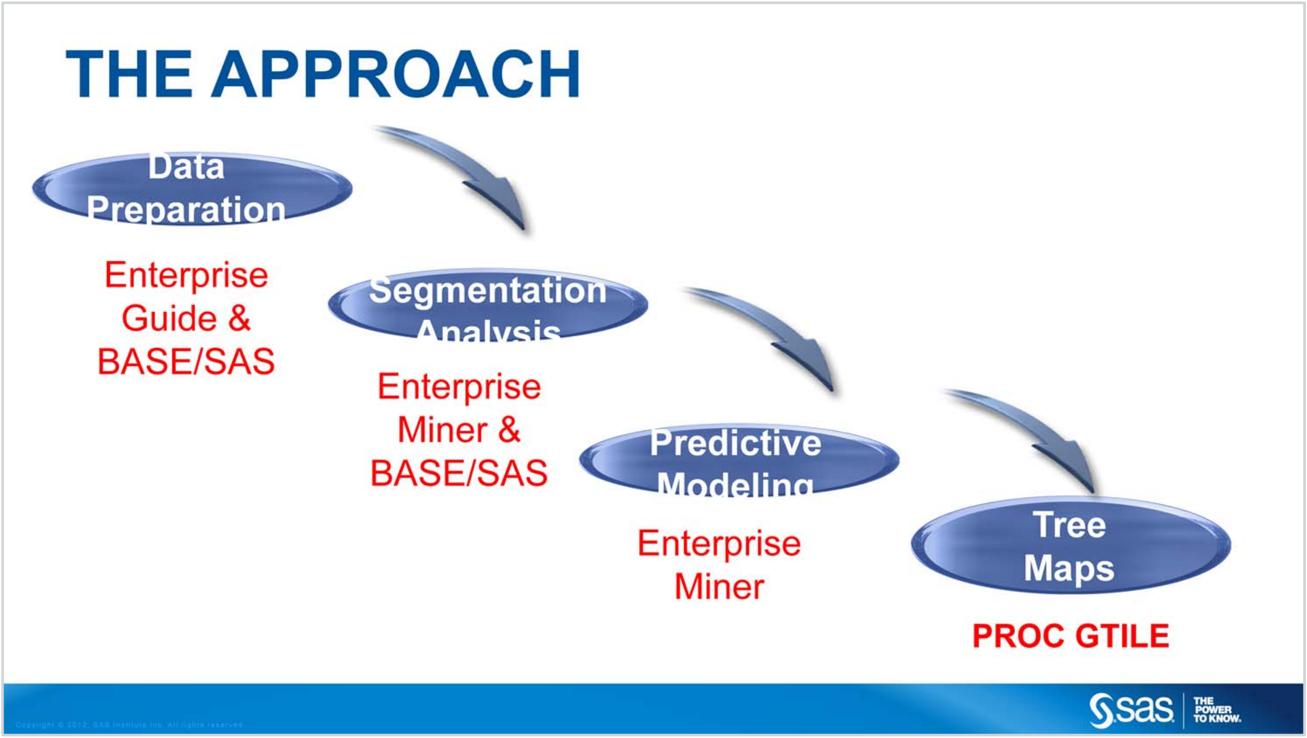
TAKING DISEASE AND HEALTH MANAGEMENT ANALYTICS INTO THE NEXT GENERATION

Wisconsin-Illinois SAS® Users Group
Des Plaines, IL
November 12, 2012

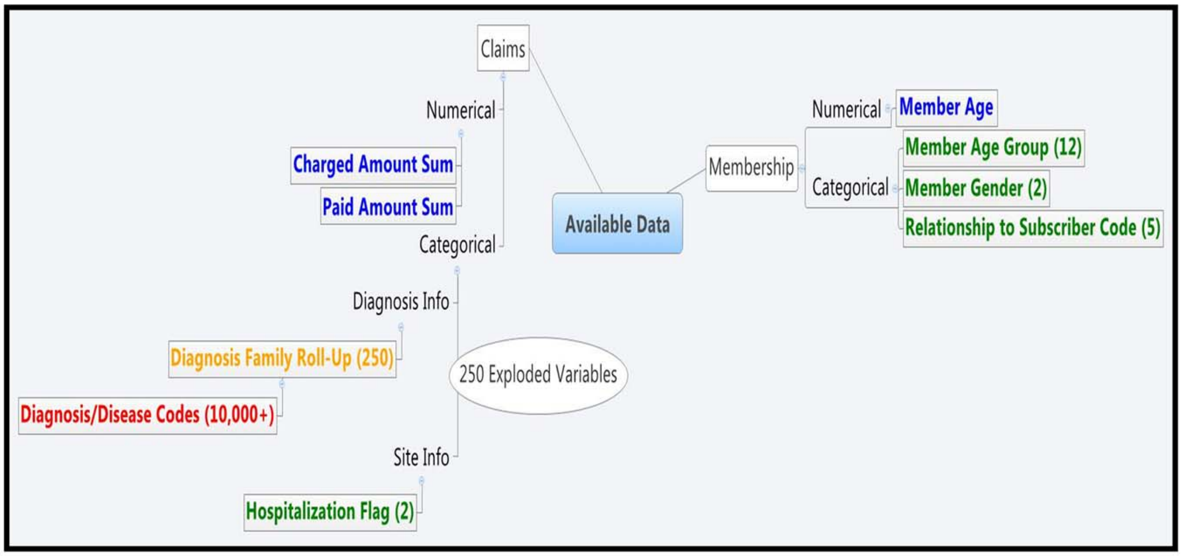
George S. Habek, M.S.,
Sr. Analytical Consultant,
New Analytical Solutions Enablement
Global Professional Services & Delivery

WHAT'S THE BIG IDEA?

- Do you struggle to visualize results of predictive modeling and segmentation within your healthcare member network?
- Would you also like to leverage new opportunities within specific patient/member segments?
- A streamlined data mining approach utilizing **PROC GTILE** to answer these questions has arrived.


THE APPROACH
Data Preparation
Enterprise Guide & BASE/SAS
Segmentation Analysis
Enterprise Miner & BASE/SAS
Predictive Modeling
Enterprise Miner
Tree Maps
PROC GTILE

PHASE I – DATA PREPARATION



- Blue represents numerical information.
- Orange represents categorical variables with many levels, which result in a cardinality issue.
- Green represents categorical attributes deemed to be acceptable for analysis.
- Red represents categorical information with too many levels, such as diagnosis codes.

PHASE II – SEGMENTATION ANALYSIS

- Four pieces of output

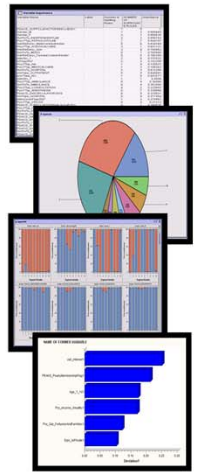
    I. Variable Importance Table

    II. Segment Size Graph

    III. Segment Plot

    IV. Cluster Profiling

        A. Descriptive Statistics
        B. Cluster Graphs



Copyright © 2012, SAS Institute Inc. All rights reserved.

PHASE II – SEGMENTATION ANALYSIS

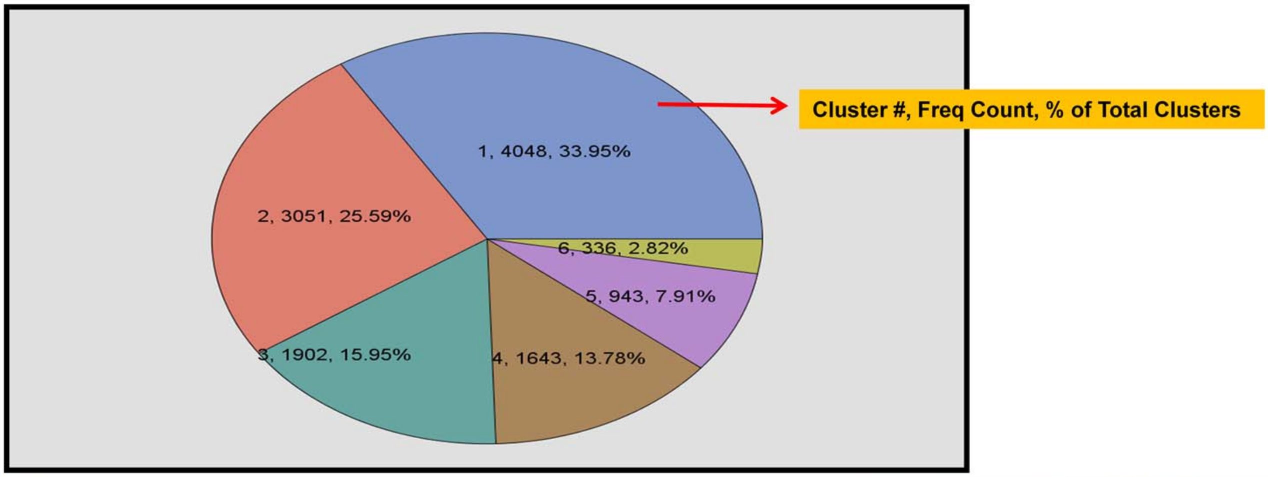
Variable Importance

NAME	IMPORTANCE
MEMBER_AGE_GROUP	1
Administrative_social_admission	0.896642401
Otitis_media_and_related_conditi	0.85285383
Other_perinatal_conditions	0.801733151
Liveborn	0.801567926
Fever_of_unknown_origin	0.801512875
Disorders_of_lipid_metabolism	0.738280497
Other_screening_for_suspected_co	0.723012037
Nonmalignant_breast_conditions	0.678949144
Medical_examination_evaluation	0.678829698
Hyperplasia_of_prostate	0.627194163
MEMBER_GENDER	0.595882802
Other_female_genital_disorders	0.567331453
Menstrual_disorders	0.564605493
Immunizations_and_screening_for	0.518495087
Other_connective_tissue_disease	0.462305147
Other_male_genital_disorders	0.448082511

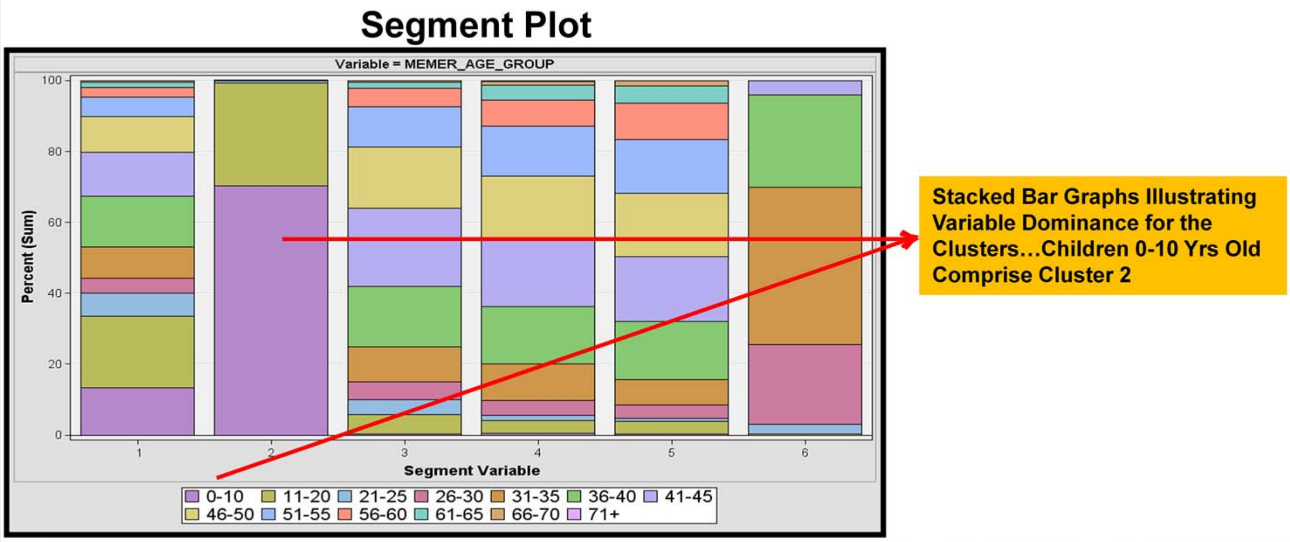
Overall measure of variable significance among the 6 clusters

PHASE II – SEGMENTATION ANALYSIS

Segment Size Graph



PHASE II – SEGMENTATION ANALYSIS



PHASE II – SEGMENTATION ANALYSIS

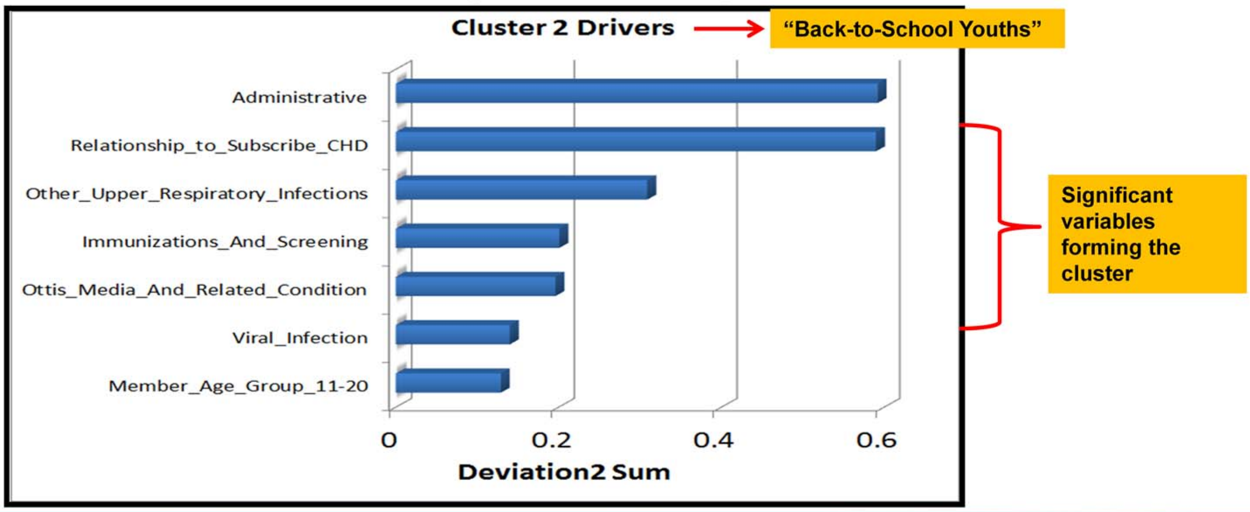
Cluster Profiling – Descriptive Statistics

Segment Id	N Obs	Variable	Mean
2	3051	member_age	7.56
		paid_amount_sum	$2,228.09
6	336	member_age	33.31
		paid_amount_sum	$10,764.86

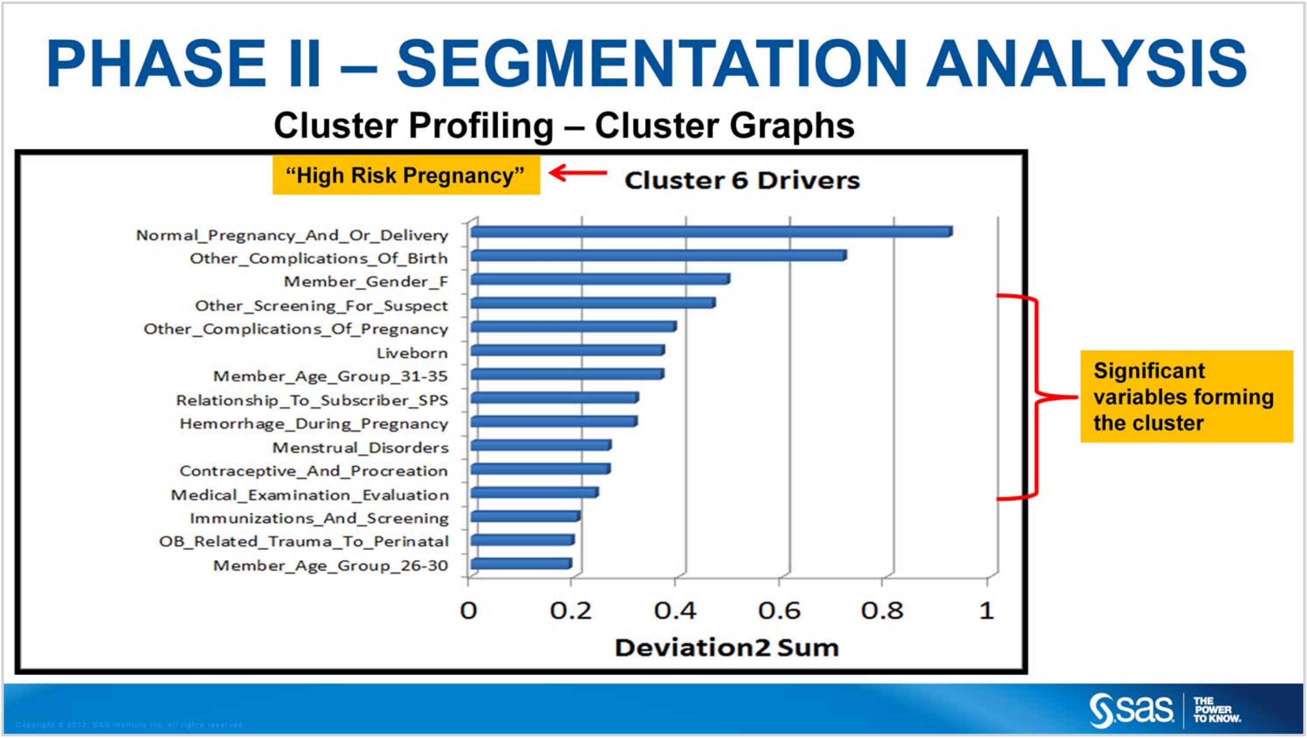
Copyright © 2012, SAS Institute Inc. All rights reserved.

PHASE II – SEGMENTATION ANALYSIS

Cluster Profiling – Cluster Graphs

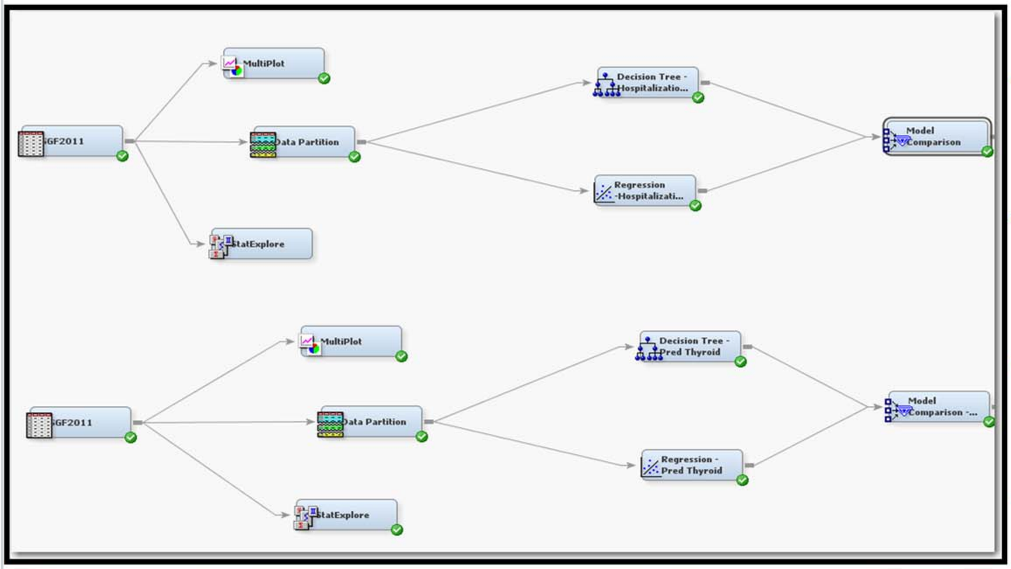


Two values are created for each cluster: 1) the average for the cluster and 2) the average for the overall population. The deviation is simply (1) – (2) → This is the true drivers for each of the clusters.



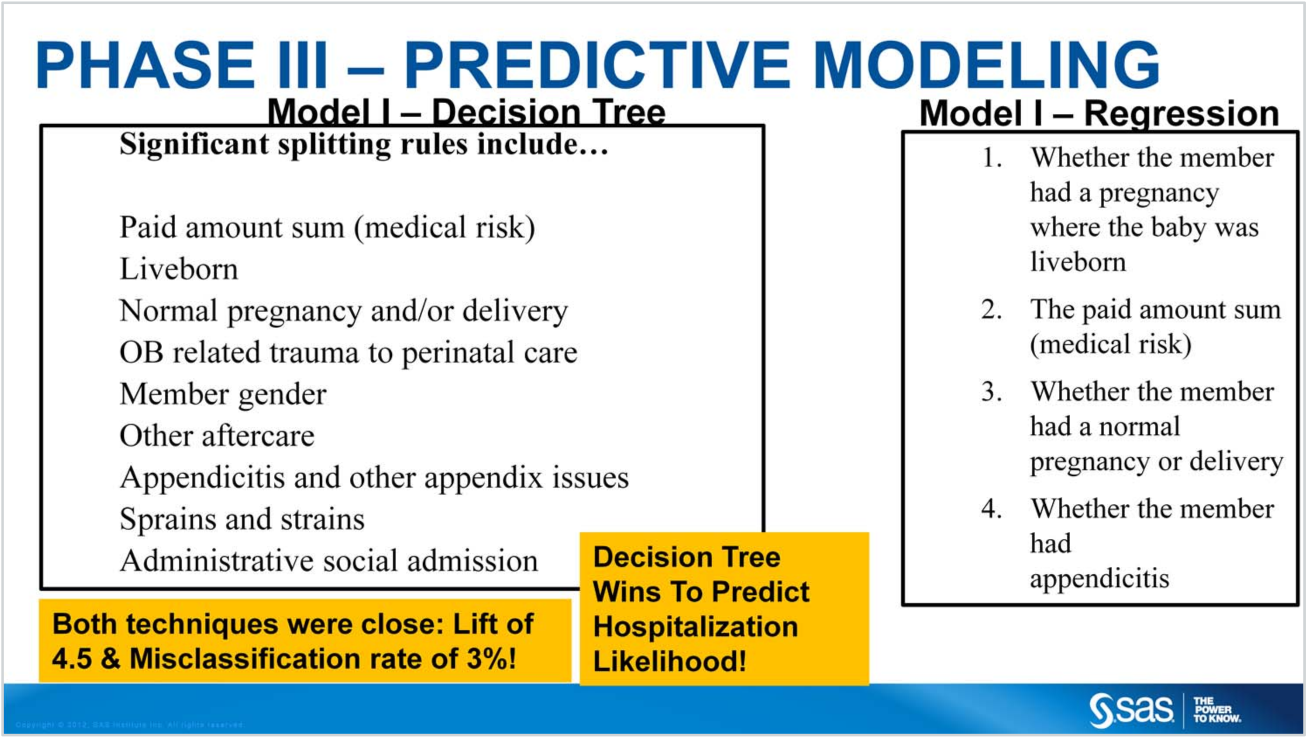
Two values are created for each cluster: 1) the average for the cluster and 2) the average for the overall population. The deviation is simply (1) – (2) → This is the true drivers for each of the clusters.

PHASE III – PREDICTIVE MODELING



Independent Model I To Clustering – Predict Hospitalization Likelihood

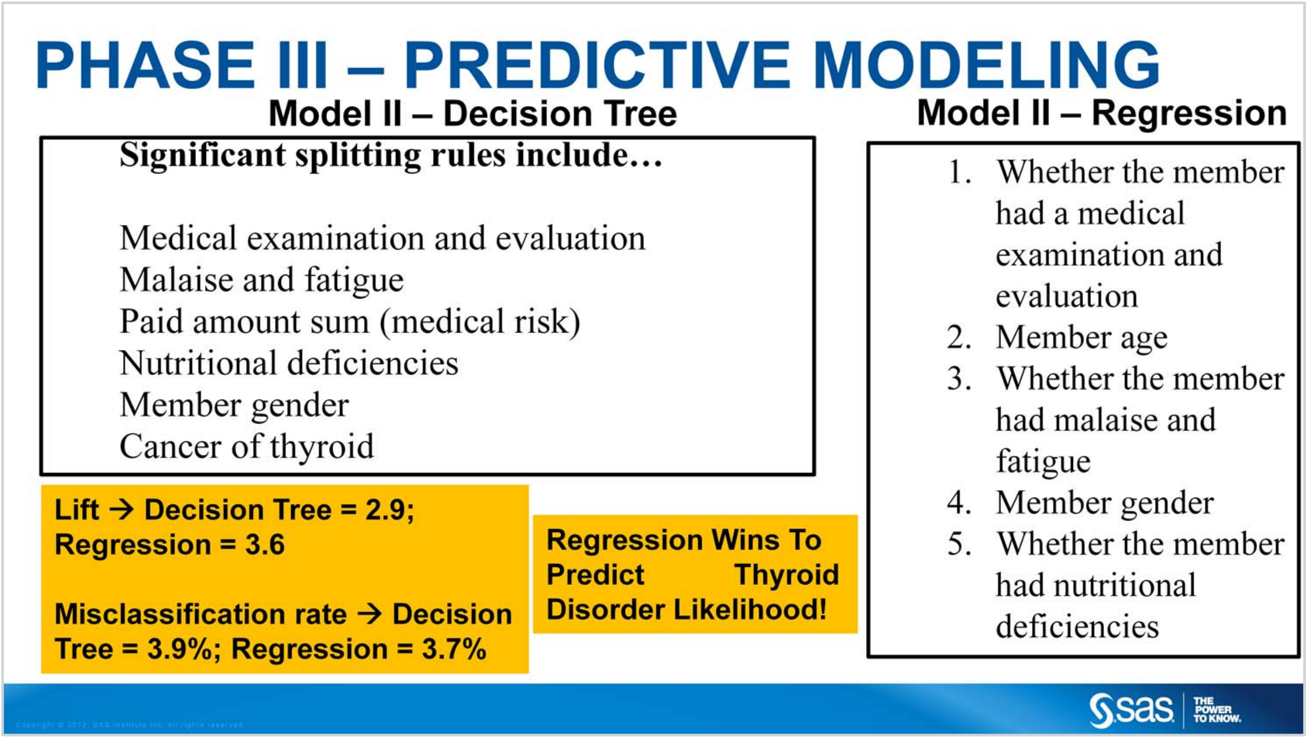
Independent Model II To Clustering – Predict Thyroid Disorder Likelihood



COST ANALYSIS

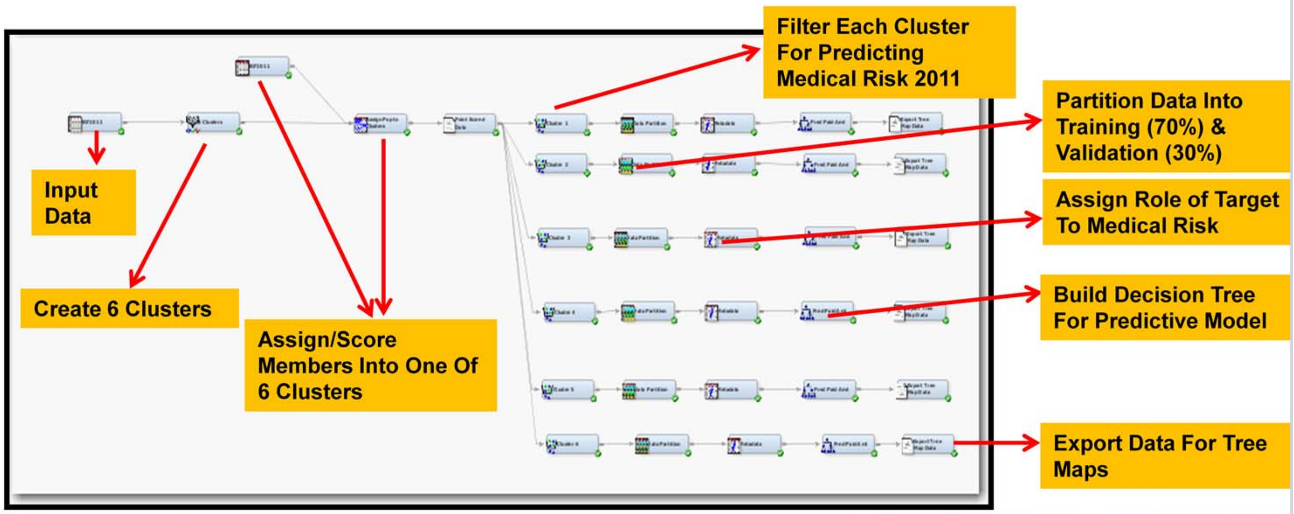
The median was chosen instead of the mean, since the median is not as sensitive to outliers within the data. In a cost analysis of Model I, the median paid amount sum (cost) for a patient/member being hospitalized is $14,078. Model I yielded a total of 201 false positives (an associated total cost of $2,829,678) and 175 false negatives (an associated total cost of $2,463,650). Although false negatives may be more of a concern because health initiative efforts must be increased in those cases, false positives and false negatives are both misclassifications from the model.

In a cost analysis of Model II, the median paid amount (cost) for a patient/member being diagnosed with a thyroid disorder is $2,129. Model II yielded a total of 13 false positives (an associated total cost of $27,677) and 438 false negatives (an associated total cost of $932,502). This is a high cost despite the fact that models I and II both have very good accuracy, so it might be useful to be rather strict when deeming what a “good” misclassification rate is (such as less than 1 percent).



In a cost analysis of Model II, the median paid amount (cost) for a patient/member being diagnosed with a thyroid disorder is $2,129. Model II yielded a total of 13 false positives (an associated total cost of $27,677) and 438 false negatives (an associated total cost of $932,502). This is a high cost despite the fact that models I and II both have very good accuracy, so it might be useful to be rather strict when deeming what a “good” misclassification rate is (such as less than 1 percent).

PHASE IV – LINKING OF SEGMENTATION ANALYSIS AND PREDICTIVE MODELING



PHASE V – TREE MAPS

Treemapping

From Wikipedia, the free encyclopedia

Treemapping is a method for displaying tree-structured data using nested rectangles.

Visualization of a directory tree (showing the size of each folder including its subfolders) by means of a treemap in TreeSize.

Contents [show]

Main idea [edit]

Treemaps display hierarchical (tree-structured) data as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing sub-branches. A leaf node's rectangle has an area proportional to a specified dimension on the data. (In the illustration, this is proportional to a waiting time). Often the leaf nodes are colored to show a separate dimension of the data.

When the color and size dimensions are correlated in some way with the tree structure, one can often easily see patterns that would be difficult to spot in other ways. A second advantage of treemaps is that, by construction, they make efficient use of space. As a result, they can legibly display thousands of items on the screen simultaneously.

PHASE V – TREE MAPS

```
LIBNAME SGF 'C:\SGF2011';
```
→ Assign Library & Path

```
DATA SGF.Finaltreemap;
  SET Treemapin;
```
→ Create Permanent Tree Map Data Set

```
ODS LISTING CLOSE;
```
→ Close Any Open Delivery System That Existed

```
ODS HTML FILE =
'C:\SGF2011\TreeMaps\clusterseg.html' GPATH = 'C:\';
```
→ Tree Map Location

```
GOPTIONS RESET = ALL DEVICE = JAVA HSIZE = 8.42
VSIZE = 5.31;
```
→ Tree Map Sizing

PHASE V – TREE MAPS

```
PROC GTILE DATA = SGF.Finaltreemap;

TILE Member_Count TILEBY=(Population_Cluster,Top_Clinical_Driver)

/COLORVAR = Predicted_Medical_Risk COLORRAMP=(green orange red);
RUN;
QUIT;

ODS HTML CLOSE;
ODS LISTING;
```

- PROC GTILE DATA = SGF.Finaltreemap; → Procedure Used & Tree Map Input Data Set
- Member_Count → Size Dimensions Of Tree Map Tiles
- Population_Cluster → Layer 1 - Cluster
- Top_Clinical_Driver → Layer 2 – Sub-group Within Cluster
- Predicted_Medical_Risk → Color Gradient Variable
- (green orange red) → Color Scale
- ODS HTML CLOSE; → Close Java Map
- ODS LISTING; → Close Output Delivery System

PHASE V – TREE MAPS

The Tree Map – Input Data

Population_Cluster	Top_Clinical_Driver	Predicted_Medical_Risk	Member_Count
Back-to-School Youths	Abdominal_pain Otitis_media_and_related_conditi	$ 3,402	29
Back-to-School Youths	Acute_and_chronic_tonsillitis Otitis_media_and_related_conditi	$ 4,535	32
Back-to-School Youths	Joint_disorders_and_dislocations MEMBER_AGE	$ 2,583	23
Back-to-School Youths	MEMBER_AGE Acute_and_chronic_tonsillitis	$ 2,518	30
Back-to-School Youths	MEMBER_AGE Fracture_of_upper_limb Otitis_media_and_related_conditi	$ 6,192	12
Back-to-School Youths	Medical_examination_evaluation Otitis_media_and_related_conditi	$ 3,818	9
Back-to-School Youths	Other_aftercare	$ 3,637	25
Back-to-School Youths	Other_gastrointestinal_disorders	$ 1,994	36
Back-to-School Youths	Other_nervous_system_disorders	$ 2,537	38
Back-to-School Youths	Otitis_media_and_related_conditi	$ 1,655	357
Female Routine Utilizers	Disorders_of_lipid_metabolism	$ 2,094	37
Female Routine Utilizers	Disorders_of_lipid_metabolism Nonmalignant_breast_conditions	$ 7,763	6
Female Routine Utilizers	Headache__including_migraine	$ 3,750	17
Female Routine Utilizers	MEMBER_AGE	$ 1,361	442
Female Routine Utilizers	Nonspecific_chest_pain	$ 3,618	27
Female Routine Utilizers	Nutritional_deficiencies Disorders_of_lipid_metabolism	$ 8,054	6
Female Routine Utilizers	Spondylosis__intervertebral_disc	$ 2,488	38
High Risk Pregnancy	Hemorrhage_during_pregnancy__abr Other_complications_of_pregnancy	$ 3,957	6
High Risk Pregnancy	Other_complications_of_pregnancy	$ 3,598	7
Male Accidental Youths	Joint_disorders_and_dislocations	$ 2,295	19
Male Accidental Youths	Other_male_genital_disorders	$ 3,046	13

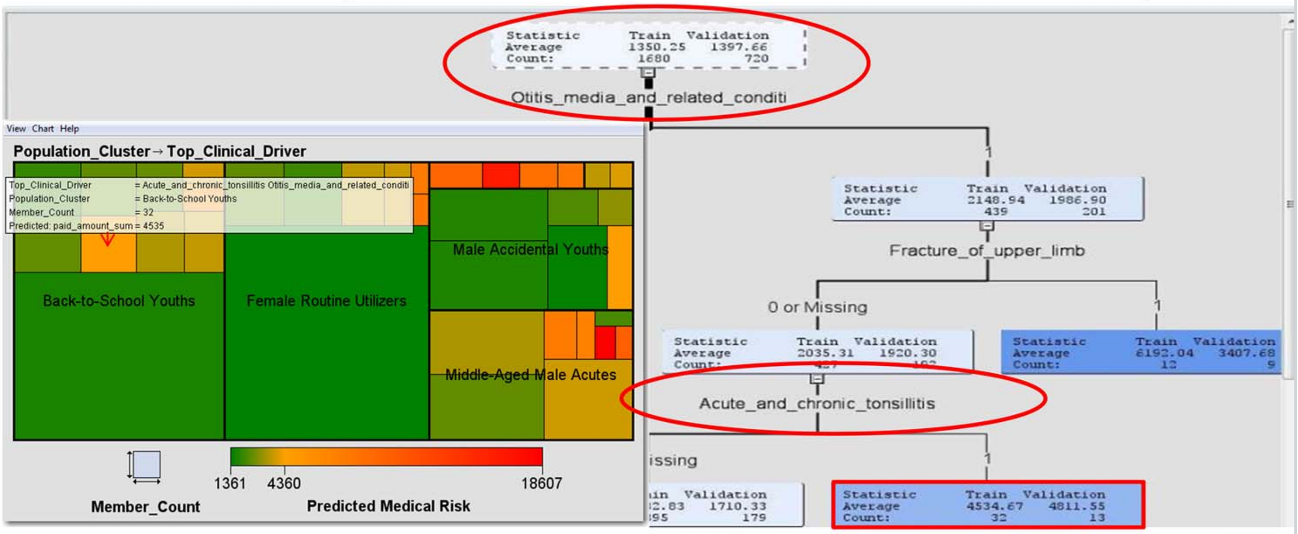
Layer 1 - Cluster

Layer 2 – Sub-group Within Cluster

Predicted Medical Risk From EM → Tree Map Color Gradient

Frequency Member Count Within Sub-group

Proactively Identified Avg Medical Risk of $145,120
For a SubGroup within the Back-to-School Youths Segment



Tree Map Demonstration


TAKING DISEASE AND HEALTH MANAGEMENT ANALYTICS INTO THE NEXT GENERATION
Wisconsin-Illinois SAS® Users Group
Des Plaines, IL
November 12, 2012
George S. Habek, M.S.,
Sr. Analytical Consultant,
New Analytical Solutions Enablement
Global Professional Services & Delivery
SAS Institute Inc.
7855 Whitetail Drive
Racine, WI 53406 USA
WWW.SAS.COM
George S. Habek, M.S.
Sr. Analytical Consultant
New Analytical Solutions Enablement
Global Professional Services & Delivery
Tel: +1 262 884 4336
Fax: +1 262 884 4336
Mobile: +1 262 506 4948
E-mail: george.habek@sas.com
THE POWER TO KNOW.
THE POWER TO KNOW.