FAVORITA: TSSI – A Time Series Sales Insight App

Enabling Time Series Clustering

Ting Yen-Hsin | He Qiyun | Caine Ng Jianyuan | Guided By: Prof Kam Tin Seong

School of Computing and Information Systems, SMU Singapore Management University

Issues and Problems

Data for transactions at convenience stores or eCommerce websites are freely available online for students and analysts to practice their analysis skills to derive insights. A common approach would be to use unsupervised learning techniques such as k-means clustering or hierarchical clustering to group data together before further analyzing the groups for the unique traits. However, possibility due to complexity or requirement for specialized enterprise software clustering of time series data is seldom done.

Yet time series data is one of the most common data forms available represented in sales transactions. With unique trends from different buying patterns, it would be a waste to not leverage it for purpose of analysis.

Motivation

The availability of R package dtwclust has allowed access to perform time series clustering without significant cost. Coupled with wrapping the package in a R-shiny user interface would allow common users to perform sophisticated time-series clustering without writing code.

In this project we wish to demonstrate how a customized R Shiny application would allow users without deep technical capabilities to perform analysis on their time series data and enable grouping of their time-series data. With time-series data among being among most commonly derived and simple data forms, the application could give users another dimension to group their data.

Approach

Exploratory

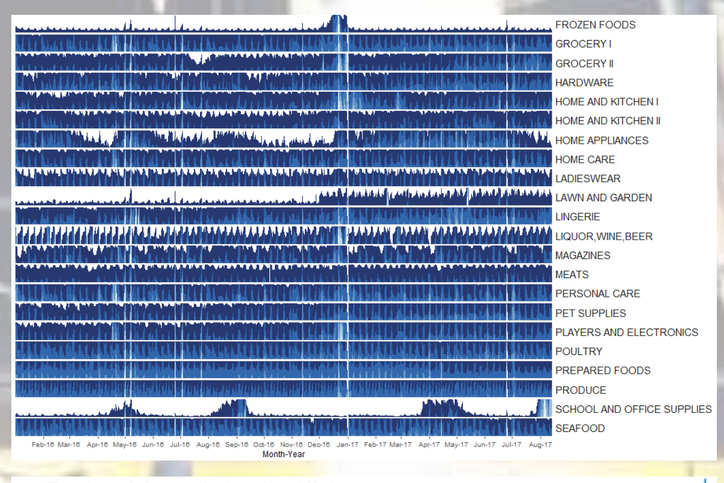
Time-Series Clustering

Results Review

Time Series Trend-Analysis

We hope to enable users to start from an overview of the data to identify trends through exploratory means. From identified trends allow them to select portions of data for time-series clustering and give options to adjust their model. A results dashboard would readily give an overview of generated cluster characteristics. Finally allowing the users to further inspect the data by overlaying specific time-series data.

Results



Exploratory

This portion of the application aims to give users an overview of the data to enable comparison of different time-series and quick identification of possible trends.

A horizon plot presents a compact view of several time series patterns to for comparison and pinpoint seasonality and similarities between different time-series data such as at different stores or product categories.

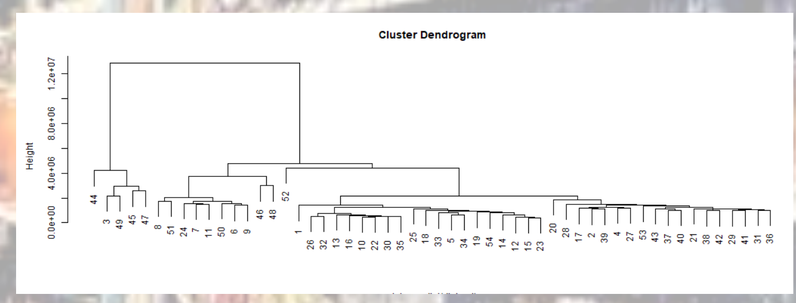
A heatmap provides an aggregate view of sales at different stores and cities on different days of the week and months.

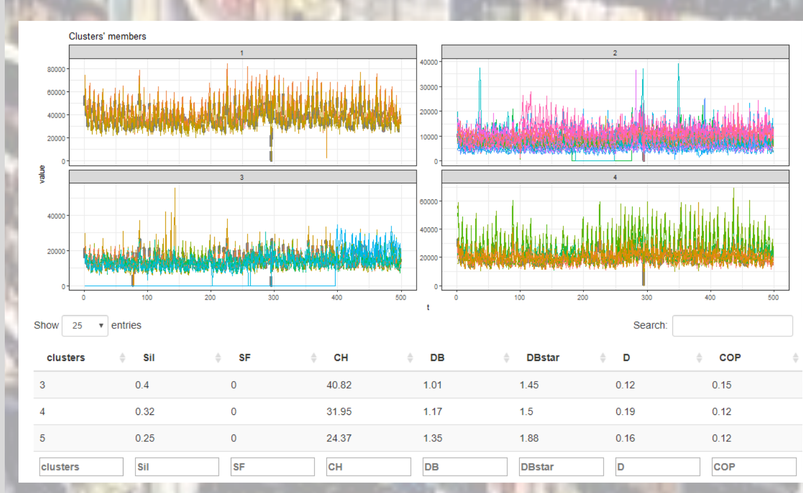
Time-Series Clustering

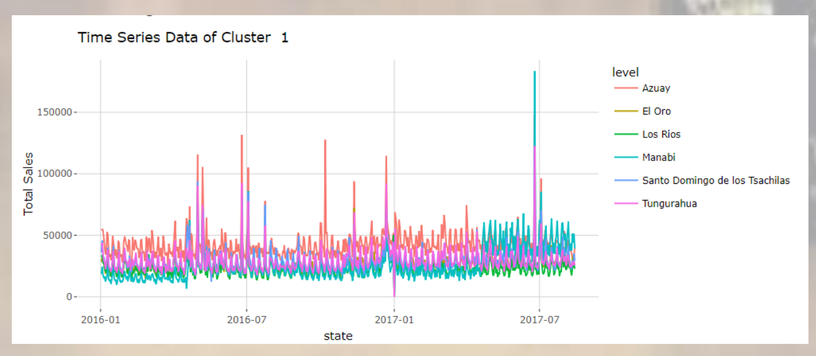
From observed trends of earlier overview, the user would be able select corresponding data to perform time-series grouping using a graphic interface to adjust different attributes of dtwclust package.

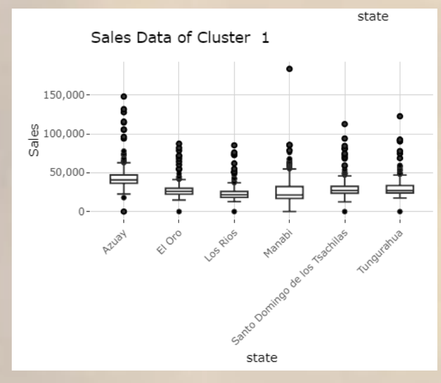
Results are shown with representative visualizations such as dendrogram or facet plot of time-series for hierarchical or partitional clustering, respectively.

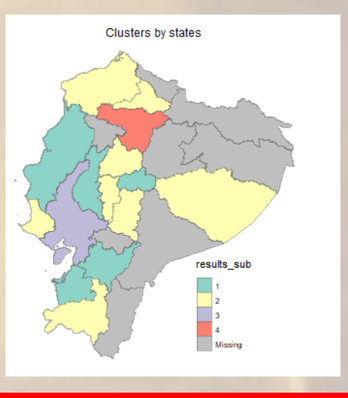
Corresponding cluster validity indexes (CVI) is also display for users to adjust the model.











Results Review

Results from time series clustering are merge with original data and presented in visualizations to allow users to identify and deep dive into reasons for the clusters.

An interactive time-series plot allow users to inspect the results. Box plot of sales for each member of the cluster is shown and for geographic attributes, locations of the grouping is shown on a map.

Future Work

The current app does allow users to group selected time-series data, these are some possible future work to improve the application:

- Option to load data of longer date ranges.
- Additional options to generalize to segregate data before clustering.
- Expand allowable options to further fine tune time-series clustering models.
- Allow forecasting after grouping of time-series data.