

# Proyecto Final Data Analytics



Profesor: Juan Creado

Tutora: Verónica Vissani

Cursante: Gabriela Ojcius

Fecha de presentación: Julio 2022

# Índice

1. Descripción de la temática de los datos .....	3
2. Hipótesis .....	4
3. Datasets .....	4
4. Diagrama Entidad-Relación .....	5
5. Listado de columnas por tablas .....	6
6. Transformaciones realizadas a los datos .....	8
8. Medidas calculadas generadas y sus fórmulas .....	10
9. Segmentaciones elegidas .....	12
10. Visualizaciones en el reporte .....	13
11. Futuras líneas .....	16

## 1. Descripción de la temática de los datos

Vivimos en un mundo digital. Desde que internet pasó a ser algo usual en nuestras vidas, muchas actividades cotidianas han cambiado, podemos comprar o vender en línea, comunicarnos con personas que están al otro lado del planeta, viéndolas y escuchándolas como si las tuviéramos al lado, investigar rápidamente, ver películas desde la comodidad de nuestro sofá, jugar en línea, por nombrar algunas.

Además, incluso la forma en que nos relacionamos con las personas ha mutado. Esta es justamente la potencia de la era digital que utilizaron algunas personas para crear aplicaciones que permitan conocerse con personas.

En este trabajo se analizarán algunos datos de usuarios de una plataforma global de citas en línea con la que se pueden conocer a personas para expandir el círculo social o conocer a la gente local. La aplicación considerada es Tinder, que cobró notoriedad en los últimos meses debido al documental *El estafador de Tinder*. Esta aplicación de citas utiliza tecnología de geolocalización para conectarse con perfiles basados en los filtros de género, distancia y orientación sexual que se establezcan.

Generar un perfil en esta aplicación es muy sencillo. Se solicitan datos como: nombre, fecha de nacimiento, género, orientación y preferencia sexual, ocupación, intereses, y finalmente imágenes. Otra opción consiste en vincular la cuenta con el perfil en otras redes como Instagram o Facebook. También se comparte una biografía en la que indicas justamente los motivos o interés de apertura de tu cuenta.

## 2. Hipótesis

El objetivo de este trabajo es indagar acerca de determinadas características, hábitos, intereses de los perfiles de los usuarios activos de la aplicación de citas Tinder. Entre las particularidades que pueden estudiarse es posible considerar:

- Cantidad de usuarios activos en la aplicación en un determinado período.
- Rango etario de los usuarios.
- Género y orientación sexual de los usuarios.
- Apertura de la cuenta por interés.
- Distribución por lugar de origen (país).

Partiendo del procesamiento y análisis de los datos, se busca obtener información más completa de los usuarios que utilizan esta red, para implementar nuevas estrategias de mercado que posibiliten el crecimiento de la empresa.

Por esta razón se implementa la construcción de un tablero de control para tomar decisiones basadas en la información que se encuentra almacenada en el dataset e identificar de forma fácil y rápida los indicadores que resulten de interés.

## 3. Datasets

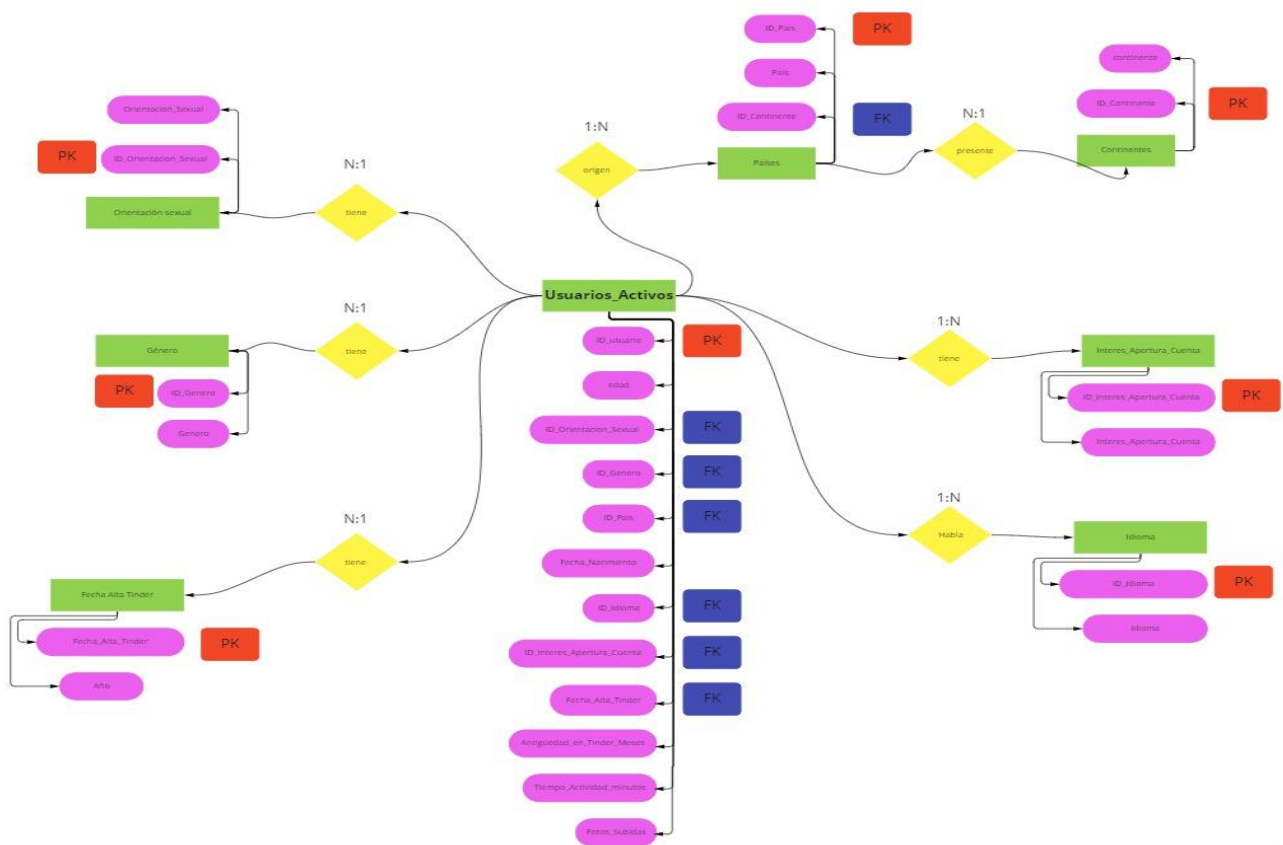
El dataset utilizado es particular, compartido por personas conocidas que tenían acceso a dicha información. A continuación, se comparte link de acceso a la base de datos utilizada para este trabajo.

<https://docs.google.com/spreadsheets/d/1FGvdm-FS1ixaOg1gTE-1gN154zcLUuTN/edit?usp=sharing&ouid=101529291502399680917&rtpof=true&sd=true>

Se aclara que no se mencionan nombres ni otros datos personales.

## 4. Diagrama Entidad-Relación

A continuación, se detalla el Diagrama de Entidad - Relación:



Obtenido desde: [https://miro.com/app/board/uXjV06ZjCss=](https://miro.com/app/board/uXjV06ZjCss=/)

miro

## 5. Listado de columnas por tablas

En este apartado se comparten las tablas y una breve descripción de estas y la definición de la clave primaria y foránea en cada una:

Tabla 1: Usuarios_activos		
Tabla con los datos de cada uno de los usuarios activos actualmente en la aplicación.		
Tipo de clave	Campo	Tipo de dato
PK - Index	ID_Usuario	Int
-	Edad	Int
FK	ID_Orientacion Sexual	Int
FK	ID_Genero	Int
FK	ID_Pais	Int
-	Fecha_Nacimiento	DateTime
FK	ID_Idioma	Int
FK	ID_Interes_Apertura_Cuenta	Int
FK	Fecha_Alta_Tinder	Datetime
-	Antigüedad_en_Tinder_Meses	Int
-	Tiempo_Actividad_minutos	Int
-	Fotos_Subidas	Int

Tabla 2: Interes_Apertura_Cuenta		
Tabla con identificador y descripción de todos los motivos de apertura de cuenta por parte de los usuarios de la aplicación.		
Tipo de clave	Campo	Tipo de dato
PK - Index	ID_Interes_Apertura_Cuenta	Int
-	Interes_Apertura_Cuenta	varchar(100)
-	Fotos_Subidas	Int

Tabla 3: Género		
Tabla con identificador y descripción de los posibles géneros de los usuarios de la aplicación.		
Tipo de clave	Campo	Tipo de dato
PK - Index	ID_Genero	Int
-	Genero	varchar(100)

Tabla 4: Orientacion_sexual		
Tabla con identificador y descripción de la orientación sexual de los usuarios de la aplicación.		
Tipo de clave	Campo	Tipo de dato
PK - Index	ID_Orientacion_Sexual	Int
-	Orientacion_Sexual	varchar(100)



**Tabla 5: Países**

Tabla con identificador y descripción de los países de residencia de los usuarios de la aplicación.

Tipo de clave	Campo	Tipo de dato
PK - Index	ID_Pais	Int
-	Pais	varchar(100)
FK	ID_Continente	Int

**Tabla 6: Continente**

Tabla con identificador y descripción de los continentes que componen a los países de residencia de los usuarios de la aplicación.

Tipo de clave	Campo	Tipo de dato
PK - Index	ID_continente	Int
-	Continente	varchar(100)

**Tabla 7: Idioma**

Tabla con identificador y descripción de los posibles idiomas de los usuarios de la aplicación.

Tipo de clave	Campo	Tipo de dato
PK - Index	ID_idioma	Int
-	Idioma	varchar(100)

**Tabla 8: Fecha Alta Tinder**

Tabla con la fecha de alta de los usuarios de la aplicación.

Tipo de clave	Campo	Tipo de dato
PK - Index	Fecha_Alta_Tinder	Int Datetime
-	Año	Int

## 6. Transformaciones realizadas a los datos

Se procedió a la carga y transformación del Dataset en Power BI.

Inicialmente de forma automática el programa generó relaciones entre las tablas, sin embargo, se presentó el inconveniente que por el formato del archivo del dataset se completaban muchas de las celdas con valores null. Esto trajo como inconveniente que las relaciones establecidas entre las entidades fuera del tipo muchos a muchos. Por esta razón, se eliminaron dichos valores y se pudo establecer relaciones adecuadas.

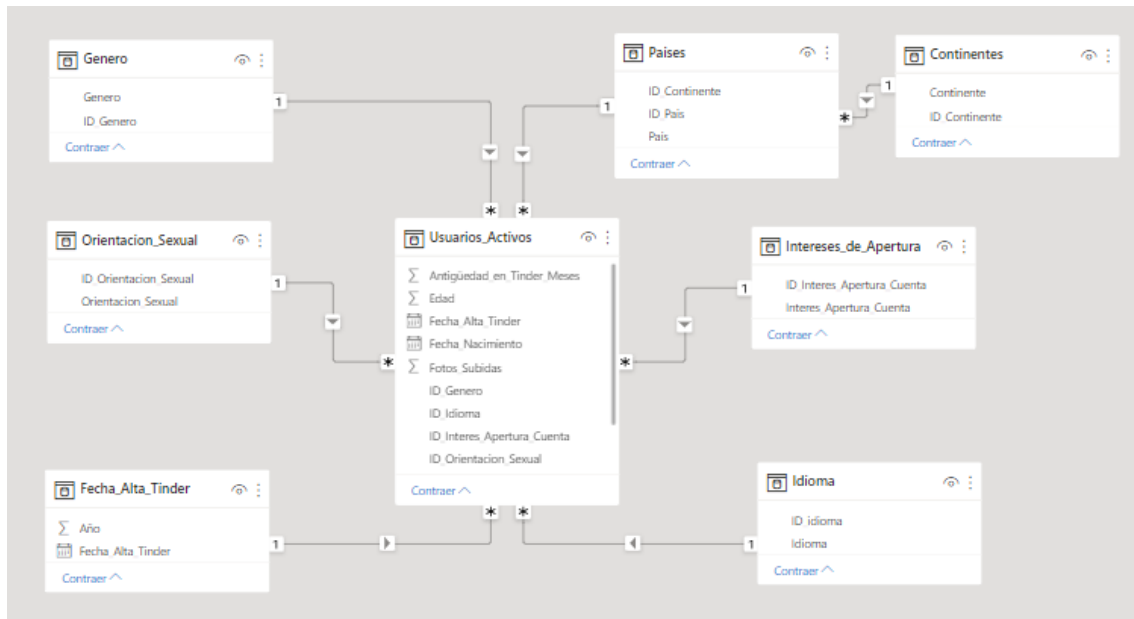
Se encontró que la tabla Fecha\_Alta\_Tinder no estaba relacionada con Usuarios\_Activos. Se procedió a realizar una transformación a través de Power Query donde se eliminaron valores duplicados permitiendo crear relaciones correctas.

También se realizaron otras transformaciones automáticas como constatar que el tipo de valores para cada tabla sea el correcto.

No fue necesario la creación de tablas puentes.



Finalmente, el modelo relacional es:



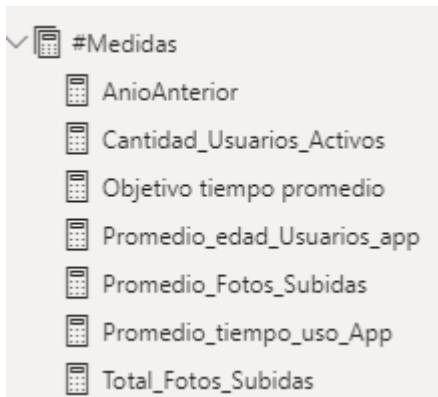
A continuación, se enumeran otras transformaciones manuales:

En la tabla "Fecha\_Alta\_Tinder" se agregaron las columnas "Mes", "Nombre del mes" y "Trimestre", se quitaron valores duplicados y se ordenaron las filas en ascendente por la columna inicial.

En la tabla "Usuarios Activos" se agregó una columna calculada "ANIO\_Alta\_Tinder"

## 8. Medidas calculadas generadas y sus fórmulas.

Se creó una tabla #Medidas para almacenar medidas calculadas.



La primera medida creada es un COUNT indicando la cantidad de usuarios de la aplicación que aparecen cargados en el Dataset considerado en el período de tiempo especificado 2017 a 2020. Se procede a realizar el cambio de tipo de dato para el ID\_Usuario a tipo texto.

`Cantidad_Usuarios_Activos = COUNT('Usuarios_Activos'[ID_Usuario])`

En el Dataset utilizado se cuenta con una tabla calendario, correspondiente a la fecha de alta de los usuarios en la aplicación. Se crearon columnas a partir de esta para realizar un análisis segmentado anual. Además, se consideraron los años y los meses con el fin de identificar los meses de mayor alta de perfiles para realizar una campaña fuerte de publicidad.

La idea obedece a la hipótesis que si las personas buscan mayor compañía en la época de festividades de fin de año (navidad, año nuevo) o en febrero por la fiesta de los enamorados.

Fecha_Alta_Tinder	Año	Mes_Alta_Tinder
viernes, 5 de junio de 2020	2020	junio
sábado, 20 de junio de 2020	2020	junio
sábado, 23 de mayo de 2020	2020	mayo
viernes, 27 de marzo de 2020	2020	marzo
lunes, 7 de septiembre de 2020	2020	septiembre
sábado, 8 de febrero de 2020	2020	febrero
sábado, 3 de octubre de 2020	2020	octubre
martes, 4 de febrero de 2020	2020	febrero
lunes, 14 de septiembre de 2020	2020	septiembre
domingo, 25 de octubre de 2020	2020	octubre
miércoles, 25 de noviembre de 2020	2020	noviembre
viernes, 8 de mayo de 2020	2020	mayo
domingo, 22 de marzo de 2020	2020	marzo
martes, 16 de junio de 2020	2020	junio
miércoles, 10 de junio de 2020	2020	junio
lunes, 27 de julio de 2020	2020	julio
sábado, 15 de agosto de 2020	2020	agosto
martes, 20 de octubre de 2020	2020	octubre
martes, 13 de octubre de 2020	2020	octubre
viernes, 6 de marzo de 2020	2020	marzo
miércoles, 5 de febrero de 2020	2020	febrero
domingo, 12 de abril de 2020	2020	abril
sábado, 22 de agosto de 2020	2020	agosto

Para conocer otras características de los usuarios se indagó sobre el promedio de edad:

```
Promedio_edad_Usuarios_app = AVERAGE ('Usuarios_Activos'[Edad])
```

Para medir el tiempo promedio de uso de la aplicación por parte de cada usuario, diariamente, medida en minutos, se procedió con la siguiente fórmula:

```
Promedio_tiempo_uso_App = AVERAGE ('Usuarios_Activos'[Tiempo_Actividad_minutos])
```

Motiva la obtención de esta medida ya que facilitará ver el grado de interés que despierta la app en el público.

Para conocer la cantidad de fotos compartidas por los usuarios en la aplicación, para eso se utilizó la siguiente medida:

```
Total_Fotos_Subidas = SUM (Usuarios_Activos[Fotos_Subidas])
```

Asimismo, se consideró el número promedio de fotos subidas por cada usuario en el período de tiempo de uso de la aplicación indicado en el data set:

```
Promedio_Fotos_Subidas = DIVIDE ([Total_Fotos_Subidas], '#Medidas'[Cantidad_Usuarios_Activos], 2)
```

## 9. Segmentaciones elegidas

El tablero cuenta con una portada que cuenta con tres botones que, al presionarlos, permiten acceder a las otras páginas con información relacionada. Asimismo, cada una de las páginas, posee botones que permiten navegar entre las demás páginas y volver a la portada.

Fueron consideradas las segmentaciones:

- En la página “Perfil de usuarios activos” se describen a los usuarios y a algunas actividades que realizan estos, segmentados por género y orientación sexual y también por año de alta.
- En la página “Localización de usuarios” se estudian la ubicación geográfica de los usuarios con un mapa interactivo. Además, se relacionan la cantidad de usuarios por continente y las altas anuales de éstos también por continente. Si se desea considerar cuáles países tienen más usuarios se visualiza un gráfico de embudo con el Top 5. También se utilizan dos segmentadores por género y por orientación sexual.
- Finalmente, en la página “Potenciales usuarios” se visualizan los intereses de apertura de las cuentas que indicaron los usuarios, las altas anuales analizadas por idioma y por género. Se agregan los filtros para visualizar por género o por intereses en la apertura.

## 10. Visualizaciones en el reporte

Se dio formato a todos los componentes del reporte, se unificaron las fuentes, los íconos y se agregaron bordes y sombras a los gráficos. Se trabajó con una paleta de color, con el fin de armonizar y que sea más agradable la información compartida.

Los gráficos considerados son los de tipo barras y de barras agrupadas, de línea para mostrar evolución a largo del tiempo, mapas, gráfico de embudo. Además en cada página se agregaron segmentadores, de texto con su menú desplegable y también con rango.

Se renombraron valores o variables en varios gráficos y se corrigieron errores ortográficos.

### Portada

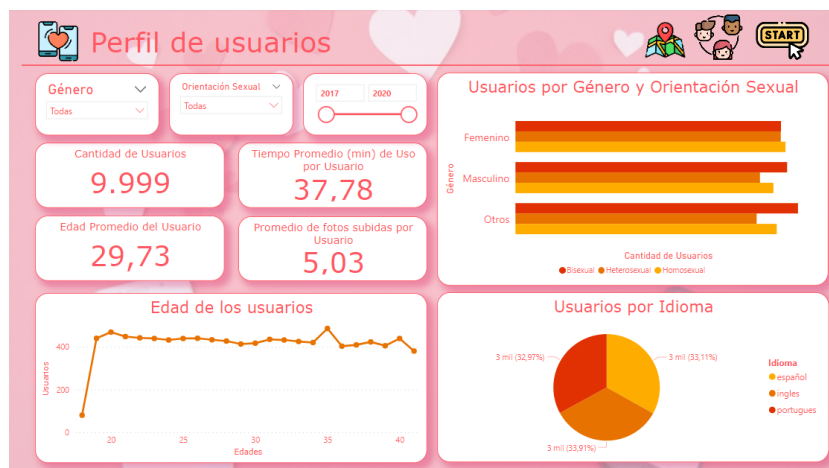
El tablero cuenta con una pantalla principal o portada "Start" que tiene tres botones que permiten navegar por las otras páginas mostrando la información relacionada a una temática particular estudiada.



Las siguientes páginas cuentan con botones de navegación en la parte superior derecha permitiendo direccionarse a cada una de las otras hojas o regresar a la portada inicial.

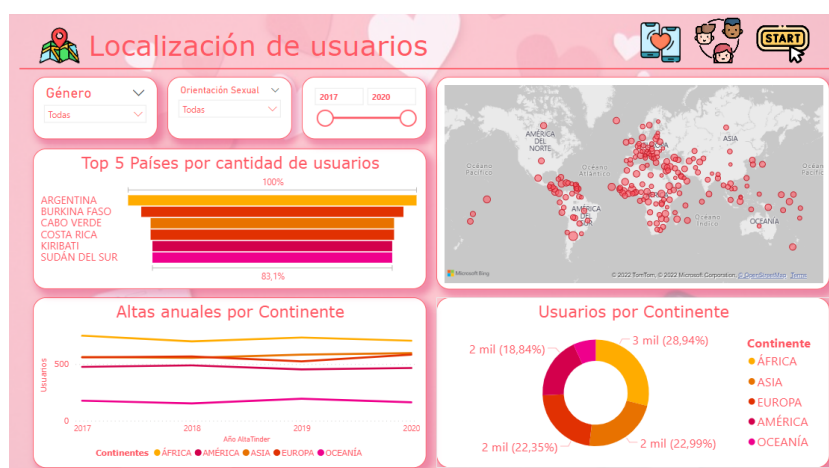
### Perfil de usuarios

En esta hoja se pueden visualizar datos específicos de los usuarios de la aplicación, como ser la cantidad estudiada, las altas anuales por mes, la cantidad de usuarios segmentados por género y orientación sexual y los por idioma. Además, se cuentan con segmentadores de orientación sexual, género y año de alta.



### Localización de Usuarios

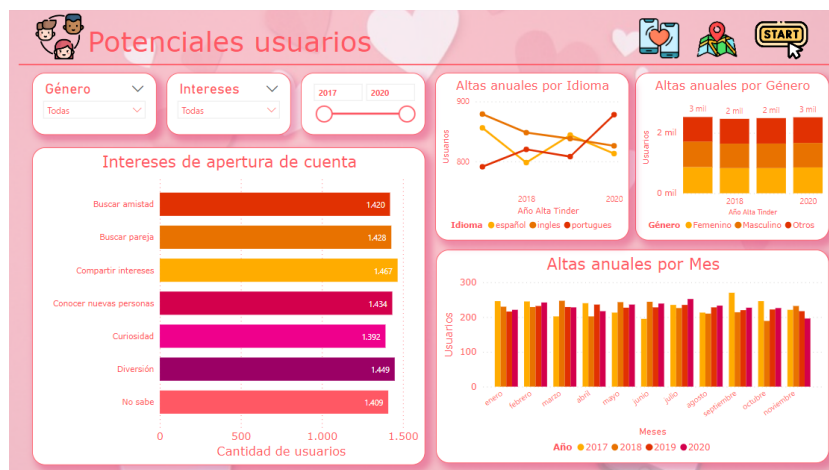
Con esta hoja se persigue conocer la distribución local de los usuarios de la aplicación. También se cuenta con gráficos que relacionan las altas anuales por continente y segmentadores por género, orientación sexual y año de alta.





Potenciales usuarios

En esta hoja se comparte información relacionada con las altas de los usuarios, segmentadas por género, idioma y por mes, en busca de algún patrón de altas de los usuarios según estos ítems, por ejemplo, si existen meses en los cuales las altas aumenten (como ser fiestas de fin de año, etc. que fue mencionado anteriormente). Asimismo, se comparte un gráfico en el que visualizan los intereses de apertura de la cuenta de los usuarios estudiados. También están presentes segmentadores por género, intereses y año de alta.



## 11. Futuras líneas

A partir del trabajo realizado, se considera que el Dashboard podría ampliarse para incorporar otras informaciones relevantes a la gerencia, entre las que podría mencionarse:

- Facturación de la aplicación generados por usuarios que poseen suscripción premium, también segmentandos por año, y por tipo de suscripción y continente de residencia de usuarios, ya que la misma puede variar según la ubicación. (Estos datos no fueron considerados en el dataset trabajado, pero se cuentan con los mismos).
- Considerar algunas métricas de tendencia de mercado, que permitan comparar la evolución de la matriculación de los usuarios en relación con lo ocurrido en otras aplicaciones similares.