**Instituto Tecnológico y de Estudios Superiores de Occidente**

**Departamento de Electrónica, Sistemas e Informática**

**Maestría en Sistemas Computacionales**

**MSC2526A – Programación para Análisis de Datos**

**Otoño 2022**



**El impacto del covid19 al   
proyecto mibici en jalisco**

Proyecto Final

Presenta:

*Cristian Emanuel Ortega Camarena,*

*743622, emanuel.ortega@hotmail.com*

Tlaquepaque, Jalisco. Noviembre de 2022.

**Resumen**

Se presenta un trabajo de comparativa y el impacto de uso en el programa de: “MiBici” a lo largo de la pandemia. Se había decidido originalmente hacer una comparativa desde el año 2018 hasta el mes de Octubre del año 2022, sin embargo las librerías de Python de las cuales nos estamos apoyando no permitieron compararlos ya que al ser demasiados datos no se pudo mas que obtener dos muestras de 15 meses cada una para el manejo de manipulación de datos. Sin embargo, incluso con este mismo problema y corriendo la implementación fue posible determinar que el impacto realmente no fue de la magnitud esperaba, ya que se pensaba que mínimo íbamos a ver 15% de impacto sin embargo en las estaciones mas utilizadas no hubo cambio si no que más ingresos de lo esperado, por lo que cualquiera de las estaciones afectadas puede ser rescatadas financieramente por las estaciones que no presentaron un cambio si no un incremento de uso.

**Tabla de Contenidos**

[1. Objetivo de Investigación 4](#_Toc51781404)

[1.1. Introducción 5](#_Toc51781405)

[1.2. Antecedentes 5](#_Toc51781406)

[1.3. Justificación 5](#_Toc51781407)

[1.4. Problema 5](#_Toc51781408)

[1.5. Hipótesis 5](#_Toc51781409)

[1.6. Objetivos 5](#_Toc51781410)

[1.6.1. Objetivo General: 5](#_Toc51781411)

[1.6.2. Objetivos Específicos: 5](#_Toc51781412)

[1.7. Novedad científica, tecnológica o aportación 5](#_Toc51781413)

[2. Obtención de los Datos 6](#_Toc51781414)

[2.1. Proceso de obtención de los datos 7](#_Toc51781415)

[2.2. Conclusiones 7](#_Toc51781416)

[3. Preparación de los Datos 8](#_Toc51781417)

[3.1. Proceso de preparación de los datos 9](#_Toc51781418)

[3.2. Conclusiones 9](#_Toc51781419)

[4. Exploración de los Datos 10](#_Toc51781420)

[4.1. Proceso de exploración de los datos 11](#_Toc51781421)

[4.2. Conclusiones 11](#_Toc51781422)

[5. Presentación de Resultados 14](#_Toc51781426)

[5.1. Presentación de resultados 15](#_Toc51781427)

[5.2. Conclusiones 15](#_Toc51781428)

[6. Conclusiones 16](#_Toc51781429)

[6.1. Conclusiones 17](#_Toc51781430)

[6.2. Trabajo a Futuro 17](#_Toc51781431)

# **Objetivo de Investigación**

**Resumen:**   
Se pretende realizar una comparación dónde se pueda ver el número de viajes y si hubo algún impacto en el número de viajes dadas las implicaciones que tuvo la enfermedad COVID19, durante toda la pandemia, esta investigación no hablará de su progreso durante la pandemia si no una comparación de 15 meses entre sí para poder determinar si durante los picos más altos de la pandemia hubo algún impacto dentro del uso total del programa.

## Introducción En este documento se analizará el crecimiento del programa gubernamental previo a la pandemia y el impacto que la pandemia COVID19 causó de impacto a dicho crecimiento, dado a que a los inicios de las pandemias se obligó a que las personas no salieran a eventos no esenciales y muchas industrias tuvieron que mover su operativa a casa o incluso minimizar la producción al mínimo. Se cree que se va a notar un impacto global de número de viajes dado a los meses que duró este evento.

## Justificación Previo a la pandemia se estaba presentando un proyecto de crecimiento del programa MiBici incluyendo sin embargo la pandemia pudo haber causado un impacto para el presupuesto o el ingreso recabado.

## Problema Sería una totalidad de millones de viajes, por lo que se desconoce si se podrá manejar dicho volumen de datos.

## Hipótesis Se cree que el manejo de datos no va a ser un problema, así como presentar caídas de uso del proyecto. Además de esto, se cree que se estará viendo un impacto de uso en todas las estaciones puesto que el encierro fue en todo el estado, impactando trabajos y escuelas. Dicho encierro duró alrededor de un mes como medida de prevención de propagación, por lo que se debería de ver un impacto total en todo.

## Objetivos

#### Objetivo General:

Obtener una comparativa de 30 meses en total, dónde se comprenderá 15 meses previos y posteriores a la declaración de la pandemia para obtener un total de las 10 estaciones mas afectadas por la pandemia.

#### Objetivos Específicos:

1.- Definir las 10 estaciones con mayor usabilidad 15 meses previos a la pandemia.

2.- Definir las 10 estaciones con mayor usabilidad 15 meses posteriores a la declaración de la pandemia.

3.- Hacer una comparativa entre las estaciones, para después obtener las 10 estaciones mas afectadas.

## Novedad científica, tecnológica o aportación

Este proyecto ofrece comparar 15 meses sin importar los años que sean, por lo que se pueden hacer comparativas directas con la única necesidad de cumplir con un formato en el nombre del archivo.

# **Obtención de los Datos**

**Resumen:**

Los datos son obtenidos directamente del portal de MiBici, ya que ellos llevan datos que pueden ser descargados de forma gratuita, estos datos están en crudo por lo cual tienen que ser trabajados y limpiados. La única otra problemática es que tienen que ser combinados entre sí para obtener el nombre de la estación a la que se está refiriendo. No se necesitó de algún permiso ya que estos son parte de la transparencia que tiene el programa mismo.

## Proceso de obtención de los datos

Como parte de transparencia, los datos pueden ser encontrados directamente en el portal del programa. En el portal nada mas se tiene que filtrar el año y escoger los meses, por lo que se descargaron un total de 30 archivos en formatos CSV más 1 extra para obtener el nombre de cada estación.

## Conclusiones

MiBici es un programa gubernamental que se encuentra documentado de forma excelente por lo que no se tiene ningún problema en la obtención de datos si no en el filtrado, ya que podría ser implementado de una forma más eficiente, sin embargo solo funcionaría para la gente que necesita grandes volúmenes de meses y no por separado por lo que el uso actual es suficiente para su acceso y disponibilidad de la información ya que cada mes se publican los resultados, sin necesidad de comprometer información de sus usuarios.

# **Preparación de los Datos**

**Resumen:**

Después de la descarga de archivos CSV se les comenzó a cambiar el formato para dividirlos entre “previo a la pandemia” y “durante la pandemia” con la nomenclatura: “pre\_covid” y “dur\_covid” seguido de un número. El numero realmente no es relevante, ni tampoco un acomodo en específico sin embargo, se recomienda que para estas comparaciones se comience de forma ascendente acorde al mes y el año por si en algún futuro las siguientes versiones requieren comparar meses en específico entre si, sin embargo para completar el alcance actual no es requerido. Una vez realizado, se organiza la agrupación y limpiado de elementos vacíos para después compararlos entre sí.

## Proceso de preparación de los datos

Una vez leído los datos se organizan en un dataframe dividiéndolos nada mas entre “previo al COVID” y “durante al COVID”. Una vez organizando el dataframe se procede a agregar a integrar la información de la estación para que en vez de que muestre el número interno de estación se pueda mostrar el nombre de la estación. Por último, se procede a realizar una resta entre sí, algo muy importante es no organizar los datos puesto que pandas no reconoce las llaves primarias entre si, y procedió a realizarlo entre celdas, por lo que si dos celdas están en la misma posición no reconoce que pueden estar siendo referenciados diferentes. Además que a lo largo de la preparación de datos se tuvo que reducir el numero de datos y cambiar el alcance a un total de 30 meses posiblemente por el volumen de información causaba un error interno que no tenía que ver con la librería en sí, posiblemente ocasionaba un “overflow” pero no era reconocido como tal.

## Conclusiones

Pandas puede ser una excelente herramienta para obtener dataframes, y trabajarlos sin embargo se requiere muchas pruebas y conocimiento de los propios datos para poder visualizar cualquier cambio ya que no necesariamente esta información se comparte en la documentación de la librería, estos errores pueden ser catastróficos para cualquier proyecto ya que si los datos no son comprobados puede ocasionar inconsistencias bastante grandes. Por ejemplo, me mostraba que hubo una perdida de 50 mil viajes en la estación mas utilizada cuando en realidad era 7 mil. Los errores no necesariamente tienen que ser los que arroje la librería ya que me daba un error similar al que no existe el dataframe o incluso que el archivo no existe, pero después de reducir el scope y los archivos pudo correr sin problema.

# **Exploración de los Datos**

**Resumen:**  
A pesar de que fueron muestras con pocas columnas pero millones de nuestras, se pudieron realizar agrupaciones que ayudaron a realizar las comparaciones pertinentes entre las estaciones mas relevantes para así poder identificar su factibilidad, por lo mismo se pudo notar durante la presentación de los datos que

## Proceso de exploración de los datos

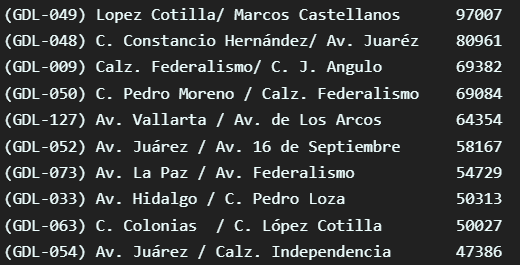
*[En esta sección* se detalla la manera como se ha realizado *una comprensión profunda de los datos. Se realiza la búsqueda de patrones, correlaciones y desvíos basados en técnicas visuales y descriptivas. Los conocimientos adquiridos en esta fase permitirán comenzar con el armado del modelo.]*

## Conclusiones

*[Se indica una breve conclusión de esta sección, donde se resalten los resultados que se obtuvieron y su relevancia para el proyecto.]*

# **presentación de resultados**

**Resumen:** En esta sección se presenta [escribir un pequeño resumen de un párrafo que explique el contenido de esta sección, donde se *obtiene una comprensión profunda de los datos. Se realiza la búsqueda de patrones, correlaciones y desvíos basados en técnicas visuales y descriptivas. Los conocimientos adquiridos en esta fase permitirán comenzar con el armado del modelo.*].

  
Text

Description automatically generated  
Text

Description automatically generated

## Presentación de resultados

*[En esta sección* se detalla la manera como se ha realizado *una comprensión profunda de los datos. Se realiza la búsqueda de patrones, correlaciones y desvíos basados en técnicas visuales y descriptivas. Los conocimientos adquiridos en esta fase permitirán comenzar con el armado del modelo.]*

## Conclusiones

*[Se indica una breve conclusión de esta sección, donde se resalten los resultados que se obtuvieron y su relevancia para el proyecto.]*

# **Conclusiones**

**Resumen:** En esta sección se presenta [escribir un pequeño resumen de un párrafo que explique el contenido de esta sección, donde se presentan las conclusiones generales del proyecto y el trabajo a futuro que puede ser explorado.].

## Conclusiones

*[En esta sección se presentan las conclusiones, las cuales deben responde a los objetivos establecidos.]*

## Trabajo a Futuro

*[Se refiere a recomendaciones o descripciones de uso, aplicaciones inmediatas que se derivan, o desarrollo de componentes o extensiones del desarrollo.]*

**Referencias Bibliográficas**

[Las referencias bibliográficas son empleando el estilo IEEE. *El orden de la bibliografía debe ser conforme a su aparición en el documento. Toda bibliografía puesta en esta sección deberá estar citada dentro del texto. La forma de hacer la cita cuando es una sola es [1], si son varias y salteadas es [3], [5], [8]. Si son varias consecutivas es [5]-[8]. Combinadas es [2]-[6], [9]. El formato de la bibliografía es como lista numerada simple, sin tabla. El formato de cada una es como se describe a continuación y dependiendo del tipo.]*

*[Sitio web]*

1. Alvarez, R. (2022, 11 enero). *Registro de usuarios de «MiBici» aumentó un 43% en 2021; la estación más usada se ubica cerca del Parque Rojo*. Milenio. <https://www.milenio.com/politica/comunidad/mibici-aumento-43-registro-2020-2021>
2. P. Curtis, "Mudding: Social Phenomena in text-based virtual realities," [Documento en línea] agosto 1992, [consultado el 30 de Agosto de 1996], Disponible en FTP: parcftp.xerox.com/pub/MOO/papers/DIAC921992.