



**Universidad Internacional de La Rioja**  
**Facultad de Ingeniería y Tecnología**

**Máster Universitario en Análisis y Visualización de Datos Masivos / Visual**  
**Analytics & Big Data**

**Actividad 2**

<b>Actividad de estudio presentado por:</b>	<b>Juan David Escobar Escobar</b>
<b>Tipo de trabajo:</b>	<b>Actividad</b>
<b>Modalidad:</b>	<b>Individual</b>
<b>Director/a:</b>	<b>Luis Miguel Garay</b>
<b>Fecha:</b>	<b>Junio 2022</b>

## Índice de contenidos

### Table of Contents

<b>MÁSTER UNIVERSITARIO EN ANÁLISIS Y VISUALIZACIÓN DE DATOS MASIVOS / VISUAL ANALITICS &amp; BIG DATA.....</b>	<b>1</b>
<b>1. EXTRACCIÓN DE DATOS .....</b>	<b>3</b>
<b>2. CARGA DE DATOS.....</b>	<b>3</b>
<b>3. CREACIÓN DIMENSIÓN DE TIEMPO.....</b>	<b>4</b>
<b>4. ESTANDARIZACIÓN DE DATOS.....</b>	<b>4</b>
<b>5. MODELO DE DATOS.....</b>	<b>4</b>
<b>6. CONSTRUCCIÓN DE INFORME INMIGRACIONES (ENTRADAS).....</b>	<b>5</b>
<b>7. CONSTRUCCIÓN DE INFORME EMIGRACIONES (SALIDAS) .....</b>	<b>6</b>
<b>8. CONSTRUCCIÓN DE INFORME INMIGRACIONES Y EMIGRACIONES (ENTRADAS Y SALIDAS) .....</b>	<b>6</b>
<b>9. MÉTRICAS Y CÓDIGO DAX .....</b>	<b>7</b>

## 1. Extracción de datos

Se tomaron datos históricos entre los años 2012 – 2021, de los registros migratorios de fuente de datos abiertos de la plataforma Colombiana (<https://www.datos.gov.co/Funci-n-p-blica/Registro-de-activos-de-informaci-n-Migraci-n-Colom/jkmc-4xas>)

### Entradas extranjeras a Colombia 2012 -2021

Total, registros: 112.295

Fecha	Fecha_Numero	Nacionalidad	Codigo_Iso_3166	Femenino	Masculino	Total_Personas	Indefinido	Total	Latitud_Longitud
1/01/2999	-99	Otros	-99	-99	-99	-99	-99	-99	-99
1/1/2012	112012	Islas Alboran y Perejil	100	1	0	1		1	(4.697144,-74.140516)
1/1/2012	112012	Albania	8	1	2	3		3	(4.697144,-74.140516)
1/1/2012	112012	Alemania	276	11	16	27		27	(-4.197703,-69.941278)
1/1/2012	112012	Alemania	276	1	1	2		2	(0.25129,-76.875963)
1/1/2012	112012	Alemania	276	63	102	165		165	(0.814836,-77.662532)
1/1/2012	112012	Alemania	276	22	23	45		45	(10.408582,-75.538003)

### Salidas desde Colombia al extranjero 2012 -2021

Total, registros: 58.426

Mes	Mes_Numero	Dí	Fecha	Fecha_Numero	Destino	Codigo_Iso_3166	Femenin	Masculi	Indefinir	Total_Person	Latitud	Longitud	Latitud_Longitud
-99	-99	1/01/2999	-99	Otros	-99	-99	-99	-99	-99	-99	-99	-99	-99
Enero1	1	1	1/01/2012	112012	Afganistan	4	0	1		1	(10.886611,-74.776708)	(10.886611,-74.776708)	
Enero1	1	1	1/01/2012	112012	Afganistan	4	1	20		21	(4.697144,-74.140516)	(4.697144,-74.140516)	
Enero1	1	1	1/01/2012	112012	Alemania	276	1	0		1	(0.814836,-77.662532)	(0.814836,-77.662532)	
Enero1	1	1	1/01/2012	112012	Alemania	276	5	2		7	(10.445761,-75.516429)	(10.445761,-75.516429)	
Enero1	1	1	1/01/2012	112012	Alemania	276	3	2		5	(10.886611,-74.776708)	(10.886611,-74.776708)	
Enero1	1	1	1/01/2012	112012	Alemania	276	22	13		35	(3.541112,-76.384584)	(3.541112,-76.384584)	
Enero1	1	1	1/01/2012	112012	Alemania	276	946	800		2	(4.697144,-74.140516)	(4.697144,-74.140516)	
Enero1	1	1	1/01/2012	112012	Alemania	276	12	6		18	(6.171601,-75.427454)	(6.171601,-75.427454)	

## 2. Carga de datos

Se tomaron dos fuentes de datos para desarrollar el informe en Power BI, las cuales son:

- [Entradas de extranjeros a Colombia.xlsx](#)
- [Salidas de colombianos desde el territorio nacional.xlsx](#)

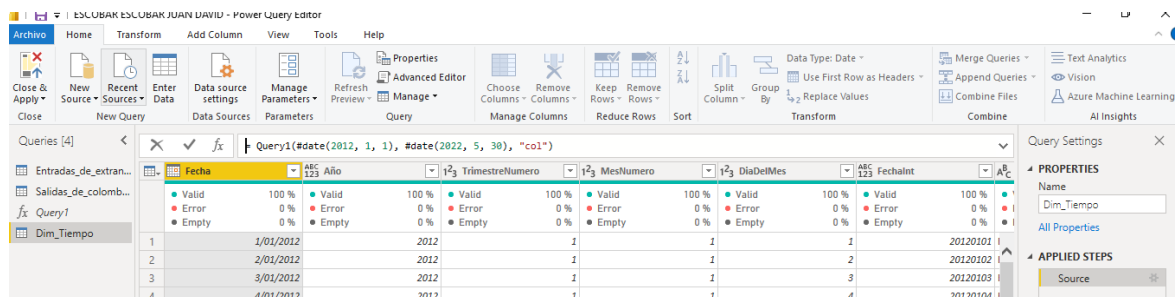
El modo de inserción de datos en el reporte fue por medio del conector a fuentes de datos tipo Excel en modo importación de datos, lo cual quiere decir que las fuentes de datos quedaran almacenadas en los metadatos del reporte, así que el reporte almacenara información de las gráficas configuradas y los datos importados, ya que al trabajar con fuentes de datos tipo Excel Power BI trabaja de manera import, esta información podría actualizarse en tiempo real, en un horario específico, para extraer más datos y actualizar el reporte. Otro punto importante por mencionar es que, al tener los datos en el mismo lugar del reporte, las consultas responden de manera más óptima que cuando se utiliza el mecanismo de consulta directa, ya que PowerBI, almacena dicha información en una infraestructura interna, en un Azure Data Lake y la sube a memoria cada vez que se consulta, mediante un cubo tabular de Microsoft Analysis Services que almacena la información en memoria.

Fecha	Año	TrimestreNumero	MesNumero	DiaDelMes	FechaInt	Mes	MesDelAño	TrimestreDelAño	DiaEnSemana	Dia	InicioSemana
domingo, 1 de enero de 2012	2012	1	1	1	20120101	Enero	Ene 2012	T1 2012	1	Domingo	domingo, 1 de enero de 2012
domingo, 8 de enero de 2012	2012	1	1	8	20120108	Enero	Ene 2012	T1 2012	1	Domingo	domingo, 8 de enero de 2012
domingo, 15 de enero de 2012	2012	1	1	15	20120115	Enero	Ene 2012	T1 2012	1	Domingo	domingo, 15 de enero de 2012
domingo, 22 de enero de 2012	2012	1	1	22	20120122	Enero	Ene 2012	T1 2012	1	Domingo	domingo, 22 de enero de 2012
domingo, 29 de enero de 2012	2012	1	1	29	20120129	Enero	Ene 2012	T1 2012	1	Domingo	domingo, 29 de enero de 2012
domingo, 5 de febrero de 2012	2012	1	2	5	20120205	Febrero	Feb 2012	T1 2012	1	Domingo	domingo, 5 de febrero de 2012
domingo, 12 de febrero de 2012	2012	1	2	12	20120212	Febrero	Feb 2012	T1 2012	1	Domingo	domingo, 12 de febrero de 2012
domingo, 19 de febrero de 2012	2012	1	2	19	20120219	Febrero	Feb 2012	T1 2012	1	Domingo	domingo, 19 de febrero de 2012
domingo, 26 de febrero de 2012	2012	1	2	26	20120226	Febrero	Feb 2012	T1 2012	1	Domingo	domingo, 26 de febrero de 2012

### 3. Creación dimensión de Tiempo

El desarrollo de una tabla tipo dimensional de tiempo, nos permite centralizar y estandarizar los datos de tiempo, como, por ejemplo, las fechas, los días, los meses, las semanas, las horas, segundos, semestres, trimestres, días feriados, entre otros. Tomaremos la dimensión de tiempo como nuestra tabla de referencia para la construcción de los filtros del informe.

#### Construcción dimensión de tiempo mediante editor de consultas (código-M)



### 4. Estandarización de datos

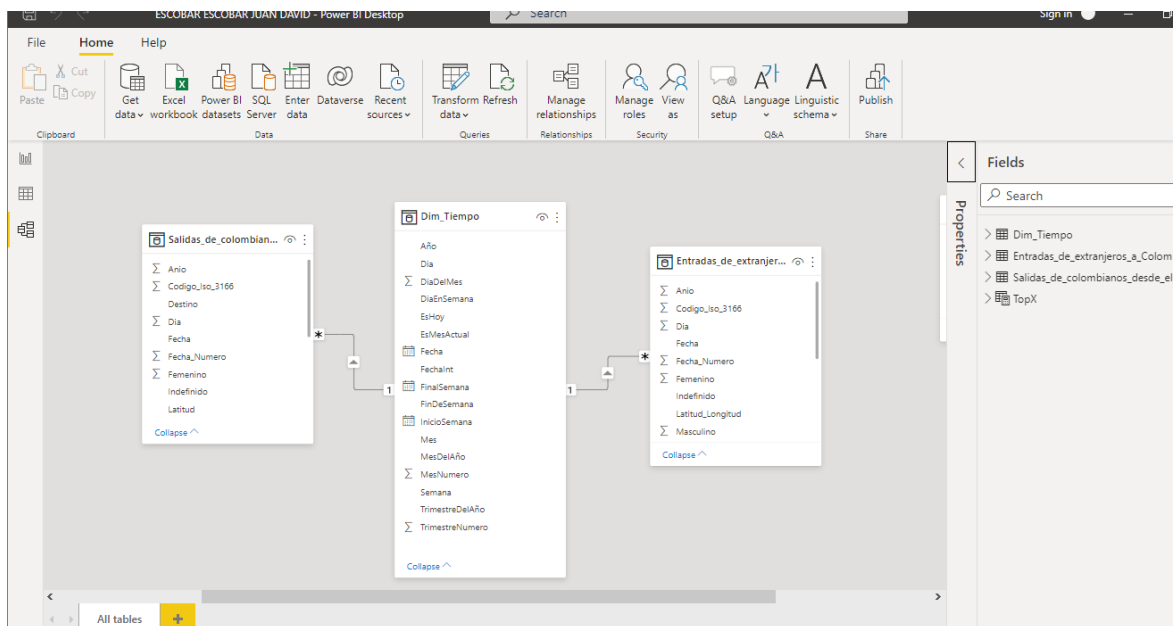
Las transformaciones o tipado de datos se optó por realizarse directamente desde las fuentes de datos en Excel, a pesar de que el editor de consultas de Power BI, nos brinda la posibilidad de efectuar dichas operaciones de transformación, tipado, ordenamiento, ETL y otros, al ser pocos datos se estandarizaron las fechas, se crearon las columnas calculadas y se tiparon los datos, ya que los tipos de datos de Excel, son fuertemente compatibles con los tipos de datos de Power BI, ya que pertenecen al mismo creador Microsoft.

Se agregó un registro Dummy en cada fuente de dato, llamado otros, el cual será utilizado en el desarrollo de algunos gráficos de Ranking o Top N.

Mes	Mes_Nume	Dí	Fecha	Fecha_Nume	Destino	Código_Iso_31	Femenin	Masculin	Indefin	Total_Person	Latitud	Longitud	Latitud_Longitud
-99	-99	-99	1/01/2999	-99	Otros	-99	-99	-99	-99	-99	-99	-99	-99

### 5. Modelo de datos

El modelo de datos es sencillo para la construcción de este informe, por lo general se acostumbra trabajar con sabanas de datos o vistas, las cuales surgen de la unión de varias tablas de una fuente de datos, en este caso contamos con 3 sencillas tablas, las cuales son las entradas, salidas del país y una dimensión de tiempo, ambas se relacionan por el campo fecha, ya que la dimensión de tiempo tiene un campo principal llamado fecha, el cual lleva el mismo formato que la tabla entrada y salida, y contiene desde la fecha mínima y máxima de las tablas entrada y salida, para poder relacionarlas y así aprovechar información adicional de tiempo, la cual puede ser útil para analizar los datos, mediante la construcción de gráficos o visualizaciones avanzadas en el informe.

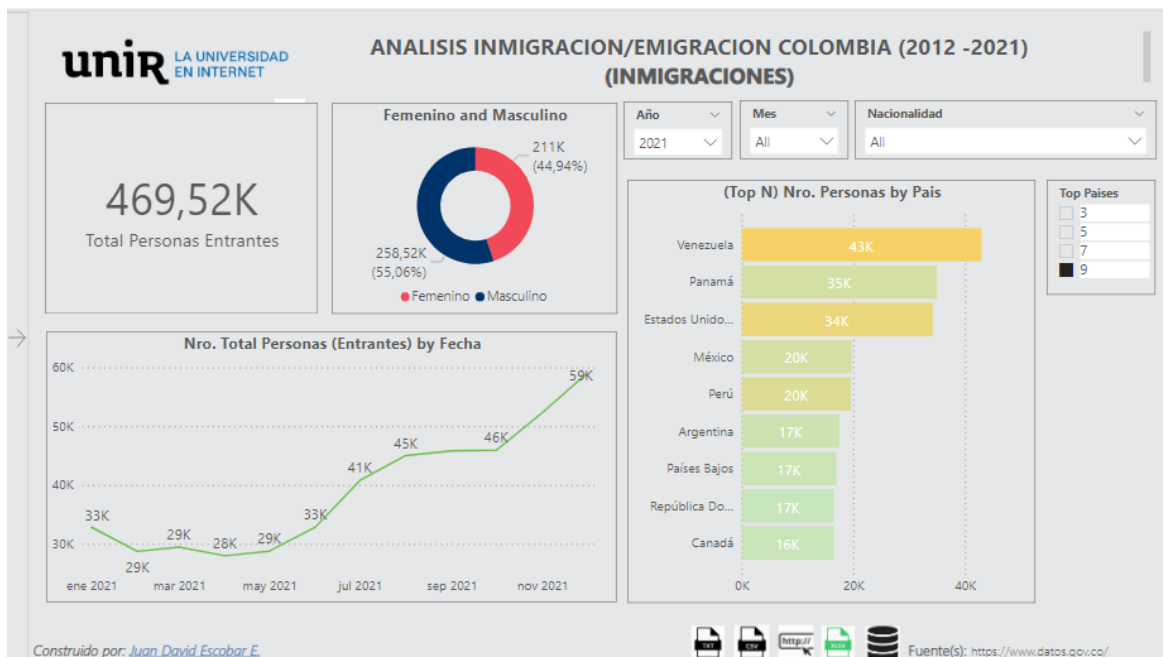


## 6. Construcción de informe Inmigraciones (Entradas)

Para la construcción del reporte, se tomó como base una plantilla que ya se tenía construida de reportes anteriores, la cual fue construida con una gama de colores, fuentes, y formatos de texto específicos, para estandarizar y mejorar la presentación del reporte, mediante un JSON Template de Power BI.

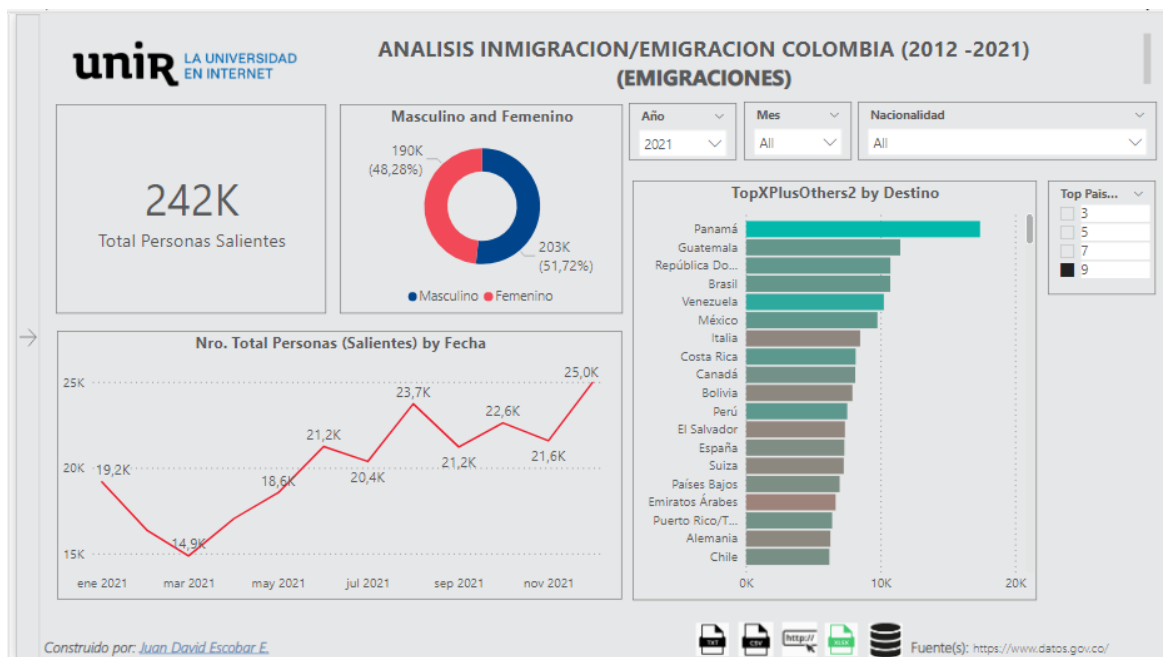
Para esta primera visualización se agregaron elementos como el Logo de la universidad, el título del Dashboard, el autor, las fuentes de datos, los filtros y las gráficas para analizar los datos.

Como dato más relevante ubicado en la parte superior izquierda se presenta una tarjeta con el conteo de personas que entraron a Colombia en un periodo de tiempo entre 2012 – 2021, adicional se agregaron filtros de tiempo, como lo es el año y el mes, y filtros por categoría nacionalidad y un selector del Top N, el cual solo afecta a la gráfica de Ranking, el reporte muestra la distinción entre hombres y mujeres que entran al país, teniendo un 10% más de hombres que de mujeres entre los años 2012 al 2021, pero este dato cambia a medida que interactuamos con los filtros y con los elementos interactivos de las gráficas. Otra grafica relevante es la gráfica de tendencia, que nos permite analizar en qué épocas o periodos de tiempo ingreso o tiende a ingresar más personas al país, por último se menciona la gráfica de ranking o top –N, donde claramente se puede evidenciar que los últimos 3 años la mayoría de inmigraciones a Colombia provienen de su país vecino Venezuela, lo cual posiblemente podemos relacionarlo a la crisis humanitaria y económica que tienen en Venezuela y a que Colombia es de los pocos países que ha abierto las puertas para brindar apoyo y ayuda humanitaria a estas personas.



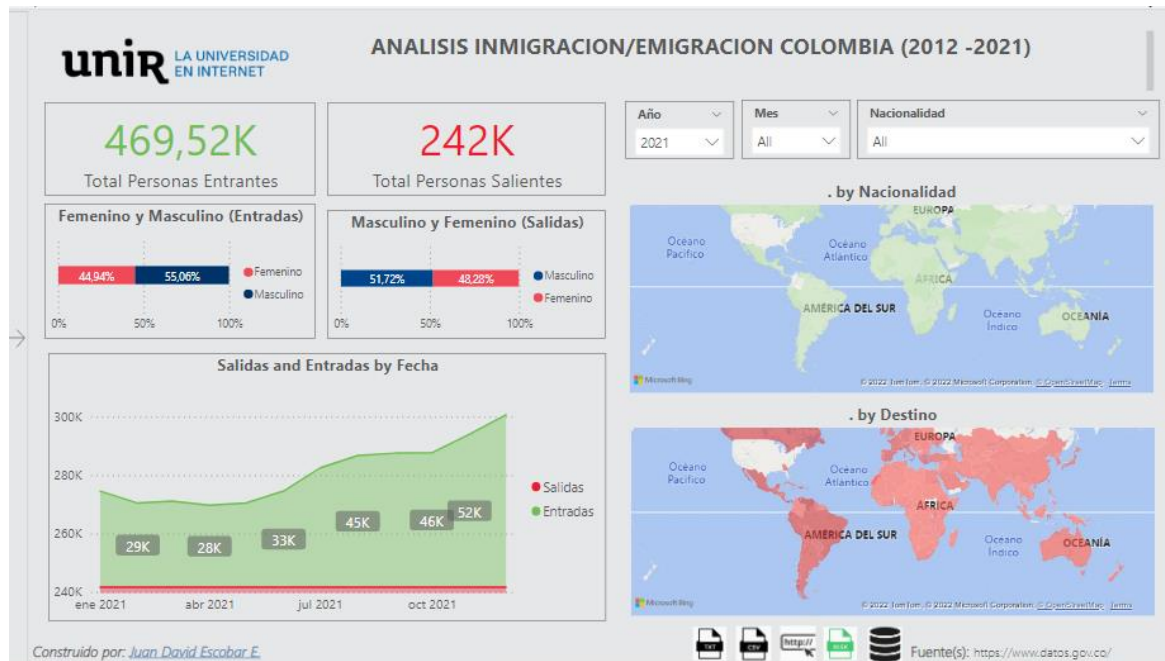
## 7. Construcción de informe Emigraciones (Salidas)

Para la construcción del reporte, se tomaron los mismos elementos que el Dashboard de entradas, ya que nos permite utilizarlo como punto de comparación. Podemos evidenciar una métrica de emigración total menor a la de inmigración entre los años 2012 – 2021, y un porcentaje de hombres y mujeres muy similar emigrando, hay una clara tendencia de aumento de emigraciones los últimos 3 años y por último se puede ver que Panamá es el país principal donde los colombianos emigran.



## 8. Construcción de informe Inmigraciones y Emigraciones (Entradas y Salidas)

Para la construcción del reporte, se tomaron elementos o gráficos de los reportes anteriores, con la finalidad de poder comparar las métricas entre las entradas y salidas del país en el tiempo. Como elemento adicional se agregó un mapa de calor, el cual nos permite visualizar aquellos países de donde provienen del exterior y de donde llegan las personas de Colombia.



## 9. Métricas y código DAX

A continuación, se detalla las métricas utilizadas y el código generado para la construcción de las mismas:

Metrica	DAX
	TopX = GENERATESERIES(3,9,2)
	SelectedTopX = MAX (TopX[Value] )
	TopX = GENERATESERIES(3,9,2)
	CantidadPersonas = SUM(Entradas_de_extranjeros_a_Colom[Total_Personas])
	Ranking = RANKX ( ALL ( Entradas_de_extranjeros_a_Colom[Nacionalidad] ), [CantidadPersonas],, DESC, DENSE )
	TopXPlusOthers = VAR SelectedTop = [SelectedTopX] RETURN IF ( [Ranking] <= SelectedTop, [CantidadPersonas], IF ( HASONEVALUE ( Entradas_de_extranjeros_a_Colom[Nacionalidad] ), IF ( VALUES ( Entradas_de_extranjeros_a_Colom[Nacionalidad] ) = "Otros", SUMX ( TOPN ( COUNTROWS ( ALL ( Entradas_de_extranjeros_a_Colom[Nacionalidad] ) ) - [SelectedTopX], ALL ( Entradas_de_extranjeros_a_Colom[Nacionalidad] ), [CantidadPersonas], 1 ), [CantidadPersonas] ) ) ) )