EDA : Exploración de Datos

ETL : Extract, Transform y Load

Area de staging, area de mi organización a donde llegan los datos de otras empresas o lugares. Se guardan en ficheros y se extraen en bruto.

Extract: de txt, Word, Excel, pdf, ..

Transform: Unir fuentes de datos, Analizar y tratar duplicados, Limpiar datos y unificar, Generar nuevas columnas, simplificar datos, tratar los faltantes

Load: Dejar los datos preparados o cargados en ficheros Bases de datos y gestionados mediante Sistemas de Gestion de Bases de Datos. Oracle, MySQL, PostgreSQL, SQL Server…

Cardinalidad es el numero diferente de valores estrictamente.

Simplificar categorizando es muy importante