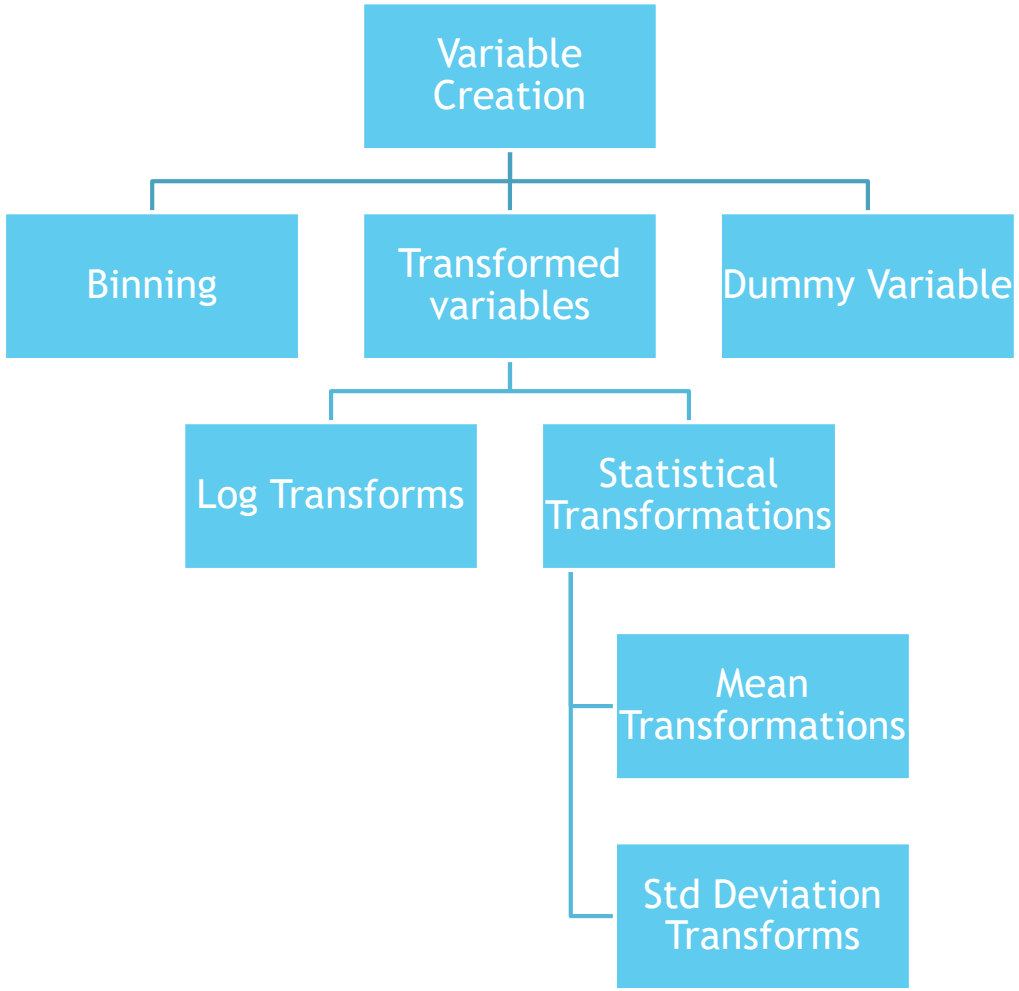
Variable Creation

DataDNA

Types of Variable Creation



Binning Variable Creation

- Combining the variable power to create a fresh variable
- Using Business Insights to create a new variable

- Numerical Variable to Categorical Example:
  - Sales/Revenue of a store containing numerical values
  - Creating a new variable which has 3-5 levels (Low, Medium, High, HNW customers)
- Multiple Categories to Powerful Categorical Example
  - Services Industry usually contain many accounts/projects under a Vertical
  - A new variable which can be Top 5, Top 10, Top 10 & Others

Binning Variable Creation – Exercise 1

Use Case: Imagine you are trying to roll out a new credit card for your bank. In this context, you like to categorise the Customer's purchasing power based on their Income so you can strategize the risk involved in the business

Exercise: Create a Customer Category from Customer's Income

Income Range	Customer Category
100 - 10000	Low Spend
10000 - 30000	Medium Spend
30000 - 50000	High Spend
50000 - 100000	High Net Worth Customer

Binning Variable Creation – Exercise 2

Use Case: Imagine you are trying to roll out a new credit card for your bank. In this context, you like to categorise the Customer's purchasing power based on their Income so you can strategize the risk involved in the business

Exercise: Create a Customer Category from Customer's Income percentile of available data

Income Range	Customer Category
Min – 20th Percentile	Lower Band
20th – 50th Percentile	Medium Band
50th – 95th Percentile	Upper Band
95th Percentile - Max	Top 5 Percentile

Evaluate the risk involved in lending the credit card to Lower Band. But we are constraining the analysis to just one variable. We also should think of other variables.

DataDNA

Transformed Variable Creation

- **Log Transforms:** When we find skewed data, log transforms work the best. Care to be taken if the values contain 1 or 0.
- **Mean Transformations:** A new variable containing the running mean of the last 5, last 10 values can be created. This reduces variance effect.
- **Std Deviation/Variance Transformations:** In order to remove variances within a variable, the transformations can be applied on top of it.

Dummy Variable Creation

- When there are variables that contain different levels, they must be transformed into Dummy variables to feed into the algorithms.
- These dummy variables also help us in extracting insights as they will be specific to those levels.
- Example:
  - Lets say we have a Category Variable which explains the car type. Call it A,B,C
  - On transformation, A new set of variables (V1,V2,V3) will be created. Hence the renaming will be as follows
    - A – 001 (V1 V2 V3)
    - B – 010 (V1 V2 V3)
    - C – 011 (V1 V2 V3)

DataDNA

Dummy Variable Creation – Exercise

- Lets try to create the dummy variables for previously created “Customer Category” Variable.

- Lets see how easily python does it for us with just line of code ☺