

# Market Clustering

# Market Clustering

- Creates small entities or groups of markets that exhibit similar behavior, which makes the Market Share algorithm more responsive to regional behavior
- With Market Clustering, an additional level of flexibility is available. The flexibility of Market Clustering allows attributes to be specific to regions of passenger types. This improves forecast accuracy

# Market Clustering – Study

- Objective:
  - Group O&D Markets into sub-groups by Seasonality and Difference in Time Zones to improve calibration of market share
- Data Sources:
  - Seasonality Data
  - Time zone differences between airports in an O&D
  - Annual Market Size

# Clustering Methodology

- Algorithm: Uses PROC FASTCLUS (a SAS procedure) to perform disjoint cluster analysis. This ensures every O&D route belongs to exactly one group.
- Distance Metric: Clusters are determined using Euclidean distances calculated from quantitative variables, specifically seasonality factors.
- Pre-Segmentation: Clustering isn't done on the whole dataset at once; it is performed within specific time zone difference **levels**.

# Definitions and Metrics

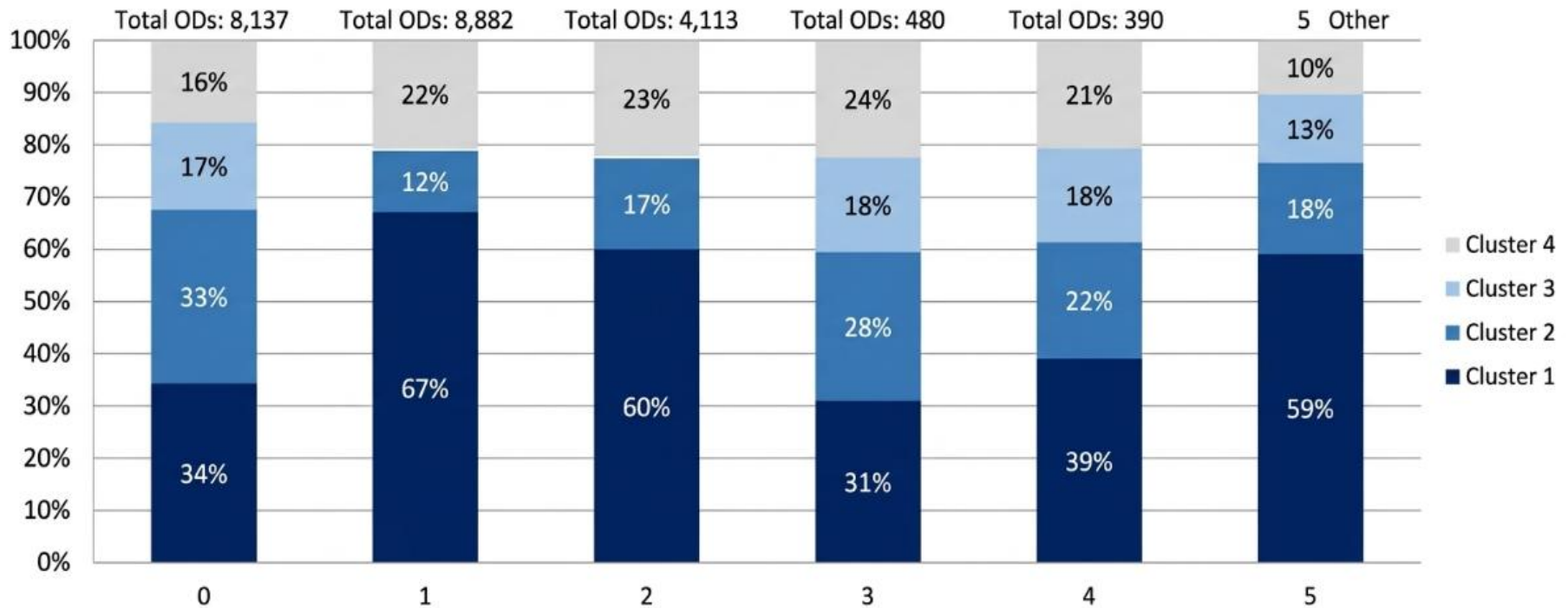
- Time Zone (TZ) Difference: Calculated as the absolute difference in minutes from Greenwich Mean Time (GMT) between the two airports.
  - *Example:* The difference between CST (-360) and EST (-300) is 60 minutes (1 hour).
- Visualization: Results are presented using passenger-weighted mean seasonality scores.
- Quality Control: The stability and reliability of the clusters are evaluated using weighted standard deviations.

# Sample Selection

- Domestic – Domestic O&Ds
  - 93,450 domestic - domestic market/point of origin(poo) combinations in the domestic seasonality factor file
  - Sample will consist of the 24,454 (26%) markets/poo with  $\geq 1$  passenger per day each way (PDEW)

# Cluster Summary for D-D Markets

## Cluster Contribution by TZ Difference (ODs/Poo %)



# Cluster Analysis

## Domestic - Domestic

TZ Difference	Cluster	Market Interpretation
0	1	Relatively flat, spring peak
	2	Inclining to October peak
	3	Extremely strong summer peak
1	1	Relatively flat, summer and October peak
	2	Strong December and winter peak
2	1	Strong summer peaks
	2	Strong December and winter peak
3	1	Extremely strong summer peak
	2	Flat
	3	Conical with August peak
4	1	Relatively flat, summer peak
	2	Extremely strong summer peak; trough in fall and winter
	3	Strong winter, low summer
5	1	Summer peaks
	2	Conical with June peak
	3	Strong winter with March peak, low summer

# Cluster Reliability and Stability Metrics

- Passenger - weighted seasonality factor

$$\bar{x}_w = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}$$

$w_i$  = market size (weight)

$x_i$  = monthly seasonality score

- Passenger - weighted standard deviations

$$sd_w = \sqrt{\frac{\sum_{i=1}^N w_i (x_i - \bar{x}_w)^2}{(N' - 1) \frac{\sum_{i=1}^N w_i}{N}}}$$

$N$  = # of ODs in the cluster

$N'$  = # of non zero weights

# Results

