

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

Índice

1. Introducción	2
2. Objetivos	2
2.1. Objetivo general	2
2.2. Objetivos particulares	3
3. Métodos y materiales	3
3.1. Base de datos	3
3.2. Promedio y varianza mensual por hora	4
3.3. Promedio y varianza diaria semanal por hora	4
3.4. Promedio por hora dada una fecha	4
4. Resultados	5
4.1. Distrución del uso de las estaciones	5
4.2. Características demográficas de usuarios	6
4.2.1. Proporción de usuarios por género	6
4.3. Distancia promedio recorrida por los usuarios	8
4.3.1. Promedio y desviación estandar mensual por hora	8
4.3.2. Promedio y desviación estandar diaria semanal por hora	8
4.3.3. Frecuencia de la distancia recorrida por viaje	9
4.4. Tiempo de uso promedio de los usuarios	10
4.4.1. Promedio y desviación estandar mensual por hora	10
4.4.2. Promedio y desviación estandar diaria semanal por hora	10
4.4.3. Distribución del tiempo de uso de una bicicleta	11
4.5. Número promedio de usuarios simultáneos	11
4.5.1. Promedio y desviación estandar mensual por hora	12
4.5.2. Promedios diarios semanales por hora	12
4.5.3. Distribución del número de usuarios simultáneo	13
5. Referencias	13

1. Introducción

Desde inicios del milenio, la población en Latinoamérica ha tenido un crecimiento semejante a una exponencial [1]. Esto genera problemas de aglomeración urbana, distribución de suelo y movilidad. Las metrópolis de Latinoamérica tienen retos difíciles. Sus problemas se ven reflejados en el tiempo y la distancia de traslado diarios que realiza cada habitante [2]. 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]. La congestión del tráfico en las grandes ciudades presenta graves problemas de movilidad. Entre las causas que crean esta congestión se encuentran la falta de planeación y la desarrollo en la infraestructura y la alta densidad poblacional.

Frente a estos problemas en el transporte público, varias metrópolis de Latinoamérica han implementado el principio de comodalidad usando a la bicicleta como medio de transporte alternativo, que dependiendo de la implementación, puede llegar a ser más rápido, cómodo y seguro en comparación a los demás medios de transporte disponibles. 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 sus 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 público de bicicletas. ECOBICI es el sistema público de bicicletas compartidas de la Ciudad de México. El sistema permite a los usuarios registrados tomar una bicicleta de cualquier estación y devolverla a otra más cercana a su destino en trayectos de 45 minutos.

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 público de bicicletas MiBici.

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

- Instalación de las estaciones en zonas propicias para el sistema.
- Delimitación de los polígonos de acción más apropiados.
- Estudio de las variables de demanda para el diseño de la red de estaciones.

2. Objetivos

2.1. Objetivo general

Encontrar patrones en el comportamiento de parámetros medibles como el tiempo de uso, distancia recorrida en un viaje y número de usuarios simultaneos en el sistema. Este análisis tiene como eje exponer aspectos del uso diario del sistema en conjunto con las características que tienen los usuarios de MiBici en el periodo 2015-2018.

2.2. Objetivos particulares

1. Analizar la distribución del uso de cada estación de bicicletas, para obtener la zona en donde existe una mayor frecuencia de uso.
2. Conocer a las características demográficas de los usuarios.
3. Obtener promedios y desviación estandar mensuales y darias semanales por cada hora entre las 6 y 24 horas de la distancia recorrida, tiempo de uso de una bicicleta y el número de usuarios simultaneos.
4. Obtener la distribución o frecuencia del promedio por hora entre las 6 y 24 horas de la distancia recorrida, tiempo de uso de una bicicleta y el número de usuarios simultaneos.

3. Métodos y materiales

3.1. Base de datos

MiBici cuenta con una página de datos abiertos en la que se puede acceder a una base de datos del uso de este sistema de transporte [4]. La base de datos está organizada en una serie de archivos CSV (valores separados por comas) con información mensual de los viajes realizados desde diciembre de 2014 hasta noviembre de 2021.

Los datos de los viajes están separados en archivos correspondientes a cada mes. Cada uno de los renglones de estos cuenta con la información de un viaje realizado. Los campos registrados para cada uno de los viajes en la base de datos son los siguientes:

- **Viaje_Id**
Número de identificación del viaje.
- **Usuario_Id**
Número único de identificación del usuario.
- **Genero**
- **Año_de_nacimiento**
- **Inicio_del_viaje**
Fecha de inicio del viaje en formato yyyy-mm-dd hh:mm.
- **Fin_del_viaje**
Fecha y hora de fin del viaje en formato yyyy-mm-dd hh:mm.
- **Origen_Id**
Número de identificación de la estación de salida.
- **Destino_Id**
Número de identificación de la estación de llegada.

Así mismo el sistema nos brinda un archivo CSV con la nomenclatura de las estaciones. Lo que nos permite utilizar los identificadores de origen y destino para acceder a información de la dirección, ubicación y estado de servicio de cada estación. Para este trabajo se utilizaron los datos contenidos en el periodo 2015-2018. Con el propósito de analizar datos en años completos y sin alteraciones realizadas por la pandemia provocada por el COVID-19 sucedida en marzo del 2020.

3.2. Promedio y varianza mensual por hora

Sea un conjunto $Q(m, h)$ de datos para un mes m entre las horas h y $h + 1$ con elementos q , entonces el promedio mensual por hora se calcula como:

$$\mu_{Q(m,h)} = \frac{1}{n(m,h)} \sum_{i=1}^{n(m,h)} q_i \quad (1)$$

y la varianza se calcula como:

$$\sigma_{Q(m,h)}^2 = \frac{1}{n(m,h) - 1} \sum_{i=1}^{n(m,h)} (q_i - \mu_{Q(m,h)})^2 \quad \begin{array}{l} m = 1, 2, \dots, 12 \\ h = 6, 7, \dots, 23 \end{array} \quad (2)$$

donde $n(m, h)$ es el número total de elementos del conjunto $Q(m, h)$.

3.3. Promedio y varianza diaria semanal por hora

Sea un conjunto $Q(d, h)$ de datos para un día de la semana d entre las horas h y $h + 1$ con elementos q , entonces el promedio diario semanal por hora se calcula como:

$$\mu_{Q(d,h)} = \frac{1}{n(d,h)} \sum_{i=1}^{n(d,h)} q_i \quad (3)$$

y la varianza se calcula como:

$$\sigma_{Q(d,h)}^2 = \frac{1}{n(d,h) - 1} \sum_{i=1}^{n(d,h)} (q_i - \mu_{Q(d,h)})^2 \quad \begin{array}{l} d = 1, 2, \dots, 7 \\ h = 6, 7, \dots, 23 \end{array} \quad (4)$$

donde $n(d, h)$ es el número total de elementos del conjunto $Q(d, h)$.

3.4. Promedio por hora dada una fecha

Sea un conjunto $Q(f, h)$ de datos para una fecha f con formato yyyy-mm-dd entre las horas h y $h + 1$ con elementos q , entonces el promedio por hora es calculado como

$$\mu_{Q(f,h)} = \frac{1}{n(f,h)} \sum_{i=1}^{n(f,h)} q_i \quad (5)$$

donde $n(f, h)$ es el número total de elementos del conjunto $Q(f, h)$.

4. Resultados

4.1. Distribución del uso de las estaciones

En la figura 1 se visualizan las líneas del metro (roja, azul marino y verde) y del metrobus (azul claro). 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 la figura 1, ya que la mayoría de las estaciones se encuentran en los polígonos de intervención urbana especial (PIUE), que son áreas enfocadas al desarrollo social, medioambiental y económico [5]. Las estaciones que están fuera de los PIUE, se encuentran alrededor de las líneas del metro o en zonas en donde el transporte público es deficiente (esquina inferior izquierda). Esto puede ser debido a que fueron instaladas para suplir el uso del metro cuando no sea eficiente para el usuario.

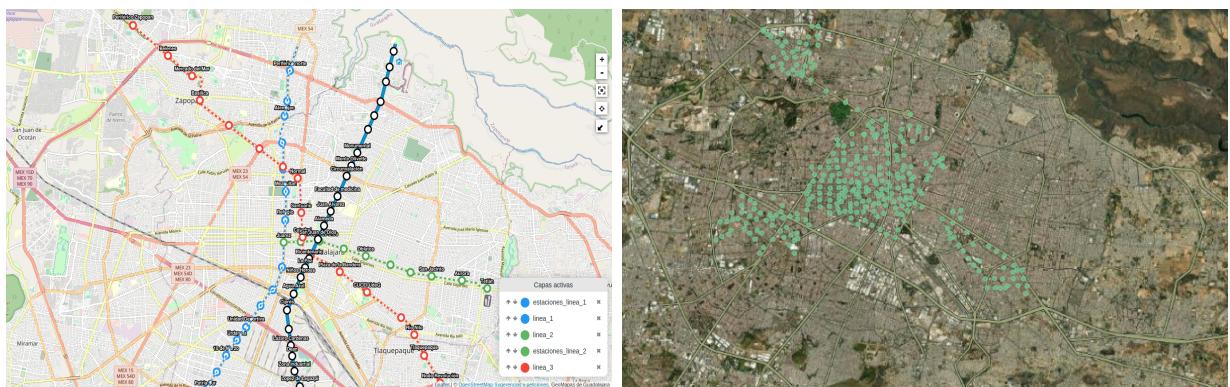
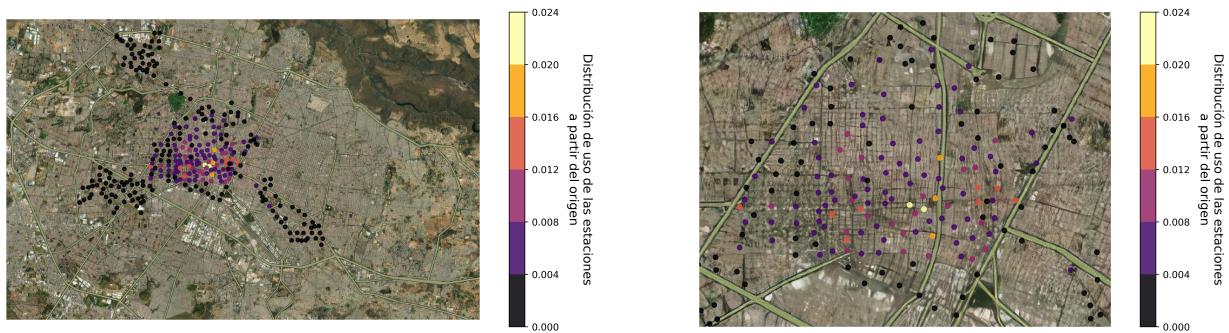


Figura 1: Mapa de la zona metropolitana de Guadalajara señalando las líneas del metro (izquierda) y las estaciones de MiBici (derecha).

En la figura 2a se puede notar que las estaciones localizadas en PIUE son las más usadas con respecto a las estaciones que se encuentran en los límites de la zona metropolitana. En la figura 2b se observa que las estaciones que se encuentran en la Avenida Ignación L. Vallarta y las que se encuentran en las intersecciones de las líneas de metro son las que reciben más uso por parte de los usuarios.



(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.

Al considerar 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 de que el servicio de

bicicletas es utilizado para transportarse en distancias cortas y como auxiliar para disminuir la aglomeración de personas en el transporte público.

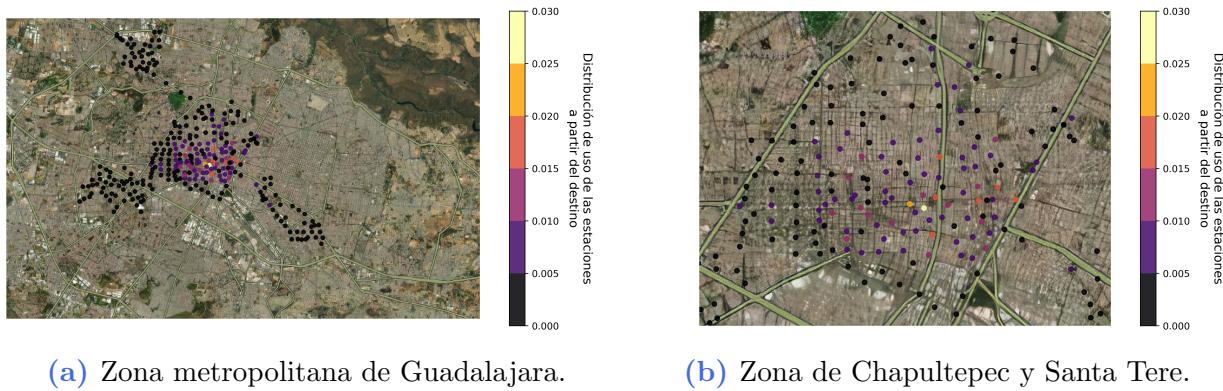


Figura 3: Distribución del uso de las estaciones de MiBici como punto de destino.

4.2. 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.

4.2.1. Proporción de usuarios por género

El número total 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. El número de usuarios de género masculino es de 27363 y 9552 son de género femenino. Esto puede ser visualizado en la figura 4.

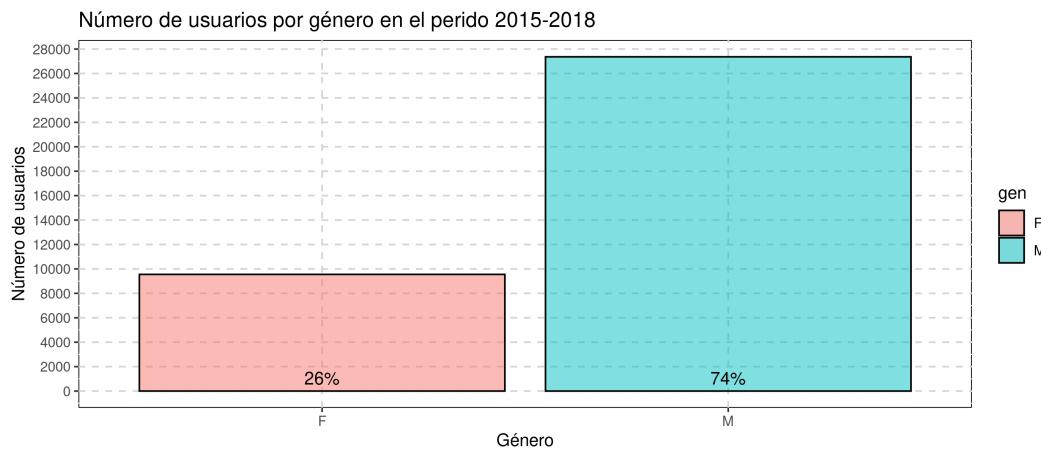


Figura 4: Número de usuarios por género de Enero 2015 a Diciembre 2018.

En donde se puede observar 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 de su 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.

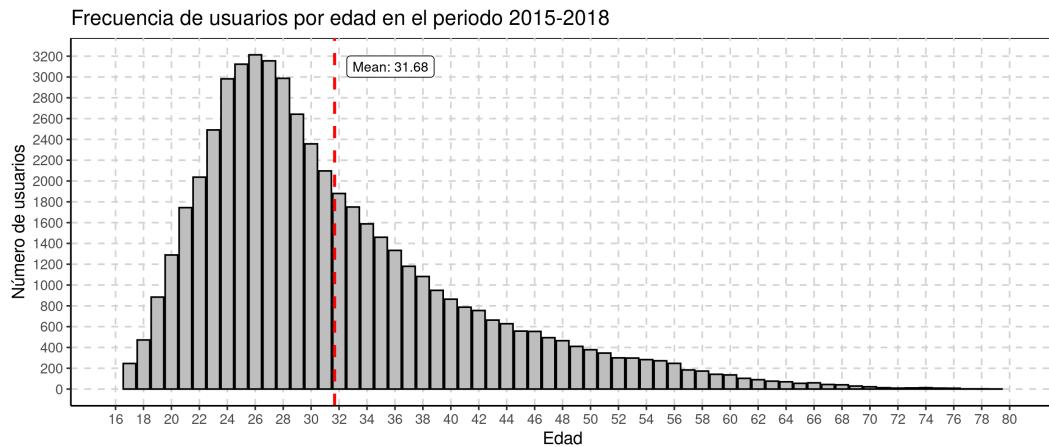


Figura 5: Distribución de usuarios por edad en el periodo de 2015 a 2018.

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.

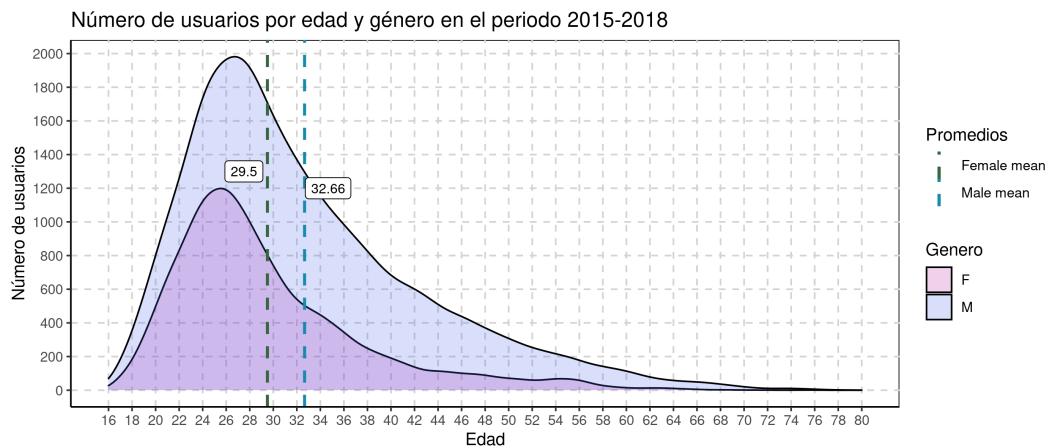


Figura 6: Distribución de edad de los usuarios, separados por género en el periodo de 2015 a 2018.

En las figuras 6 y 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.

4.3. Distancia promedio recorrida por los usuarios.

Para calcular la distancia entre estaciones se uso la ecuación 6.

$$\beta = \sin^2\left(\frac{|\Delta\theta|}{2}\right) + \cos(\theta_1)\cos(\theta_2)\sin^2\left(\frac{|\Delta\phi|}{2}\right)$$

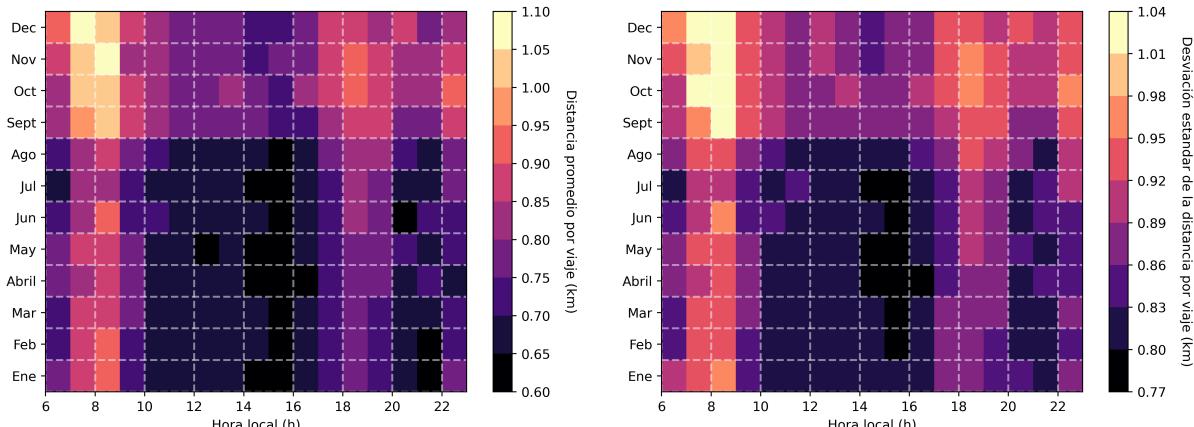
$$D = 2R \arcsin(\sqrt{\beta}) \quad (6)$$

donde R es el radio de la tierra, θ_i es la latitud y ϕ_i es la longitud.

En base a eso, se usaron las ecuaciones 1, 2, 3 y 4. Los resultados obtenidos son los siguientes:

4.3.1. Promedio y desviación estandar mensual por hora

En la figura 7a se visualiza que existe una disminución del uso de las bicicletas entre las 10 y 16 horas a lo largo del año. Además, entre los meses mayo y julio existe una disminución de manera general. Este periodo de meses coincide con la temporada del año donde se presentan con mayor frecuencia lluvias [6]. 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 [6]. 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 los usuarios calculadas con las ecuaciones 1 y 2.

4.3.2. Promedio y desviación estandar diaria semanal por hora

En la figura 8 se aprecia que el uso de la bicicleta se incrementa entre semana a las 7 y 15 horas. Existe una disminución de la distancia recorrida entre las 12 y 16 horas, algo que ya se había previsto en la figura 8a.

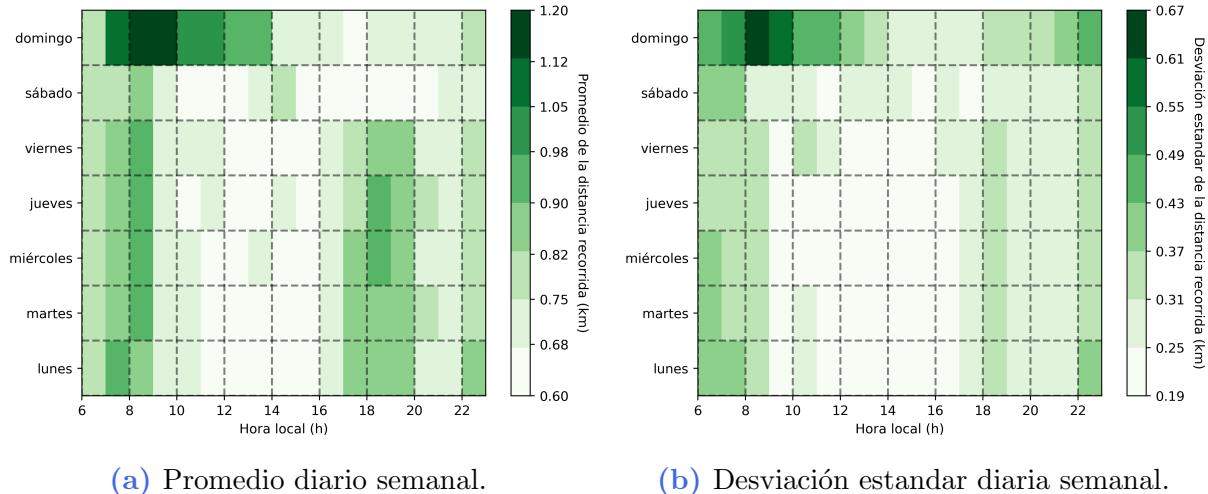


Figura 8: Distancia promedio y desviación estandar diaria semanal por hora recorrida por los usuarios calculadas con las ecuaciones 3 y 4.

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. Esta suposición también puede ser realizada con 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.

4.3.3. Frecuencia de la distancia recorrida por viaje

A partir del promedio por hora definido en la ecuación 5 se obtuvo la distribución mostrada en la figura 9. Al existir estaciones de origen y destino, las distancias pueden ser tratadas como una variable discreta, ya que solo existe un conjunto finito. Se observa que existen dos máximos en las estaciones utilizadas. Las cuales corresponden a las distancias entre las estaciones más usadas como se mostro en la figuras 2 y 3.

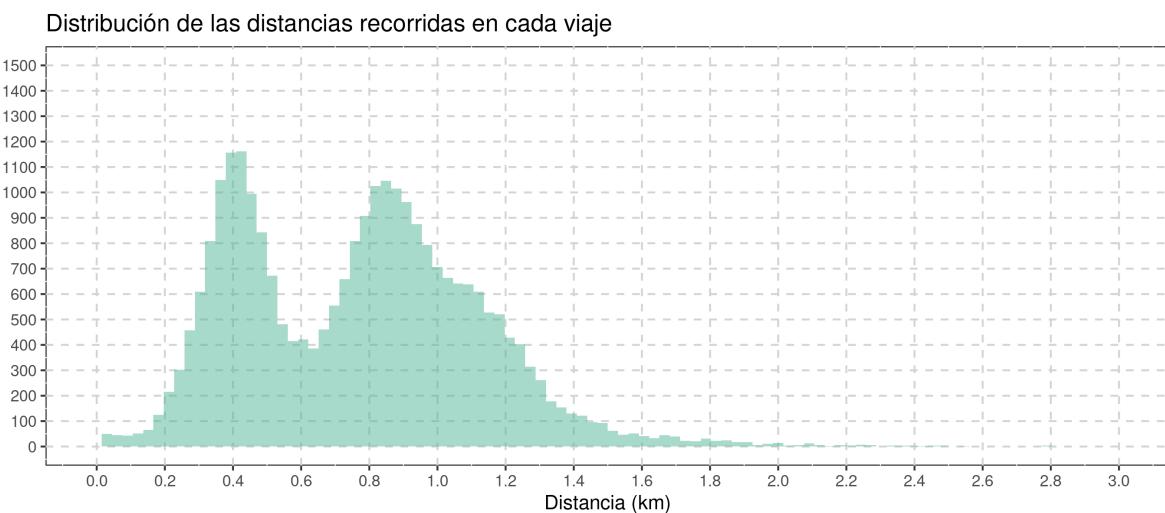


Figura 9: Frecuencia de las distancias recorridas en el periodo 2015-2018.

4.4. Tiempo de uso promedio de los usuarios

En base a los campos Inicio_del_viaje y Fin_del_viaje se calculo el tiempo de uso de una bicicleta en minutos. Con estos datos se usaron las ecuaciones 1, 2, 3 y 4. Los resultados obtenidos son los siguientes:

4.4.1. Promedio y desviación estandar mensual por hora

En la figura 10a 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 [6].

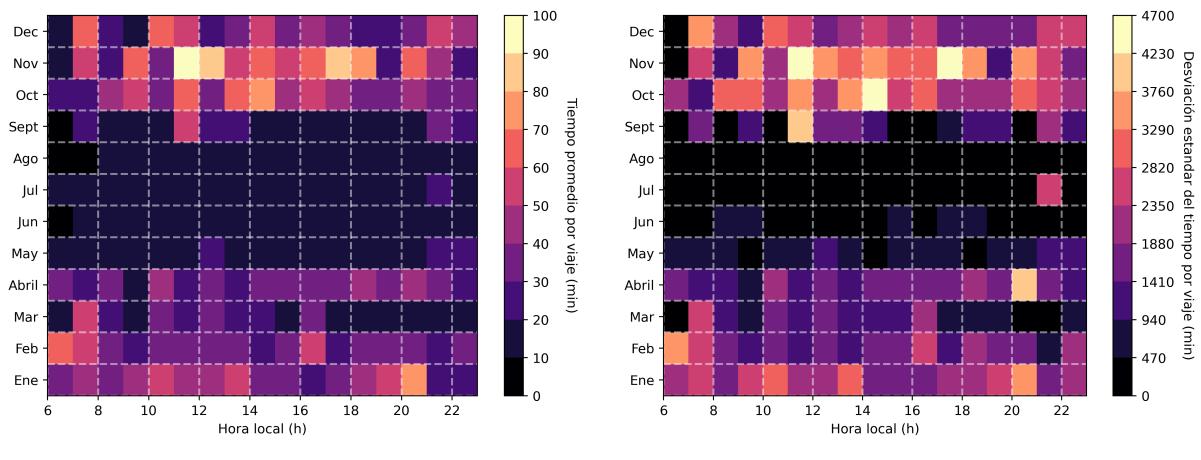


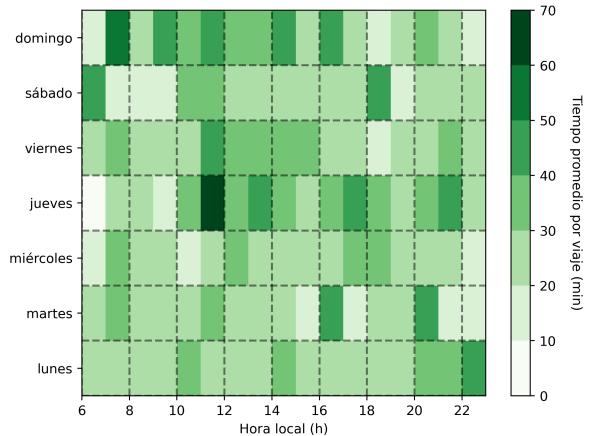
Figura 10: Tiempo de uso promedio y desviación estandar mensual por hora de lo usuarios calculadas con las ecuaciones 1 y 2.

En la figura 10a se presenta un mínimo entre los meses de mayo y agosto. Con ayuda de la figura 10b 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 [6], ya que algunos usuarios pueden dejar de usarlas al sentir una incomodidad y otros prefieran usar más el servicio. Además, los usuarios pueden sentir menos preocupación de las vías públicas y el usar la bicicleta representa más un momento de descanso de su rutina.

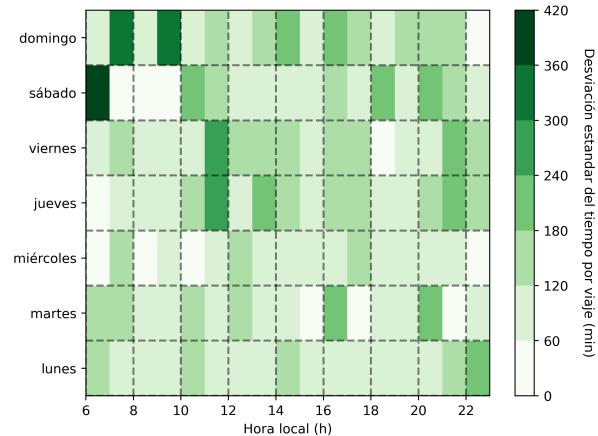
4.4.2. Promedio y desviación estandar diaria semanal por hora

En la figura 11 se aprecia que no hay alguna preferencia en los tiempos de uso en los días entre semana, por lo que se podria estimar un tiempo promedio de uso para toda la semana.

El factor más importante a tomar en cuenta es el mes del que se trate, ya que como se describió en la figura 10, sí existen variaciones a lo largo del año. En la figura 11b 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



(a) Promedio diario semanal.



(b) Desviación estandar diaria semanal.

Figura 11: Tiempo de uso promedio y desviación estandar diaria semanal por hora de los usuarios calculadas con las ecuaciones 3 y 4.

4.4.3. Distribución del tiempo de uso de una bicicleta

A partir de los tiempos calculados para todos los viajes realizados en el periodo 2015-2018 se usó la ecuación 5. La distribución de estos datos se muestran en la figura 12. Aquí se observa como la correspondencia de la figura 9 se ve representada en los tiempos de uso. Ya que al recorrer una misma distancia esta tiene un tiempo semejante cada vez que se realiza.

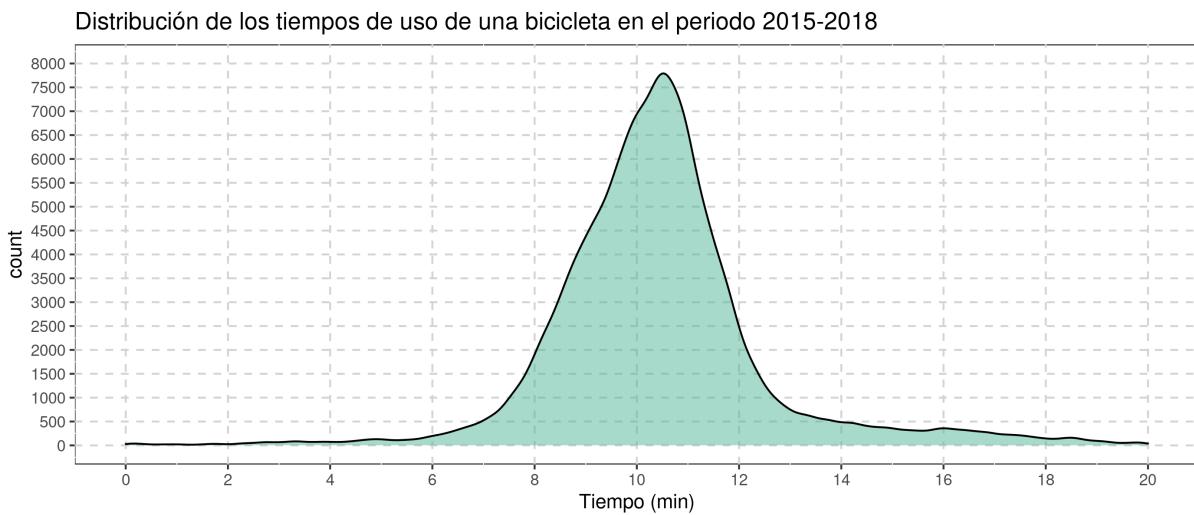


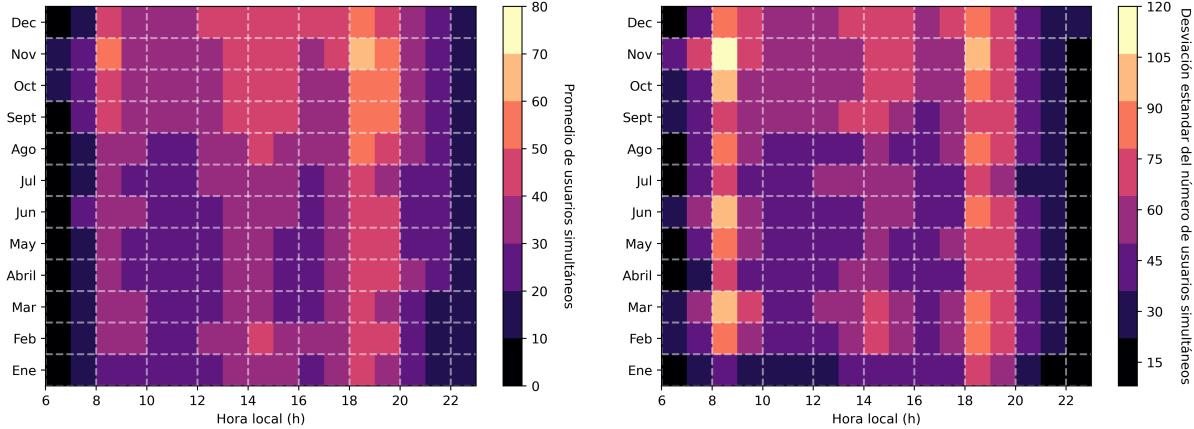
Figura 12: Distribución del tiempo de uso de una bicicleta en el periodo 2015-2018.

4.5. Número promedio de usuarios simultáneos

Con base en este tiempo calculado en minutos, se usaron las ecuaciones 1, 2, 3 y 4. Los resultados obtenidos son los siguientes:

4.5.1. Promedio y desviación estandar mensual por hora

En la figura 13 se visualiza que existe un aumento en el número de usuarios entre los meses septiembre y diciembre. Lo cual coincide con lo antes planteado con las figuras 7 y 10. En la figura 13b 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 cual servicio de transporte público les han mejor tomado en un día.

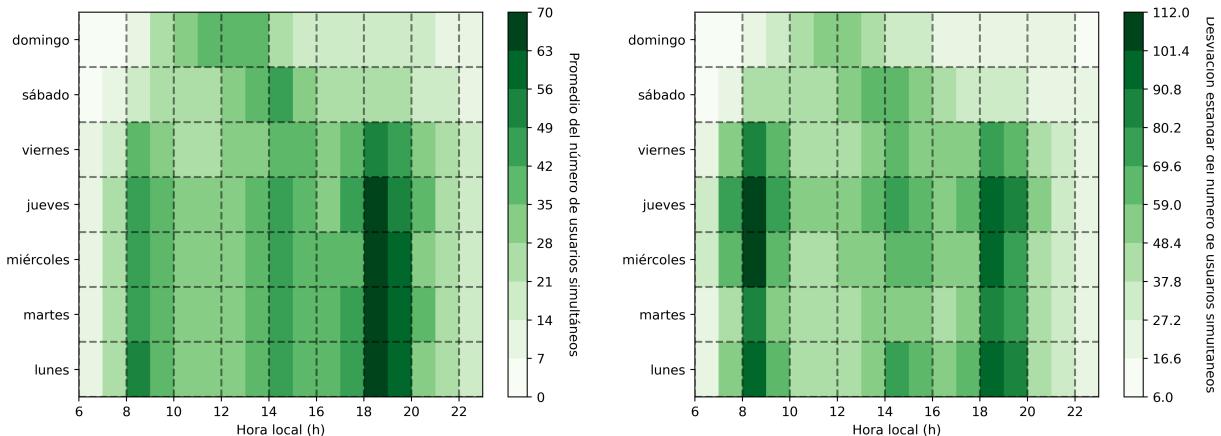


(a) Promedio mensual del número de usuarios. (b) Desviación estandar del número de usuarios.

Figura 13: Tiempo de uso promedio y desviación estandar mensual por hora de los usuarios calculadas con las ecuaciones 1 y 2.

4.5.2. Promedios diarios semanales por hora

En la figura 14a 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.



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

Figura 14: Número de usuarios promedio y desviación estandar diaria semanal por hora calculados con las ecuaciones 3 y 4.

Con la figura 14b se observa que en el periodo donde existe un máximo, este presenta una gran

variación, que concuerda con lo encontrado en la figura 13b. Por ende, se puede suponer que los usuarios que utilizan el servicio en ese periodo de tiempo pueden no ser recurrentes y lo usan cuando les sea necesario, más no es su primera opción.

4.5.3. Distribución del número de usuarios simultáneo

Con el conteo de números de viajes, se realizó el cálculo del promedio usando la ecuación 5. La distribución de los datos obtenidos se representa en la figura 15. En donde se visualiza que existe una mayor frecuencia para pocos usuarios simultáneos y los casos en donde una gran cantidad de bicicletas siendo usadas de manera simultánea es poco común.

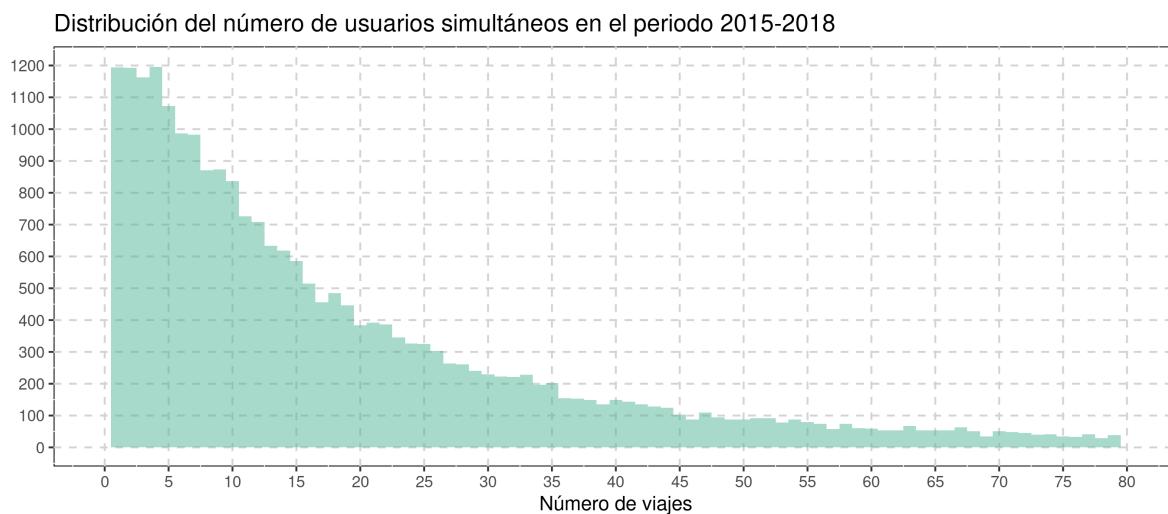


Figura 15: Distribución del número de usuarios simultáneos en el periodo 2015-2018.

5. 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, Maffi 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] Datos abiertos MiBici;. Available from: <https://www.mibici.net/es/datos-abiertos/>.
- [5] 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>.
- [6] 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>.

El presente documento, los archivos de la base de datos utilizada, archivos de apoyo y programas realizados se encuentran en el repositorio https://github.com/giovannilopez9808/Analisis_de_datos_proyecto.