

Entrega del Trabajo Final del Curso de R aplicado al Análisis Cualitativo

Nora Gesto

30/9/2021

Introducción

El presente documento se realiza en el marco del trabajo final del curso de R para el análisis cualitativo. Se presentará en diferentes apartados la aplicación de las variedad de técnicas vistas en el curso, manteniendo como hilo conductor el tema de las **estadísticas educativas**.

Procesamiento de archivos pdf desde la web

En primer instancia se utiliza la técnica de procesamiento de datos proveniente de archivos en pdf, digitalizados y extrayendo las tablas e información contenida en los mismos para sistematizar esos datos en indicadores y graficarlos.

Para esta parte del trabajo se van a utilizar informes del **Monitor Educativo del CES**, en el cual se presentan los informes de cada liceo con información de matriculación y aprobación en cada año.

Se toma a los efectos de hacer un ejercicio de ejemplo simple solo al departamento de Flores.

La complejidad está en que cada liceo presenta la información en un archivo pdf con la información mezclada para varones y mujeres y para procesarla de forma separada es necesario extraerla de las tablas. Una vez extraída la información de las tablas, se puede presentar la información por liceo y por ciclo diferenciado para mujer y varón y la evolución que ha tenido en la matrícula la participación de las mujeres en los últimos años.

Table 1: Matrícula total en Ciclo Básico en Flores

| | x |
|-------|-----|
| X2018 | 842 |
| X2019 | 764 |
| X2020 | 751 |
| X2021 | 764 |

Table 2: Matrícula de Mujeres

| liceo | ciclo | X2018 | X2019 | X2020 | X2021 |
|-----------------|-------|-------|-------|-------|-------|
| Flores dptal | 1CB | 68 | 71 | 82 | 76 |
| Flores dptal | 2CB | 84 | 59 | 73 | 74 |
| Flores dptal | 3CB | 98 | 84 | 66 | 76 |
| Flores dptal | CBT | 250 | 214 | 221 | 226 |
| Ismael Cortinas | 1CB | 9 | 8 | 10 | 14 |
| Ismael Cortinas | 2CB | 7 | 8 | 6 | 10 |

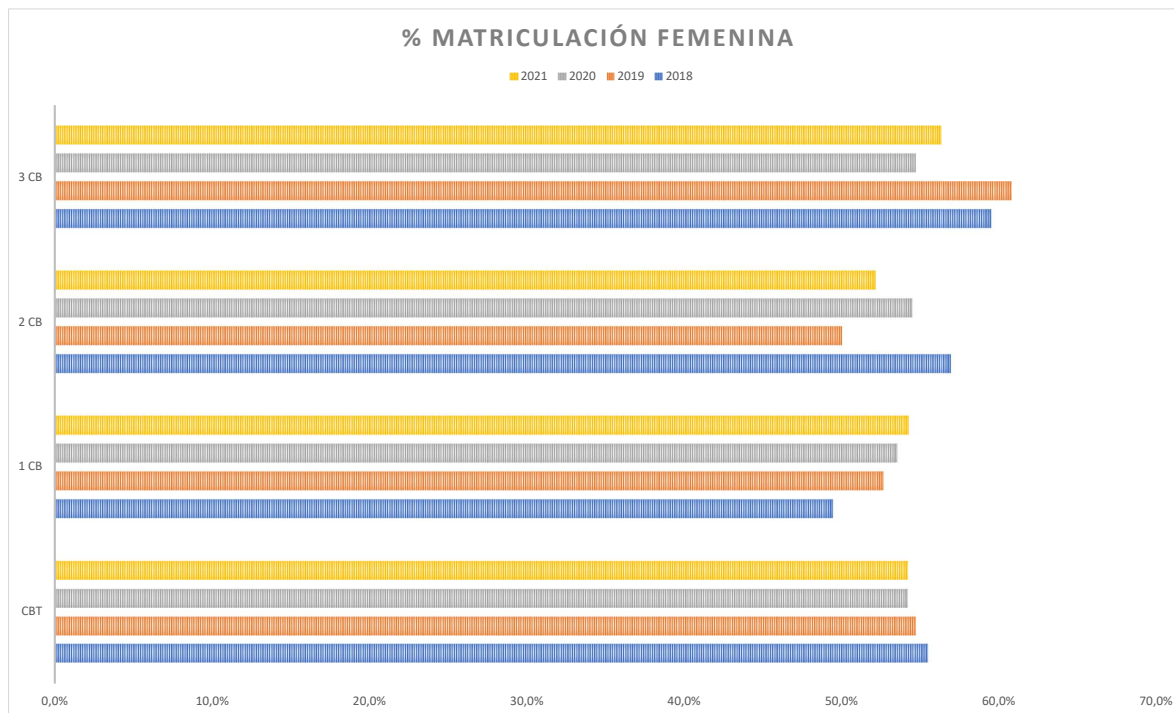
| liceo | ciclo | X2018 | X2019 | X2020 | X2021 |
|-----------------|-------|-------|-------|-------|-------|
| Ismael Cortinas | 3CB | 20 | 6 | 10 | 8 |
| Ismael Cortinas | CBT | 36 | 22 | 26 | 32 |
| Flores nro2 | 1CB | 53 | 51 | 53 | 51 |
| Flores nro2 | 2CB | 77 | 55 | 49 | 51 |
| Flores nro2 | 3CB | 51 | 76 | 58 | 54 |
| Flores nro2 | CBT | 181 | 182 | 160 | 156 |

Table 3: Matrícula de Varones

| liceo | ciclo | X2018 | X2019 | X2020 | X2021 |
|-----------------|-------|-------|-------|-------|-------|
| Flores dptal | 1CB | 76 | 60 | 76 | 56 |
| Flores dptal | 2CB | 65 | 71 | 54 | 72 |
| Flores dptal | 3CB | 57 | 52 | 67 | 55 |
| Flores dptal | CBT | 198 | 183 | 197 | 183 |
| Ismael Cortinas | 1CB | 21 | 9 | 12 | 8 |
| Ismael Cortinas | 2CB | 10 | 16 | 9 | 13 |
| Ismael Cortinas | 3CB | 9 | 6 | 13 | 7 |
| Ismael Cortinas | CBT | 40 | 31 | 34 | 28 |
| Flores nro2 | 1CB | 36 | 48 | 38 | 55 |
| Flores nro2 | 2CB | 52 | 35 | 44 | 39 |
| Flores nro2 | 3CB | 49 | 49 | 31 | 45 |
| Flores nro2 | CBT | 137 | 132 | 113 | 139 |

Table 4: % de Matriculación Femenina

| | X2018 | X2019 | X2020 | X2021 |
|-----|-------|-------|-------|-------|
| CBT | 55.46 | 54.71 | 54.19 | 54.19 |
| 1CB | 49.43 | 52.63 | 53.51 | 54.23 |
| 2CB | 56.95 | 50.00 | 54.47 | 52.12 |
| 3CB | 59.51 | 60.81 | 54.69 | 56.33 |



El dato más destacable es el decenso de la matriculación de las mujeres medido respecto a los varones en el ciclo básico en los últimos cuatro años en el departamento de Flores.

Si bien es positivo que haya menos deserción por parte de los varones, es un fenómeno peculiar que las mujeres aumentan su matriculación respecto a los varones en primer año de ciclo básico y luego disminuye en los siguientes años de ciclo básico.

Análisis de Titulares de la web

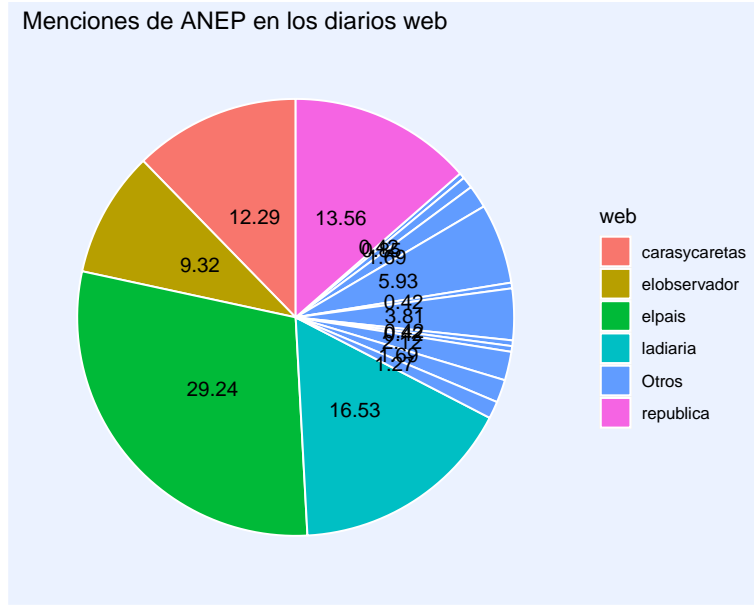
Utilizando el **proyecto GDELT** se recopilan los titulares de la web de los últimos 365 días en los cuales se hacen menciones de ANEP y de Robert Silva, y se identifica el tono de los mismos (si son menciones positivas o negativas).

Se capturan un total de 236 menciones de ANEP siendo la mayoría en **elpais.com.uy** y **ladiaria.com.uy** y 97 menciones de Robert Silva.

Table 5: Cantidad de menciones de ANEP

| | Menciones | % |
|----------------------|-----------|-------|
| 180.com.uy | 1 | 0.42 |
| brecha.com.uy | 2 | 0.85 |
| carasycaretas.com.uy | 29 | 12.29 |
| carve850.com.uy | 4 | 1.69 |
| diariocambio.com.uy | 14 | 5.93 |
| elobservador.com.uy | 22 | 9.32 |
| elpais.com.uy | 69 | 29.24 |

| | Menciones | % |
|------------------------|-----------|--------|
| infonegocios.biz | 1 | 0.42 |
| ladiaria.com.uy | 39 | 16.53 |
| lr21.com.uy | 9 | 3.81 |
| negocios.elpais.com.uy | 1 | 0.42 |
| primerahora.com.uy | 1 | 0.42 |
| radiomontecarlo.com.uy | 5 | 2.12 |
| republica.com.uy | 32 | 13.56 |
| sarandi690.com.uy | 4 | 1.69 |
| subrayado.com.uy | 3 | 1.27 |
| Sum | 236 | 100.00 |



En cuanto al tono, el proyecto GDELT con su algoritmo le asigna un tono, positivo o negativo al artículo encontrado, con lo cual, genera un índice para cada artículo en estos 365 días de búsqueda solicitados.

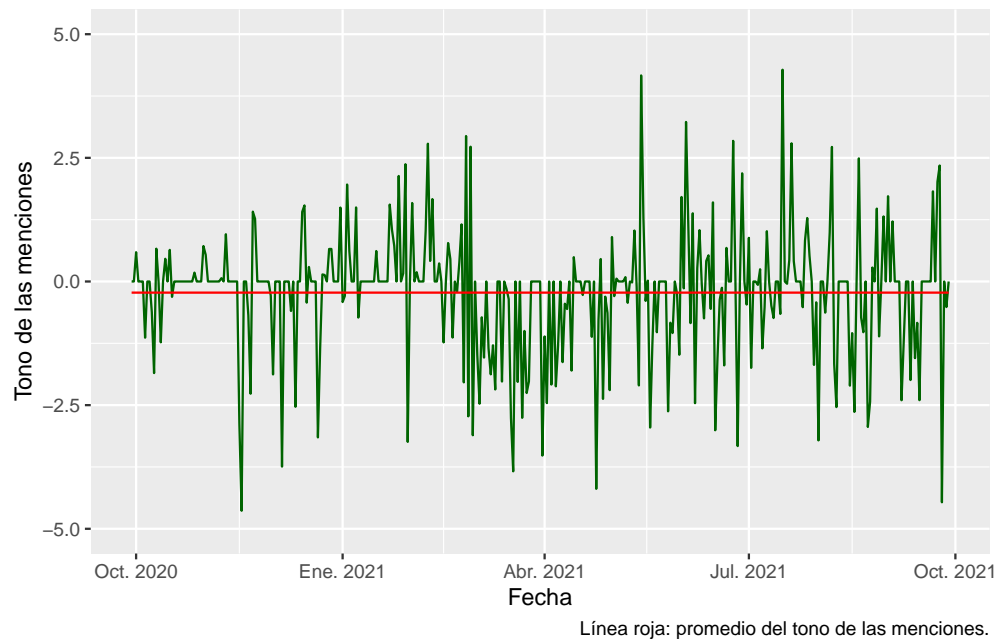
El Índice promedio vinculado al tono de los titulares de los artículos buscados para **ANEP** tiene un valor de -0.414 y los vinculados a **Robert Silva** de -0.225. Siendo su interpretación que si el valor es menor a cero es un tono negativo y si es superior a cero es positivo.

Table 6: Estadísticas descriptivas del índice medio del tono en las menciones

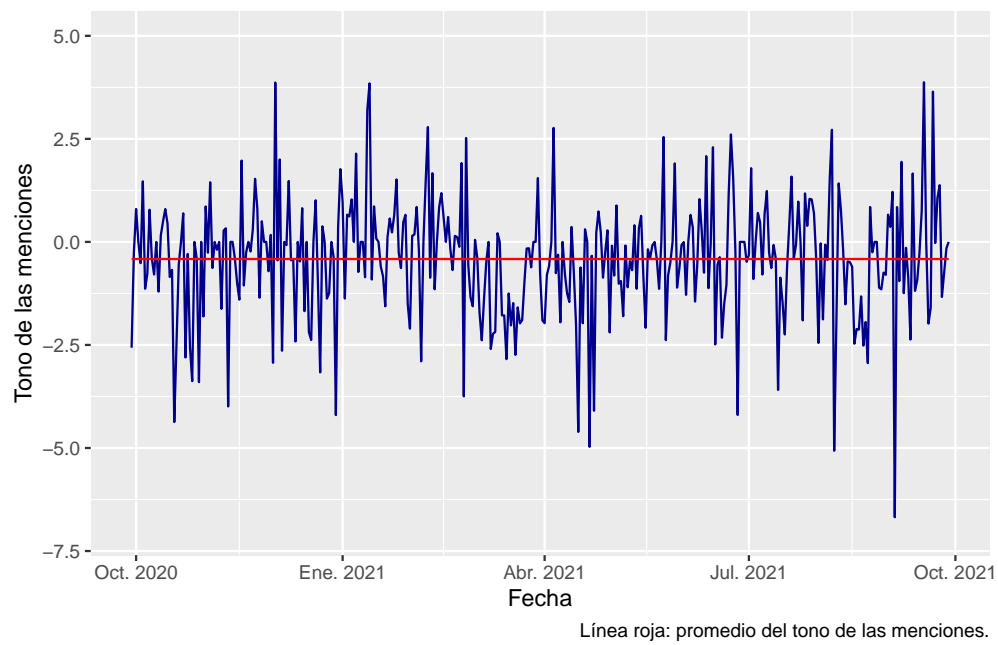
| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|--------|---------|--------|--------|---------|-------|
| RS | -4.636 | -0.555 | 0.000 | -0.225 | 0.000 | 4.282 |
| ANEP | -6.680 | -1.142 | -0.262 | -0.414 | 0.331 | 3.873 |

Se observa en las gráficas una marcada negatividad en los índices de ambos en el entorno del comienzo de clases y abril 2021, período en que hubo conflictos por el inicio de clases, la alimentación en las escuelas y un paro.

Tono Promedio de las menciones a Robert Silva en Titulares



Tono Promedio de las menciones a ANEP en Titulares



A continuación se presenta una **nube de palabras** construída con los titulares recogidos en las 236 menciones de ANEP, lo cual permite destacar las palabras más repetidas y relevantes en los mismos.

Análisis comparado de los sentimientos de los Titulares de la web con otros paquetes.

Se utiliza el paquete **syuzhet** para detectar sentimientos en los titulares que se descargaron de la web y hacer el comparado con el algoritmo propio del proyecto GDELT.

El paquete **syuzhet** clasifica las palabras de los titulares en 10 dimensiones y les asigna un puntaje si lo tiene mayor a cero o nada. En función de esas dimensiones, para este trabajo se reclasificaron las mismas en positivas, negativas y neutrales, no se encontraron menciones neutrales con lo cual se grafican solo los valores de la suma de las positivas y las negativas.

Se agrupan como dimensiones positivas a:

- joy
- trust
- positive

Dimensiones neutrales:

- anticipation
- surprise

Dimensiones negativas:

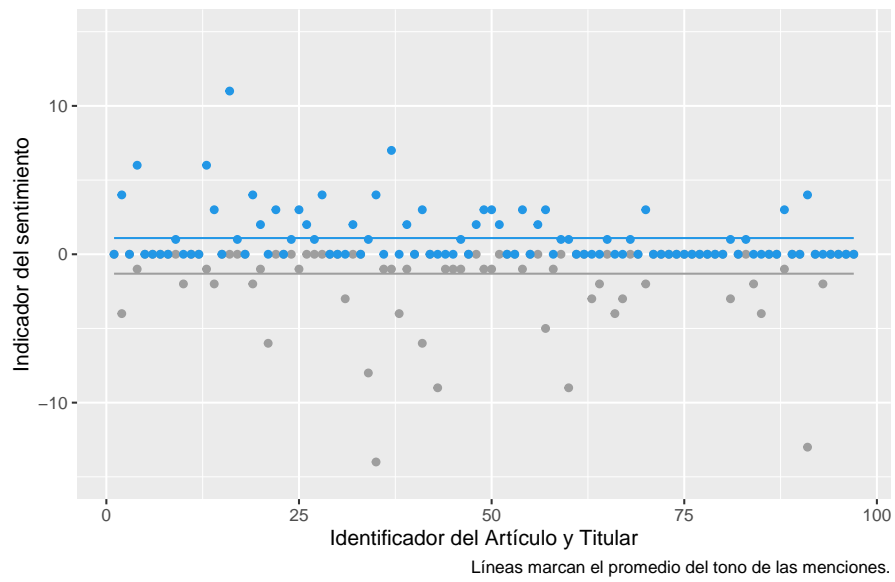
- anger
- disgust
- fear
- sadness
- negative

En función de estas agrupaciones, se llega a una conclusión similar a la del algoritmo del proyecto GDELT en que el promedio para Robert Silva es negativo.

El valor promedio de los sentimientos negativos es -1.309 y el de los sentimientos positivos 1.093, por lo tanto, en los titulares asociados a Robert Silva predominan los sentimientos negativos.

En los titulares asociados a ANEP, el valor promedio de los sentimientos negativos es -1.055 y el de los sentimientos positivos 1.568, con lo que predominan los sentimientos positivos en los titulares de ANEP en promedio.

Análisis de sentimientos de las menciones a Robert Silva en Titulares



Análisis de sentimientos de las menciones a ANEP en Titulares

