review_count` junto con aumentos en `funny` o `cool`, podrías analizar si estas interacciones están relacionadas con reseñas positivas o negativas, y cómo varía esta proporción trimestralmente.

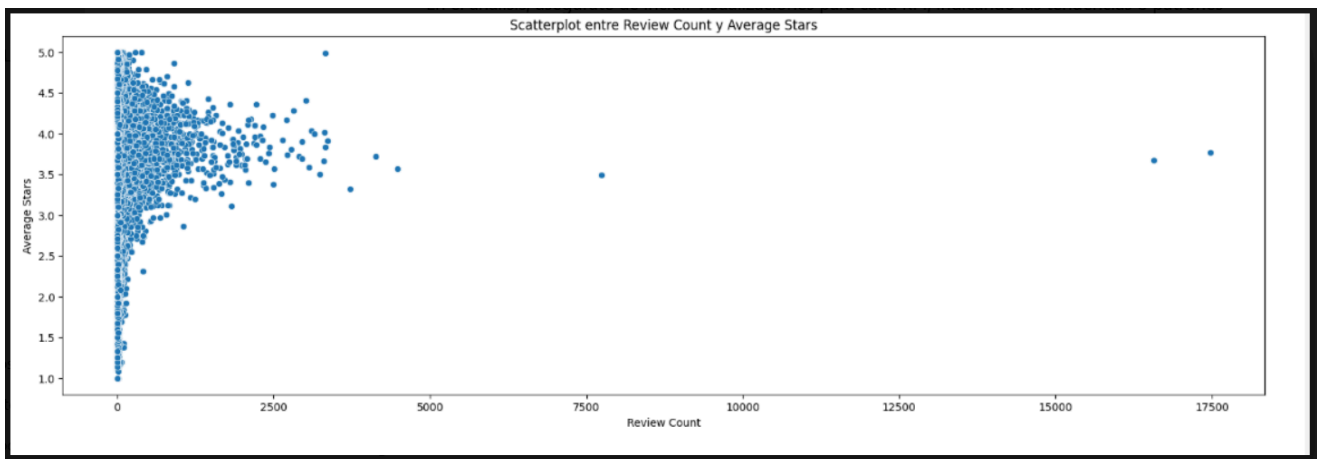
## 3. Índice de sentimiento positivo superior a 3.5 respecto al año anterior:

- o Aunque `average_stars` no muestra una correlación fuerte con las demás variables, sí podrías considerar el conjunto de variables de "compliment" (`compliment_hot`, `compliment_cool`, etc.), que tienen una fuerte correlación entre sí. Esto sugiere que cuando un usuario hace un cumplido, es probable que esté satisfecho, y esto podría ser un indicador positivo en el sentimiento general.
- o Analizar el crecimiento de estas variables de cumplidos respecto al año anterior puede ayudar a entender si el índice de sentimiento positivo está mejorando o si hay una necesidad de intervención.

## Conclusiones

Esta matriz de correlación indica que:

- La satisfacción del usuario en términos de estrellas no está directamente vinculada a la cantidad de reseñas o a otras métricas de interacción, sugiriendo que puedes enfocar las mejoras en otros aspectos del producto/servicio.
- La tasa de crecimiento de reseñas y cumplidos podría ser una buena señal indirecta de satisfacción.
- Un análisis temporal de los cumplidos y reseñas te permitirá evaluar si el sentimiento positivo está aumentando año a año, lo cual es relevante para tus KPIs.



Este scatterplot muestra la relación entre la cantidad de reseñas (eje x) y el puntaje promedio (eje y), en relación a los KPIS

1. **Concentración en Reseñas con Pocas Opiniones:** La mayoría de los puntos están agrupados a la izquierda, donde el número de reseñas es bajo, lo cual indica que muchos negocios o productos reciben pocas opiniones. Esto podría impactar en la **satisfacción mensual del usuario**, ya que aquellos con pocas reseñas podrían no representar bien la experiencia real.
2. **Puntaje Alto con Pocas Reseñas:** Se observa que muchos negocios con pocas reseñas tienden a tener puntajes altos (4-5 estrellas). Esto puede sesgar el **índice de sentimiento positivo**, especialmente si solo un grupo pequeño de usuarios está calificando positivamente.
3. **Distribución de Altos Conteos de Reseñas:** Los puntos aislados en el lado derecho muestran negocios con muchas reseñas. Si estos tienen puntajes consistentes, podrían estar contribuyendo significativamente al KPI de **tasa de crecimiento de reseñas positivas y negativas**.

4. **Patrones en Reseñas Muy Populares:** Los pocos puntos con altos conteos de reseñas (de más de 5000, por ejemplo) sugieren que estos negocios o productos son muy populares. Sería interesante observar si estos puntos tienen variaciones en los trimestres para ver su influencia en el crecimiento de reseñas.

Este gráfico puede ayudarte a detectar áreas en las que podrías mejorar la satisfacción y monitorear tendencias en reseñas positivas y negativas.