

**Proyecto Final - Análisis de datos**  
**Edgar Osvaldo López Zúñiga**  
**Giovanni Gamaliel López Padilla**

# Índice

<b>1. Introducción</b>	<b>1</b>
1.1. MiBici . . . . .	2
<b>2. Resultados</b>	<b>2</b>
2.1. Distrución del uso de las estaciones . . . . .	2
2.2. Características demográficas de usuarios . . . . .	4
2.2.1. Proporción de usuarios por género . . . . .	4
2.3. Distancia promedio recorrida por los usuarios. . . . .	6
2.4. Promedio mensual por hora . . . . .	6
2.5. Promedio diario semanal por hora . . . . .	7
2.6. Tiempo de uso promedio de los usuarios . . . . .	8
2.6.1. Promedios mensuales por hora . . . . .	8
2.6.2. Promedio diario semanal por hora . . . . .	9
2.7. Número promedio de viajes . . . . .	10
2.7.1. Promedio mensual por hora . . . . .	10
2.7.2. Promedios diarios semanales por hora . . . . .	11

## 1. Introducción

Desde inicios del milenio, la población en América Latina ha tenido un crecimiento semejante a una exponencial [1]. Esto genera problemas de aglomeración urbana, distribución de suelo, movilidad urbana privdo y público, entre otros. Las metrópolis de América Latina tienen retos difíciles, ya que sus problemas se ven reflejados en el tiempo y distancia de traslado que realiza cada habitante [2]. Por lo que el problema de transporte y movilidad urbana es uno de los factores más importantes para las administraciones, siendo un pilar fundamental en el desarrollo social y económico. La aplicación del principio de comodalidad plantea favorecer la promoción e implementación de distintas alternativas que satisfagan las necesidades de transporte, garantizando cobertura, conectividad, flujos continuos, seguridad y eficiencia [3].

Frente a estos problemas en el transporte público, varias metrópolis de América Latina han comenzado a implementar el principio de comodalidad. La bicicleta es un medio de transporte

alternativo, que dependiendo la implementación, puede llegar a ser más rápido, comodo y seguro en comparación a los demás medios de transporte disponibles.

La congestión del tráfico en las grandes ciudades presentan graves problemas de movilidad. Entre las causas que crean esta congestión son la falta de planeación y la desarrollo en la infraestructura y la alta densidad poblacional. Una correcta implementación de un sistema de bicicletas aporta a la disminución de la congestión del tráfico.

En el mundo existen alrededor de 400 sistemas de bicicletas disponibles al público. Cada sistema tiene particularidades y tecnologías que se ajustan a las necesidades de la región y habitantes. La implementación de estos sistemas se debe realizar sobre un estudio que incluye diferentes factores para que se aporte de una manera eficiente hacia la disminución del tráfico.

El área metropolitana de la Ciudad de México tiene un sistema de transporte público que integra 11 líneas de metro, 7 de autobuses (Metrobus), autobuses no integrados y sistema de bicicleta pública. ECOBICI es el sistema de bicicletas públicas compartidas de la Ciudad de México. El sistema permite que a los usuarios registrados tomar una bicicleta de cualquier estación y devolverla a la más cercana a su destino en trayectos de 45 minutos.

## 1.1. MiBici

Guadalajara es la segunda metrópolis más importante de México. Su sistema de transporte incluye 2 líneas de tren ligero, una línea de autobuses integrados, autobuses no integrados y el sistema de bicicletas públicas *MiBici*.

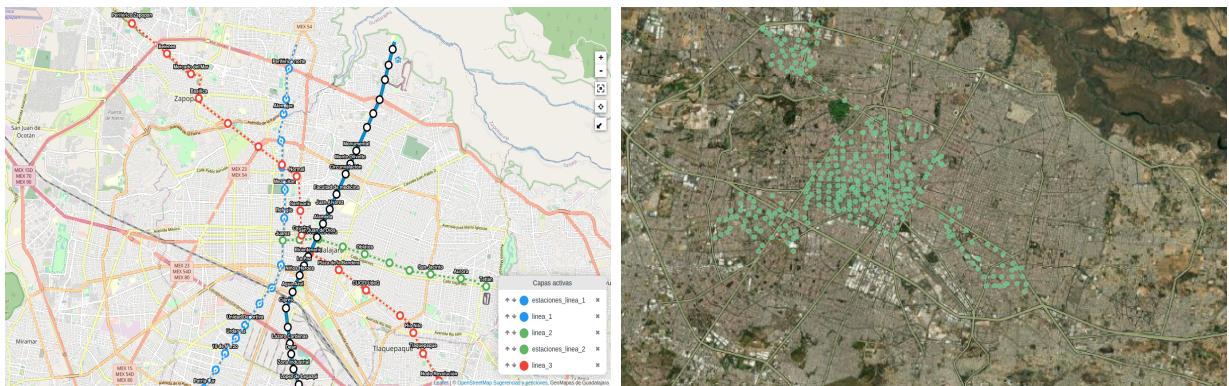
MiBici es el sistema público de bicicletas de la ciudad de Guadalajara. Es un sistema el cual está diseñado para realizar recorridos cortos de manera eficiente tomando en cuenta los siguientes puntos:

- Instalaciones de las estaciones en zonas propicias para el sistema.
- Delimitando los polígonos de acción más apropiados.
- Estudiando las variables de demanda para el diseño de la red de estaciones.

## 2. Resultados

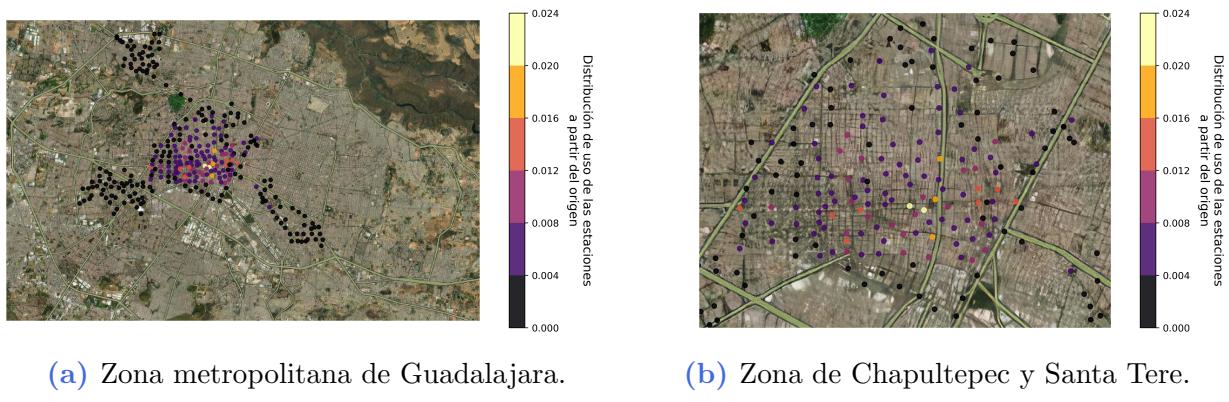
### 2.1. Distribución del uso de las estaciones

En la figura 1 se visualiza las líneas del metro y del metrobus. Como se mencionó en la introducción, el sistema público de bicicletas está enfocado en disminuir la congestión del tráfico usando vías adecuadas y en puntos apropiados. Esto se ve reflejado en las estaciones de MiBici, esto es debido a que la mayoría se encuentra en los polígonos de intervención urbana especial (PIUE), las cuales son áreas enfocadas al desarrollo social, medioambiental y económico [4]. Aunado a esto, las estaciones que están fuera de estos polígonos, se encuentran alrededor de las líneas del metro. Esto puede ser debido a que fueron instaladas para suplir el uso del metro cuando no sea eficiente para el usuario.



**Figura 1:** Mapa de la zona metropolitana de Guadalajara señalando las líneas del metro (izquierda) y las estaciones de MiBici (derecha).

En la figura 2a se puede notar como las estaciones localizadas en los PIUE son las más usadas con respecto a las estaciones que se encuentran en las orillas de la zona metropolitana. En la figura 2b se visualiza que las estaciones que se encuentran sobre la Avenida Ignación L. Vallarta y las que se encuentran en las intersecciones de las líneas de metro.

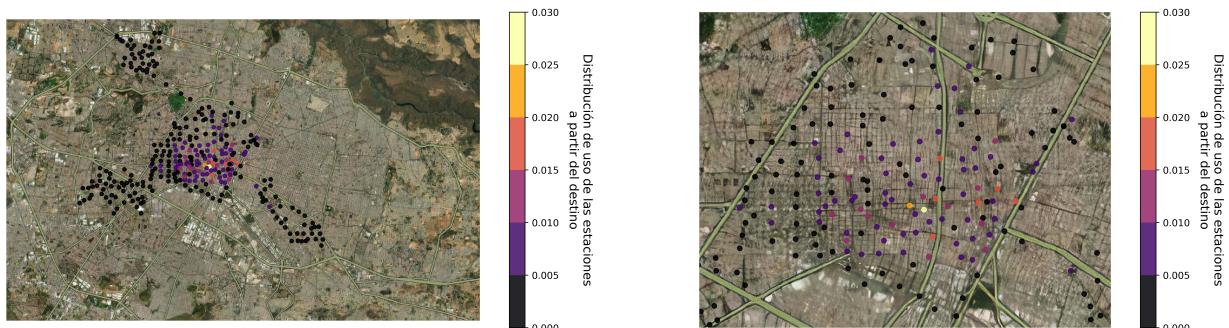


(a) Zona metropolitana de Guadalajara.

(b) Zona de Chapultepec y Santa Tere.

**Figura 2:** Distribución del uso de las estaciones de MiBici como punto de origen.

En los casos de la distribución de las estaciones que son usadas como destino del viaje se obtiene la figura 3. Comparando las figuras 2 y 3 se visualiza que las estaciones que se encuentran en las intersecciones de las líneas del metro son las más usadas. Esto puede ser un reflejo que el servicio de bicicletas esta siendo usado para transportarse en distancias cortas y auxiliando para disminuir la aglomeración de personas en el transporte público.



(a) Zona metropolitana de Guadalajara.

(b) Zona de Chapultepec y Santa Tere.

**Figura 3:** Distribución del uso de las estaciones de MiBici como punto de destino.

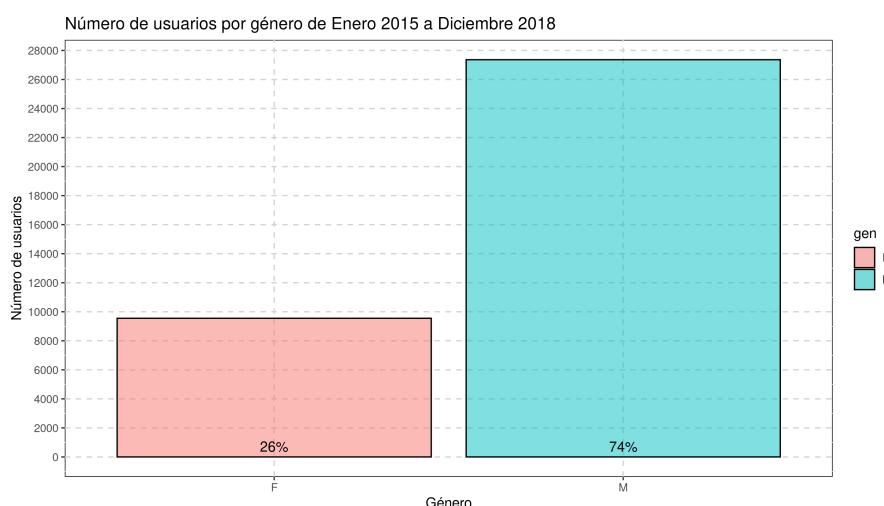
## 2.2. Características demográficas de usuarios

En esta sección se hará un análisis de las características demográficas de los usuarios del sistema MiBici. En particular se estudiará la distribución de edades del conjunto de todos los usuarios, la proporción de usuarios por cada género, así como las distribuciones de edades para los usuarios de género masculino y femenino por separado.

Al igual que en el resto del reporte, se limitó el estudio de los datos al intervalo de tiempo desde enero del año 2015 hasta diciembre de 2018.

### 2.2.1. Proporción de usuarios por género

En total, el número de usuarios diferentes que utilizaron el sistema en el intervalo de enero de 2015 a diciembre 2018 fue 36917 de los cuales 2 no registraron su género, 27363 son de género masculino y 9552 son de género femenino. Esto puede ser visualizado en la figura 4.

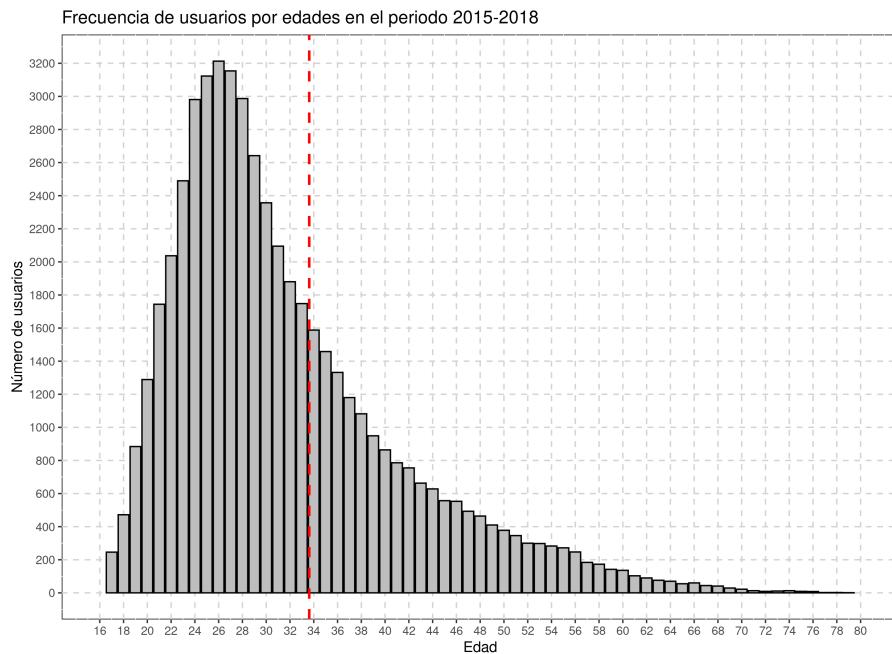


**Figura 4:** Número de usuarios por género de Enero 2015 a Diciembre 2018.

En donde se puede ver que el 74 % de los usuarios son hombres y 26 % son mujeres. Se deja fuera de la gráfica a los dos usuarios que no registraron su género debido a que su contribución a las proporciones es muy pequeña.

### Número de usuarios por edad

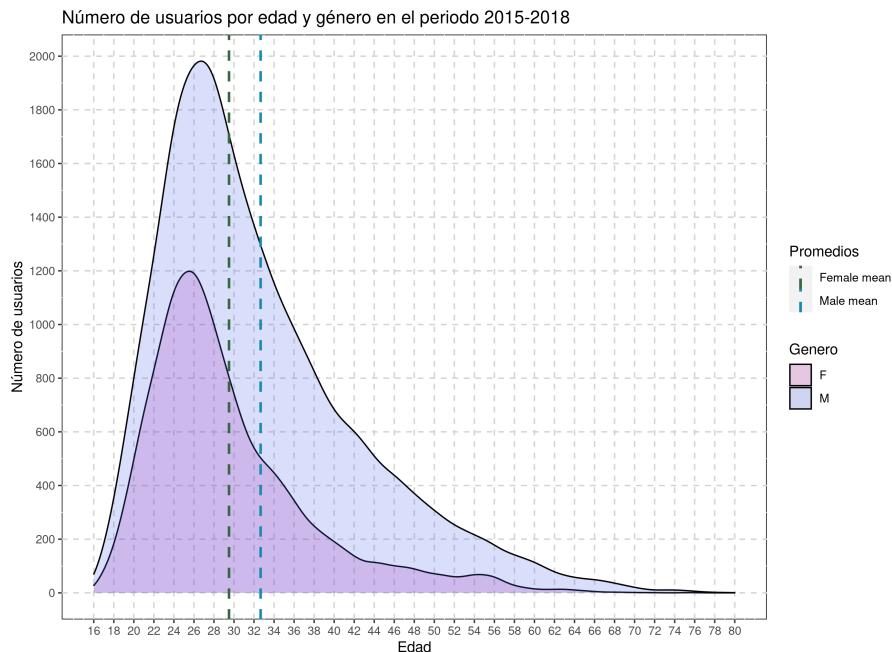
En la figura 5 se puede observar la distribución de los usuarios por edades al momento del primer registro en el sistema en el intervalo de 2015 a 2018. Para la realización de esta figura se tomó en cuenta solo la primera aparición de todos los usuarios distintos en ese intervalo de tiempo.



**Figura 5:** Distribución de usuarios por edad en el periodo de 2015 a 2018.

Uno de los requisitos para poder adquirir una suscripción al sistema MiBici es ser mayor de 16 años, y en caso de ser menor de edad, ser acompañado por un parent o tutor. Es por esto que la edad mínima registrada es de 16 años.

Si separamos a los usuarios por género, las distribuciones de edades serían como en la figura 6. En donde se puede observar que para ambos géneros, se tiene una distribución similar.



**Figura 6:** Distribución de edad de los usuarios, separados por género en el periodo de 2015 a 2018.

En la figura 6 como en la figura 4 se puede notar que el número de usuarios de género masculino es mayor que los de género femenino, además, es posible ver que hay una diferencia en la edad

promedio de los usuarios de cada género. Siendo la media de la edad para el género femenino de 29.50 años y para el género masculino de 32.66 años.

### 2.3. Distancia promedio recorrida por los usuarios.

Los datos abiertos contiene la localización geográfica de cada estación de MiBici. Para calcular entre estaciones se uso la ecuación 1.

### 2.4. Promedio mensual por hora

$$\begin{aligned}\beta &= \sin^2\left(\frac{|\Delta\theta|}{2}\right) + \cos(\theta_1)\cos(\theta_2)\sin^2\left(\frac{|\Delta\phi|}{2}\right) \\ D &= 2R \arcsin(\sqrt{\beta})\end{aligned}\quad (1)$$

donde R es el radio de la tierra,  $\theta_i$  es la latitud y  $\phi_i$  es la longitud.

En base a las distancias calculadas en cada viaje realizado en el periodo 2015 al 2018. El calculo del promedio mensual por hora de las distancias recorridas en cada viaje se realizó con la ecuación 2.

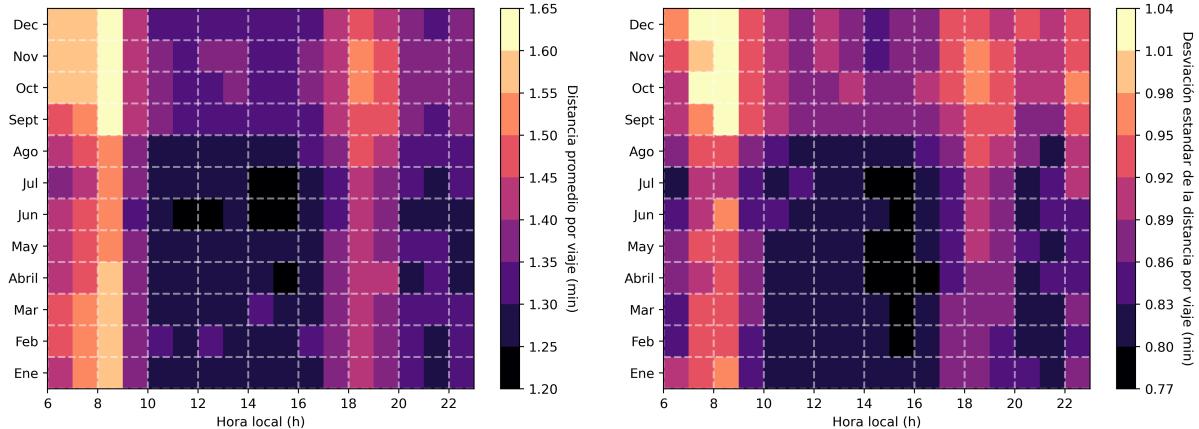
$$\bar{D}_{m,h} = \frac{1}{n_{(m,h)}} \sum_{i=1}^{n_{(m,h)}} \hat{D}_i \quad \begin{array}{l} m = 1, 2, \dots, 12 \\ h = 6, 7, \dots, 23 \end{array} \quad (2)$$

donde  $\hat{D}$  y  $n_{(m,h)}$  es el conjunto y la cantidad total de viajes realizados para el mes  $m$  realizados entre las  $h$  y  $h+1$  horas respectivamente. De una manera semejante, se calculo la varianza de este conjunto con la ecuación 3.

$$\sigma_{D_{(m,h)}}^2 = \frac{1}{n_{(m,h)} - 1} \sum_{i=1}^{n_{(m,h)}} (D_i - \bar{D}_{(m,h)})^2 \quad \begin{array}{l} m = 1, 2, \dots, 12 \\ h = 6, 7, \dots, 23 \end{array} \quad (3)$$

En la figura 7 se muestran los resultados de aplicar las ecuaciones 2 y 3 a los datos de las distancias.

En la figura 7a se visualiza que existe una disminución de su uso entre las 10 y 16 horas a lo largo del año y entre los meses mayo y julio en cualquier horario. Este último coincide con la temporada del año donde se presentan más lluvias [5]. Con esto, tenemos que existe un mínimo de la distancia recorrida entre las 14 y 16 horas en los meses de junio y julio. Con ayuda de la figura 7b se obtiene que la distancia recorrida por los usuarios entre los meses de septiembre y diciembre tienen una mayor variación y esto puede deberse a la baja temperatura [5]. Ya que los usuarios al sentir menos presión en sus cuerpos por la temperatura pueden desviarse de sus caminos usuales.



(a) Promedio mensual de la distancia recorrida. (b) Desviación estandar de la distancia recorrida.

**Figura 7:** Distancia promedio y desviación estandar mensual por hora recorrida por lo usuarios calculadas con las ecuaciones 2 y 3.

## 2.5. Promedio diario semanal por hora

Otra manera de abordar los datos es realizando un promedio del día semanal por las horas del día. Esto se obtuvo aplicando la ecuación 4 a los datos.

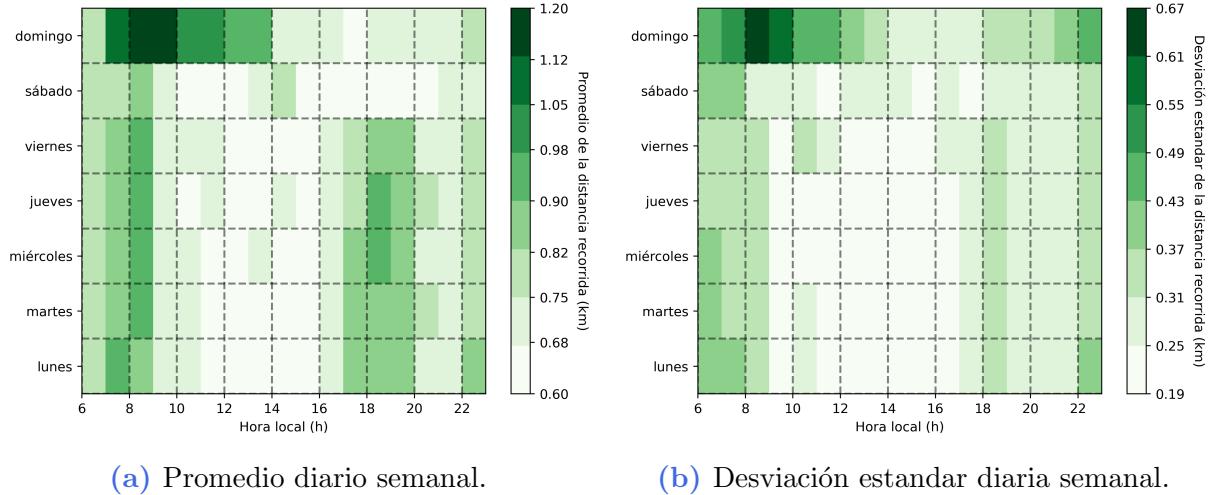
$$\bar{D}_{d,h} = \frac{1}{n_{(d,h)}} = \sum_{i=1}^{n_{(d,h)}} \hat{D}_i \quad d = 1, 2, \dots, 7 \quad h = 6, 7, \dots, 23 \quad (4)$$

donde  $D_i$  y  $n_{(d,h)}$  es el conjunto y la cantidad total de viajes realizados para el día de la semana  $d$  realizados entre las  $h$  y  $h + 1$  horas respectivamente. De una manera semejante, se calculo la varianza de este conjunto con la ecuación 5.

$$\sigma_{D_{(d,h)}}^2 = \frac{1}{n_{(d,h)} - 1} \sum_{i=1}^{n_{(d,h)}} (D_i - \bar{D}_{(d,h)})^2 \quad d = 1, 2, \dots, 7 \quad h = 6, 7, \dots, 23 \quad (5)$$

En la figura 8 se muestran los resultados de aplicar las ecuaciones 4 y 5 a los datos de las distancias.

En la figura 8 se aprecia que el uso de la bicicleta se incrementa entre semana a las 7:00 horas y 15:00 horas. Existe una disminución de la distancia recorrida entre las 12 y 16 horas, algo que ya se habia previsto en la figura 8a. Se presenta en el periodo de las 8 y 10 horas los días domingo, dando indicios que en ese periodo se usan de manera recreativa. Este indicio tambien nos lo da la figura 8b, ya que la variación en ese periodo de tiempo es mayor, y esto puede ser debido a que son usadas en diversas actividades. En cambio entre semana la varianza es poca, dando la impresión que los usuarios realizan una actividad semejante como lo es el transporte de un lugar hacia otro.



**Figura 8:** Distancia promedio y desviación estandar diaria semanal por hora recorrida por los usuarios calculadas con las ecuaciones 4 y 5.

## 2.6. Tiempo de uso promedio de los usuarios

Los datos abiertos de MiBici contienen el tiempo inicial y final del uso de una bicicleta a través de dos estaciones, las cuales pueden ser la misma.

### 2.6.1. Promedios mensuales por hora

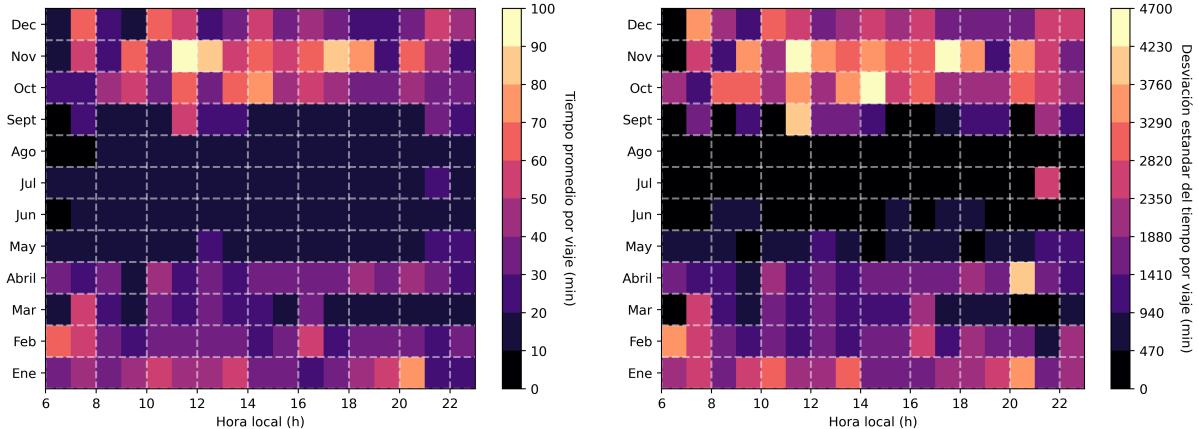
$$\bar{T}_{m,h} = \frac{1}{n_{(m,h)}} \sum_{i=1}^{n_{(m,h)}} \hat{T}_i \quad m = 1, 2, \dots, 12 \quad h = 6, 7, \dots, 23 \quad (6)$$

donde  $\hat{T}$  y  $n_{(m,n)}$  es el conjunto de tiempos y la cantidad total de viajes realizados para el mes  $m$  realizados entre las  $h$  y  $h + 1$  horas respectivamente. De una manera semejante, se calculó la varianza de este conjunto con la ecuación 7.

$$\sigma_{T_{(m,h)}}^2 = \frac{1}{n_{(m,h)} - 1} \sum_{i=1}^{n_{(m,h)}} (T_i - \bar{T}_{(m,h)})^2 \quad m = 1, 2, \dots, 12 \quad h = 6, 7, \dots, 23 \quad (7)$$

En la figura 9 se muestran los resultados de aplicar las ecuaciones 6 y 7 a los datos de los tiempos de uso.

En la figura 9a se visualiza que existe un aumento en el tiempo de uso entre los meses octubre y diciembre. Este último coincide con el fin de la temporada con más frecuencia de lluvias y una disminución de la temperatura [5]. En la misma figura se presenta un mínimo entre los meses de mayo y agosto. Con ayuda de la figura 9b se obtiene que el tiempo de uso de los usuarios entre los meses de octubre y diciembre tienen una mayor variación y esto puede deberse a que la baja temperatura y baja probabilidad de lluvia [5]. Ya que los usuarios pueden sentir menos preocupación de las vías públicas y el usar la bicicleta representa más un momento de descanso que estar alerta de su alrededor.



(a) Promedio mensual del tiempo de uso.

(b) Desviación estandar del tiempo de uso.

**Figura 9:** Tiempo de uso promedio y desviación estandar mensual por hora de los usuarios calculadas con las ecuaciones 6 y 7.

### 2.6.2. Promedio diario semanal por hora

Otra manera de abordar los datos es realizando un promedio del día semanal por las horas del día. Esto se obtuvo aplicando la ecuación 4 a los datos.

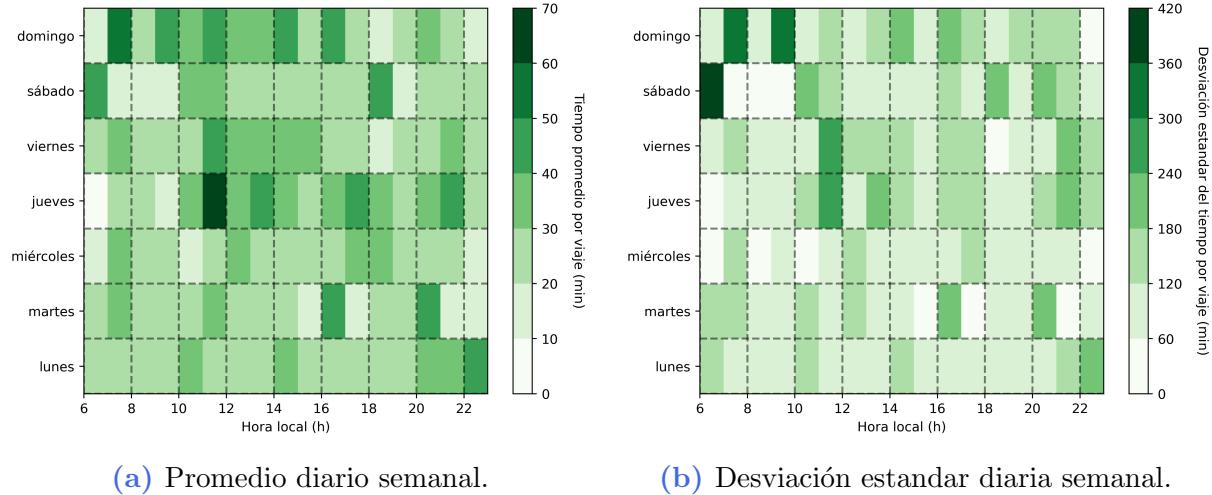
$$\bar{T}_{d,h} = \frac{1}{n_{(d,h)}} = \sum_{i=1}^{n_{(d,h)}} \hat{T}_i \quad d = 1, 2, \dots, 7 \quad h = 6, 7, \dots, 23 \quad (8)$$

donde  $T_i$  y  $n_{(d,h)}$  es el conjunto y la cantidad total de viajes realizados para el día de la semana  $d$  realizados entre las  $h$  y  $h + 1$  horas respectivamente. De una manera semejante, se calculó la varianza de este conjunto con la ecuación 9.

$$\sigma_{T_{(d,h)}}^2 = \frac{1}{n_{(d,h)} - 1} \sum_{i=1}^{n_{(d,h)}} (T_i - \bar{T}_{(d,h)})^2 \quad m = 1, 2, \dots, 12 \quad h = 6, 7, \dots, 23 \quad (9)$$

En la figura 10 se muestran los resultados de aplicar las ecuaciones 8 y 9 a los datos de las distancias.

En la figura 10 se aprecia que no hay alguna preferencia en los tiempos de uso en los días entre semana, por lo que se podría estimar un tiempo promedio por día para toda la semana. Con esto el factor más importante a tomar en cuenta es el mes del que se trate, ya que como se describió en la figura 9, sí existen variaciones a lo largo del año. En la figura 10b entre las 8 y 10 am se ve que existe una mayor variación en el tiempo de uso. Lo cual concuerda con lo antes mencionado en la figura 8b.



**Figura 10:** Tiempo de uso promedio y desviación estandar diaria semanal por hora de los usuarios calculadas con las ecuaciones 8 y 9.

## 2.7. Número promedio de viajes

Los datos abiertos de MiBici contienen un identificador de usuario, por lo que se puede saber el número de viajes realizados en un día por todos los usuarios.

### 2.7.1. Promedio mensual por hora

Se procedió a obtener un promedio mensual del número de viajes realizados por hora. El promedio mensual por hora se realizó usando la ecuación 10.

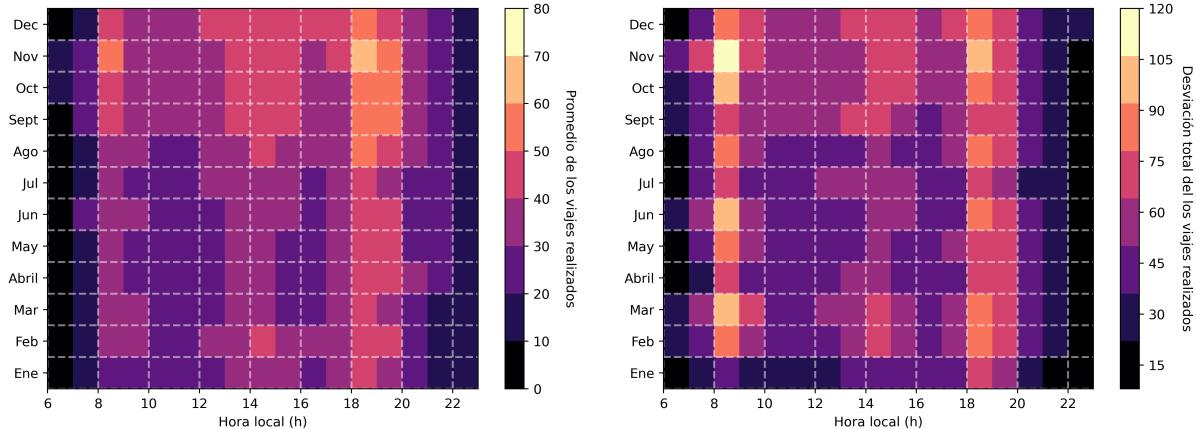
$$\bar{V}_{m,h} = \frac{1}{n_{(m,h)}} \sum_{i=1}^{n_{(m,h)}} \hat{V}_i \quad \begin{matrix} m = 1, 2, \dots, 12 \\ h = 6, 7, \dots, 23 \end{matrix} \quad (10)$$

donde  $\hat{V}$  y  $n_{(m,h)}$  es el conjunto y la cantidad total de viajes realizados para el mes  $m$  realizados entre las  $h$  y  $h + 1$  horas respectivamente. De una manera semejante, se calculó la varianza de este conjunto con la ecuación 11.

$$\sigma_{V_{(m,h)}} = \frac{1}{n_{(m,h)} - 1} \sum_{i=1}^{n_{(m,h)}} (V_i - \bar{V}_{(m,h)})^2 \quad \begin{matrix} m = 1, 2, \dots, 12 \\ h = 6, 7, \dots, 23 \end{matrix} \quad (11)$$

En la figura 11 se muestran los resultados de aplicar las ecuaciones 10 y 11 a el número de viajes.

En la figura 11 se visualiza que existe un aumento en el número de viajes entre los meses septiembre y diciembre. Lo cual coincide con lo antes planteado con las figuras 7 y 9. En la figura 11b se observa como el número de usuarios tiene una variación mayor a las 8, y 18 horas. Esto puede deberse a la elección de los usuarios por que servicio de transporte público les mejor tomar en un día.



(a) Promedio mensual del número de usuarios. (b) Desviación estandar del número de usuarios.

**Figura 11:** Tiempo de uso promedio y desviación estandar mensual por hora de los usuarios calculadas con las ecuaciones 10 y 11.

### 2.7.2. Promedios diarios semanales por hora

Otra manera de abordar los datos es realizando un promedio del día semanal por las horas del día. Esto se obtuvo aplicando la ecuación 12 a los datos.

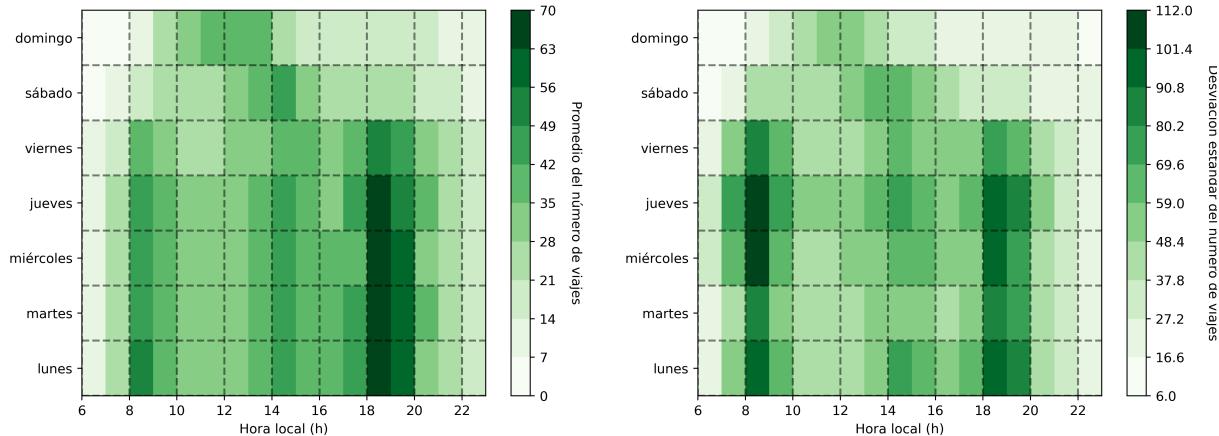
$$\bar{t}_{d,h} = \frac{1}{n_{(d,h)}} = \sum_{i=1}^{n_{(d,h)}} \hat{t}_i \quad d = 1, 2, \dots, 7 \quad h = 6, 7, \dots, 23 \quad (12)$$

donde  $V_i$  y  $n_{(d,h)}$  es el conjunto y la cantidad total de viajes realizados para el día de la semana  $d$  realizados entre las  $h$  y  $h + 1$  horas respectivamente. De una manera semejante, se calculó la varianza de este conjunto con la ecuación 13.

$$\sigma_{V_{(d,h)}} = \frac{1}{n_{(d,h)} - 1} \sum_{i=1}^{n_{(d,h)}} (V_i - \bar{V}_{(d,h)})^2 \quad m = 1, 2, \dots, 12 \quad h = 6, 7, \dots, 23 \quad (13)$$

En la figura 12 se muestran los resultados de aplicar las ecuaciones 12 y 13 a los datos de las distancias.

En la figura 12a se aprecia que existe una predilección en utilizar el servicio de bicicleta entre semana a las 8 y 16 horas. Existiendo un mínimo en el horario de 17 a 23 horas los días sábados y domingos. Con la figura 12b se observa que en el periodo donde existe un máximo, este presenta una gran variación, lo cual concuerda con lo encontrado en la figura 11b. Por ende, se puede suponer que los usuarios que utilizan el servicio en ese periodo de tiempo pueden no ser recurrentes y usarlo cuando les sea necesario, más no es su primera opción.



(a) Promedio diaria del número de usuarios semanal.  
(b) Desviación estandar del número de usuarios semanal.

**Figura 12:** Número de usuarios promedio y desviación estandar diaria semanal por hora calculados con las ecuaciones 12 y 13.

## Referencias

- [1] Comisión Económica para América Latina y el Caribe. Observatorio Demográfico de América Latina 2014: Proyecciones de población. CEPAL. 2015; Available from: <http://hdl.handle.net/11362/39228>.
- [2] Hall ET. La dimensión oculta. México : Siglo XXI; 1978.
- [3] Pastori E, Brambilla M, Maffii S, Vergnani R, Gualandi E, Skinner I. Modal shift in European transport: a way forward. TRAN Committee. 2018; Available from: [https://www.europarl.europa.eu/thinktank/en/document/IPOL-STU\(2018\)629182](https://www.europarl.europa.eu/thinktank/en/document/IPOL-STU(2018)629182).
- [4] Gobierno de Guadalajara. Plan Rector Polígono de Intervención Urbana Especial (PIUE) 13 Industrial; 2018. Available from: <https://transparencia.guadalajara.gob.mx/sites/default/files/GacetaTomoVEjemplar21Secc1aOctubre22-2018.pdf>.
- [5] Weather spark. El clima y el tiempo promedio en todo el año en Guadalajara;. Available from: <https://es.weatherspark.com/y/3866/Clima-promedio-en-Guadalajara-M%C3%A9jico-durante-todo-el-a%C3%B1o>.