The freedom to better

market — sell — serve — operate

the way you want

Airline Market Segmentation

Himadri Mukherjee, Lakshmi Manasa Kasivajjula, Vineeth Patapati

Sabre / Airline Solutions

Presentation outline

- Background, problem overview
- Market segmentation methodology
- Results
- Directions for future work

Planning and Scheduling – Profitability Forecasting & Calibration

Long-Term Planning	Medium-Term Planning And Scheduling	Short-Term Scheduling	Day Of Departure

Profitability Forecasting
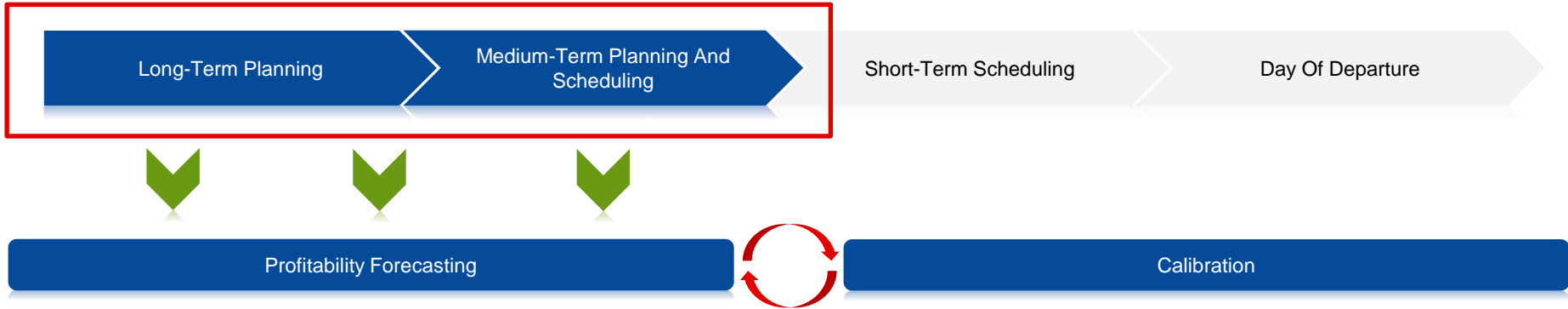
Decision-support tool used to forecast

- Network profitability
- Schedule evaluation for new destinations, hub structures, alliances/code-shares
- Gauge competitor impact

Calibration

Calibration is a seasonal and iterative process to

- Define and refine connection, market share and other parameters
- Impact of calibrated parameters varies with the network size and structures.

- Calibration parameters are currently generated either at entity (region pair/market group) level or market level
- Input data for the analysis is also currently picked at entity level or market level

Sabre / Airline Solutions

# Market Segmentation - Problem Background

Motivations

Current state:

- Entities (Region Pairs) are heterogeneous, purely geographical concept based & not dynamic or data driven.

- This leads to increased calibration time due to market specific adjustments
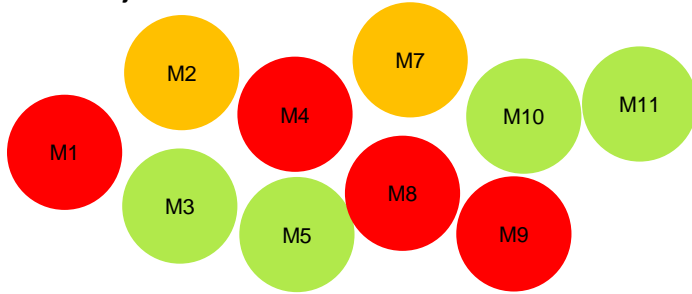
Expected state:

- Market groups which are homogeneous, capture passenger behavior, dynamic and data driven
- Less or no need for market specific over rides
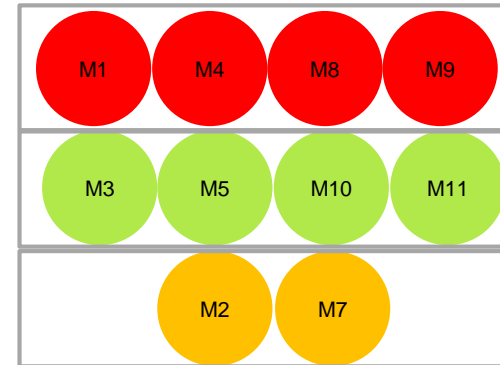
Expectations from Optimum Market Grouping

- Lesser number of market-groups

- Higher calibration accuracies

- Lesser number of overrides

- Accurate allocation of new markets to market groups

- Faster calibration time

- Improved future forecasting capability with dynamic allocation or markets based on attribute changes

Segmentation

A **meaningful** grouping of **entities** to enable focused business **decisions**

- **Distinct** groups of similar entities

- **Intuitive**

- **Verifiable** against business knowledge

- **Relevant** to the decision-making context

CUSTOMER

TRIP

FLIGHT

MARKET

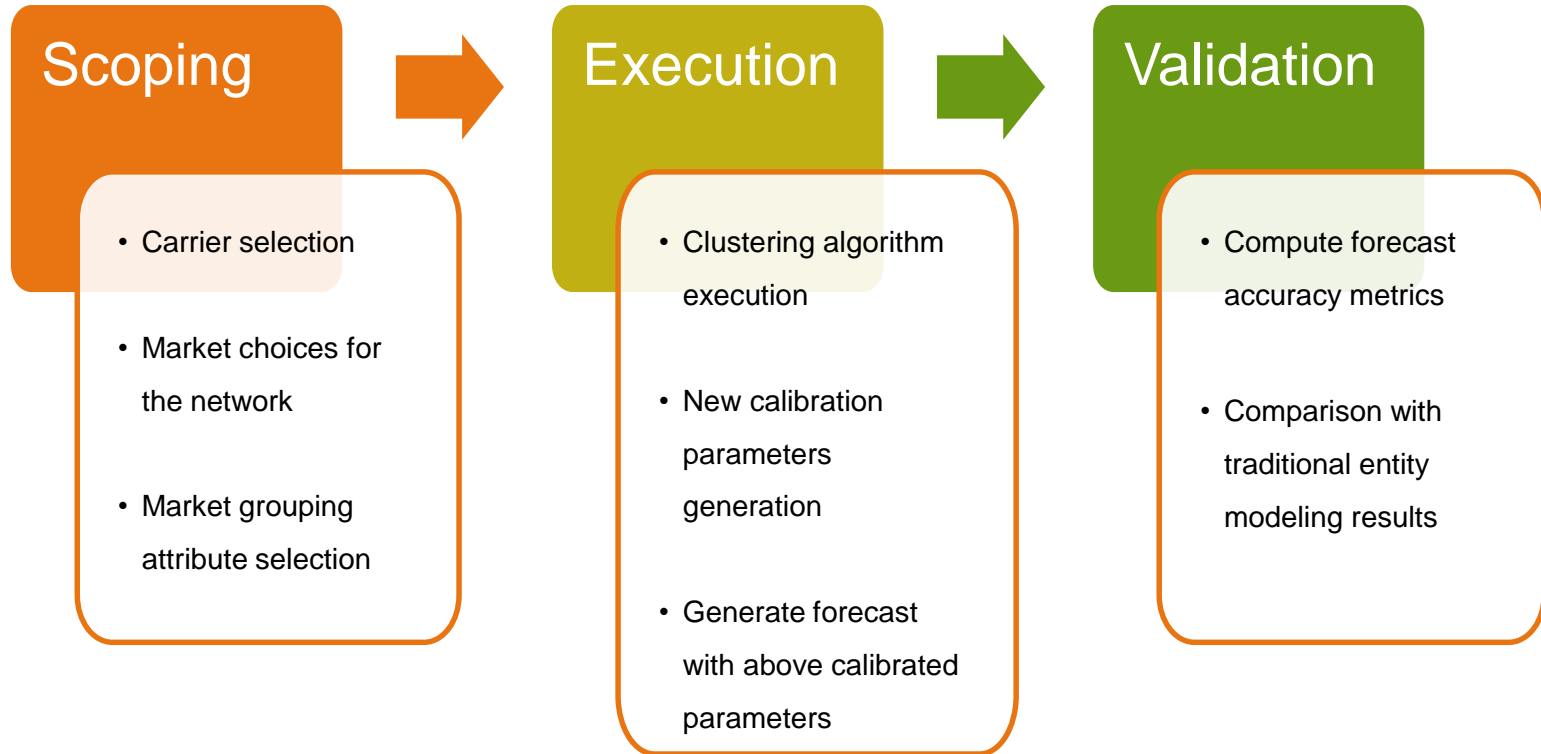Customer Management

Network Management
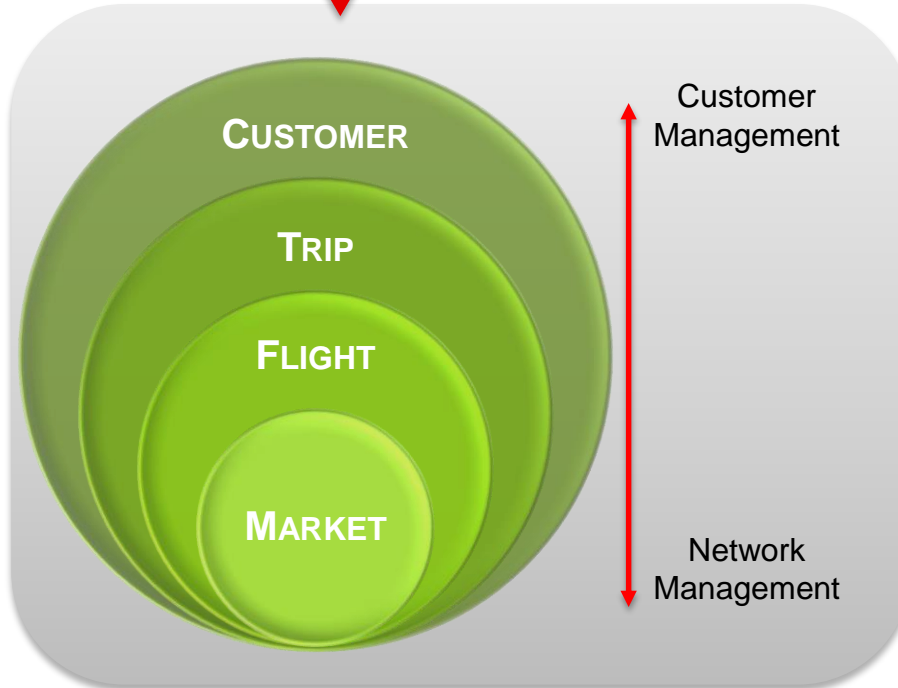
- **Describe** group-level behavior

- **Predict** behavior by group with what-if analysis

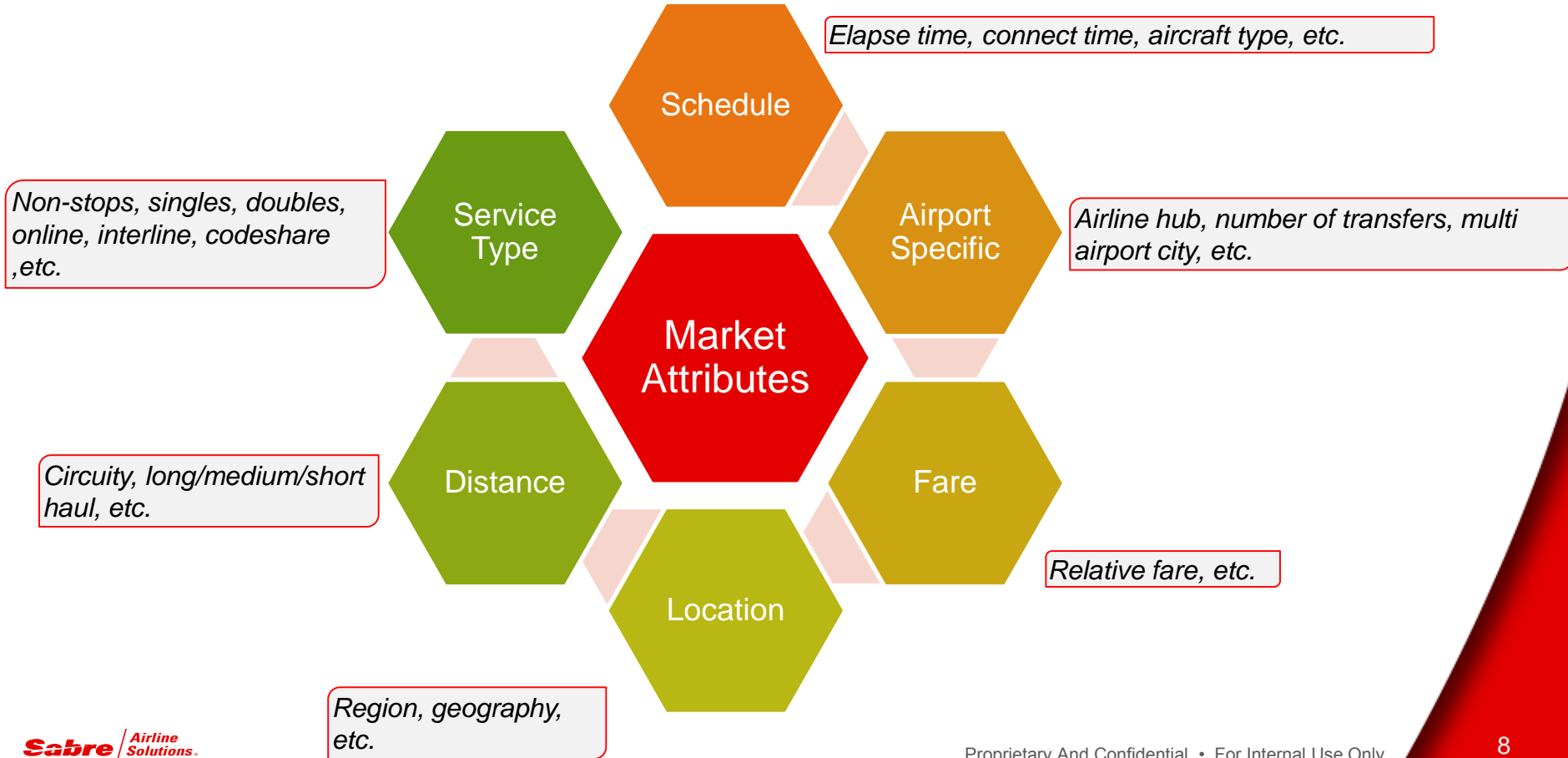- **Influence** group behavior through customized decisions

Segmentation Use Cases

- Shopping requests

- Trips

- Hotel reservations

- Customers

- Destinations

- Markets

Market Segmentation - Benchmark Methodology

Scoping

- Carrier selection

- Market choices for the network

- Market grouping attribute selection

Execution

- Clustering algorithm execution

- New calibration parameters generation

- Generate forecast with above calibrated parameters

Validation

- Compute forecast accuracy metrics

- Comparison with traditional entity modeling results

Market Attributes

Schedule — Elapse time, connect time, aircraft type, etc.

Airport Specific — Airline hub, number of transfers, multi airport city, etc.

Service Type — Non-stops, singles, doubles, online, interline, codeshare ,etc.

Market Attributes

Distance — Circuity, long/medium/short haul, etc.

Fare — Relative fare, etc.

Location — Region, geography, etc.

Sabre / Airline Solutions

Shortlisted Market Characteristics

- Market circuitry (ratio of total distance travelled by the itinerary to the non stop distance of the market)

- Market elapsed time ratio (ratio of elapsed time of itinerary to the shortest elapsed time possible)
- Market connect time
- Departure TOW

- Number of distinct itineraries in a market
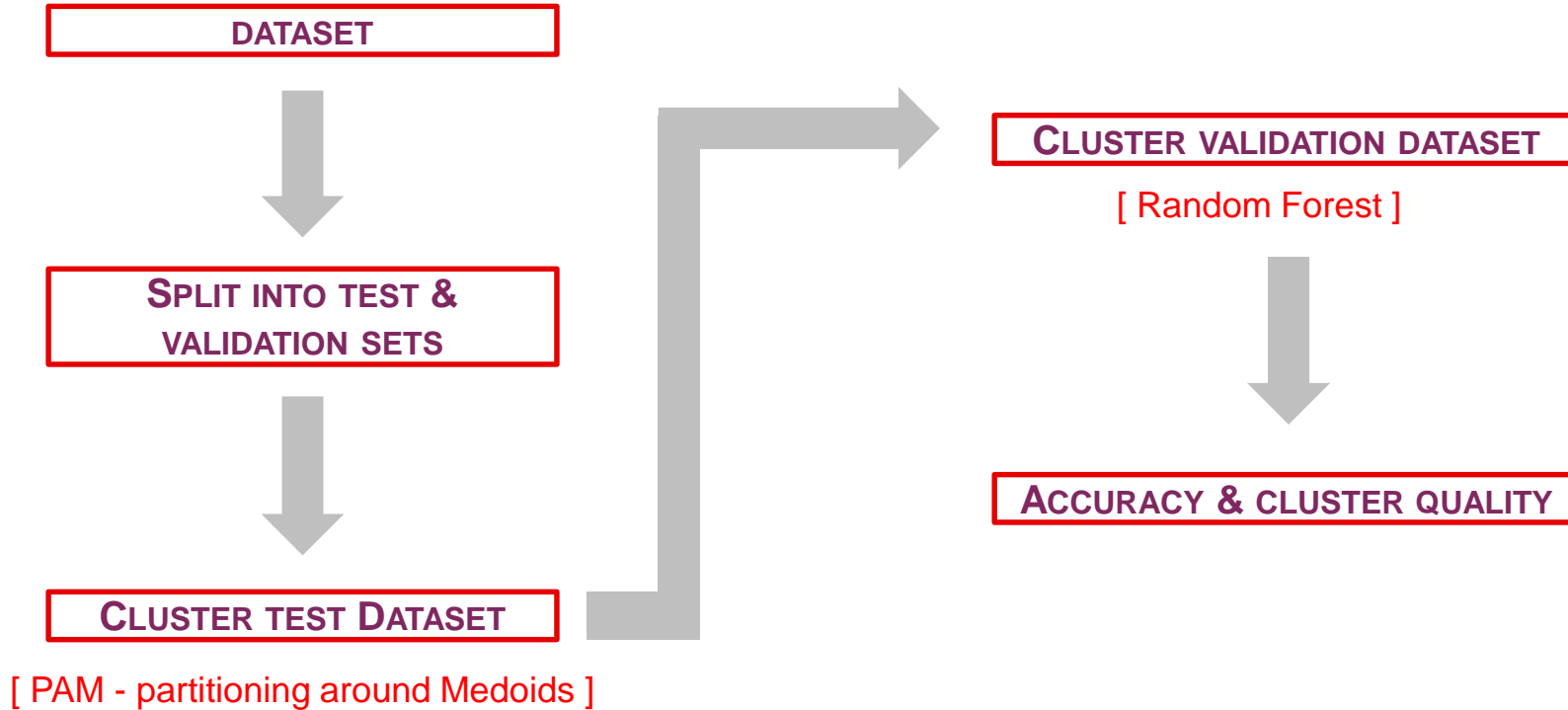- Market size

- Aircraft type
- Service types (traffic share of service types)
  - Non stop, Single online, Double Online, Single Interline, Double Interline

- Geography
- LCC presence

Clustering Methodology

DATASET

SPLIT INTO TEST & VALIDATION SETS

CLUSTER TEST DATASET

[PAM - partitioning around Medoids]

CLUSTER VALIDATION DATASET

[Random Forest]

ACCURACY & CLUSTER QUALITY

PAM (Partitioning around Medoids)

- o More robust to noise and outliers as compared to k-means because it minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances

- o Applicable for data sets which has continuous as well as categorical variables

- o Dissimilarity matrix is constructed using Gower's general similarity coefficient

1. A set of medoids is chosen at random

2. Compute distances to other points

3. Data is clustered according to the medoid they are most similar to

4. Medoid set is optimized by iteration

Where cost between any two points is

$$\text{cost}(x, c) = \sum_{i=1}^{d} |x_i - c_i|$$

where x is any data object, c is the medoid, and d is the dimension of the object

Sabre / *Airline Solutions*

Optimal Number of Clusters

To find optimal number of clusters: Silhouette width

High value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

a(i) be the average dissimilarity of i with all other data within the same cluster

b(i) is the lowest average dissimilarity of i to any other cluster, where i is not a member

Silhouette Width – average of silhouettes over the entire dataset
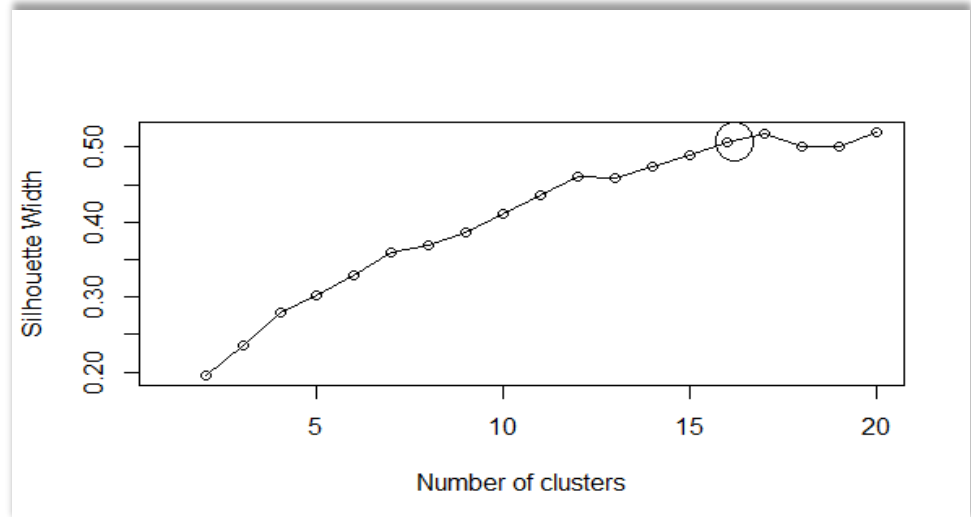
Cluster Characteristics

Cluster	Characteristics
1	*Urgent Traveler Markets*
2	*Budget Traveler Markets*
3	*Loyalty Traveler Markets*
4	*Business Traveler Markets*
5	*Traditional Traveler Markets*
6	*Leisure Traveler Markets*
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
16	*Relaxer Traveler Markets*

2st principle component
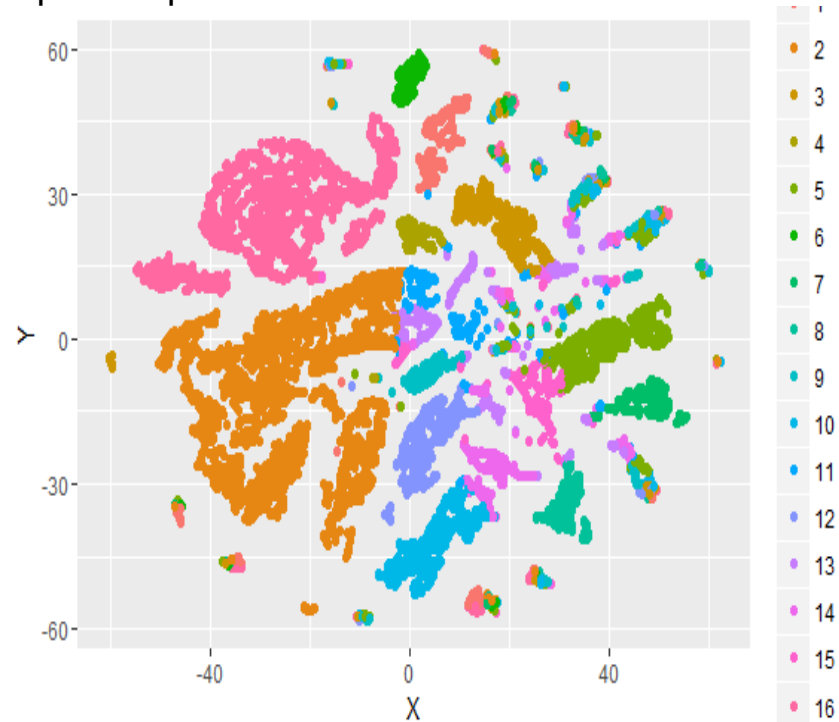
1st principle component

Examples

By standard entity-based method all the markets are grouped in the same entity (WEU_SCAND),

but by the new method these markets fall under different market-groups as:

- TXL_ARN has higher pax,non stop and high distinct itins
- TXL_KTT has low pax, low distinct itin and mostly single online service type
- TXL_GOT has medium pax and medium distinct itins compared to other two markets

By standard entity-based method the above markets are grouped into different entities like CNAM_SSAM/UK_IND/SME_EASIA based on geography,

but by the new method these markets fall under one single market-group due to similar market characteristics like :

higher pax, similar departure times and higher distinct itineraries

Calibration KPIs – Entity parameters vs. Cluster parameters

Dimension	Standard Entity-Based Method	New Clustering/Segmentation Method(Only connection attributes)	New Clustering/Segmentation Method(connection + other attributes)
Number of market groups	100+	16	16
OA Match Percentage	88%	86%	90%
OA Overbuild Ratio	1.68	1.8	1.69
Host Match Percentage	93%	94%	95%
Host Overbuilt Ratio	2.91	3.8	3.31
Market-Share Error	2.7%	2.9%	2.8%
Number of overrides	0	0	0
Automation	*Static/Manual*	*Completely automated*	
Time to calibrate	*Estimated saving of 1 week's effort (13%)*		
Attributes considered	*Only geography*	*Geography + several other*	

Except 'Host Overbuild' the new method is superior or similar in all the counts

Improvement in all the counts by increasing attributes

Next Steps

Model Enhancement

- More market attributes into the model

- Robust Modeling methods:
  - ✓ Ensemble approaches for accuracy improvement in grouping

Model Validation

- Tests across various kinds of airline networks

- Model validation for calibration process improvement checks

- Model validation for decision support improvement checks