Exploratory Data Analysis EDA

**Why?**

1. To determine if there are any problems with your dataset.
2. To determine whether the question you are asking can be answered by the data that you have.
3. To develop a sketch of the answer to your question.

**How?**

1. Collect/Read your data
   * Any bias in sampling, or Survey Design etc.
2. Descriptive Statistics
   * Mean, median, sd, min, max, quartiles, outliers, etc
3. Simple/Basic charts and graphs
   * Boxplot, Histogram, Quartile plot, etc
4. Develop a sketch of the answer to your question.

For our dataset:

1. Read data
   1. Any missing value?
   2. Factor or numeric?
2. descriptive analysis
   1. mean/median
   2. distribution
   3. outliers
3. Correlation
   1. Z/T-test
   2. chi-square
   3. ANOVA
4. Model selection
   1. Different model
   2. Model comparison
   3. Goodness of fit (chi-square / anova)
   4. Confidence interval for regression
5. Prediction
   1. Confidence interval for prediction
   2. standard error
6. Diagnosing
   1. Non-linearity between y and x’s – residuals vs fitted values
   2. Normality of errors
   3. Independence of errors – DW test
   4. Inhomogeneity of variance
   5. Outliers and high leverage points

Text

Description automatically generated

Text

Description automatically generated with medium confidence

1. Fixing up models

Table

Description automatically generated with medium confidence