

A continuación se presentan 3 temas con sus respectivas opciones de datasets e indicadores para su consideración.

## Migration

Me llamó la atención encontrar un [dataset](#) con décadas de información, en este caso sobre la población mundial por país y año, además que se podría complementar o tratar de encontrar correlación con la data de [indicadores de desarrollo mundial](#), siendo la motivación o utilidad de este tipo de análisis identificar los focos de migración y sus posibles razones.

Algunas métricas que se podrían obtener son:

1. Países con mayor cantidad de población y densidad de población.
2. Saldo migratorio por país y año.
3. Niveles de ingreso por país.
4. Correlación entre los niveles de ingreso de un país y su saldo migratorio.
5. Correlación entre el nivel de ingreso de un país y sus principales indicadores de desarrollo.

También vi un dataset de migración enfocados a Estados Unidos y sus fronteras ([US border crossing](#)) el cual me parece interesante cruzarlo o complementarlo con los arrestos a inmigrantes en dichas fronteras ([Illegal Immigrants Arrested by US Border Patrol](#)) para obtener métricas como:

1. Frontera con más tránsito.
2. Principales formas de cruzar las fronteras.
3. Promedio diario de personas cruzando la frontera.
4. Días o fechas con mayor concurrencia.
5. Cantidad de inmigrantes arrestados por frontera y año.
6. Proporción de arrestos del total de transeúntes por año y frontera.

La motivación o utilidad de este tipo de análisis sería para monitorear el tránsito en las fronteras y así distribuir recursos de seguridad acorde.

## Police Shooting

Me parece muy interesante esta colección de datos debido a la relevancia del tema en la actualidad sobre la brutalidad policial, en este caso en Estados Unidos ([Data Police Shootings](#)), el análisis sería útil para monitorear los indicadores más descriptivos de estos hechos por demografía y así alertar sobre las zonas más propensas a esto y tomar medidas.

Algunas de las métricas que se podrían obtener son:

1. Cuantas personas fueron disparadas sin estar armadas o presentando síntomas de enfermedad mental.
2. Distribución de los disparados por estado, edad y género.
3. Totalidad de disparados por año y raza.
4. Proporción de los disparados según su nivel de amenaza o de fuga.

## Services

A continuación se presentan opciones de servicio entretenimiento y recreación, que podría resultar interesante medir su desempeño y demanda para así personalizar su oferta, ampliar su catálogo de productos, mejorar sus sistemas de recomendación o experiencia de usuario.

### Netflix

Me pareció interesante un dataset sobre el contenido de [Netflix](#) para observar la distribución del mismo y su crecimiento durante los años, identificando las siguientes métricas:

1. Promedio de tiempo que tarda netflix en ofertar contenido de terceros.
2. Proporción de contenido original y de terceros en el tiempo.
3. Distribución de contenido por país y audiencia.
4. Distribución de las series por duración.
5. Proporción de series posiblemente canceladas (sólo 1 temporada).

### Hotel Booking

Me pareció interesante este [dataset](#) por la cantidad de indicadores que registra, los cuales se podrían utilizar para destacar las siguientes métricas:

1. Proporción de reservas canceladas en el tiempo.
2. Distribución de la estadía: en días de la semana o en fin de semana.
3. Cantidad de huéspedes que repitieron la experiencia.
4. Proporción de asignación de habitación diferente a la reservada.
5. Promedio de cambio en las reservaciones.
6. Promedio de peticiones especiales por usuario.
7. Relación entre el cambio en la reservación o peticiones especiales y el tipo de usuario.

### BikeSharing

El tema de mi TFM (Trabajo Final de Master) era realizar un estudio sobre el uso de CitiBike, el programa de intercambio de bicicletas más grande en los Estados Unidos, por lo que poder analizar la data de un sistema parecido en otras ciudades me llamó la atención, ya que tendría a CitiBike como referencia y trataría de obtener las siguientes métricas:

1. Distribución de los viajes por tipo de usuario.
2. Estaciones más usadas.
3. Rutas más usadas.
4. Relación del servicio con el clima y los accidentes de bicicletas. (depende si encuentro datos del clima y los accidentes para cada ciudad, pero me encantaría encontrar factores externos que influyan en el servicio y ver su correlación).