

The Role of Predictive Analytics

Wu-Chyuan (Gary) Gau, PhD, ASA, MAAA

05/16/2017

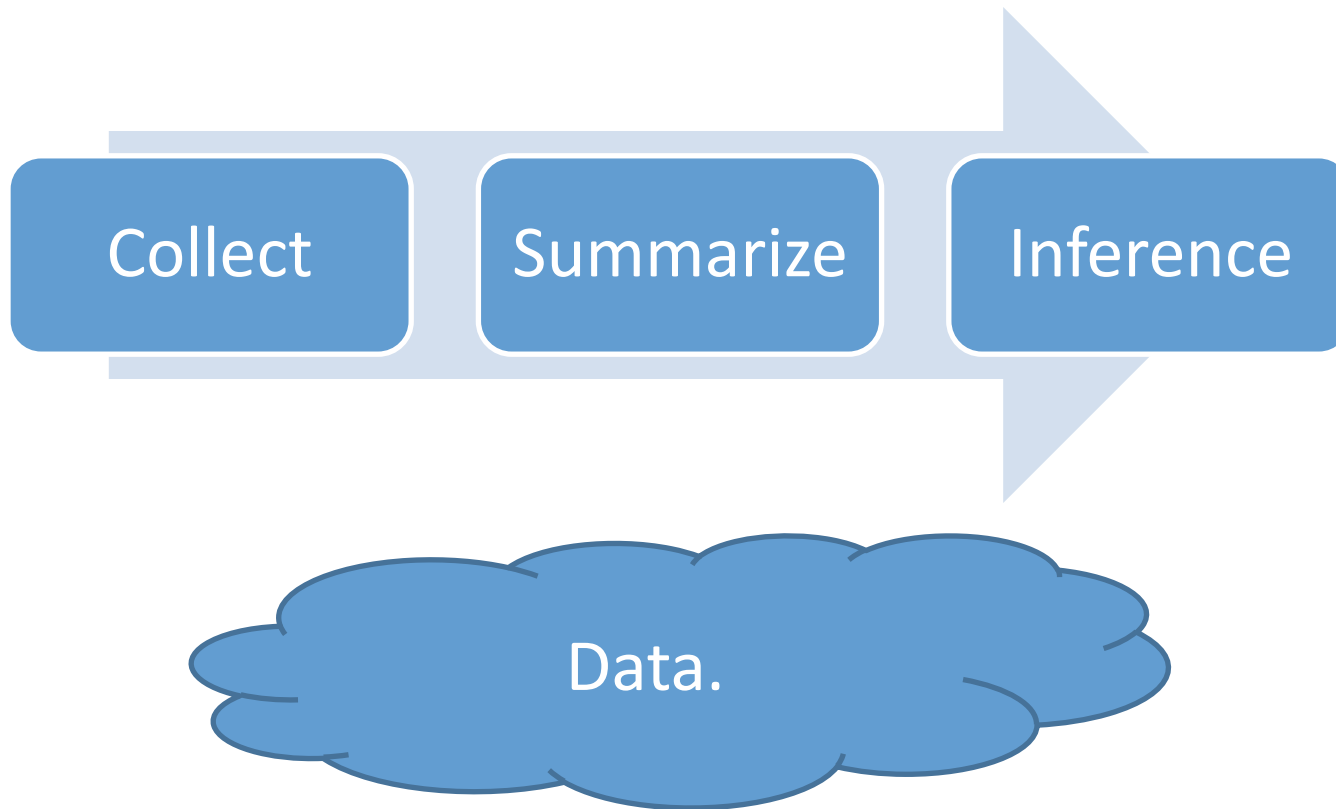
Agenda

- What Is Predictive Analytics?
- Actuarial Traditional Functions
- Post-Reform Marketplace Background
- Predictive Modeling Methodology
- Impact of Risk Modeling
- Data Massage
- Competing on Analytics
- Healthcare Predictive Modeling Opportunities
- Predictive Analytics in Other Actuarial Areas
- Machine Learning Algorithms Cheat Sheet

What is Predictive Analytics?

Traditional Statistics

- Statistics is a scientific way of studying how to:



Predictive Analytics

- **Predictive analytics** encompasses a variety of statistical techniques from predictive modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events

Source: https://en.wikipedia.org/wiki/Predictive_analytics

Advanced Analytics

Gartner defines Advanced Analytics as: **“the analysis of all kinds of data using sophisticated quantitative methods (for example, statistics, descriptive and predictive data mining, simulation and optimization) to produce insights that traditional approaches to business intelligence (BI) — such as query and reporting — are unlikely to discover.” ***

*<http://www.applieddatalabs.com/content/new-gartner-magic-quadrant-advanced-analytics-platforms>

Actuarial Traditional Functions

Actuarial Traditional Functions

- Pricing
- Reserving
- Financial Performance

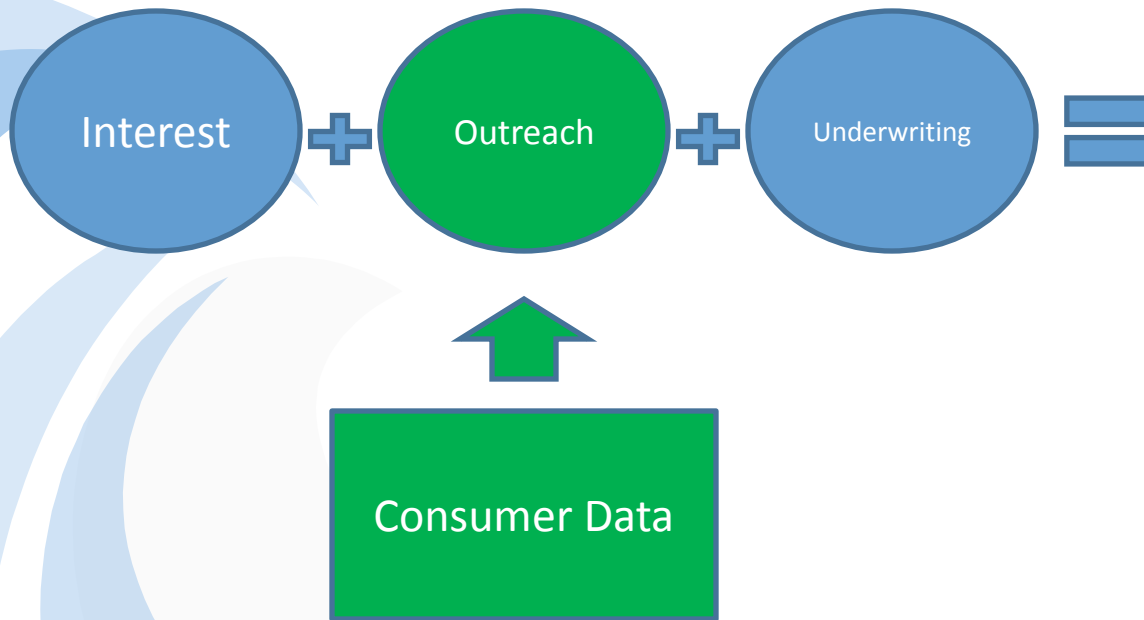
Financial Projection

Incurred Claim

Incurred Date	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16	Oct-16	Nov-16	Dec-16
Jan-16	\$14,081,127											
Feb-16												
Mar-16												
Apr-16												
May-16												
Jun-16												
Jul-16												
Aug-16												
Sep-16												
Oct-16												
Nov-16												
Dec-16												

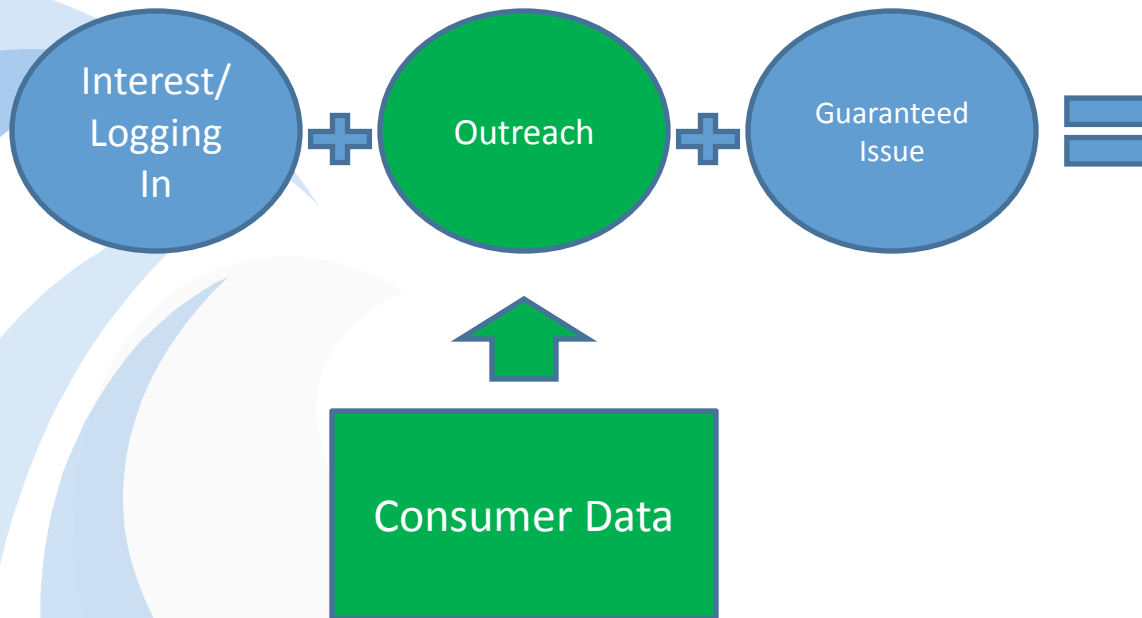
Post-Reform Marketplace Background

Where We Were



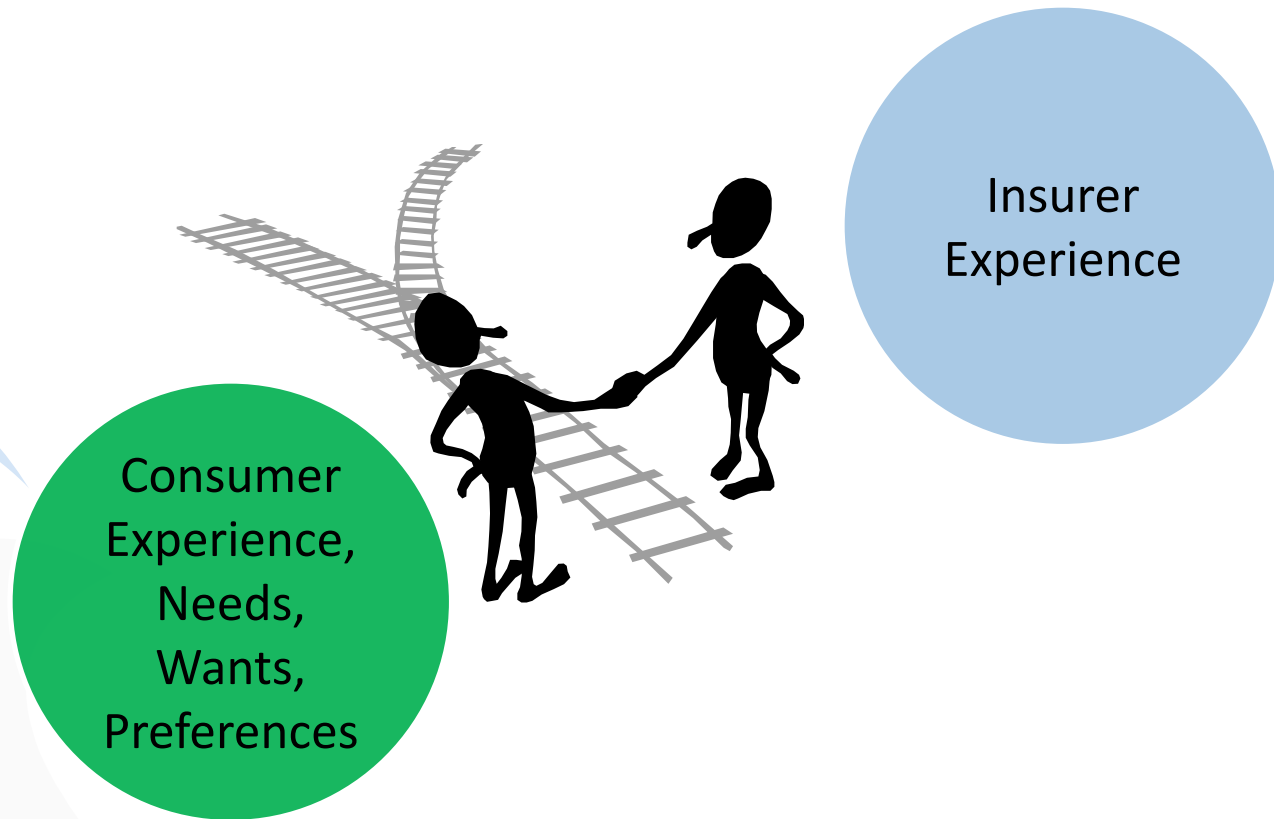
Risk Selection

Where We Are



Risk Management

Where We Are Going



2010 FL Health Insurance Coverage

Sector	Population (million)
Uninsured	4.3
Medicaid/SCHIP (Non-Risk)	1.3
Medicaid/SCHIP (Risk)	1.3
TRICARE/Other	1.2
Medicare	2.4
Under 65	0.8
Small Group	2.2
Large Group	5.4

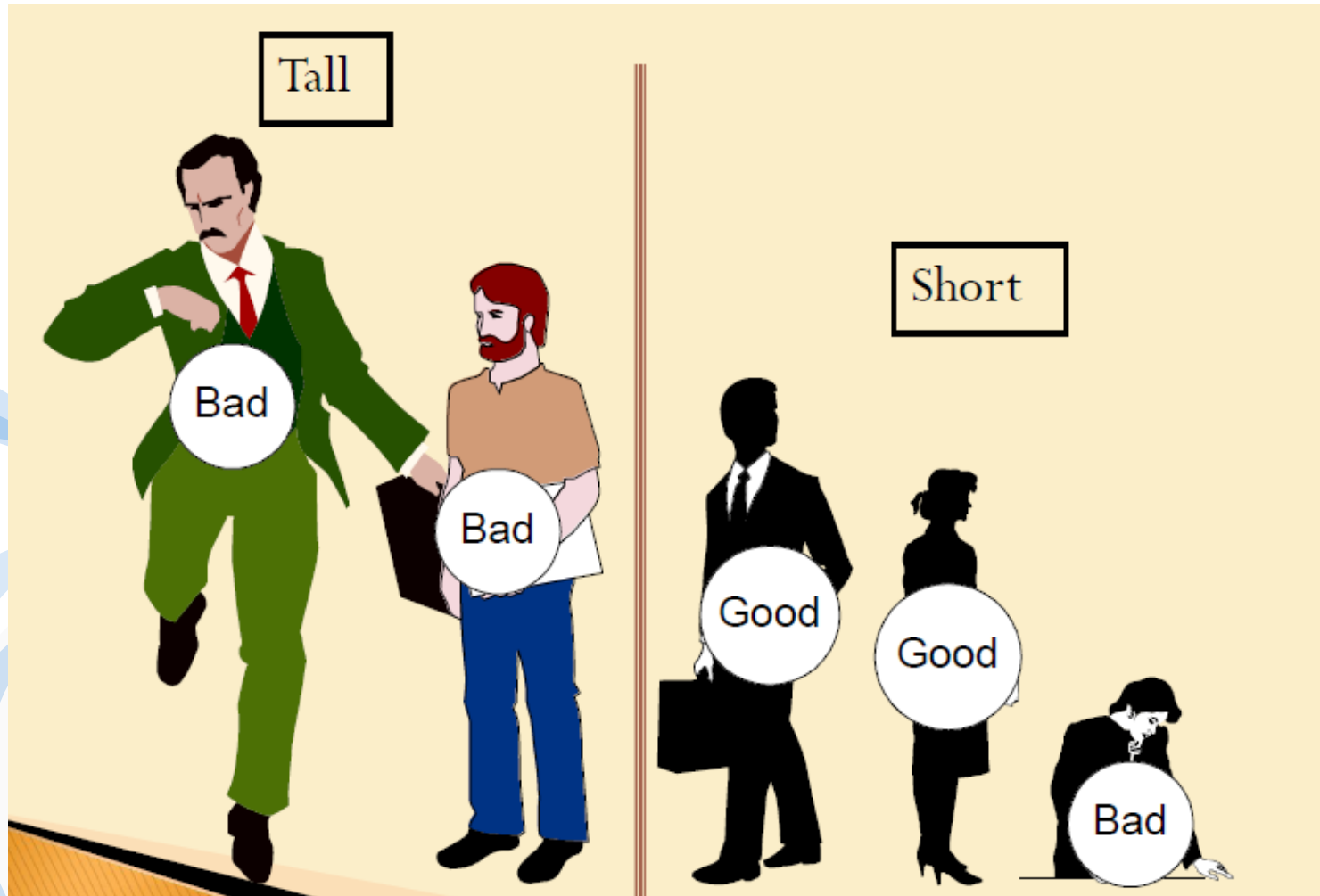
Predictive Modeling Methodology

1. Define the business problem.
2. Translate business problems into predictive modeling problems.
3. Select appropriate data.
4. Get to know the data.
5. Create a model set.
6. Fix problems with the data.
7. Transform the data.
8. Build models.
9. Assess models.
10. Deploy models.
11. Assess results.

Consumer Data

- Age
- Gender
- Household income
- Hobby
- Music
- Pets
- Sports
- Read

Classification and Segmentation



Impact of Risk Modeling

Loss Ratio	Risk 1	Risk 2	Risk 3	Overall
LBA 1	79%	73%	73%	74%
LBA 2	81%	79%	75%	77%
LBA 3	99%	93%	82%	87%
Overall	87%	82%	77%	80%

Ensemble

Proxy Model

	1	2	3	4	5	Total
1	66.7%	75.5%	76.8%	82.0%	79.3%	73.0%
2	68.8%	72.7%	77.1%	79.6%	87.6%	75.6%
3	67.9%	73.4%	72.6%	80.2%	82.7%	75.6%
4	69.9%	83.1%	78.9%	81.7%	83.3%	80.8%
5	84.8%	81.1%	88.9%	89.5%	93.2%	91.1%
Total	68.4%	76.1%	77.8%	82.8%	89.1%	80.0%

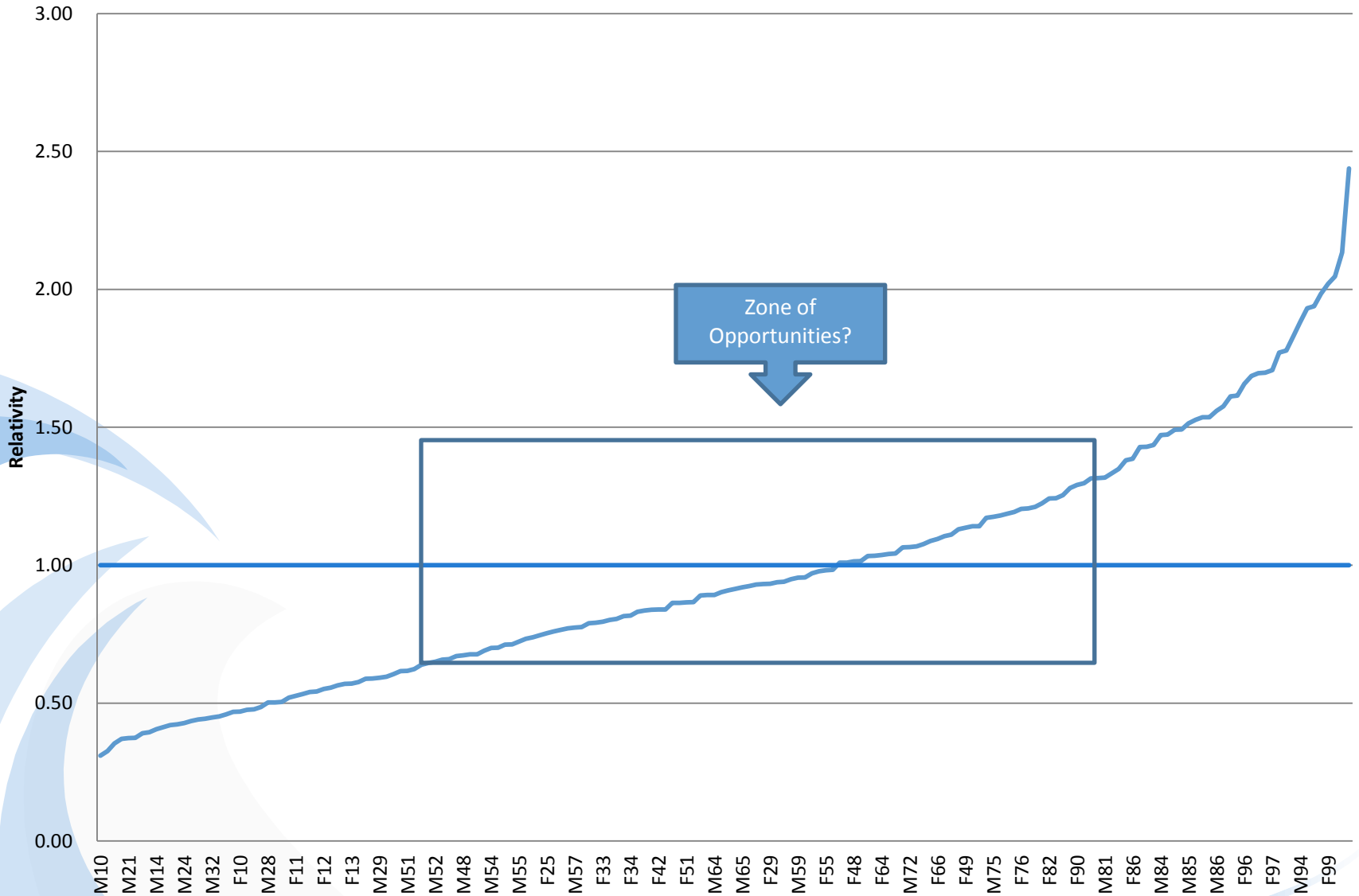
Build Up Model

Profitability

Profit Decile	Mean Risk Score	Mean Allowable Rating Factor	Loss Ratio
1	1.83	2.048	97.1%
2	1.71	1.823	87.0%
3	1.51	1.729	84.9%
4	1.52	1.739	82.1%
5	1.47	1.659	81.8%
6	1.33	1.539	77.7%
7	1.38	1.534	77.3%
8	1.39	1.576	72.7%
9	1.55	1.621	69.3%
10	2.07	1.995	66.9%
Total	1.58	1.726	80.0%

For Illustrative Purpose Only

Gender, Age Band, and Lifestyle Based Risk Profile



For Illustrative Purpose Only

Data Massage

Categorical Variables

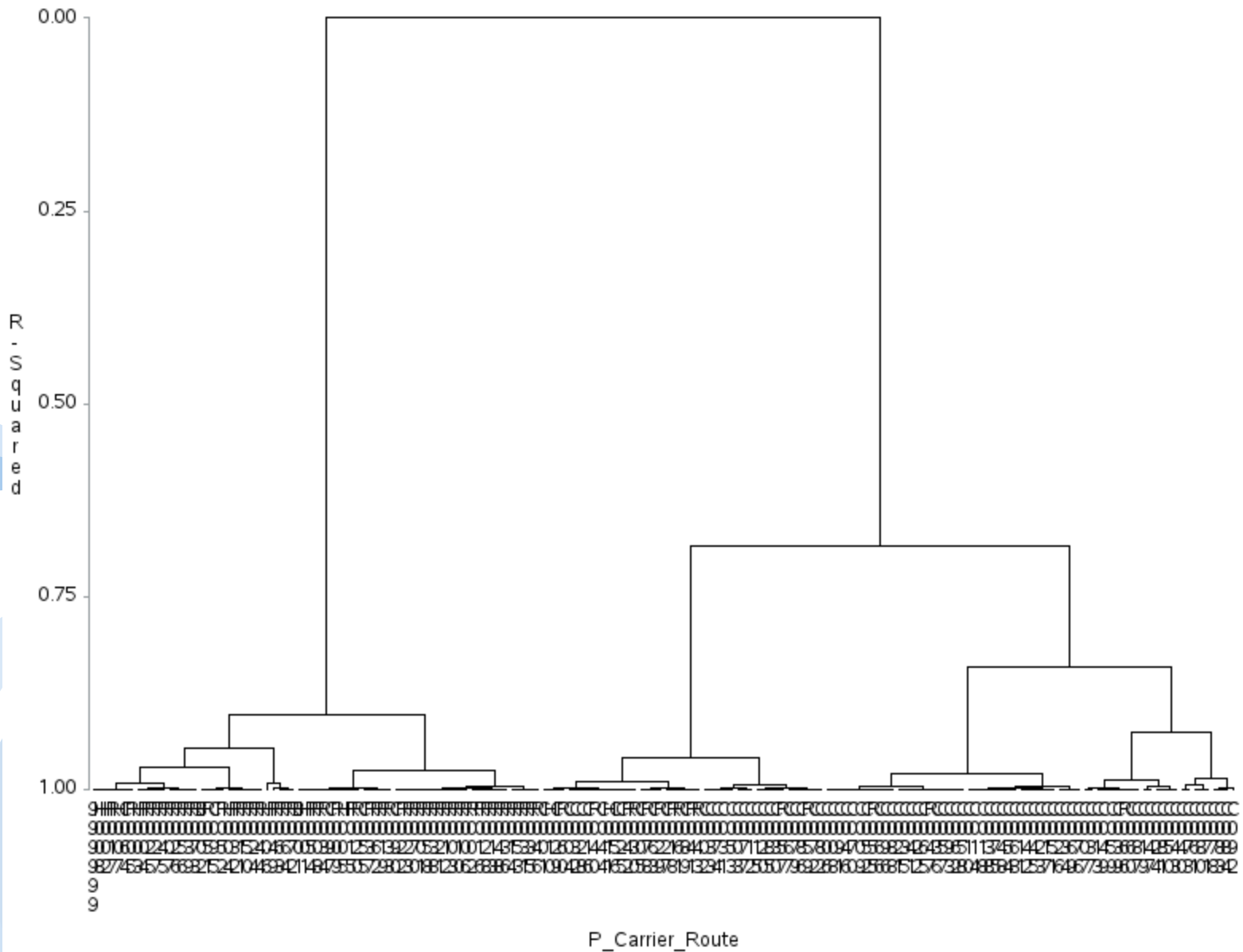
- Not all data mining tools can handle categorical inputs well.
 - Decision tree – OK
 - NN – not Ok
 - Logistic regression – not OK
- **FL Census data**
 - Population: 18,951,066**
 - 67 counties**
 - 983 ZIPs**
 - 4,245 Tracts**

Postal Carrier Route

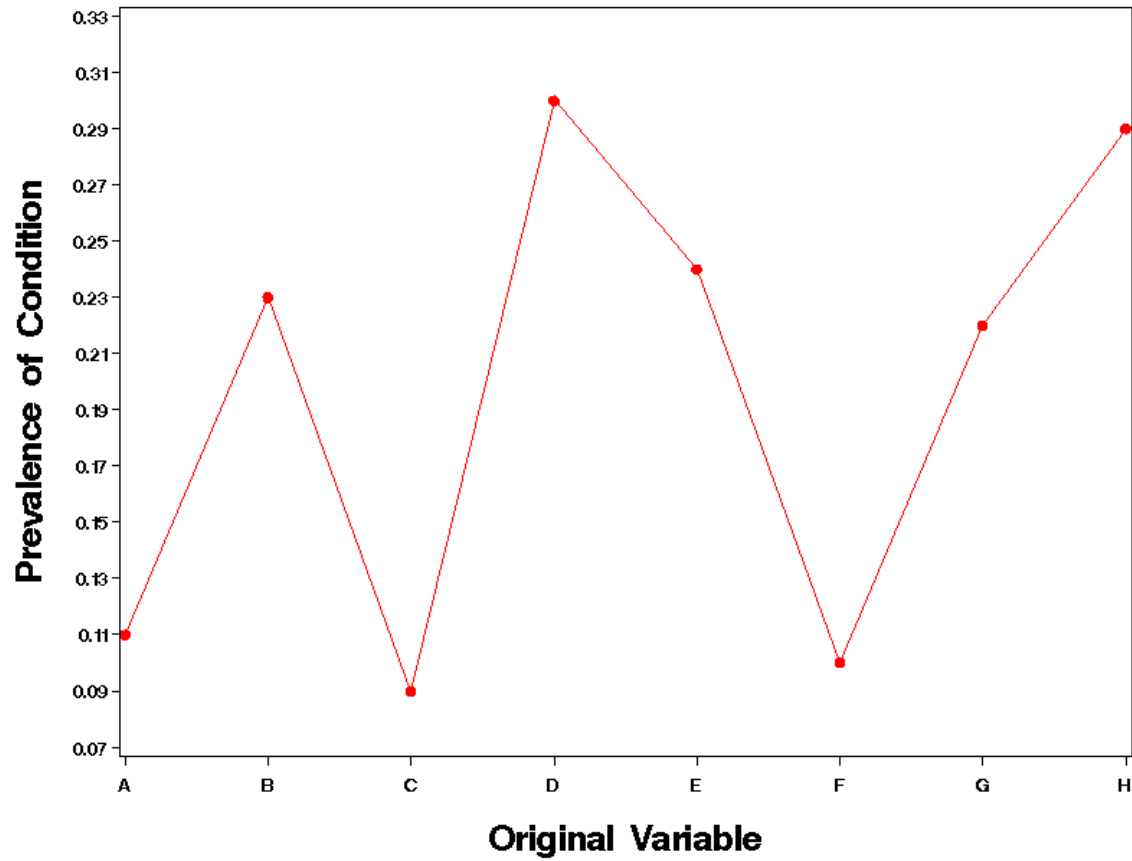
Postal Carrier Routes in ZIP Code

32246 Jacksonville, Florida

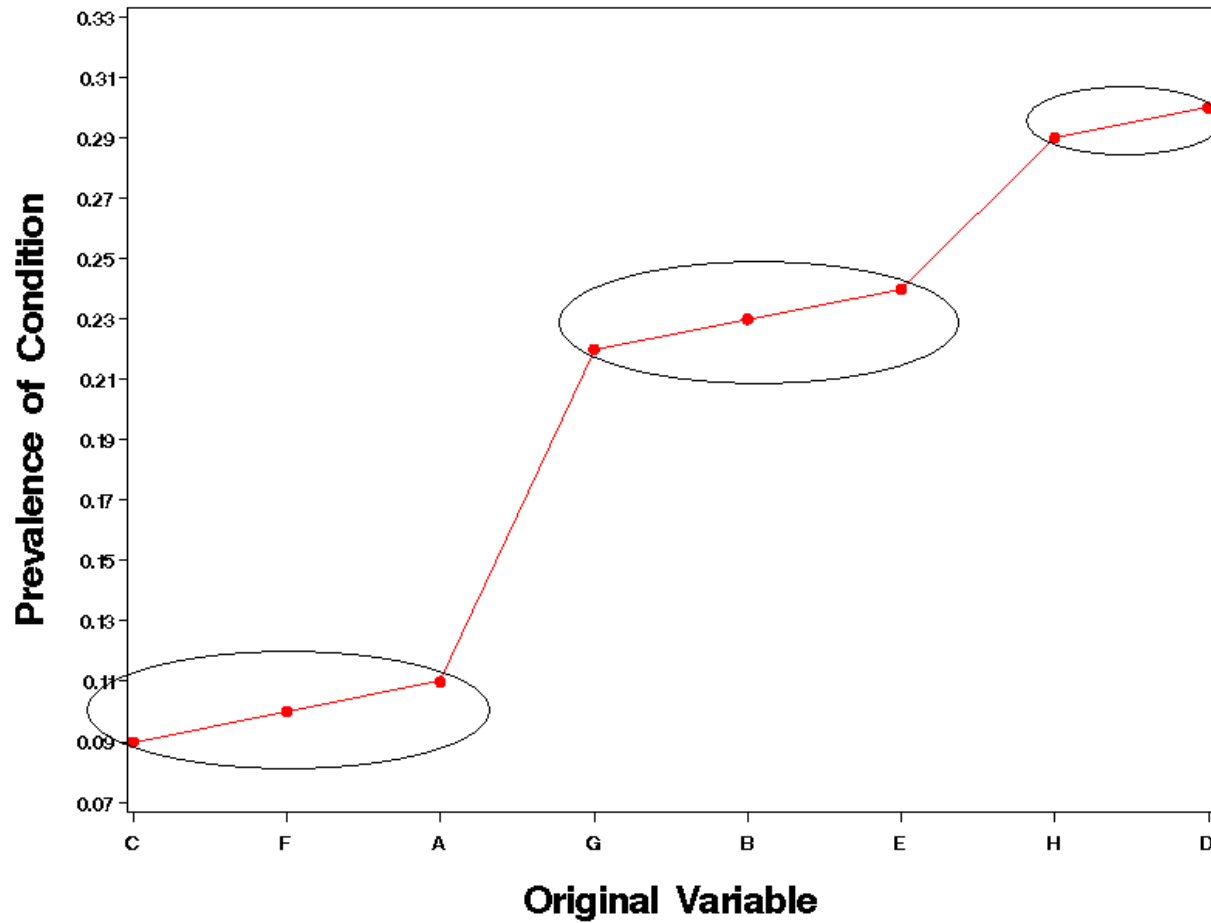
Route	Type of Route	Business Count	Apartment Count	Residential Count	Avg Household Income	Avg Property Value
C001	City	104	0	402	\$53,906	\$87,604
C003	City	16	332	652	\$38,909	\$72,367
C008	City	1	0	635	\$66,709	\$231,474
C013	City	1	0	614	\$62,982	\$99,833
C023	City	57	383	483	\$49,286	\$107,979
C026	City	24	965	1	\$63,789	\$325,942
C029	City	12	160	615	\$41,921	\$100,865
C033	City	69	200	454	\$45,507	\$79,672
C035	City	0	0	877	\$64,216	\$129,577
C036	City	75	585	754	\$50,719	\$141,914
C050	City	0	0	745	\$53,179	\$106,004
C053	City	104	471	428	\$62,547	\$106,338
C056	City	24	1037	3	\$65,055	\$146,883
C057	City	76	325	592	\$64,857	\$0
C058	City	239	780	219	\$56,239	\$236,262
C059	City	64	148	474	\$45,998	\$108,723
C063	City	197	223	170	\$45,397	\$119,388
C066	City	7	0	707	\$54,047	\$84,645
C068	City	20	86	778	\$51,498	\$161,574
C069	City	44	0	601	\$48,736	\$72,014
R036	Rural Route	1	0	500	\$74,220	\$227,251
R094	Rural Route	0	0	589	\$63,689	\$226,449
R107	Rural Route	12	0	860	\$71,118	\$192,187
R112	Rural Route	3	421	549	\$58,061	\$176,949
R116	Rural Route	2	0	648	\$57,329	\$196,408
R142	Rural Route	0	0	781	\$60,788	\$176,989
R158	Rural Route	1	0	874	\$67,407	\$191,619
R182	Rural Route	3	203	496	\$68,592	\$251,456
Overall		1,156	6,319	15,501	\$57,580	\$142,343



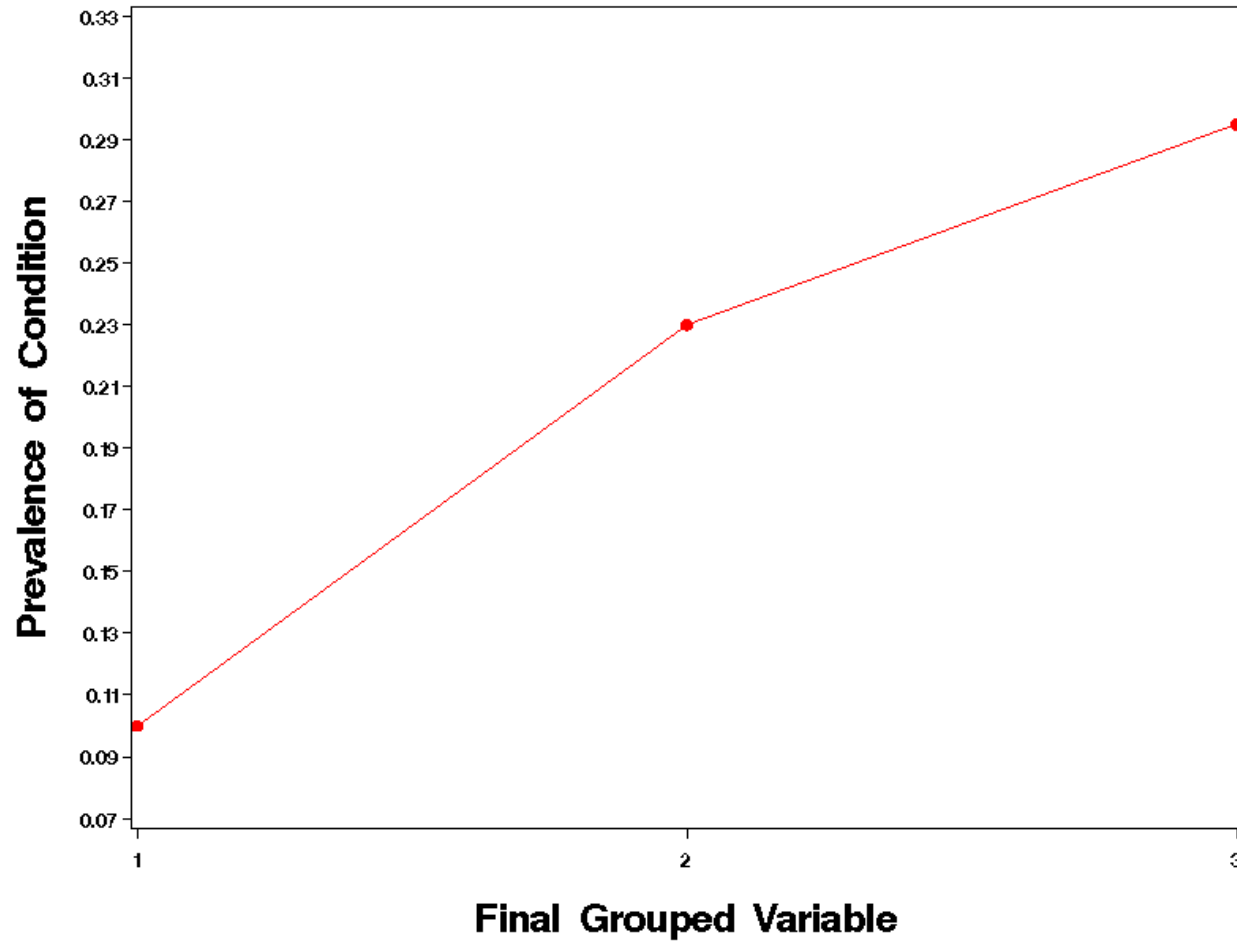
Data Preparation I



Data Preparation II



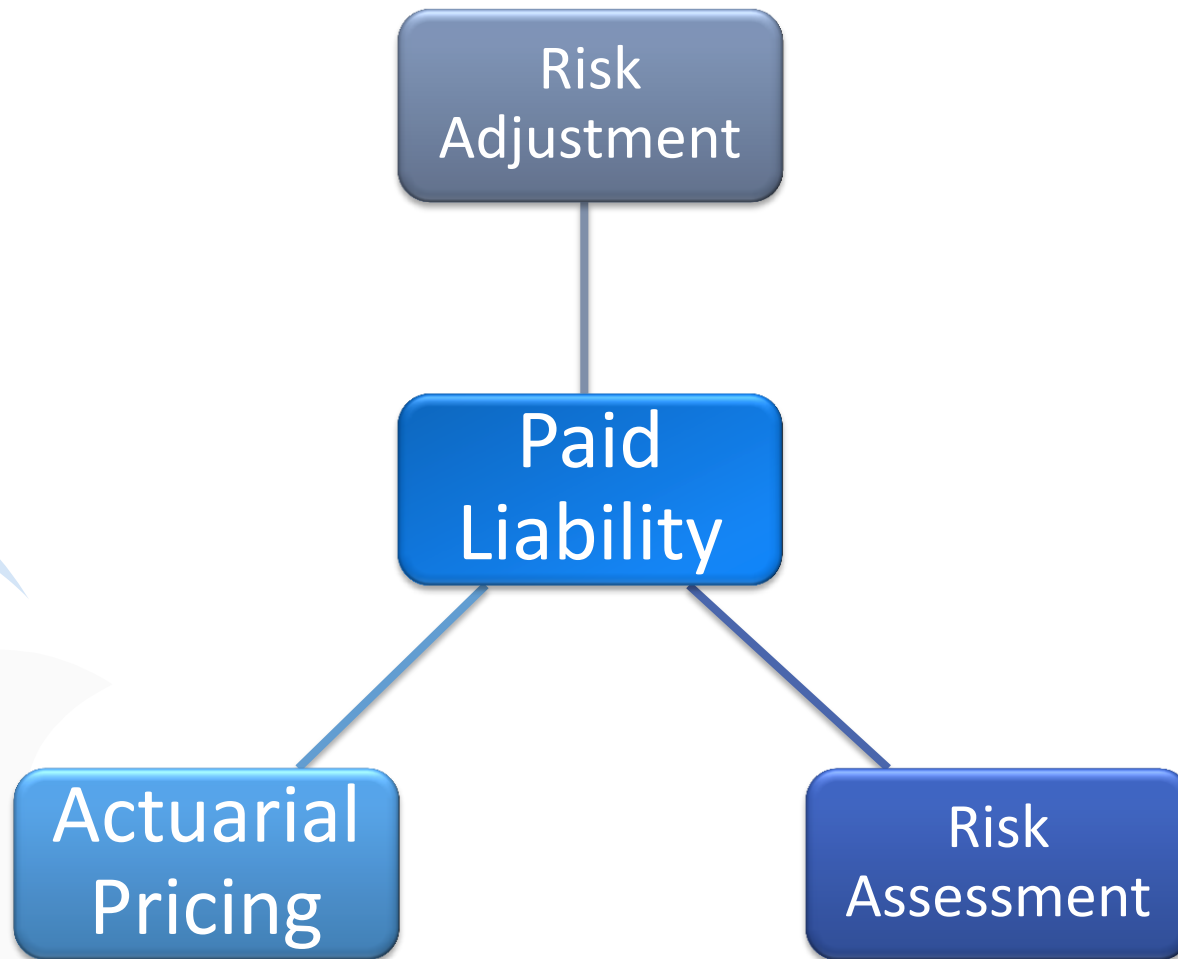
Data Preparation III



Missing Values of Input Variables

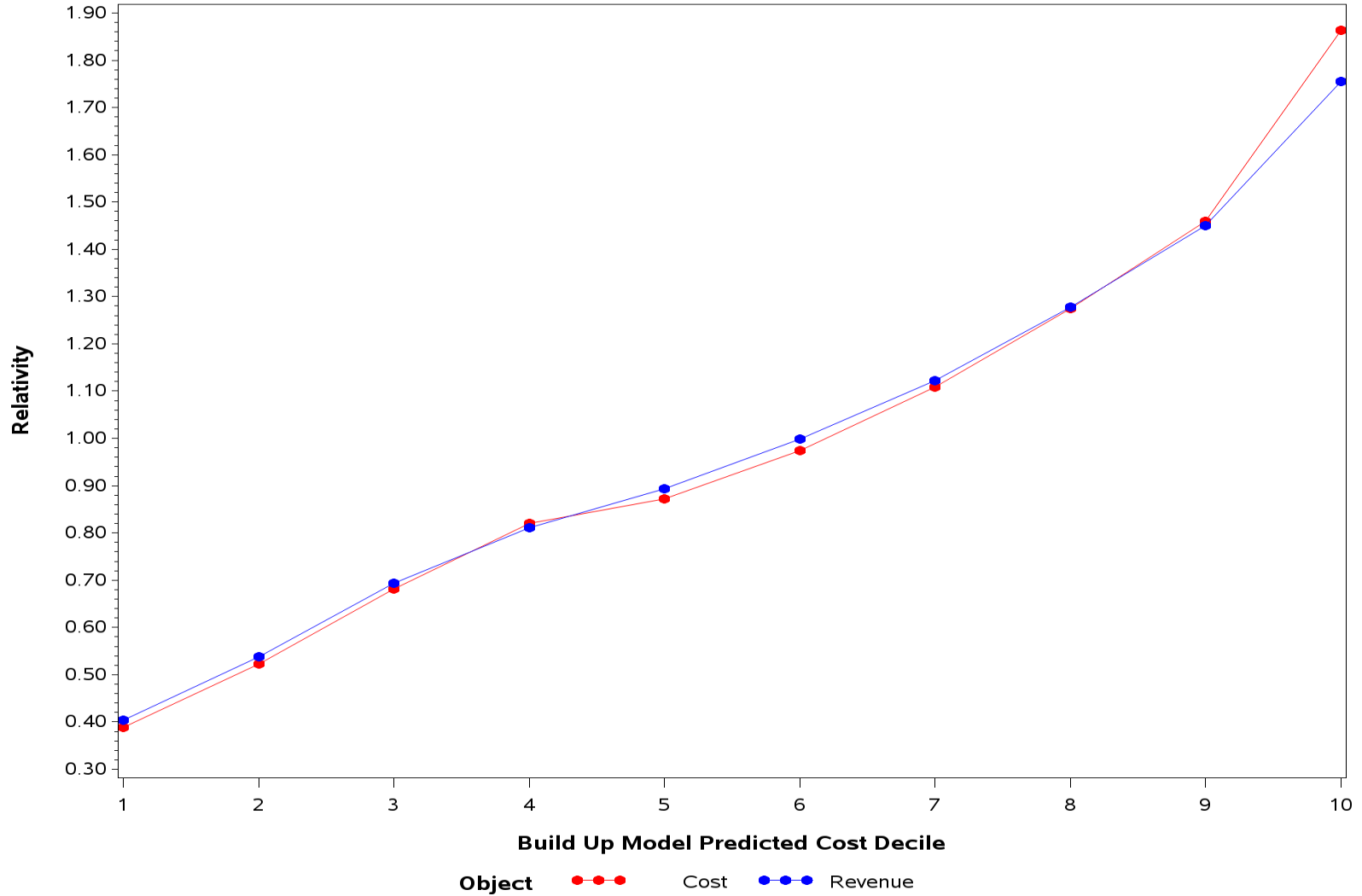
- Missing completely at random (MCAR)?
- Missing value imputation
- Missing value indicator
- Missing value patterns

Competing on Analytics

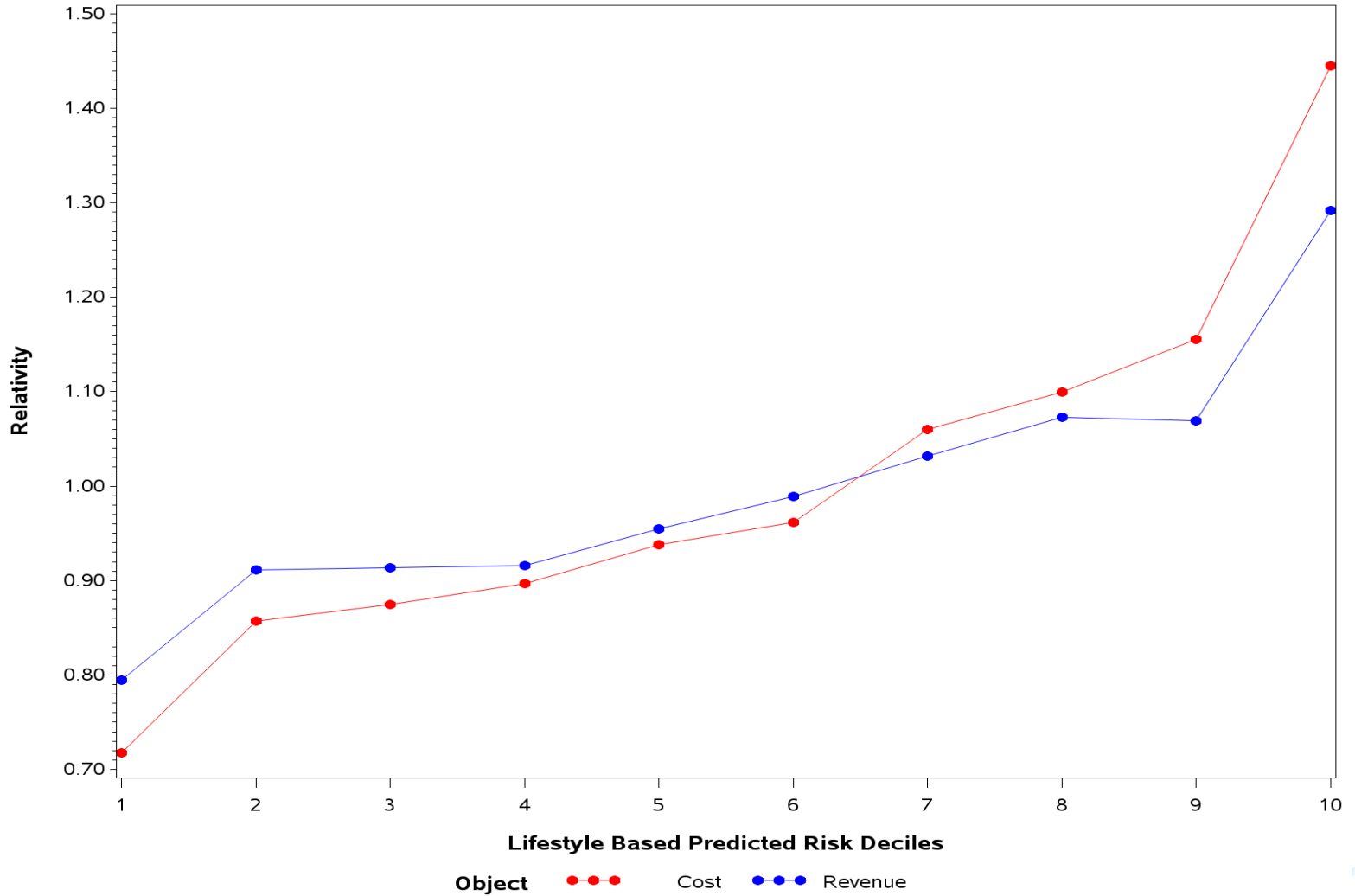


For Illustrative Purpose Only.

Margin 1



Margin 2



Example: Diabetes

X	Description	Prevalence
A	Health focused	6.66%
B	Success-oriented	4.99%
C	DIY	8.11%
D	Undisciplined	15.40%
E	Family centered	5.58%
F	Poor health	15.53%

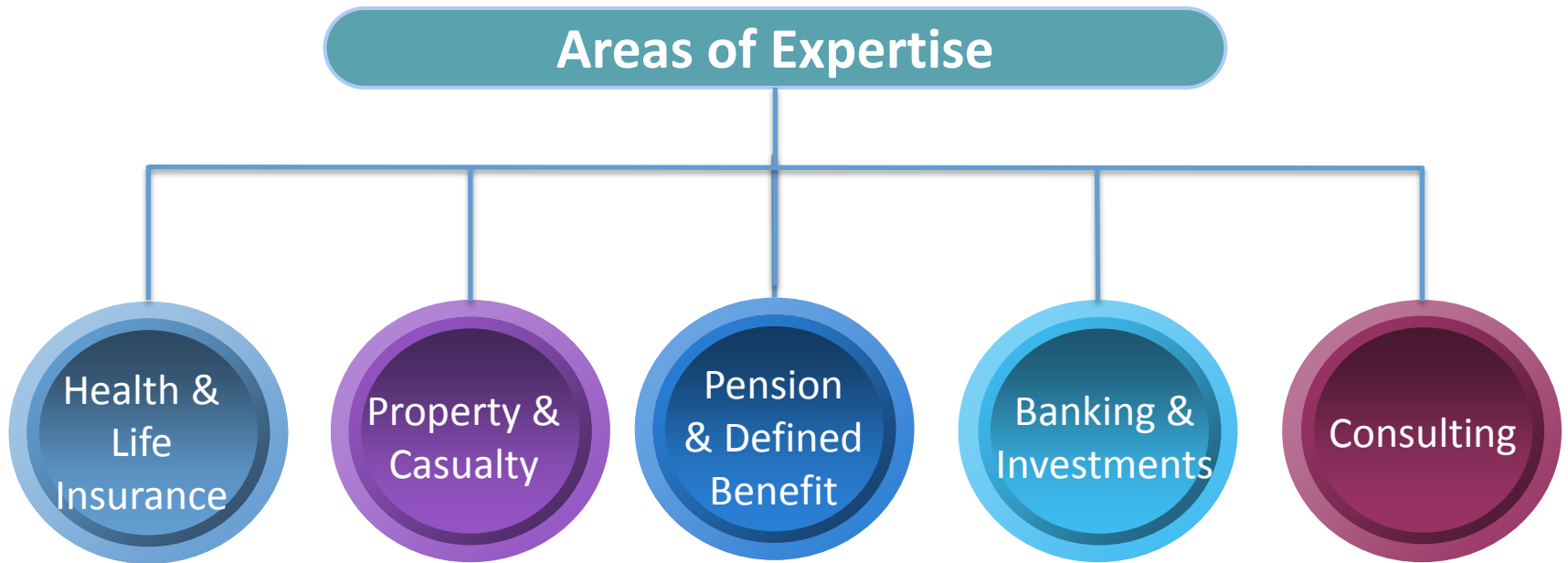
Example: Diabetes

Segmentation	Prevalence
1	0.98%
2	1.74%
3	2.49%
4	3.40%
5	4.68%
6	6.39%
7	8.13%
8	10.78%
9	15.00%
10	26.67%

Healthcare Predictive Modeling Opportunities

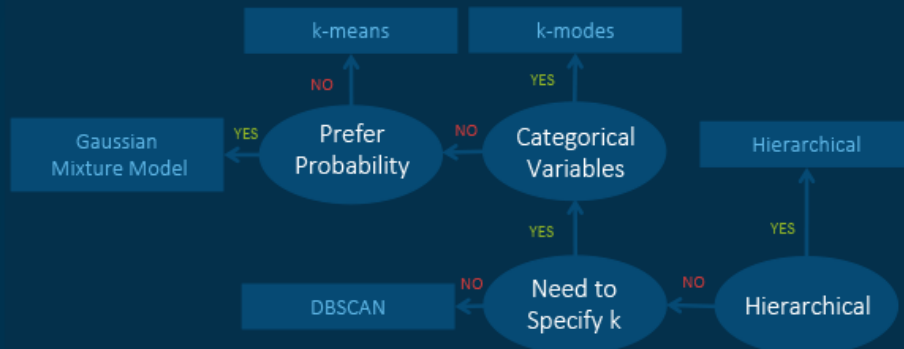
- Targeted Marketing, Outreach, and Retention Strategies through Predictive Modeling
- Risk Adjustment – How to Achieve Optimal Risk Scores and Avoid Transfer Payments
- Optimizing Care Management with Innovation - Leveraging Predictive Modeling for Proactive Identification & Micro-segmentation
- Using Predictive Analytics to Optimize Product Design and Pricing Strategies
- Applying Predictive Models to Improve Quality Measures
- Using Predictive Modeling to Prevent Fraud, Waste, and Abuse
- Disease Mappings
- Methods for Evaluation of Intervention Programs

Predictive Analytics in Other Actuarial Areas



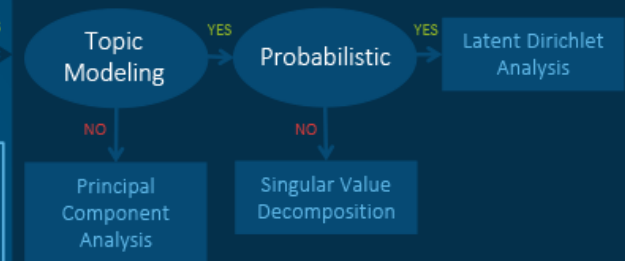
Machine Learning Algorithms Cheat Sheet

Unsupervised Learning: Clustering

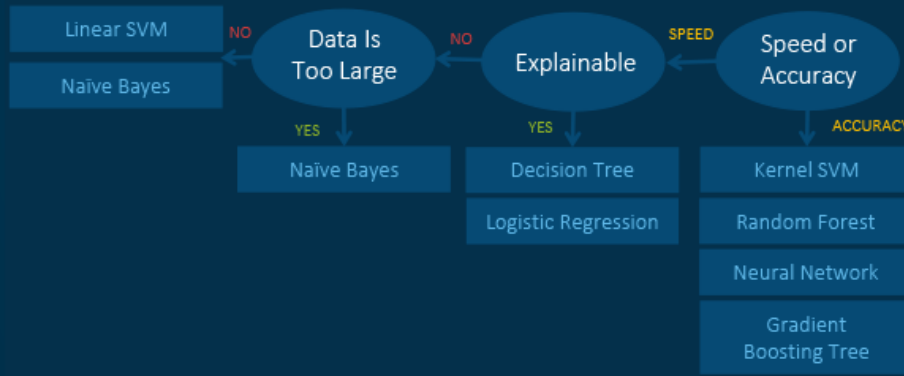


Unsupervised Learning: Dimension Reduction

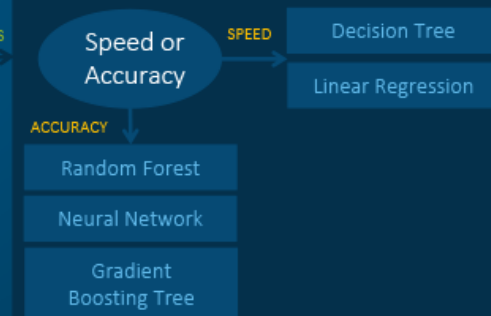
START



Supervised Learning: Classification



Supervised Learning: Regression



Source:

<http://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/> by Dr. Li

IT IS DIFFICULT TO MAKE
PREDICTIONS,
ESPECIALLY ABOUT THE FUTURE
-Danish Proverb