# Objetivos de ETL:

Carga de datos en crudos.
Archivos CSV a parquet.
Sacar nulos.
Duplicados.
Datos mal cargados por el usuario.
Validar outliers en variables numéricas.
Separar las fechas en columnas (año, mes, día , hora, duración de viaje) en archivos taxis_green, taxis_yellow.
Cambiar los tipos de datos.
Consumir APIs externas para agregar información geográfica y generar un mapa interactivo.
Guardar en archivos SQL para consultas operativas.

# Definir funciones:

Ingresos.
Horarios y días con mayor demanda.
duración de viajes.
taxis con mayor eficiencia por combustibles.
modelos de autos utilizados.
zonas con mayor cantidad de viajes.
zonas más contaminadas por ruidos y CO2.

# Fundamentación de la elección de la nube:

Evaluamos usar Amazon AWS o Google cloud. Analizando ambas nubes obtenemos algunas diferencias:

En materia de seguridad ambos nos ofrecen un sistema robusto ya sea para el cifrado de datos, módulos hardware de seguridad, etc.

Amazon nos ofrece más de 50 servicios.
Amazon DynamoDB: Base de datos NoSQL totalmente gestionada que ofrece rendimiento de milisegundos de un solo dígito.
Amazon RDS: Relational Database Service ofrece bases de datos relacionales alojadas en servidores virtuales seguros. Las actualizaciones, respaldo y seguridad es llevada a cabo por AWS por lo que el usuario sólo tendrá que gestionar la base de datos, con los beneficios que ello conlleva.
Amazon Redshift: Servicio de almacenamiento de datos en la nube que permite ejecutar consultas analíticas complejas.
Amazon Glue

Google Cloud proporciona servicios en 5 áreas:
Computing: Son servicios escalables donde, si se aumentan los requerimientos de las aplicaciones, Google automáticamente aumenta la capacidad de la infraestructura.

Networking: Ofrece servicios relacionados con la configuración de una red que permite conectar diversas máquinas virtuales.
Almacenamiento: Cloud Storage o Datastore.
Big Data: Bigquery o Dataflow.
Machine Learning: Proporciona una serie de librerías o API online, para que los desarrolladores puedan trabajar sobre ellas. Algunas de estas librerías son Visión API o Speech API, además nos permite crear con la misma infraestructura que utiliza Google.
En conclusión ambas plataformas destacan por una gran variedad de servicios en la nube de gran calidad. en materia de seguridad ambos son muy sofisticados. En cuanto a infraestructura tendremos más potencia de procesamiento en Google Cloud Platform y más memoria en AWS. Aunque ambas plataformas son personalizables, almacenamiento de objetos ambas plataformas ofrecen servicios propios como Google Cloud Storage y Amazon S3.
La gran ventaja de AWS es la variedad de servicios disponibles,Quizás el punto más débil de AWS es su estructura de precios, algo confusa a la hora de calcular los costos.
Google Cloud ofrece una gran escalabilidad y equilibrio de carga, nos permite ajustar la infraestructura al uso que se le esté dando. Los precios de Google Cloud son más competitivos a pesar de que ofrece menos servicios de los proporcionados por AWS.

```
Por lo tanto decimos que vamos a usar _____, ya que se adecua al proyecto
actual.
```