

# Estudio de INP en México

Sarahi García, Ramón Ruíz\*  
Centro de Investigación en Matemáticas CIMAT\*\*

11 de Diciembre de 2023

## I. PREGUNTAS DE INTERES

- ¿Cómo se relaciona la frecuencia de viajes de un usuario con su edad?
- ¿El género tiene un efecto significativo en la duración y/o frecuencia de los viajes?

## II. INTRODUCCIÓN

El Índice Nacional de Precios al Consumidor (INPC) es un indicador de la evaluación y comprensión de la dinámica económica des país. Esta diseñado para estimar la evolución de los precios de los bienes y servicios consumidos por las familias de México.

Existe complejidad inherente a la medición de las variaciones de precios pues, entre otros factores, existe una amplia gama de artículos y servicios de consumo así como una constante fluctuación en los precios. Además, la asincronía en los cambios de precios y sus variaciones de velocidad agregan un nivel adicional de complejidad al análisis.

En este contexto, el presente se enfoca en analizar los precios de una amplia gama de productos en diversas ciudades de México. Utilizando datos muestrales, exploremos las fluctuaciones en los precios de nueve productos a lo largo del tiempo y en diferentes ubicaciones geográficas del país. Empleamos técnicas de análisis estadístico y visualización de datos, incluyendo dendrogramas y heatmaps, para identificar patrones y relaciones significativas en los datos.

La información recopilada para este análisis proviene de la base de datos proporcionada por una de las herramientas de del INPC del INEGI: **Consulta precios promedio**, que contiene registros de precios de distintas marcas y/o proveedores de una serie de productos y servicios en 55 ciudades que se consumen en todo el país. Las cotizaciones de los productos en esta base de datos son principalmente mensuales y se encuentran desde el mes de Agosto del año 2018.

Debido a la extensa cantidad de productos y servicios nos centraremos exclusivamente en los siguientes productos

mostrados en 1, desde el mes de agosto de 2018 hasta el mes de marzo de 2024 de las 55 ciudades. Para un mejor manejo y representación visual de los datos se modificaron, y/o quitaron algunas de las filas, de modo que para cada producto/servicio tenemos un único proveedor/marca.

| Anio | Mes | Clave ciudad | limon | leche | camaron | tortillas | electricidad | alquiler | internet | frijol | pollo |
|------|-----|--------------|-------|-------|---------|-----------|--------------|----------|----------|--------|-------|
| 2018 | 8   | 1            | 33.68 | 19.00 | 175.00  | 14.00     | 417.35       | 100.24   | 5.0      | 30.00  | 80.75 |
| 2018 | 8   | 2            | 14.23 | 16.58 | 178.00  | 16.00     | 192.79       | 100.03   | 25.0     | 22.78  | 90.00 |
| 2018 | 8   | 3            | 13.75 | 16.00 | 135.00  | 11.50     | 417.35       | 100.08   | 399.0    | 28.67  | 81.00 |
| 2018 | 8   | 4            | 20.40 | 11.17 | 322.00  | 10.05     | 417.35       | 100.24   | 8.0      | 30.29  | 35.00 |
| 2018 | 8   | 5            | 20.68 | 15.47 | 225.48  | 16.00     | 192.79       | 100.16   | 299.0    | 38.78  | 59.24 |
| ...  | ... | ...          | ...   | ...   | ...     | ...       | ...          | ...      | ...      | ...    | ...   |
| 2024 | 2   | 51           | 37.50 | 25.00 | 190.00  | 18.00     | 540.59       | 106.18   | 10.0     | 50.00  | 79.75 |
| 2024 | 2   | 52           | 39.15 | 27.50 | 170.00  | 23.00     | 540.59       | 111.11   | 5.0      | 55.00  | 89.00 |
| 2024 | 2   | 53           | 35.00 | 23.00 | 324.00  | 24.50     | 540.59       | 116.90   | 299.0    | 48.78  | 68.00 |
| 2024 | 2   | 54           | 19.00 | 21.50 | 140.00  | 23.50     | 443.56       | 105.68   | 499.0    | 38.00  | 50.00 |
| 2024 | 2   | 55           | 36.75 | 18.89 | 225.00  | 20.00     | 540.59       | 111.03   | 20.0     | 37.50  | 80.00 |

Figura 1. Datos utilizados para este trabajo

## III. ANÁLISIS EXPLORATORIO

### Usuarios

A continuación se presentan algunas de las carterísticas de los usuarios de MiBici del conjunto de datos mencionado hay **17282 distintos usuarios** que utlizaron, de los cuales 11668 son hombres y 5384 son mujeres cantidad-genero.

Se puede apreciar en edad que la mayoría de los usuarios son adultos entre 25 y 40 años. La edad mínima registrada es de 20 años y ls máxima de 106.

Por otro lado, si hacemos la distinción de edades por género, como muestran los siguientes boxplots edadygen podemos observar que en el género masculino el rando intercuantil abarca desde 30 hasta más de 40 años mientras que el de las mujeres inicia casi igual pero no alcanza los 40 años.

### Viajes

Algunas características que podemos obtener sobre los viajes a partir de los datos son los horarios y días de la semana más concurridos, la duración de los viajes y las estaciones más y menos utilizadas. A continuación dare-

\* yesenia.garcia@cimat.mx, ramon.ruiz@cimat.mx

\*\* REP, a cargo de Dr. Johan Van horebeek



mos un panorama general sobre los viajes que se realizan con MiBici.

De las gráficas día, hora y duracion podemos observar que la cantidad de viajes disminuye significativaente los fines de semana, la hora en que los usuarios más utilizan Mibici es las 18:00 y desde esa hora donde se alcanza el máximo, el número de viajes decrece hasta su mínimo a las 23:00. Además la mayoría dura entre 6 y 14 minutos, en promedio, duran 11.

En la figura día, la diferencia más relevante se aprecia entre el fin de semana y el resto de los días, aunque sí se observa un crecimiento de lunes a viernes.

Los máximos más relevantes de la figura hora, se encuentran a las 8:00 y a las 18:00 y por la noche, después de las 20:00 se tiene concurrencia. No hay datos desde las 12:00 hasta las 6:00.

En la primer gráfica de la figura duracion, se aprecia la distribución de duración de los viajes, sin embargo, debido a que hay algunos datos que indican viajes de hasta 500 minutos, no se puede apreciar el lugar donde se encuentra la mayor concentración de datos, por esta razón en la gráfica de abajo se realizó un zoom al intervalo [0,40], de esta manera puede observarse mejor el histograma. Hay un pico significativo al inicio, lo que indica que una alta cantidad de viajes dura menos de uno o dos minutos.

Podemos extraer un poco más de información si formamos un nuevo DataFrame donde agrupamos los datos por usuario y agregamos una columna que cuente el número de viajes de cada usuario que haya utilizado MiBici en Enero de 2019 y otra que indique el grupo de edad al que pertenece cada usuario (rangos=[20,25, 30,35, 40,45, 50,55, 60,65, 70,75, 80]).

De esta manera podemos entender mejor el comportamiento general por usuario, por ejemplo, la figura cantidad nos muestra que la mayoría de los usuarios sólo realizó un viaje en el mes elegido. De hecho, observamos que la frecuencia de cantidad de viajes vs disminuyendo. Aplicamos la función describe() al DataFrame anterior y obtuvimos, además, que el promedio de viajes de un usuario fue de 20 y cerca de la mitad de los usuarios realizaron entre 5 y 29 viajes.

También podemos saber cómo se distribuye el número de viajes por usuario que se realizan en un mes de acuerdo con el grupo de edad o del género de éste:

La Figura porgenero es interesante, pues en contraste con la cantidad de usuarios por género cantidadgenero (donde hay una gran diferencia entre la cantidad de hombre y mujeres registrados), la distribución de cantidad de viajes que se realizan por usuario no parece ser significativamente distinta entre hombres y mujeres.

Finalmente, graficamos un pairplot para darnos una idea de cómo se relacionan las variables y si es que hay correlación entre algunas de ellas.

Como puede observarse no se aprecia ninguna relación clara entre ningún par de variables y tampoco parece haber agrupamientos por género.

## IV. RESULTADOS

Se busca responder si hay alguna relación entre la frecuencia y/o duración de viajes con la edad y/o el género, por lo que a continuación probaremos con una regresión lineal.

Primero veamos un scatterplot de estas variables:

Ninguno de los dos gráficos parece mostrar una relación entre las variables, sin embargo, probaremos una regresión lineal utilizando dos distintos enfoques para determinar qué sucede.

Resalta que  $R^2 = 0$ , lo que indica que nuestro modelo no explica la variabilidad de nuestros datos, de hecho, da lo mismo hacer la regresión que simplemente tomar la media de nuestros datos, cosa que se aprecia en el siguiente gráfico:

Se intentó una segunda manera de probar si existe alguna relación entre la edad y la duración de los viajes. En lugar de utilizar todas las edades, éstas se agruparon en 5 categorías (las utilizadas en Fig porgrupo) y utilizando la clasificación onehot t2 se realizó nuevamente la regresión lineal.

A continuación se muestra el resumen de esta regresión. Si bien  $R^2$  ya no es cero sigue siendo un valor muy bajo, por lo que el modelo no es adecuado para la distribución de nuestros datos.

Las dos anteriores maneras de encontrar una relación con la edad se utilizaron también para la duración de los viajes, aunque los resultados son análogos a los ya presentados.

## V. CONCLUSIONES

Con un modelo lineal no puede establecerse si hay o no una relación entre la edad y la frecuencia y/o duración de los viajes. En cuanto al género, desde el análisis exploratorio pudimos apreciar que no parece un factor realmente importante para estas dos variables, a pesar de que haya menos mujeres registradas, la distribución de los tiempos y cantidad de viajes es similar a la de los hombres.