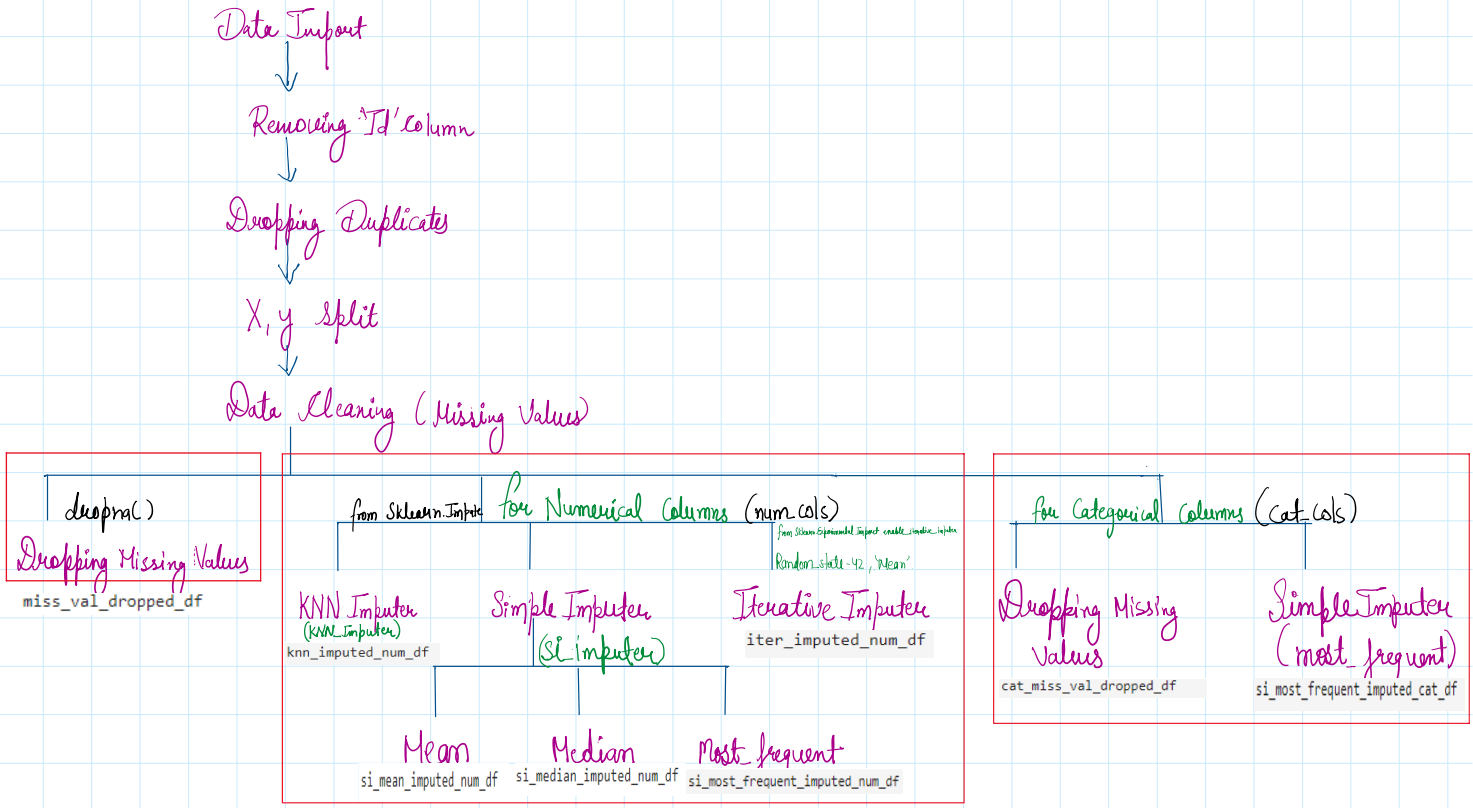
Kaggle project process flow using linear Regression.



final df after Missing Values treated.

```
miss_val_dropped_df
```
← Df with Missing Values Dropped

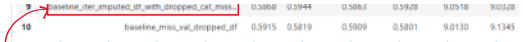
```
knn_imputed_df
si_mean_imputed_df
si_median_imputed_df
si_most_frequent_imputed_df
iter_imputed_df
```
← Df with Numerical Columns & Most frequent Categorical Columns

```
knn_imputed_df_with_dropped_cat_missing_val
si_mean_imputed_df_with_dropped_cat_missing_val
si_median_imputed_df_with_dropped_cat_missing_val
si_most_frequent_imputed_df_with_dropped_cat_missing_val
iter_imputed_df_with_dropped_cat_missing_val
```
← Df with Numerical Columns & Categorical Columns Missing Value Dropped

One Hot Encoding & Ordinal Encoding. (df names not changed)
Model Selection → LinearRegression()
Baseline Model Evaluation (RMSE, R^2, Adjusted R^2)

	method	r2_train	r2_test	adj_r2_train	adj_r2_test	train_rmse	test_rmse
0	baseline_knn_imputed_df	0.5887	0.5713	0.5882	0.5698	9.1402	9.0492
1	baseline_si_mean_imputed_df	0.5892	0.5716	0.5887	0.5700	9.1349	9.0463
2	baseline_si_median_imputed_df	0.5891	0.5714	0.5886	0.5698	9.1359	9.0483
3	baseline_si_most_frequent_imputed_df	0.5891	0.5714	0.5886	0.5698	9.1359	9.0483
4	baseline_iter_imputed_df	0.5898	0.5725	0.5893	0.5709	9.1281	9.0368
5	baseline_knn_imputed_df_with_dropped_cat_missi...	0.5863	0.5944	0.5857	0.5928	9.0578	9.0325
6	baseline_si_mean_imputed_df_with_dropped_cat_m...	0.5863	0.5940	0.5858	0.5924	9.0570	9.0371
7	baseline_si_median_imputed_df_with_dropped_cat...	0.5864	0.5938	0.5858	0.5921	9.0567	9.0397
8	baseline_si_most_frequent_imputed_df_with_drop...	0.5864	0.5938	0.5858	0.5921	9.0567	9.0397
9	baseline_iter_imputed_df_with_dropped_cat_miss...	0.5868	0.5944	0.5863	0.5928	9.0518	9.0328
10	baseline_miss_val_dropped_df	0.5915	0.5819	0.5909	0.5801	9.0130	9.1345

Best Baseline Model → No Difference → Data transformation (MinMax Scaler / Standard Scaler)



Best Baseline Model → No Difference → Data transformation (MinMax Scaler / Standard Scaler)

Outliers Treatment (for Numerical Columns)

Zscore & IQR (Skewness between ± 0.5 Then Method "Z.score" else IQR)

- Removing Outlier
- Winsorization
- Not treating Outliers

Feature Engineering (polynomial features ✓ / log transformation)

Feature Selection

p-value, Lasso, mi, corr/vif..

Model Evaluation & Selecting Best Model

Test prediction

Exporting to CSV. & Submit prediction.