AED

Carlos Galan, Lucia Ardila, Valentina Lopez

Table of Contents

Introduccion

Para las aerolíneas, no solo es importante ofrecer un buen servicio, sino también entender las opiniones de sus usuarios. A través de las reseñas, pueden identificar áreas de mejora, analizar comentarios positivos y negativos, y evaluar el impacto de las calificaciones que reciben. En este informe, exploraremos una base de datos de reviews dejadas por los

pasajeros, con el objetivo de comprender mejor la percepción del servicio ofrecido por distintas aerolíneas.

La base de datos utilizada en este proyecto proviene de Kaggle:https://www.kaggle.com/datasets/juhibhojani/airline-reviews, y originalmente no contenía información cuantitativa continua. Por esta razón, se decidió incorporar columnas adicionales como vader_pos, roberta_pos, vader_neg, entre otras, generadas mediante modelos de análisis de sentimientos como VADER y Roberta. Estos modelos permiten capturar de manera precisa el tono emocional de las reseñas, agregando valor al análisis al ofrecer una interpretación más detallada de las experiencias de los usuarios y su influencia en las calificaciones generales. Sin embargo, cabe resaltar que, aunque estas variables aportan contexto, el enfoque central del proyecto es estrectamente estadístico, no de análisis de sentimientos.

Diccionario de variables

- **vader_neg**: Variable cuantitativa continua que indica la proporción de negatividad en un comentario según el modelo VADER, con un rango de 0 a 1.
- **vader_neu**: Variable cuantitativa continua que refleja la proporción de neutralidad de un comentario basado en el modelo VADER; valores cercanos a 1 indican ausencia de emociones marcadas.
- **vader_pos**: Variable cuantitativa continua que mide la proporción de positividad en un comentario usando el modelo VADER, donde valores más altos indican un tono predominantemente positivo.
- **vader_compound**: Variable cuantitativa continua que representa un índice de sentimiento en un rango de -1 a 1, calculado por VADER.
- **roberta_neg**: Variable cuantitativa continua que mide la proporción de negatividad en un comentario según el modelo Roberta, con un rango de 0 a 1.
- **roberta_neu**: Variable cuantitativa continua que indica la proporción de neutralidad de un comentario usando el modelo Roberta; valores cercanos a 1 sugieren una ausencia de emociones marcadas.
- **roberta_pos**: Variable cuantitativa continua que refleja la proporción de positividad en un comentario basado en el modelo Roberta, con valores más cercanos a 1 señalando un tono positivo.
- **Airline Name**: Variable cualitativa nominal que indica el nombre de la aerolínea sobre la cual se realiza la reseña.
- **Overall_Rating**: Variable cuantitativa discreta que representa la calificación global otorgada por el usuario a la experiencia de vuelo, con un rango de 1 a 5.
- **Verified**: Variable cualitativa binaria que indica si el usuario que escribió la reseña fue verificado como pasajero real del vuelo.
- **Review**: Variable cualitativa textual que contiene el texto de la reseña que describe la experiencia del usuario.
- **Type of Traveller**: Variable cualitativa nominal que clasifica al usuario según el tipo de viaje realizado.

- **Seat Type**: Variable cualitativa ordinal que indica el tipo de clase en la que viajó el pasajero.
- **Seat Comfort**: Variable cuantitativa discreta que representa la puntuación otorgada por el usuario a la comodidad del asiento, en una escala de 1 a 5.
- **Cabin Staff Service**: Variable cuantitativa discreta que califica el servicio brindado por el personal de cabina, en una escala de 1 a 5.
- **Food & Beverages**: Variable cuantitativa discreta que evalúa la calidad de la comida y las bebidas ofrecidas durante el vuelo, en una escala de 1 a 5.
- **Value for Money**: Variable cuantitativa discreta que mide la percepción del usuario sobre la relación calidad-precio del vuelo.
- **Recommended**: Variable cualitativa binaria que indica si el usuario recomienda o no la aerolínea.
- **Review Length**: Variable cuantitativa discreta que mide la longitud del texto de la reseña.
- **Subjectivity**: Variable cuantitativa continua que representa la proporción de subjetividad en el comentario, calculada utilizando el modelo TextBlob.Valores cercanos a 1 indican una opinión altamente personal(emocional), mientras que valores bajos sugieren un enfoque más objetivo.
- **Average_Rating** Variable cuantitativa continua que muestra el promedio de los ratings Seat Comfort,Cabin Staff Service, Food and Beverages y Value for Money para cada resgistro.

Estadisticas descriptivas

Insights relevantes.

- VADER, aunque eficiente y rápido, utiliza un enfoque basado en reglas y un diccionario léxico estático, lo que lo limita al procesar matices complejos como sarcasmo o emociones implícitas. Por otro lado, RoBERTa, basado en transformadores y aprendizaje profundo, interpreta el contexto lingüístico de manera robusta, siendo más adecuado para analizar sentimientos en textos complejos y sutiles.

Vader

- Sentimiento Neutro (vader_neu): La media de 0.808 muestra que la mayoría de las reseñas contienen una gran cantidad de contenido neutral. Esto sugiere que, aunque los clientes mencionan aspectos tanto positivos como negativos, una gran parte del contenido no refleja un sentimiento fuerte. Y esto pudo ser una gran dificultad para los algoritmos de clasificación de sentimiento.

- Sentimiento Negativo (vader_neg): Con una media de 0.093 y una mediana de 0.086, se puede concluir que el contenido negativo es bastante bajo, pero aún está presente.Ademas el máximo de 0.461 muestra que algunas reseñas tienen un contenido negativo significativo, pero este no es el caso general.

- Sentimiento Positivo (vader_pos): La media de 0.099 muestra que, en promedio, las reseñas tienen una pequeña proporción de contenido positivo, lo cual es relativamente bajo. El máximo de 0.555 indica que algunas reseñas son muy positivas, aunque representan una minoría tambien.

- VADER Compound (vader_compound): El promedio de -0.071 indica que el sentimiento general de las reseñas tiende a ser ligeramente negativo. Esto puede significar que la mayoría de las reseñas están más inclinadas hacia una insatisfacción leve o un sentimiento neutro-negativo. La mediana de -0.328 refuerza esta idea, ya que la mayoría de las puntuaciones de sentimiento compuesto tienden a estar en el rango negativo. Los extremos (máximo de 0.998 y mínimo de -0.997) indican que, aunque en general las reseñas son neutrales o negativas, existen algunas con sentimientos extremos, tanto positivos como negativos.

- Se observa que el analisis individual de cada variable de vader termina generalizandose en la variable vader compound la cual debe decidir donde clasificar el comentario a pesar de su falta de polaridad.

Roberta

- roberta_neg:Con una media de 0.586, RoBERTa detecta un mayor porcentaje de contenido negativo en comparación con VADER. Esto podría deberse a diferencias en cómo los dos modelos interpretan el tono de las reseñas. Pues roberta es mucho mas robusto y vader está diseñado para textos no muy complejos.

- roberta_neu: La media de 0.144 muestra una baja proporción de contenido neutral según RoBERTa, lo que es bastante diferente del análisis de VADER, que detectaba mucho más contenido neutral.

- roberta_pos: La media de 0.270 indica que RoBERTa encuentra más contenido positivo en comparación con VADER, pero aún es menos común que el contenido negativo. Ademas el máximo de 0.993 sugiere que algunas reseñas son extremadamente positivas, similar a lo observado en VADER.

- Las valoraciones generales son bajas, y una gran parte de los usuarios no recomendarían la aerolínea.

- Las reseñas son en su mayoria largas y subjetivas.

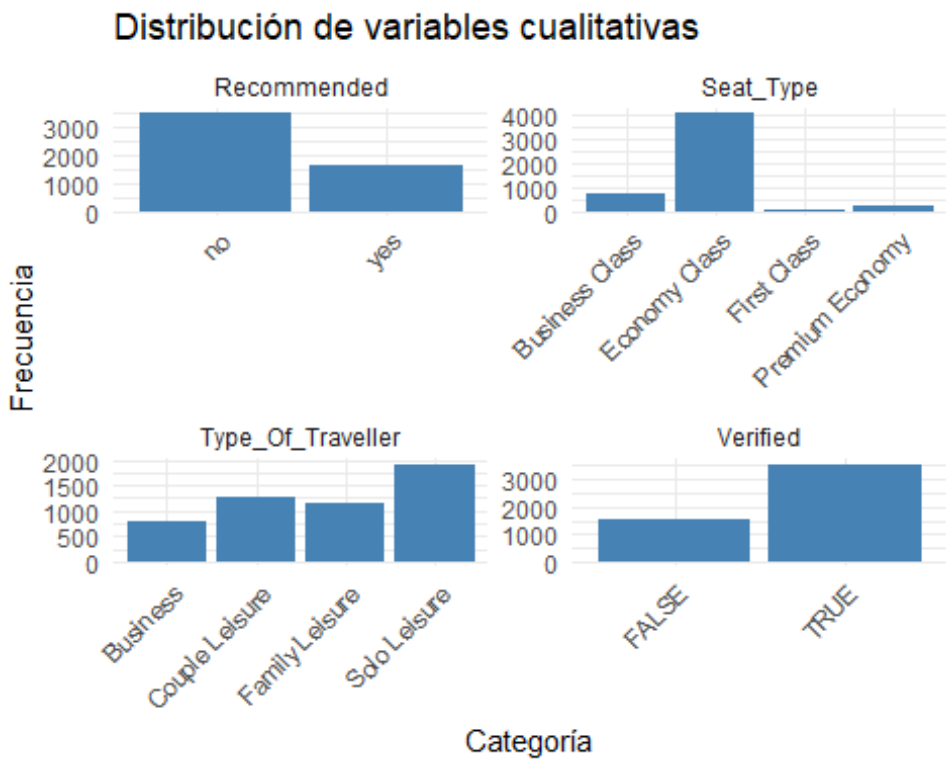
- VADER tiende a clasificar muchas reseñas como neutrales y tiene dificultades para capturar sentimientos más complejos, especialmente en reseñas largas o complejas. Esto lo hace menos preciso en este tipo de textos.

- RoBERTa, por otro lado, detecta una mayor proporción de contenido negativo y positivo, con menos contenido neutral, lo que lo convierte en un modelo más robusto para analizar reseñas que contienen emociones más complejas.

Graficas

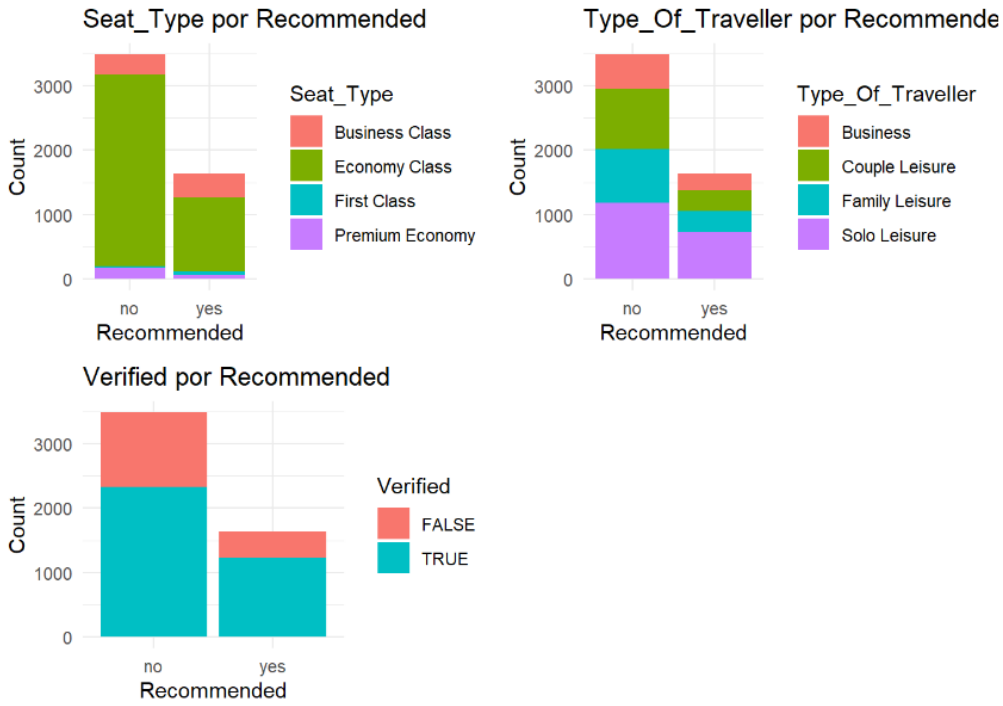
Cualitativas

Diagramas de barras



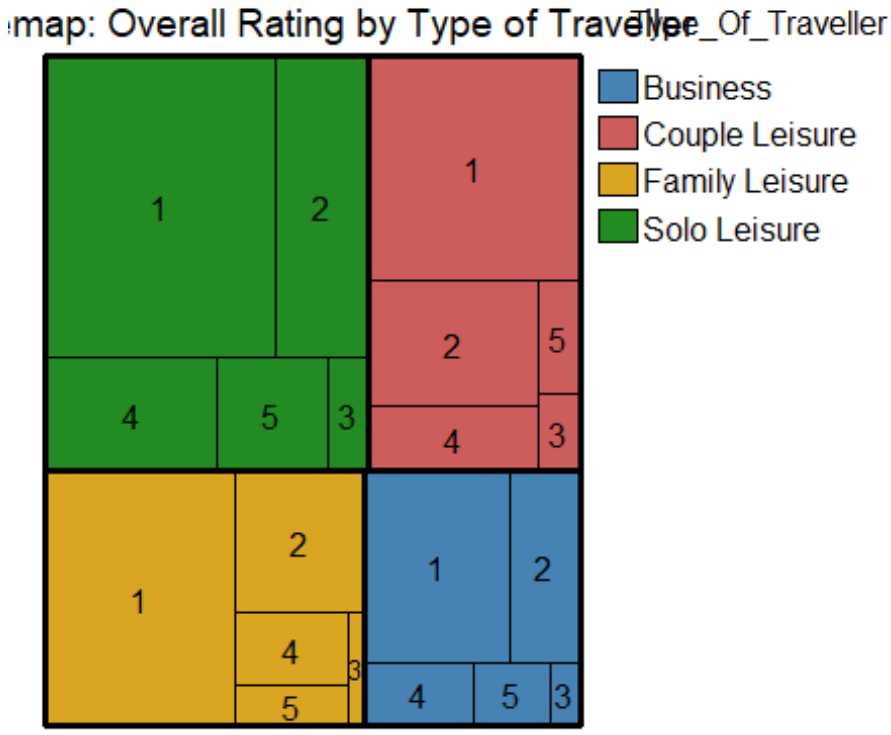
- En la primera grafica muestra que la mayoría de los usuarios no recomendaron el servicio, por lo que en teoria los modelos de predicción de vader y roberta deberían haber clasificado un sentimiento en su mayoría negativo.
- El gran sesgo hacia la “Economy Class” es coherente puesto que es la clase con mas limitaciones debido a su precio.
- La predominancia de viajeros especialmente los que viajan solos, puede sugerir que el servicio tiene más demanda entre turistas o viajeros ocasionales, en lugar de ejecutivos o viajeros de negocios. Esto puede ser de para estrategias de segmentación de mercado.
- la mayoría de las opiniones en este conjunto de datos son de usuarios cuya identidad o compra ha sido confirmada, lo que podría darle más credibilidad al análisis de las opiniones.

diagramas de barras apilados

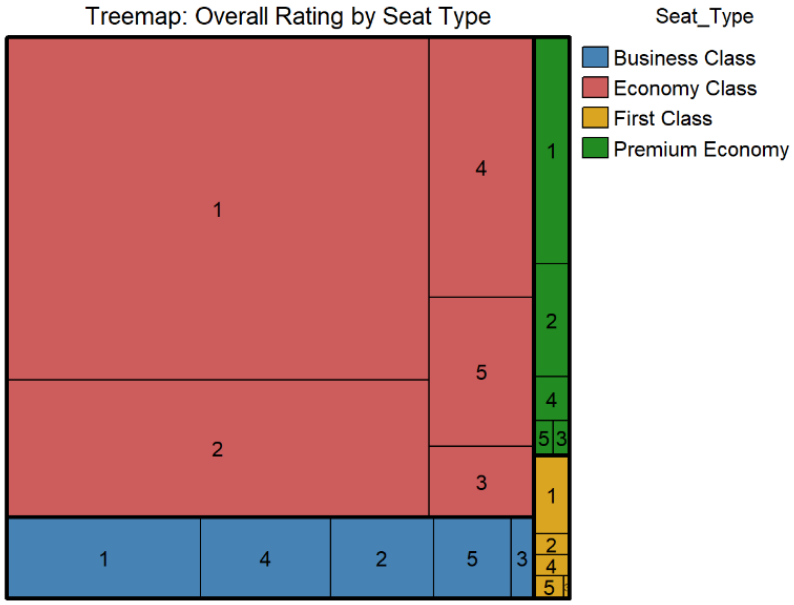


- Los pasajeros de Economy Class y los viajeros Solo Leisure son menos propensos a recomendar el servicio, mientras que los viajeros de Business Class y los verificados tienen una mayor probabilidad de dar recomendaciones positivas.

Treemaps



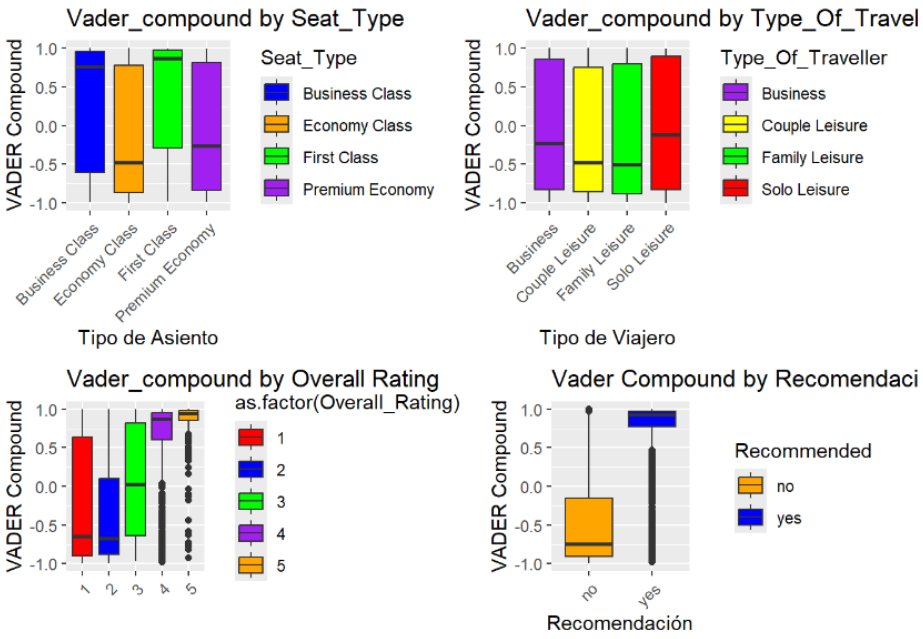
- El treemap muestra que los viajeros solos y las familias tienen una tendencia a calificaciones más bajas, mientras que los viajeros de negocios y las parejas tienen experiencias más variadas.



- Los pasajeros en Economy Class tienden a dar las calificaciones más bajas, lo cual podría reflejar una menor satisfacción en comparación con otras clases, mientras que Business Class muestra una experiencia más equilibrada.

Boxplots

Boxplots comparando vader_compound con cuantitativas



1. Distribución de VADER Compound por Tipo de Viajero:

- Family Leisure parece tener una distribución más amplia, con valores más dispersos hacia el lado positivo y ligeramente menos hacia el negativo, sugiriendo que los viajes en familia pueden generar una mayor variedad de experiencias emocionales.

- Solo Leisure y Couple Leisure muestran distribuciones similares, con la mayoría de los valores concentrados en la parte positiva del VADER Compound. Los viajeros solos y en pareja parecen tener una experiencia más homogénea, con emociones predominantemente positivas.

2. Distribución de VADER Compound por Overall Rating:

- Las calificaciones más bajas, como 1 y 2, tienen una mayor proporción de sentimientos negativos, ya que sus distribuciones se inclinan hacia valores bajos (incluso negativos) en el VADER Compound.
- A medida que sube la calificación (de 3 a 5), la proporción de emociones positivas aumenta, como es de esperarse, y la distribución se desplaza hacia valores más altos en el eje del VADER Compound.

3. Distribución de VADER Compound por Recomendación:

- Los usuarios que no recomiendan la aerolínea muestran una tendencia hacia valores negativos en el VADER Compound, lo cual tiene sentido, ya que se esperaría que las emociones en estas reseñas fueran predominantemente negativas

Boxplots comparando Average_Rating con cuantitativas

1. Distribución de Average Rating por Tipo de Asiento (Seat Type)

- Los pasajeros en clases más caras como Business y First Class tienden a otorgar calificaciones más altas, lo que podría indicar una mejor experiencia en esos tipos de asiento en comparación con Economy Class.

2. Distribución de Average Rating por Tipo de Viajero

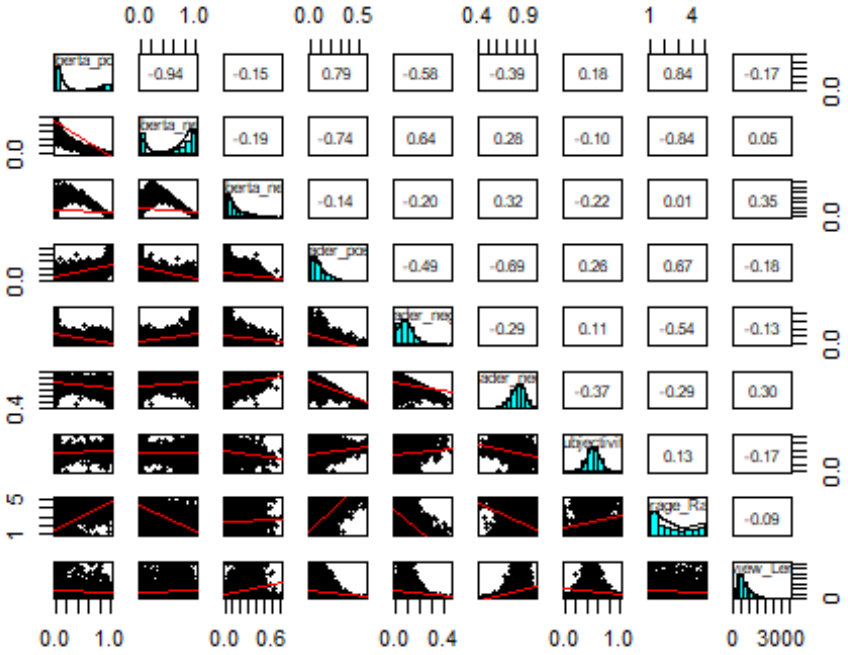
Los viajeros de negocios y aquellos que viajan solos parecen más satisfechos en general, mientras que aquellos que viajan en pareja o en familia podrían tener experiencias más diversas, lo que se refleja en calificaciones más variadas.

3. Distribución de Average Rating por Puntuación Total

- La calificación promedio está alineada con la puntuación global que los usuarios dan a la experiencia, lo cual es esperable. Esto sugiere consistencia en la percepción general del servicio y las subcategorías evaluadas.

4. Distribución de Average Rating por Recomendación Como es lógico, quienes recomiendan la aerolínea han tenido una experiencia más satisfactoria en general, reflejada en su Average Rating elevado.

Cuantitativas



- Gráficos de dispersión y coeficiente de pearson:

1. roberta_pos y roberta_neg (-0.94): Existe una fuerte correlación negativa entre las proporciones de comentarios positivos y negativos según el modelo Roberta, lo que tiene sentido dado que a mayor positividad, menor negatividad en los comentarios, y viceversa.

2. roberta_pos y Overall_Rating (0.84): Fuerte correlación positiva entre la positividad medida por Roberta y la calificación general (Overall Rating). Esto indica que los comentarios positivos tienden a estar relacionados con calificaciones más altas.

3. vader_neg y Overall_Rating (-0.54): La negatividad según VADER tiene una correlación negativa moderada con las calificaciones globales, similar a la tendencia observada con Roberta.

4. Los gráficos entre roberta_pos y vader_pos (0.79) muestran una clara relación directamente proporcional, confirmando que ambos modelos tienden a estar de acuerdo al medir positividad. Sin embargo en cuanto a vader_neg y robarta_neg no se evidencia alguna relación, mostrando que ambos modelos si suelen diferir en los resultados (al ser roberta un modelo mas robusto que vader).

5. Adicionalmente entre roberta_neu y vader_neu (0.32) la dispersion de puntos es aleatoria pues los comentarios neutrales no deberian seguir alguna tendencia.

6. Contrastando la subjetividad con vader neu y roberta_neu se observa que se logra observar una relacion inversa, indicando que en algunos casos los comentarios mas neutrales son los que menos emociones sucitan y mas objetivos son.

7. En su mayoria las variables tienen una distribución asimétrica y parecen sesgadas hacia un lado (positivamente o negativamente). Además, a simple vista se observa que la variable continua 'subjectivity' podría seguir una distribución normal.

Diferencia de medias

Evaluar si los modelos son significativamente diferentes haciendo una prueba de hipotesis para las diferencias de medias entre (vader_pos,roberta_pos),(vader_neu,roberta_neu),(vader_neg,roberta_neu)

1. Prueba t para vader_pos vs roberta_pos
- Hipótesis nula (H_0): No hay diferencia en las medias de vader_pos y roberta_pos ($\mu_1 - \mu_2 = 0$).
- Hipótesis alternativa (H_1): Hay una diferencia en las medias de vader_pos y roberta_pos ($\mu_1 - \mu_2 \neq 0$).
2. Prueba t para vader_neu vs roberta_neu
- Hipótesis nula (H_0): No hay diferencia en las medias de vader_neu y roberta_neu ($\mu_1 - \mu_2 = 0$).
- Hipótesis alternativa (H_1): Hay una diferencia en las medias de vader_neu y roberta_neu ($\mu_1 - \mu_2 \neq 0$).
3. Prueba t para vader_neg vs roberta_neg
- Hipótesis nula (H_0): No hay diferencia en las medias de vader_neg y roberta_neg ($\mu_1 - \mu_2 = 0$).
- Hipótesis alternativa (H_1): Hay una diferencia en las medias de vader_neg y roberta_neg ($\mu_1 - \mu_2 \neq 0$).

Analisis

- Dado que todos los p-valores son muy bajos (menores que 0.05), puedes rechazar la hipótesis nula en todos los casos, lo que indica que hay diferencias significativas entre las medias de las comparaciones realizadas.

- Los intervalos de confianza te indican que la diferencia media entre las columnas de vader y roberta es significativa, con los intervalos no incluyendo cero.

Normalidad

Univariado(Se encuentra en el html)

- Prueba de normalidad univariada para cada variable usando Anderson-Darling debido a que la muestra es muy grande.

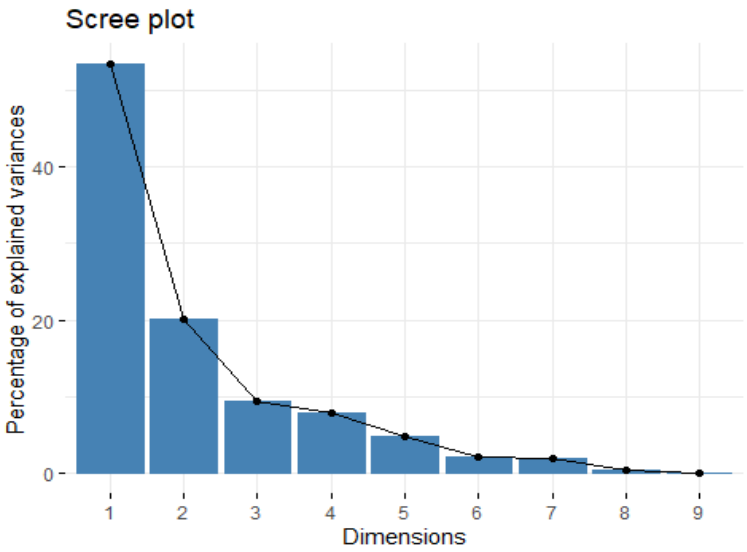
Bivariado (Se encuentra en el html)

Las variable NO se distribuye normal ni de forma univariada ni bivariada.

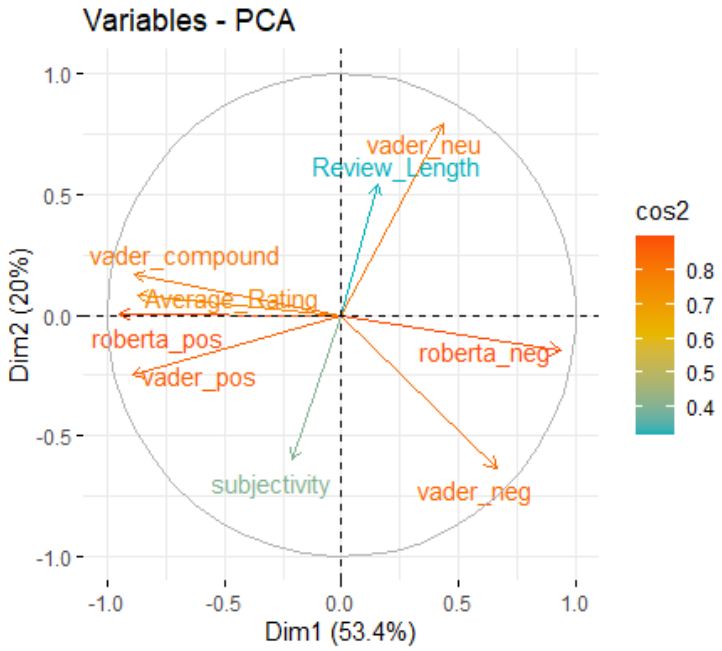
KMO y Barlet

- KMO dió un resultado de 0.7 lo cual indica que se pasó este test para hacer analisis de PCA para las variables elegidas.
- Es importante resaltar que se tuvo que quitar roberta neu, debido a que generaban valores bajos de KMO pues casi no hubo reviews en roberta que hayan tenido una representación alta de comentarios neutrales.
- Barlet pasó tambien la prueba pues rechazó la hipotesis nula de que son variables independeintes. Es decir las variables son significativamente dependientes y podemos usar PCA para reducir su dimensionalidad.

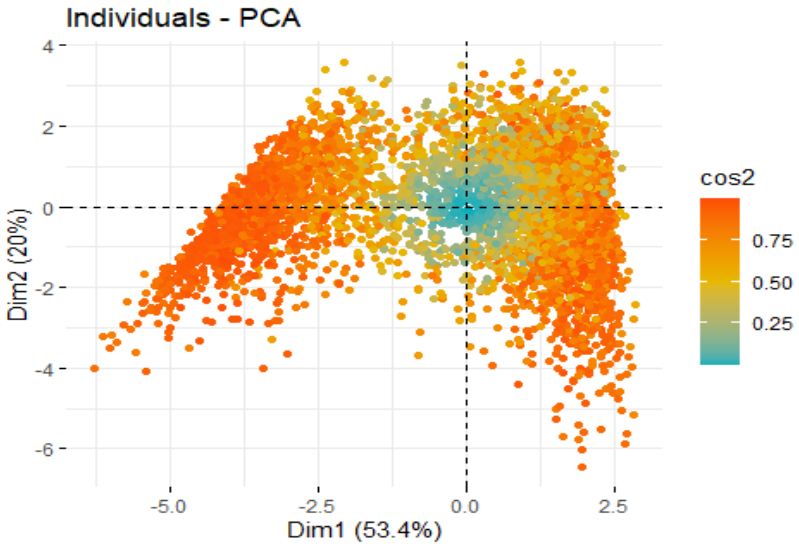
PCA



- Con base en el punto de codo, se consideró usar 2 componentes principales para capturar la mayor parte de la varianza y reducir la dimensionalidad en la que nos encontramos.



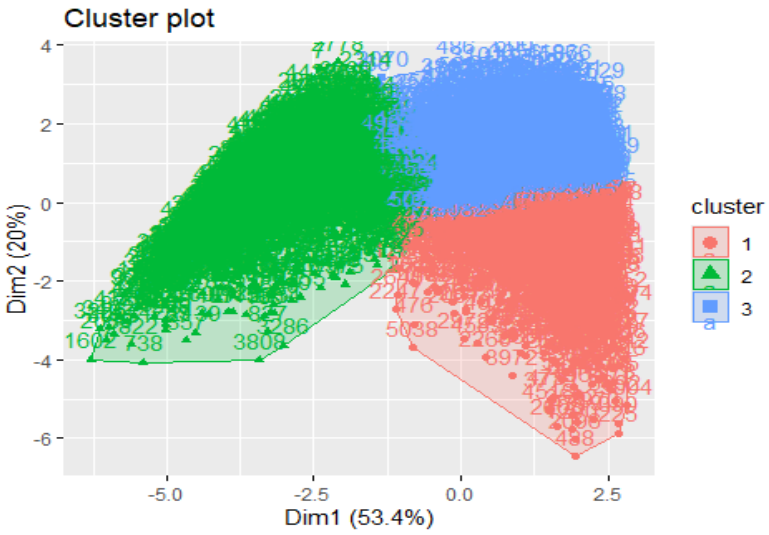
- Las variables positivas (vader_pos, roberta_pos) y las negativas (vader_neg, roberta_neg) están claramente separadas, indicando que los componentes principales capturan esta polaridad.
- Dim1 parece estar relacionado con la polaridad del sentimiento, mientras que Dim2 captura la neutralidad y la longitud del comentario.



- Los datos están separados en función de la polaridad (Dim1) y otras características secundarias como longitud y neutralidad (Dim2).
- La mayoría de los individuos están dispersos a lo largo de este eje(dim1), reflejando que la polaridad es la característica más importante en los datos.

K-means

Se quiso probar k-means para k=3 para mirar si podemos encontrar infomacion sobre la neutralidad de los comentarios.



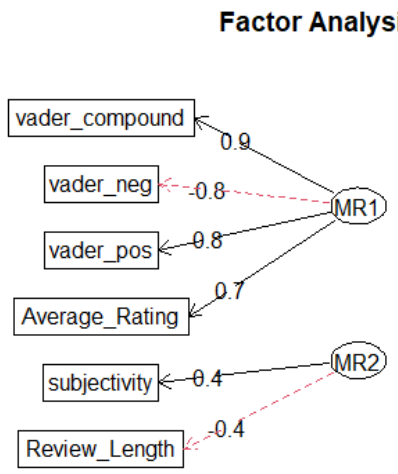
- Dim1: Es el principal diferenciador entre los clústeres 1 (positivo) y 3 (negativo).
- Dim2:Diferencia a los comentarios más neutros y largos, agrupados en el clúster 2 (verde).
- El clustering basado en los componentes principales logra una separación clara entre sentimientos positivos, negativos y neutros.

Analisis factorial

Se decidió hacer 2 modelos para analisis factorial:

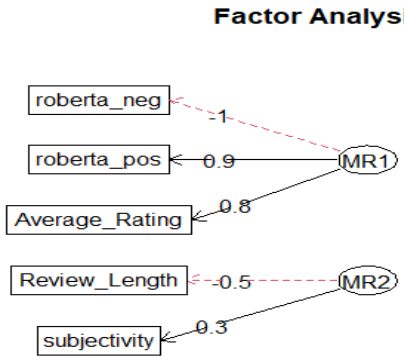
modelo 1: solo involucra variables respecto a Vader y otras variables relevantes para el analisis.

modelo 2: solo involucra variables respecto a Roberta y otras variables relevantes para el analisis.



Analisis facorial para el modelo VADER:

- Factor 1 (MR1): Representa el eje emocional (positividad/negatividad) que explica la percepción general y las calificaciones.
- Factor 2 (MR2): Captura características relacionadas con la subjetividad y la longitud de los comentarios.



- En este caso, RoBERTa muestra una mayor separación entre sentimientos positivos y negativos, con roberta_neg teniendo una carga perfecta (-1) en MR1.
- Las variables de subjetividad y longitud tienen menor influencia comparativa en RoBERTa que en VADER.

```
## Precisión del modelo QDA (Modelo 1): 0.9382
```

- El modelo tiene un mejor desempeño del esperado en clasificar correctamente la variable “Recomended”. Sin embargo, hay un número similar de errores en ambas direcciones (48 y 47), lo que indica un modelo equilibrado en términos de sensibilidad y especificidad.

- Esto indica que vader a pesar que no es un modelo muy robusto, lográ aportar infomación adecuada sobre los sentimientos positivos y negativos de reviews. Aunque es importante observar que conociamos las probabilidades a priori puesto que la base de datos está desequilibrada a los comentarios negativos.

```
## Precisión del modelo QDA (Modelo 2): 0.9479
```

Conclusiones

- Dado que las variables no pasaron las pruebas de normalidad (univariada y multivariada), fue necesario realizar un modelamiento complementario utilizando técnicas como análisis de componentes principales (PCA), análisis factorial, análisis de clústeres y análisis discriminante cuadrático (QDA). El objetivo principal de este enfoque fue evaluar el comportamiento de los modelos y su capacidad para clasificar adecuadamente las opiniones de los usuarios.

- En el análisis descriptivo, se observó que VADER, a pesar de ser un modelo menos robusto que RoBERTa, presentó dificultades para diferenciar entre comentarios positivos y negativos, clasificando una gran cantidad de reseñas como neutrales. Esto generaba dudas sobre su desempeño en tareas más complejas de clasificación.

- PCA permitió reducir la dimensionalidad del conjunto de datos al identificar los ejes principales que explican la mayor parte de la varianza de las variables originales. y observar la clara separación que lograron generar los modelos de vader y roberta sobre los reviews.

- Analisis factorial: Ambos análisis muestran que la polaridad (positividad/negatividad) es el factor más importante (MR1). El segundo factor (MR2) está relacionado con características textuales como longitud y subjetividad.

- Aunque RoBERTa es más robusto y redujo errores, los modelos basados en VADER se desempeñaron sorprendentemente bien, probablemente porque el conjunto de datos es explícito en términos de polaridad del sentimiento y calificaciones.

- En general, los análisis realizados permitieron confirmar que tanto VADER como RoBERTa son herramientas útiles, y que VADER puede ser adecuado para tareas donde la interpretación de sentimientos sea menos compleja. Esto resalta la importancia de complementar los análisis descriptivos con modelamientos más avanzados para evaluar objetivamente el rendimiento de los modelos.