Data cleaning and transformation