CARLTON MEZETIN

FEATURE SELECTION

WHAT IS FEATURE SELECTION

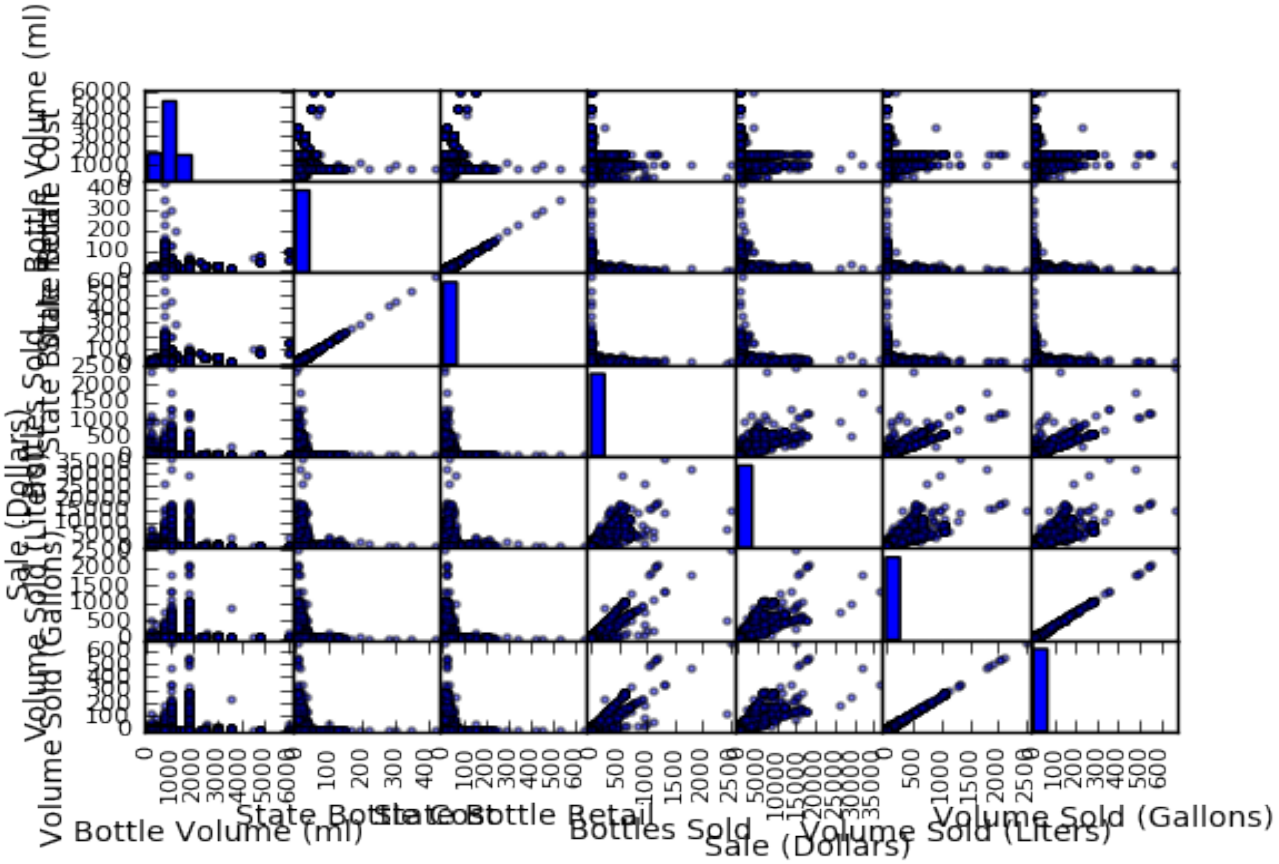
- **Selecting attributes that are most relevant to predictive modeling**

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p + \varepsilon$$

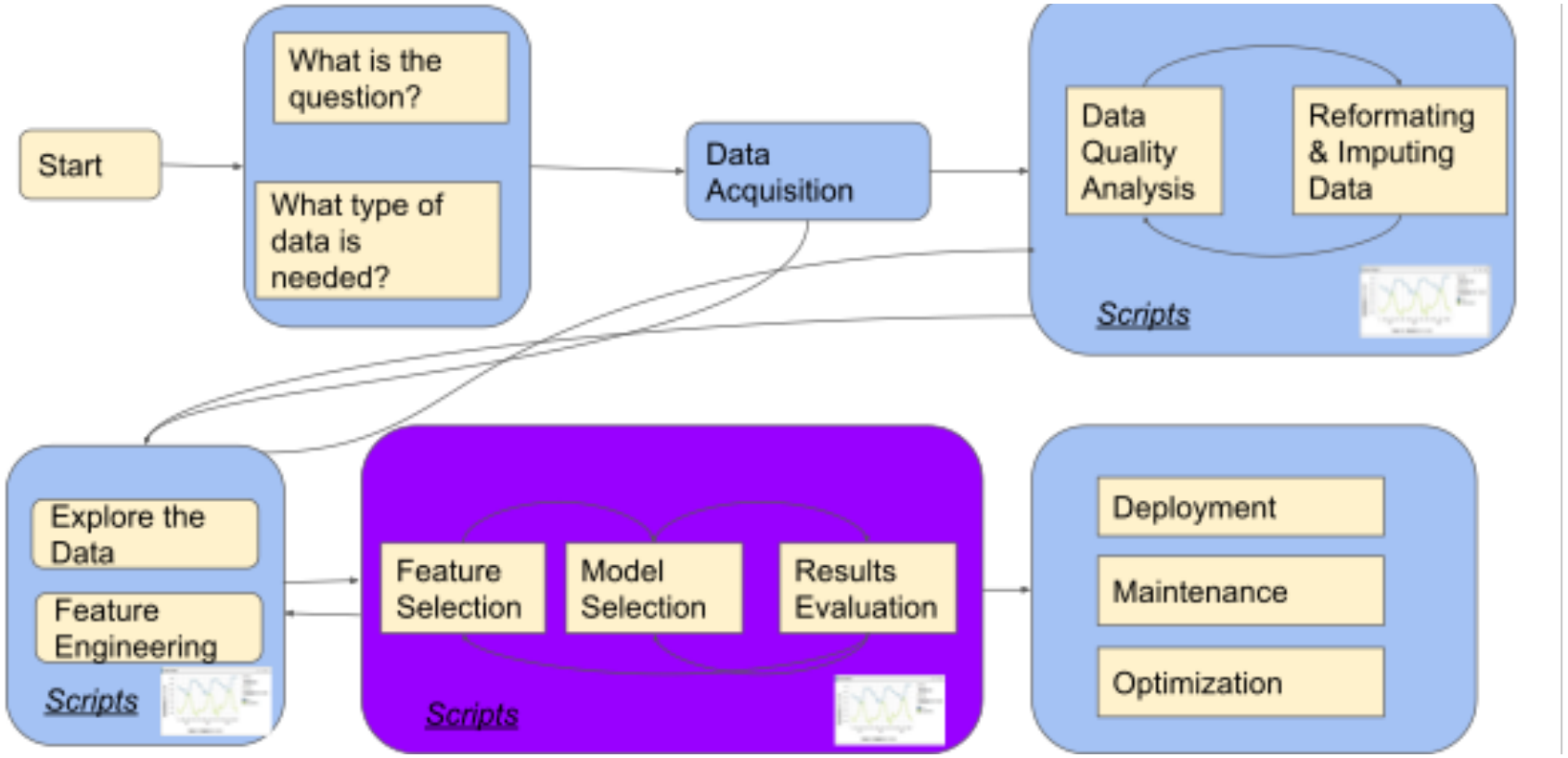
- **"feature selection… is the process of selecting a subset of relevant features for use in model construction"**
- **The objective of variable selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data.**
  - -Guyon and Elisseeff in "An Introduction to Variable and Feature Selection" (PDF)
- **The first step in the modeling process**

	Bottle Volume (ml)	State Bottle Cost	State Bottle Retail	Bottles Sold	Sale (Dollars)	Volume Sold (Liters)	Volume Sold (Gallons)
Bottle Volume (ml)	1.000000	0.312841	0.313819	-0.012476	0.082446	0.156258	0.156090
State Bottle Cost	0.312841	1.000000	0.999960	-0.062980	0.135931	0.009296	0.009264
State Bottle Retail	0.313819	0.999960	1.000000	-0.062831	0.136114	0.009736	0.009705
Bottles Sold	-0.012476	-0.062980	-0.062831	1.000000	0.825446	0.883348	0.883378
Sale (Dollars)	0.082446	0.135931	0.136114	0.825446	1.000000	0.846420	0.846432
Volume Sold (Liters)	0.156258	0.009296	0.009736	0.883348	0.846420	1.000000	1.000000
Volume Sold (Gallons)	0.156090	0.009264	0.009705	0.883378	0.846432	1.000000	1.000000

SIMPLE FEATURE SELECTION



WHERE DOES FEATURE SELECTION OCCUR



WHY IS FEATURE SELECTION IMPORTANT

- **Signal vs. Noise**
  - We only want to know the variables that are most predictive and not explanatory
- **Reducing complexity of our models**
  - Interpretability
  - Computational expense
- **Model accuracy**
  - Filtering out

FEATURE SELECTION METHODS

- **Top Down-** Imposes a global constraint that feature must meet in order to be kept.
  - Variance Threshold

- **Bottom Up-**
  - Sequential forward selection : iteratively adds best performing features.
  - Sequential backward selection : iteratively delete least significant features

FEATURE SELECTION ALGORITHMS

- **Embedded Methods –** Learn which features best contribute to the accuracy of the model while the model is being created.
  - i.e. Regularization techniques (Lasso, Ridge)

- **Filter Methods –** applies a statistical measure to scoring to each feature. Features with higher ranked scores are kept
  - i.e. Chi squared test

- **Wrapper Methods -** selects a subset of features then using a predictive model evaluates and scores. Then repeats for every subset combination of features.
  - i.e. Recursive feature elimination

EXAMPLES

- **Select K Best**
- **Variance Threshold**
- **Select Percentile**
- **Recursive Feature Selection**
- **Regularization**

CONCLUSION

The objective of variable selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data.

— Guyon and Elisseeff in "An Introduction to Variable and Feature Selection" (PDF)