1. How is the characteristics of the data? Do exploratory data analysis, calculate descriptive statistics, and visualize it.

CustomerID Churn Tenure CityTier WarehouseToHome \

count 1896.000000 1896.000000 1739.000000 1896.000000 1744.000000

mean 51668.825949 0.500000 7.347901 1.719409 15.922018

std 1598.949360 0.500132 8.149302 0.936148 8.498368

min 50001.000000 0.000000 0.000000 1.000000 5.000000

25% 50474.750000 0.000000 1.000000 1.000000 9.000000

50% 50948.500000 0.500000 4.000000 1.000000 14.000000

75% 52809.250000 1.000000 13.000000 3.000000 22.000000

max 55622.000000 1.000000 50.000000 3.000000 36.000000

HourSpendOnApp DeviceRegistered SatisfactionScore NumberOfAddress \

count 1766.000000 1896.000000 1896.000000 1896.000000

mean 2.682899 3.543776 3.275316 4.088080

std 0.679286 1.015023 1.269551 2.694888

min 0.000000 1.000000 1.000000 1.000000

25% 2.000000 3.000000 2.000000 2.000000

50% 3.000000 3.000000 3.000000 3.000000

75% 3.000000 4.000000 4.000000 6.000000

max 4.000000 6.000000 5.000000 21.000000

Complain OrderIncreaseFromLastYear CouponUsed OrderCount \

count 1896.000000 1855.000000 1817.000000 1837.000000

mean 0.385549 15.391914 1.470556 2.694066

std 0.486853 3.695976 1.862077 2.866878

min 0.000000 11.000000 0.000000 1.000000

25% 0.000000 12.000000 0.000000 1.000000

50% 0.000000 14.000000 1.000000 2.000000

75% 1.000000 18.000000 2.000000 3.000000

max 1.000000 26.000000 16.000000 16.000000

DaySinceLastOrder CashbackAmount

count 1812.000000 1896.000000

mean 3.786976 164.907252

std 3.540237 44.698011

min 0.000000 0.000000

25% 1.000000 132.940000

50% 3.000000 150.870000

75% 7.000000 181.610000

max 46.000000 323.590000

1. Please Do preprocessing data. Is there any missing values or outliers? If yes, solve it and give some explanation. Do variable selection or dimension reduction if needed and give some explanation.

Data memiliki missing values pada kolom Tenure, WarehouseTohome, HourSpendOnApp,OrderIncreaseFromLastYear,CouponUse,OrderCount,DaySinceLastOrder.

Cara yang dilakukan untuk mengatasi missing values adalah dengan menggantikan nilai missing values dengan mean.

1. Find the best model and evaluate the model.