Supply Chain Management Car Sales Prediction

Problem Statement:

1. Perform Basic EDA
    a. Boxplot
    b. Histogram – Distribution of Target Variable
    c. Distribution Plot – Target Variable
    d. Aggregation for all numerical Columns
    e. Unique Values across all columns
    f. Duplicate values across all columns
    g. Correlation – Heatmap
    h. Regression Plot
    i. Bar Plot
    j. Pair plot

    P.s: All the above charts should be plotted against every independent variable vs Target Variable. (Except b and c)
2. Drop all duplicate rows
3. Drop all non-essential features
4. Replace outliers with Nulls (if you find it essential) and replace all the nulls with respective approach of central tendencies (Mean/Median/Mode).
5. Calculate Z score to validate whether outliers are still present or not.
6. Clean the data with formatting issues if any. (converting datatypes, replacing dollars, etc.)
7. Add your view of EDA to enhance understanding of data. i.e., Grouping data and observing the way data is distributed. Try to add as many layers of EDA as possible.
8. Build a model of choice – Regression problem statement, hence build a regression model first and calculate MAE, MSE, RMSE, MPE and MAPE. Also see the R2 score.
9. Build at least a minimum of 4 different Regression models. All the models should use K-Fold cross Validation to train the model with at least 5-fold cross validation.
10. Compare the error and pick the ideal one with least errors.
11. Run hyperparameter tuning on all the models and pick the best parameters (A minimum of 2 Parameters should be tuned) and picked.
12. Now, compare the models and pick the ideal one.
13. Try to Predict the target with maximum independent features.
14. Write the insights and your findings (in PPT).
15. Plot all independent variable against Target Variable in Tableau (Once the data is cleaned) and build a Dashboard and Story out of it.