DATA QUALITY PLAN:

Feature	Data Quality Issue	Handling Strategy
ConditionNotes	Missing Values (47%)	Do nothing
ShipsFromCountry	Missing Values (37%)	Complete-case analysis
ShipsFromState	Missing Values (41%)	Complete-case analysis
ListingPrice	Outliers	Drop rows
SellerFeedbackRating	Outliers	Do nothing
SellerFeedbackCount	Outliers	Do nothing
ShippingPrice	Outliers	Drop rows
ShippingTime_minHours	Outliers	Do nothing
ShippingTime_maxHours	Outliers	Do nothing

ConditionNotes:
47% of the dataset have null values for ConditionNotes. I am not removing the data from the set as I don't believe it is a necessary field in very instance, therefore having null values here is not an issue.

ShipsFromCountry and ShipsFromState:
ShipsFromCountry and ShipsFromState have 37% and 41% null values respectively. I am leaving this data in, as I am not convinced at this stage if and how this information impacts 'IsWinner' target outcome. Further analysis may shed more light.

ListingPrice and ShippingPrice:
My analysis of outliers leads me to believe that ListingPrice and ShippingPrice outliers are incorrect data - perhaps due to error or data corruption of the csv file I am using. I am removing the outliers relating to both columns as I do not want this data to skew my later analysis. As explained in the Outliers section above, the majority of the prices falls within a narrow band, and these outliers look unrealistic in comparison.

SellerFeedbackRating and SellerFeedbackCount:
I think both SellerFeedbackRating and SellerFeedbackCount may be very relevant in the 'IsWinner' target outcome. I am not removing the outliers in both cases because I am interested to see whether there is a correlation between the data here and 'IsWinner'.

ShippingTime_minHours and ShippingTime_maxHours:
I am uncertain as to whether this information impacts the target outcome. I have decided to leave the outlier rows in until I can do some further analysis to understand whether or not there is a relationship here.