



**COMILLAS**  
UNIVERSIDAD PONTIFICIA

ICAI

MASTER UNIVERSITARIO EN INGENIERÍA DE  
TELECOMUNICACIONES

TRABAJO FIN DE MASTER

ESTUDIO COMPARATIVO DE LA MOVILIDAD EN  
BICICLETAS ELECTRICAS EN CAPITALS  
MUNDIALES

Autor: Simón Díaz Vigón

Director: Carlos Miguel Vallez Fernández

Madrid



Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título  
Estudio comparativo de la movilidad en bicicletas eléctricas en capitales mundiales  
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el  
curso académico 2019/20 es de mi autoría, original e inédito y  
no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido  
tomada de otros documentos está debidamente referenciada.



Fdo.: Simón Díaz Vigón

Fecha: 27/08/2020

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Carlos Miguel Vallez Fernández

Fecha: 27/08/2020



## **AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN RED DE PROYECTOS FIN DE GRADO, FIN DE MÁSTER, TESIS O MEMORIAS DE BACHILLERATO**

### **1º. Declaración de la autoría y acreditación de la misma.**

El autor D. Simón Díaz Vigón DECLARA ser el titular de los derechos de propiedad intelectual de la obra: Estudio comparativo de la movilidad en bicicletas eléctricas en capitales mundiales, que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual.

### **2º. Objeto y fines de la cesión.**

Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad, el autor **CEDE** a la Universidad Pontificia Comillas, de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución y de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra a) del apartado siguiente.

### **3º. Condiciones de la cesión y acceso**

Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia habilita para:

- a) Transformarla con el fin de adaptarla a cualquier tecnología que permita incorporarla a internet y hacerla accesible; incorporar metadatos para realizar el registro de la obra e incorporar "marcas de agua" o cualquier otro sistema de seguridad o de protección.
- b) Reproducir la en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato.
- c) Comunicarla, por defecto, a través de un archivo institucional abierto, accesible de modo libre y gratuito a través de internet.
- d) Cualquier otra forma de acceso (restringido, embargado, cerrado) deberá solicitarse expresamente y obedecer a causas justificadas.
- e) Asignar por defecto a estos trabajos una licencia Creative Commons.
- f) Asignar por defecto a estos trabajos un HANDLE (URL *persistente*).

### **4º. Derechos del autor.**

El autor, en tanto que titular de una obra tiene derecho a:

- a) Que la Universidad identifique claramente su nombre como autor de la misma
- b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.
- c) Solicitar la retirada de la obra del repositorio por causa justificada.
- d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

### **5º. Deberes del autor.**

El autor se compromete a:

- a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
- b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.
- c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.

- d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

**6º. Fines y funcionamiento del Repositorio Institucional.**

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

- La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.
- La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusiva del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.
- La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.
- La Universidad se reserva la facultad de retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a 27 de agosto de 2020

**ACEPTA**



Fdo..... Simón Díaz Vigón.....

Motivos para solicitar el acceso restringido, cerrado o embargado del trabajo en el Repositorio Institucional:



**COMILLAS**

UNIVERSIDAD PONTIFICIA

ICAI

MASTER UNIVERSITARIO EN INGENIERÍA DE  
TELECOMUNICACIONES

TRABAJO FIN DE MASTER

ESTUDIO COMPARATIVO DE LA MOVILIDAD EN  
BICICLETAS ELECTRICAS EN CAPITALS  
MUNDIALES

Autor: Simón Díaz Vigón

Director: Carlos Miguel Vallez Fernández

Madrid



# **ESTUDIO COMPARATIVO DE LA MOVILIDAD EN BICICLETAS ELECTRICAS EN CAPITALS MUNDIALES**

**Autor: Díaz Vigón, Simón.**

Director: Vallez Fernández, Carlos Miguel.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

## **RESUMEN DEL PROYECTO**

Estudio comparativo sobre el uso del servicio de bike sharing en varias capitales mundiales. Abarca variables como la duración de los viajes, la edad y el género de los usuarios, o si el día estudiado es o no laborable. Se obtienen modelos de comportamiento basados en dichas variables.

**Palabras clave:** Python, Power BI, Bike sharing, Comparación

### **1. Introducción**

En la sociedad actual se presenta un problema muy real que es el sobre uso de combustibles fósiles y el abuso de métodos de transporte no sostenibles. Por consiguiente, en este trabajo se realiza un estudio comparativo del uso que se le da a un medio de transporte sostenible como son las bicicletas eléctricas en distintas capitales mundiales, mediante el cual una ciudad en la que no se haya instalado el servicio pueda realizar una previsión de su posible uso o una ciudad en la que ya exista pueda evaluar de manera eficaz los perfiles de uso del servicio.

### **2. Definición del proyecto**

En el presente trabajo se presenta un sistema capaz de analizar y comparar el comportamiento de los usuarios del servicio de bike sharing de diversas capitales mundiales. El objetivo principal es encontrar la capacidad de clasificar dicho comportamiento en función de diversos factores como el clima de la ciudad, la edad de los usuarios, su género, si están suscritos al servicio o no y si el día es laborable o no lo es.

### **3. Descripción del sistema**

Para lograr el objetivo propuesto anteriormente, el desarrollo se basa en dos segmentos principales.

El primer segmento se trata de una función de analítica y limpieza de datos que se desarrolla en lenguaje Python y cuya utilidad es recibir los datos en bruto facilitados por las ciudades bajo estudio y revolver una base de datos en formato de modelo canónico, unificada e igualada para cada uno de los casos, lo cual permite un análisis comparativo simultaneo para el total de las ciudades bajo estudio. Un punto para destacar es que esta función se ha preparado con el objetivo de ser elástica y fácilmente ampliable a nuevas ciudades que se quieran estudiar a posteriori.

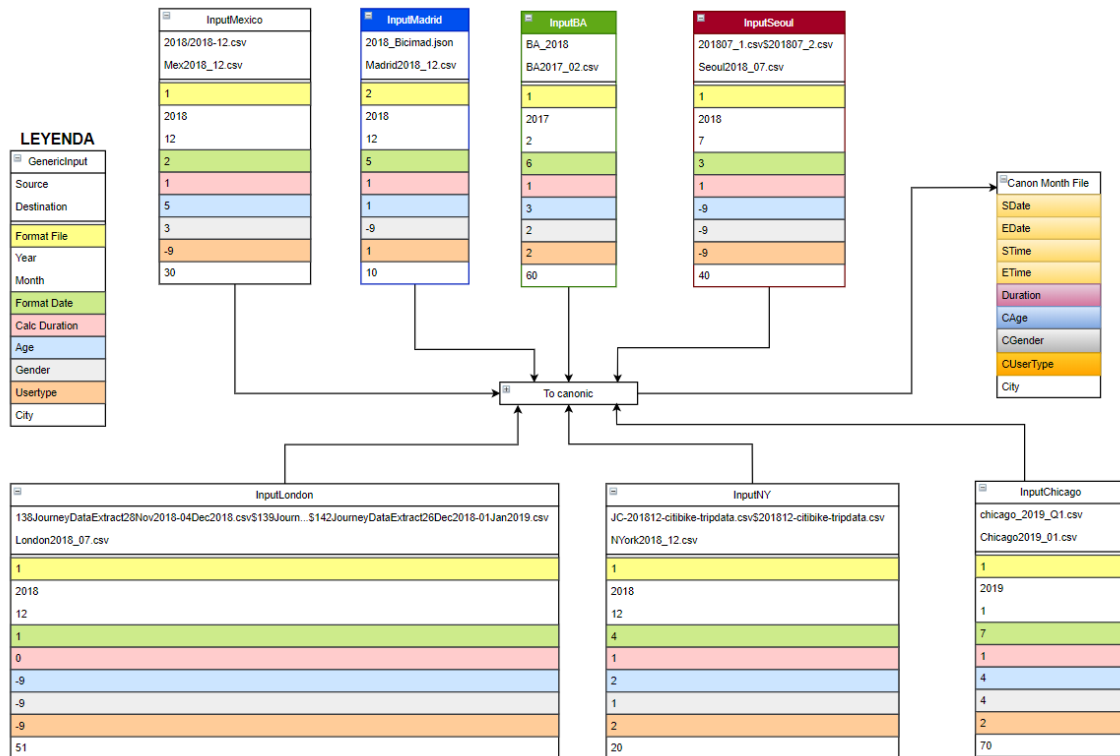


Ilustración 1 - Esquema de la función de tratamiento de datos

El segundo segmento es la parte de análisis y representación de esta base de datos ya en el modelo canónico que se obtiene como salida de la función. Para esta parte se desarrolla una utilidad en Power BI capaz de, de manera automatizada, beber de la base de datos generada y presentar un análisis comparativo en forma de gráficos previamente diseñado. Cabe destacar que, por limitaciones técnicas, el análisis se ha realizado año a año, y por tanto las comparativas y conclusiones extraídas del global se han hecho de manera externa a Power BI.

#### 4. Resultados

Los resultados obtenidos de cara al análisis son los gráficos de salida de la utilidad en Power BI, se representan factores como la duración media de los viajes o el conteo total de viajes en función de la ciudad, la fecha, la edad de los usuarios, el género de los usuarios o si estos están suscritos al servicio o no.

A partir de dichas representaciones, se realiza una comparación entre los resultados de cada ciudad y de cada periodo bajo estudio, llegándose a distintas conclusiones que se comentaran más adelante.

Un ejemplo de las figuras utilizadas para este análisis se presenta a continuación.

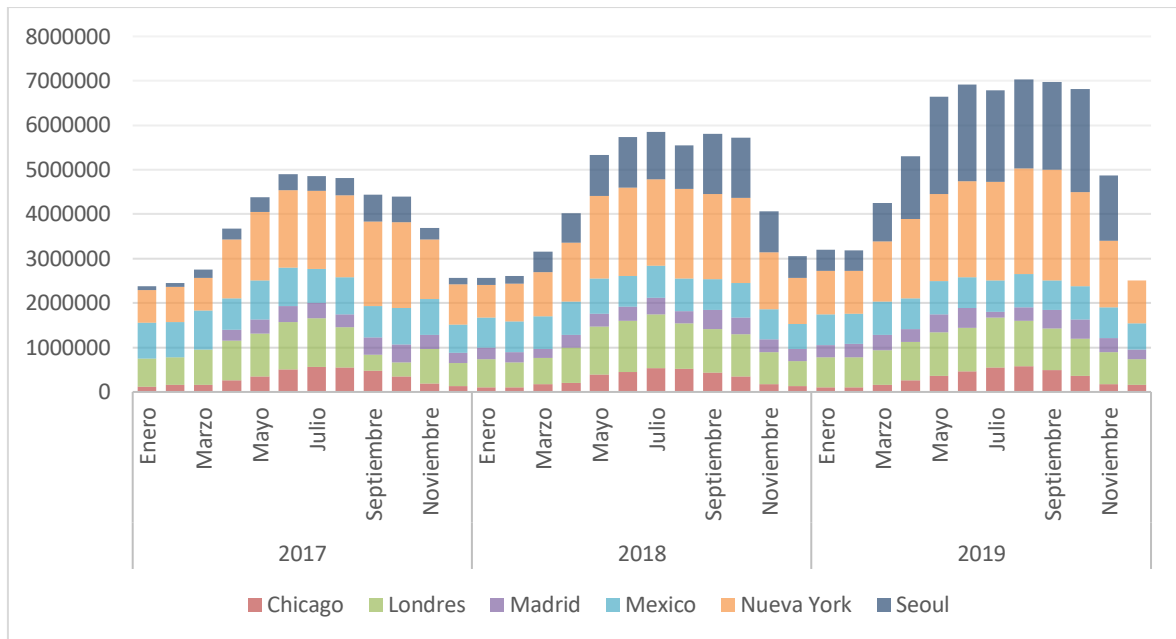


Ilustración 2 – Ejemplo de resultados de conteo de viajes anual

## 5. Conclusiones

El tratamiento de datos de fuentes tan dispares resulta altamente complejo hasta que se logra una generalización.

Las condiciones medioambientales extremas como nieve o monzones afectan en mucha mayor medida al uso del servicio que la variación de temperatura.

El rango de edad de mayor uso del servicio está entre los 27 y los 40 años.

Los usuarios masculinos hacen un uso mucho mayor del servicio, pero sus viajes tienen una duración más reducida que los de los usuarios femeninos.

Los usuarios que están registrados en el servicio hacen un uso predominantemente en días laborables, mientras que el uso en días no laborables es principalmente por cliente esporádico.

## 6. Referencias

- [1] ODS: <https://verdesdigitales.com/2019/09/03/ods-en-comunicacion/>
- [2] Metodología scrum: <https://www.scrum.org/resources/blog/que-es-scrum>
- [3] Bike share map: <https://bikesharemap.com/#/4/10.9035/48.7522/>

# COMPARATIVE STUDY ON THE MOBILITY VIA BIKE SHARING IN A VARIETY OF WORLD CAPITALS

**Author: Díaz Vigón, Simón.**

Supervisor: Vallez Fernández, Carlos Miguel.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

## ABSTRACT

Comparative study on the usage of bike sharing services in a variety of world capitals. Encases variables such as the trip duration, age and gender of the users or if the day is a work day or not. Behaviour models are obtained based on those variables.

**Keywords:** Python, Power BI, Bike sharing, Comparison

### 1. Introduction

Current society presents a very real problem which is the over usage of fossil fuels and the abuse of non-renewable energy sources. Thus, this paper focuses on the creation of a comparative study on the use of a sustainable means of transportation, which is the bike sharing, in a variety of world capitals. With this study a city which has not yet implemented this service, can make a simulation of its possible usage or a city in which the service is currently available can evaluate in an efficient way the profiles of usage of the service.

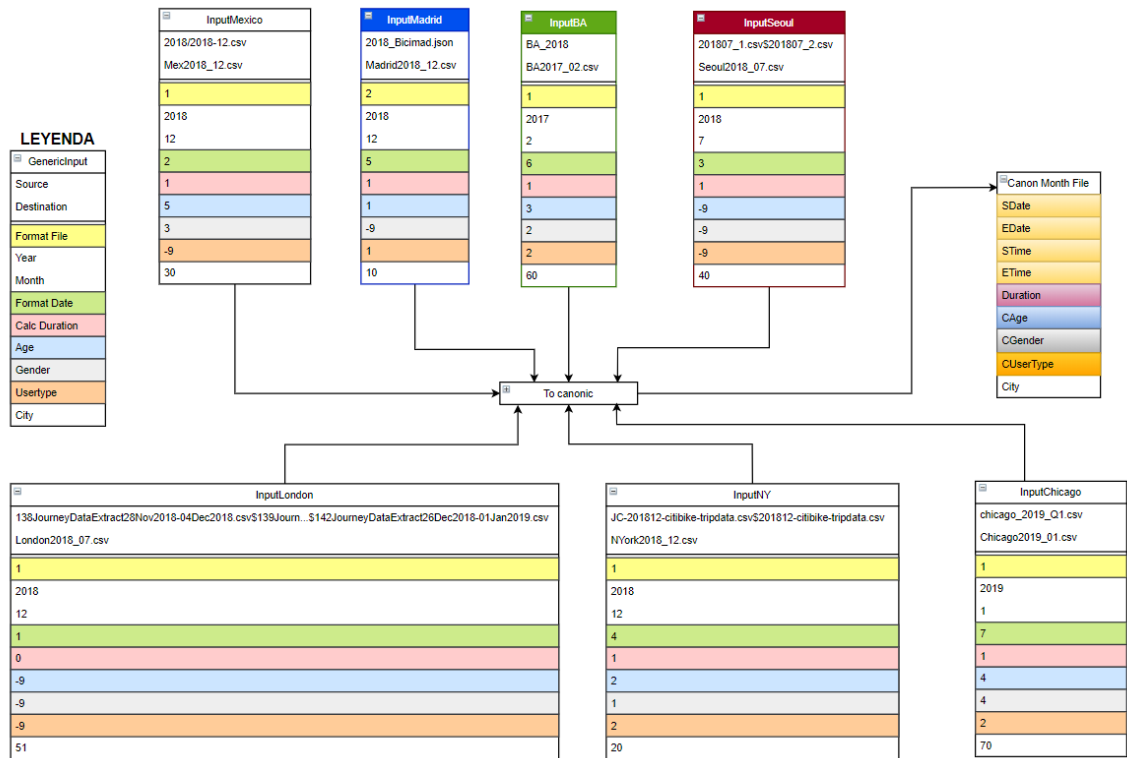
### 2. Project definition

The following paper presents a system capable of analysing and comparing the behaviour of the users of the bike sharing system in a variety of world capitals. The main objective is achieving the ability to classify said behaviour depending on several factors, such as the weather on the city, the ages of the users, their gender, if there are currently subscribed to the service or not and if the day is a work day or not.

### 3. Description of the system

In order to achieve the objective previously mentioned, the development of the system is separated in two parts.

The first part is a data analytics and cleaning function developed in the Python language which receives the raw data given by the cities under study and returns a database in a cannon format. This database is unified and equalized for each one of the cases, which allows for a simultaneous comparative analysis for all the cities under study. A point to clarify is that this function has been developed with the objective of being flexible in mind, which means that at any given time, new cities with different data formats may be added with ease.



*Illustration 3 - Scheme of the data treatment function*

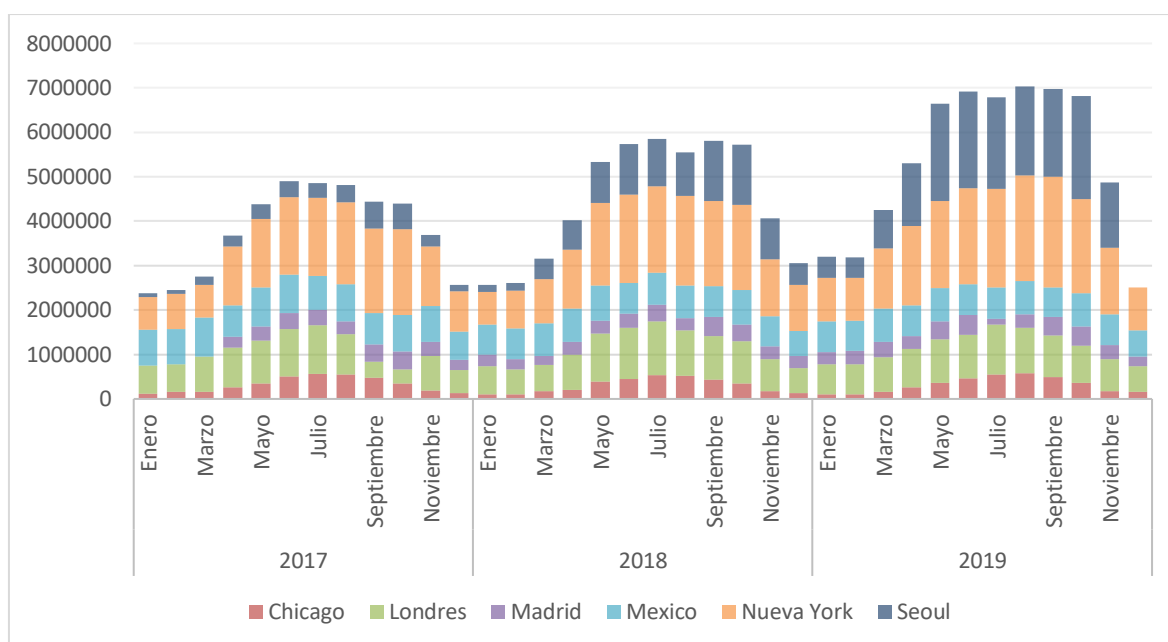
The second part contains the analysis and representation of the database in its cannon form previously generated by the function. For this part, a utility in Power BI is developed, which is able to, in an automated way, use the data from the database to present a comparative analysis in the form of several graphs. It is worth noting that due to technical limitations this part has been limited to yearly analysis, and such the comparisons between the results obtained for each year have been done outside of Power BI.

#### 4. Results

The obtained results from the analysis are composed by the graphs generated in Power BI. Attributes such as the average duration of the trips, or the total count of trips in regard to the city, date, age of the user, gender of the user or if the user is subscribed to the service or not are represented.

Taking into account said graphs, a comparison is made for each city and each period of time under study, reaching the conclusions explained in the next section.

The following graph shows an example of the utility from which the analysis is done.



*Illustration 4 – Example on the annual count of trips*

## 5. Conclusions

Working with data from different sources in different formats proves to be quite complicated until a generalization is reached.

Extreme environmental conditions such as typhoons and snow affect greatly the usage of the service, whereas temperature variation does not.

The ages of the users which use the service the most is between 27 to 40 years.

Male users utilise the service more than female users, however the trips of the male users are shorter.

Registered users make use of the service mainly on work days, whereas the usage on non-work days is mainly made by sporadic users.

## 6. References

- [1] ODS: <https://verdesdigitales.com/2019/09/03/ods-en-comunicacion/>
- [2] Metodología scrum: <https://www.scrum.org/resources/blog/que-es-scrum>
- [3] Bike share map: <https://bikesharemap.com/#/4/10.9035/48.7522/>

## *Índice de la memoria*

<i>Capítulo 1. Introducción</i> .....	<b>7</b>
<i>Capítulo 2. Descripción de las Tecnologías</i> .....	<b>9</b>
<i>Capítulo 3. Estado de la Cuestión</i> .....	<b>10</b>
<i>Capítulo 4. Definición del Trabajo</i> .....	<b>11</b>
4.1 Justificación .....	11
4.2 Objetivos.....	11
4.3 Metodología .....	12
4.4 Planificación.....	13
4.5 Estimación económica .....	14
<i>Capítulo 5. Modelo Desarrollado</i> .....	<b>15</b>
5.1 Desarrollo de la función.....	15
5.1.1 Entrada de la función.....	16
5.1.2 Formato del archivo.....	17
5.1.3 Fecha y hora.....	18
5.1.4 Cálculo de la duración.....	19
5.1.5 Género.....	20
5.1.6 Edad.....	21
5.1.7 Tipo de usuario.....	21
5.1.8 Salida de la función .....	22
5.1.9 Agrupación de las salidas .....	23
5.2 Configuración y herramientas Power BI.....	23
5.3 Representación de los datos .....	25
5.3.1 Primera iteración.....	26
5.3.2 Segunda iteración .....	29
5.3.3 Iteración final .....	30
<i>Capítulo 6. Análisis de Resultados</i> .....	<b>36</b>
6.1 Estudio de las variaciones de duración a nivel anual.....	36

6.2	Estudio de la duración por edades .....	40
6.3	Estudio de la duración por géneros.....	45
6.4	Estudio de la duración por tipo de usuario.....	51
6.5	Estudio de la duración según el día de la semana.....	59
6.5.1	<i>Duraciones medias 2017</i> .....	64
6.5.2	<i>Conteo de viajes 2017</i> .....	66
6.5.3	<i>Duraciones medias 2018</i> .....	68
6.5.4	<i>Conteo de viajes 2018</i> .....	70
6.5.5	<i>Duraciones medias 2019</i> .....	72
6.5.6	<i>Conteo de viajes 2019</i> .....	74
6.5.7	<i>Resultados</i> .....	76
<b>Capítulo 7. Conclusiones y Trabajos Futuros .....</b>		<b>78</b>
<b>Capítulo 8. Bibliografía.....</b>		<b>82</b>
<b>Anexo A – Código .....</b>		<b>84</b>
<b>Anexo B – Cronograma .....</b>		<b>85</b>
<b>Anexo C – Tabla inputs.....</b>		<b>86</b>
<b>Anexo D – Meteorología .....</b>		<b>87</b>
<b>Anexo E – ODS.....</b>		<b>94</b>

## *Índice de figuras*

Figure 5-A Metodología Scrum [3] .....	13
Figure 6-A Input .....	15
Figure 6-B Inputs de la función .....	17
Figure 6-C Función, formato del archivo.....	18
Figure 6-D Función, formato de la fecha .....	19
Figure 6-E Función, cálculo de la duración .....	20
Figure 6-F Función, género .....	20
Figure 6-G Función, edad.....	21
Figure 6-H Función, tipo de usuario .....	22
Figure 6-I Función, datos de salida.....	22
Figure 6-J Queries Power BI .....	23
Figure 6-K Script duración Power BI .....	24
Figure 6-L Relaciones Power BI .....	25
Figure 6-M Comprobación Enero 2018 .....	26
Figure 6-N Comprobación Febrero 2018.....	27
Figure 6-O Comprobación Marzo 2018.....	28
Figure 6-P Calendario 1 .....	29
Figure 6-Q Comprobación duración media por mes 2018.....	30
Figure 6-R Calendario 2.....	31
Figure 6-S Comprobación media duración por género y ciudad.....	32
Figure 6-T Comprobación media duración por edad y ciudad.....	33
Figure 6-U Comprobación media duración por tipo de usuario y ciudad.....	33
Figure 6-V Comprobación media duración en función del día del año .....	34
Figure 7-A Duración a nivel anual 2017.....	36
Figure 7-B Duración a nivel anual 2018.....	37
Figure 7-C Duración a nivel anual 2019.....	38

Figure 7-D Duración media edad 2017.....	41
Figure 7-E Porcentaje conteo edad 2017 .....	41
Figure 7-F Duración media edad 2018 .....	42
Figure 7-G Porcentaje conteo edad 2018.....	42
Figure 7-H Duración media edad 2019.....	43
Figure 7-I Porcentaje conteo edad 2019.....	43
Figure 7-J Duración media por genero 2017.....	45
Figure 7-K Conteo genero por ciudad 2017.....	46
Figure 7-L Conteo genero por mes 2017 .....	46
Figure 7-M Duración media por genero 2018.....	47
Figure 7-N Conteo genero por ciudad 2018.....	47
Figure 7-O Conteo genero por mes 2018.....	48
Figure 7-P Duración media por genero 2019 .....	48
Figure 7-Q Conteo genero por ciudad 2019.....	49
Figure 7-R Conteo genero por mes 2019 .....	49
Figure 7-S Duración media por tipo de usuario 2017.....	52
Figure 7-T Duración media por tipo de usuario y ciudad 2017 .....	53
Figure 7-U Conteo registrados en laborable vs no laborable 2017 .....	53
Figure 7-V Duración media por tipo de usuario 2018 .....	54
Figure 7-W Duración media por tipo de usuario y ciudad 2018 .....	55
Figure 7-X Conteo registrados en laborable vs no laborable 2018 .....	55
Figure 7-Y Duración media por tipo de usuario 2019 .....	56
Figure 7-Z Duración media por tipo de usuario y ciudad 2019 .....	57
Figure 7-AA Conteo registrados en laborable vs no laborable 2019.....	57
Figure 7-BB Duración media Buenos Aires 2018.....	60
Figure 7-CC Duración media Chicago 2018.....	60
Figure 7-DD Duración media Londres 2018.....	61
Figure 7-EE Duración media Madrid 2018.....	61
Figure 7-FF Duración media México 2018.....	62
Figure 7-GG Duración media Nueva York 2018 .....	62

Figure 7-HH Duración media Seúl 2018 .....	63
Figure 7-II Media Enero-Marzo 2017 Dia .....	64
Figure 7-JJ Media Abril-Junio 2017 Dia .....	64
Figure 7-KK Media Julio-Septiembre 2017 Dia .....	65
Figure 7-LL Media Octubre-Diciembre 2017 Dia.....	65
Figure 7-MM Conteo Enero-Marzo 2017 Dia.....	66
Figure 7-NN Conteo Abril-Junio 2017 Dia.....	66
Figure 7-OO Conteo Julio-Septiembre 2017 Dia .....	67
Figure 7-PP Conteo Octubre-Diciembre 2017 Dia.....	67
Figure 7-QQ Media Enero-Marzo 2018 Dia .....	68
Figure 7-RR Media Abril-Junio 2018 Dia .....	68
Figure 7-SS Media Julio-Septiembre 2018 Dia.....	69
Figure 7-TT Media Octubre-Diciembre 2018 Dia.....	69
Figure 7-UU Conteo Enero-Marzo 2018 Dia.....	70
Figure 7-VV Conteo Abril-Junio 2018 Dia.....	70
Figure 7-WW Conteo Julio-Septiembre 2018 Dia .....	71
Figure 7-XX Conteo Octubre-Diciembre 2018 Dia .....	71
Figure 7-YY Media Enero-Marzo 2019 Dia .....	72
Figure 7-ZZ Media Abril-Junio 2019 Dia.....	72
Figure 7-AAA Media Julio-Septiembre 2019 Dia.....	73
Figure 7-BBB Media Octubre-Diciembre 2019 Dia.....	73
Figure 7-CCC Conteo Enero-Marzo 2019 Dia.....	74
Figure 7-DDD Conteo Abril-Junio 2019 Dia.....	74
Figure 7-EEE Conteo Julio-Septiembre 2019 Dia.....	75
Figure 7-FFF Conteo Octubre-Diciembre 2019 Dia.....	75
Figure 8-A Duración media 2017, 2018 y 2019 .....	79
Figure 8-B Conteo de duración 2017, 2018 y 2019.....	80
Figure 0-A Objetivos de Desarrollo Sostenible [2] .....	94

## *Índice de tablas*

Table 1 Disponibilidad de datos .....	9
Table 2 Presupuesto .....	14
Table 3 Rangos de edad .....	40

## Capítulo 1. INTRODUCCIÓN

El servicio de bikesharing surge en Ámsterdam en 1965 con la intención de prestar bicicletas gratuitamente para el uso público en las que se pretendía que se hiciese un viaje y después se dejasen para el siguiente usuario. Lamentablemente esta iniciativa pereció al mes de su instauración, todas las bicicletas fueron o bien robadas o tiradas a los canales.

No es hasta la mitad de los años noventa que se instauran servicios funcionales en Copenhague y Londres para Europa, donde se establecía un servicio de pago por uso, y Oregón en los Estados Unidos, donde el servicio era gratuito, pero de bicicletas de segunda mano. Actualmente, el grueso del uso de este servicio se encuentra en las ciudades asiáticas, de los quince mayores programas de bikesharing, trece están en ciudades chinas.

En definitiva, el servicio se basa en ofrecer el uso al público, que pueden ser tanto residentes locales que utilizan el servicio de manera habitual como visitantes o turistas que harán un uso esporádico, de bicicletas mediante un sistema de alquileres a un precio asequible.

De esta reflexión sale la idea de realizar un estudio sobre la utilización de servicios de bikesharing en este tipo de ciudades. Si se consigue profundizar lo suficiente en esta iniciativa y demostrar con números tangibles que efectivamente se llega a una mejora de la calidad de vida de la población, se podría conseguir que el servicio se generalizase aún más.

Con estas premisas en mente, el siguiente proyecto tiene como objetivo el estudio y caracterización de los viajes realizados por los usuarios de servicios de bikesharing en varias capitales mundiales para su posterior análisis y comparación.

La idea fundamental es conseguir entender el uso del servicio en función de diversos factores como son la edad de los usuarios, su género, la duración de los viajes o la época del año en la que se producen.

Para lograr este objetivo se desarrollará una herramienta de programación capaz de recibir datos en bruto como entrada y devolver los mismos en un formato de modelo canónico para facilitar el trabajo sobre ellos y su estudio.

Con esos datos ya refinados se realiza un análisis exhaustivo de los trayectos mediante el uso de numerosas representaciones que permiten crear una relación entre el uso del servicio en las distintas ciudades y la extracción de diversas conclusiones que podrían servir en un futuro para mejoras del servicio o nuevas instauraciones en localidades que en este momento carecen del mismo.

## Capítulo 2. DESCRIPCIÓN DE LAS TECNOLOGÍAS

Para hacer los análisis previamente descritos y de manera sucinta, las herramientas principales empleadas son el software de programación basado en Python 3, Jupyter notebooks y el software de representación de datos de Microsoft Power BI.

Las versiones correspondientes a las herramientas anteriormente citadas utilizadas para el desarrollo del trabajo son 3.7.6 de Python, conda 4.8.3 para el manejo de packages y 2.83.5894.961 para Power BI.

Los datos iniciales se analizan en herramientas como Excel o Notepad++ para comprobaciones básicas como la correcta presentación de cabeceras o la eliminación de datos corruptos.

El factor limitante en este caso es que la realización del proyecto es desde un ordenador personal, lo cual ya descarta la posibilidad de llevar el proyecto a un nivel de alto procesamiento de datos. Por tanto, los análisis se orientan de una manera más horizontal, en la que se busca más el análisis exhaustivo en franjas cortas de tiempo que una observación global a lo largo de lo que podrían ser años. Además, los datos disponibles de las ciudades bajo estudio también limitan en cierto modo el trabajo.

Ciudad	Fecha de inicio	Fecha fin	Anotaciones
Madrid	Abril 2017	Mayo 2020	Formato .json
Nueva York	Junio 2013	Mayo 2020	Split en Jersey City y resto de ciudad
Londres	Enero 2015	Mayo 2020	Split semanal
Mexico	Enero 2010	Mayo 2020	Formato .csv archivo unico annual
Seoul	Enero 2015	Noviembre 2019	Empiezan con datos por Q, luego mensuales
Buenos Aires	Enero 2010	Julio 2019	Sin datos de 09/2017 a 01/2018
Chicago	Enero 2015	Marzo 2020	Formato .csv con split por Q

*Table 1 Disponibilidad de datos*

## Capítulo 3. ESTADO DE LA CUESTIÓN

Si bien es cierto que hasta la fecha se han realizado múltiples estudios relacionados con la cuestión del bikesharing, la mayoría tienden a estar basados en una única localidad, o en localidades centradas en un único país. Este estudio busca ir un paso más allá y llegar a conclusiones de una mayor relevancia y amplitud al comparar ciudades tan diversas como Nueva York, Madrid y Seúl.

De nuevo observando los estudios realizados anteriormente sobre el tema, también cabe destacar que se hace más hincapié en el control en tiempo real de los usuarios y las propias bicicletas, sin profundizar tanto en los patrones de comportamiento a largo plazo. En este estudio se tratará de caracterizar a las distintas poblaciones con el fin de encontrar ligaduras y similitudes en función de factores como las estaciones o la edad de los usuarios en poblaciones muy diferentes.

Por otro lado, este trabajo no se centra únicamente en lo que es el estudio de los resultados y del comportamiento de los usuarios, gran parte del valor reside en la función para el tratamiento de datos desarrollada. Se trata de una herramienta con la que se ha logrado una capacidad de adaptación y generalización que permite añadir al sistema o de forma directa o con una necesidad mínima de modificación cualquier nueva ciudad no contemplada en un principio.

## Capítulo 4. DEFINICIÓN DEL TRABAJO

### 4.1 JUSTIFICACIÓN

A la vista de los argumentos expuestos anteriormente, se resuelve que el mercado de las tecnologías de bike sharing es uno en auge y que no solo tiene un atractivo económico sino también social, como se explicaba en el capítulo 2, se trata de un servicio en línea con los objetivos de desarrollo sostenible.

El desarrollo de este proyecto facilita el estudio de los condicionales que marcan el uso del servicio, por tanto, resulta una facilidad altamente importante a la hora de decidir si resulta una buena opción instalarlo en una nueva ciudad, o si es necesaria una ampliación o revaluación en una ciudad existente.

Por otro lado, se trata de un proyecto en constante crecimiento, la primera versión es de carácter limitado, pero con sencillas ampliaciones se podría llegar a unos niveles de análisis más altos.

### 4.2 OBJETIVOS

El objetivo principal de este trabajo es estudiar el comportamiento de los usuarios del servicio de bikesharing de las ciudades seleccionadas en función de varios parámetros. Por tanto, de este objetivo principal derivan varios puntos que serán los que se estudiarán para obtener la caracterización de la población deseada, entre ellos se encuentran:

- Estudio del comportamiento en función del clima de cada una de las ciudades.
- Estudiar cómo afecta la edad a la duración de los viajes en cada una de las ciudades que tenemos en consideración.
- Diferencias en el comportamiento entre hombres y mujeres, ya sea por horas del día a las que utilizan el servicio o la duración de sus viajes.

- Diferencias en el comportamiento entre laborables o fines de semana, principalmente ligado al volumen de viajes y su duración.
- Comparaciones entre viajes según el tipo de usuario, existen usuarios suscritos al servicio y usuarios que no lo están.
- Tipo de usuario con hora y día de la semana, para observar si los usuarios que por ejemplo utilizan el servicio para ir a trabajar o a clase suelen suscribirse al servicio o no.

Además, de los apartados anteriores también saldrán combinaciones con otros factores como puede ser la edad, es decir, realizar fragmentaciones de la población. Para el caso por ejemplo de la distinción por género, hemos visto que no se llega a un resultado representativo, por lo que para enriquecer ese dato se ha juntado con la edad para estudiar duraciones de viajes en función de las franjas de edad y el género.

### **4.3 METODOLOGÍA**

La metodología llevada a cabo para el desarrollo del proyecto ha sido la metodología Scrum con la variación de reuniones semanales en lugar de diarias. Esta metodología posiciona al tutor del proyecto como el Scrum Master y a su vez Product Owner, y al realizador o realizadores del mismo como Team. La idea detrás de este modelo se basa en el fijado de objetivos a corto y largo plazo de manera que el logro de los objetivos a corto plazo se regula mediante las reuniones periódicas preestablecidas en las que además se realizan puntos de corrección y reorientación del trabajo.

En la figura 4-A se presenta un esquema del funcionamiento de esta metodología.

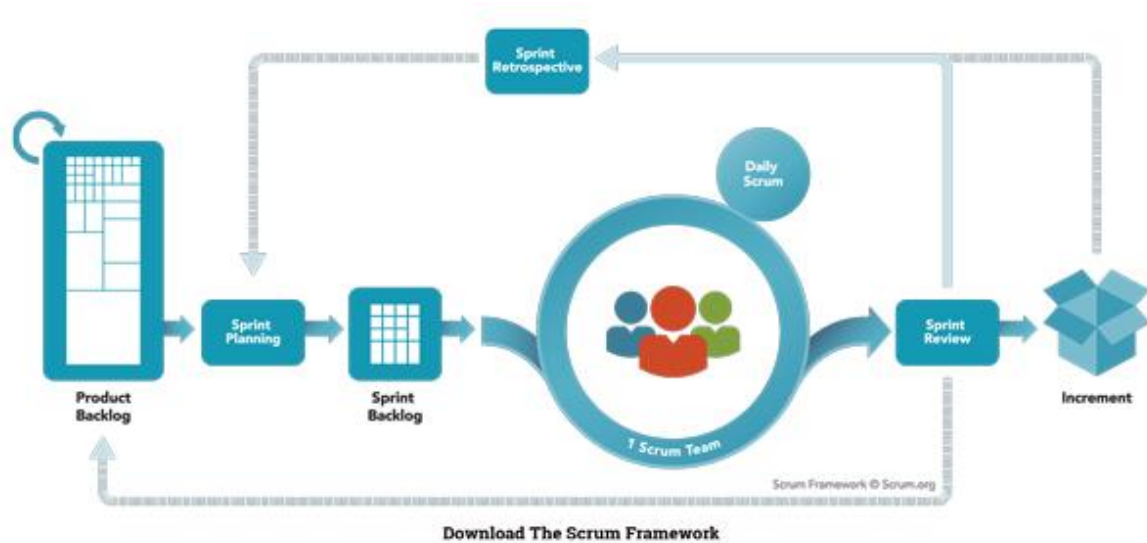


Figure 4-A Metodología Scrum [3]

#### 4.4 PLANIFICACIÓN

Se adjunta en el Anexo 2 – Cronograma, una representación del desarrollo temporal del proyecto en el que se muestran las distintas partes del mismo.

Se comienza con un apartado de prerequisites en la que se describe el tiempo empleado para la preparación y formación previa antes de abordar la cuestión. A continuación, el segmento de la gestión de los datos, donde se adquieren y estudian las distintas ciudades y los datos que ofrecen para el posterior trabajo sobre ellos. La parte siguiente comprende la composición de la función, lo cual está descrito en profundidad en el apartado 5.1 – Desarrollo de la función. Del mismo modo, la parte siguiente de Power BI, se describe en el apartado 5.2 – Representación de los datos. Finalmente se expone un último apartado de conclusiones y finalización.

En concordancia con la metodología SCRUM seguida para el desarrollo del proyecto, se representan con rombos azules las reuniones de seguimiento semanales realizadas con el tutor en las que se exponen los problemas encontrados y el avance conseguido hasta la fecha de la reunión.

Por último, existen otros dos hitos a considerar. Primero, la finalización de la función, momento en el cual ya se puede proceder al análisis de los datos, ya que previo a este punto al no haberse desarrollado una base de datos lista para su estudio, no es posible pasar de fase, y segundo, la existencia de una cantidad de contenido suficiente para el análisis, ya que del mismo modo carece de sentido comenzar un análisis de los datos teniendo una cantidad reducida de los mismos.

#### **4.5 ESTIMACIÓN ECONÓMICA**

Análogamente al apartado anterior, se adjunta el presupuesto considerado para la realización de proyecto.

Item	Cantidad	Precio Ud	Precio Total	Total (+21% IVA)
Hora experto Python	9sem a 10h/sem	[25 - 40] €	[2250 - 3600] €	[2722.5 - 4356] €
Hora experto PowerBI	7sem a 10h/sem	[25 - 40] €	[2250 - 3600] €	[2722.5 - 4356] €
Hora experto Analítica de datos	8sem a 10h/sem	[20 - 30] €	[1600 - 2400] €	[1936 - 2904] €
Licencia PowerBI	1	0.00 €	0.00 €	0.00 €
Licencia Python	1	0.00 €	0.00 €	0.00 €
Amortizacion PC	1	50.00 €	50.00 €	60.50 €
Gastos generales	150	3.00 €	450.00 €	544.50 €
			[6600 - 10100] €	[7986 - 12221] €

*Table 2 Presupuesto*

Los precios calculados para el presupuesto se han decidido dar en intervalos, dada la naturaleza fluctuante de los costes en este tipo de servicios. De este modo, se consideran un total de 240 horas de trabajo totales divididas entre la parte de programación, la parte de representación y la parte de análisis. Además, se añaden como ítems las licencias de los softwares utilizados. En este caso, y dado que estas licencias son o bien gratuitas o facilitadas por la universidad, el coste es cero. Por último, se añaden como costes extra la amortización del PC utilizado para el desarrollo del proyecto y un ítem de gastos generales que comprende mayormente los gastos de electricidad derivados del trabajo.

## Capítulo 5. MODELO DESARROLLADO

De cara a explicar el desarrollo del modelo que conforma la parte técnica del trabajo, se ha decidido realizar una división en dos partes del mismo. Por un lado, se presenta un sistema esquemático sobre la función utilizada para el tratamiento de los datos. Para facilitar su comprensión y evitar la necesidad de entrar de lleno en el código puro, se ha resuelto la presentación de diagramas de flujo. Por otro lado, se expone la metodología seguida para la representación de los datos, sin entrar aun en su análisis.

### 5.1 DESARROLLO DE LA FUNCIÓN

En un primer punto, se hará foco en la entrada de la función. Para simplificar el trabajo con la misma, la forma que se ha elegido para la construcción es la de un solo input, en un formato csv, el cual contiene toda la información que la función requiere para operar. Un ejemplo de este archivo formato csv se presenta a continuación en la figura 5-A.

```
1 variables, valores
2 source, ./TFM/bdata/2019/chicago_2019_Q1.csv
3 destination, ./TFM/mdata/Chicago2019_01.csv
4 format_file, 1
5 year, 2019
6 month, 1
7 format_date, 7
8 calc_duration, 1
9 age, 4
10 gender, 4
11 usertype, 2
12 ciudad, 70
```

Figure 5-A Input

Para un mejor entendimiento del funcionamiento de los inputs a recibir por la función, se adjunta en el Anexo C – Tabla inputs una tabla que describe en profundidad los parámetros que componen el fichero y como deben ser configurados en función de la ciudad que se vaya a tratar. Esta tabla también podrá ser utilizada como guía para una futura ampliación de la

función en la que se quieran estudiar nuevas ciudades, sirviendo como plantilla para resolver si los métodos utilizados para casos existentes se pueden extrapolar al nuevo dataset o si es necesario programar una nueva instrucción.

Para los casos en los que el dataset bajo estudio carezca de valores de age, gender y usertype, se utiliza el valor -9, lo cual da instrucciones a la función de generar la columna correspondiente adherida a todos los viajes del periodo temporal correspondiente rellena con el valor -99. Esta operación es necesaria para poder conseguir un modelo canónico que permita la combinación de todas las ciudades estudiadas.

### **5.1.1 ENTRADA DE LA FUNCIÓN**

Una vez generado el fichero de entrada, y preparado el dataset correspondiente, se realiza una llamada a la función, cuyo funcionamiento será explicado a continuación. Las siguientes imágenes presentan el funcionamiento del sistema para tres inputs distintos y como se irían clasificando dentro de cada punto de transformación.

Como primer punto, y de cara al ejemplo, elegimos los ficheros de input, uno para Madrid, otro de Buenos Aires y el ultimo de Seúl. Como se puede apreciar, cada uno tiene unos valores distintos en los campos que se describían anteriormente, ya que las transformaciones a realizar no son coincidentes.

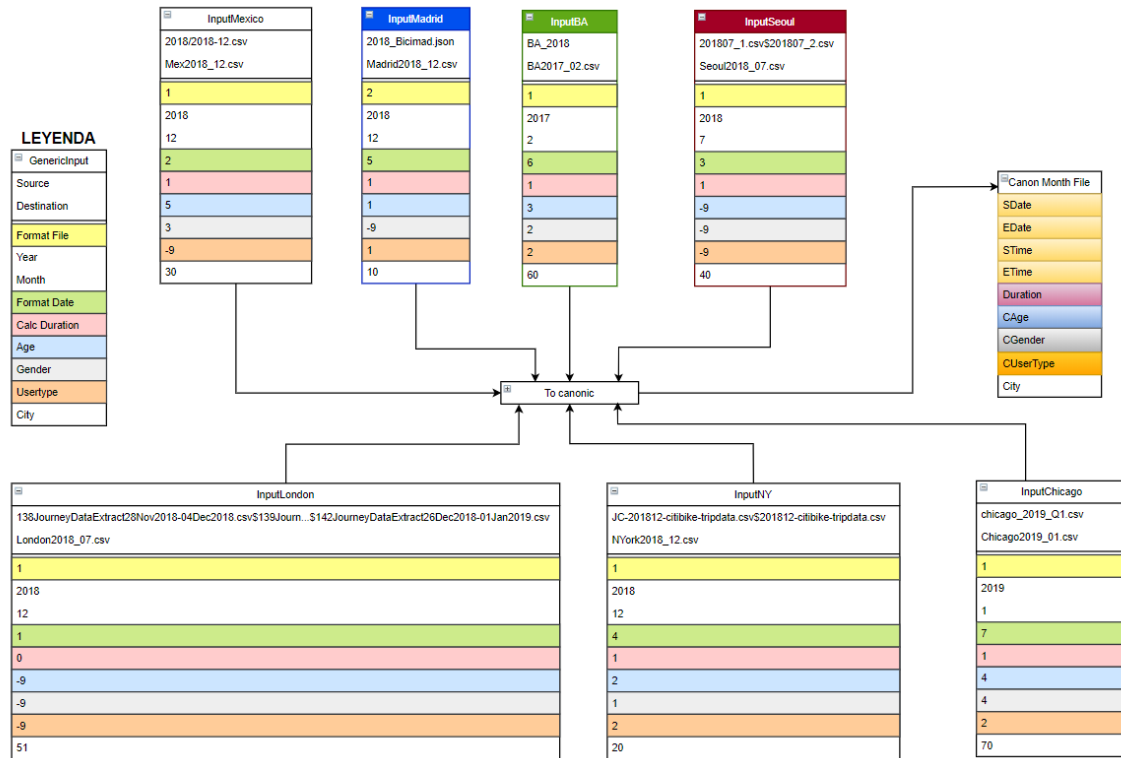


Figure 5-B Inputs de la función

### 5.1.2 FORMATO DEL ARCHIVO

El primer punto por el que debe pasar el dataset es la carga en la función, que realmente en términos computacionales no es más que una manera de convertir el archivo de datos inicial en un dataframe formato pandas, de manera que resulte sencilla la operativa. En este primer modelo, se contemplan dos posibilidades, en las cuales una de ellas a su vez se bifurca en otras dos posibles opciones.

Primero, y con una asignación de valor 1 en el campo format\_file, está la recepción de un fichero en formato csv. Dentro de esta sección, se puede dar el caso de que sea un solo fichero, el caso más sencillo ya que realmente la transformación es prácticamente directa, o que se trate de múltiples entradas. Este caso viene derivado de los datasets que dan la información fragmentada por ejemplo por semanas o quincenas en lugar de por mes. Para su

tratamiento, se pasa como input una concatenación de las direcciones de los ficheros y la propia función es capaz de reconocer y separar los mismos.

Segundo, y con una asignación de valor 2 en el campo format\_file, se trata la recepción del fichero en formato json, de lo cual cabe destacar que es computacionalmente más caro que el caso anterior.

Como salvaguarda, se genera una excepción con un mensaje de error en el caso de que el valor del campo format\_file no sea ninguno de los anteriores.

En la Figure 5-C Función, formato del archivo se muestra una representación del flujo que seguirían dentro de la función los inputs generados anteriormente.

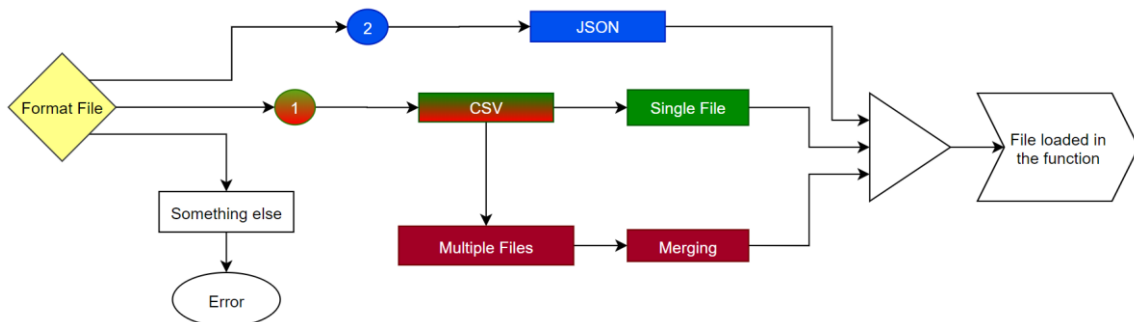


Figure 5-C Función, formato del archivo

### 5.1.3 FECHA Y HORA

El siguiente paso describe las operaciones necesarias para transformar la fecha y hora a la que sale la bicicleta de la estación de origen y a la que llega a la estación de destino. De nuevo se hace un Split en función del formato de los datos originales, y por ejemplo para un valor de format\_date igual a 3, se sabe que se ha de convertir del formato coreano al formato datetime que es el que la librería pandas utiliza para representar este tipo de datos.

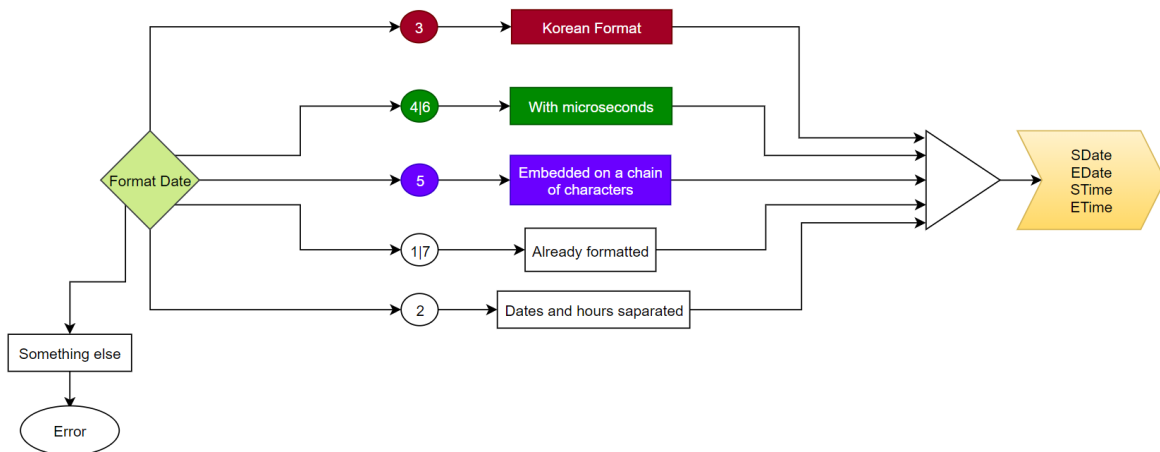
Existen casos como el 4 y el 6 o el 1 y el 7, en el que las operaciones realizadas son prácticamente las mismas con mínimas diferencias. Para la representación en el diagrama se

ha decidido combinarlas ya que no presenta nueva información el dejarlas separadas y en el fondo los métodos 6 y 7 no son más que el 1 y el 4 con una leve modificación.

De nuevo, al igual que en el apartado anterior, se contempla el caso en el que se pase un formato invalido de fecha.

La salida de este segmento produce cuatro de las columnas finales que se recogen en el modelo canónico, las fechas de inicio y final de viaje y las horas de inicio y final de viaje.

La Figure 5-D Función, formato de la fecha de nuevo representa la evolución por esta parte de la función de los inputs anteriores.

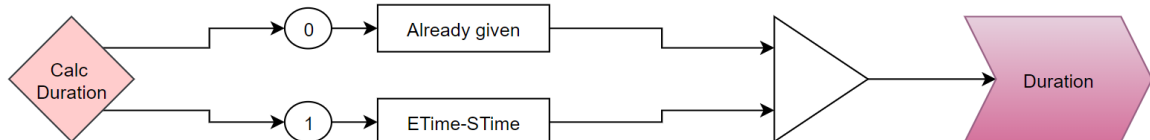


*Figure 5-D Función, formato de la fecha*

### 5.1.4 CÁLCULO DE LA DURACIÓN

El apartado del cálculo de la duración es probablemente el más sencillo de toda la función. Se contemplan dos casos, en el primero, la duración del viaje ya está definida en el propio dataset de origen, por lo que tan solo se extrae de ahí. El segundo caso es trivial una vez se tienen calculados los datos del apartado anterior, ya que sin mayor complicación se realiza una resta entre el valor de fin del viaje menor el valor de inicio.

De cara a la representación en la Figure 5-E Función, cálculo de la duración, dado que para los 3 casos que se estudian el cálculo es mediante la segunda opción, solo se dibuja una flecha en el flujo que engloba a las tres.

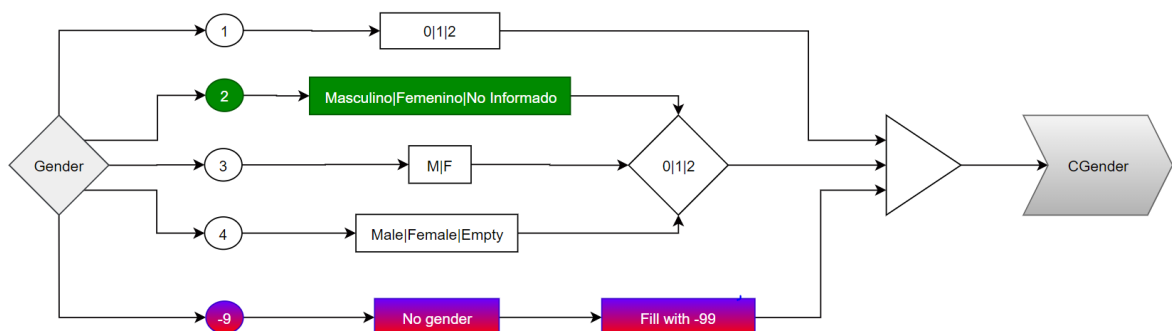


*Figure 5-E Función, cálculo de la duración*

### 5.1.5 GÉNERO

Para el cálculo del género se ha decidido utilizar una estructura que transforma los valores en cuatro dígitos distintos, -99 indica que no se facilita información de género, 0 indica que el género no es determinado, 1 indica usuario masculino y 2 indica usuario femenino.

De este modo, surgen varias transformaciones a realizar dependiendo de la base de datos, tal y como se muestra en la Figure 5-F Función, género, en la cual para los casos 2, 3 y 4 se ha de realizar una transformación extra en la que se convierten los valores de, por ejemplo, Masculino, Femenino y No Informado a los dígitos anteriormente citados.



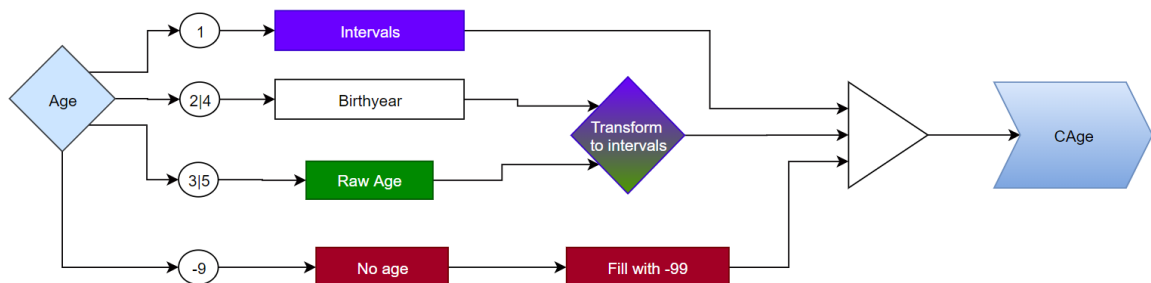
*Figure 5-F Función, género*

### 5.1.6 EDAD

Análogamente al cálculo del género, se realiza el cálculo de la edad de los usuarios. Dado que Madrid presta los datos de edad en forma de los siguientes intervalos a los cuales se les asignan valores. 1 para edades entre 10 y 16, 2 para 19 a 26, 3 para 27 a 40, 4 para 41 a 65 y 6 para 66 a 100.

Para poder mantener el formato homogéneo, se ha decidido pasar todos los datos de edad a esos mismos intervalos. De este modo, las tres formas en las que se reciben los datos son mediante los intervalos citados anteriormente, mediante la fecha de nacimiento del usuario o directamente mediante la edad pura del mismo.

De nuevo en el diagrama de la Figure 5-G Función, edad observamos los flujos de cada uno de los inputs y como para el caso de Buenos Aires donde se recibe puramente la edad del usuario es necesario realizar la transformación a intervalos.



*Figure 5-G Función, edad*

### 5.1.7 TIPO DE USUARIO

Por último, se tiene la clasificación según el tipo de usuario, en la que se ha decidido simplificar a de nuevo -99 para los casos en los que no se contempla el campo, 0 para usuario no suscrito al servicio y 1 para usuario suscrito al servicio.

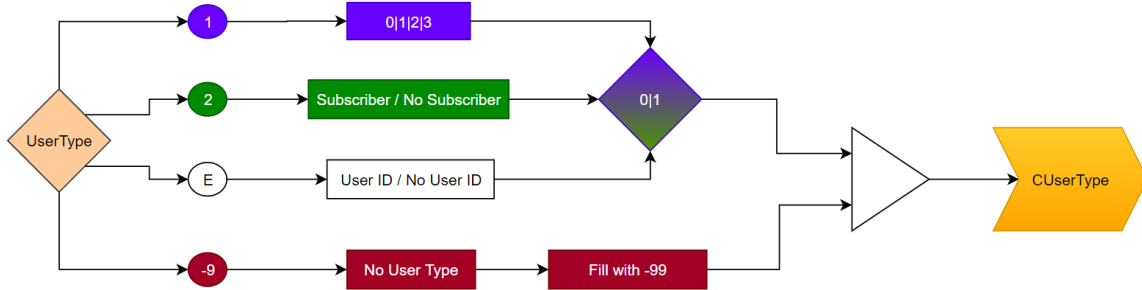


Figure 5-H Función, tipo de usuario

### 5.1.8 SALIDA DE LA FUNCIÓN

Finalmente, se obtiene una salida en el formato canónico como la que se presenta a continuación en la Figure 5-I Función, datos de salida, correspondiente a los datos de Nueva York en diciembre de 2017.

	SDate	EDate	STime	ETime	Duration	Age	Gender	User Type	City
0	2017-12-01	2017-12-01	00:00:59	00:08:32	453	4	1	1	NewYork
1	2017-12-01	2017-12-01	00:04:42	00:07:21	159	3	1	1	NewYork
2	2017-12-01	2017-12-01	00:17:00	00:19:55	175	3	1	1	NewYork
3	2017-12-01	2017-12-01	00:50:22	00:57:57	455	4	1	1	NewYork
4	2017-12-01	2017-12-01	00:55:03	01:06:13	670	5	1	1	NewYork
...	...	...	...	...	...	...	...	...	...
889962	2017-12-31	2018-01-01	23:54:22	00:01:00	398	5	2	1	NewYork
889963	2017-12-31	2018-01-01	23:54:44	00:00:16	332	4	1	1	NewYork
889964	2017-12-31	2018-01-01	23:56:07	00:05:33	566	0	1	1	NewYork
889965	2017-12-31	2018-01-01	23:57:16	00:24:56	1660	4	2	1	NewYork
889966	2017-12-31	2018-01-01	23:58:56	00:20:16	1280	5	2	1	NewYork

Figure 5-I Función, datos de salida

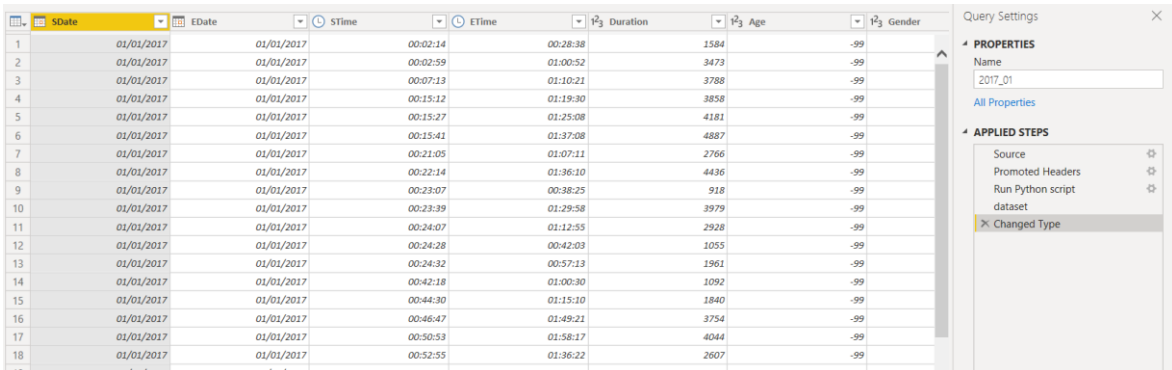
### 5.1.9 AGRUPACIÓN DE LAS SALIDAS

Una vez de obtienen las salidas correspondientes a todas las ciudades para el mes del año con el que se esté trabajando se ejecuta un script que combina esas mismas salidas en un único fichero, con el objetivo de crear una simulación de una base de datos que se actualiza periódicamente. Estos archivos finales serán los que se utilizarán en Power BI para la representación y el estudio. No se adjunta una imagen de este fichero final ya que realmente no es más que el fichero presentado como ejemplo anteriormente, pero con el resto de las ciudades concatenadas.

## 5.2 CONFIGURACIÓN Y HERRAMIENTAS POWER BI

Una vez se tienen los datos de salida de la función, antes de realizar la representación son necesarios una serie de pasos intermedios en Power BI para obtener el resultado deseado.

Primero, y como el resultado que se busca es un análisis prácticamente automático año a año, se genera un proyecto genérico capaz de realizar el análisis para cualquier año que se necesite. Para esto, el primer paso a realizar es hacer el chequeo por duración, se ignorarán viajes con duración menor de 60 segundos y mayor de 7200. La forma de llevar esto a cabo será mediante un script de Python que se ejecutará en Power BI para cada uno de los ficheros de entrada.



	SDate	EDate	STime	ETime	Duration	Age	Gender
1	01/01/2017	01/01/2017	00:02:14	00:28:38	1584	-99	
2	01/01/2017	01/01/2017	00:02:59	01:00:52	3473	-99	
3	01/01/2017	01/01/2017	00:07:13	01:10:21	3788	-99	
4	01/01/2017	01/01/2017	00:15:12	01:19:30	3858	-99	
5	01/01/2017	01/01/2017	00:15:27	01:25:08	4181	-99	
6	01/01/2017	01/01/2017	00:15:41	01:37:08	4887	-99	
7	01/01/2017	01/01/2017	00:21:05	01:07:11	2766	-99	
8	01/01/2017	01/01/2017	00:22:14	01:36:10	4436	-99	
9	01/01/2017	01/01/2017	00:23:07	00:38:25	918	-99	
10	01/01/2017	01/01/2017	00:23:39	01:29:58	3979	-99	
11	01/01/2017	01/01/2017	00:24:07	01:12:55	2928	-99	
12	01/01/2017	01/01/2017	00:24:28	00:42:03	1055	-99	
13	01/01/2017	01/01/2017	00:24:32	00:57:13	1961	-99	
14	01/01/2017	01/01/2017	00:42:18	01:00:30	1092	-99	
15	01/01/2017	01/01/2017	00:44:30	01:15:10	1840	-99	
16	01/01/2017	01/01/2017	00:46:47	01:49:21	3754	-99	
17	01/01/2017	01/01/2017	00:50:53	01:58:17	4044	-99	
18	01/01/2017	01/01/2017	00:52:55	01:36:22	2607	-99	

Figure 5-J Queries Power BI

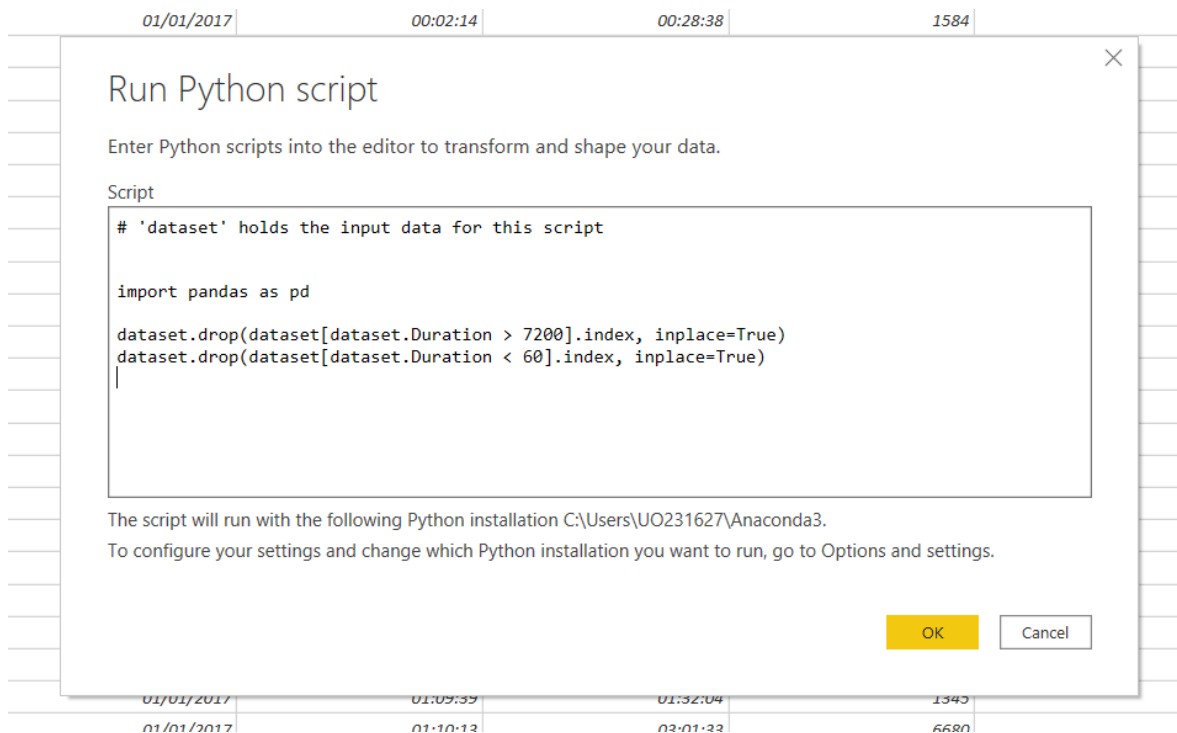
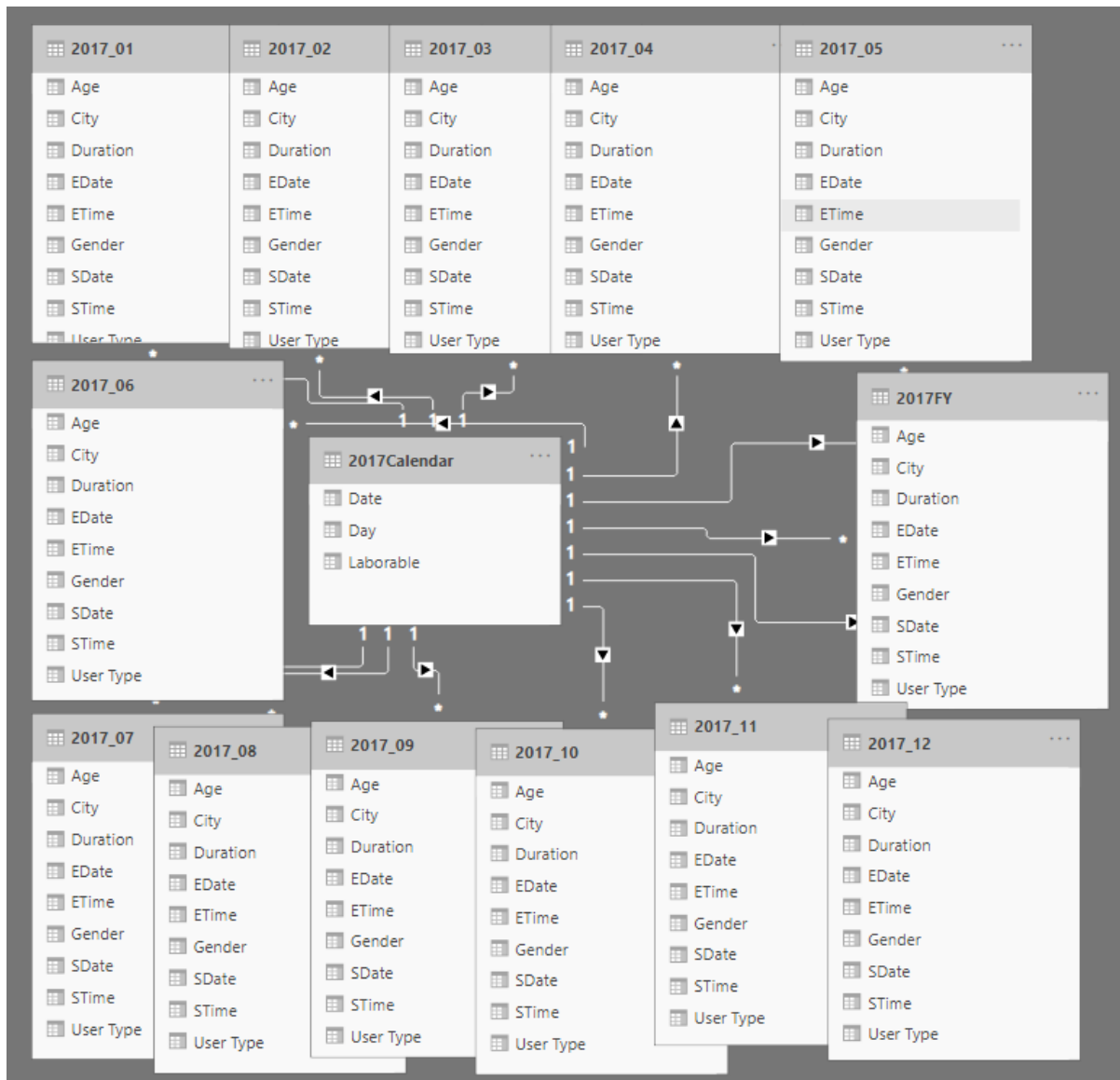


Figure 5-K Script duración Power BI

Una vez completado este paso, se tienen los doce archivos que han sido generados por la función, a los que se les ha hecho la transformación para eliminar las duraciones fuera de intervalo, y el archivo calendario. El siguiente paso consiste en unificar los doce archivos y generar uno nuevo que será el equivalente al año entero. Una vez realizado este paso, y para que los ejes de tiempo coincidan, se ponen en común las fechas de todos los meses con el calendario.



*Figure 5-L Relaciones Power BI*

### **5.3 REPRESENTACIÓN DE LOS DATOS**

Como ya se ha comentado en apartados anteriores, la herramienta elegida para realizar la representación de los datos es Power BI, y para ello el punto de partida es la utilización del archivo descrito en el punto anterior que es finalmente generado por la función y comprende todos los viajes realizados en el periodo de un mes para todas las ciudades bajo estudio.

### 5.3.1 PRIMERA ITERACIÓN

El primer paso realizado para la representación comprende unos ejercicios muy básicos de comprobación de resultados. Para el ejemplo propuesto se representa la media de la duración de los viajes por día y ciudad de los tres primeros meses del año individualmente para asegurar que los datos generados son correctos. Así, los resultados obtenidos para el año 2018 se muestran a continuación:

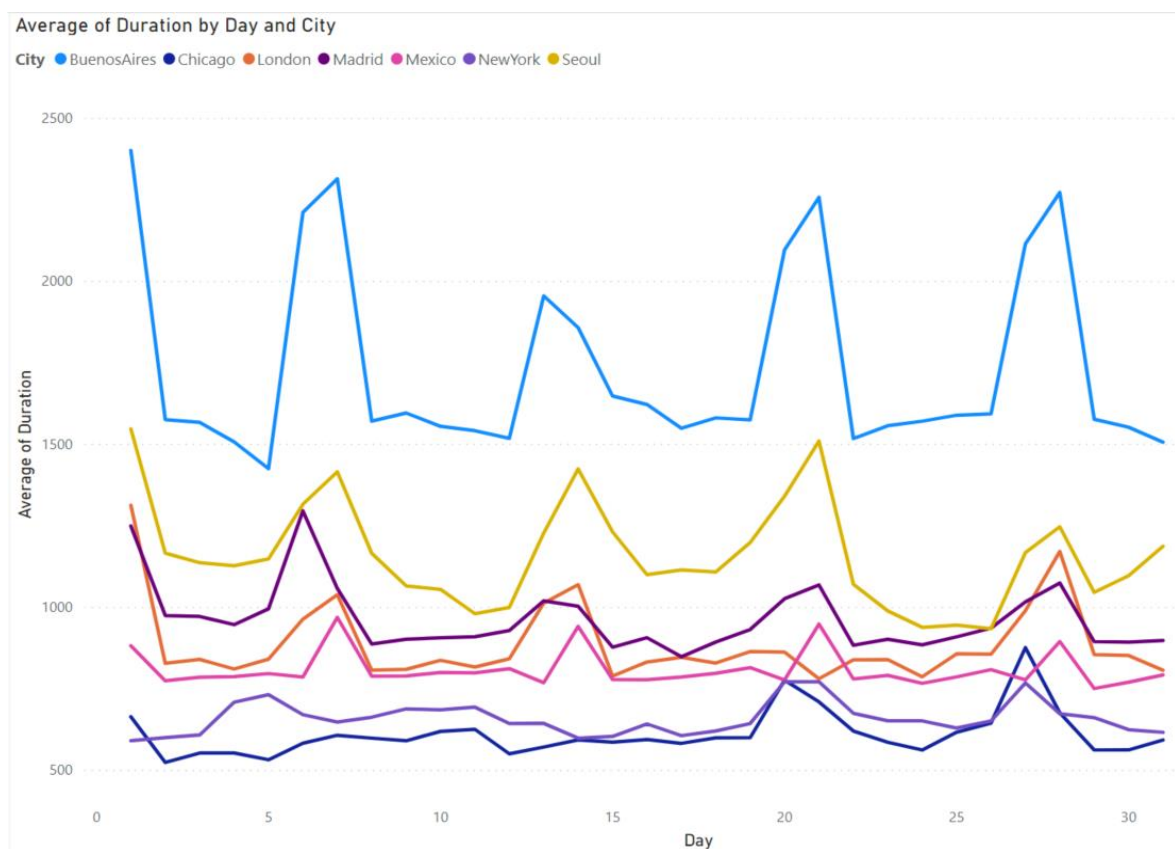
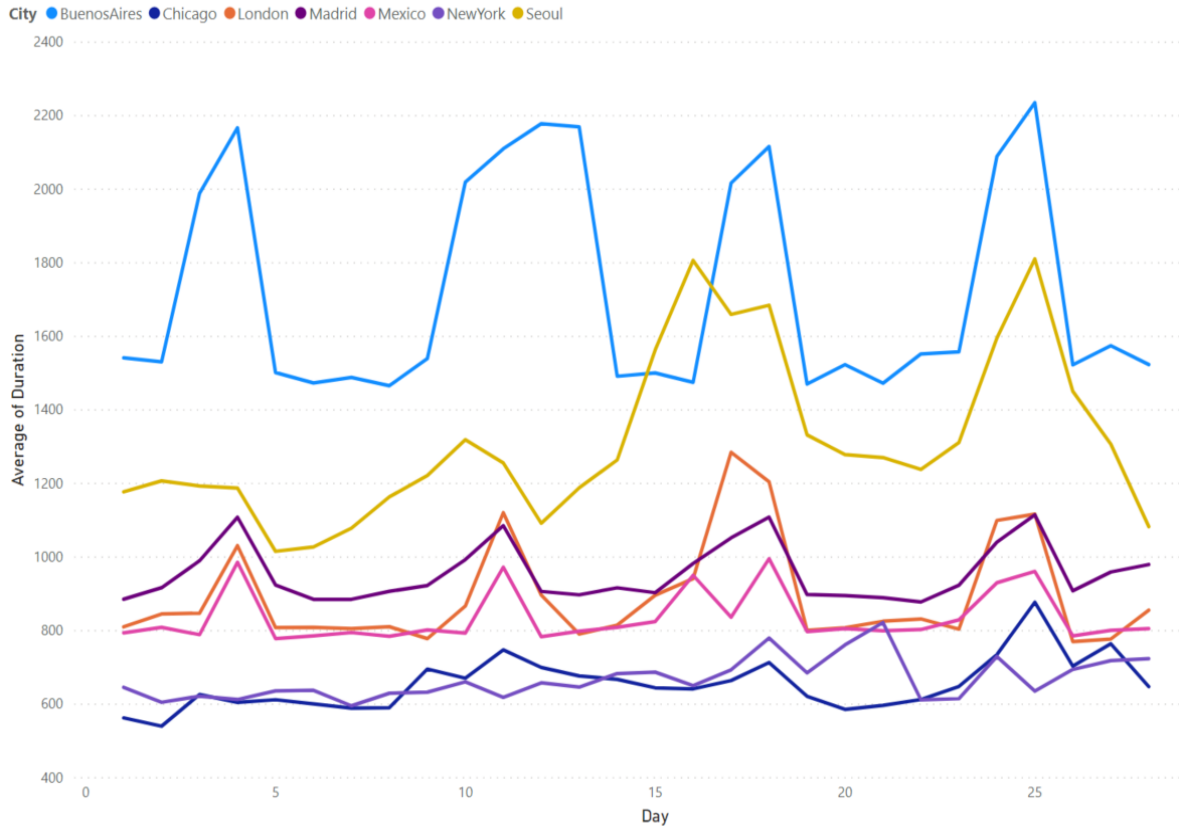


Figure 5-M Comprobación Enero 2018

Average of Duration by Day and City



*Figure 5-N Comprobación Febrero 2018*

Average of Duration by Day and City

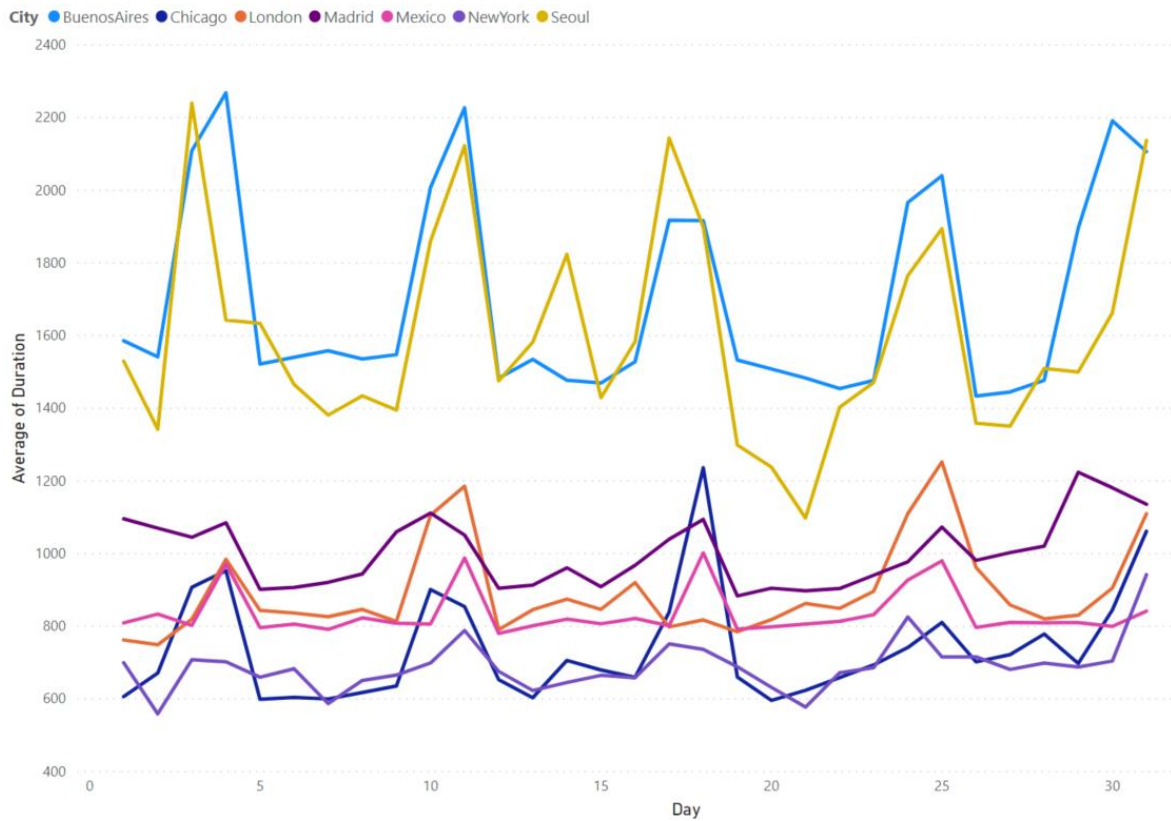


Figure 5-O Comprobación Marzo 2018

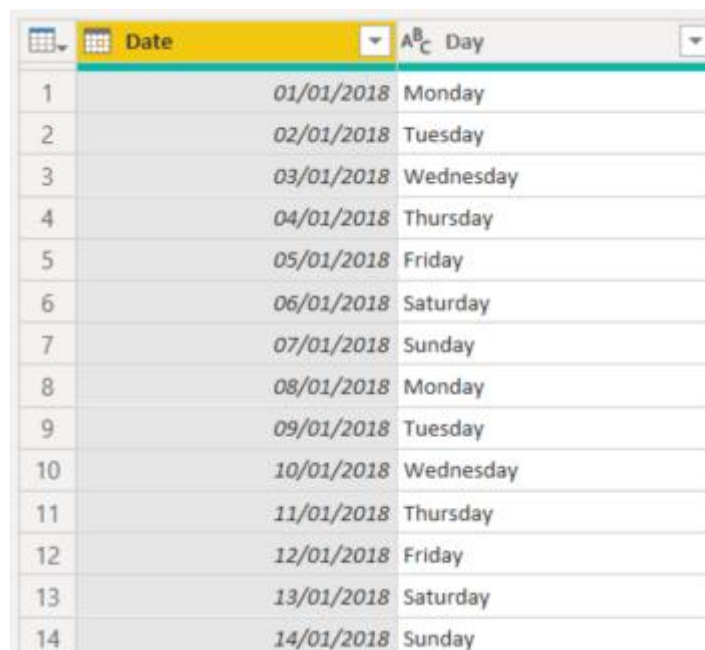
Se ha de matizar que en este paso no se trata de sacar ninguna información o conclusión de los datos más allá del saber que han sido generados correctamente, no hay huecos vacíos y los valores se encuentran dentro del intervalo esperado. Se podría decir que esta primera interacción es una fase de comprobación y corrección previa al trabajo de modificación e interpretación de los datos.

Este proceso se repite para cada uno de los archivos generados por la función antes de continuar al apartado siguiente.

### 5.3.2 SEGUNDA ITERACIÓN

En esta segunda iteración y mediante herramientas propias de Power BI, se genera un eje de tiempos que comprende la extensión completa del año en el que se está trabajando con el fin de poder representar los datos que anteriormente estaban separados por mes de una manera completa.

A continuación, se presenta el fichero Calendario, cuya construcción se realiza de forma automática en función del año que se quiera tratar. De cara a los datos que lleva, no son más que las fechas numéricas con su día de la semana correspondiente.

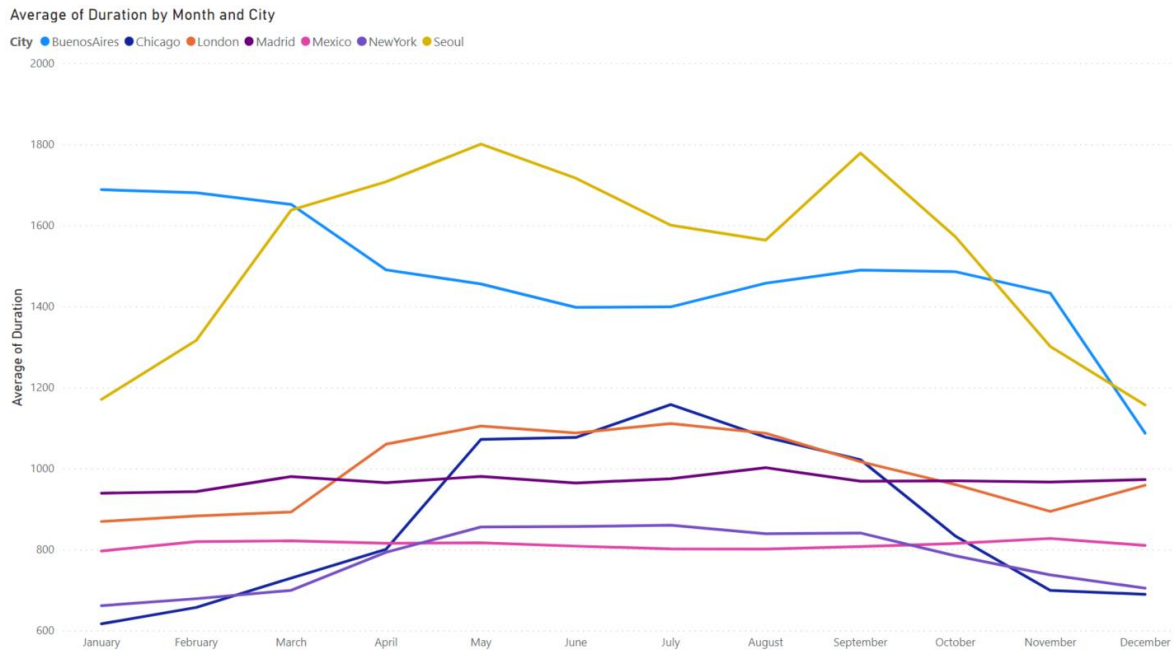


	Date	Day
1	01/01/2018	Monday
2	02/01/2018	Tuesday
3	03/01/2018	Wednesday
4	04/01/2018	Thursday
5	05/01/2018	Friday
6	06/01/2018	Saturday
7	07/01/2018	Sunday
8	08/01/2018	Monday
9	09/01/2018	Tuesday
10	10/01/2018	Wednesday
11	11/01/2018	Thursday
12	12/01/2018	Friday
13	13/01/2018	Saturday
14	14/01/2018	Sunday

Figure 5-P Calendario 1

Con la utilidad que presta el calendario y los ficheros de todos los meses del año se compone finalmente el fichero final sobre el que se realizarán las pruebas y del que se extraerán las conclusiones. Este fichero es sencillamente una combinación mediante transposición de los doce ficheros anteriores. Técnicamente, se podría hacer una composición de mayor tamaño y no frenar en un año completo, no obstante, por limitaciones técnicas, no se ha podido sobrepasar ese umbral.

Se muestra a continuación una representación de la duración media anual por ciudad tras haberse realizado los pasos anteriores.



*Figure 5-Q Comprobación duración media por mes 2018*

### 5.3.3 ITERACIÓN FINAL

Para completar el último paso con respecto a la representación de los datos, se ha decidido establecer un modelo de funcionamiento en el que la representación es totalmente ajena a la generación de los datos. De este modo, los archivos generados por la función se almacenan en una carpeta de la que bebe el set up preparado en Power BI, simulando una base de datos de actualización periódica.

Dado que la función únicamente transforma los datos al modelo canónico sin ningún tipo de inteligencia adicional, es posible que se cuelen en los datos finales algunos viajes erróneos que no deberían ser tenidos en cuenta, como movimientos de las bicicletas por la propia empresa o situaciones en las que el posible usuario desengancha la bicicleta sin querer y la vuelve a enganchar. Para solucionar este problema, se desarrolla una función aplicada en

Power BI que filtra la duración de los viajes y elimina todos aquellos con una duración inferior a 60 segundos y todos aquellos con una duración superior a 7200 segundos.

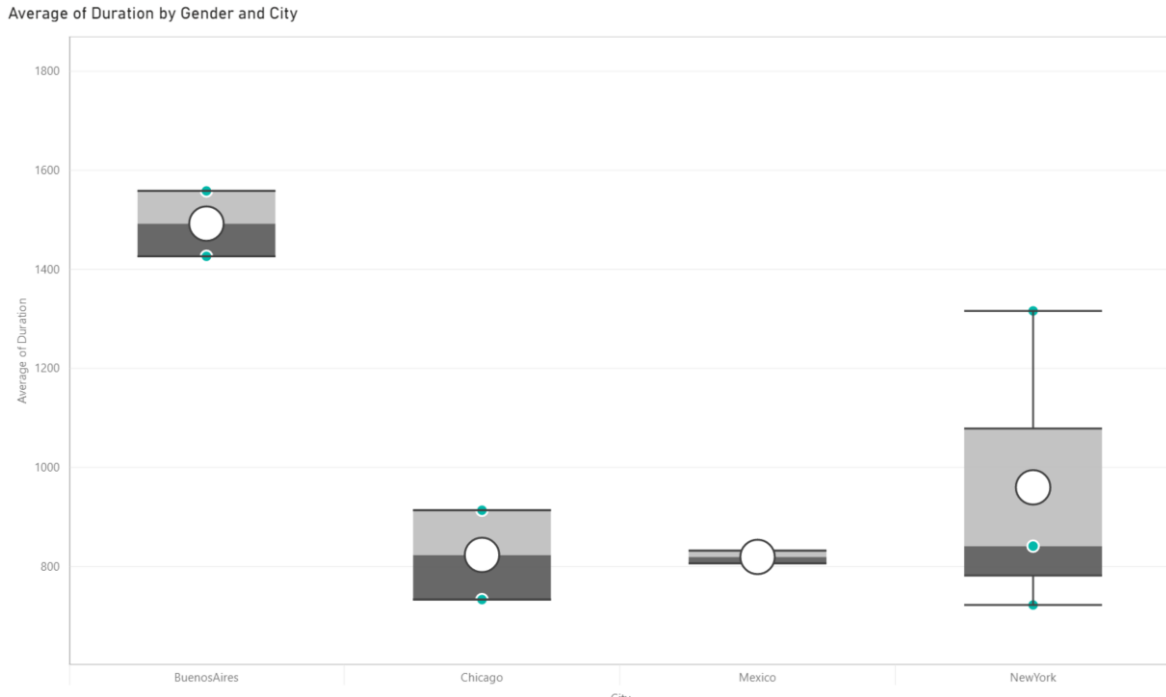
Además, mediante la inteligencia del propio programa Power BI, únicamente se tiene un archivo sobre el que el trabajo a realizar no es más que cargar los datos correspondientes al año que se quiera estudiar. Siempre que se mantenga el formato original, el proyecto ejecutara las instrucciones de forma automática y se obtendrán los resultados deseados son necesidad de trabajar sobre el mismo.

Por otro lado, se añade al calendario una columna adicional llamada “Laborable” que tiene valor cero cuando los días son de lunes a viernes y valor uno cuando es sábado o domingo. La utilidad de esta nueva columna será explicada más adelante.

	Date	Day	Laborable
1	01/01/2018	Monday	0
2	02/01/2018	Tuesday	0
3	03/01/2018	Wednesday	0
4	04/01/2018	Thursday	0
5	05/01/2018	Friday	0
6	06/01/2018	Saturday	1
7	07/01/2018	Sunday	1
8	08/01/2018	Monday	0
9	09/01/2018	Tuesday	0
10	10/01/2018	Wednesday	0
11	11/01/2018	Thursday	0
12	12/01/2018	Friday	0
13	13/01/2018	Saturday	1
14	14/01/2018	Sunday	1
15	15/01/2018	Monday	0
16	16/01/2018	Tuesday	0
17	17/01/2018	Wednesday	0
18	18/01/2018	Thursday	0
19	19/01/2018	Friday	0

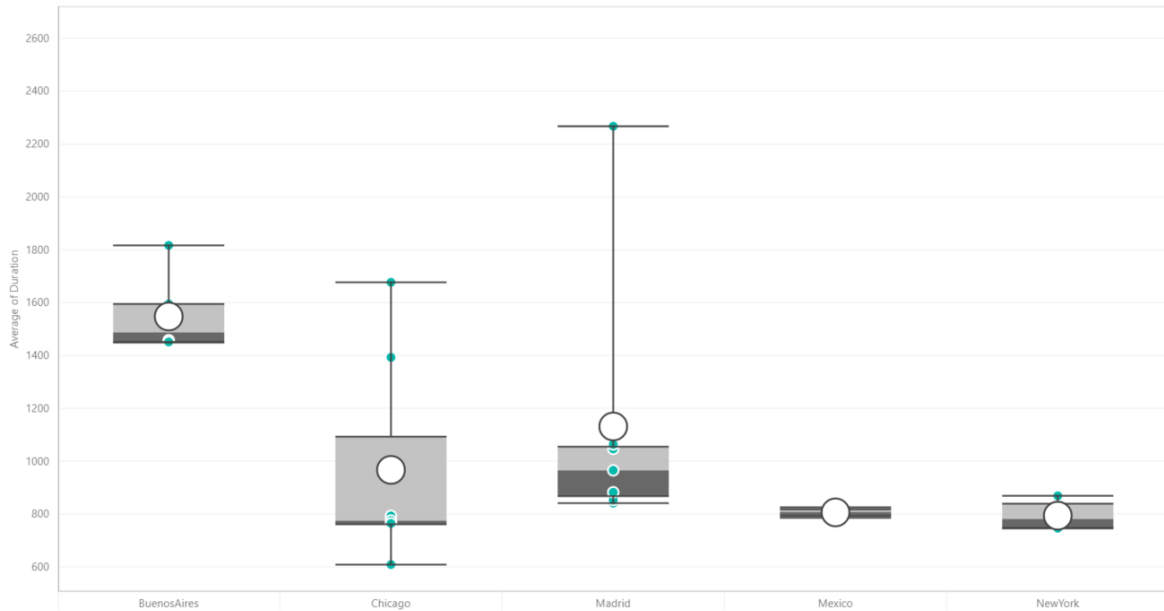
*Figure 5-R Calendario 2*

De este modo, los resultados que se obtienen de cara únicamente a la representación son los siguientes:



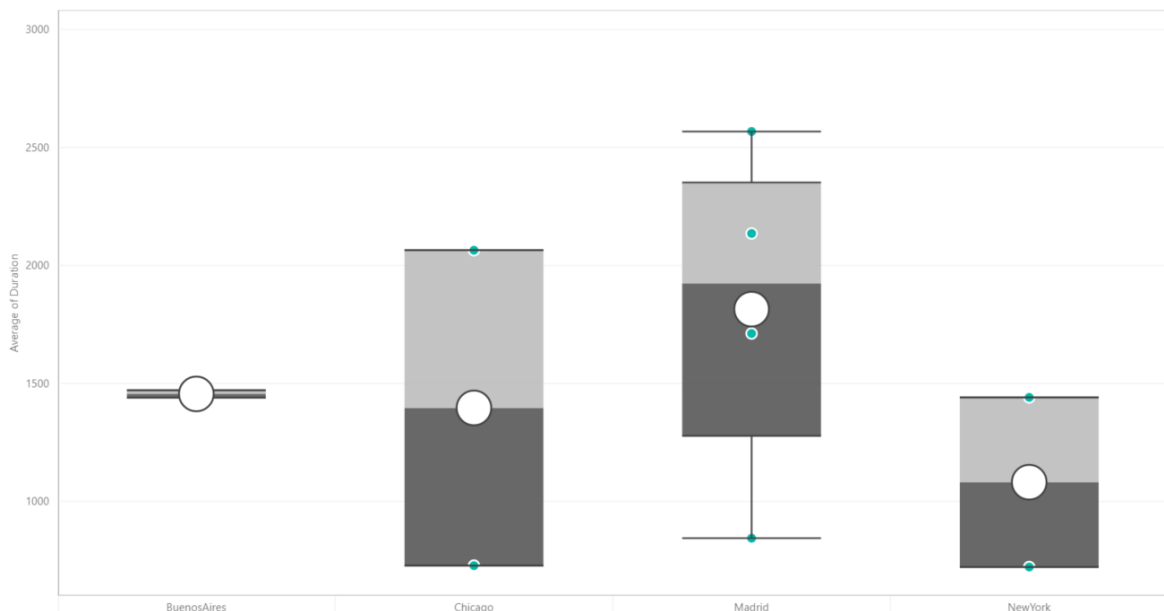
*Figure 5-S Comprobación media duración por género y ciudad*

Average of Duration by Age and City



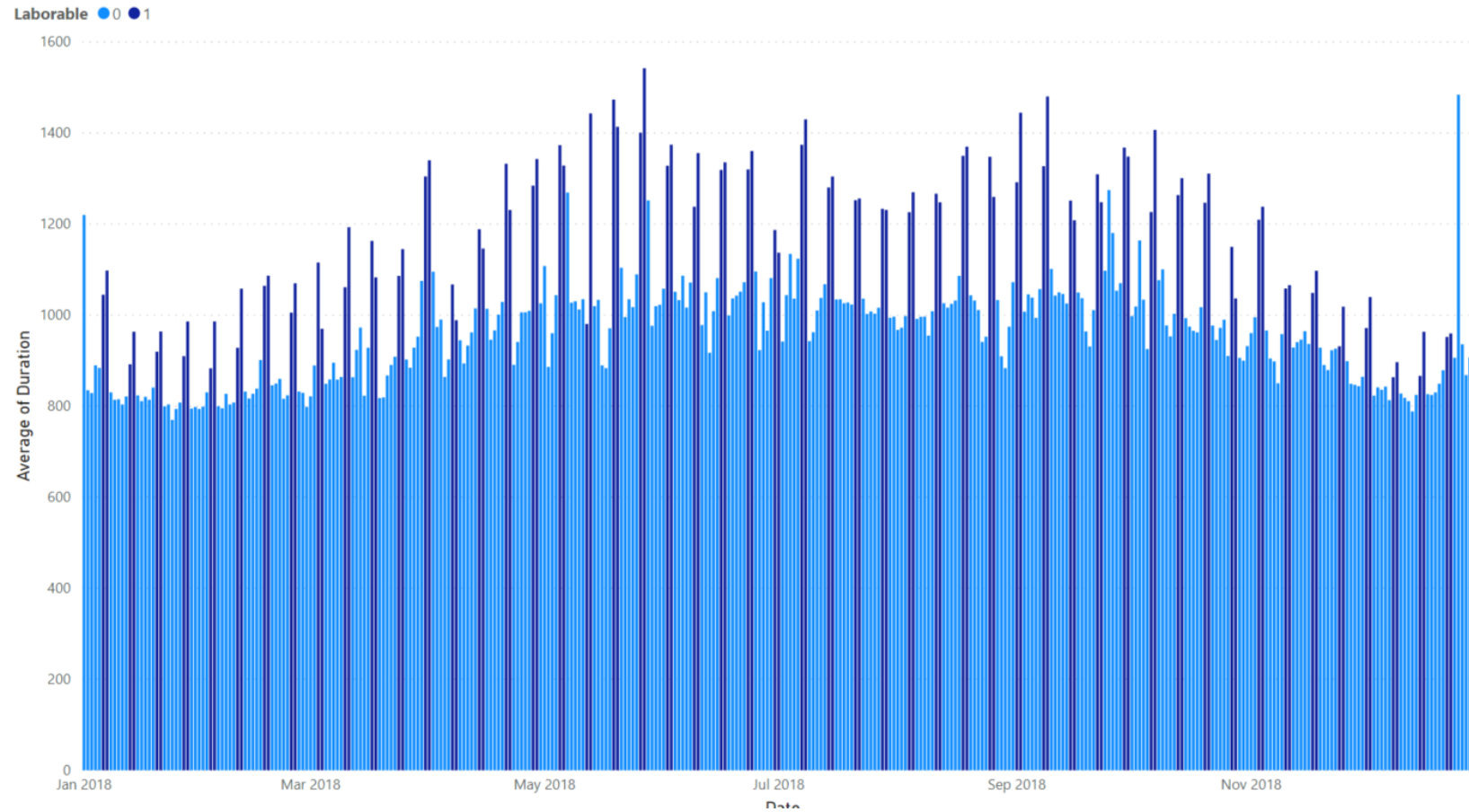
*Figure 5-T Comprobación media duración por edad y ciudad*

Average of Duration by User Type and City



*Figure 5-U Comprobación media duración por tipo de usuario y ciudad*

Average of Duration by Date and Laborable



*Figure 5-V Comprobación media duración en función del día del año*

Se utilizan diagramas de bloques para tener una visión clara de cómo se distribuyen los datos de la muestra, esta representación muestra claramente la media, las distribuciones en cuartiles y la distancia de los outliers.

Además de ese formato, también se utilizan gráficos de barras cuando es necesario hacer un muestreo temporal en lugar de divisiones por ciudades, ya que se entiende que esa representación con diagramas de bloque no tiene sentido.

Por último, y con respecto a la ampliación del segundo calendario, se puede observar en la figura 6-S la utilidad de la misma, permite realizar una división en día laborable y no laborable de los viajes anuales.

## Capítulo 6. ANÁLISIS DE RESULTADOS

En esta sección de la memoria se describen con detalle los resultados obtenidos mediante las operaciones y las representaciones anteriormente descritas. Esta sección se compone a su vez de varias partes, cada una de ellas dedicada al estudio concreto y resolución de las cuestiones planteadas inicialmente al comienzo del trabajo.

### 6.1 ESTUDIO DE LAS VARIACIONES DE DURACIÓN A NIVEL ANUAL

El primer apartado de esta sección presenta un estudio relativamente sencillo pero que responde a diversas preguntas que fueron surgiendo durante la fase de la analítica de datos.

Las siguientes figuras muestran la evolución de la media de las duraciones por mes de los viajes a lo largo de los años bajo estudio.

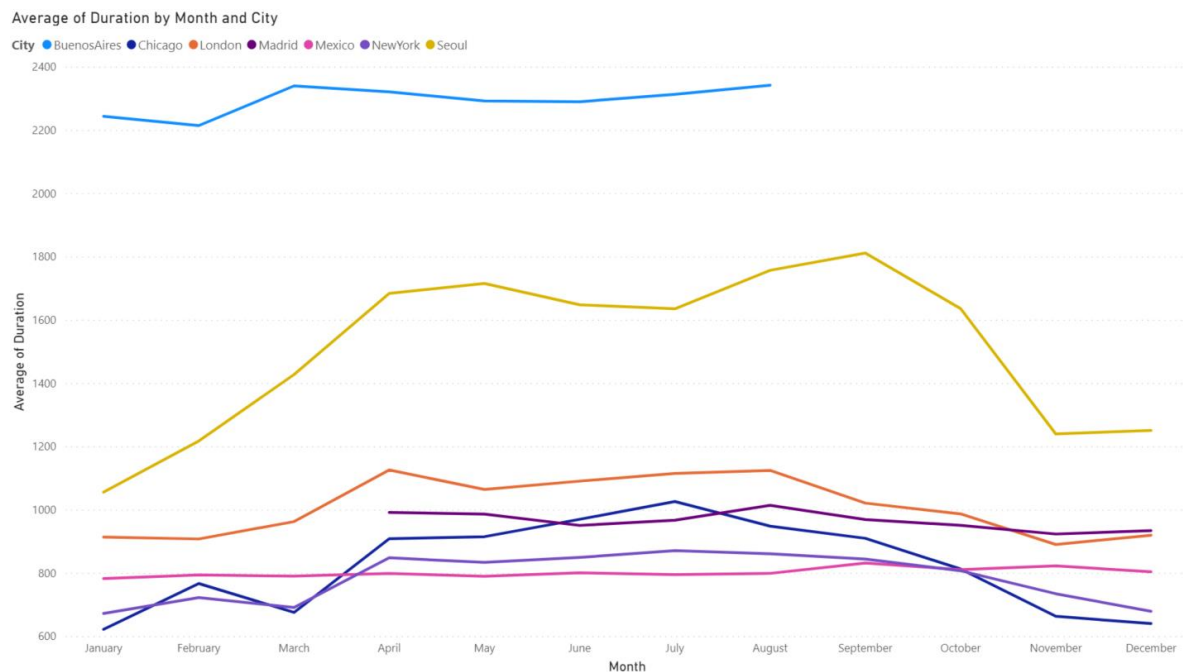
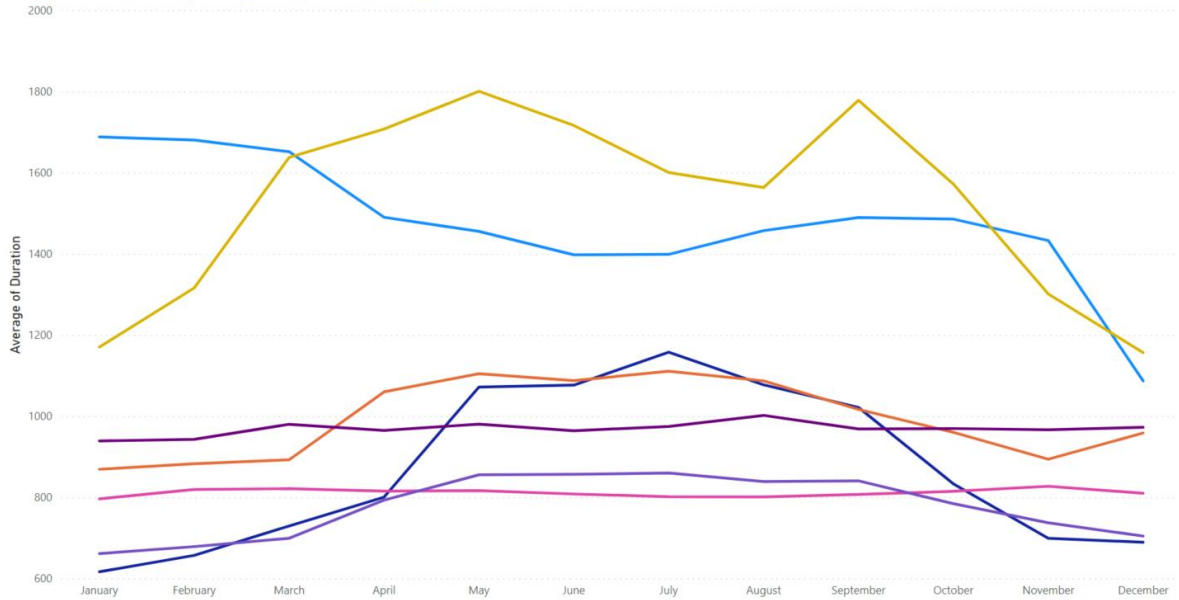


Figure 6-A Duración a nivel anual 2017

Average of Duration by Month and City

City ● BuenosAires ● Chicago ● London ● Madrid ● Mexico ● NewYork ● Seoul



*Figure 6-B Duración a nivel anual 2018*

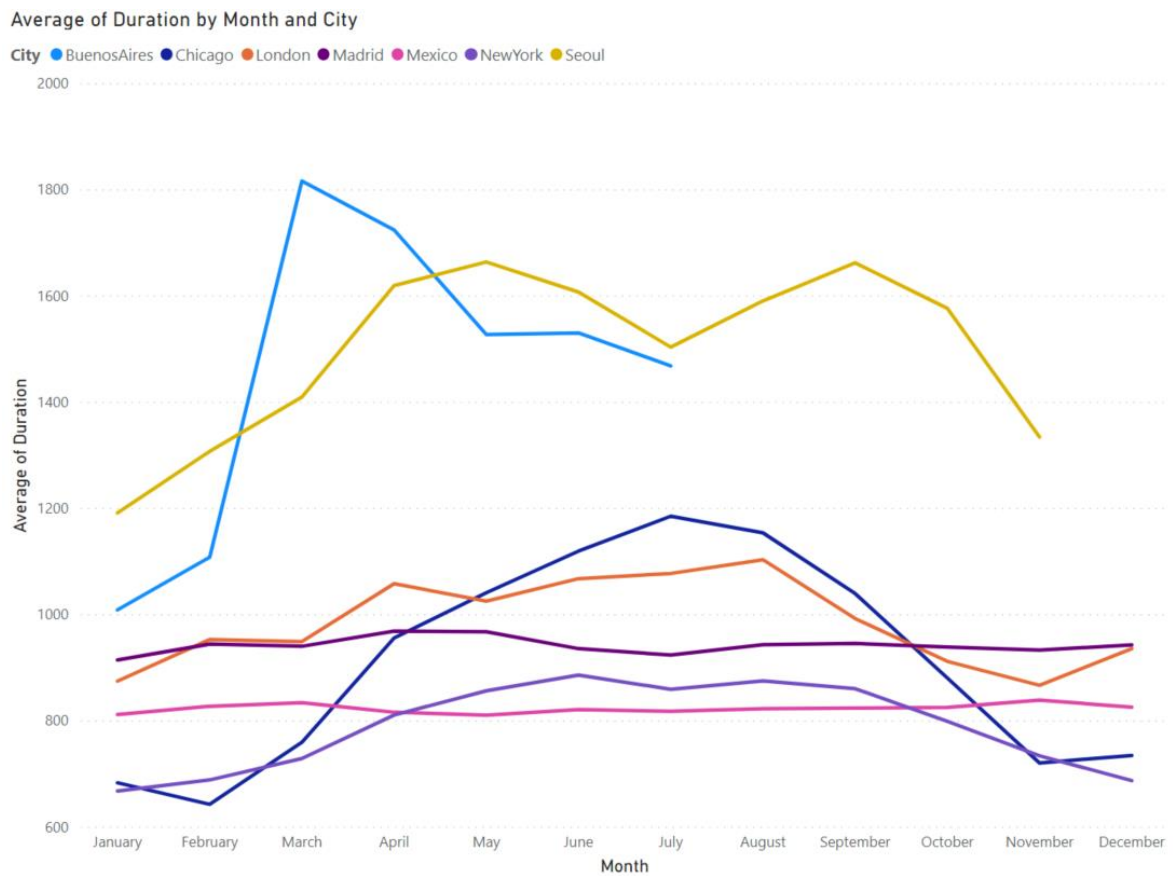


Figure 6-C Duración a nivel anual 2019

Parece lógico que la duración y la frecuencia de los viajes realizados por los habitantes de las ciudades bajo estudio presenten un aumento durante los meses de verano, como sucede por ejemplo en Buenos Aires, es por esto por lo que resulta curioso ver como para los meses de junio, julio y agosto se puede observar perfectamente cómo la duración media de los viajes realizados en Seúl es notablemente inferior a meses como abril o septiembre.

Del mismo modo, resulta interesante observar cómo en Chicago y Nueva York el volumen y la duración de los viajes realizados presenta una disminución más que notable durante los meses de enero, febrero, marzo, noviembre y diciembre.

Otro punto por estudiar es la constancia en la duración de los viajes en Madrid y México, sitios en los que apenas se observa variación a lo largo de los meses.

Con el objetivo de entender los factores anteriormente citados, se ha hecho un estudio del comportamiento de la meteorología en las ciudades con las que se trabaja, obteniéndose los resultados que se adjuntan en el Anexo D - Meteorología.

Las conclusiones extraídas por tanto y que responden directamente a las preguntas planteadas inicialmente son las siguientes.

Para el caso de Seúl, los meses de junio, julio y agosto coinciden con la época de monzones en la zona, lo que explica el no aumento de la duración de los viajes en esas fechas. Del mismo modo, las bajas temperaturas en diciembre, enero y febrero también encajan con el resto de los resultados.

En Chicago y Nueva York, los meses de uso reducido del servicio son diciembre, enero y febrero, coincidentes con las épocas de nevada de la zona. De nuevo resultados coherentes ya que se entiende que la nieve es un gran obstáculo para el uso de las bicicletas para el desplazamiento.

Por otro lado, Madrid, México y, aunque con cierta salvedad Londres, presentan unos datos meteorológicos bastante constantes a lo largo del año en comparación con las ciudades anteriores, lo que se refleja en un uso sin apenas oscilaciones durante los tiempos bajo estudio.

De esta parte se extrae por tanto que los factores que más limitan el uso del servicio con respecto a la meteorología no son tanto las oscilaciones de temperatura si no sucesos más agravantes como el alto volumen de precipitaciones tanto en forma de lluvia como en forma de nieve.

## 6.2 ESTUDIO DE LA DURACIÓN POR EDADES

Uno de los valores que se obtiene para la base de datos es la edad del usuario que utiliza el servicio. No obstante, no es un dato que faciliten todas las ciudades, concretamente Seúl y Londres y en un aparte Buenos Aires para el año 2017.

Por otro lado, y como se comentaba al principio de la memoria, debido a que los datos inicialmente dados por la ciudad de Madrid presentan la edad de los usuarios en forma de intervalo, se ha decidido utilizar ese formato como el estándar y traducir a ese estilo las edades del resto de ciudades.

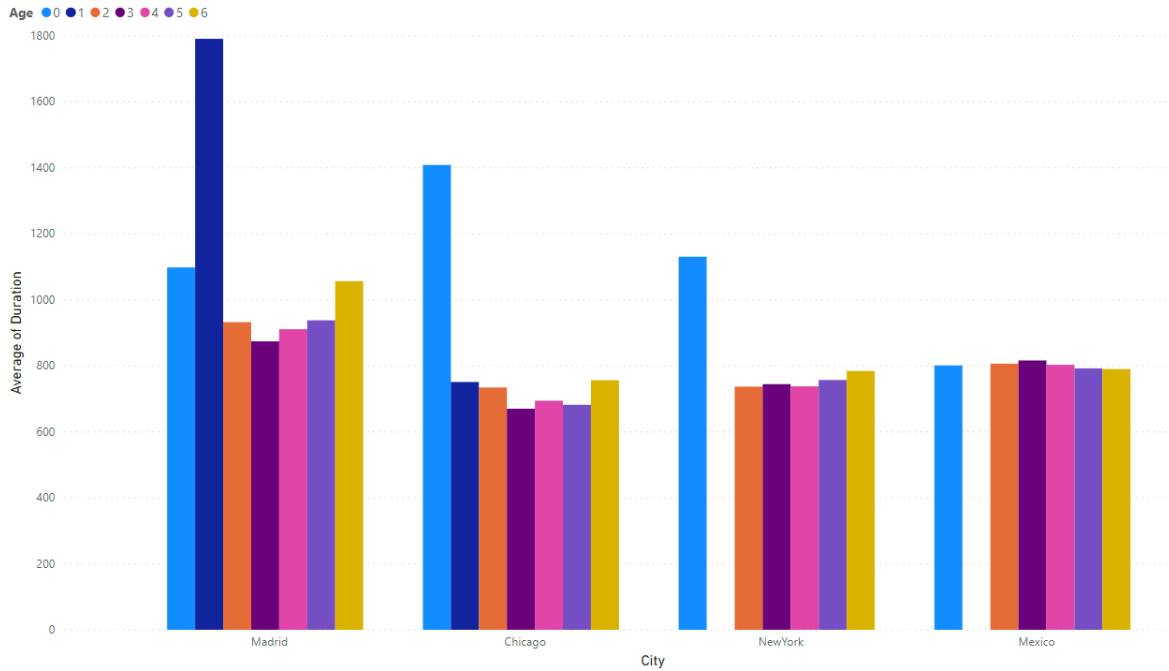
De tal manera, en la tabla siguiente se muestra el criterio seguido para dar los valores pertinentes a los intervalos de las edades.

Valor	Rango de edad
0	No se ha podido determinar el rango de edad del usuario
1	El usuario tiene entre 0 y 16 años
2	El usuario tiene entre 17 y 18 años
3	El usuario tiene entre 19 y 26 años
4	El usuario tiene entre 27 y 40 años
5	El usuario tiene entre 41 y 65 años
6	El usuario tiene 66 años o más

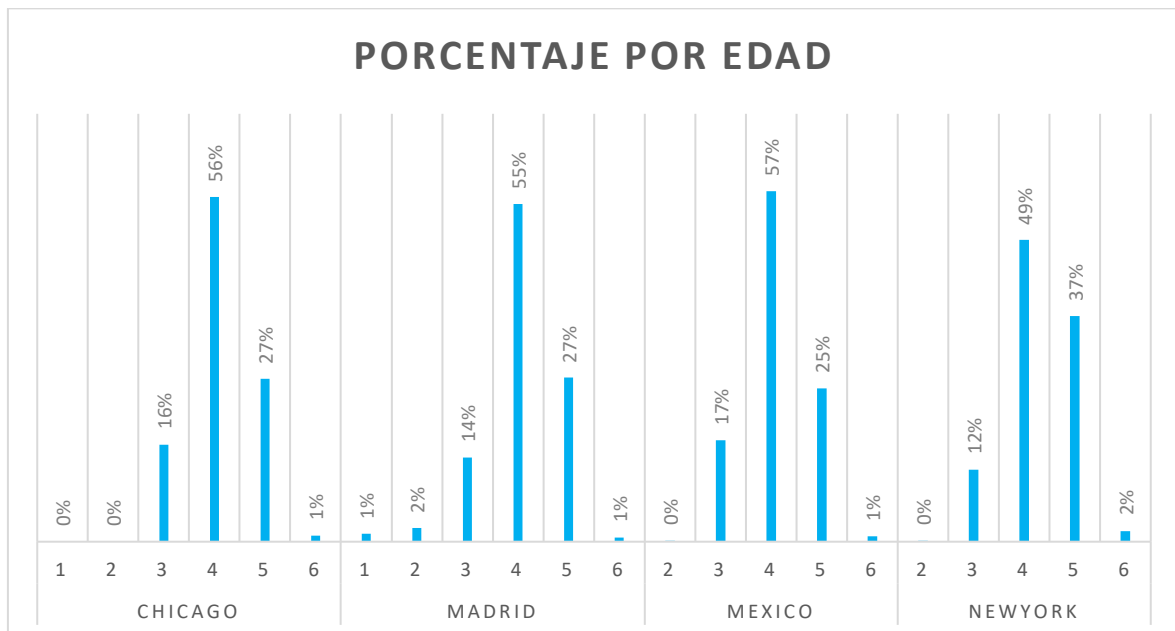
Table 3 Rangos de edad

A continuación, se muestran varias figuras que representan por un lado la duración media de los viajes de los usuarios por año en función de su edad por ciudad y por otro lado el conteo de viajes para cada franja de edad en cada ciudad en porcentaje.

Average of Duration by City and Age



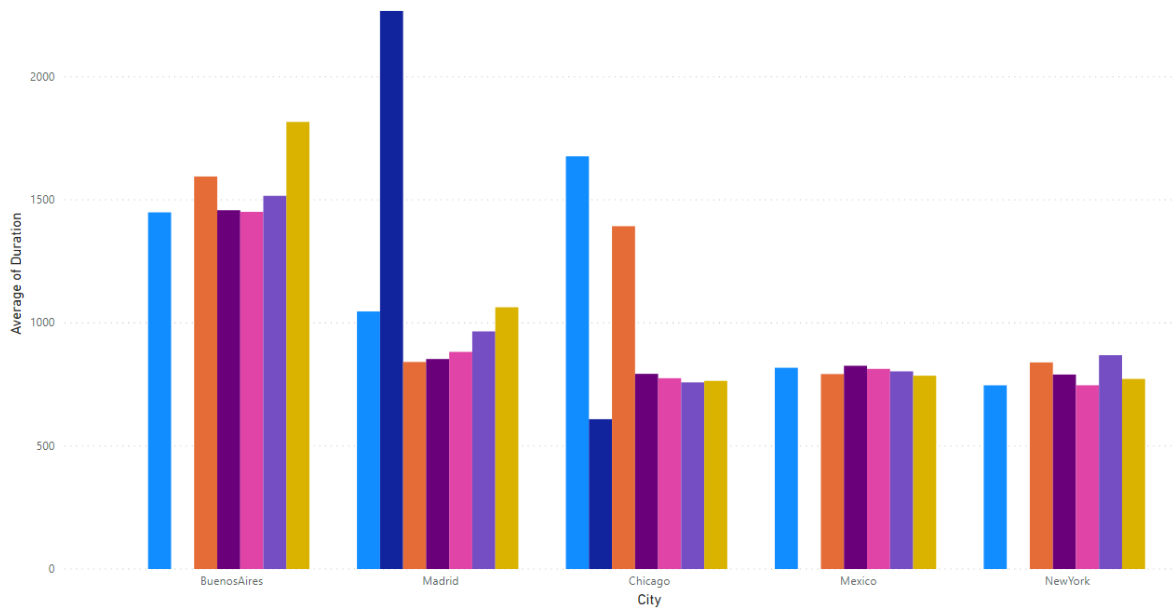
*Figure 6-D Duración media edad 2017*



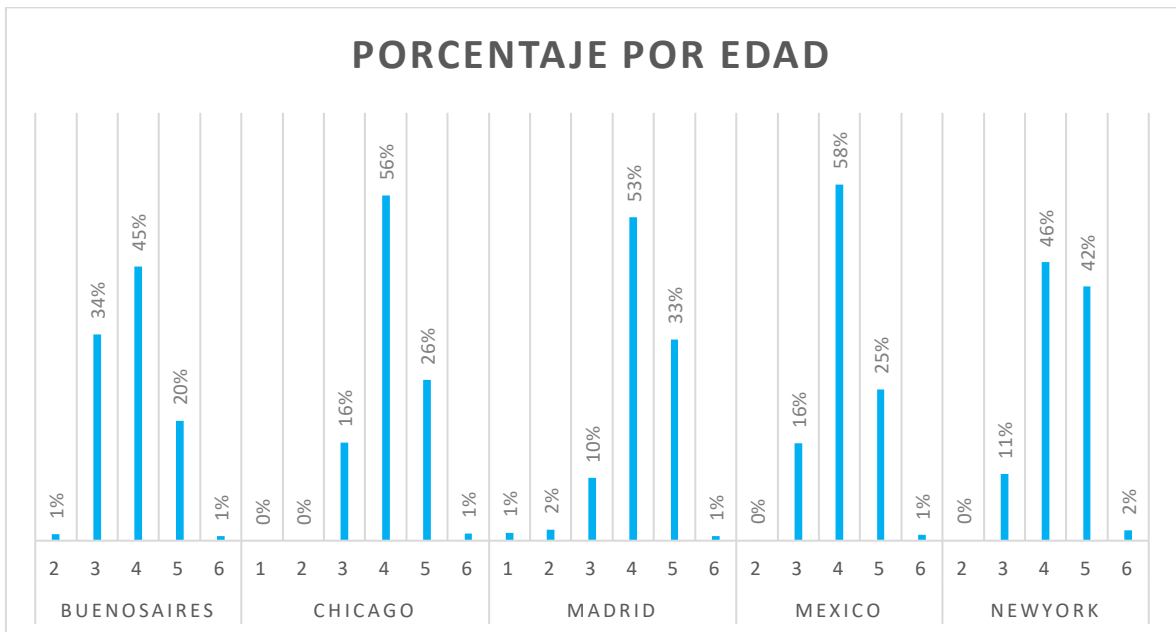
*Figure 6-E Porcentaje conteo edad 2017*

Average of Duration by City and Age

Age 0 1 2 3 4 5 6



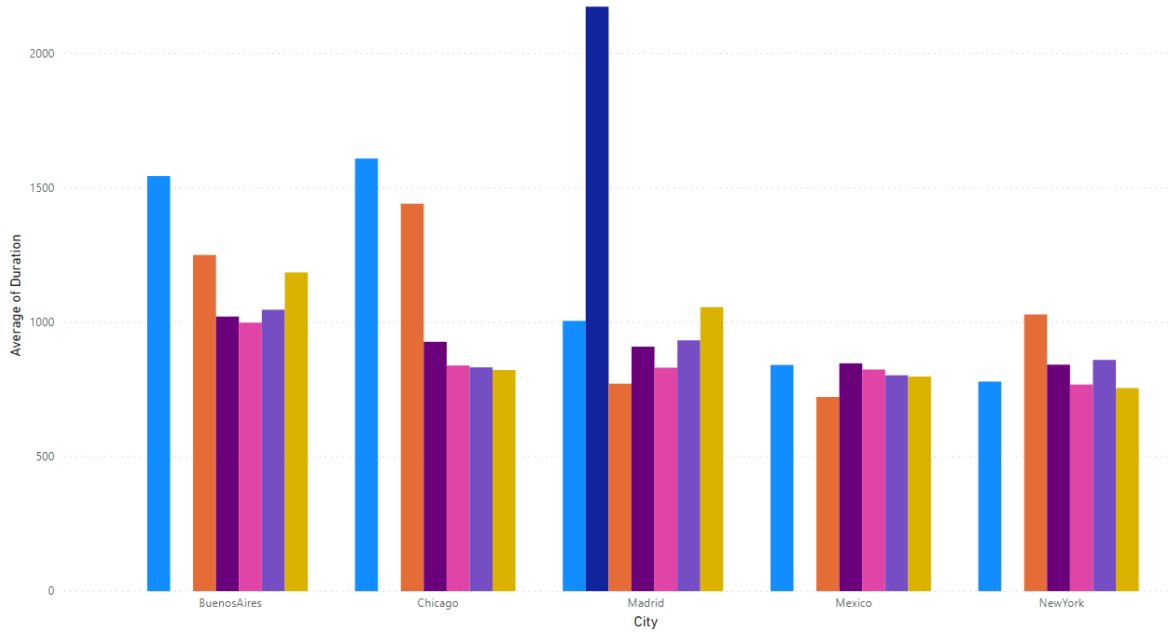
*Figure 6-F Duración media edad 2018*



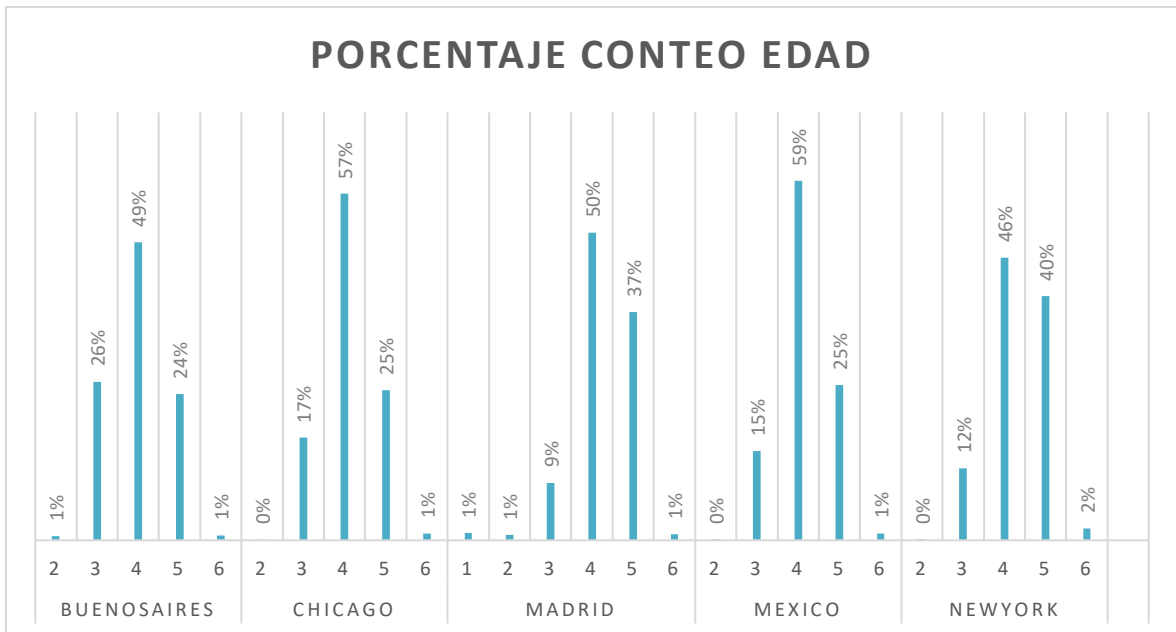
*Figure 6-G Porcentaje conteo edad 2018*

Average of Duration by City and Age

Age 0 1 2 3 4 5 6



*Figure 6-H Duración media edad 2019*



*Figure 6-I Porcentaje conteo edad 2019*

Al estudiar únicamente la duración media de los viajes se observan ciertos despuntes que podrían resultar interesantes, como por ejemplo las altas medias en edades de 0 a 16 años en Madrid o lo mismo para edades de 17 a 18 en Chicago. No obstante, estos resultados quedan parcialmente ofuscados al tener en cuenta el conteo de viajes. Si bien es cierto que los usuarios dentro de esos rangos de edad utilizan el servicio para trayectos más largos, dentro de la foto completa, el número de viajes que realizan frente al total es ínfimo.

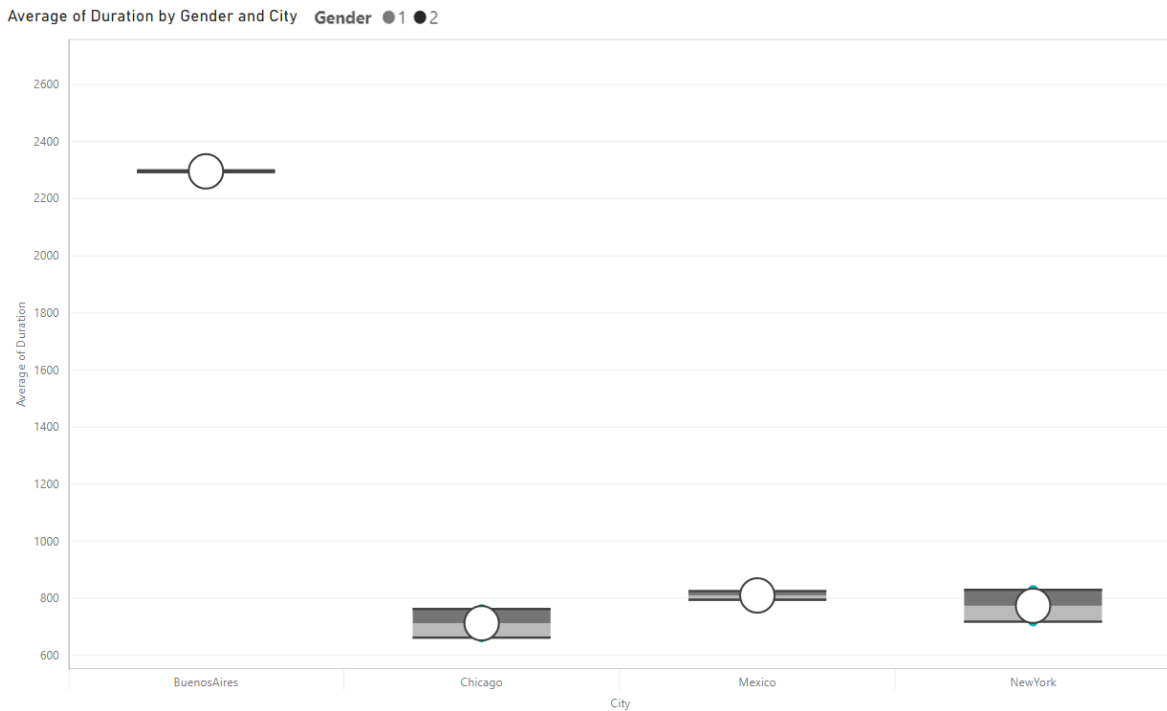
De hecho, el usuario predominante del servicio para todas las ciudades estudiadas tiene entre 27 y 40 años, seguido relativamente de cerca por aquel con edades entre 41 y 65.

Como datos de interés cabe también destacar que en Buenos Aires como excepción se da un uso bastante amplio del servicio por usuarios entre 19 y 26 años y la duración media en general es notablemente más elevada que en el resto de las ciudades. Una posible explicación de este fenómeno es que en Buenos Aires el servicio de bike sharing, Ecobici, es gratuito para todos los usuarios, lo cual parece propulsar a la población más joven al uso del mismo. Por esta razón, se extrae que los usuarios más jóvenes parecen no hacer un uso habitual del servicio más por cuestiones monetarias que logísticas.

### 6.3 ESTUDIO DE LA DURACIÓN POR GÉNEROS

Otro de los factores bajo estudio en este trabajo es el género del usuario que hace uso del servicio. Para este análisis, análogamente al caso anterior, no se tienen valores de genero para todas las ciudades, si no que únicamente Buenos Aires, Chicago, México y Nueva York los facilitan.

De una manera similar a los apartados anteriores, para realizar el estudio de este valor se realizan varias representaciones en función del año concreto bajo estudio. Por consiguiente, se tienen dos figuras en formato box plot que representan la duración media de los viajes según el género y el conteo de viajes total también según género, siendo la parte oscura la equivalente al género femenino y la clara al masculino y un gráfico de barras que hace el mismo conteo, pero por mes del año en lugar de por ciudad.



*Figure 6-J Duración media por genero 2017*

Count of Duration by Gender and City Gender ●1 ●2

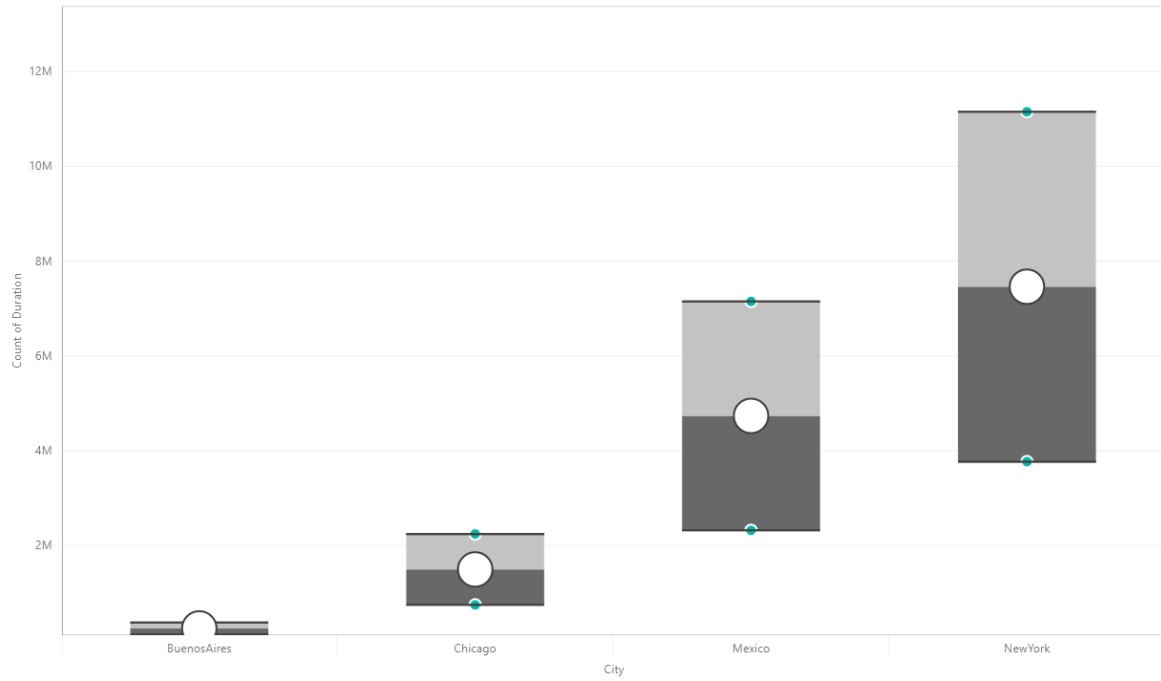


Figure 6-K Conteo genero por ciudad 2017

Count of Duration by Month and Gender

Gender ●1 ●2

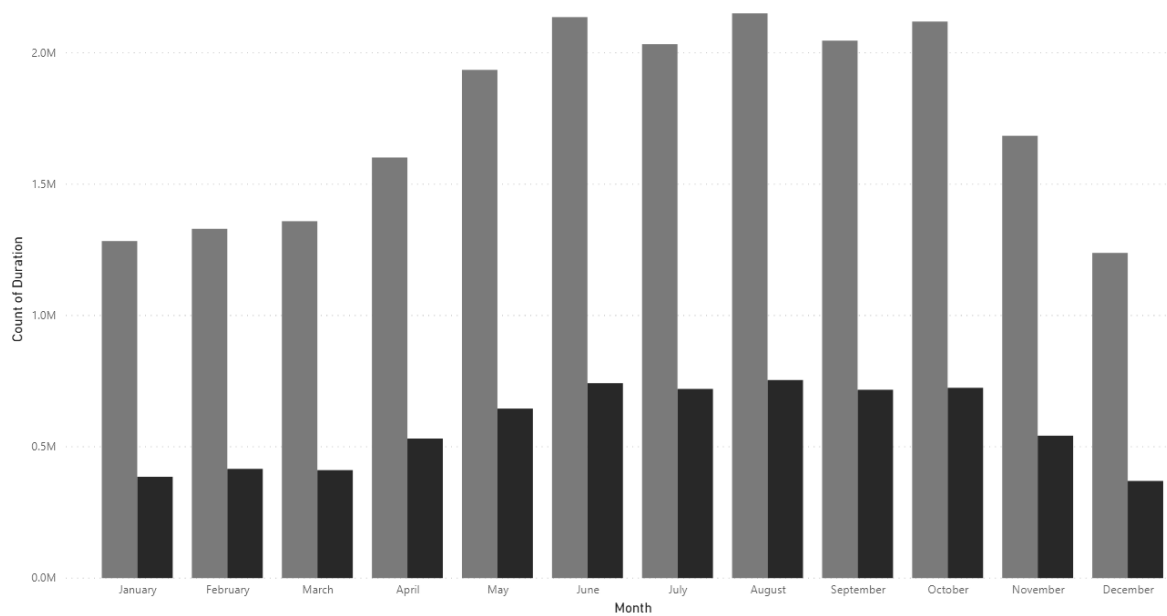
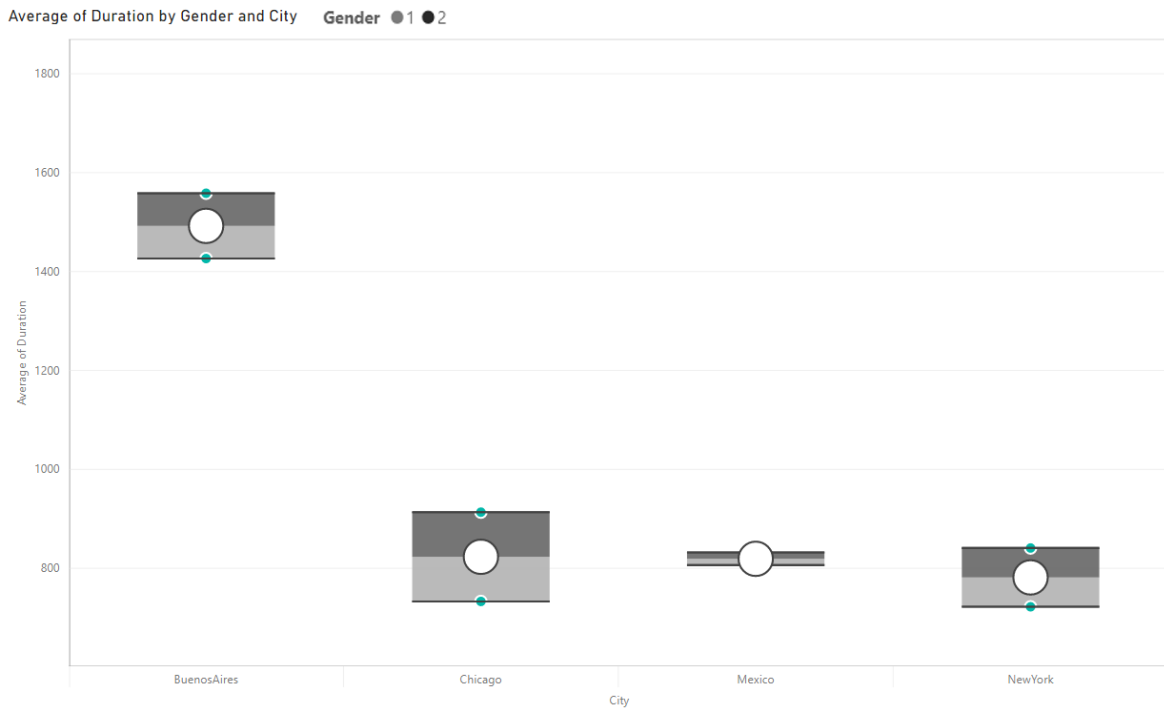
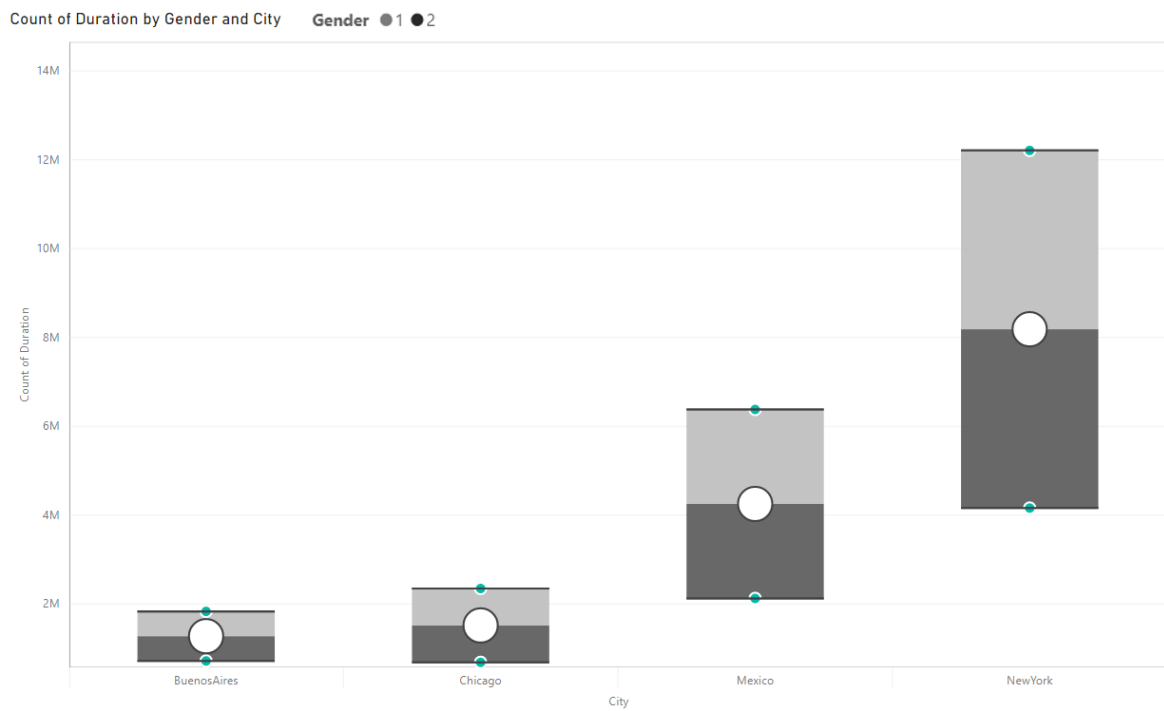


Figure 6-L Conteo genero por mes 2017

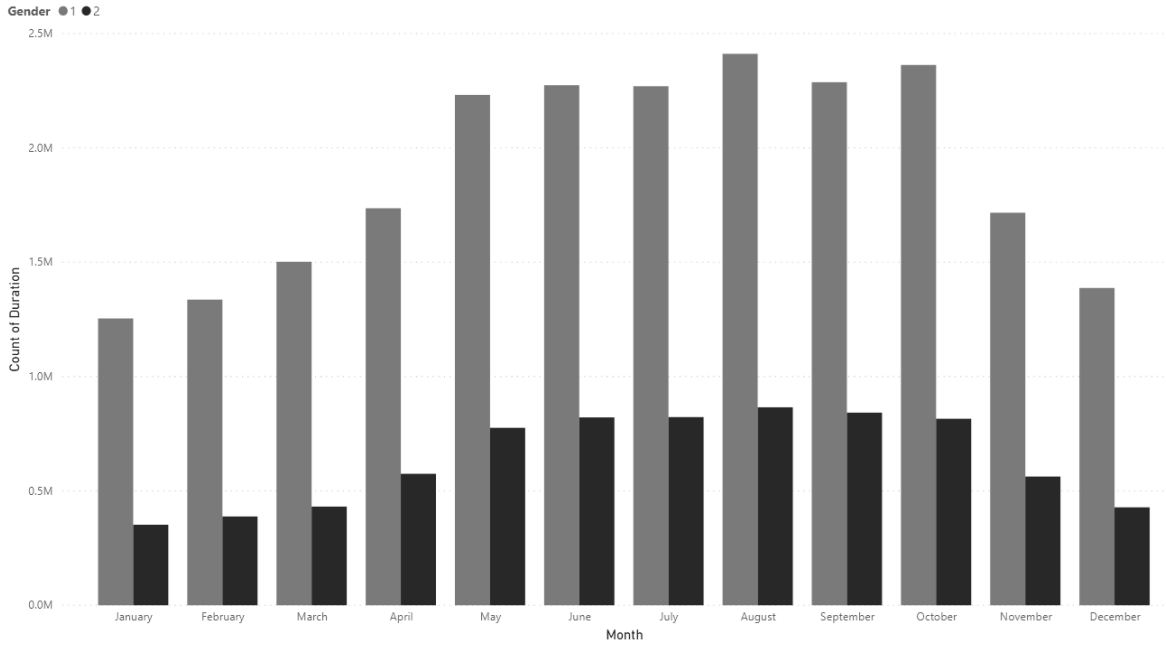


*Figure 6-M Duración media por genero 2018*



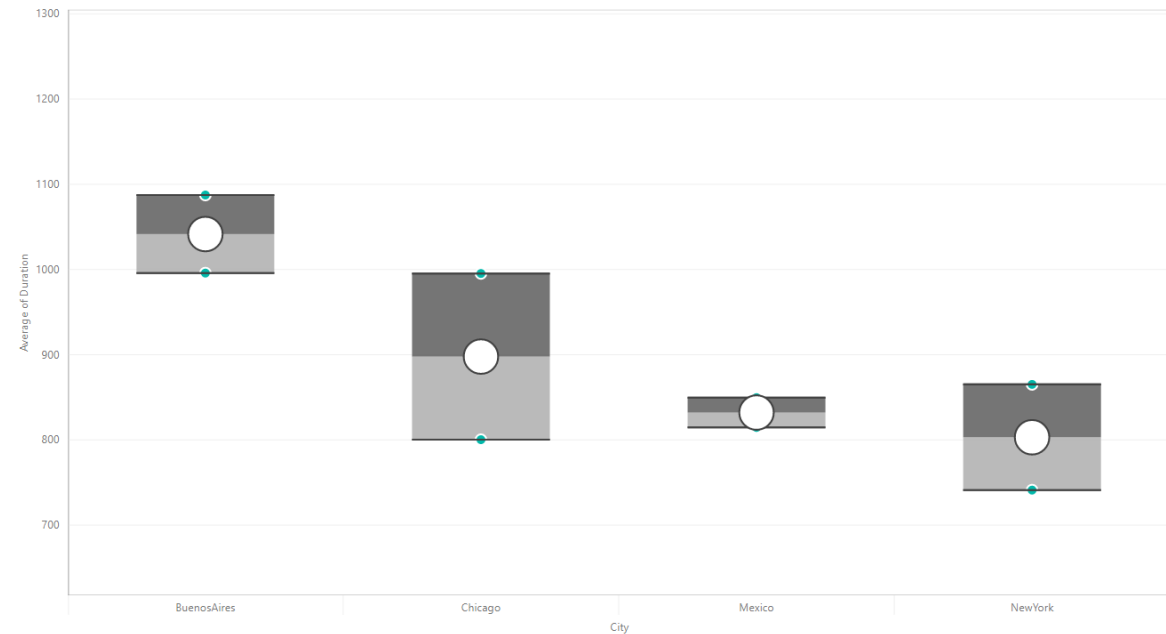
*Figure 6-N Conteo genero por ciudad 2018*

Count of Duration by Month and Gender



*Figure 6-O Conteo genero por mes 2018*

Average of Duration by Gender and City



*Figure 6-P Duración media por genero 2019*

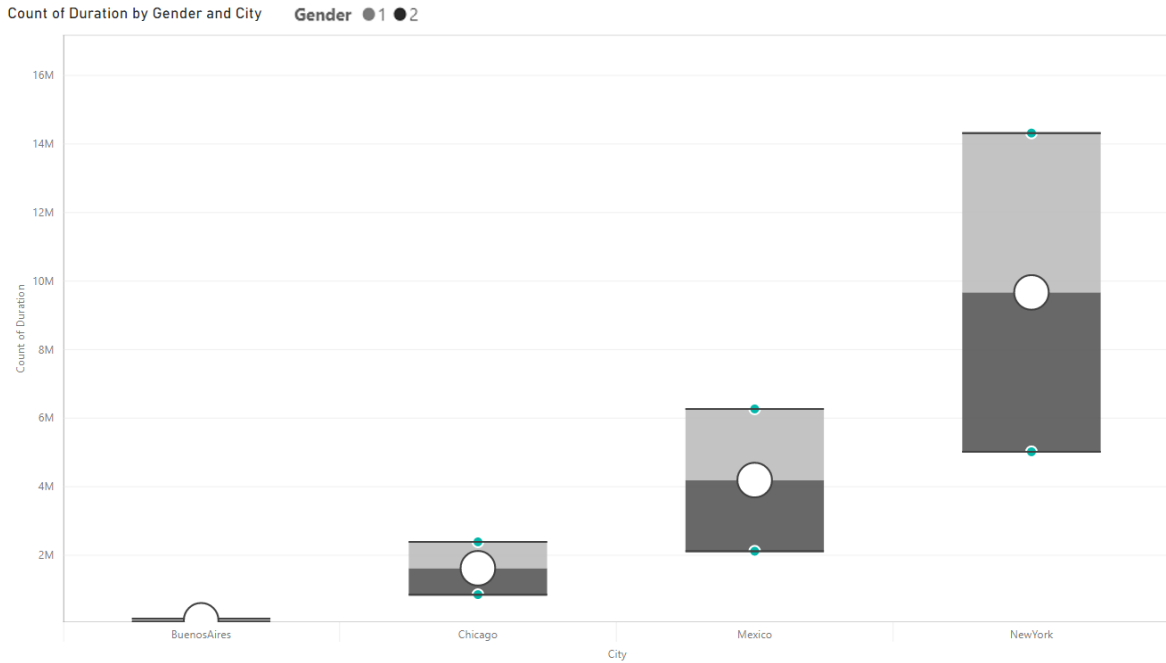


Figure 6-Q Conteo genero por ciudad 2019

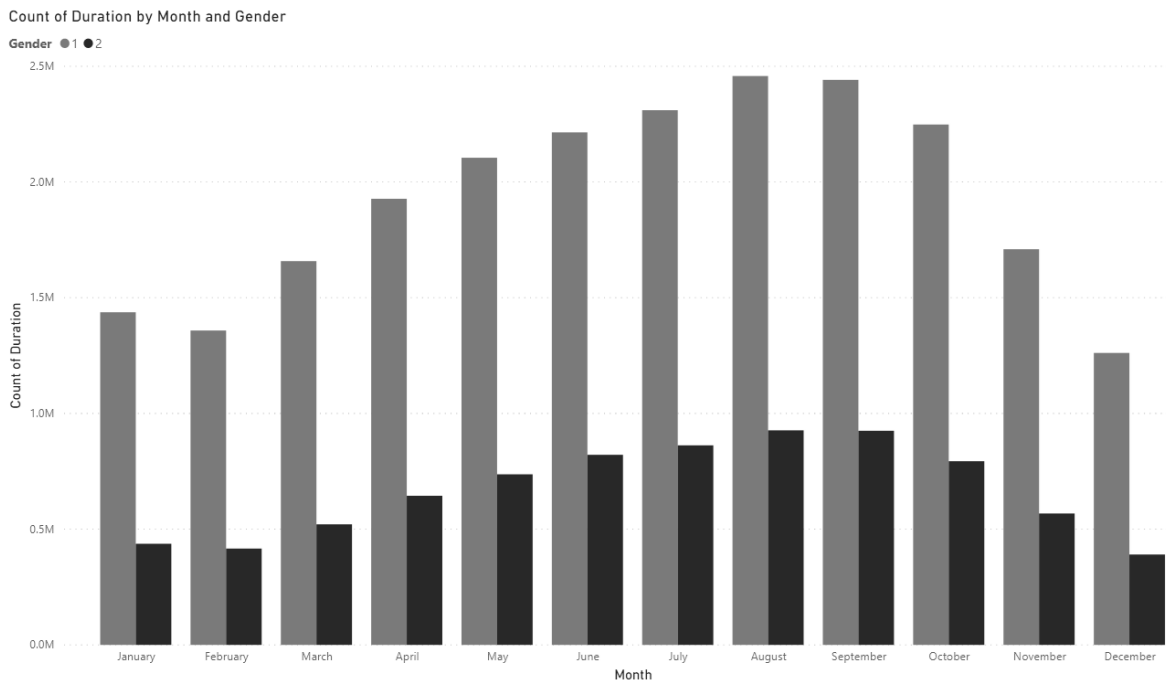


Figure 6-R Conteo genero por mes 2019

Tras una observación concienzuda de los resultados obtenidos, se puede llegar a la conclusión de que las variaciones anuales en función del género son muy bajas, por lo que esta sección se analiza directamente como un compendio para el cual las conclusiones extraídas son válidas independientemente del año.

Las conclusiones más claras a las que se llegan es que el servicio es principalmente utilizado por personas de género masculino, comprendiendo estas entre el 70% y el 80% del total de los viajes realizados en función de la época del año en la que se mida. Este resultado viene principalmente de los datos de Nueva York, que al tener un volumen notablemente más alto de viajes que el resto de las ciudades y una diferencia abismal entre el uso de las personas de género masculino frente a las de género femenino, genera esta diferenciación.

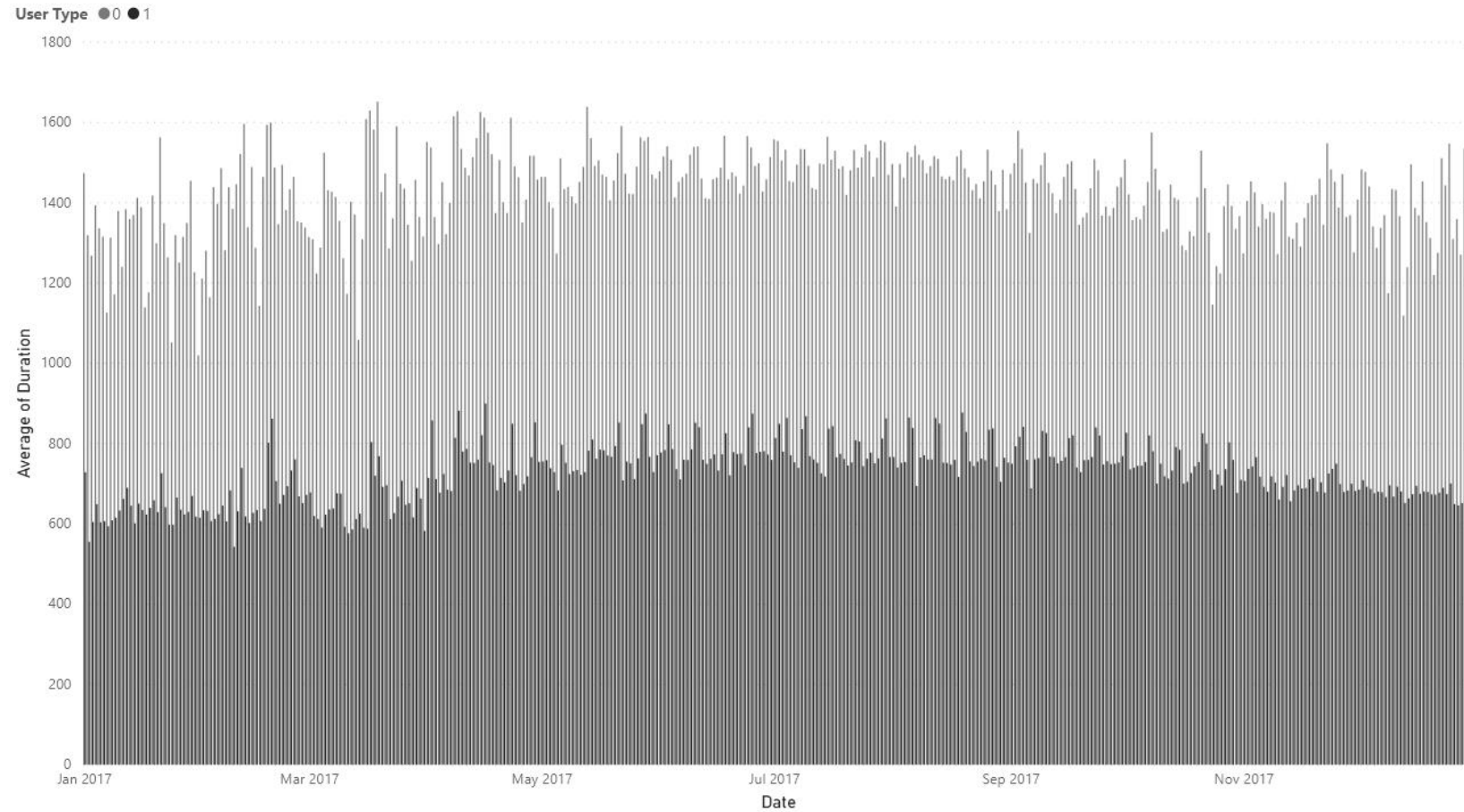
Por último, resulta interesante el análisis de las duraciones medias. En todas las ciudades la duración media de los viajes realizados por personas de género femenino es notablemente mayor que las de los realizados por personas de género masculino, habiendo hasta una diferencia superior a dos minutos en Chicago. La explicación de esta diferencia parece radicar en el estilo de conducción menos agresivo que suelen presentar las personas de género femenino.

## **6.4 ESTUDIO DE LA DURACIÓN POR TIPO DE USUARIO**

Otro de los parámetros bajo estudio que contiene la base de datos generada es el tipo de usuario que utiliza el servicio. Esto quiere decir si la persona se encuentra registrada en el sistema de alquileres o no. Sobre este dato, solo Chicago, Madrid, Nueva York y Buenos Aires para el año 2018 facilitan valores.

A la hora de hacer en análisis en esta sección, se hace una representación por día de las duraciones medias en función del tipo de usuario para las ciudades descritas anteriormente, con el fin de tratar de encontrar algún patrón representativo. Por otro lado, se hace también un estudio individual por ciudad de la duración media por tipo de usuario, para el que se utilizan box plots en los que la parte oscura corresponde a los usuarios registrados. Finalmente, se realiza un conteo de los viajes realizados por usuarios registrados en días laborables frente a no laborables.

Average of Duration by Date and User Type



*Figure 6-S Duración media por tipo de usuario 2017*

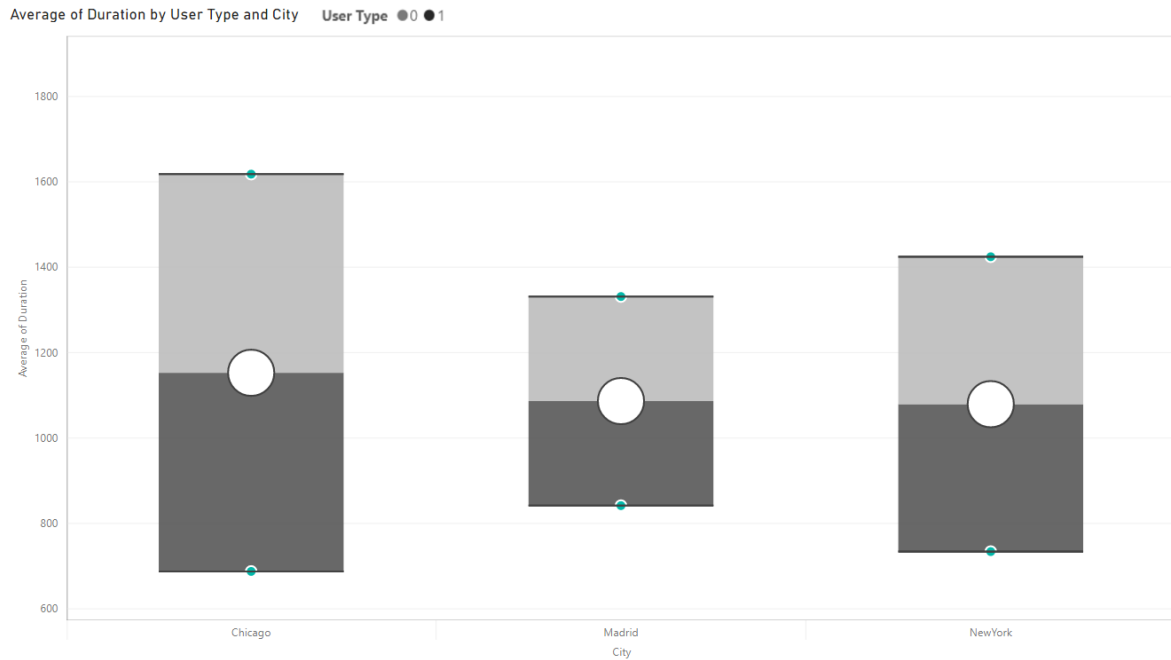


Figure 6-T Duración media por tipo de usuario y ciudad 2017

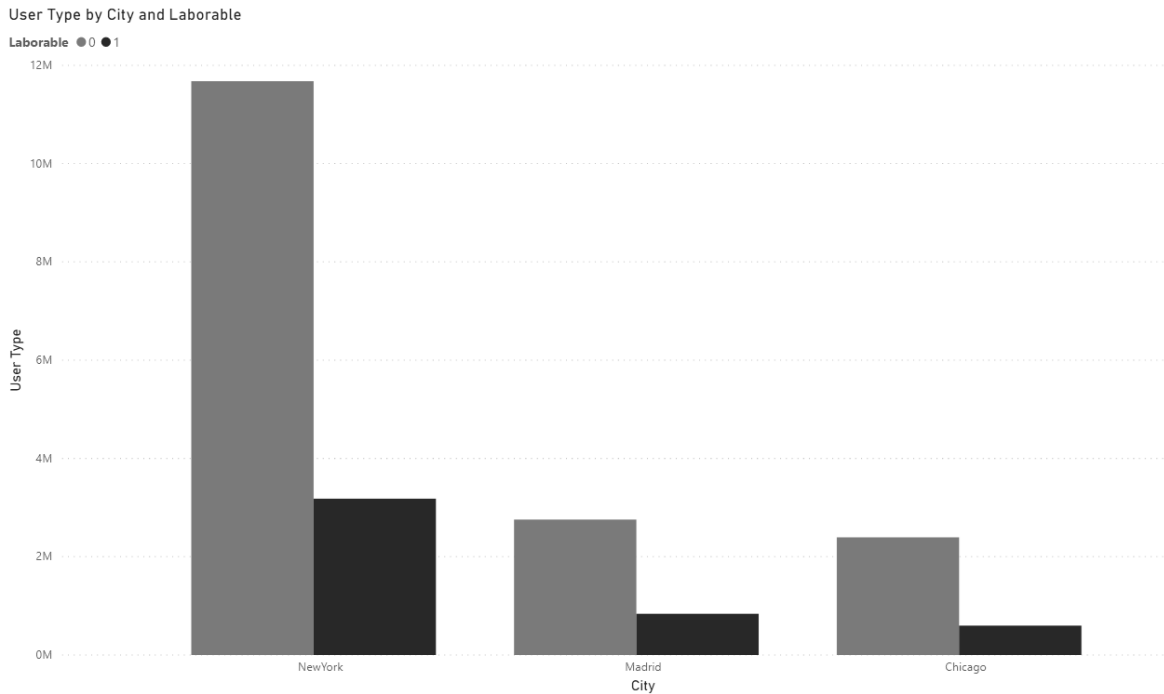
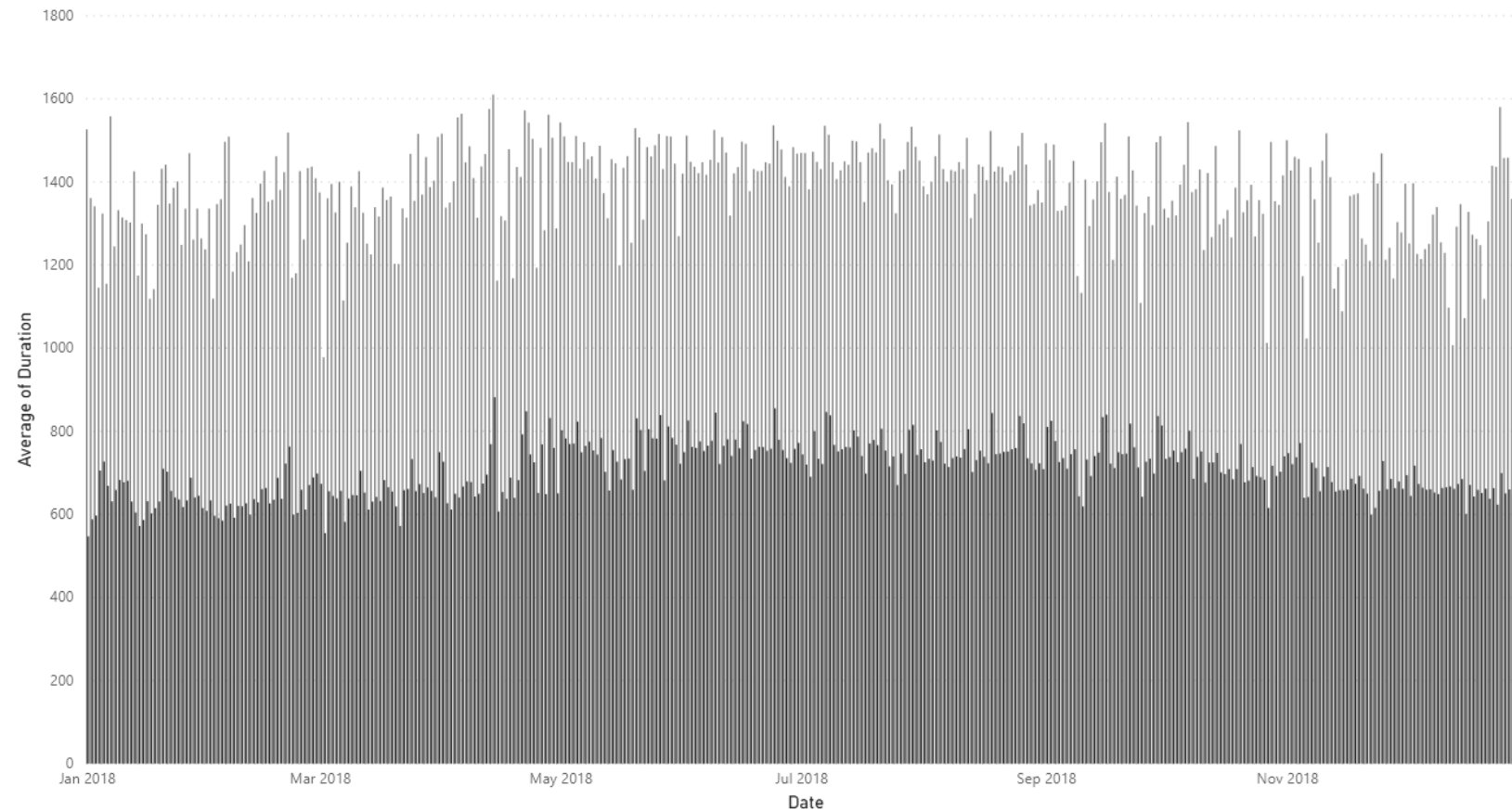


Figure 6-U Conteo registrados en laborable vs no laborable 2017

Average of Duration by Date and User Type

User Type ● 0 ● 1



*Figure 6-V Duración media por tipo de usuario 2018*

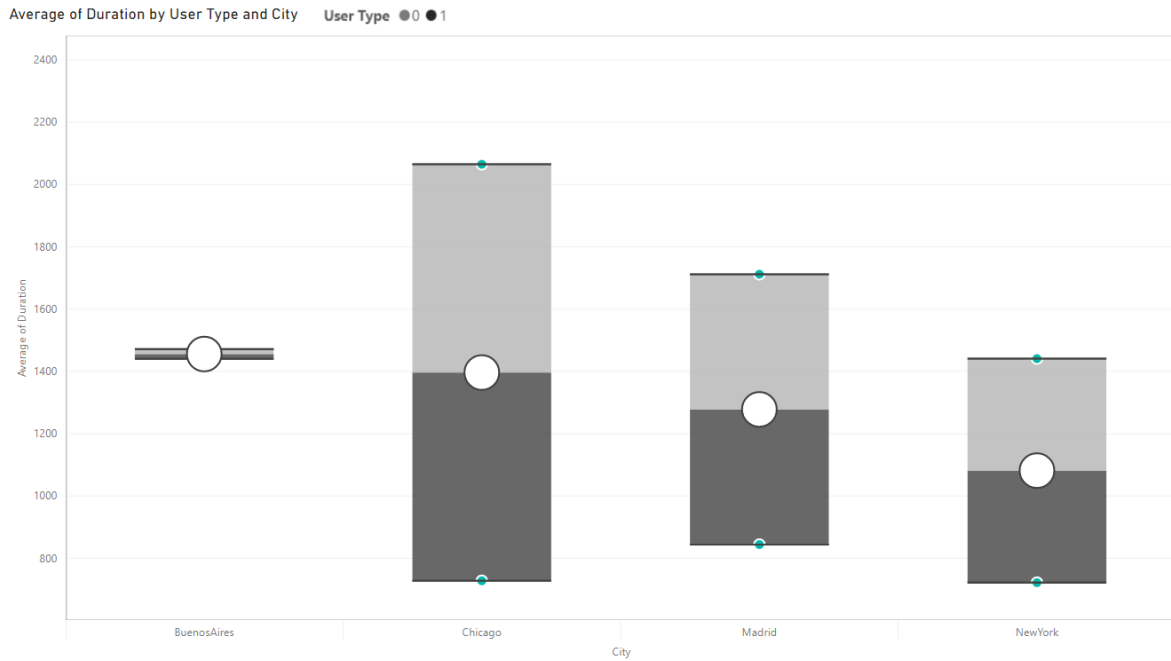


Figure 6-W Duración media por tipo de usuario y ciudad 2018

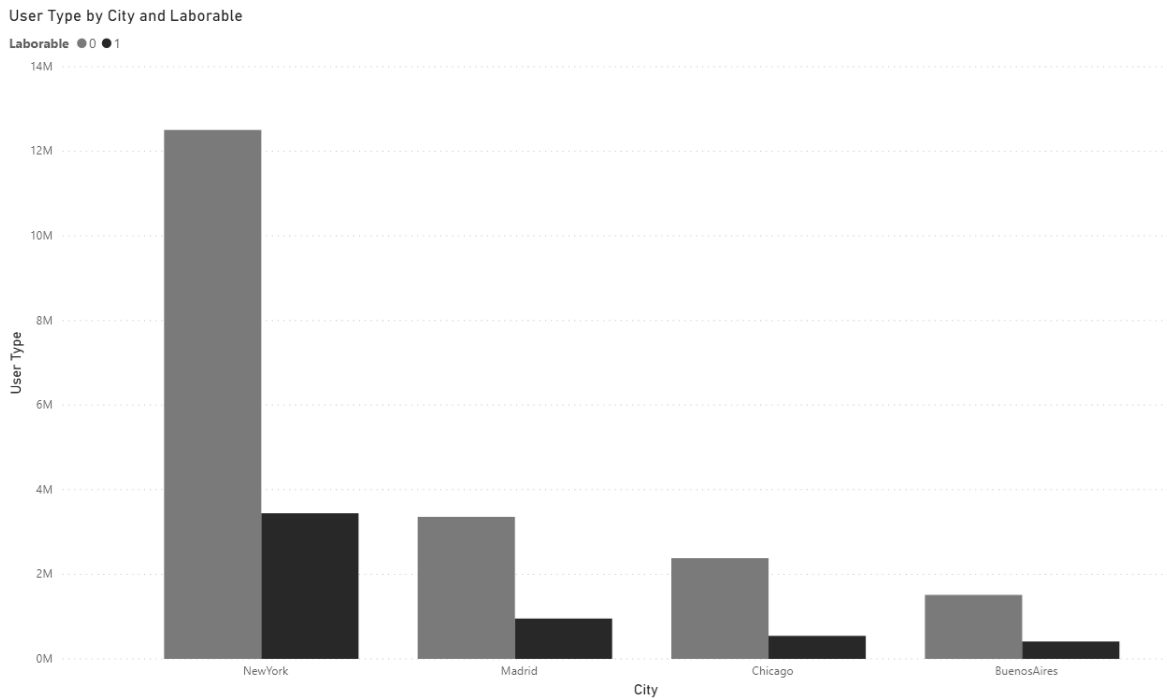


Figure 6-X Conteo registrados en laborable vs no laborable 2018

Average of Duration by Date and User Type

User Type ● 0 ● 1

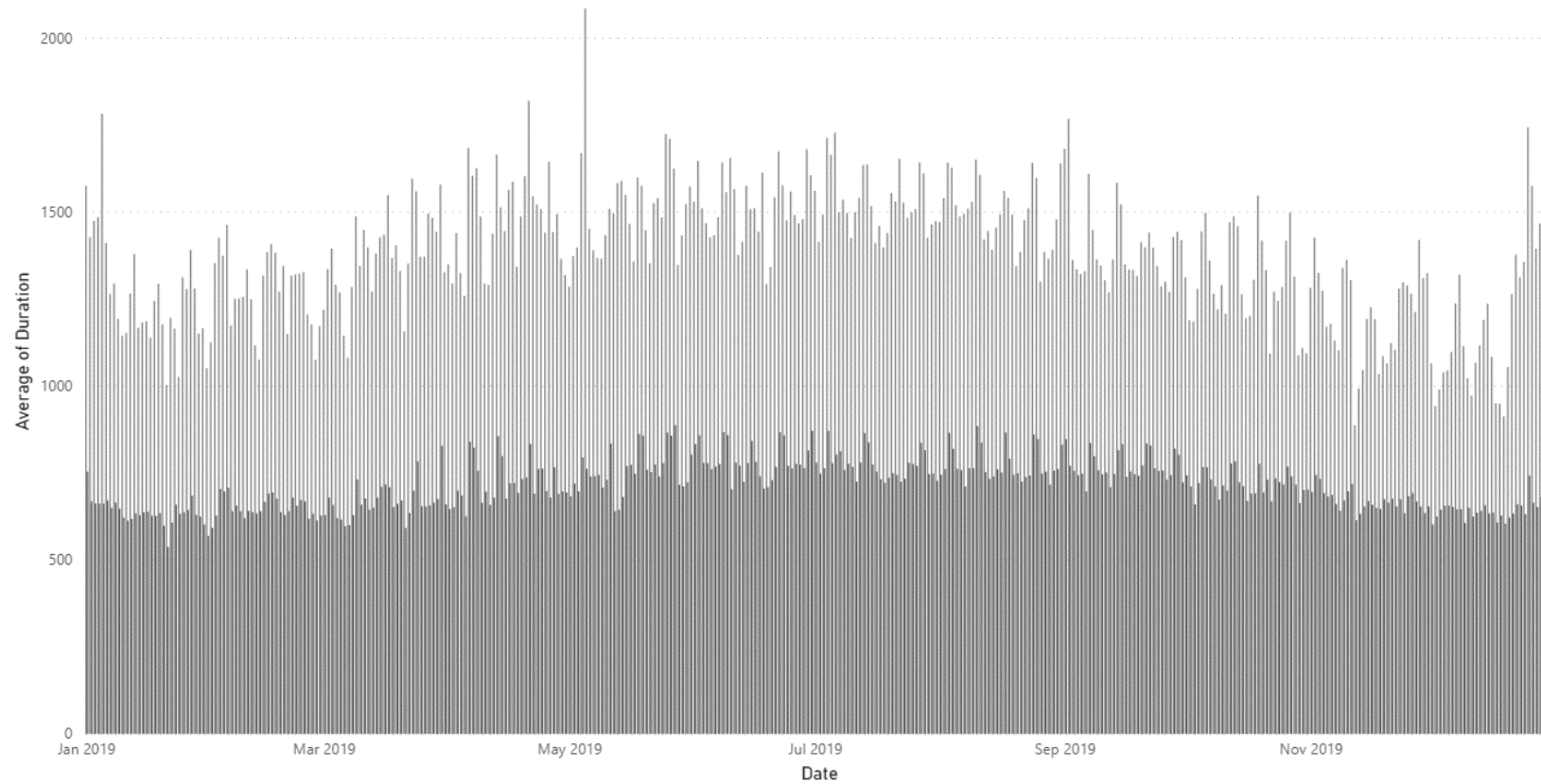
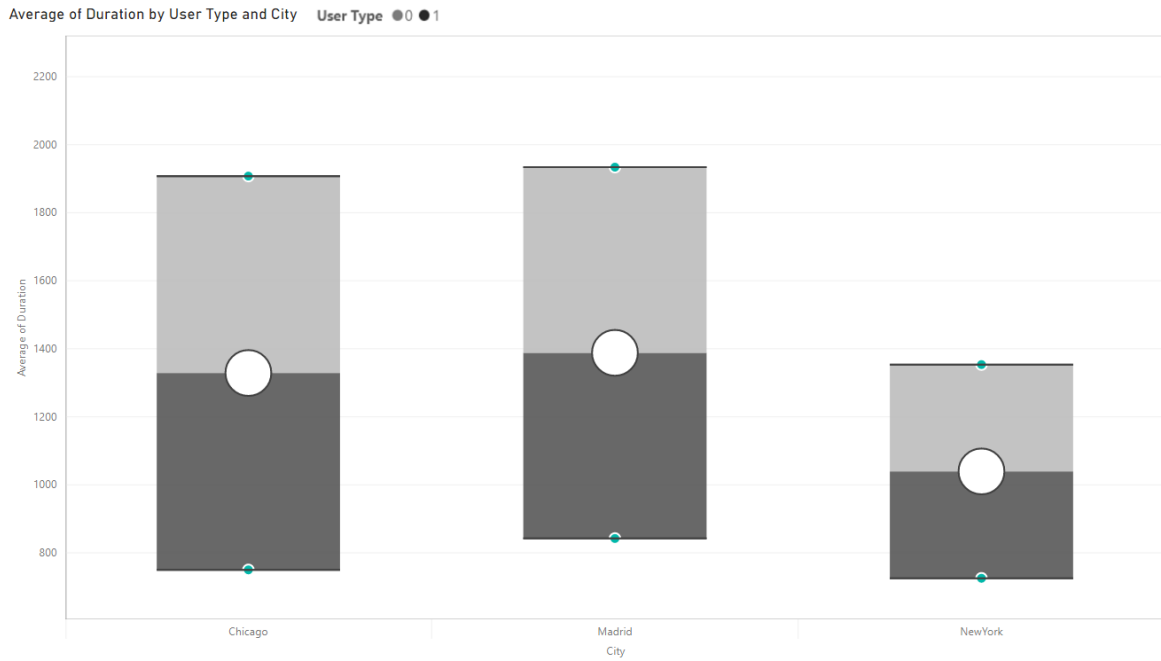
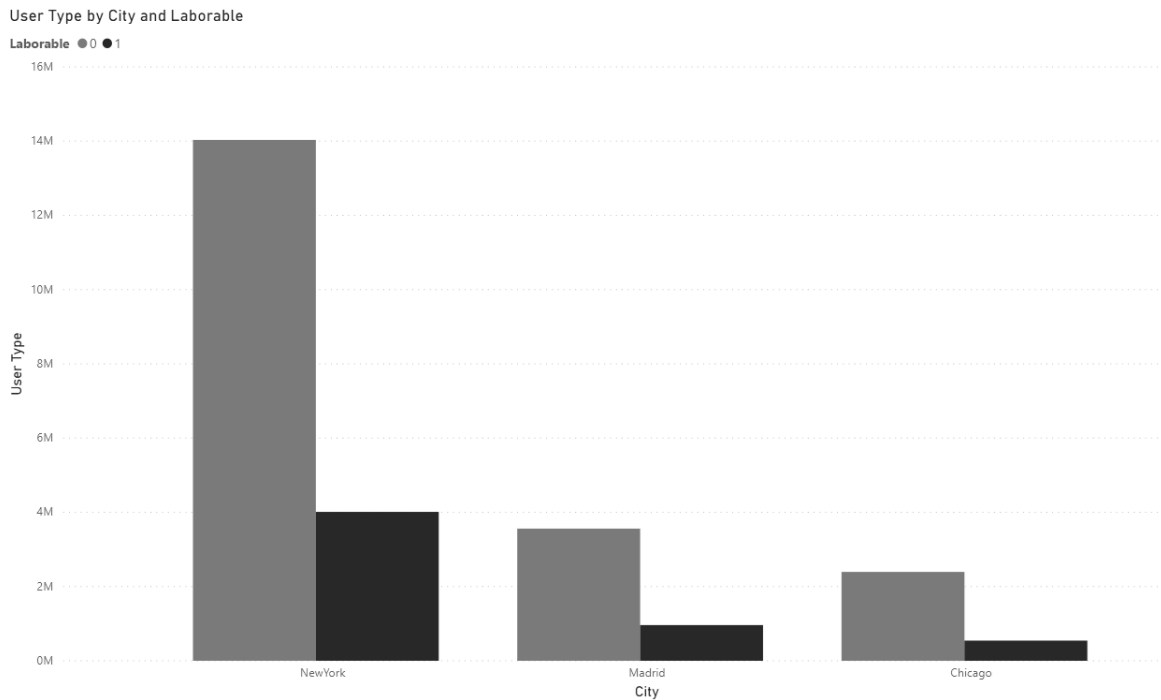


Figure 6-Y Duración media por tipo de usuario 2019



*Figure 6-Z Duración media por tipo de usuario y ciudad 2019*



*Figure 6-AA Conteo registrados en laborable vs no laborable 2019*

Los valores obtenidos en cuanto a la duración media según el tipo de usuario se mantienen bastante homogéneos a lo largo de los años estudiados, viéndose claramente una diferenciación entre las duraciones, para los usuarios registrados la duración media por viaje se encuentra en los 600 segundos, mientras que para los usuarios no registrados tenemos un valor próximo al doble, 1200 segundos. Esta reflexión concuerda con los valores medios obtenidos al hacer el análisis individualmente por ciudades. Para todos los casos los usuarios registrados presentan una media de duración de viaje notablemente inferior a la de los usuarios no registrados.

Además de los resultados anteriores, merece la pena el estudio de la tercera grafica para cada año, en la que se hace una representación de la distribución entre los viajes realizados por los usuarios registrados en días laborables y no laborables. Los números resultan homogéneos en todos los casos y muestran que el 75% de los viajes realizados por usuarios registrados se centran en días laborables, lo cual lleva a la conclusión de que los usuarios habituales del servicio lo contratan para desplazarse mayormente a su puesto de trabajo o estudio.

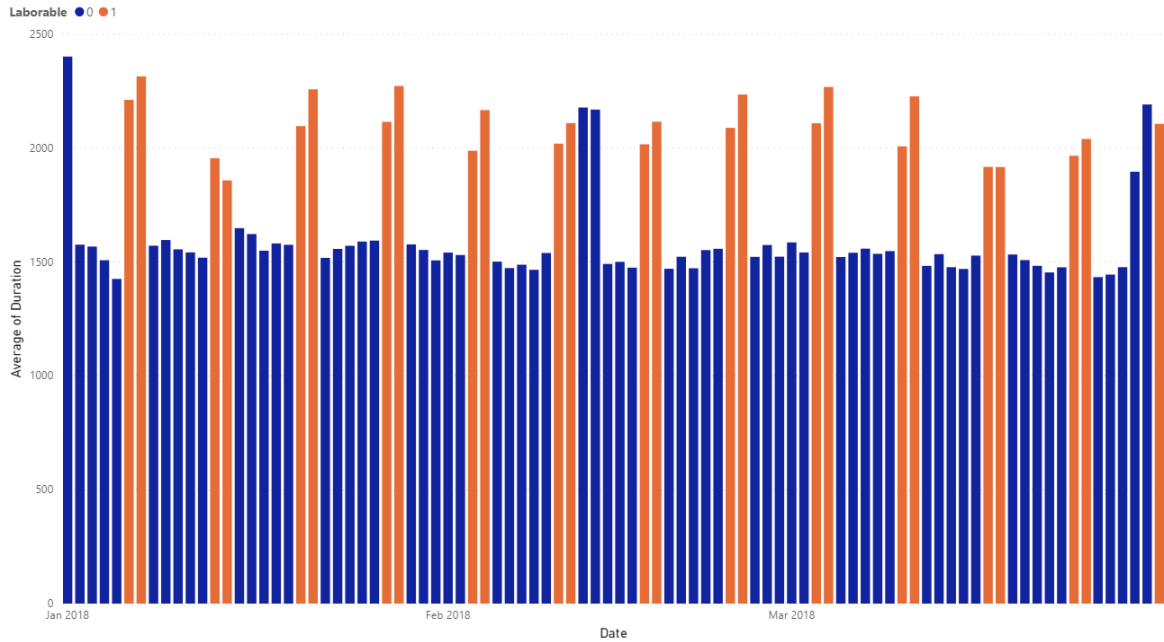
## **6.5 ESTUDIO DE LA DURACIÓN SEGÚN EL DÍA DE LA SEMANA**

En este apartado se estudia una comparación frente a, por un lado, las duraciones medias en función del día del año, haciendo hincapié en la diferenciación entre día laborable y día no laborable y, por otro lado, el conteo de viajes realizados igualmente en cada día del año de nuevo diferenciando especialmente días laborables de no laborables.

De manera análoga a los apartados anteriores, se estudian los años 2017, 2018 y 2019, y para realizar la muestra de los datos se hace una división en subapartados primero separando por el año correspondiente a la realización del viaje y segundo por media de duraciones o conteo total. Además, para una correcta visualización de los datos, se ha decidido también hacer un corte en la representación de tres meses, de tal manera que, para cada uno de los casos en cada año, se muestran cuatro figuras, la primera correspondiente a las fechas de enero a marzo, la segunda para abril a junio, la tercera para julio a septiembre y la última para octubre a diciembre.

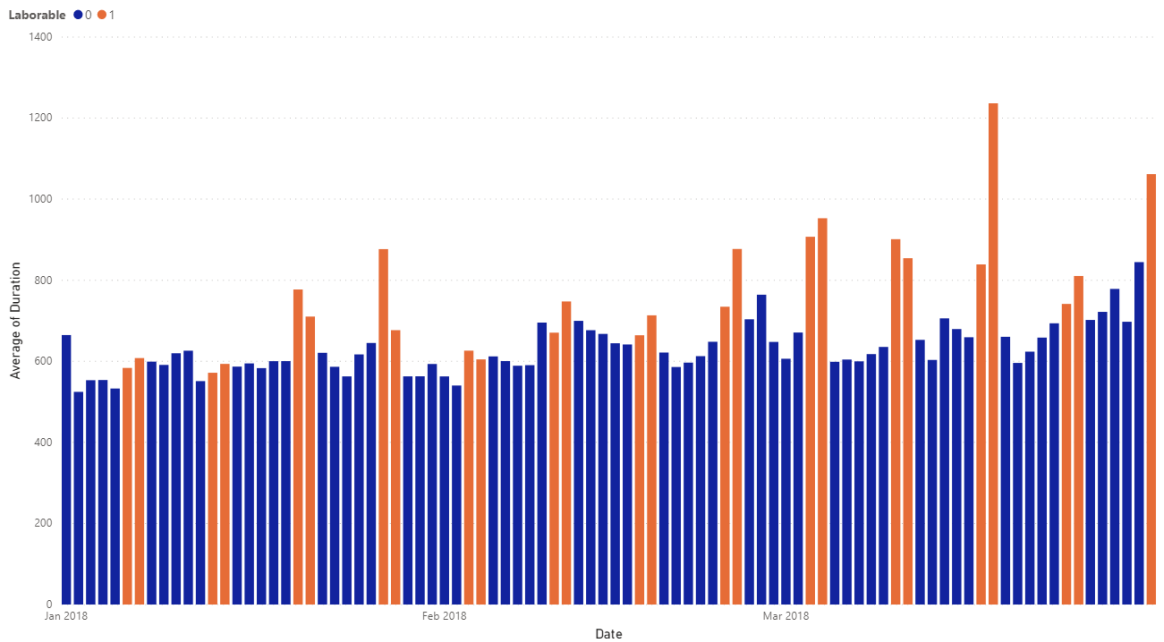
Estas representaciones se realizan en un conjunto que es la media total de todas las ciudades bajo estudio. Esta decisión se ha tomado en base a que tras estudiar individualmente cada una de las ciudades se llega a la conclusión de que los resultados son tan similares que no tiene sentido hacer la distinción, no obstante, se presentan las siguientes figuras como demostración de lo aquí indicado.

Average of Duration by Date and Laborable



*Figure 6-BB Duración media Buenos Aires 2018*

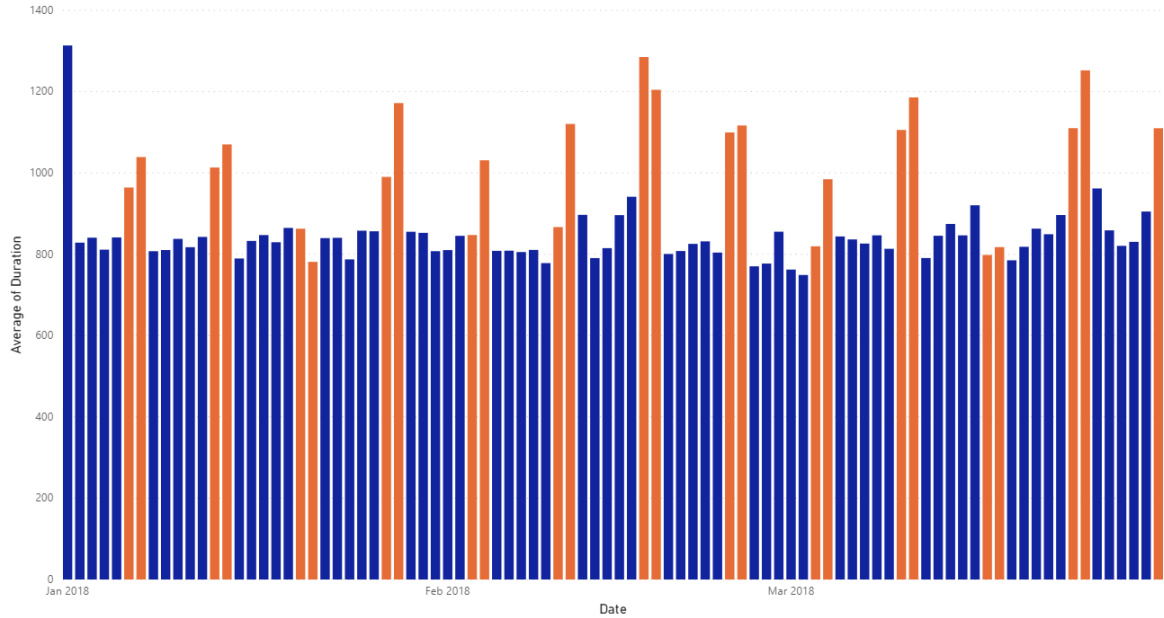
Average of Duration by Date and Laborable



*Figure 6-CC Duración media Chicago 2018*

Average of Duration by Date and Laborable

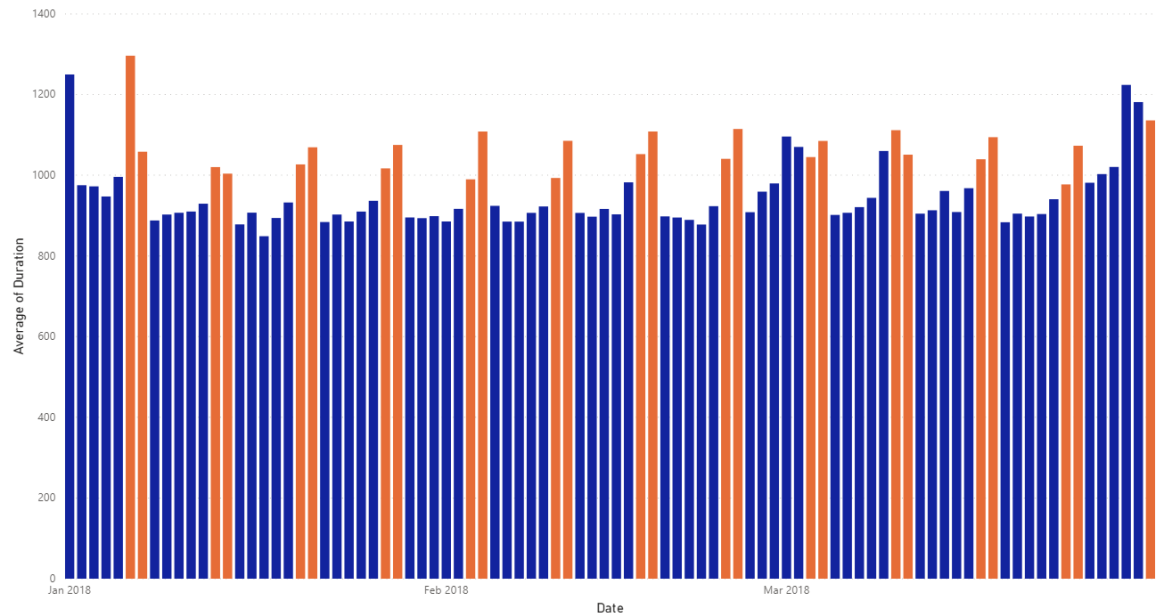
Laborable ● 0 ● 1



*Figure 6-DD Duración media Londres 2018*

Average of Duration by Date and Laborable

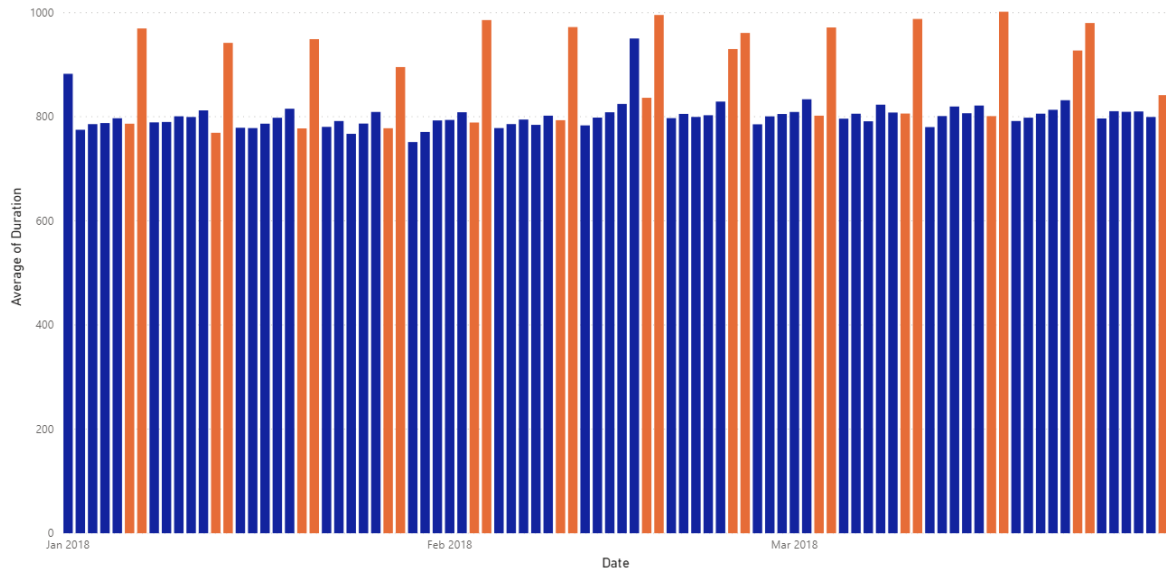
Laborable ● 0 ● 1



*Figure 6-EE Duración media Madrid 2018*

Average of Duration by Date and Laborable

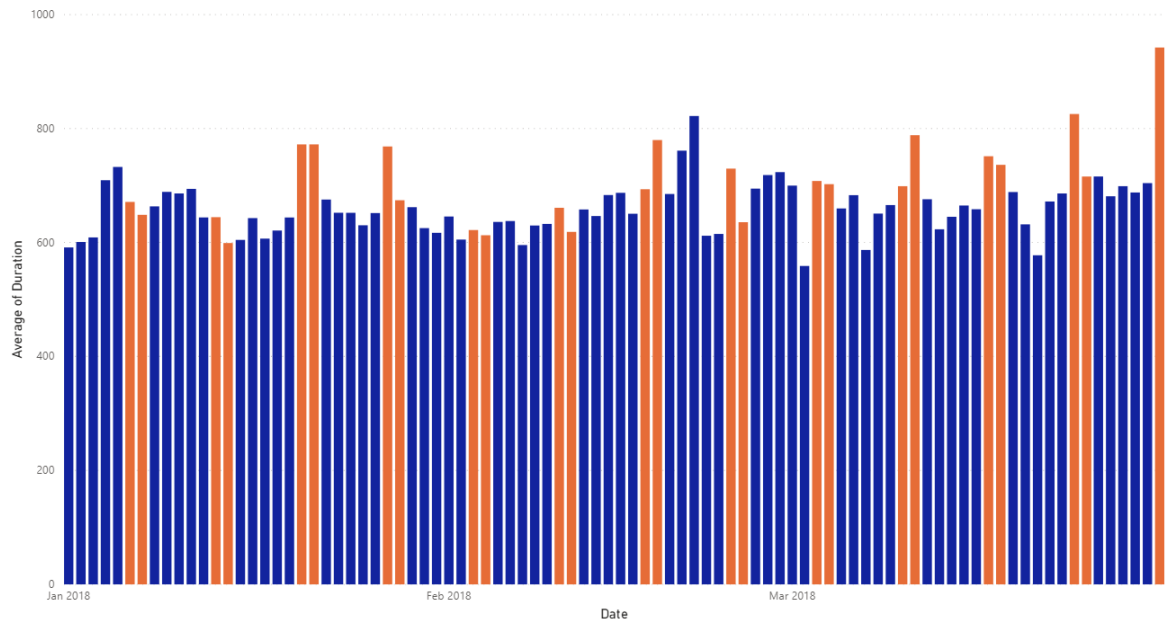
Laborable ● 0 ● 1



*Figure 6-FF Duración media México 2018*

Average of Duration by Date and Laborable

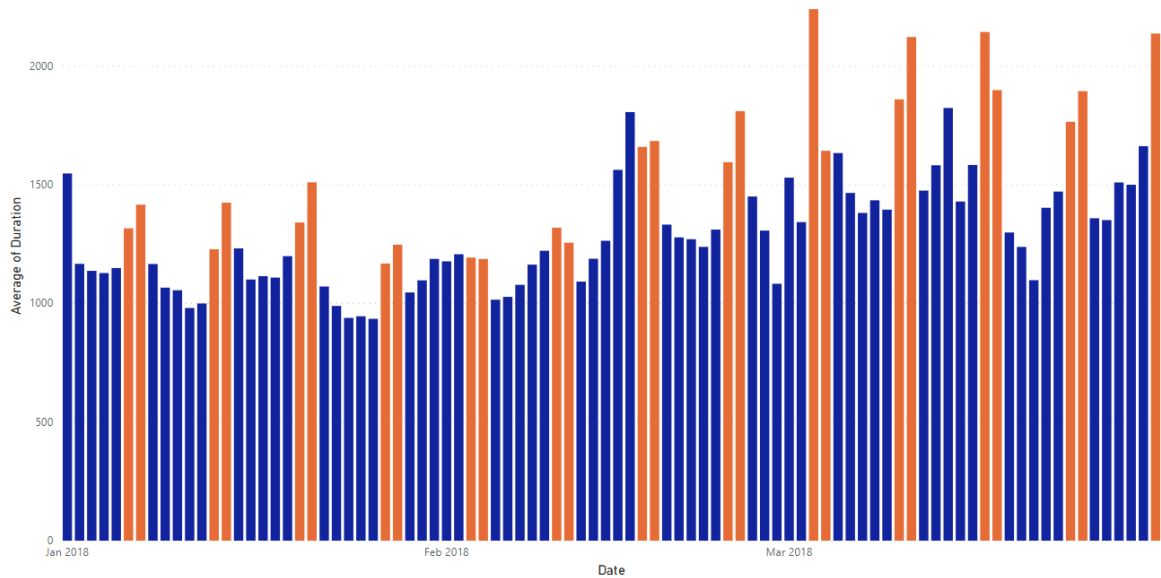
Laborable ● 0 ● 1



*Figure 6-GG Duración media Nueva York 2018*

Average of Duration by Date and Laborable

Laborable ● 0 ● 1



*Figure 6-HH Duración media Seúl 2018*

Como se puede observar en las figuras siguientes, las conclusiones obtenidas al realizar el estudio individualmente por ciudad son extrapolables al realizado con el compendio de las siete, por lo que es este último el que se analiza a continuación.

### 6.5.1 DURACIONES MEDIAS 2017

Average of Duration by Date and Laborable

Laborable ● (Blank) ● 0 ● 1

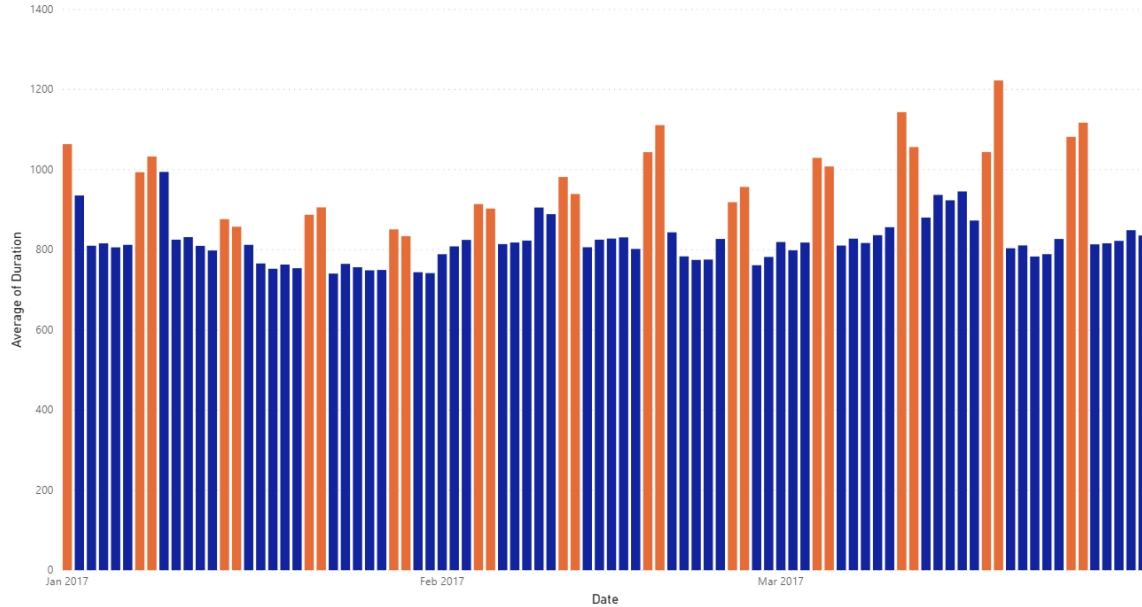


Figure 6-II Media Enero-Marzo 2017 Dia

Average of Duration by Date and Laborable

Laborable ● 0 ● 1

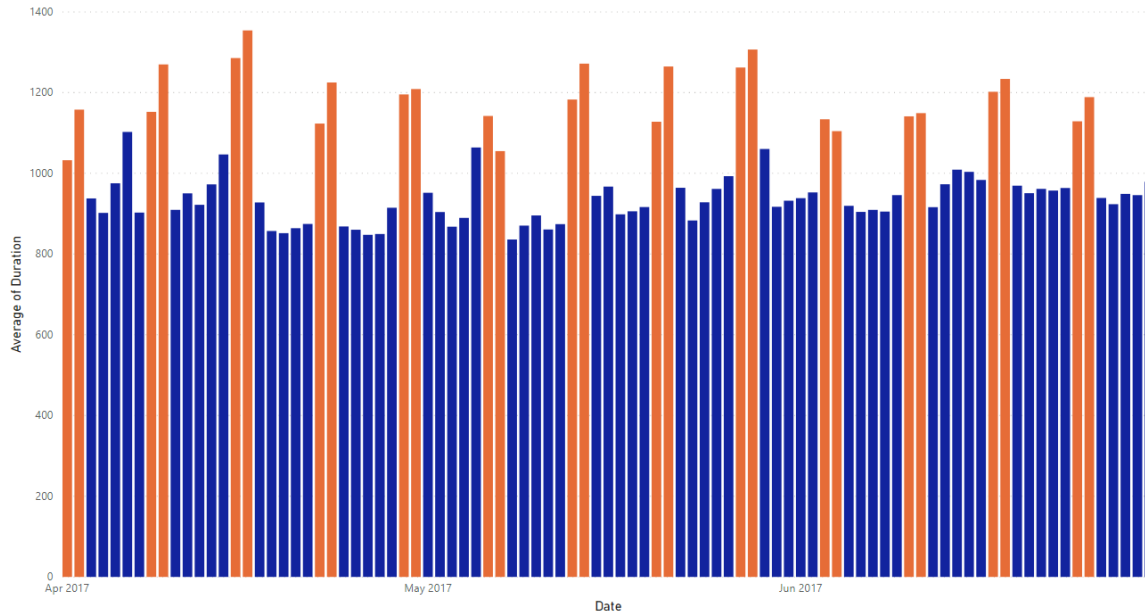
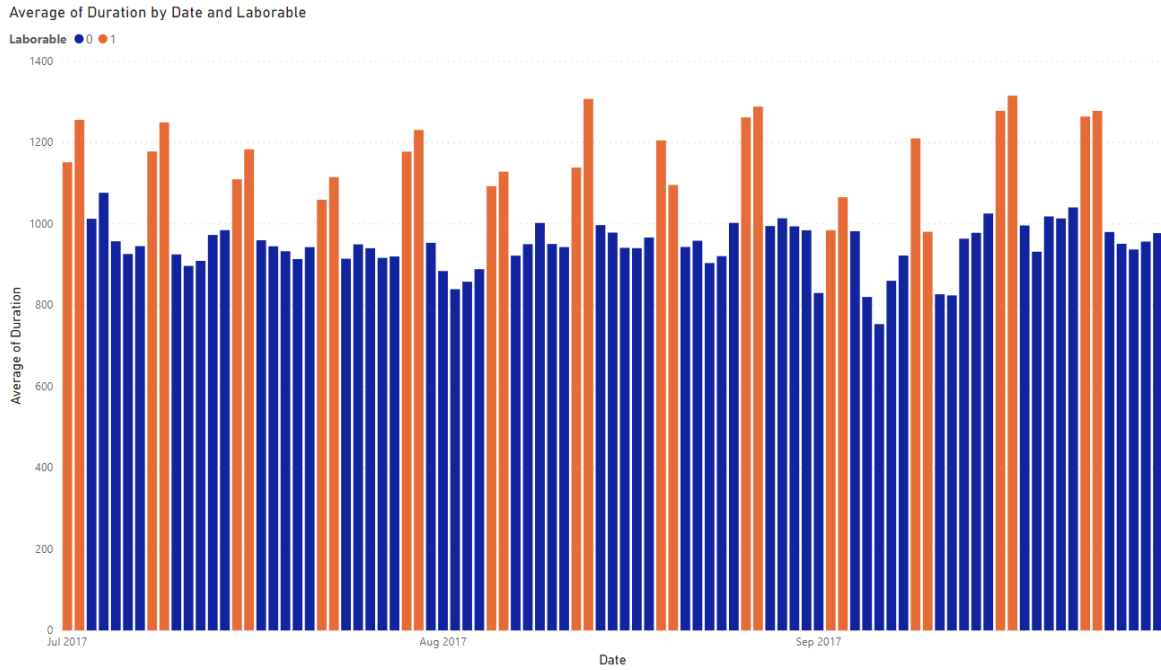
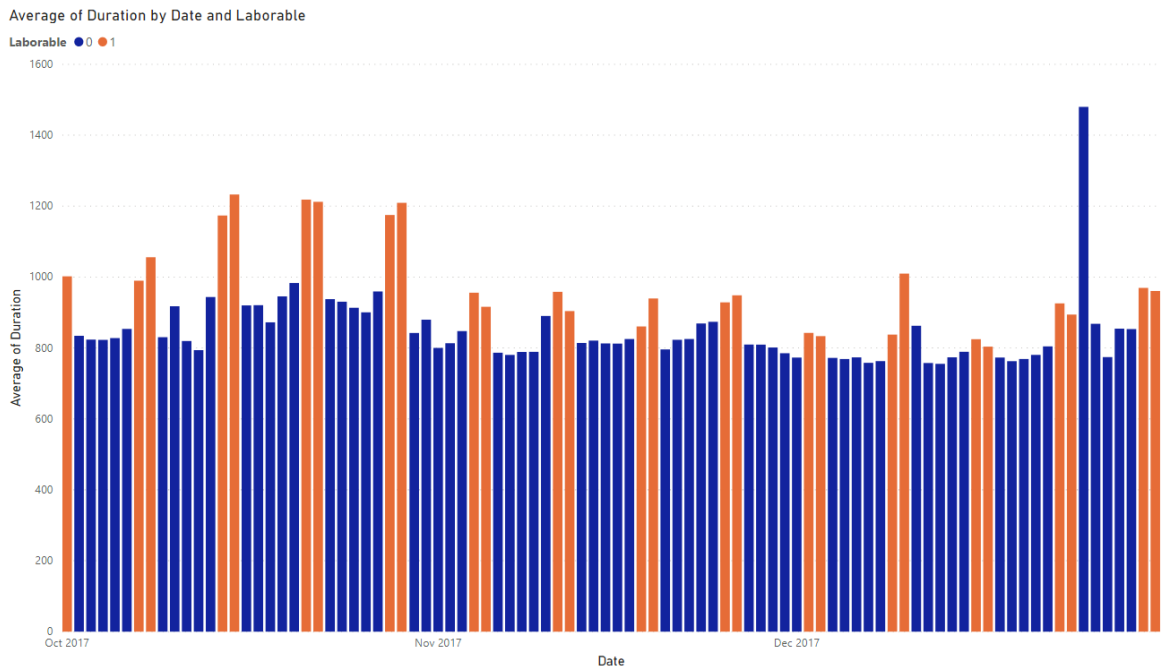


Figure 6-JJ Media Abril-Junio 2017 Dia



*Figure 6-KK Media Julio-Septiembre 2017 Dia*



*Figure 6-LL Media Octubre-Diciembre 2017 Dia*

## 6.5.2 CONTEO DE VIAJES 2017

Count of Duration by Date and Laborable

Laborable (Blank) ● 0 ● 1

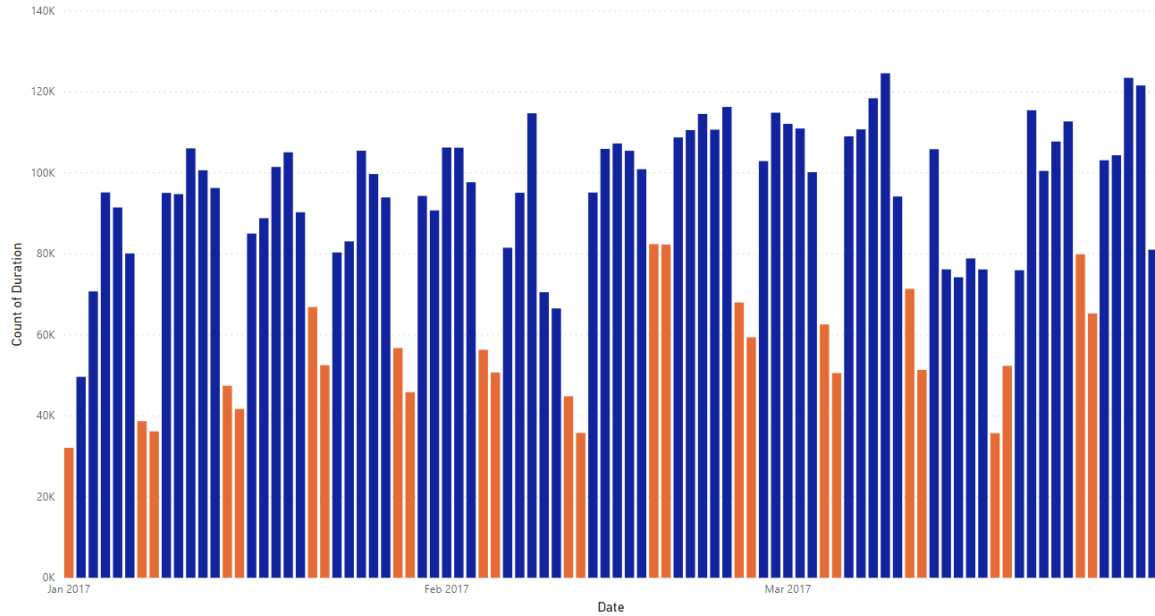


Figure 6-MM Conteo Enero-Marzo 2017 Dia

Count of Duration by Date and Laborable

Laborable ● 0 ● 1

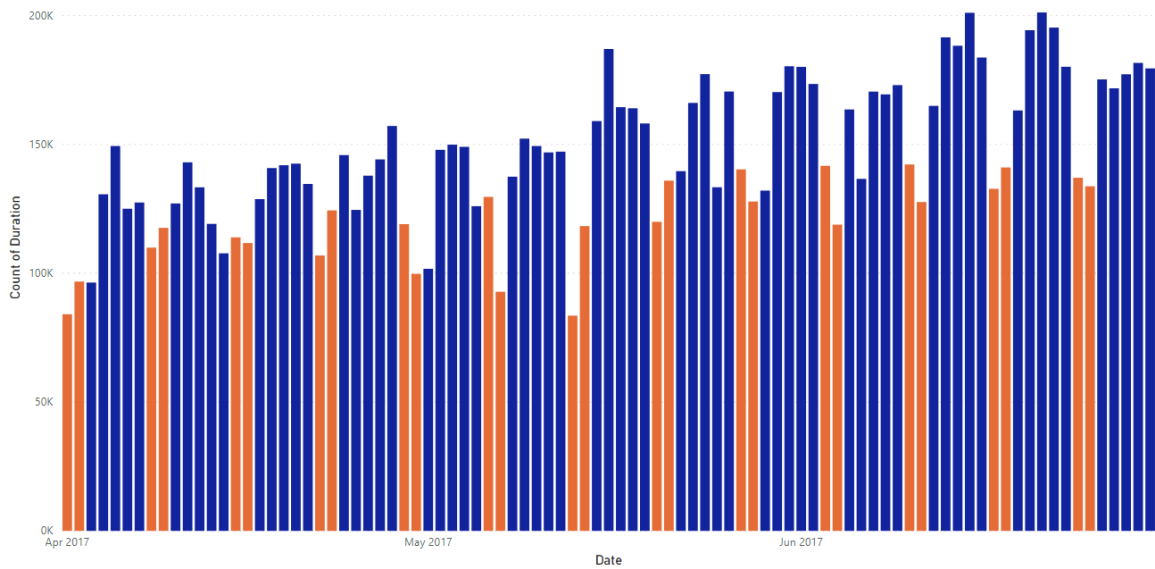
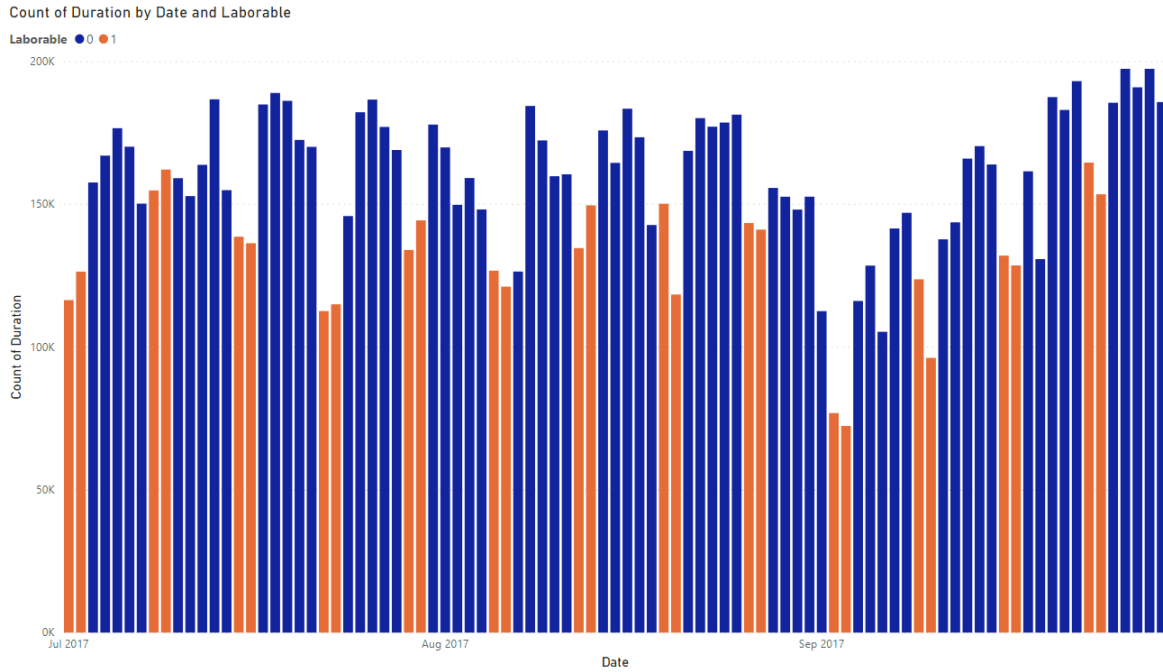
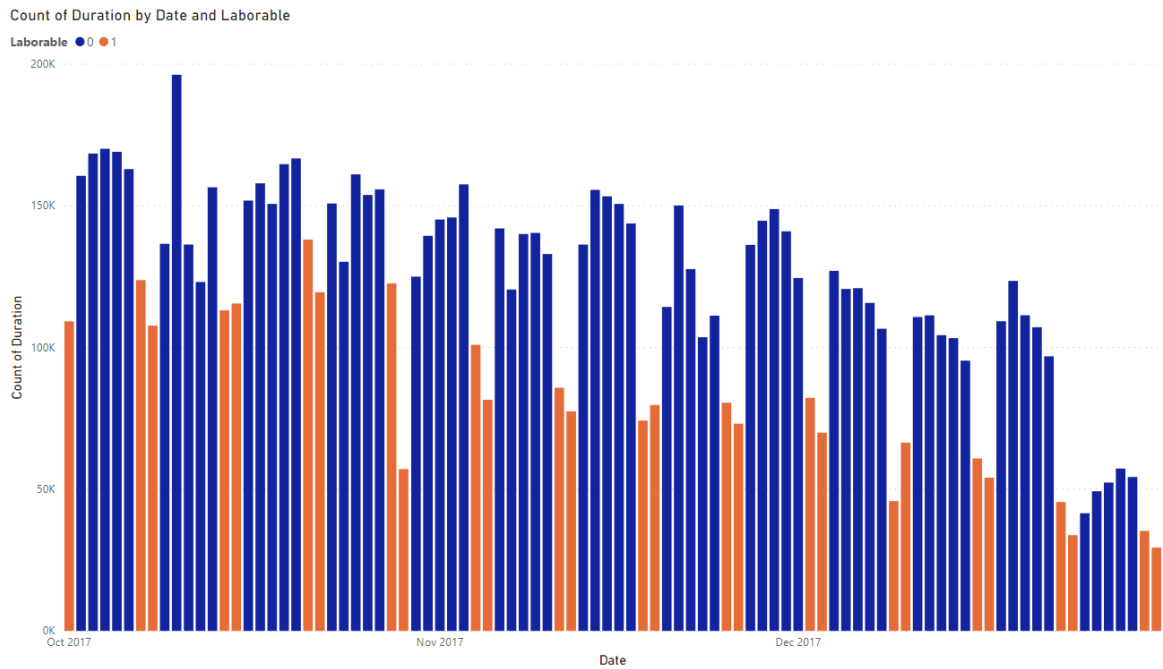


Figure 6-NN Conteo Abril-Junio 2017 Dia



*Figure 6-OO Conteo Julio-Septiembre 2017 Dia*



*Figure 6-PP Conteo Octubre-Diciembre 2017 Dia*

### 6.5.3 DURACIONES MEDIAS 2018

Average of Duration by Date and Laborable

Laborable ● 0 ● 1

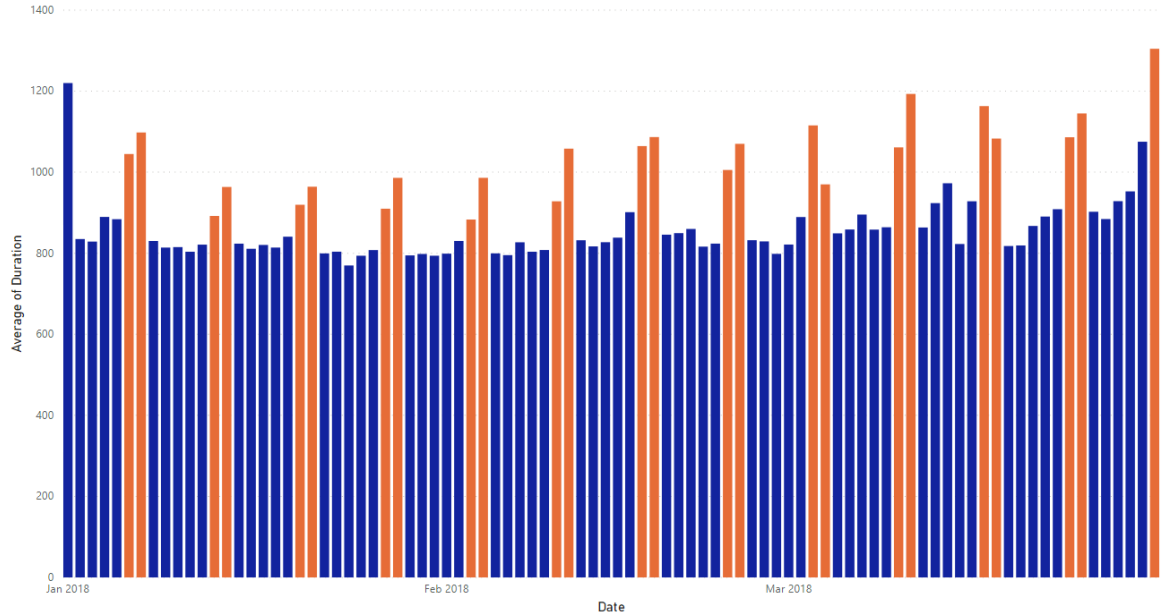


Figure 6-QQ Media Enero-Marzo 2018 Dia

Average of Duration by Date and Laborable

Laborable ● 0 ● 1

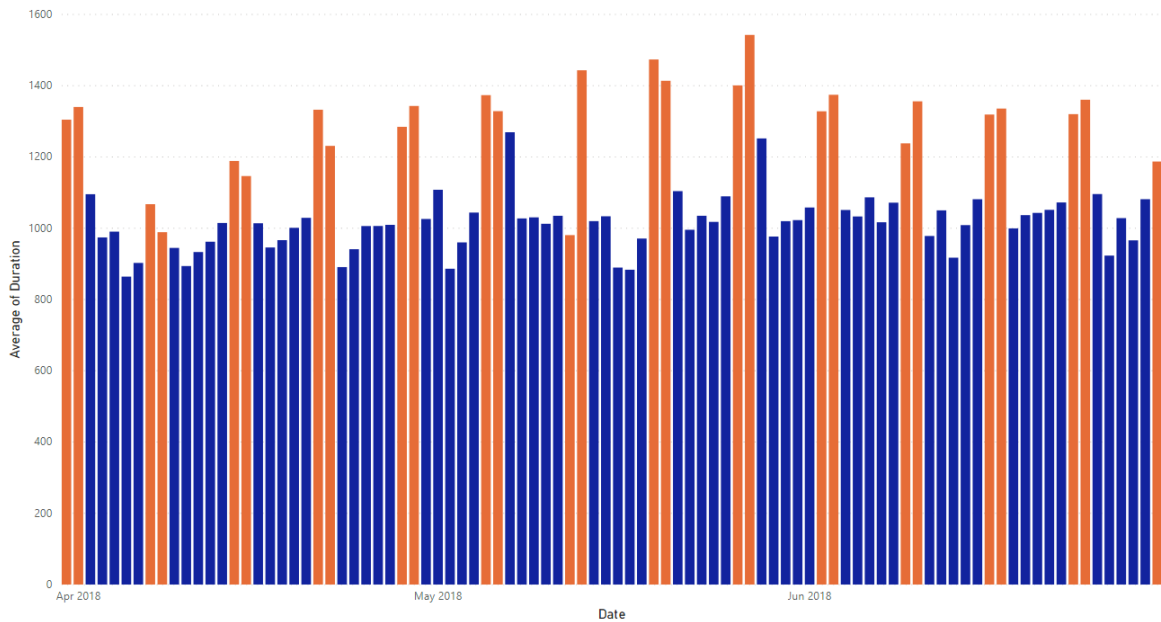
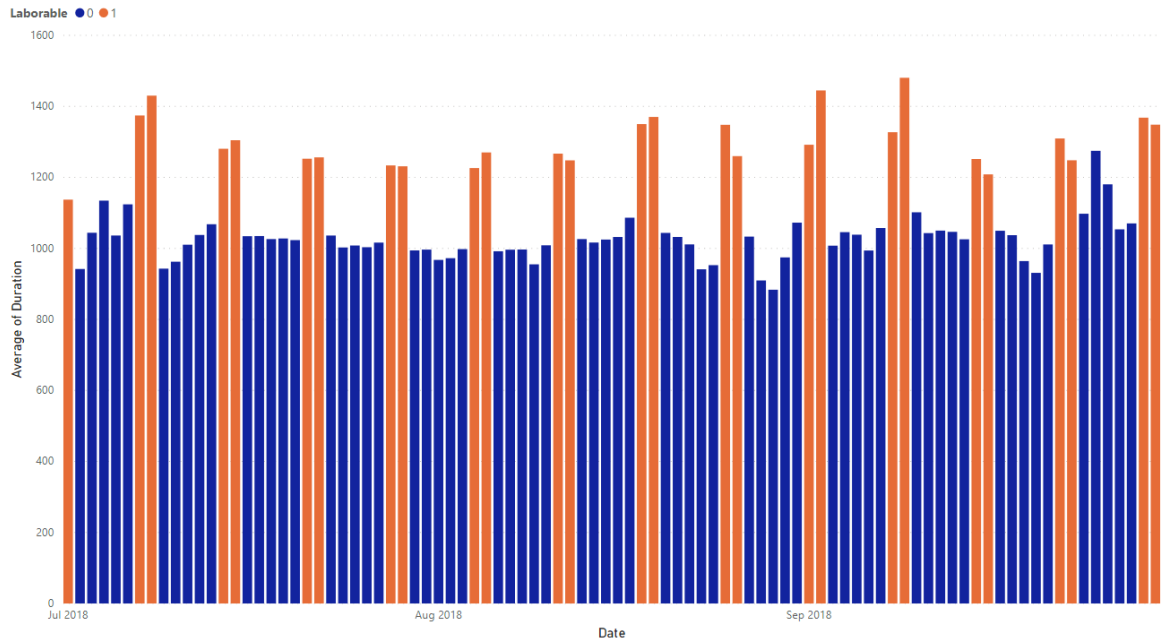


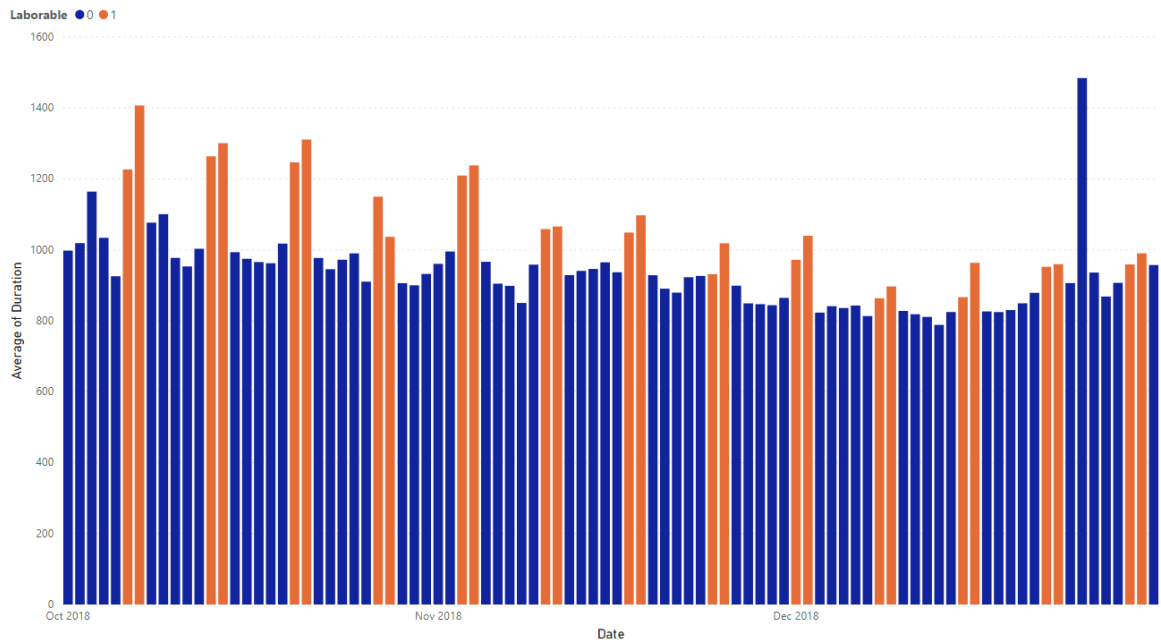
Figure 6-RR Media Abril-Junio 2018 Dia

Average of Duration by Date and Laborable



*Figure 6-SS Media Julio-Septiembre 2018 Dia*

Average of Duration by Date and Laborable



*Figure 6-TT Media Octubre-Diciembre 2018 Dia*

## 6.5.4 CONTEO DE VIAJES 2018

Count of Duration by Date and Laborable

Laborable ● 0 ● 1

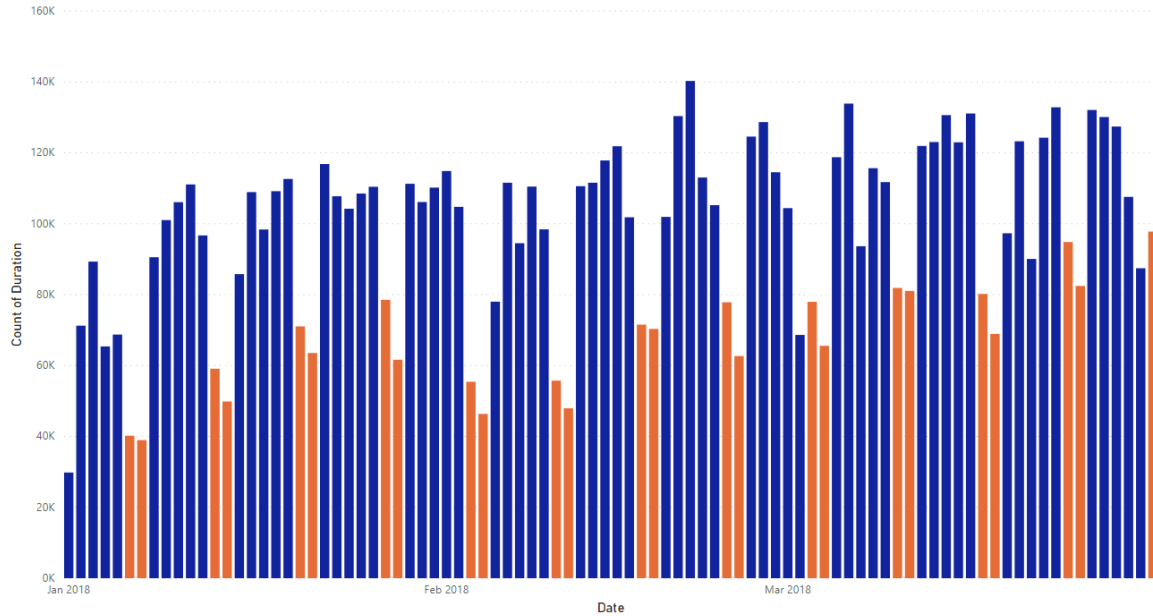


Figure 6-UU Conteo Enero-Marzo 2018 Dia

Count of Duration by Date and Laborable

Laborable ● 0 ● 1

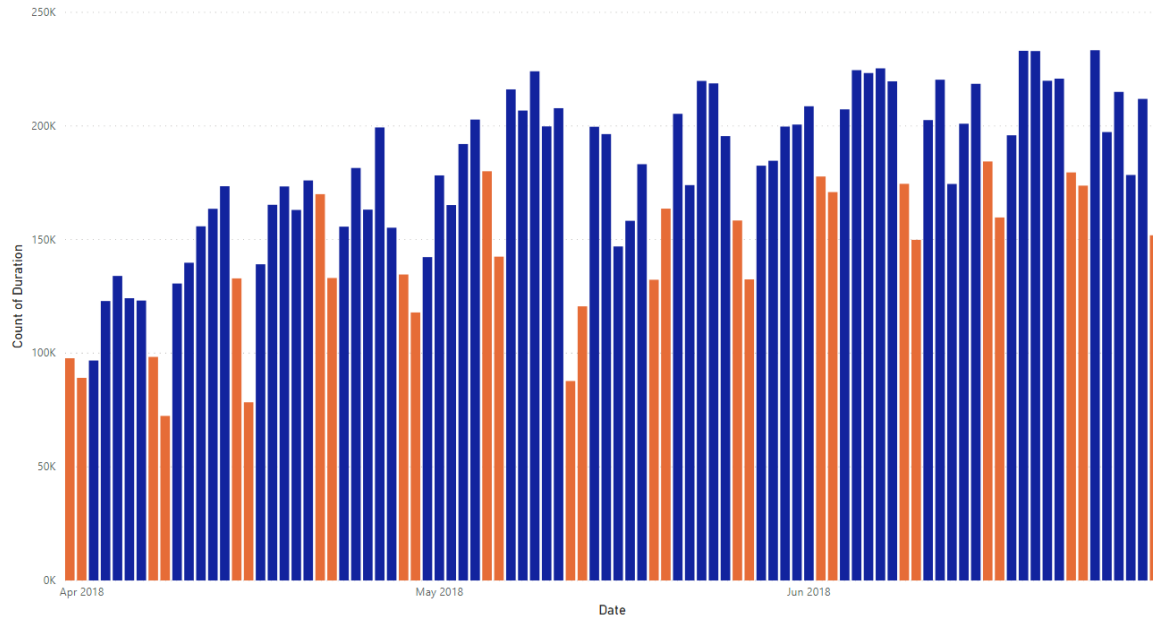
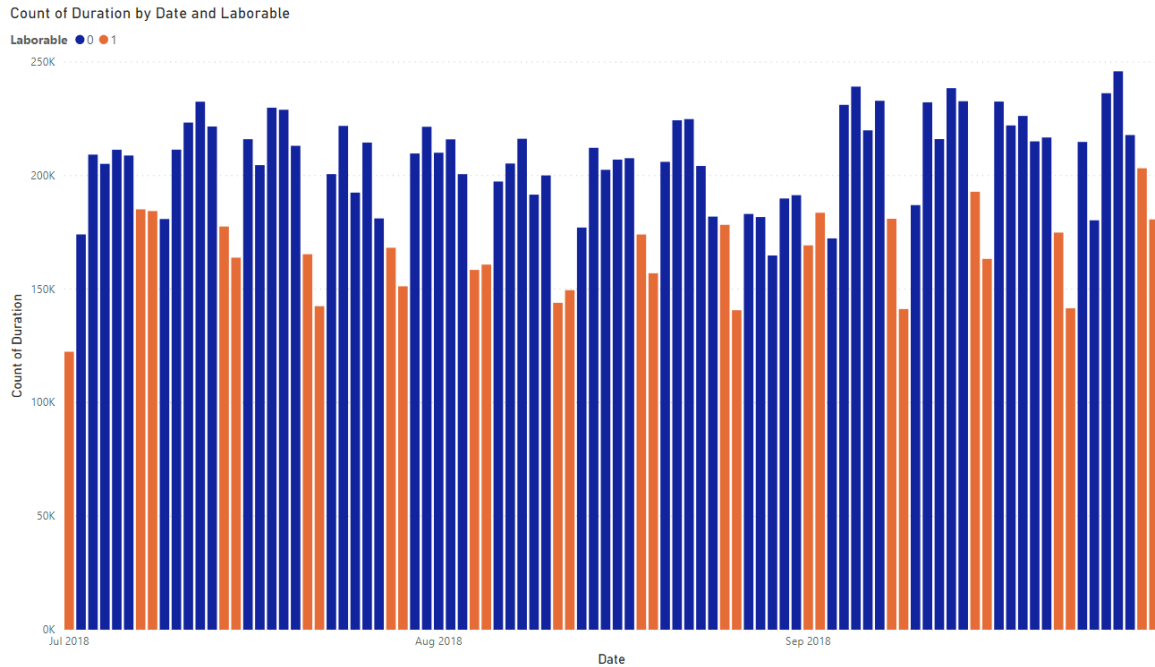
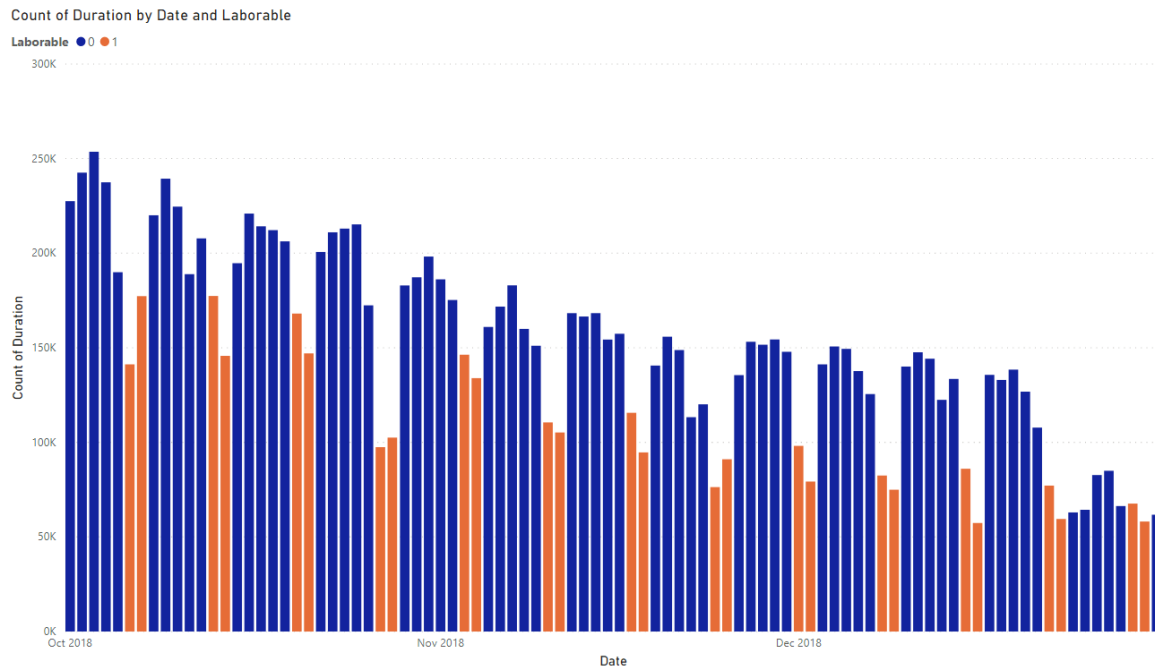


Figure 6-VV Conteo Abril-Junio 2018 Dia



*Figure 6-WW Conteo Julio-Septiembre 2018 Dia*



*Figure 6-XX Conteo Octubre-Diciembre 2018 Dia*

### 6.5.5 DURACIONES MEDIAS 2019

Average of Duration by Date and Laborable

Laborable ● 0 ● 1

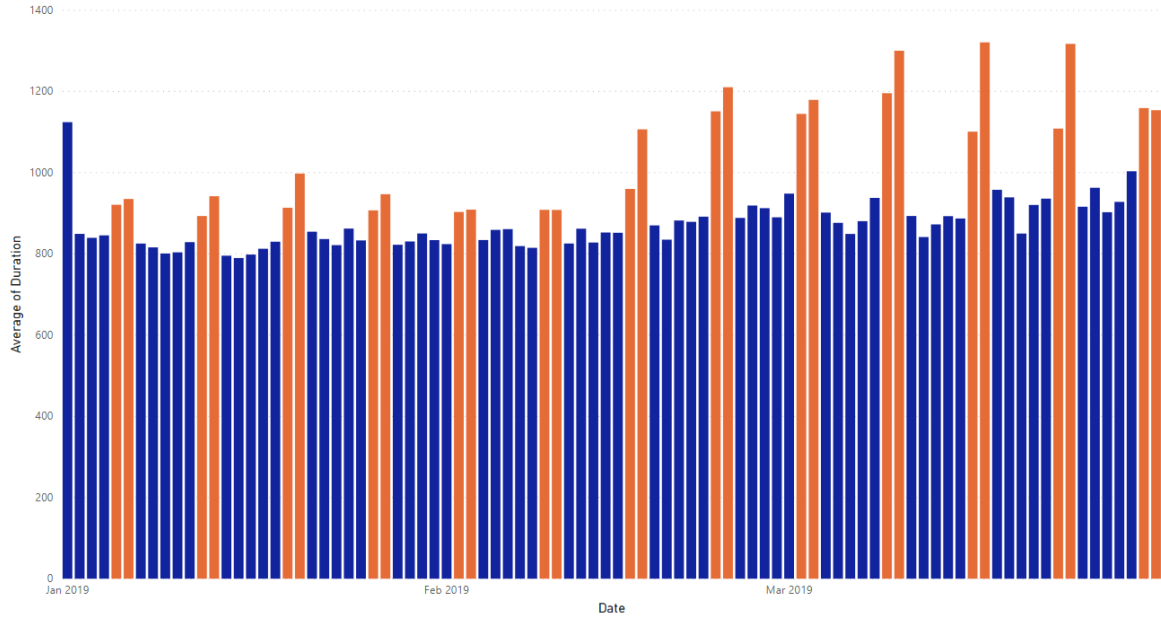


Figure 6-YY Media Enero-Marzo 2019 Dia

Average of Duration by Date and Laborable

Laborable ● 0 ● 1

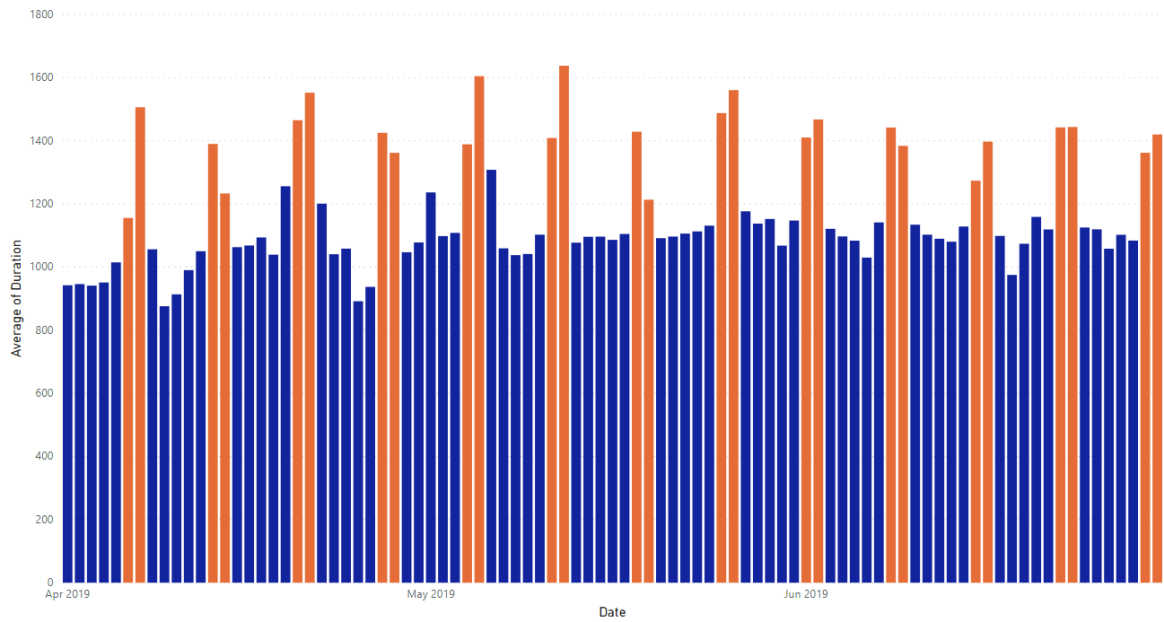
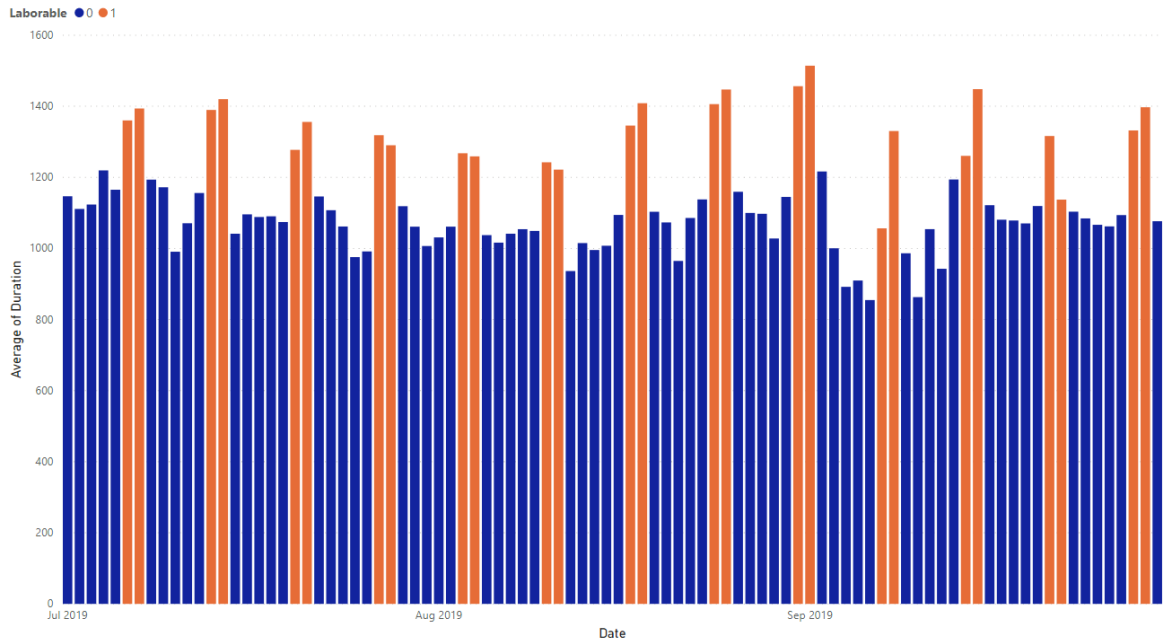


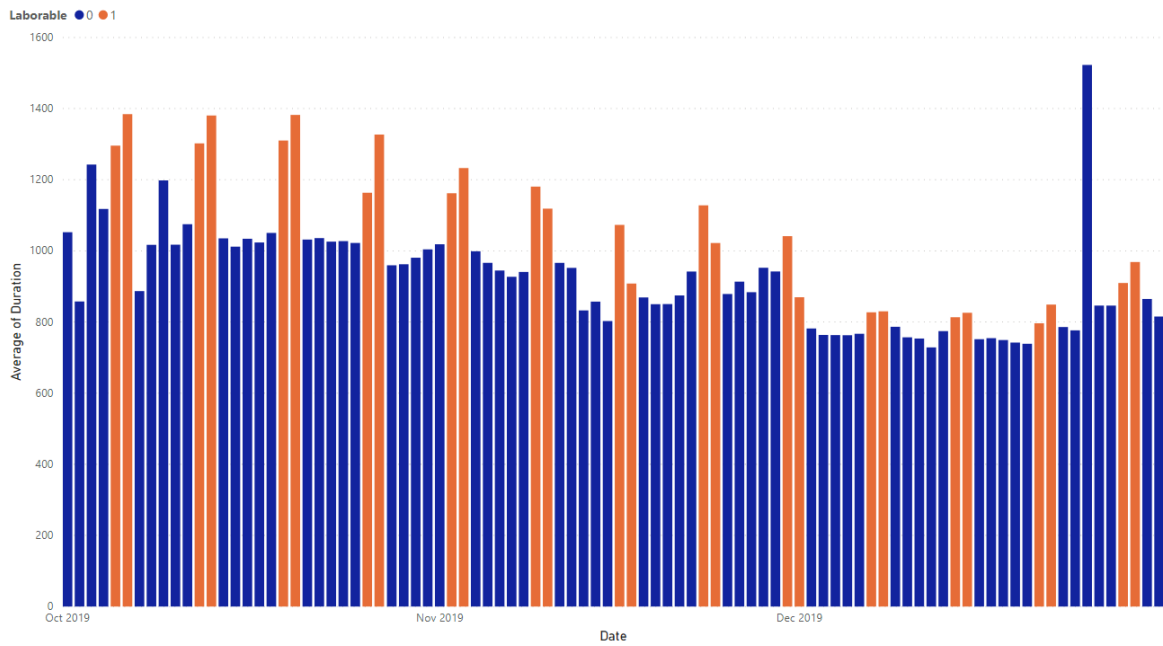
Figure 6-ZZ Media Abril-Junio 2019 Dia

Average of Duration by Date and Laborable



*Figure 6-AAA Media Julio-Septiembre 2019 Dia*

Average of Duration by Date and Laborable



*Figure 6-BBB Media Octubre-Diciembre 2019 Dia*

## 6.5.6 CONTEO DE VIAJES 2019

Count of Duration by Date and Laborable

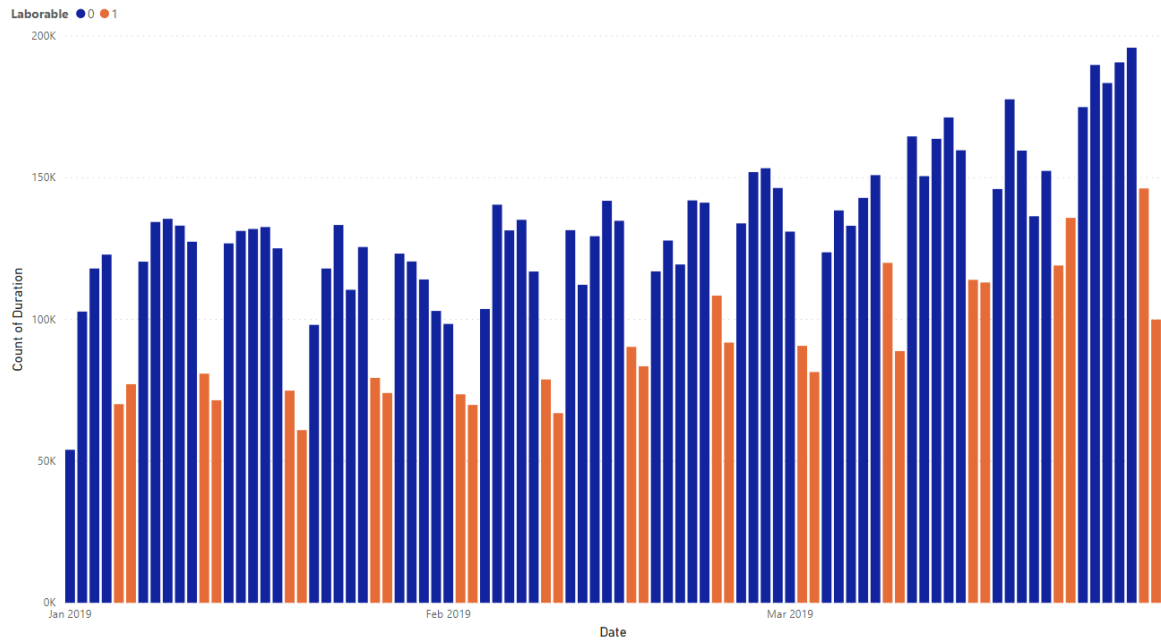


Figure 6-CCC Conteo Enero-Marzo 2019 Dia

Count of Duration by Date and Laborable

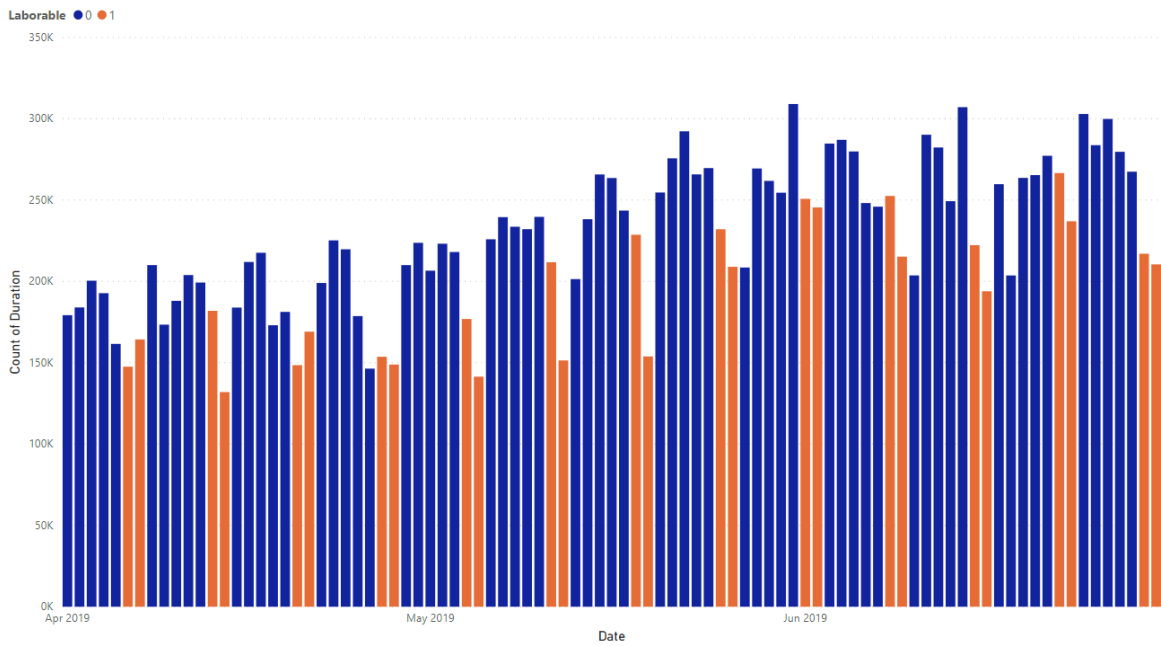
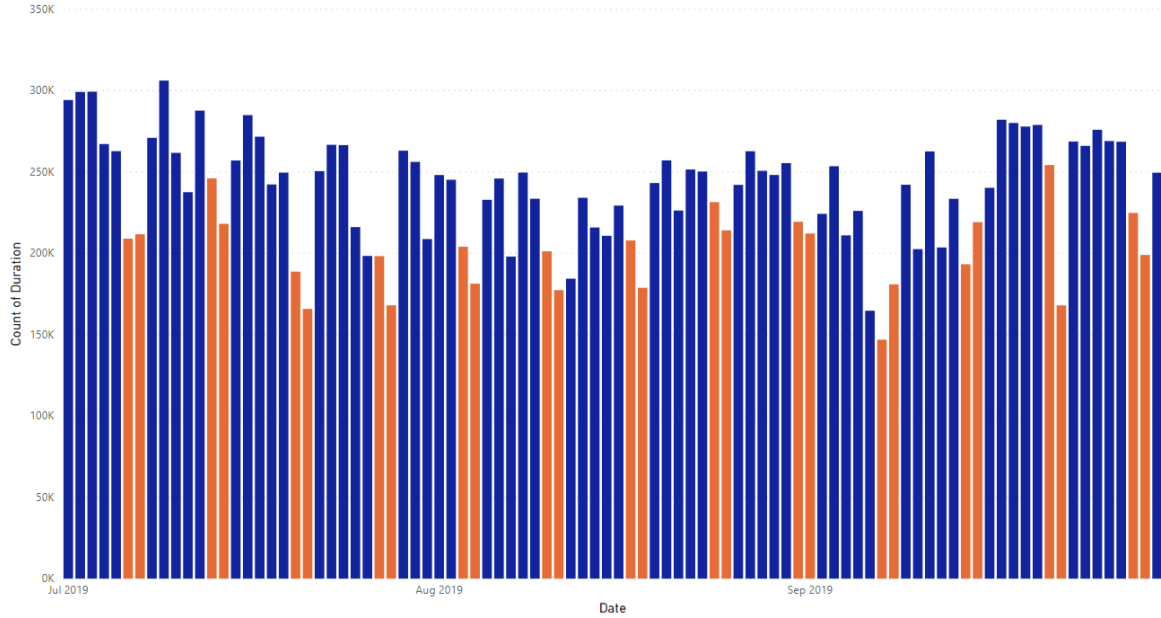


Figure 6-DDD Conteo Abril-Junio 2019 Dia

Count of Duration by Date and Laborable

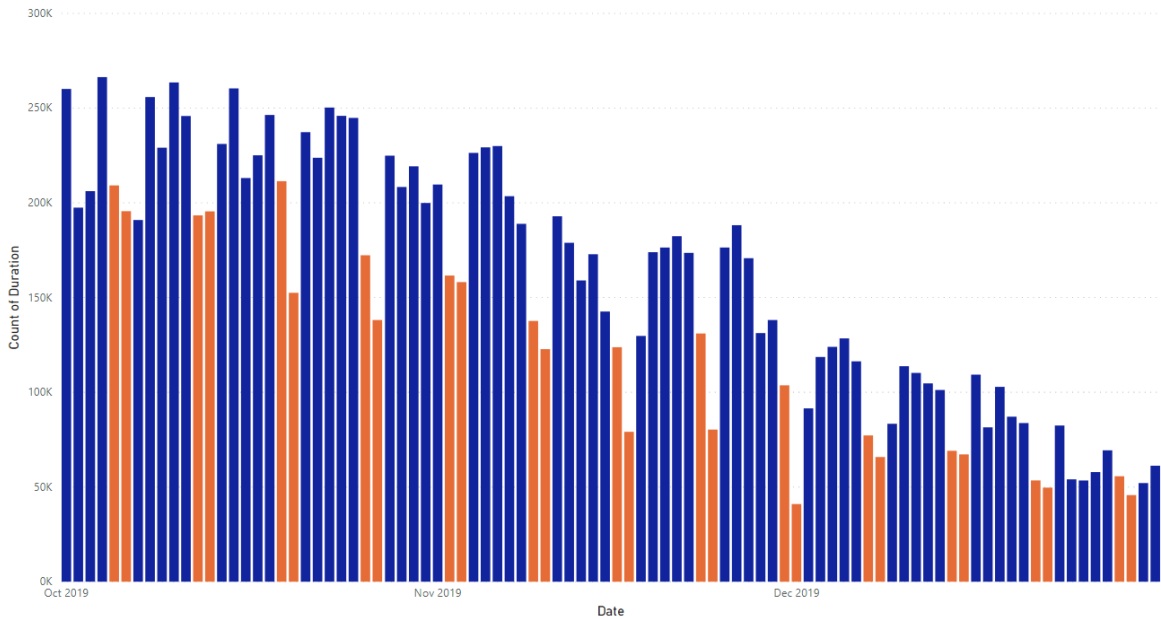
Laborable ● 0 ● 1



*Figure 6-EEE Conteo Julio-Septiembre 2019 Dia*

Count of Duration by Date and Laborable

Laborable ● 0 ● 1



*Figure 6-FFF Conteo Octubre-Diciembre 2019 Dia*

## **6.5.7 RESULTADOS**

A la hora de la observación y evaluación de resultados para este apartado, la primera conclusión a la que se llega es que el comportamiento, tanto en el formato de duraciones medias como en el del conteo de viajes es muy similar año a año. Si bien cabe destacar que los valores del conteo de viajes diarios experimentan una tendencia creciente, esto es casi con toda probabilidad mayoritariamente debido al aumento en la popularidad y el uso del servicio.

Por otro lado, al centrarse en la parte de las duraciones medias, cabe destacar que los resultados obtenidos son coherentes con lo que se había extraído del apartado 6.1. La duración media en los meses más fríos se sitúa en torno a los 800 segundos mientras que para los meses de temperaturas más apacibles se tienen valores más cercanos a los 1200.

También hay que destacar, aun observando las duraciones medias, que los picos en los cuales estas duraciones tienen mayor valor se sitúan en los días no laborables del año, lo que reafirma la conclusión extraída en el apartado 6.4 de que los usuarios que utilizan este servicio por cuestiones más orientadas al ocio tienen a realizar viajes más largos que los usuarios que lo utilizan en ocasiones de diario. También hay que recalcar que un día festivo en todas las localidades estudiadas, el día 25 de diciembre, Navidad, la duración media obtenida pese a ser un día de semana, encaja con los valores obtenidos para días normalmente no laborables, lo que reafirma aún más esta conclusión.

Dejando ahora de lado el estudio de las duraciones medias, resulta interesante comprobar como prestando atención al conteo de viajes se observa una foto opuesta a la obtenida en el estudio anterior. Los números presentan un valor notablemente superior los días laborables, existiendo una diferencia de hasta 75.000 viajes para los meses fríos y de 100.000 viajes para los meses cálidos, lo cual solo refuerza la extracción anterior de que el uso durante días laborables corresponde a trabajadores o jóvenes estudiantes que utilizan el servicio como método de transporte y el uso durante días no laborables corresponde a visitantes o viajes de ocio.

En resumen, el uso del servicio se concentra en los días laborables, pero las duraciones en proporción son altamente superiores los días no laborables.

## Capítulo 7. CONCLUSIONES Y TRABAJOS FUTUROS

El presente trabajo engloba una solución de estudio para cualquier posible ciudad que disponga actualmente de servicio de bike sharing y busque realizar una mejora de sus prestaciones o directamente no disponga del servicio y se plantee su instalación.

Mediante las herramientas de análisis de datos y programación, es posible la realización sencilla de un estudio de mejora o éxito de la iniciativa basado en los datos actuales o en una modificación o interpolación de los mismos.

De cara a los resultados obtenidos, cabe destacar que las mayores limitantes en el uso de este servicio son las condiciones medioambientales extremadamente adversas como los tifones o las nevadas, un nivel de poder adquisitivo muy bajo o una edad muy avanzada.

Por otro lado, los mayores usuarios del servicio son aquellos jóvenes o adultos que diariamente utilizan estos sistemas para el transporte sencillo hacia su puesto de trabajo o de estudio.

Se llega también a la conclusión de que los usuarios registrados en el servicio hacen un uso predominante en los días laborables mientras que los usuarios esporádicos o no registrados utilizan el servicio en días no laborables.

Resulta interesante la interpretación de los datos vistos a largo plazo, ya que se observa claramente como el comportamiento de los usuarios tanto en el conteo de viajes como en las duraciones medias de los mismos siguen una distribución senoidal con mayor amplitud en los meses de verano y menor en los meses de invierno. En las siguientes figuras se muestra este fenómeno, dejando de lado a Buenos Aires al ser una ciudad del hemisferio sur.

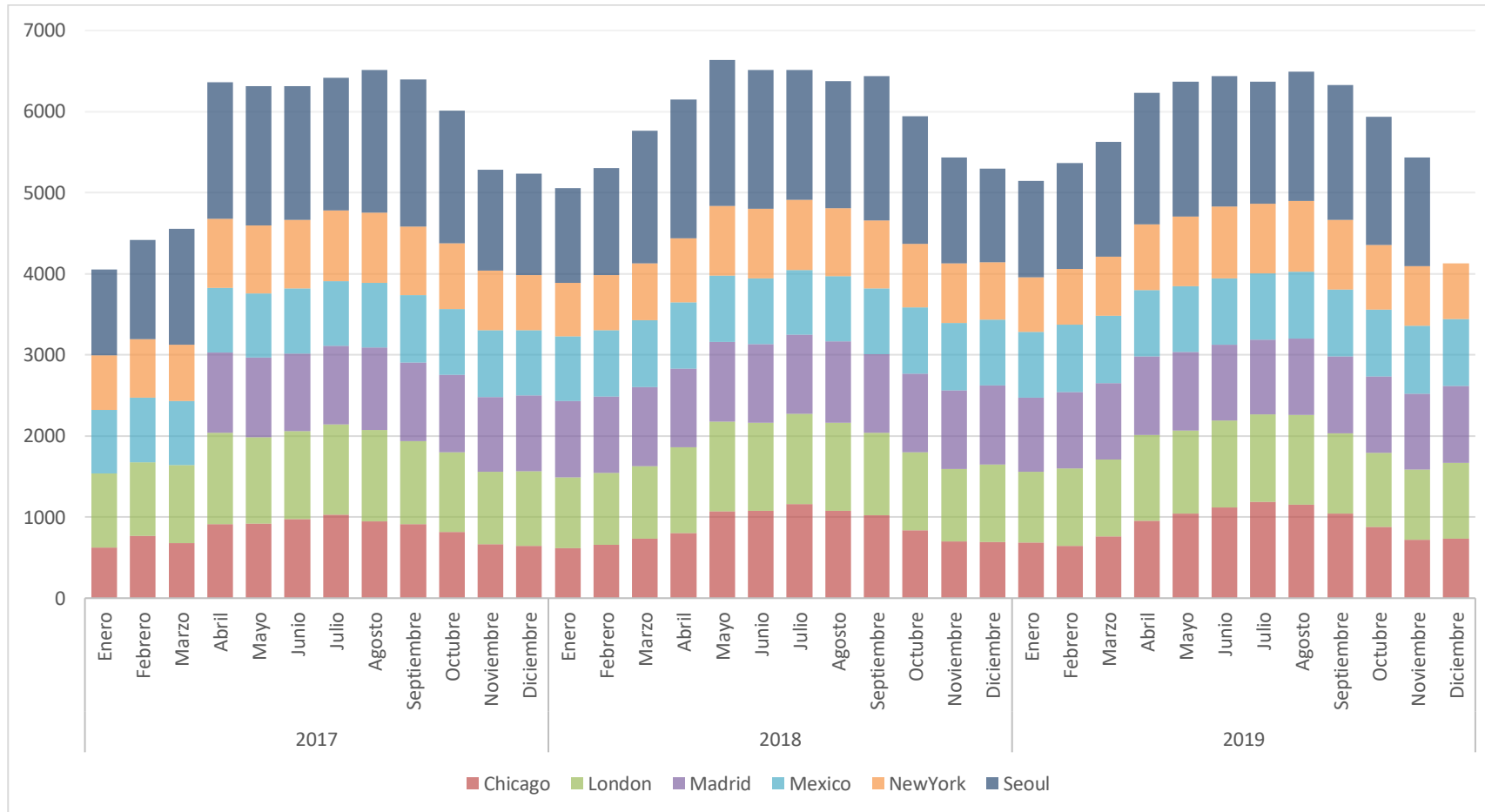


Figure 7-A Duración media 2017, 2018 y 2019

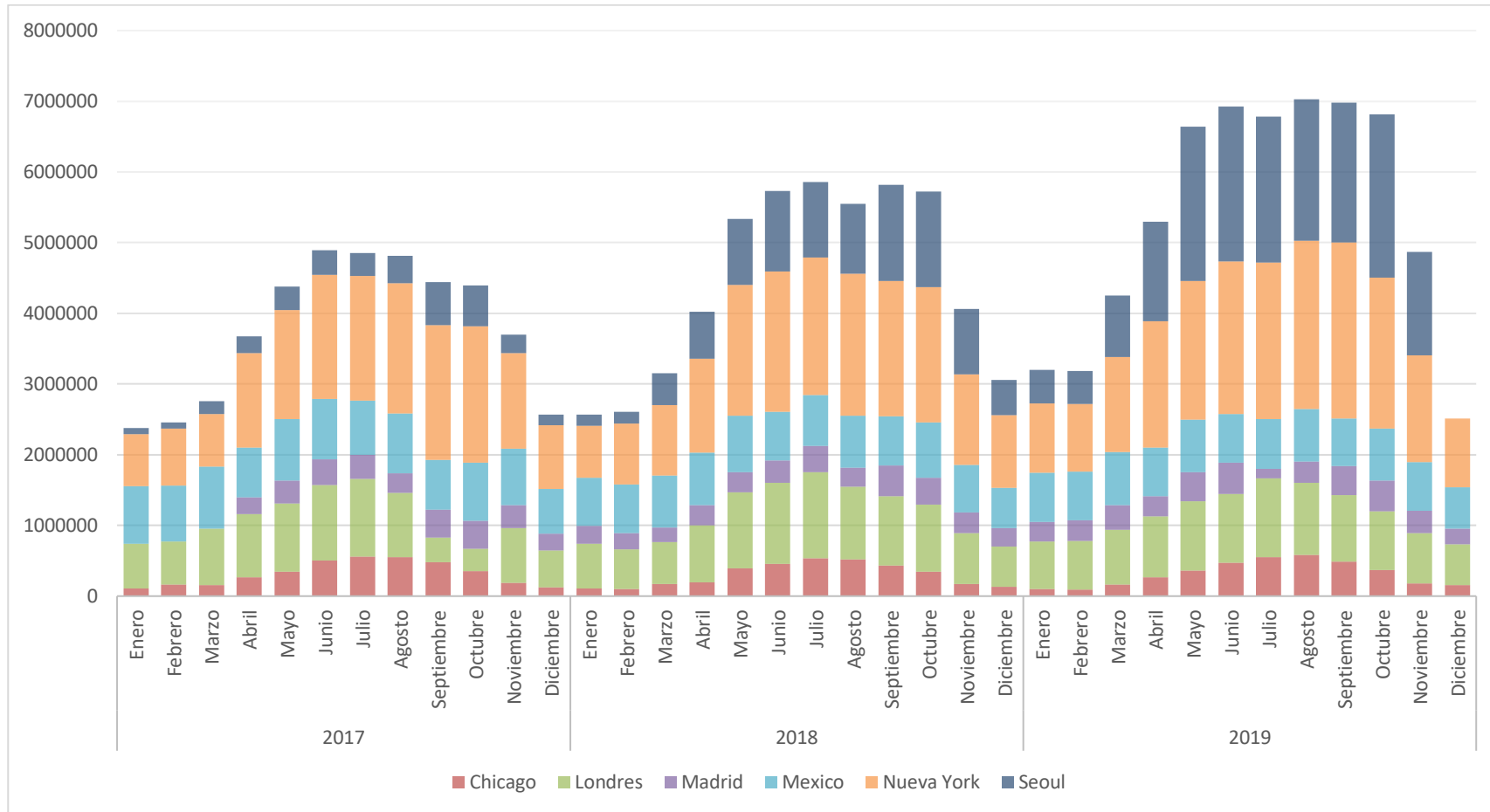


Figure 7-B Conteo de duración 2017, 2018 y 2019

No obstante, este fenómeno podría ser caso de una coincidencia en los datos de los años estudiados, por lo que sería necesario un estudio más amplio y con más ciudades para la realización de comprobaciones adicionales.

Enlazando con la reflexión anterior, contando con el trabajo realizado hasta el momento, resulta sencillo realizar una ampliación a futuras localidades, aunque llevar a cabo el análisis de las contempladas hasta ahora haya sido complicado, se entiende que la muestra es lo suficientemente grande para englobar las posibles diferencias que pueda haber entre las distintas formas de almacenamiento de datos, y en el hipotético caso de que no hayan sido contempladas aun, el cambio a realizar para la nueva adaptación debería ser sencillo. Además, la herramienta de representación está diseñada de tal forma que, con la capacidad computacional necesaria, es capaz de tratar un numero indefinido de nuevas ciudades que se quieran considerar sin modificación alguna.

Finalmente, como posibles ampliaciones al planteamiento inicial, se podría valorar la utilización de un servicio de hosting en la nube para permitir el acceso remoto y una mayor capacidad de procesamiento del sistema, es decir, mover la parte de trabajo computacional a lo que podría ser una arquitectura big data. Por otro lado, como se comentaba anteriormente, esta capacidad extra de procesamiento permitiría aumentar el espectro de ciudades bajo estudio y por tanto una mayor exactitud a la hora de la obtención de conclusiones.

## Capítulo 8. BIBLIOGRAFÍA

- [1] Meteorología: <https://www.ncdc.noaa.gov/>
- [2] ODS: <https://verdesdigitales.com/2019/09/03/ods-en-comunicacion/>
- [3] Metodología scrum: <https://www.scrum.org/resources/blog/que-es-scrum>
- [4] Bike share map: <https://bikesharemap.com/#/4/10.9035/48.7522/>
- [5] Descripción datos Madrid:  
<https://opendata.emtmadrid.es/Documentos/Servicios-y-estructuras-Bicimad-V1-1.aspx>
- [6] Datos Madrid: [https://opendata.emtmadrid.es/Datos-estaticos/Datos-generales-\(1\)](https://opendata.emtmadrid.es/Datos-estaticos/Datos-generales-(1))
- [7] Descripción datos Nueva York: <https://www.citibikenyc.com/system-data>
- [8] Datos Nueva York: <https://s3.amazonaws.com/tripdata/index.html>
- [9] Descripción datos Londres: <https://tfl.gov.uk/info-for/open-data-users/our-open-data?intcmp=3671#on-this-page-5>
- [10] Datos Londres: <https://cycling.data.tfl.gov.uk/>
- [11] Datos y descripción México:  
<https://www.ecobici.cdmx.gob.mx/en/informacion-del-servicio/open-data>
- [12] Datos y descripción Seúl: <http://data.seoul.go.kr/dataList/OA-15182/F/1/datasetView.do?tab=F#>
- [13] Datos y descripción Buenos Aires:  
<https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>

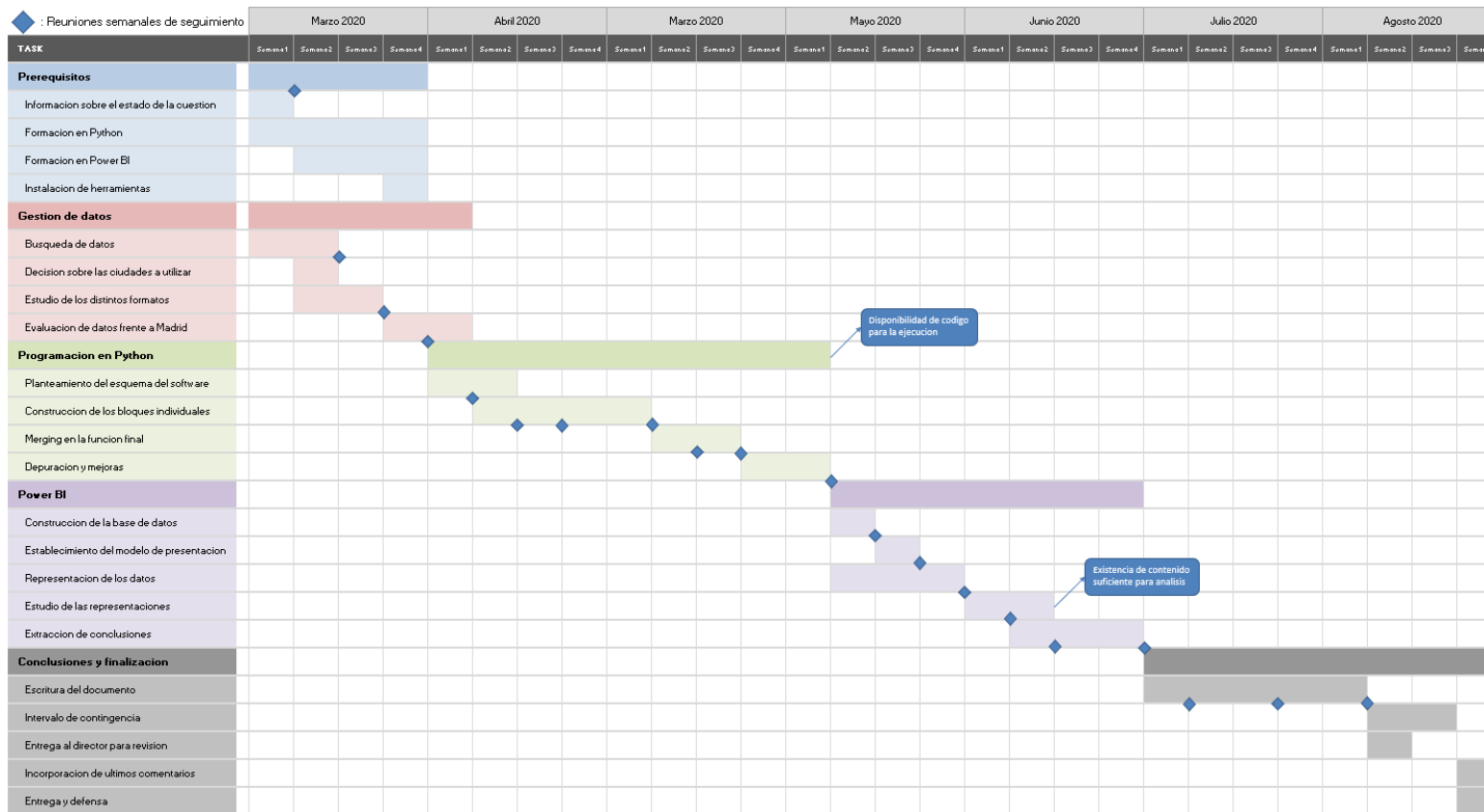
- [14] Datos y descripción Chicago: <https://divvy-tripdata.s3.amazonaws.com/index.html>
- [15] Elliot Fishman, “Bike share”. Julio, 2019.  
[https://books.google.es/books?id=nYWiDwAAQBAJ&pg=PT58&hl=es&source=gbs\\_selected\\_pages&cad=2#v=onepage&q&f=false](https://books.google.es/books?id=nYWiDwAAQBAJ&pg=PT58&hl=es&source=gbs_selected_pages&cad=2#v=onepage&q&f=false)
- [16] Christopher Moon-Miklaucic, Anna Bray-Sharpin, Ivan de la Lanza, Azra Khan, Luca Lo Re, and Anne Maassen, “The evolution of bike sharing: 10 questions on the emergence of new technologies, opportunities and risks”. World Resources Institute. Enero 2019.  
<https://www.retocdmx.com/documents/wri3.pdf>
- [17] John Shovic, Alan Simpson, “Python for dummies”. John Wiley & Sons. 2019.  
[http://dl.booktolearn.com/ebooks2/computer/python/9781119557593\\_Python\\_All\\_in\\_One\\_For\\_Dummies\\_b106.pdf](http://dl.booktolearn.com/ebooks2/computer/python/9781119557593_Python_All_in_One_For_Dummies_b106.pdf)
- [18] Michael Heydt, “Learning Pandas”. Packt Publishing Ltd. Junio 2017.  
<https://www.packtpub.com/product/learning-pandas/9781783985128>
- [19] Brett Powell, “Mastering Microsoft Power BI”. Packt Publishing Ltd. Marzo 2018. <https://www.packtpub.com/product/mastering-microsoft-power-bi/9781788297233>
- [20] Brian Larson, “Data Analysis with Microsoft Power BI”. McGraw Hill. Diciembre 2019. <https://www.mheducation.com/highered/product/data-analysis-microsoft-power-bi-larson/9781260458619.html>

## ANEXO A – CÓDIGO

El código de la función utilizada en Python se recoge en el siguiente link de Github para una mayor facilidad de visualización y utilización:

<https://github.com/SimonDiazVigon/TFM/blob/master/TFM.ipynb>

## ANEXO B – CRONOGRAMA

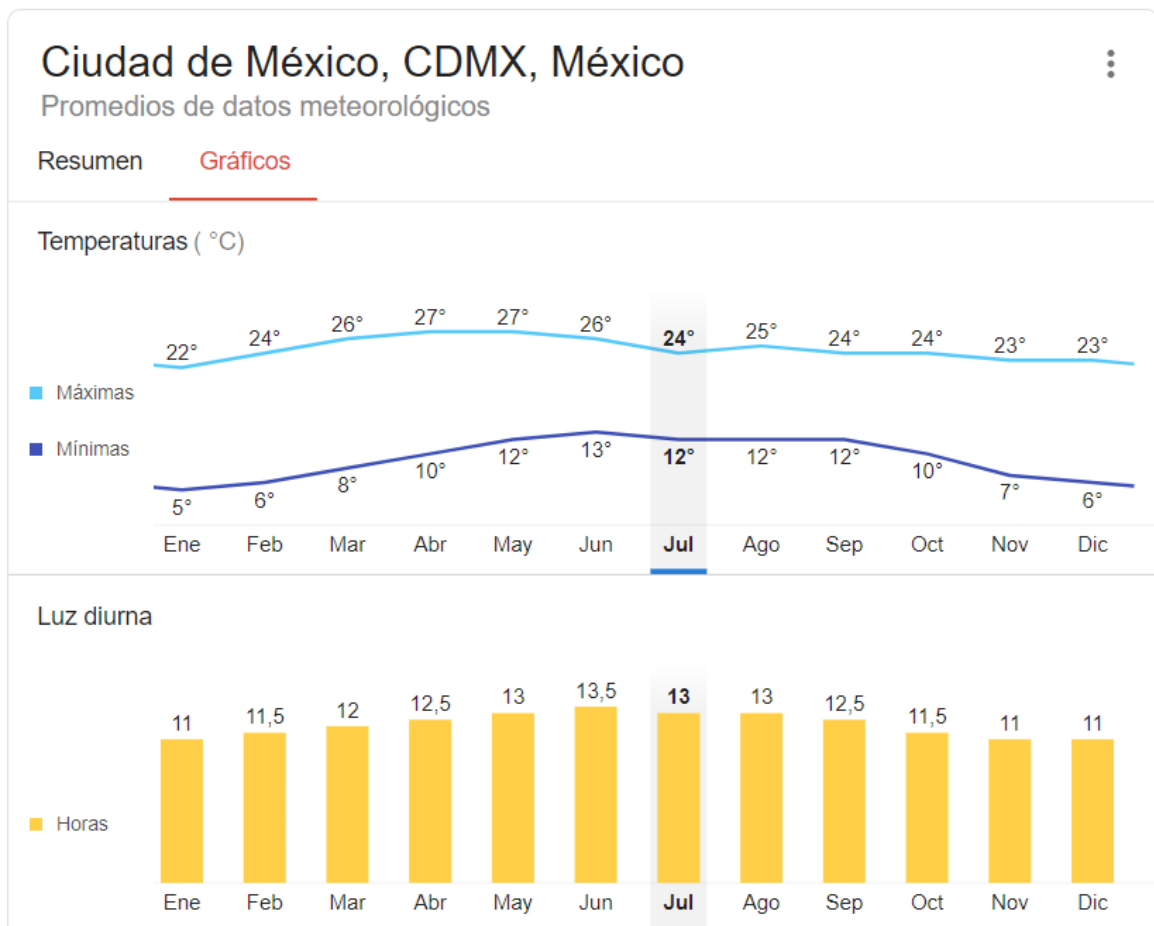


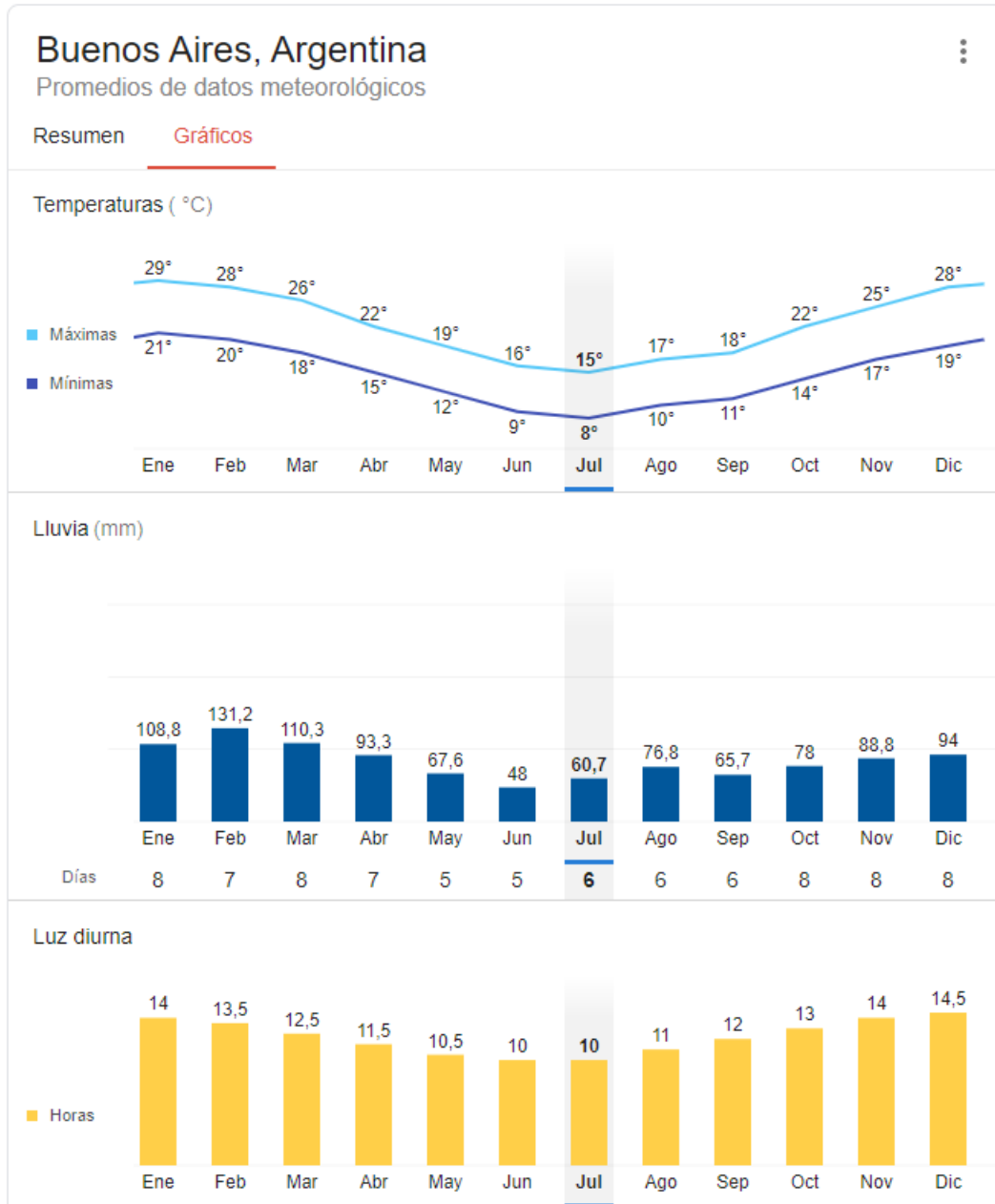
**ANEXO C –  
TABLA  
INPUTS**

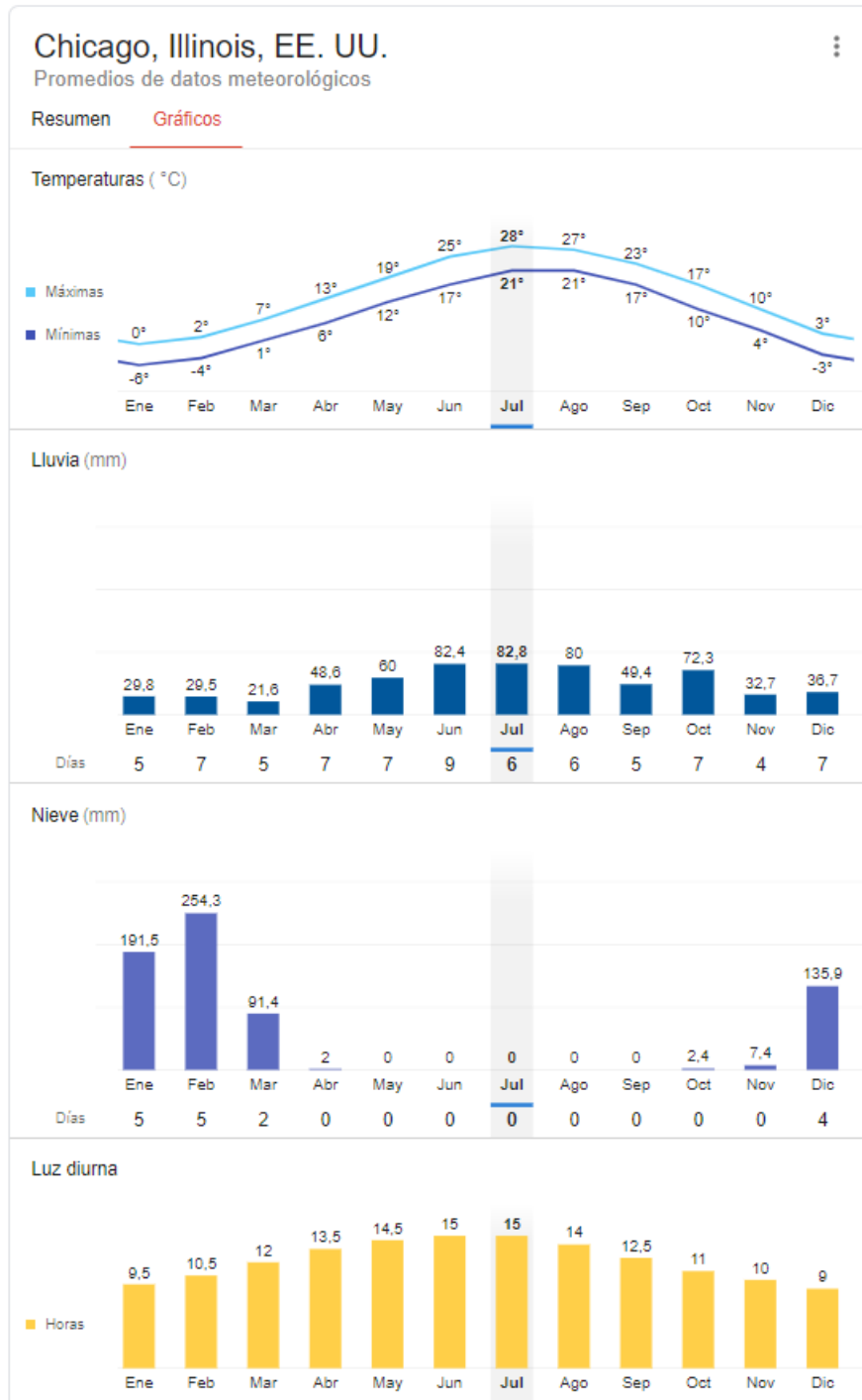
Parametro	Valor	Descripcion Especifica	Descripcion Generica
Source	./TFM/BDData/2018/seoul_201807_1M.csv\$./TFM/BDData/2018/seoul_201807	-	Archivo fuente desde el que se va a realizar la conversión de los datos.
	./TFM/BDData/2018/2018/2018-12.csv	-	
	./TFM/BDData/2018/138JourneyDataExtract28Nov2018-04Dec2018.csv\$138	-	
	./TFM/BDData/2018/JC-201812-citibike-tripdata.csv\$201812-citibike-tripdat	-	
	./TFM/BDData/2018/chicago_2019_Q1.csv	-	
	./TFM/BDData/2018/buenos_aires_CE_2018.csv	-	
Destination	./TFM/BDData/2018/201812_Usage_Bicimad.json	-	Directorio y nombre de archivo de destino, el cual se genera en el último paso de ejecución de la función.
	./TFM/BDData/Seoul2018_07.csv	-	
	./TFM/BDData/Mex2018_12.csv	-	
	./TFM/BDData/London2018_07.csv	-	
	./TFM/BDData/NYork2018_12.csv	-	
	./TFM/BDData/Chicago2019_01.csv	-	
Formato	1	Formato .csv	Formato en el que se pasa el archivo de origen.
	2	Formato .json	
Año	2017	-	Año al que corresponden los datos del archivo de origen.
	2018	-	
	2019	-	
Mes	1 2 3 4 5 6 7 8 9 10 11 12	-	Mes correspondiente.
Formato fecha	1 7	Ya viene en un formato admitido por la librería	Formato en el que está grabada la fecha en el archivo original.
	2	Fechas y horas segmentadas	
	3	Formato coreano	
	4 6	Con microsegundos	
	5	Dentro de una cadena de caracteres	
Calculo duracion	0	Ya viene dada	Como ha de calcularse la duración del viaje.
	1	Resta entre final de viaje y principio de viaje	
Edad	1	Viene dado por intervalos	Formato de la edad.
	2 4	Viene el año de nacimiento	
	3 5	Viene la edad como numero	
	-3	No hay	
Genero	1	Valores 0,1 y 2	Formato del género.
	2	Masculino o femenino	
	3	M o F	
	4	Male, Female o Empty	
	-3	No hay	
Tipo de usuario	1	Valores 0, 1, 2 y 3	Formato del tipo de usuario.
	2	Subscriber o No Subscriber	
	3	Existe un User ID o no	
	-3	No hay	
Ciudad	10	Madrid	Ciudad con la que se está trabajando.
	20	New York	
	30	Mexico	
	40	Seoul	
	50	London	
	60	Buenos Aires	
70	Chicago		

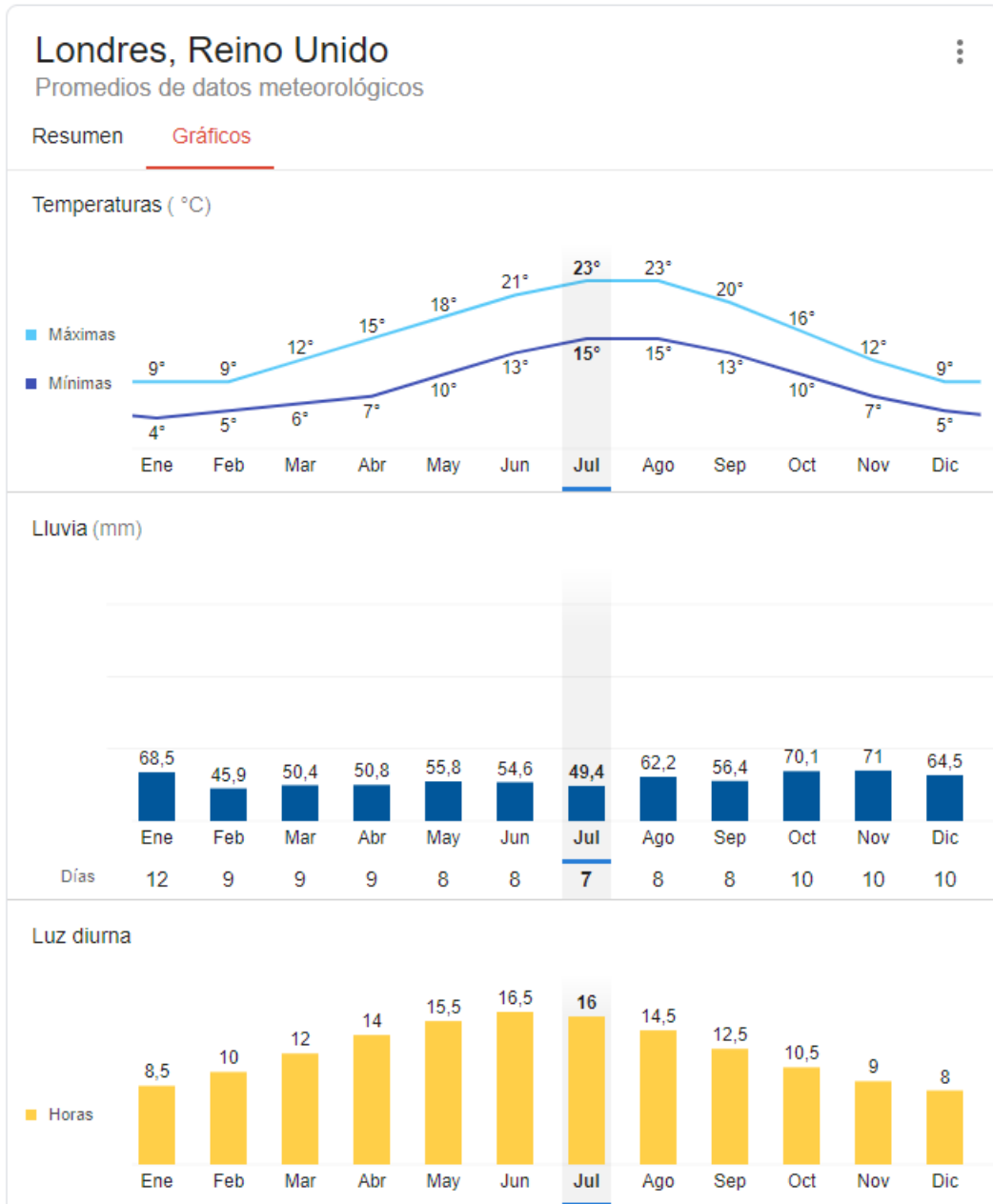
## ANEXO D – METEOROLOGÍA

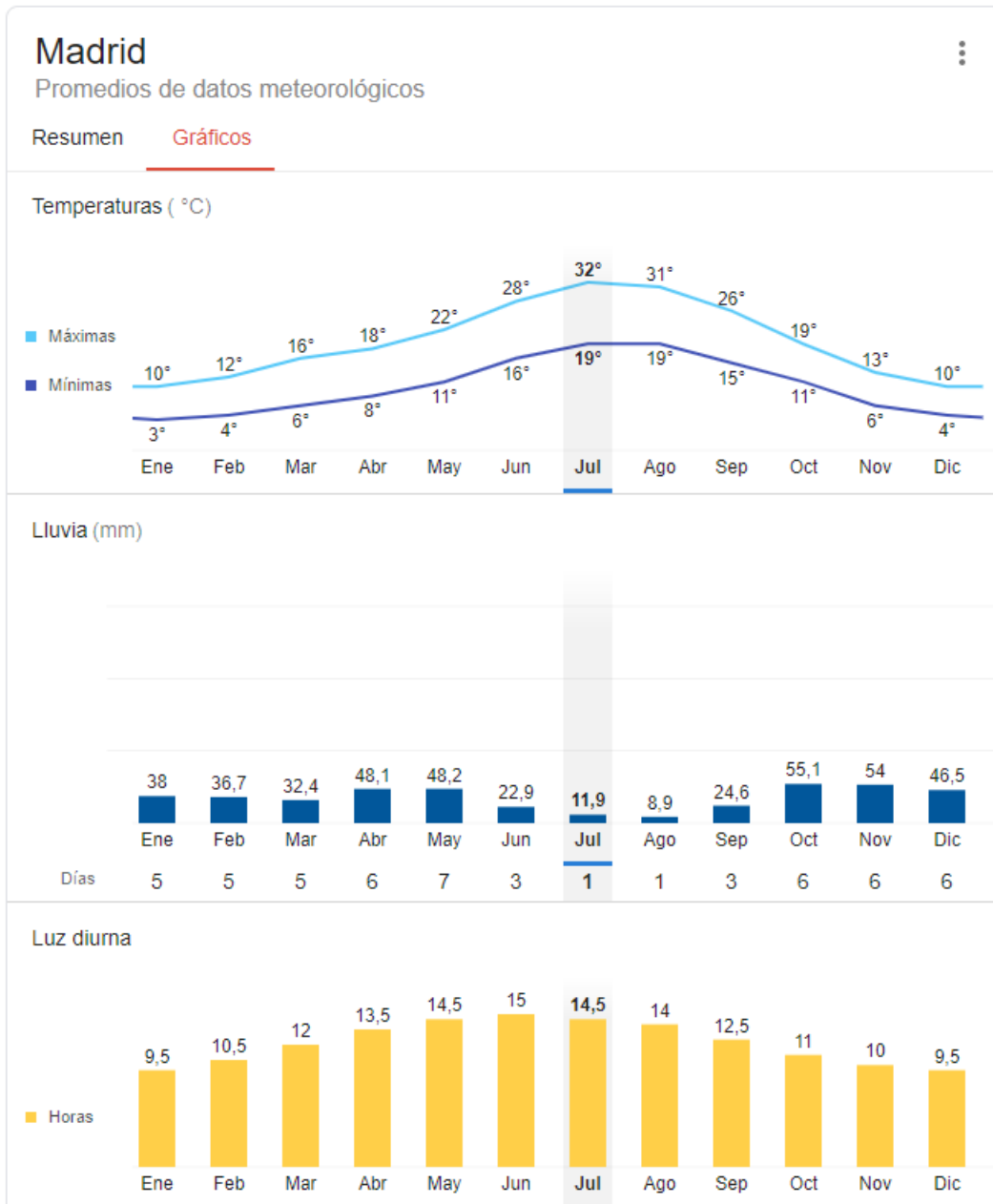
En este anexo se muestran las meteorologías pertinentes a cada una de las ciudades estudiadas según los datos extraídos de la web: <https://www.ncdc.noaa.gov/>

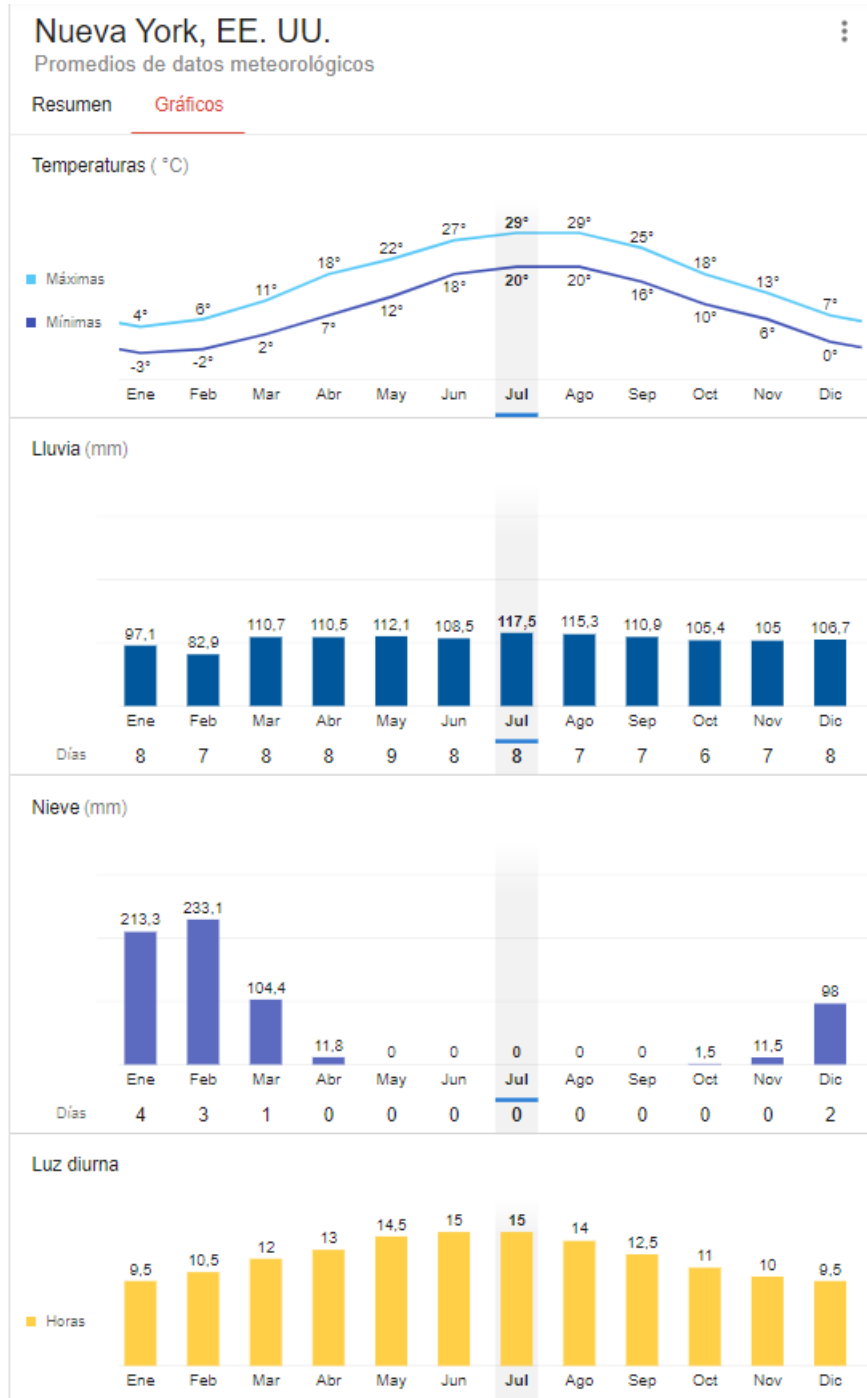


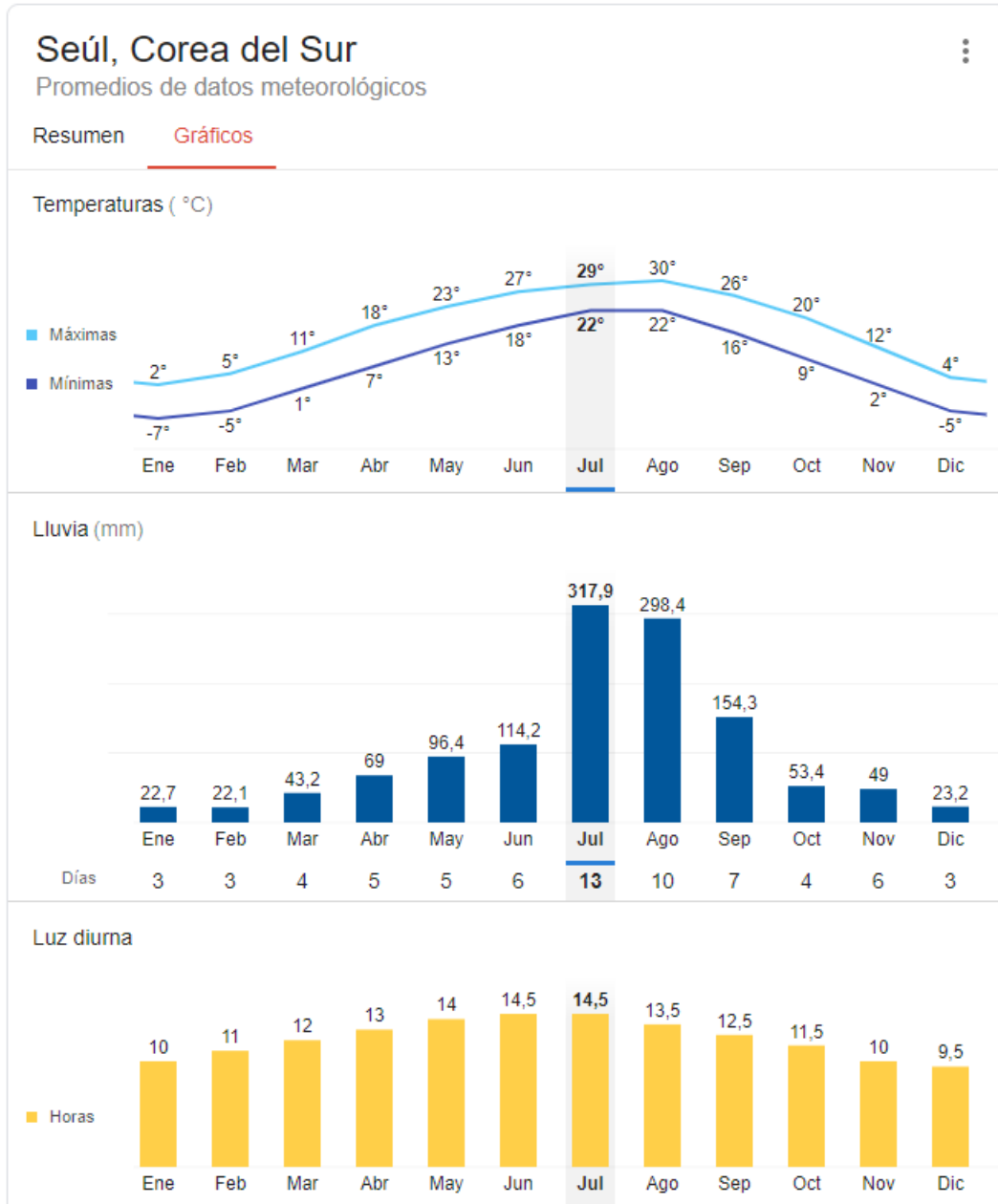












## ANEXO E – ODS

Una de las razones por las que resulta interesante el desarrollo de este estudio es la estrecha relación del servicio con los objetivos de desarrollo sostenible que tanto hemos de tener en cuenta en este sector.



Figure 0-A Objetivos de Desarrollo Sostenible [2]

Numerándolos individualmente, los objetivos con los que más se alinea el trabajo son el 3 salud y bienestar, 7 energía asequible y no contaminante, 11 ciudades y comunidades sostenibles, 13 acción por el clima y 15 vida de ecosistemas terrestres.

Desarrollando el punto, en la realidad actual, la polución provocada por el transporte a motor no eléctrico en los grandes núcleos urbanos es un gran problema, por lo que cualquier tipo de alternativa a esos servicios debería tener una alta prioridad.

Por otro lado, fomentar el movimiento y el ejercicio de los usuarios del servicio desemboca en una mejora de la salud de los mismos y a su vez del resto de la población al reducir las posibles emisiones causadas por el uso de vehículos no eléctricos, sobre todo en los desplazamientos individuales.