

Análisis de datos de INE y de COVID

Emilio Villa Cueva

Proyecto final del curso de Herramientas Informáticas y Gestión de la Información

14 de junio de 2021

Índice

1. Pronóstico de ocupación de casillas con datos de INE	2
1.1. Problema a resolver	2
1.2. Procedimiento seguido	2
1.3. Resultados	3
2. Análisis de percepción de afectación de COVID por género alrededor del mundo	3
2.1. Sobre el objetivo del proyecto	3
2.2. Procedimiento seguido	3
2.3. Panorama general de afectaciones	4
2.4. Contraste entre afectación a hombres y a mujeres	6
2.5. Sobre el notebook	6
Referencias	9

1. Pronóstico de ocupación de casillas con datos de INE

1.1. Problema a resolver

Con base en los datos recopilados por el INE sobre las listas nominales alrededor del país, disponibles en [1], se buscó realizar una estimación de dicha lista nominal para el mes de febrero de 2021.

Una vez teniendo una estimación de la lista nominal, el objetivo final del trabajo consistió en estimar el número de casillas necesario por sección dentro del municipio de Guanajuato, bajo la premisa de que por cada 750 mexicanos en la lista nominal se requería una casilla. De manera que:

$$\text{Núm de casillas} = \lceil \frac{\text{lista nominal}}{750} \rceil \quad (1)$$

Entonces usando la estimación para febrero de la lista nominal por sección, se podrá calcular el número de casillas requeridas por sección usando la Ecuación 1.

1.2. Procedimiento seguido

Se trabajó con datos obtenidos la página de internet de [1], los datos se encontraban como archivos formato .txt con los valores mensuales de las listas nominales para todo el país.

Se trabajo con datos desde marzo de 2019 hasta diciembre de 2020, para poder realizar una estimación lo más exacta posible. Dentro de dichos datos se separaba la lista en mujeres, hombres, nacionales y extranjeros, también se proporcionaba la lista combinada (mujeres y hombres) nacional.

Debido a que para los archivos de los diferentes meses en ocasiones se tenían nombres diferentes para la misma variable, lo primero que se hizo fue cambiar el nombre en todos a *LISTA_N* para la lista nominal, filtrando el estado de Guanajuato y concatenando todos los meses.

Posteriormente se agruparon todas las observaciones por municipio para realizar la regresión por municipio, esto también permitió realizar visualizaciones del crecimiento en la lista nominal por municipio como la que se muestra en la figura 1.

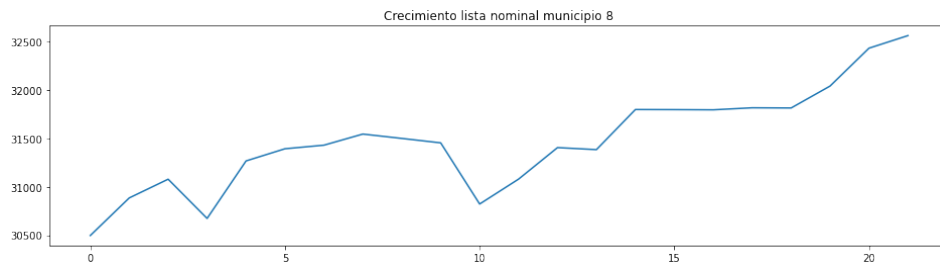


Figura 1: Crecimiento de lista nominal

La regresión lineal se realizó usando la implementación del paquete *Scikit-Learn*, usando la función *LinearRegression()*, y usando 22 meses para ajustar los datos (hasta diciembre de 2020), se realizó el pronóstico al mes de marzo 2021, calculando la tasa de cambio por municipio de la forma que se muestra en la ecuación 2 y guardándola en un arreglo.

$$\Delta_{nom}^{\text{municipio}} = \frac{LN_{\text{mar2021}}}{LN_{\text{dic2020}}} \quad (2)$$

Teniendo estas tasas de cambio, se procedió a calcular el valor estimado de la lista nominal por sección, usando los valores de las tasas del municipio al que corresponden, de la forma que se muestra en la figura 3.

$$LN_{\text{sección x mar2021}}^{\text{sección x}} = \Delta_{\text{nom}}^{\text{municipio}} \times LN_{\text{sección x dic2020}}^{\text{sección x}} \quad (3)$$

De manera que se pudieron calcular estimaciones para todas las casillas. Finalmente se uso la ecuación 1 para calcular el número necesario de casillas por sección y se sumaron todas las casillas requeridas.

1.3. Resultados

Siguiendo este procedimiento, se logró realizar una estimación adecuada para el número de casillas requeridas por sección. Los resultados individuales por casilla se tienen en el archivo, pero en todo el estado se espera que el número total de casillas ronde las **7620**, que fue el resultado obtenido.

Todo se realizó en el *notebook INEforecast.ipynb* adjunto usando *Python 3.8*. Revisar para ver lo que se describe aquí.

2. Análisis de percepción de afectación de COVID por género alrededor del mundo

2.1. Sobre el objetivo del proyecto

Se trabajó con los datos recopilados por [2], a inicios de la pandemia COVID19 durante julio de 2020 que buscaban dar luz a las dinámicas de género que durante la pandemia se vieron (posiblemente) modificadas alrededor del mundo.

A diferencia del reporte anterior, esta vez (además de haber realizado el análisis con Python), se buscó aprovechar más las gráficas de barras para denotar las **diferencias en la percepción de afectación durante la pandemia** tomando en cuenta las diferentes regiones: Asia del Este, Europa y Asia del Este, Latinoamérica y el Caribe, Medio Oriente y Norte de África, Norteamérica, Sur de Asia y África Sub-Sahariana. Se decidió trabajar con regiones en vez de países ya que se tienen más de 120 países en el conjunto de datos. Este tipo de gráficos (gráficos de barra), aunque sencillos, son capaces de capturar diferencias importantes entre regiones respecto a las afectaciones en diferentes ámbitos, como se verá a continuación.

2.2. Procedimiento seguido

Primero conviene describir el tipo de datos con los que se trabajó. Consisten en porcentajes de respuestas **afirmativas** a preguntas específicas realizadas por Facebook a personas en todo el mundo respecto a las afectaciones de la pandemia en diferentes ámbitos, los datos se separan en respuestas de mujeres, de hombres y combinadas. Se tienen las respuestas a más de 80 interrogantes distintas, que buscar dar luz a diferentes aspectos de la vida de la gente alrededor del mundo. En el presente proyecto se decidió trabajar en particular con 8 preguntas de la encuesta que se consideraron de interés:

- **a1_agree**: De acuerdo con que hombres y mujeres deben de tener las mismas oportunidades

- **c2_clean**: El encuestado se encarga de la limpieza de la casa
- **c2_cook**: El encuestado se encarga de cocinar en la casa
- **c3_agree**: El encuestado está de acuerdo con que el rol más importante de una mujer es cuidar la casa y su familia.
- **c2d_increase**: Durante la pandemia aumentó el tiempo que se dedica al hogar
- **d5_agree**: Se ha sentido incómodo o en riesgo en su propio hogar
- **d7_money**: Su principal preocupación durante la pandemia es tener dinero suficiente para subsistir
- **d6_school**: Se vió imposibilitado de ir a la escuela.

Por ejemplo, una observación típica¹ sería como la que se muestra en la figura 2.

	Region	Country	Gender	a1_agree	c2_clean	c2_cook	c3_agree	c2d_increase	d5_agree	d7_money	d6_school
0	North America	Canada	Female	93	87	77	30	45	17	35	33
1	North America	Canada	Male	92	74	65	23	35	16	33	27

Figura 2: Observación típica de los datos

Tras haber seleccionado las variables de interés, se procedió a calcular la media por región de cada una de estas variables, generando un *dataframe* para hombres y otro para mujeres, el cual se muestra en el cuadro 1.

	Region	a1_agree	c2_clean	c2_cook	c3_agree	c2d_increase	d5_agree	d7_money	d6_school
1	East Asia & Pacific	90.416	66.1667	50.9167	51.1667	42.583	33.5	32.5	36.75
2	Europe and Central Asia	91.846	83.0	70.795	35.538	38.564	23.1026	35.82	30.256
3	Latin America and Caribbean	91.762	80.62	62.5714	35.095	56.143	38.7143	54.143	44.714
4	Middle East and North Africa	89.706	74.0	58.7647	55.294	53.882	43.1176	36.529	35.471
5	North America	91.5	87.5	77.5	33.5	44.0	15.5	37.0	34.5
6	South Asia	89.7143	64.43	58.7143	62.714	53.857	29.714	28.428	41.571
7	Sub-Saharan Africa	86.44	67.28	58.04	64.92	49.12	50.64	36.12	51.68

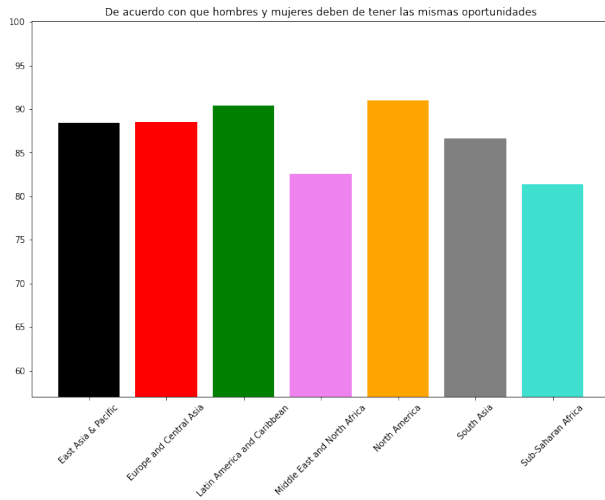
Cuadro 1: Dataframe con media de respuestas afirmativas a diferentes variables de interés

De la información de estos *dataframes* es posible generar visualizaciones muy interesantes que sirven para realizar un análisis.

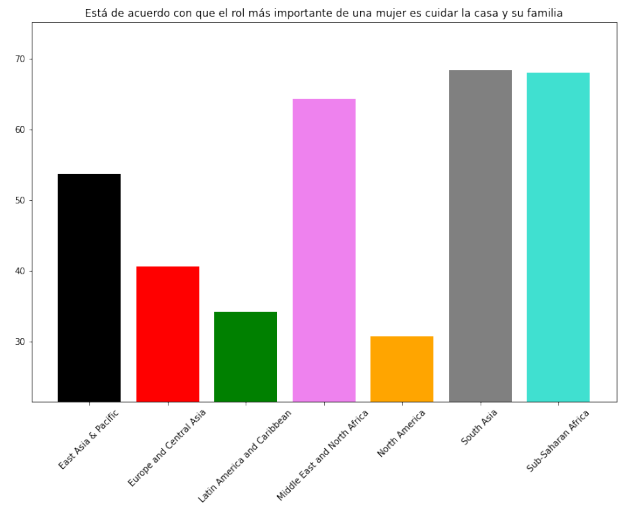
2.3. Panorama general de afectaciones

Antes de contrastar entre hombres y mujeres, se tomaron las respuestas de ambos hacia algunas de las preguntas que se consideraron de interés. Las respuestas se ilustran en la figura 3. En la primera parte, más que contrastar diferencias de afectación por género, busca analizar cómo se vieron afectadas diferentes

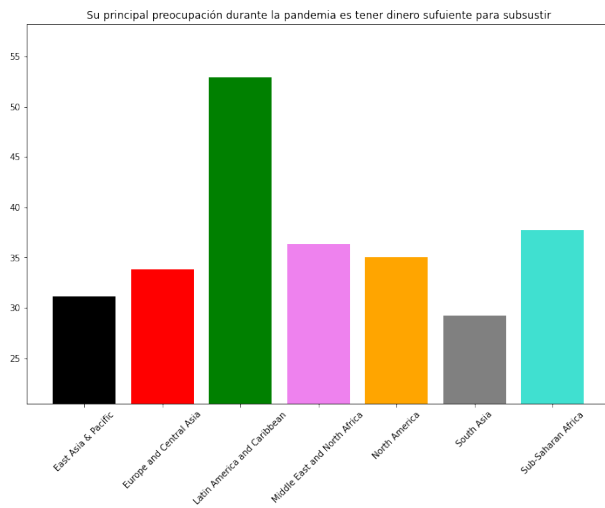
¹Habiendo filtrado ya las variables de interés de la lista.



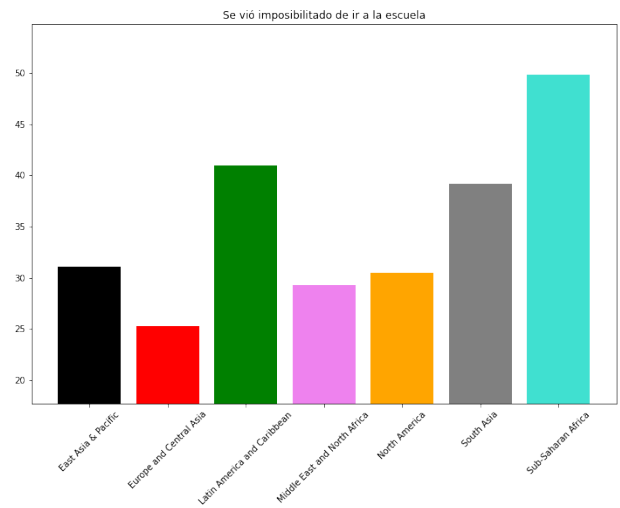
(a)



(b)



(c)



(d)

Figura 3: Afectaciones en diferentes ámbitos

regiones del mundo por efecto de la pandemia, así como las percepciones (tanto de hombres como de mujeres) del rol de las mujeres en la sociedad.

Por ejemplo, se observan regiones donde aún no se considera que la mujer y el hombre deban de tener las mismas oportunidades (figura 3a). Y aún se considera que el rol principal de la mujer es el cuidar de la casa (figura 3b), habiendo regiones como el sur de Asia, África y Medio Oriente donde más de la mitad de la población opina eso.

Respecto a la afectación por efecto de la pandemia en un aspecto más general, se puede observar que en los ámbitos económicos (figura 3c) y de educación (figura 3d) América Latina y el sur de África destacan como los más afectados.

2.4. Contraste entre afectación a hombres y a mujeres

En esta sección el procedimiento seguido fue que para cada una de las preguntas de interés, y por cada región, se calculó la diferencia entre las respuestas² de las mujeres y de los hombres, como se muestra en la ecuación 4.

$$\Delta \% = \%_F - \%_M \quad (4)$$

Esto con el interés de mostrar de una manera más clara preguntas donde alguno de los géneros se vio más afectado por región. Los resultados se ilustran entonces en las figuras 4 y 5.

De las visualizaciones podemos puntualizar algunas observaciones interesantes. Primero, en la figura 4a notamos que en Medio Oriente, una zona donde existe una brecha grande aún entre los derechos del hombre y de la mujer, la proporción de mujeres que considera que se necesitan derechos iguales es considerablemente mayor que los hombres. De las figuras 4b, 4c y 5a, notamos dos cosas: durante la pandemia fue a mujer quien **en todo el mundo** se encargó mayoritariamente del hogar (o al menos de la cocina y la limpieza), y al observar la figura 4d notamos que si bien son los hombres quien en la mayoría de las regiones consideran que las tareas del hogar son responsabilidad de la mujer, en América, es mayor la cantidad de mujeres que opinan que las responsabilidades del hogar son de ellas.

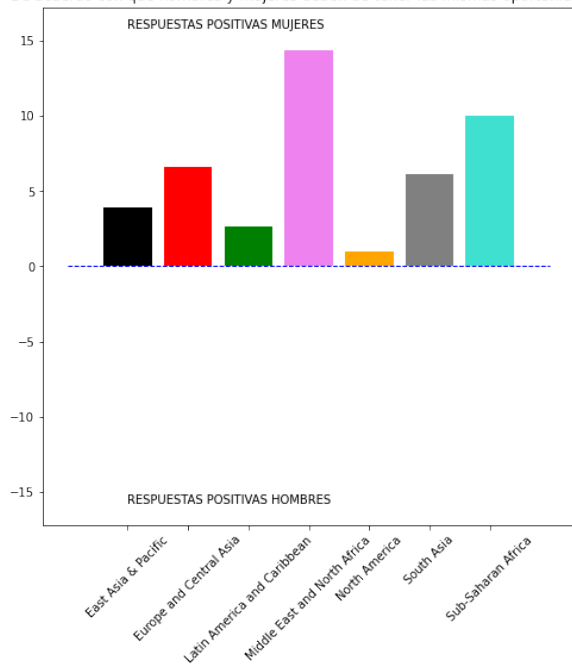
Un resultado curioso es el de la figura 5b, que hace alusión a si la gente se sintió insegura en su hogar, donde por algún motivo en el sur de Asia fueron en su mayoría los hombres quienes se sintieron inseguros, aunque en el resto del mundo (y probablemente por casos de violencia doméstica que escalaron durante las cuarentenas) fueron las mujeres quienes se sintieron más inseguras. Un resultado similar se obtiene en la figura 5c, donde si bien en la mayoría de las regiones fueron las mujeres quienes se preocuparon más por la economía del hogar, en Asia y África los hombres se encontraban más preocupados. Finalmente, se observa en la figura 5d que fueron las mujeres quienes faltaron más a la escuela por efecto de la pandemia.

2.5. Sobre el notebook

Se anexa el notebook *covidAnalysis.ipynb*, donde se realizó todo lo mencionado en ésta segunda sección relativo a los datos de COVID, los datos también se encuentran en *rawData.csv* dentro de la carpeta *data*.

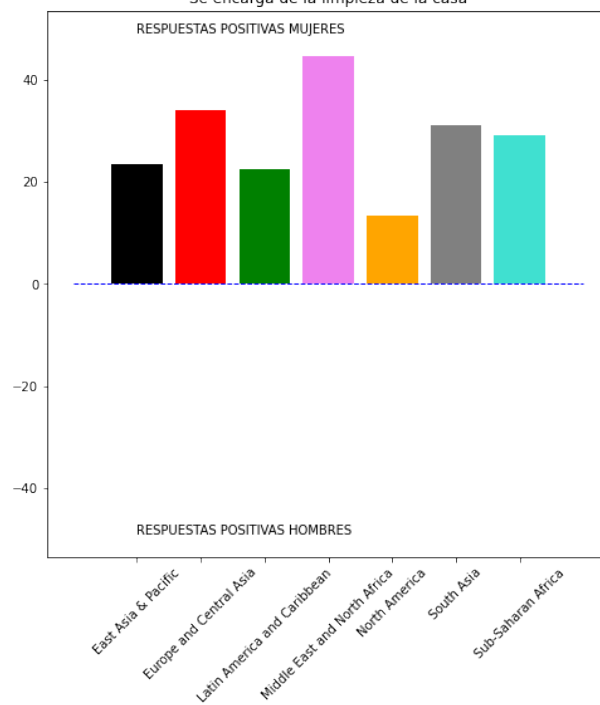
²Las respuestas afirmativas, es decir, el porcentaje de encuestados que contestó afirmativamente a la pregunta que se le hizo

De acuerdo con que hombres y mujeres deben de tener las mismas oportunidades



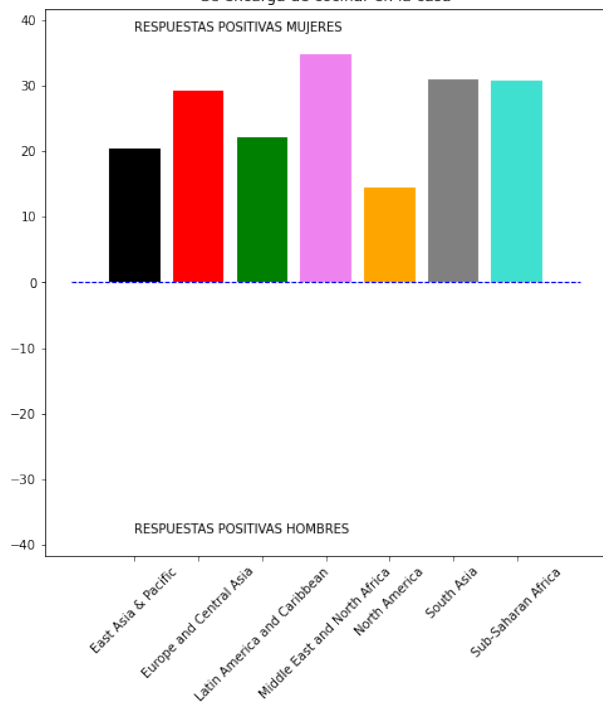
(a)

Se encarga de la limpieza de la casa



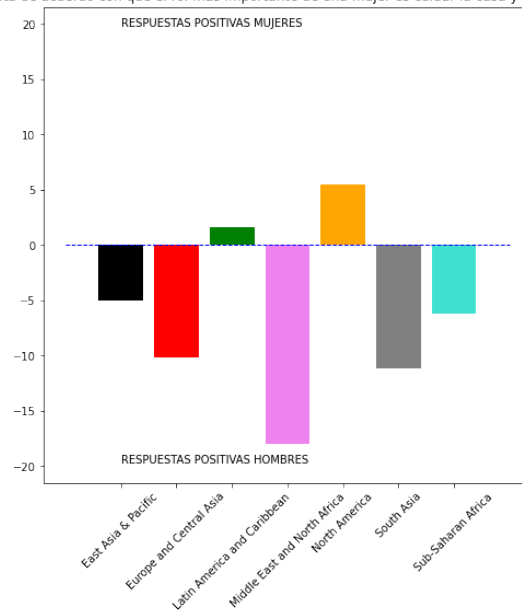
(b)

Se encarga de cocinar en la casa



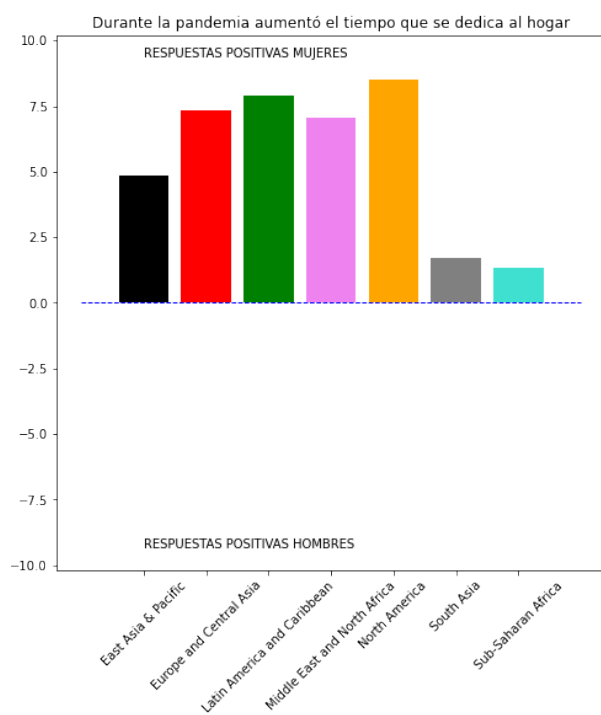
(c)

Está de acuerdo con que el rol más importante de una mujer es cuidar la casa y su familia

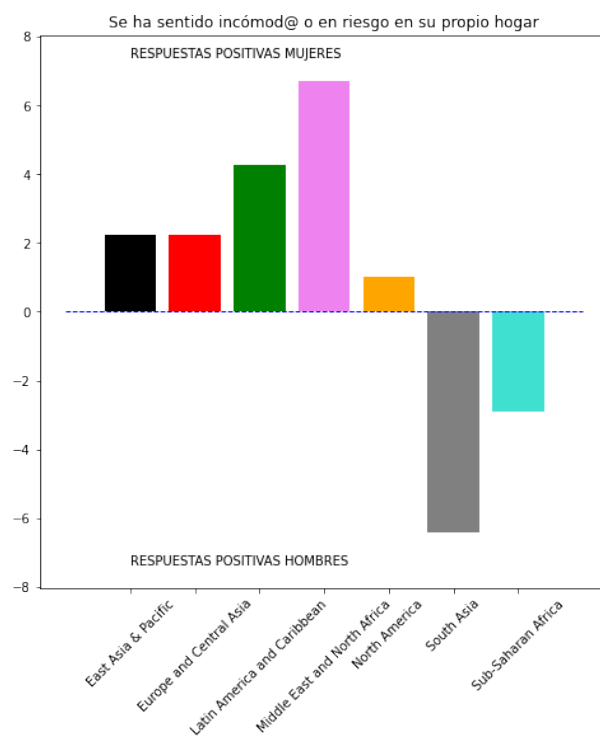


(d)

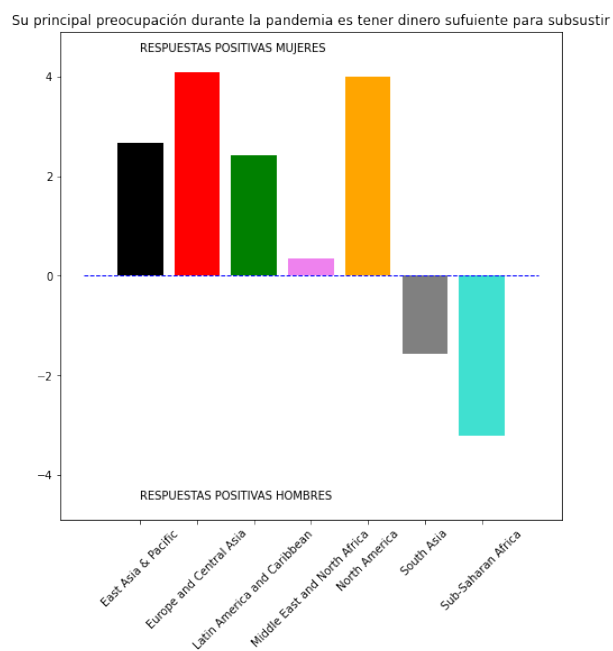
Figura 4: Afectaciones en diferentes ámbitos



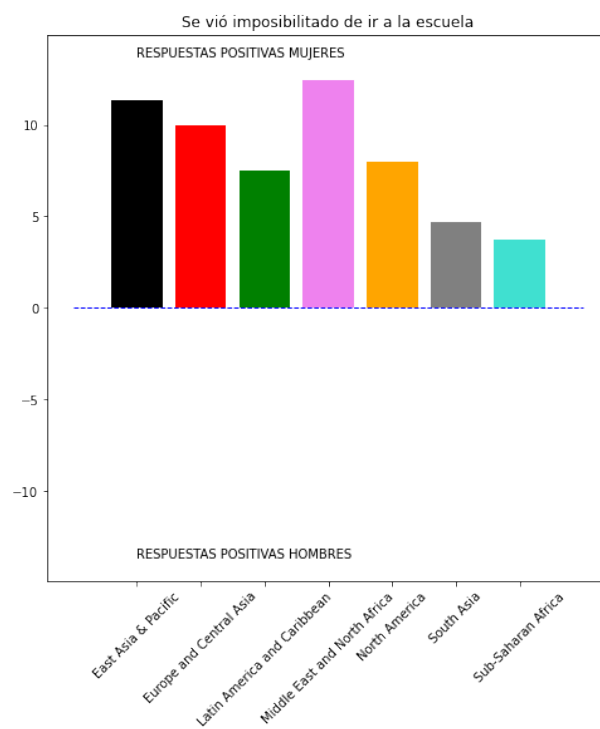
(a)



(b)



(c)



(d)

Figura 5: Afectaciones en diferentes ámbitos

Referencias

- [1] *Estadística de Padrón Electoral y Lista Nominal de Electores*. URL: <https://www.ine.mx/transparencia/datos-abiertos/#/archivo/estadistica-padron-electoral-lista-nominal-electores>.
- [2] *Survey on Gender Equality at Home Report*. 2020. URL: <https://dataforgood.fb.com/docs/gendersurveyreport/>.