

Proyecto 1

Fernando Stein Vallarta 165455

Pablo López Landeros 178863

Manuel García Garduño 162136

## Introducción:

Para el proyecto utilizamos la base de datos de Servicio Línea de Mujeres que se encuentra disponible en la página de datos abiertos de la CDMX. En esta base de datos podemos encontrar la información de las llamadas realizadas al servicio Línea de Mujeres de LOCATEL desde noviembre de 2016. Sin embargo, para este proyecto únicamente consideramos los datos del año 2020.

La captura de la información se realiza a través de lo que el gobierno de la CDMX ha denominado Sistema de Registro de Información de Locatel (SIRILO). Cabe resaltar que la información obtenida es otorgada de forma opcional por las mujeres que llaman y requiere del consentimiento de las usuarias para ser difundida.

Para obtener la información, las operadoras realizan una entrevista a las mujeres que se comunican al servicio de asesorías. La atención que brindan las operadoras depende del motivo de la llamada ya que este permite a la operadora elaborar un plan de acción. Cabe resaltar que la información obtenida es otorgada de forma voluntaria por las mujeres que llaman y requiere del consentimiento de las usuarias para ser difundida.

En esta base podrás encontrar toda la información de las llamadas realizadas al servicio Línea de Mujeres de LOCATEL desde noviembre de 2016. Esta información se actualizará semanalmente a partir del 6 de abril de 2020 La Línea mujeres brinda atención enfocada a las mujeres que sufren de violencia, inequidad y desigualdad de género en los siguientes rubros:

1. Información sobre programas y acciones del gobierno de la CDMX
2. Asesoría psicológica
3. Asesoría Jurídica
4. Orientación médica

```
#Paquetes requeridos
```

```
library(tidyverse)
library(cowplot)
library(kableExtra)
library(knitr)
library(lubridate)
library(dplyr)
library(moments)
library(readr)
```

```
#Lectura de Datos y Diccionario
```

```
linea.mujeres <- read_csv("https://raw.githubusercontent.com/pablolopez2733/Aplicada1/master/Bases%20de%20datos/linea.mujeres.csv")
diccionario.linea <- read_csv("https://github.com/pablolopez2733/Aplicada1/raw/master/Bases%20de%20Datos/diccionario.linea.csv")
```

## Descripción de los datos

Para realizar una breve descripción de los datos, empleamos el diccionario proporcionado por la página de datos abiertos de la CDMX. Observamos que la base de datos se compone de la siguiente forma:

```
#Descripción de la base de datos usando el Diccionario
```

```
glimpse(linea.mujeres)
```

```
## Rows: 66,315
## Columns: 27
## $ FOLIO <dbl> 656650, 656651, 656654, 656659, 656663, 656666, 6...
## $ FECHA_HORA_ALTA <dtm> 2020-02-11 14:01:00, 2020-02-11 14:01:00, 2020-0...
## $ DÍA_ALTA <dbl> 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 1...
## $ MES_ALTA <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
## $ AÑO_ALTA <dbl> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2...
## $ SEXO <chr> "FEMENINO", "FEMENINO", "FEMENINO", "FEMENINO", "...
## $ EDAD <dbl> 83, 29, 13, 67, 39, 56, 60, 40, 68, 20, 69, 45, 6...
## $ ESTADO_CIVIL <chr> "SOLTERO (A)", "SOLTERO (A)", "SOLTERO (A)", "SOL...
## $ OCUPACION <chr> "HOGAR", "EMPLEADO (A)", "HOGAR", "EMPLEADO (A)",...
## $ ESCOLARIDAD <chr> "BACHILLERATO", "BACHILLERATO", "PRIMARIA", "SECU...
## $ ESTADO_USUARIA <chr> "CIUDAD DE MÉXICO", "ESTADO DE MÉXICO", "CIUDAD D...
## $ MUNICIPIO_USUARIA <chr> "CUAUHTÉMOC", "NEZAHUALCÓYOTL", "GUSTAVO A. MADER...
## $ COLONIA_USUARIA <chr> "ROMA NORTE", "EL SOL", "GUADALUPE PROLETARIA", "...
## $ CP_USUARIA <dbl> 6700, 57200, 7670, 1470, 9660, 7320, 9030, 57410,...
## $ ESTADO_HECHOS <chr> NA, NA, NA, NA, NA, "CIUDAD DE MÉXICO", NA, NA, N...
## $ MUNICIPIO_HECHOS <chr> NA, NA, NA, NA, NA, "GUSTAVO A. MADERO", NA, NA, ...
## $ COLONIA_HECHOS <chr> NA, NA, NA, NA, NA, "FOVISSSTE", NA, NA, NA, NA, ...
## $ CP_HECHOS <dbl> NA, NA, NA, NA, NA, 7320, NA, NA, NA, NA, NA, NA,...
## $ ORIGEN <chr> "ASESORÍA", "ASESORÍA", "ASESORÍA", "ASESORÍA", "...
## $ SERVICIO <chr> "PSICOLÓGICO", "PSICOLÓGICO", "PSICOLÓGICO", "JUR...
## $ TEMATICA_1 <chr> "VIOLENCIA", "CONFLICTOS EMOCIONALES", "SERVICIOS...
## $ TEMATICA_2 <chr> "OTRA", "ESTADOS DE ANIMO", "INFORMACION", NA, "D...
## $ TEMATICA_3 <chr> "SOLICITUD DE SERVICIOS DE APOYO", "ENFRENTAR SIT...
## $ TEMATICA_4 <chr> NA, NA, NA, NA, NA, "RELACIÓN PADRES E HIJOS", NA...
## $ TEMATICA_5 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "URGENCIAS Y ...
## $ TEMATICA_6 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "URGENCIAS Y ...
## $ TEMATICA_7 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "URGENCIAS Y ...
```

```
variables <- diccionario.linea %>% select(`Nombre de la variable`, Definición)
kable(variables, booktabs = T) %>% kable_styling(latex_options = "striped")
```

Nombre de la variable	Definición
FOLIO	Identificador único de la llamada
FECHA_ALTA	Fecha en la que se generó la llamada de la usuaria (día, mes y año)
AÑO_ALTA	Año en que se generó la llamada de la usuaria.
MES_ALTA	Mes en que se generó la llamada de la usuaria.
DIAS_ALTA	Día en que se generó la llamada de la usuaria.
HORA_ALTA	Año en que se generó la llamada de la usuaria.
SEXO	Sexo de la usuaria que realizó la llamada.
EDAD	Edad de la usuaria que realizó la llamada.
ESTADO CIVIL	Estado civil de la usuaria que realizó la llamada.
OCUPACION	Ocupación de la usuaria que realizó la llamada.
ESCOLARIDAD	Escolaridad de la usuaria que realizó la llamada.
ESTADO_USUARIA	Entidad Federativa de la usuaria que realizó la llamada.
MUNICIPIO_USUARIA	Municipio de la usuaria que realizó la llamada.
COLONIA_USUARIA	Colonia de la usuaria que realizó la llamada.
CP_USUARIA	Código Postal de la usuaria que realizó la llamada.
ESTADO_HECHOS	Estado en el que se registraron los hechos.
MUNICIPIO_HECHOS	Municipio en el que se registraron los hechos.
COLONIA_HECHOS	Colonia en la que se registraron los hechos.
CP_HECHOS	Código Postal en el que se registraron los hechos.
ORIGEN	Plataforma desde la que se inicio la llamada.
SERVICIO	Servicio que se le proporcionó a la usuaria.
TEMATICA_1	Primera temática en la que se categoriza la llamada.
TEMATICA_2	Segunda temática en la que se categoriza la llamada.
TEMATICA_3	Tercera temática en la que se categoriza la llamada.
TEMATICA_4	Cuarta temática en la que se categoriza la llamada.
TEMATICA_5	Quinta temática en la que se categoriza la llamada.
TEMATICA_6	Sexta temática en la que se categoriza la llamada.
TEMATICA_7	Séptima temática en la que se categoriza la llamada.

Ahora bien, en el presente proyecto buscamos estudiar múltiples situaciones que rodean a la realidad de discriminación por género en la CDMX.

En primer lugar, nos gustaría analizar si desde la cuarentena ha existido un incremento en las llamadas realizadas a la Línea Mujeres. La Jornada Nacional de Sana Distancia (JNSD), o cuarentena comenzó el 23 de marzo del presente año. Luego entonces, construimos un histograma en donde observamos el número de llamadas por mes, contemplando que los meses de cuarentena son abril, mayo y algunos días de marzo y junio:

#### #ANALISIS ESTADISTICO-DESCRIPTIVO

*#En primer lugar, realizaremos un análisis de casos por mes*

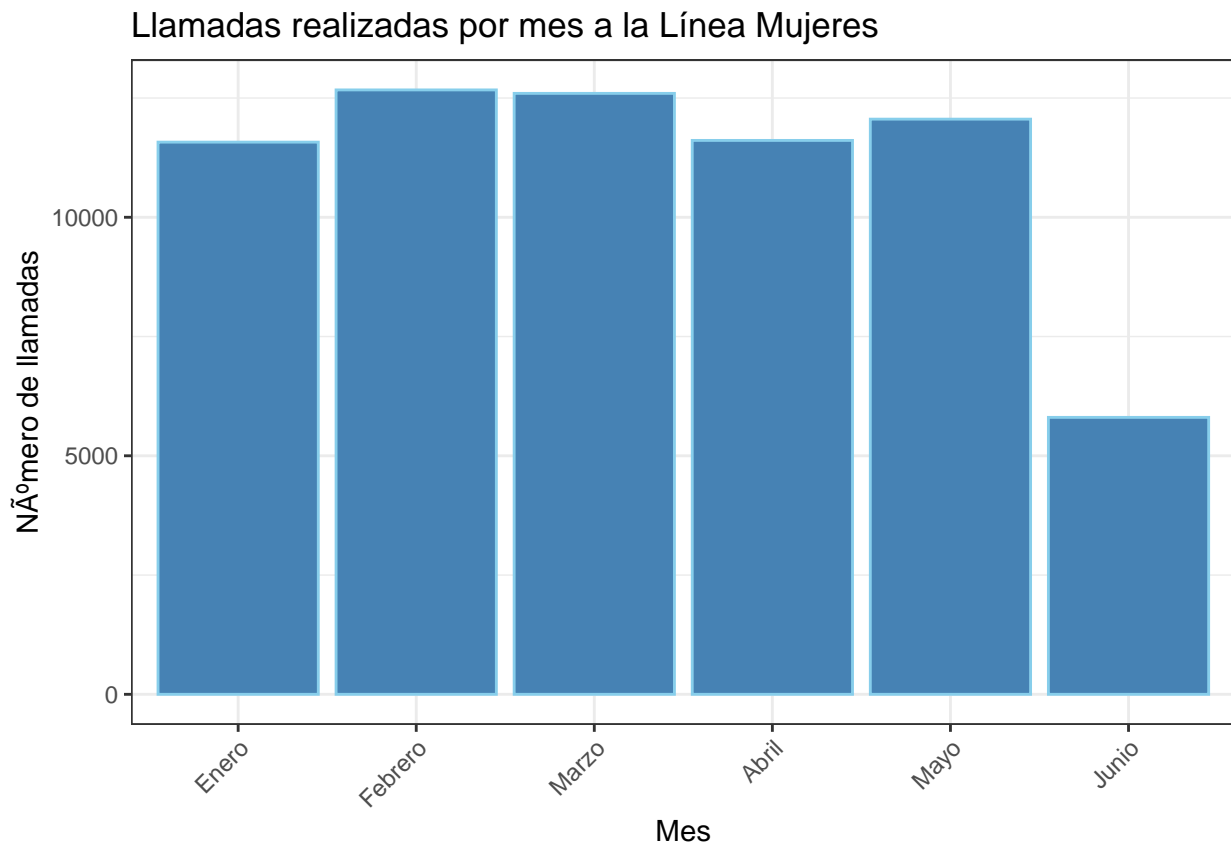
```
llamadas.mes <- linea.mujeres %>% group_by(MES_ALTA) %>% count()
llamadas.mes$mes<-c("Enero", "Febrero", "Marzo", "Abril", "Mayo", "Junio")
```

```
ggplot(llamadas.mes) +
  geom_col(aes(x = reorder(mes, MES_ALTA), y = n), color='skyblue',fill='steelblue') +
  theme_bw() +
  labs(
    title = "Llamadas realizadas por mes a la Línea Mujeres",
    x = "Mes",
```

```

y = "Número de llamadas"
) +
theme(axis.text.x=element_text(angle=45, hjust=1))

```



En la anterior gráfica notamos que para los meses de abril, mayo y junio no se nota un aumento en el número de llamadas, de hecho, las llamadas han disminuido. Sin embargo, el mes de junio aún no ha concluido, de ahí que la gráfica anterior no parece ser contundente.

Ahora bien, en busca de un resultado más claro sobre si existe un aumento en las llamadas a la línea mujeres a partir del cuarentena realizamos otra gráfica, ahora de líneas. En esta gráfica es más clara la tendencia a la baja a raíz del comienzo de la cuarentena.

```

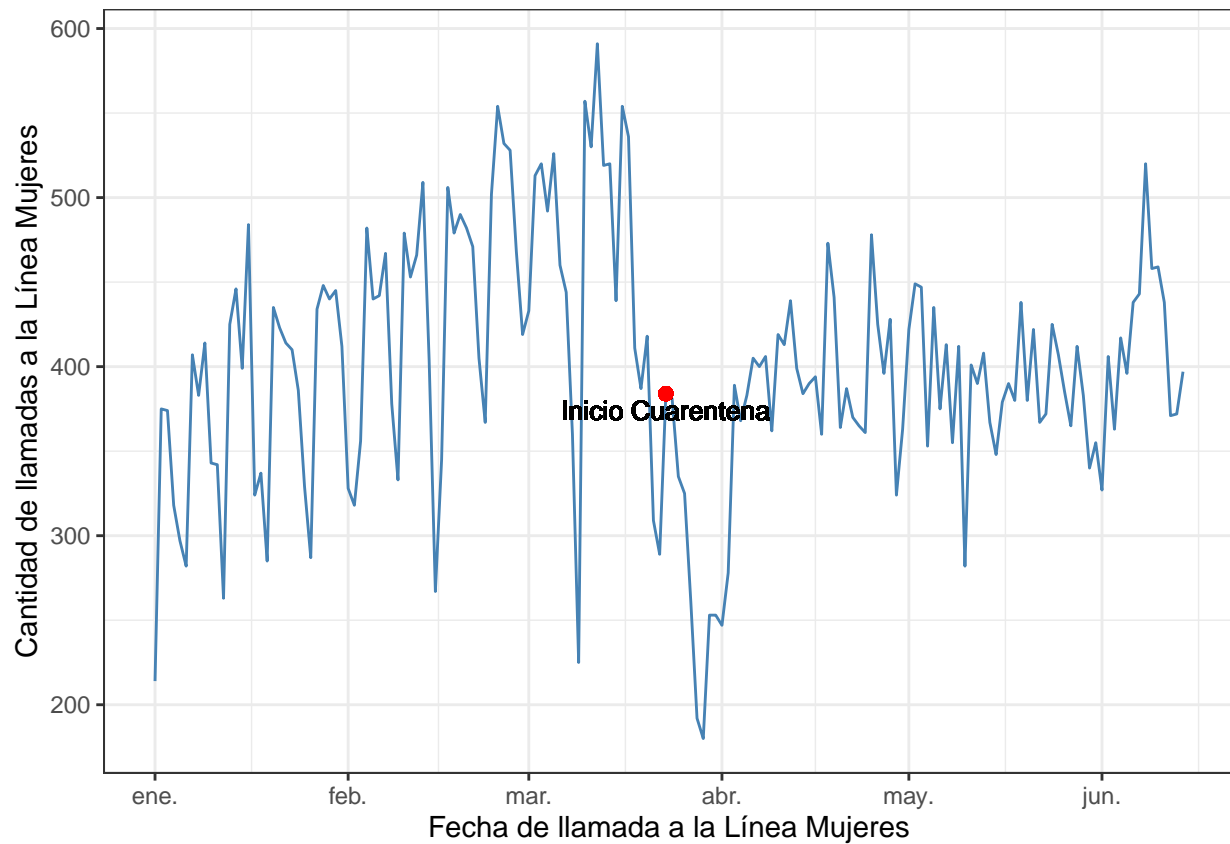
#PreCovid - PostCovid

#Inicia 23 - Marzo - 2020
linea.mujeres <- linea.mujeres %>% mutate(FECHA_HORA_ALTA = ymd_hms(FECHA_HORA_ALTA))
linea.mujeres <- linea.mujeres %>% mutate(fecha = date(FECHA_HORA_ALTA))
conteo_llamadas<-linea.mujeres %>% group_by(fecha) %>% tally() #llamadas por día

#Gráfica de Líneas por fecha
ggplot(conteo_llamadas) +
  geom_line(aes(x = fecha, y = n), colour = "steelblue") +
  theme_bw() +
  labs(
    x = "Fecha de llamada a la Línea Mujeres",
    y = "Cantidad de llamadas a la Línea Mujeres"
  ) +

```

```
geom_point(aes(x = dmy("23/03/2020"), y = 384), color = "red", size = 2) +
geom_text(aes(x = dmy("23/03/2020"), y = 384), label = "Inicio Cuarentena", size= 3.5, nudge_y= -10
```



*#OJO : No se nota ninguna tendencia a partir del inicio de la cuarentena*

De igual forma, nos gustaría realizar un análisis por edad. En primera instancia, calculamos algunas estadísticas de resumen de la variable edad como su: desviación media absoluta, media, mediana, rango intercuartílico y varianza.

*#Análisis por edades*

*#Tabla de resumen*

```
llamadas.tabla.edad <- linea.mujeres %>%
```

```
  summarise(MAD = mad(EDAD),
            Promedio = mean(EDAD),
            Mediana = median(EDAD),
            IQR = IQR(EDAD),
            Varianza = var(EDAD))
```

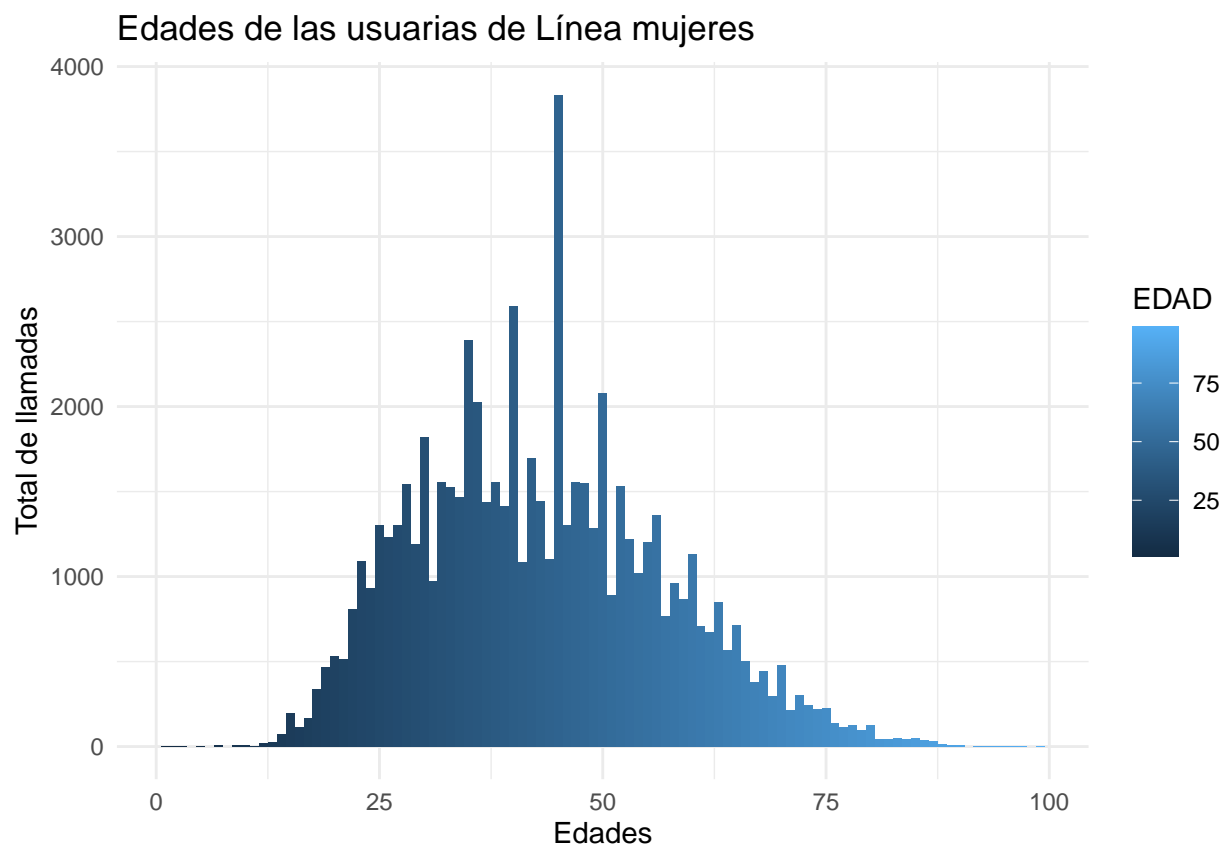
```
kable(llamadas.tabla.edad, booktabs = T) %>% kable_styling(latex_options = "striped")
```

MAD	Promedio	Mediana	IQR	Varianza
14.826	43.11089	42	19	192.6623

Con los anteriores resultados observamos que el promedio de edad esta alrededor de los 43 años, que el 50%

de la población es menor a los 42 años, que, en promedio, el resto de las personas están a 14 años de 42 años, y que el 50% de la población centrada se encuentra en el intervalo [23, 61] años, y finalmente la variabilidad de las edades es mayor a 192. Si quisiéramos representar lo anterior en una gráfica, podemos usar una gráfica de barras.

```
llamadas.edad <- linea.mujeres %>% group_by(EDAD) %>% count()
ggplot(llamadas.edad)+
  geom_col(aes(x = EDAD, y = n, fill = EDAD)) +
  ggtitle("Edades de las usuarias de Línea mujeres") +
  theme_minimal() +
  labs(
    x = "Edades",
    y = "Total de llamadas"
  )
)
```



En la anterior gráfica, observamos que la distribución de las edades pareciera ser normal. De ahí que decidimos realizar un diagrama de densidad que valida nuestra hipótesis de normalidad, a pesar de tener algunos picos.

```
ggplot(linea.mujeres) +
  geom_density(aes(x=EDAD), color="darkblue", fill="lightblue")
```

