# Data Cleaning and Transformation Report

## a) Load employee\_data.csv

Sample data:

Employee ID Full Name Department Salary  
 101 Alice Johnson HR 50000  
 102 Bob Smith NaN 60000  
 103 Charlie Lee Engineering not available  
 104 Dana White Finance 70000

## b) Convert Salary to Numeric

df['Salary'] = pd.to\_numeric(df['Salary'], errors='coerce')

After conversion:

Full Name Salary  
Alice Johnson 50000.0  
 Bob Smith 60000.0  
 Charlie Lee NaN  
 Dana White 70000.0

## c) Remove Rows with Missing Department

df\_cleaned = df.dropna(subset=['Department'])

Data after removing rows with missing Department:

Employee ID Full Name Department Salary First Name  
 101 Alice Johnson HR 50000.0 Alice  
 103 Charlie Lee Engineering NaN Charlie  
 104 Dana White Finance 70000.0 Dana

## d) Extract First Name

df\_cleaned['First Name'] = df\_cleaned['Full Name'].apply(lambda x: x.split()[0])

Final cleaned data:

Employee ID Full Name Department Salary First Name  
 101 Alice Johnson HR 50000.0 Alice  
 103 Charlie Lee Engineering NaN Charlie  
 104 Dana White Finance 70000.0 Dana