

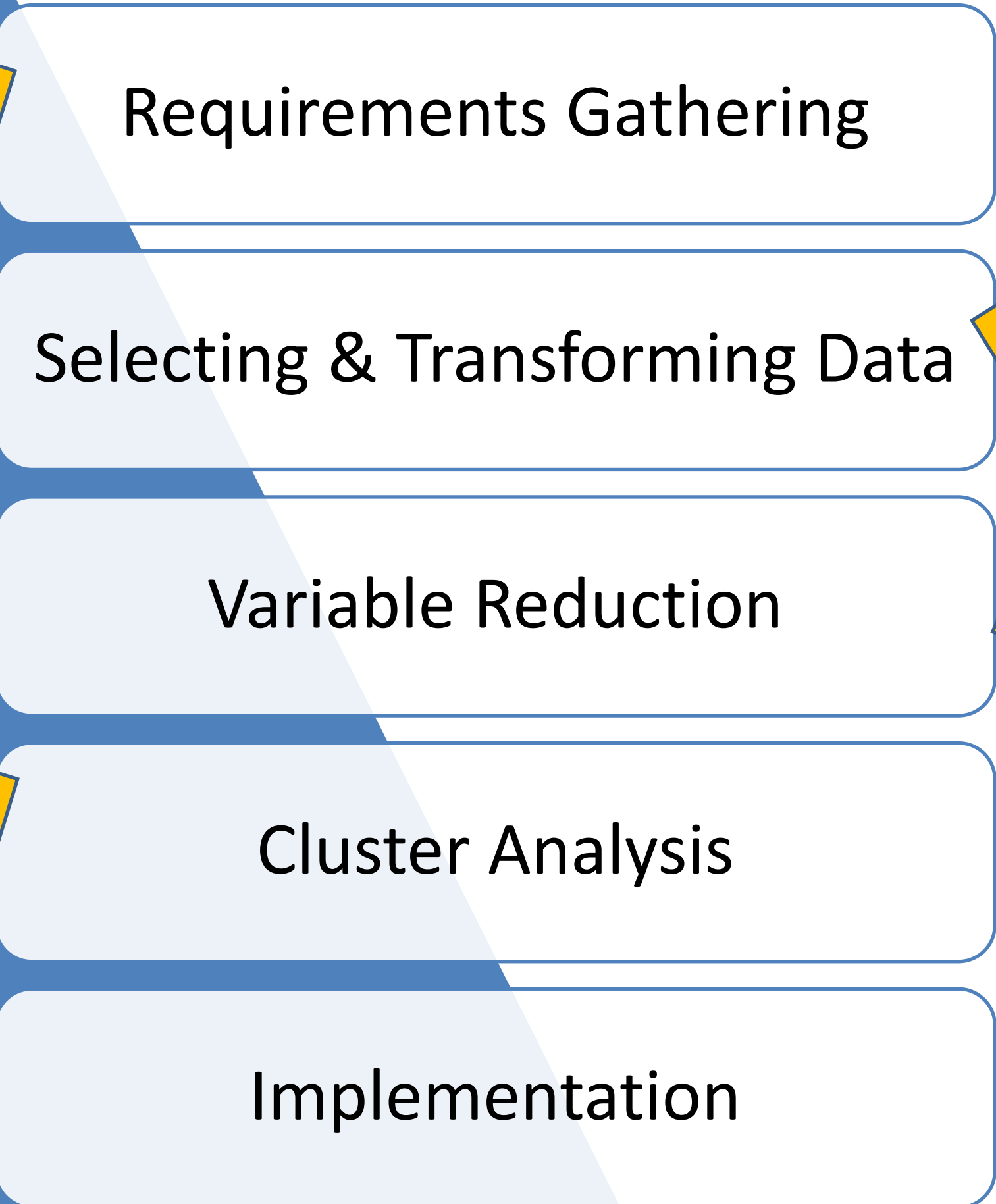
Population Segmentation in a Healthcare Environment

Enactment of the Affordable Care Act (ACA) impacted the United States health insurance industry in 2014. Previously, the insurer determined who their member population were based on actuarial practices, medical underwriting and product development. Now, member characteristics will vary and insurers will need to maintain a balance of high and low risk members to balance their portfolio of products to support business objectives.

MaryAnne DePesquo, BlueCross BlueShield of Arizona, USA



Project Overview



Requirements Gathering

- Target Population
- Project Plan
- Resource Allocation
- Stakeholders Engagement
- Actionable Outcomes

Selecting & Transforming Data

- Database 150k
- 2,600 Fields
- Exploration
 - PROC FREQ
 - PROCMEANS
- Normality
 - Log Transform
- Remove Outliers
- Group to Deciles
- Re-categorize
 - Missing Levels
 - Explode Levels
 - Combine Levels

Cluster Analysis
PROC CLUSTER
PROC FASTCLUS

PROC CLUSTER

- Hierarchal methodology
- Need to know underlying distribution
- Three methods - Average, Centroid, Ward's
- Used with survey data due to the ability to capture non-spherical clusters
- Determine number of clusters best fits data

```
PROC CLUSTER
DATA =mydat.azpop
METHOD=WARD
SIMPLE
CCC
PSEUDO;
VAR Insurvar1 Insurvar2 Channel Age
Income Dwelling Tenure;
RUN;
```

Variable Reduction Methods
PROC VARCLUS
PROC LOGISTIC

PROC VARCLUS

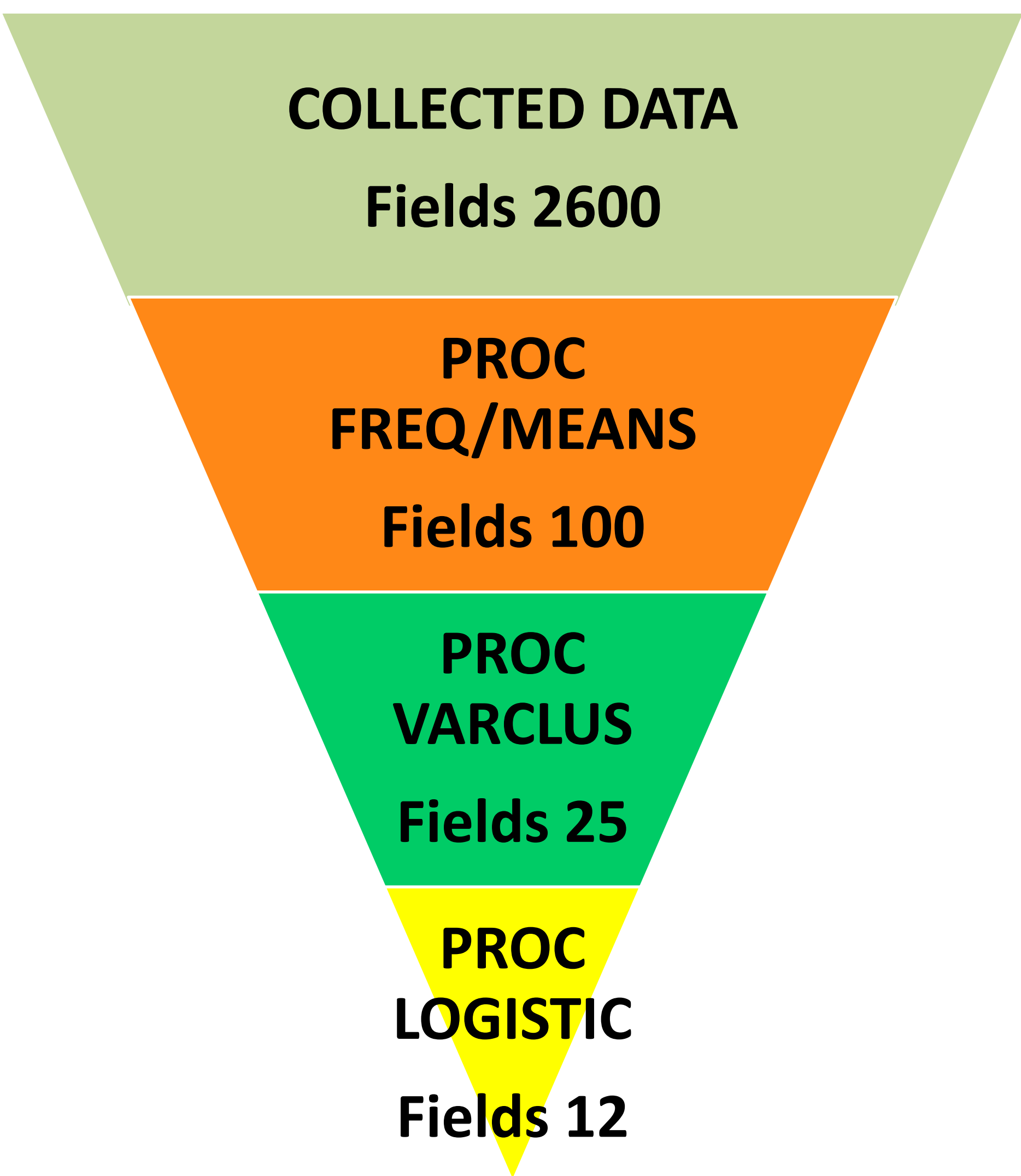
```
PROC VARCLUS
MAXEIGEN =0.7
DATA=mydat.analytic;
VAR age gender tenure
education ethnicity income
dwelling num_persons;
RUN;
```

Clusters		R-squared with		1-R**2
Cluster	Variable	Own Cluster	Next Closest	Ratio
Cluster 1	Age	0.6376	0.0073	0.3651
	Tenure	0.5176	0.0143	0.5942
Cluster 2	Gender	0.7754	0.0338	0.4108
	Education	0.0329	0.0038	0.1936
Cluster 3	Dwelling	0.5977	0.0214	0.2091
	Ethnicity	0.8909	0.0837	0.9164
Cluster 4	Num_persons	0.8252	0.0338	0.2385
	Income	0.7500	0.0056	0.4385

PROC LOGISTIC

```
PROC LOGISTIC DATA= mydat.analytic
DESCENDING;
MODEL Target_LR = Num_Persons
Gender Dwelling Prev_Insur
Tenure Age Ethnicity Deductible
Marital_Status Education /
SELECTION=BACKWARD SLS=0.05
CORRB RSQ LACKFIT
OUT= mydat.Logistic_out
Prob=pred_prob;
RUN;
```

Variable Reduction



Cluster Results
7 Clusters
Primary Drivers 10



Campaign Implementation

- ✓ Target 3 Clusters
- ✓ Matched 'Desired' Characteristics
- ✓ Top 10% in Each Cluster Selected
- ✓ Direct Mail Campaigns
- ✓ Monitor Response to Conversion