

# Overview of Standard Transformations

Print to PDF ►

Let us understand standard transformations we perform on top of data in collections.

- Filtering
  - Get all the revenue generating orders considering only with order\_status as COMPLETE or CLOSED.
  - Get all the sales details of all the dealers with in a state.
- Row level transformations such as standardization, masking, cleansing etc.
  - Converting names to upper or lower cases.
  - Standardizing Phone numbers or Addresses.
  - Encrypting Phone Numbers or Social Security Numbers.
  - Masking Phone Numbers, Social Security Numbers or Date of Births and provide limited information such as last 4 digits of social security number.
  - Removing unwanted spaces and other special characters.
- Total Aggregations
  - Revenue generated by a product
  - Revenue generated for a given day
- Grouped Aggregations
  - Revenue generated by each product
  - Revenue generated on daily basis
  - Revenue generated by each car dealer
  - Revenue generated by each retail store
  - Sales Commission for each channel partner
- Sorting and Ranking
  - Top 5 stores by revenue
  - Top 5 car dealers by sales
  - Top 10 sales agents by the commission

Typically we use external libraries such as Pandas, Pyspark etc to perform these standard transformations. However, we will try to develop using conventional loops to understand how they are implemented and also to get better with respect to programming.