



ANALÍTICA Y VISUALIZACIÓN DE DATOS

Transporte Urbano de Pasajeros

Profesor:
Alejandro López Gómez
Integrantes:

- Mendiola Gómez Juan
Carlos
- Molina Montaño Fidel
Uriel

ÍNDICE

1. INTRODUCCIÓN	2
2. OBJETIVOS	3
2.1 Objetivo General	3
2.2 Objetivos Específicos	3
3. METODOLOGÍA	5
3.1 Extracción de Datos	5
3.2 Transformación de Datos.....	5
4. PREPROCESAMIENTO DE DATOS	7
4.1 Análisis Exploratorio Inicial.....	7
4.2 Depuración de Variables.....	8
4.3 Etiquetado de Entidades y Municipios.....	9
4.4 Eliminación de Campos No Relevantes	10
4.5 Tratamiento de Valores Nulos	12
5. VISUALIZACIÓN DE DATOS EN TABLEAU	13
5.1 Preparación de Datos para Tableau.....	13
5.2 Creación del Dashboard	14
5.3 Aplicación de Filtros y Diseño Visual.....	15
5.4 Gráficas Complementarias (Barras, Mapa, Pastel).....	16
6. RESULTADOS Y ANÁLISIS	19
6.1 Hallazgos Principales	19
6.2 Análisis por Entidad Federativa	19
6.3 Evolución Temporal del Transporte.....	20
7. CONCLUSIONES Y RECOMENDACIONES	20

INTRODUCCIÓN

Dentro de este proyecto de la materia de Análitica y visualización de datos se planea realizar la búsqueda de un dataset de interés, que contenga suficientes datos y columnas, con el objetivo de conocer el conjunto de datos inicial, para definir el propósito por el cual fue elegido, posteriormente realizar el preprocesamiento necesario para la elaboración de un dashboard interactivo en alguna herramienta de visualización de preferencia que permita alojar el dashboard en la nube y finalmente realizar las conclusiones obtenidas a partir del dashboard interactivo final.

Objetivo general:

Analizar la evolución del transporte público en distintos estados de la República Mexicana, con el fin de identificar el incremento anual en el uso del servicio. Asimismo, se pretende determinar cuáles estados transportan mayor número de pasajeros y el tipo de transporte.

Objetivos específicos:

El transporte urbano de pasajeros representa un componente vital de la movilidad en las principales entidades federativas de México. Sin embargo, su evolución a lo largo del tiempo requiere ser analizada para entender mejor su comportamiento, evaluar políticas públicas de movilidad y tomar decisiones más informadas sobre la planificación del servicio.

En este estudio, se utilizarán datos públicos abiertos del INEGI para analizar la evolución del transporte urbano de pasajeros en distintos estados de la República Mexicana. Se identificarán patrones anuales, diferencias entre entidades, tipos de transporte más utilizados y se presentarán los resultados en un dashboard interactivo. Este análisis permitirá no solo visualizar la situación actual, sino también histórica que puede generar información útil para tomadores de decisiones, investigadores y ciudadanos interesados en la movilidad urbana.

Para ello, se plantean los siguientes objetivos específicos:

1. **Analizar la evolución temporal** del transporte público urbano desde 1986 hasta 2025, identificando incrementos o disminuciones en el número de pasajeros transportados por año, mes y por tipo de transporte.
2. **Identificar las entidades federativas** con mayor y menor volumen de pasajeros transportados, comparando su comportamiento a lo largo de los años y los tipos de transporte más relevantes en cada caso.

3. **Depurar y transformar los datos brutos**, incluyendo la integración de etiquetas a partir de catálogos complementarios (entidades, municipios y fechas), y eliminar campos irrelevantes para mejorar la calidad del análisis.
4. **Visualizar los datos de manera clara e interactiva**, mediante dashboards elaborados en Tableau, que incluyan mapas, gráficas de barras y gráficos de pastel que faciliten la comprensión por parte de distintos usuarios.
5. **Proponer recomendaciones con base en los hallazgos**, dirigidas a mejorar la planificación del transporte urbano, optimizar la cobertura de rutas y asignar recursos de forma más eficiente en las entidades con mayores retos de movilidad.

Extracción de datos:

Para este proyecto de Analítica y Visualización de Datos, se recopilaron datos públicos del Instituto Nacional de Estadística y Geografía (INEGI), específicamente del apartado de Información Económica. Se seleccionó el conjunto de datos correspondiente al transporte público, el cual contiene un total de 29,739 registros desde 1986 hasta principios de 2025. Este conjunto incluye información sobre número de pasajeros, rutas, unidades, entre otros aspectos relevantes. La fuente utilizada fue: <https://www.inegi.org.mx/datosabiertos/>.

Transformación de datos:

El primer conjunto de datos no cuenta con etiquetas descriptivas para las columnas, por lo que fue necesario identificar mediante el diccionario los valores que representa cada una. Estas variables como *mes*, *entidad* y *municipio*, se dispone de datasets complementarios que incluyen las etiquetas correspondientes, lo que facilita su interpretación y posterior análisis por lo que se hará el etiquetado de los datasets con el original.

Se tiene como primera instancia el dataset original, las primeras observaciones son la falta de etiquetas, pero igual se tiene distinto tipo de información registrada para cada tipo de transporte en la columna de 'VARIABLE' son agrupaciones de información por

cada año y tipo de transporte, la cual no es la misma para cada transporte, es información distinta, lo que a su vez implica cantidades distintas en la columna ‘VALOR’ por lo que se muestra una distribución muy sesgada.

Figura 1 dataset original con distinta información para columna ‘VARIABLE’



Se analizaron los valores únicos que tiene el dataset por cada columna, se observan los años desde 1986 hasta 2025, todos los posibles transportes, las variables que describen actividades, las cuales se descartarán algunas ya que no son de interés para el análisis, por otro lado la columna ‘ESTATUS’, contiene 3 posibles valores, pero son muy pocos datos en cifras preliminares y estimadas, por lo que esta columna no se tomará en cuenta y será eliminada.

Figura 2 valores unicos iniciales

	Data type	# Amount	Unique values	...
Ausente:	0 (0%)	Ausente:	0 (0%) Ausente:	0 (0%)
Distinto:	3 (38%)	Distinto:	8 (100%) Distinto:	8 (100%)
int64	50%		[1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...]	13%
object	38%		[1 2 3 4 5 6 7 8 9 10 11 12]	13%
float64	13%		['Red de Transporte de Pasajeros' 'Sistema de Transporte Colectivo Metro...' 'Otros']	63%
ANIO	int64	40	[1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000]	
ID_MES	int64	12	[1 2 3 4 5 6 7 8 9 10 11 12]	
TRANSPORTE	object	27	['Red de Transporte de Pasajeros' 'Sistema de Transporte Colectivo Metro' 'Tren Lige...	
VARIABLE	object	19	['Autobuses en operación de lunes a viernes' 'Autobuses en operación de sábado a...	
ID_ENTIDAD	int64	10	[9 19 14 15 22 13 8 11 21 12]	
ID_MUNICIPIO	int64	8	[999 39 14 48 19 20 114 1]	
VALOR	float64	15256	[4.20100e+03 2.60100e+03 2.71994e+07 ... 4.97095e+05 2.83552e+05 2.13543e+05]	
ESTATUS	object	3	['Cifras Definitivas' 'Cifras Preliminares' 'Cifras Estimadas']	

Catalogo para ID_entidad:

- 8 = Chihuahua
- 9 = Ciudad de México
- 11 = Guanajuato
- 12 = Guerrero
- 13 = Hidalgo
- 14 = Jalisco
- 15 = Estado de México
- 19 = Nuevo Leon
- 22 = Queretaro

Catalogo para ID_municipio:

Para el caso del catálogo de municipios, se requiere saber a que entidad esta asociado para concentrarnos en esos municipios, en todos los casos 999 se maneja como "otros municipios"

En el caso del catálogo para fechas, es tal cual un valor numérico ordinal, por lo que su etiquetado es el mismo que para entidades.

Figura 3 Se realizó un análisis sobre la cantidad de municipios con '999' ya que estos representan a municipios desconocidos.

```
municipios = df["ID_MUNICIPIO"].value_counts()[999]
print(f'Municipios sin nombre = {municipios}')
print(f'La suma de ambos = {round((municipios/total_filas)*100),2}%' )
✓ 0.0s
Municipios sin nombre = 12869
La suma de ambos = 43.27%
```

El valor 999 aparece 12869 representando el 43% casi el 50% de los municipios son desconocidos ahora se quiere conocer aquellas combinaciones para conocer de que entidades provienen estos municipios que no se les asignó un nombre.

Figura 4 se hizo un ciclo para conocer las combinaciones que contienen el valor de 999 ya que de ser muchos, se tomará la desición de no tomar en cuenta los municipios.

```
var_ENT = df["ID_ENTIDAD"].unique()

for ent in var_ENT:
    count = ((df['ID_ENTIDAD'] == ent) & (df['ID_MUNICIPIO'] == 999)).sum()
    print(f"Entidad {ent} con municipio 999: {count} fila(s)")

✓ 0.0s

Entidad 9 con municipio 999: 11145 fila(s)
Entidad 19 con municipio 999: 0 fila(s)
Entidad 14 con municipio 999: 0 fila(s)
Entidad 15 con municipio 999: 1724 fila(s)
Entidad 22 con municipio 999: 0 fila(s)
Entidad 13 con municipio 999: 0 fila(s)
Entidad 8 con municipio 999: 0 fila(s)
Entidad 11 con municipio 999: 0 fila(s)
Entidad 21 con municipio 999: 0 fila(s)
Entidad 12 con municipio 999: 0 fila(s)
```

Se observa que realmente estos valores desconocidos pertenecen solamente a la ciudad de Mexico con valor 9 y el estado de Mexico con valor de 15, por lo que se dejarán los municipios para el análisis ya que si dan una idea más acertada para las demás entidades de donde proviene el transporte, pues en la ciudad de México tiene transporte en toda la ciudad al igual que el Estado por lo que para estos dos simplemente se dejará el municipio como el mismo nombre del estado.

Ahora se realizará el etiquetado, una vez conociendo los valores del dataset.

Figura 5 etiquetado con diccionario para mes y entidad

```

# 2. Cargar el archivo con nombres de los meses
df_meses = pd.read_csv("tc_periodo_mes.csv")

# 3. Crear el diccionario de mapeo: {1: 'Enero', 2: 'Febrero', ...}
dic_meses = df_meses.set_index("ID_MES") ["MES"].to_dict()

# 4. Mapear los números a nombres en una nueva columna
df["MES"] = df["ID_MES"].map(dic_meses)
[9]    ✓  0.0s

df_entidad = pd.read_csv("tc_entidad.csv")

dic_entidad = df_entidad.set_index("ID_ENTIDAD") ["NOM_ENTIDAD"].to_dict()

df["ENTIDAD"] = df["ID_ENTIDAD"].map(dic_entidad)
[10]   ✓  0.0s

```

Para el caso de entidades y los periodos fue sencillo ya que solo se creo un diccionario de mapeo y se agregaron al dataset original

En el caso de municipios se utilizo la tecnica de merge porque los municipios dependian de la etiqueta de entidad es por esto que fue un proceso diferente.

Figura 6 merge para municipios

```

#ID_ENTIDAD, ID_MUNICIPIO, NOM_MUNICIPIO
df_municipio = pd.read_csv("tc_municipio.csv")

# Haz el merge sobre las dos llaves: ID_ENTIDAD e ID_MUNICIPIO
df = df.merge(
    df_municipio[["ID_ENTIDAD", "ID_MUNICIPIO", "NOM_MUNICIPIO"]],
    on=["ID_ENTIDAD", "ID_MUNICIPIO"],
    how="left"
)

# Renombra la columna para tener "MUNICIPIO"
df = df.rename(columns={"NOM_MUNICIPIO": "MUNICIPIO"})

# para "Otros municipios" se remplaza por el nombre de la entidad completa
mask = df["MUNICIPIO"] == "Otros municipios"
df.loc[mask, "MUNICIPIO"] = df.loc[mask, "ENTIDAD"]

```

Ahora se tiene la columna 'VARIABLE' la cual como se mencionó tiene campos que no nos interesan para el análisis, se realizó un análisis de las siguientes:

Figura 7 valores que toma la columna 'VARIABLE'

	<code>variables_del_dataset = pd.DataFrame(df["VARIABLE"].unique())</code>
✓	0.0s Abrir "variables_del_dataset" en Data Wrangler
	0
0	Autobuses en operación de lunes a viernes
1	Autobuses en operación de sábado a domingo
2	Kilómetros recorridos
3	Pasajeros transportados
4	Personal ocupado
5	Rutas
6	Energía eléctrica consumida
7	Longitud de servicio
8	Trenes en servicio
9	Ingresos por pasaje
10	Unidades en operación
11	Lineas en servicio
12	Pasajeros transportados con boleto pagado
13	Pasajeros transportados con cortesía
14	Unidades en existencia
15	Pasajeros transportados con descuento
16	Cabinas en operación de lunes a viernes
17	Cabinas en operación de sábado a domingo
18	Pasaieros transportados con transbordo

Son 18 campos, de los cuales 10 no son de interés para el análisis:

- * personal ocupado
- * longitud de servicio
- * trenes en servicio
- * Pasajeros trasportados con cortesía
- * Pasajeros transportados con boleto pagado

- * Unidades en existencia
- * Pasajeros transportados con descuento
- * Cabinas en operación de lunes a viernes
- * Cabinas en operación de sábado a domingo
- * Pasajeros transportados con transbordo

Figura 8 eliminación de filas con información que no nos interesa

```

df = df.drop(['ID_MES', 'ID_ENTIDAD', 'ID_MUNICIPIO', 'ESTATUS'], axis=1)

# Lista de variables que no nos interesan
variables_a_eliminar = [
    'Personal ocupado',
    'Longitud de servicio',
    'Trenes en servicio',
    'Pasajeros trasportados con cortesía',
    'Pasajeros transportados con boleto pagado',
    'Unidades en existencia',
    'Pasajeros transportados con descuento',
    'Cabinas en operación de lunes a viernes',
    'Cabinas en operación de sábado a domingo',
    'Pasajeros transportados con transbordo',
    'Pasajeros transportados con cortesía'
]

# Opción 2: Eliminar las filas que SÍ están en la lista (equivalente a la opción 1)
df = df.drop(df[df['VARIABLE'].isin(variables_a_eliminar)].index)
df

```

Despues de eliminar las filas que contenian estos campos, quedaron 23182 registros, por lo que se eliminaron al rededor de 6000 registros, las columnas de 'ESTATUS' y de IDs ya que ya teniamos los valores de estos.

Figura 9 dataset con filas y columnas eliminadas

	ANIO	TRANSPORTE	VARIABLE	VALOR	MES	ENTIDAD	MUNICIPIO
0	1986	Red de Transporte de Pasajeros	Autobuses en operación de lunes a viernes	4201.0	Enero	Ciudad de México	Ciudad de México
1	1986	Red de Transporte de Pasajeros	Autobuses en operación de sábado a domingo	2601.0	Enero	Ciudad de México	Ciudad de México
2	1986	Red de Transporte de Pasajeros	Kilómetros recorridos	27199400.0	Enero	Ciudad de México	Ciudad de México
3	1986	Red de Transporte de Pasajeros	Pasajeros transportados	172000400.0	Enero	Ciudad de México	Ciudad de México
5	1986	Red de Transporte de Pasajeros	Rutas	219.0	Enero	Ciudad de México	Ciudad de México
...
29732	2025	Qrobús	Autobuses en operación de sábado a domingo	781.0	Marzo	Querétaro	Querétaro
29733	2025	Qrobús	Kilómetros recorridos	213581.0	Marzo	Querétaro	Querétaro
29734	2025	Qrobús	Pasajeros transportados	497095.0	Marzo	Querétaro	Querétaro
29736	2025	Qrobús	Pasajeros transportados con cortesía	0.0	Marzo	Querétaro	Querétaro
29738	2025	Qrobús	Rutas	114.0	Marzo	Querétaro	Querétaro

23182 rows x 7 columns

Figura 10 valores únicos despues de etiquetado

ANIO	int64	40	[1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010]
TRANSPORTE	object	27	['Red de Transporte de Pasajeros' 'Sistema de Transporte Colectivo Metro' 'Tren Ligero' 'Trolebús' 'Sistema de Transporte Colectivo Metro']
VARIABLE	object	10	['Autobuses en operación de lunes a viernes' 'Autobuses en operación de sábado a domingo' 'Kilómetros recorridos' 'Pasajeros transportados' 'Rutas']
VALOR	float64	13020	[4.20100e+03 2.60100e+03 2.71994e+07 ... 7.81000e+02 2.13581e+05 4.97095e+05]
MES	object	12	['Enero' 'Febrero' 'Marzo' 'Abril' 'Mayo' 'Junio' 'Julio' 'Agosto' 'Septiembre' 'Octubre' 'Noviembre' 'Diciembre']
ENTIDAD	object	10	['Ciudad de México' 'Nuevo León' 'Jalisco' 'México' 'Querétaro' 'Hidalgo' 'Chihuahua' 'Guanajuato' 'Puebla' 'Guerrero']
MUNICIPIO	object	10	['Ciudad de México' 'Monterrey' 'Guadalajara' 'México' 'Querétaro' 'Pachuca de Soto' 'Chihuahua' 'León' 'Puebla' 'Acapulco de Juárez']

Se revisan los valores nulos dentro del dataset:

Figura 11 cantidad y porcentaje de valores nulos

```

missing_values = df.isnull().sum()
print(missing_values)
✓ 0.0s Abrir "missing_values" en Data Wrangler

ANIO      0
TRANSPORTE 0
VARIABLE   0
VALOR     156
MES        0
ENTIDAD    0
MUNICIPIO  0
dtype: int64


porcentaje_nulos = (missing_values / total_filas) * 100
porcentaje_nulos
✓ 0.0s Abrir "porcentaje_nulos" en Data Wrangler

ANIO      0.000000
TRANSPORTE 0.000000
VARIABLE  0.000000
VALOR     0.524564
MES        0.000000
ENTIDAD    0.000000
MUNICIPIO  0.000000
dtype: float64

```

Se tienen 156 valores nulos pertenecientes a la columna ‘VALOR’, representando tan solo el 0.52 % de los registros, para este caso son muy pocos nulos, además como se mencionó anteriormente la columna ‘VALOR’ depende de la columna ‘TRANSPORTE’ y la columna ‘VARIABLE’ por lo que no combiene simplemente llenar estos valores con la media ya que pueden salir los datos erroneos por lo que se opta eliminar estos nulos.

Por ultimo una vez realizado los procesos necesarios para el análisis, nos quedaron 23026 registros con 7 columnas el cual será transformado a un documento .xlsx para poder trabajar con la herramienta de TABLEAU y realizar el dashboard interactivo.

```

# Guardar como archivo Excel
df.to_excel("transporte_definitivo.xlsx", index=False)

0.0s

```

Figura 12 dataset final

	ANIO	TRANSPORTE	VARIABLE	VALOR	MES	ENTIDAD	MUNICIPIO
0	1986	Red de Transporte de Pasajeros	Autobuses en operación de lunes a viernes	4201.0	Enero	Ciudad de México	Ciudad de México
1	1986	Red de Transporte de Pasajeros	Autobuses en operación de sábado a domingo	2601.0	Enero	Ciudad de México	Ciudad de México
2	1986	Red de Transporte de Pasajeros	Kilómetros recorridos	27199400.0	Enero	Ciudad de México	Ciudad de México
3	1986	Red de Transporte de Pasajeros	Pasajeros transportados	172000400.0	Enero	Ciudad de México	Ciudad de México
5	1986	Red de Transporte de Pasajeros	Rutas	219.0	Enero	Ciudad de México	Ciudad de México
...
29732	2025	Qrobús	Autobuses en operación de sábado a domingo	781.0	Marzo	Querétaro	Querétaro
29733	2025	Qrobús	Kilómetros recorridos	213581.0	Marzo	Querétaro	Querétaro
29734	2025	Qrobús	Pasajeros transportados	497095.0	Marzo	Querétaro	Querétaro
29736	2025	Qrobús	Pasajeros transportados con cortesía	0.0	Marzo	Querétaro	Querétaro
29738	2025	Qrobús	Rutas	114.0	Marzo	Querétaro	Querétaro

23026 rows × 7 columns

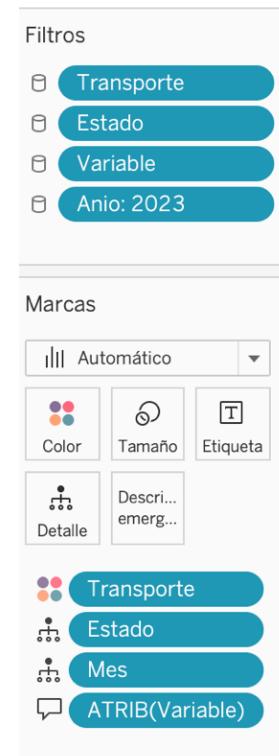
Elaboración del dashboard.

Figura 13 Se cargó el conjunto de datos:

Tablas	Tablas
Abc Entidad	Abc Anio
Abc MES	Abc Estado
Abc Municipio	Abc Mes
Abc Transporte	Abc Municipio
Abc Variable	Abc Transporte
Abc Nombres de medidas	Abc Variable
# Anio	Abc Nombres de medidas
# Valor	# Valor
# Sheet1 (Recuento)	# Sheet1 (Recuento)
# Valores de medidas	# Valores de medidas

Pero se cambiaron los nombres y el año lo pasamos a string para que no hiciera operaciones tableau.

Figura 14 Filtrado

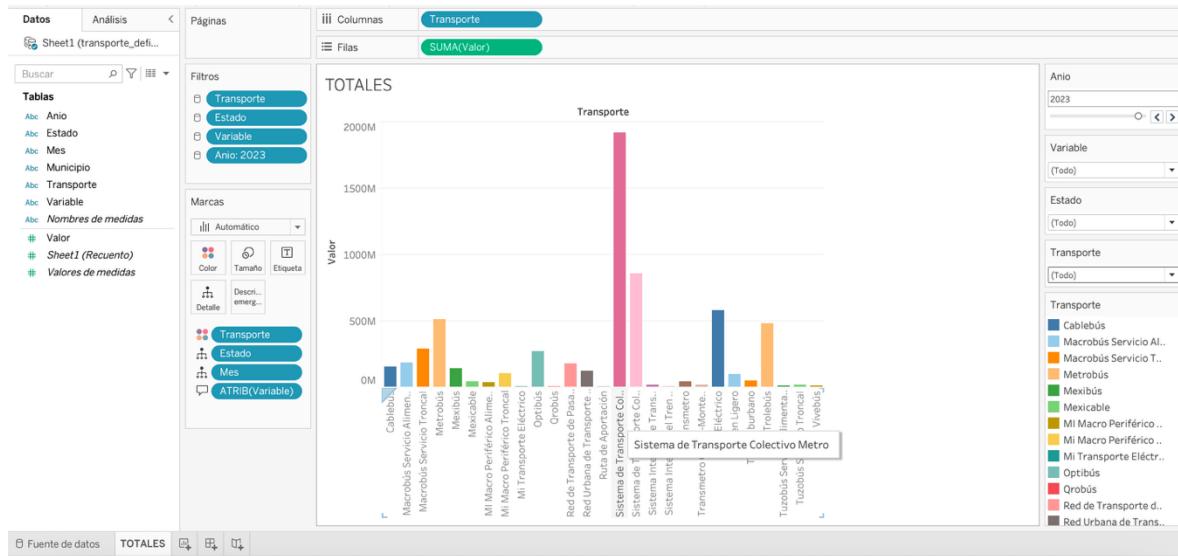


Los filtros ayudarán a conocer información específica del conjunto de datos, además los filtros serán agregados a la parte derecha del dashboard con diferente diseño dependiendo el dato que se desea seleccionar, el simbolo a su izquierda significa que afectarán a las graficas y el mapa que se mostrará proximamente.

Las marcas darán color a las gráficas y nivel de detalle, el simbolo jerárquico significa que se modificarán los resultados de la gráfica, siendo estos el mes y estado, la variable estará estática ya que solo mostrará la de interés.

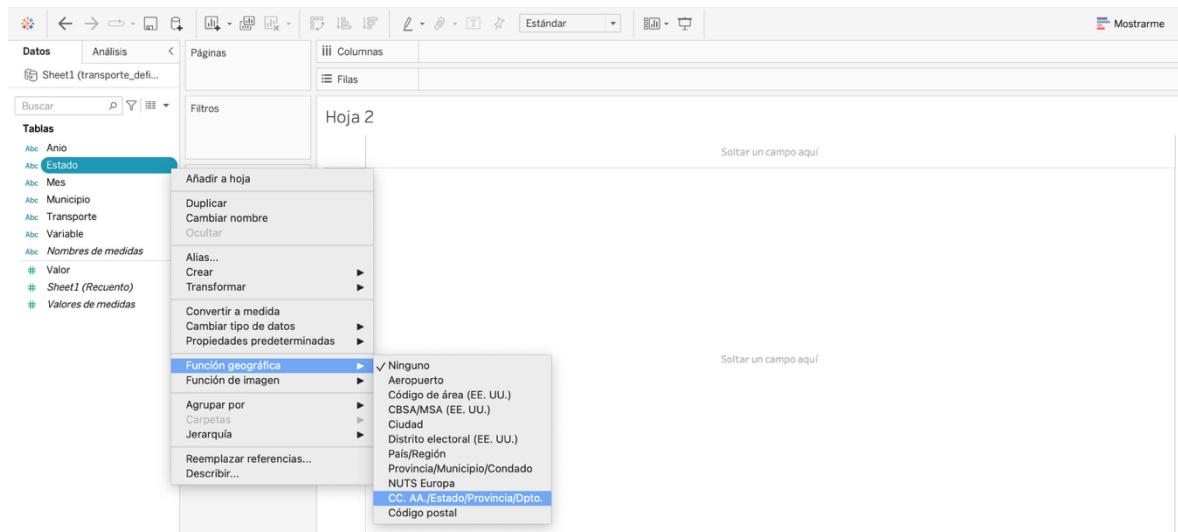
Así quedaría la primera gráfica la cual contiene todos los transportes, seleccionados en el año 2023 y la variable tiene todos los posibles campos, por lo que por ello muestra varias barras, algunas con mucha diferencia de valor pero es precisamente porque esta mostrando todo.

Figura 15 primer grafica con filtros en tableau



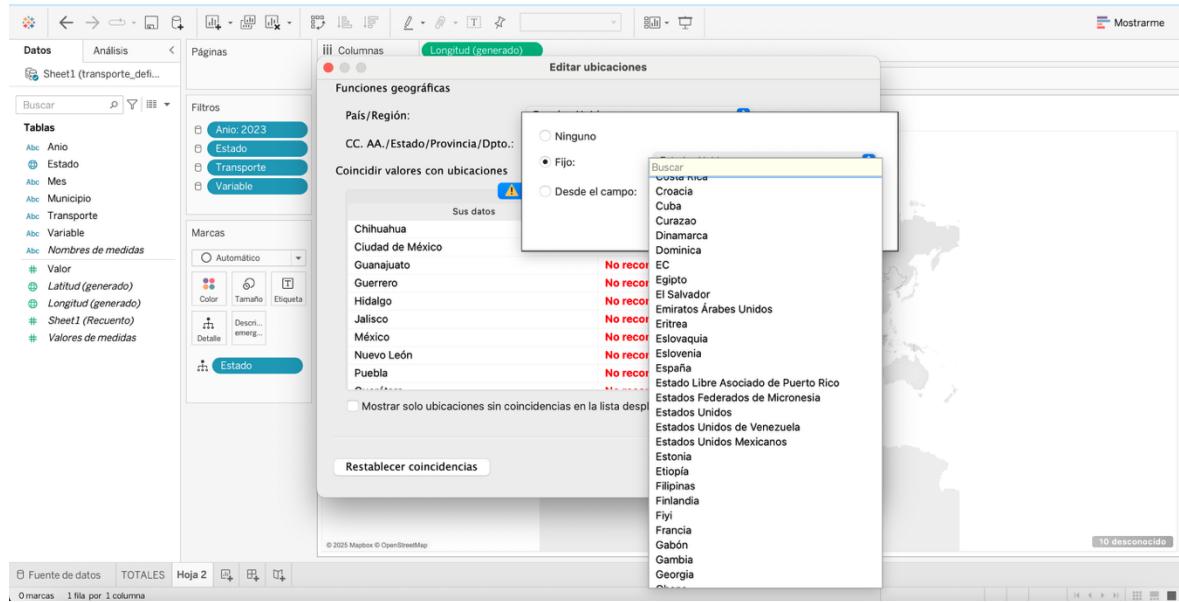
Como segundo paso se creó el mapa, pero para esto se tiene que cambiar el formato de la variable para que tableau pueda generar el mapa.

Figura 16 configurando el mapa en tableau



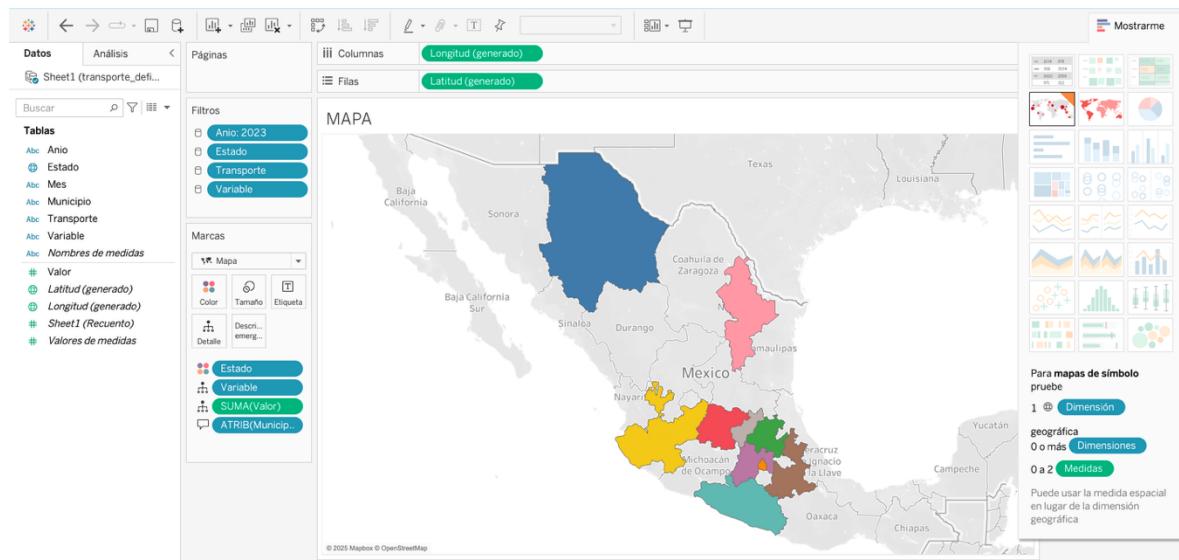
Se generó un problema ya que no mostraba nada, la razón era porque tenía asignado por predeterminado que los estados pertenecían a Estados Unidos y simplemente se hizo el cambio.

Figura 17 configuración para estados de la República Mexicana



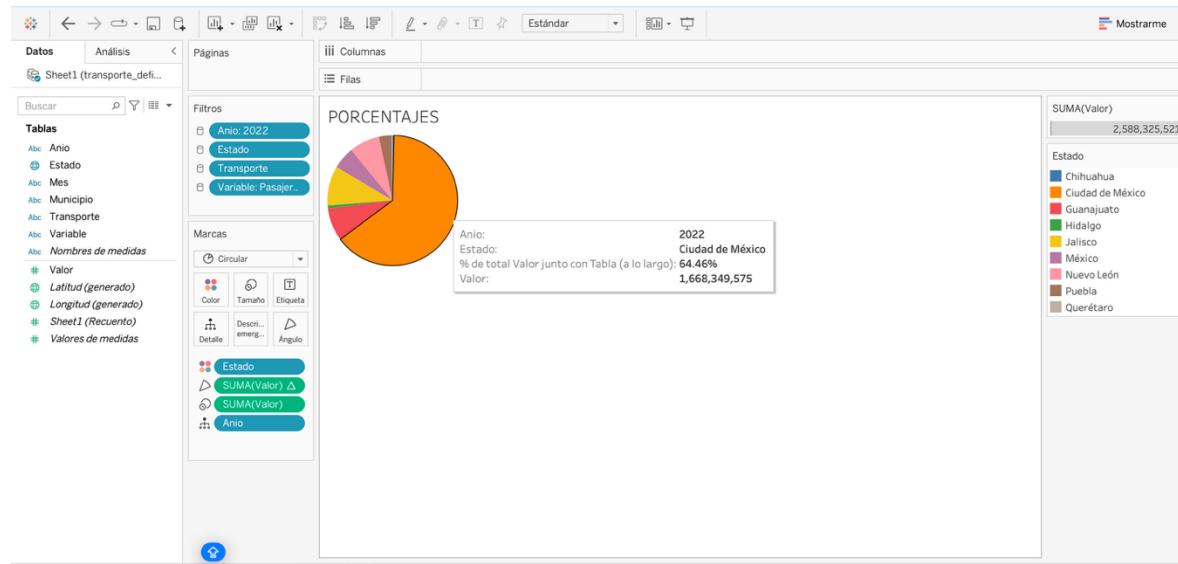
Este es el resultado final, el cual ya contiene los filtros por la condición mencionada anteriormente y de igual manera se ajustan los colores y nivel de detalle que mostrará, este es el resultado.

Figura 18 mapa final para el dashboard



Se realizó una gráfica de pastel con el fin de mostrar el porcentaje de la variable que queramos conocer, para este caso particular se elije el año 2022, nos muestra el porcentaje de personas transportadas en ese año, siendo la Ciudad de México la que mayor parte transportó.

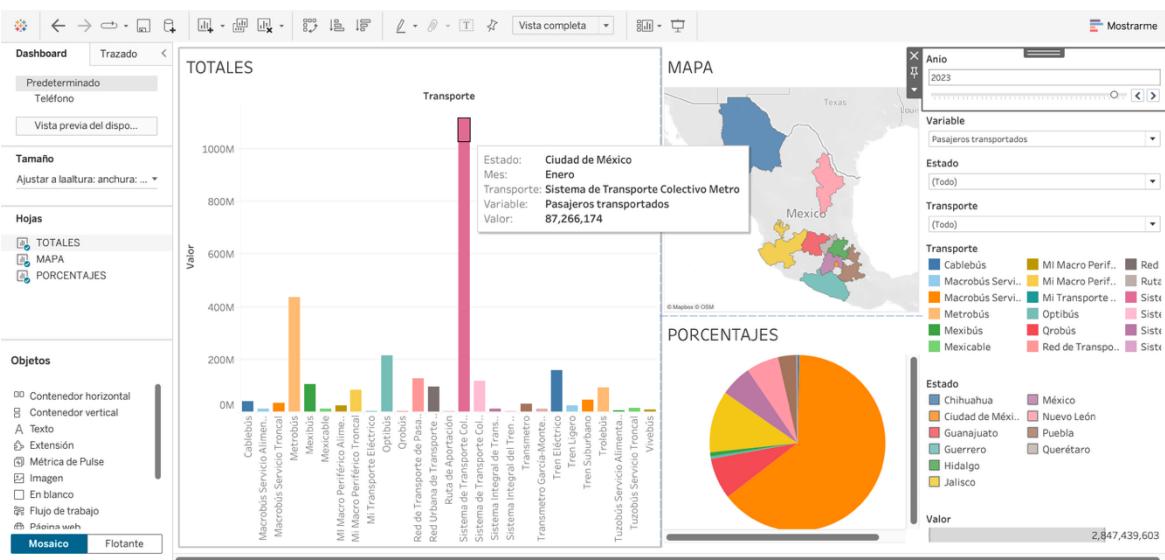
Figura 19 Creación de gráfica de pastel para dashboard



Por ultimo todos estos gráficos y el filtrado se agrega a una hoja para el dashboard y se ajusta su tamaño para darle espacio a cada gráfica y sea entendible.

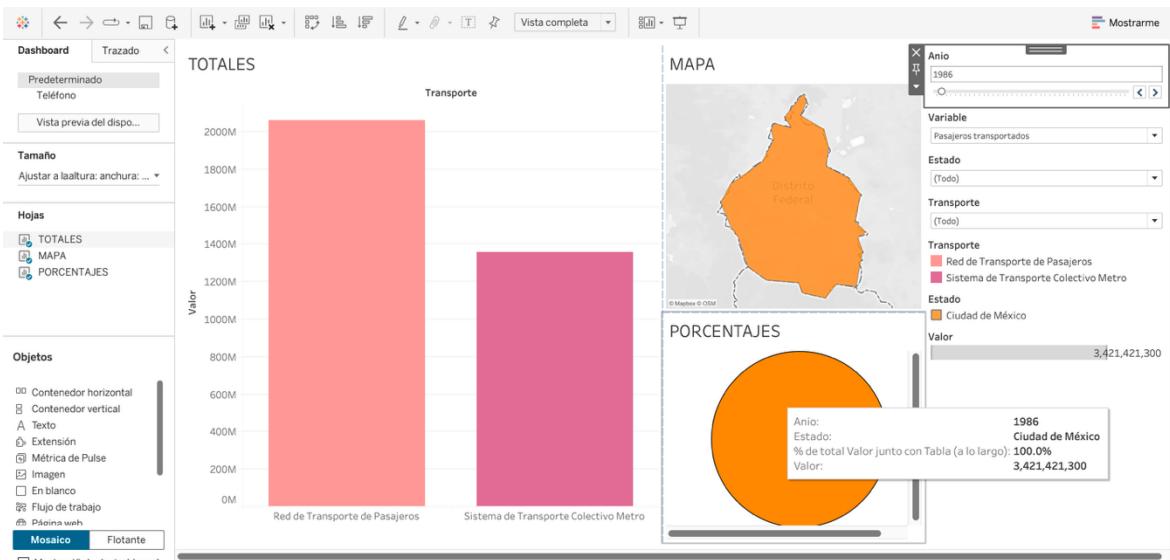
Para este caso se evalua la cantidad de personas transportadas, esto para el año 2023, la grafica de barras conforme se recorre va mostrando el mes, el nivel de detalle es la cantidad, el estado, el transporte, si el estado está en este intervalo de tiempo se pintatará el mapa ya que para el año 1986 solo hay registros de transporte para la ciudad de México.

Figura 20 Dashboard completado



Aquí está un ejemplo para el año 1986 solo hay registros de dos transportes y solo se ha marcado la ciudad en el mapa al igual que la gráfica de pastel ya que esta se cambia conforme a los estados.

Figura 21 Información particular en el dashboard



Conclusiones

El proyecto logró analizar la evolución del transporte urbano de pasajeros en varios estados de México a través de un dashboard interactivo en Tableau. Se utilizaron 29,739 registros públicos del INEGI, abarcando desde 1986 hasta principios de 2025. El preprocesamiento de datos fue crucial, incluyendo el etiquetado de columnas como "mes", "entidad" y "municipio", y la eliminación de la columna "ESTATUS" por su poca relevancia.

Un desafío significativo fue el 43% de municipios desconocidos (identificados como '999'), los cuales se determinó que pertenecían solamente a la Ciudad de México y el Estado de México, por lo que se mantuvieron para el análisis. Se eliminaron alrededor de 6,000 registros con información irrelevante de la columna 'VARIABLE', como "personal ocupado" o "longitud de servicio". También se eliminaron 156 valores nulos en la columna 'VALOR' (0.52% de los datos) debido a su dependencia de otras

columnas, lo que hizo que la imputación fuera inadecuada. El dataset final para Tableau quedó con 23,026 registros y 7 columnas.

En Tableau, se crearon filtros para año, estado y tipo de transporte para facilitar la exploración de los datos. El dashboard final integra un gráfico de barras, un mapa y un gráfico de pastel, permitiendo una visualización clara e interactiva de los datos.

Como punto final, el conocimiento extraído a partir del dashboard interactivo con respecto a datos históricos.

Se observó que la Ciudad de México es la única entidad con registros de transporte público desde 1986 hasta 1994 con la red de transporte de pasajeros, siendo estos los camiones verdes ahora con la simbología (RTP), junto con el sistema de transporte colectivo metro, transportando mensualmente alrededor de 172 millones de pasajeros y 110 millones respectivamente, notando que RTP era el que más transportaba, investigando a parte se encontró que el metro se inauguró en septiembre de 1969 por lo que son algunos años donde no se registraron datos y cabe destacar que existen hasta la fecha en que se hace el análisis.

Para el año 1995 se registró el estado de Nuevo León, en el municipio de Monterrey, con su transporte colectivo metro, con un aproximado de 3.4 millones de pasajeros mensualmente, indagando más, se encontró que se inauguró en 1991 por lo que para el año en que se registró en la base de datos ya tenía 4 años funcionando, igual para este mismo intervalo de tiempo la ciudad de México ya contaba con Trolebus y Tren ligero que transportaban aprox. 2.5 millones de pasajeros mensualmente cada uno.

Hasta el año 1997 se registró el estado de Jalisco, en el municipio de Guadalajara con el Trolebus y Tren ligero, con un total mensual de aprox. 4 millones y 1.2 millones respectivamente.

Para el año 2000 los pasajeros en RTP bajó, distribuyéndose quizás en el Tren ligero y Metrobus ya que de los 172 millones, ahora rondaba entre 10 y 12 millones y aún así no se equipara ya que para Tren ligero y Metrobus no pasa de los 2 y 8 millones respectivamente.

Para 2008 se registró el Estado de México con el Tren suburbano que investigando más a fondo su ruta era desde Buenavista hasta Cuautitlán y efectivamente se inauguró en Mayo de 2008 transportando en Junio aprox 900 mil pasajeros y a finales de ese año aprox 1.5 millones de pasajeros mensualmente.

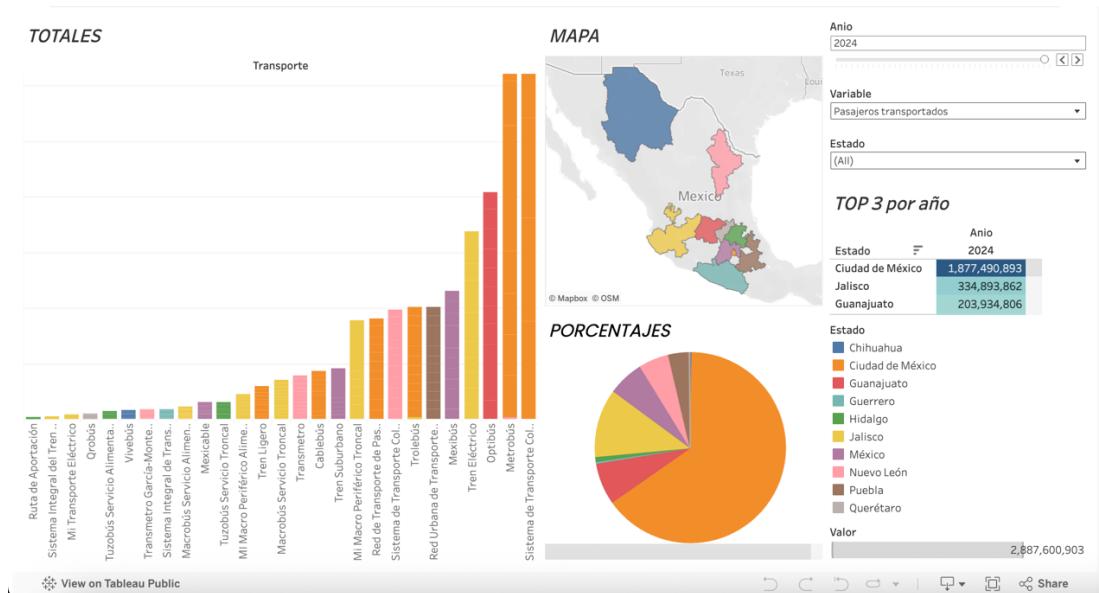
Para 2017 se registra el estado de Querétaro con el Querobus transportando cerca de 1 millón de pasajeros. En el siguiente año se registra Hidalgo en el municipio de Pachuca de Soto con el Tuzobus con servicio alimentador y troncal ambos transportando alrededor de 1.5 millones de pasajeros mensualmente.

Para 2021 se registraron los estados de Chihuahua y Guanajuato en el municipio de Leon, el Vivebus con 700 mil pasajeros mensuales y Optibus con hasta 16 millones respectivamente, quizás debiendo a que Chihuahua es un estado con mayor población casi el doble en 2020, en este intervalo de tiempo se registró el Mexibus en el estado de México transportando hasta 7 millones de personas mensualmente.

En 2022 se registro Puebla, con la red urbana de transporte articulado y en 2023 Guerrero en el municipio Acapulco de Juarez igualmente con Qrobus con un promedio de 200 mil pasajeros mensuales, hasta esta fecha ya estan todos los estados registrados, los 2 estados que mas transportan en 2023 son:

1. Ciudad de Mexico: ya que ocupa un 64% de la gráfica de pastel con un total de 1,828 millones de pasajeros en 2023.
 2. Jalisco ocupando un 11.34 % con 322 millones de pasajeros en 2022.

En 2024 se mantuvo parecido los numeros en pasajeros transportados, para 2025 aun en los primeros meses se sigue ocupando el mismo porcentaje siendo CDMX, Jalisco y Guanajuato los que mas pasajeros transportan actualmente.



Link de Dashboard interactivo:

https://public.tableau.com/app/profile/juan.mendiola/viz/proyecto_AyVD/Dashboard1