

Tarea

Análisis exploratorio de datos: Fecha de entrega 22 de marzo del 2021

Instrucciones: Genera un reporte en formato pdf dando respuesta a cada una de las siguientes preguntas.

El reporte deberá cumplir con los siguientes requisitos para tener derecho a ser calificado: 1) nombres de los integrantes del equipo (máximo 3 alumnos) en la parte superior derecha de la primera hoja sin ocupar mucho espacio, 2) tener excelente presentación, 3) máximo 10 hojas, 4) sin desperdiciar espacios, 5) gráficas bien diseñadas, cuidando escalas y espacios y 6) justificando con prosa de forma concisa, bien pensada y escrita sus respuestas a cada pregunta.

1. **Censo de Población y Vivienda 2020.** Cada 10 años el Instituto Nacional de Estadística y Geografía (INEGI) levanta el Censo de Población y Vivienda con el objetivo de conocer varias características de los habitantes de México y sus viviendas a nivel nacional, estatal, municipal, por localidad, por grupos de manzanas y hasta por manzana. Descarga la base de datos con los Principales resultados por localidad (ITER) de 2020 de los [Estados Unidos Mexicanos](#) y genera la siguiente información:

- 1.1. Una base de datos a nivel municipio extrayendo directamente los sub-totales para cada municipio y usa esta base para responder las siguientes preguntas.

- ¿De qué tamaño es tu base de datos? (*renglones × columnas* y peso en MB).
- No filtres y agregues sobre las localidades para cada municipio. ¿Por qué esto sería una pésima idea?
- No pongas la base en el reporte, sólo contesta las dos preguntas anteriores.

- 1.2. El porcentaje de población económicamente activa¹ para cada una de las 32 entidades federativas, así como para las 16 alcaldías que componen la CDMX. Presenta los resultados mediante gráficas de barras y describe lo que te parezca más importante. **Importante:** Calcula estos porcentajes tomando como base la población de 12 años y más en cada estado u alcaldía.

- 1.3. El porcentaje de población de 5 años y mas que habla alguna lengua indígena así como el porcentaje de población de 15 años y más analfabeta. Lo anterior para cada uno de municipios que componen nuestro país. Los porcentajes se tienen que calcular con respecto a las poblaciones correspondientes de cada municipio (de forma análoga a lo que se hizo en el inciso anterior). ¿Existe alguna relación entre ambas variables? (realiza un diagrama de dispersión)

2. Pandemia de COVID-19 en la CDMX.

2.1. Hospitalizaciones

- 1) A partir de la fecha de ingreso (variable `fecha_ingreso` en formato `as.Date`) primero junio de 2020 (un lunes, esto es importante) calcula el

$$\text{porcentaje de hospitalizaciones por día} = \frac{\text{hospitalizados por día (de los infectados)}}{\text{infectados de COVID-19} \times \text{día}}$$

Genera una gráfica de serie de tiempo con esta información (en el eje x la `fecha_ingreso` y en el eje y los %).

¹Población de 12 años y más económicamente activa

- 2) Utiliza la función `frollmean` de la librería `data.table` para obtener una media móvil con una ventana de 7 días. Genera una gráfica de serie de tiempo con esta información.
- 3) Comenta y compara las dos gráficas anteriores (es fácil interpretar la primer gráfica, la segunda, etc).
- 4) Observando con mucha atención la gráfica con los promedios móviles, quizá tengas que modificar tu escala (hacer un acercamiento sobre el eje y para ver claramente lo que sucede). Obtén cada una de las fechas en las que se observan crestas (pequeñas montañas). Investiga en internet y responde ¿qué sucedió en esas fechas? (fueron días festivos ¿cuáles?, periodos vacacionales ¿cuáles?, ¿cambios de color del semáforo?, etc.)

2.2. Comorbilidades en los infectados por COVID-19

- 1) Calcula la probabilidad condicional de fallecer por COVID-19 dado cada una de las comorbilidades en la base de datos ($P(F|C_j)$, $j = 1, 2, \dots$). Obviamente sólo considerando los enfermos de COVID-19. Haz una gráfica de barras y comenta al respecto.
- 2) Piensa muy bien la siguiente pregunta, ¿qué comorbilidad esta asociada a un mayor número de fallecimientos por COVID-19 en la CDMX? Justifica tu respuesta de la siguiente manera: sí en la CDMX hubiera 100,000 infectados por COVID-19, obtén el número promedio de fallecimientos asociados a cada comorbilidad.