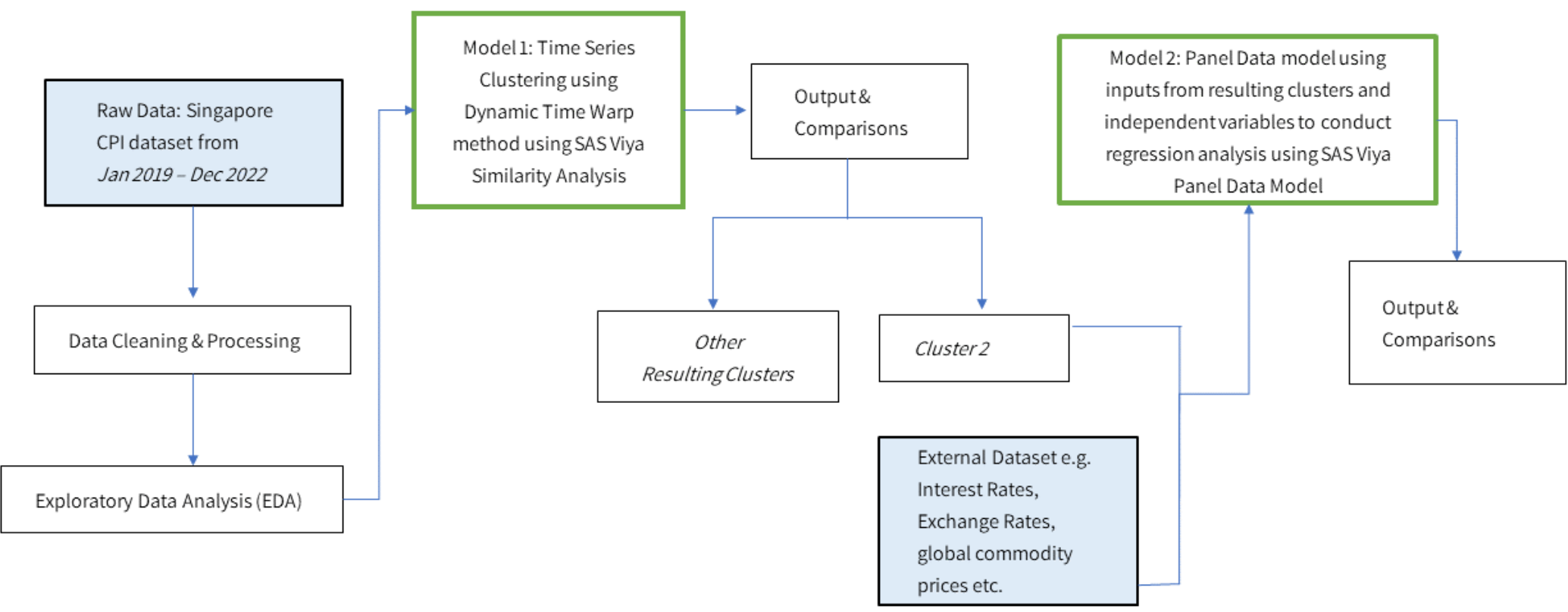


## Motivation

**Motivation** of this project stems from the undiscovered pricing patterns of the items within the CPI that are being concealed by virtue of using the CPI as the main metric to discuss consumer price inflation. Using statistical data analysis techniques, we aim to explore the patterns of the price movements at the items level, conduct analysis on similar time series as well as relationship of key financial service metrics on the items in the CPI.

**Objectives** of this project are: (1) Extract clusters through time series clustering using Dynamic Time Warp (DTW) method. (2) Derive possible reasons for similarity in price movements in conjunction with global events to identify possible trigger points in time. (3) Detect the key financial services metrics that contribute to the CPI using regression analysis.

## Research Project Flow



**The research design of this project is shown in this flow chart and described below:**

- (1) Raw dataset on Singapore CPI (Blue) is cleaned and processed, followed by EDA
- (2) Prepared dataset is brought into SAS Viya for Time Series Clustering
- (3) Obtain results and conduct comparisons from the resulting clusters
- (4) Using Cluster 2 as a sample, data is combined with external dataset (Blue) to run Panel Data model to find the variables which contribute to the changes in the CPI.
- (5) Conclude insights & analysis from the regression model.

## Data & Data Preparation

### Data Sets

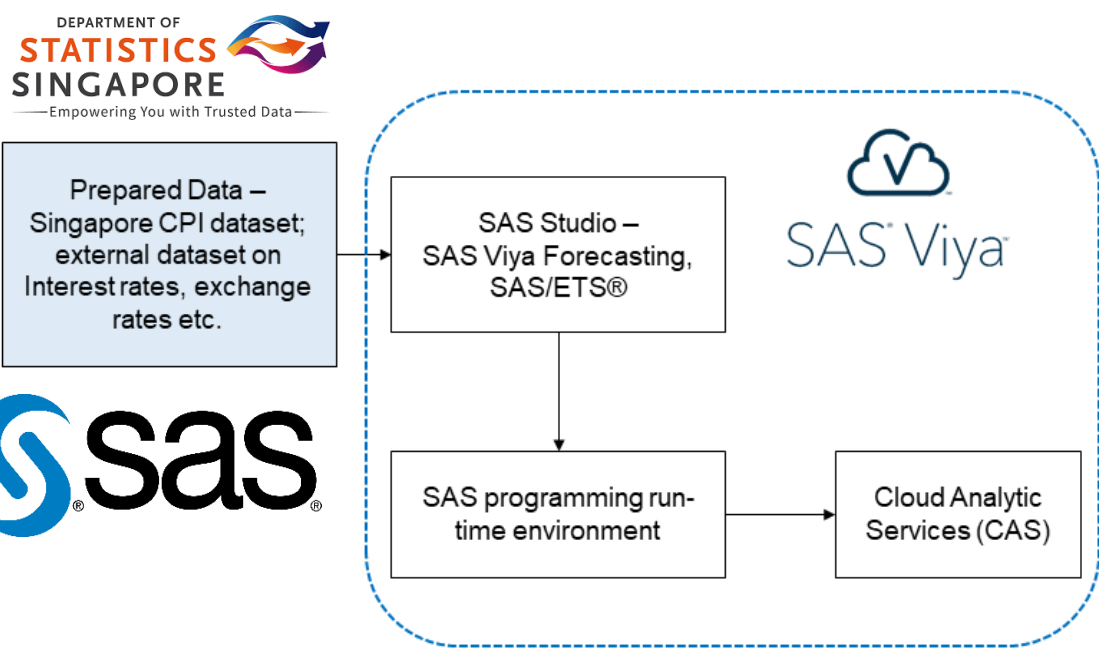
Singapore CPI data is available on Singstat.

In this project the sources of data sets are:

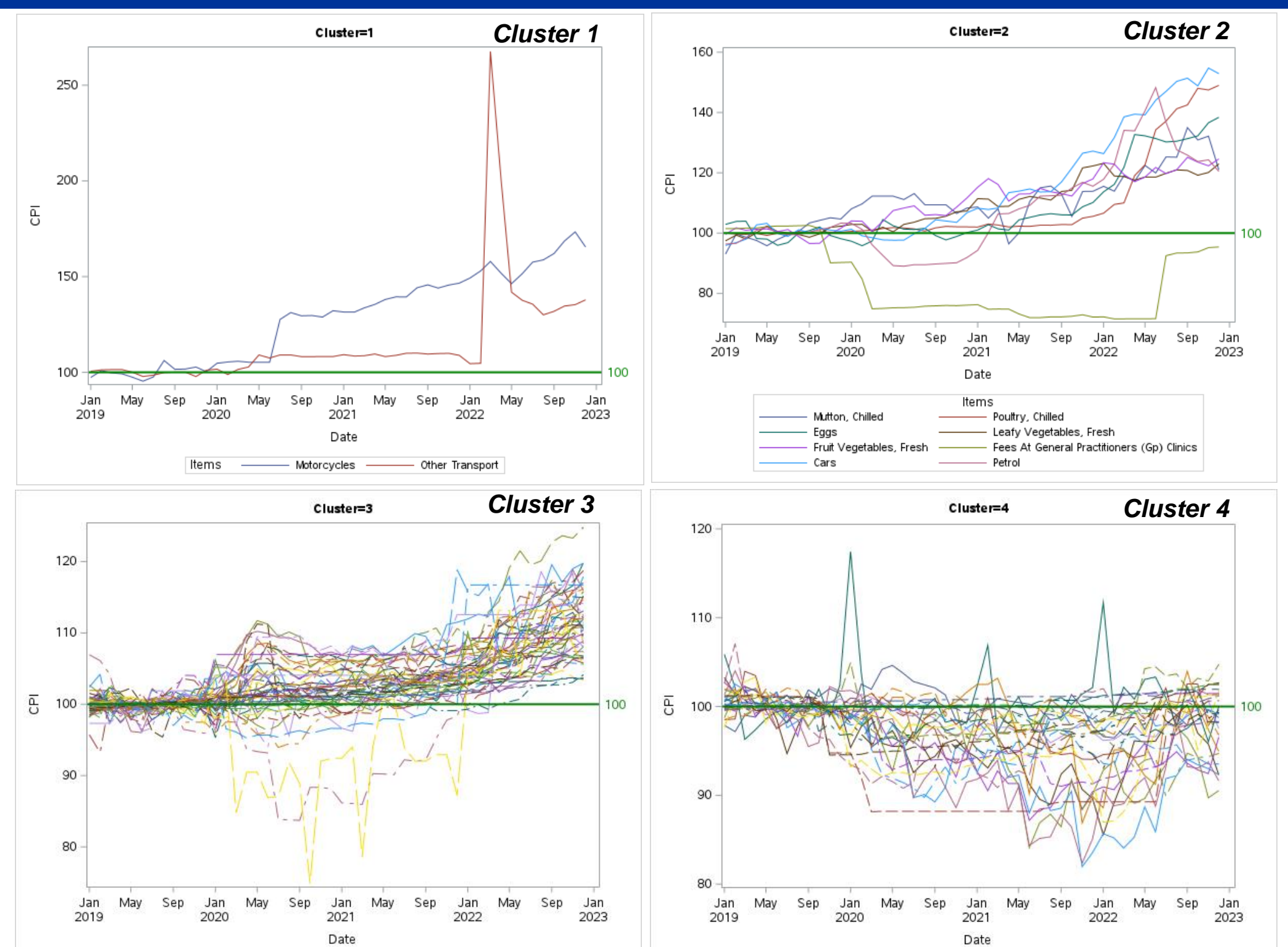
- CEIC – for CPI, commodity prices (Gold price and Oil price), volume of import & export in Singapore
- MAS website – for exchange rates (USD/SGD, SGD/RMB), average SORA

**Analytical Software used – SAS®**

SAS Viya is an analytical software that allows faster model building and processing with access to data through cloud. In this project, SAS Studio was accessed through a browser. Aside from the data visualization tasks, two main SAS tasks are used in this project – (1) Similarity Analysis under SAS Viya Forecasting Software, Time Series Mining and (2) Panel Data Models under Econometrics or SAS/ETS® module.

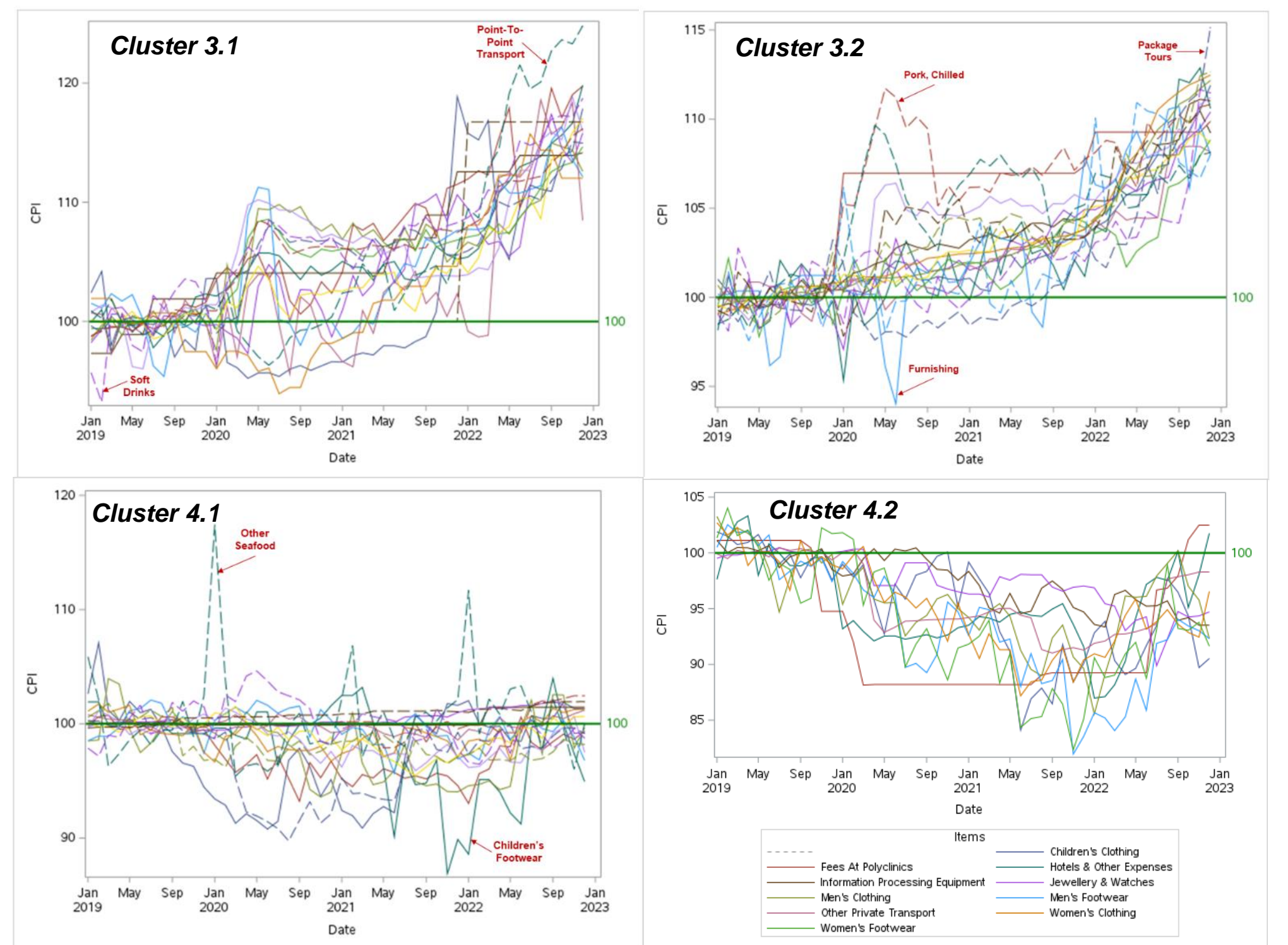


## Clustering Results I

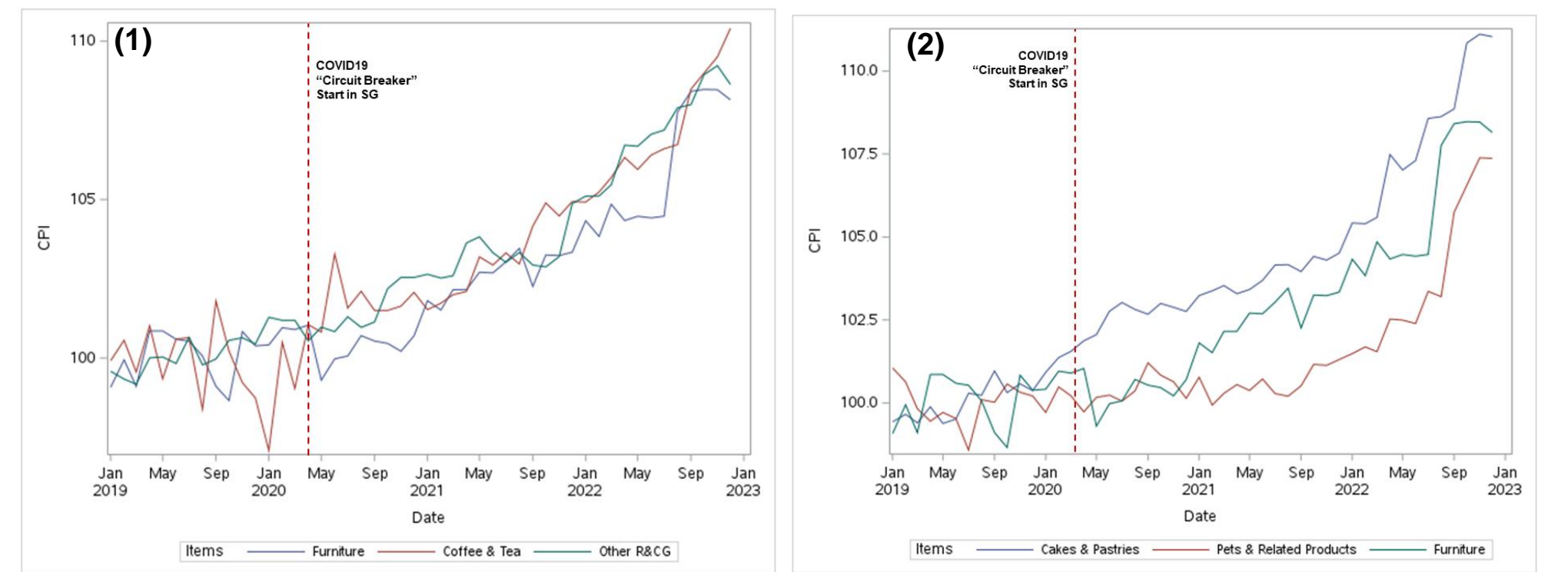


Initial findings show that items across divisions were clustered together indicating cross divisional similarities. This would otherwise be masked within the CPI. Based on the dendrogram, 0.1 Semi-Partial R-Squared was used as the cut off for the number of clusters. The result is that we have 4 clusters (shown above), from 2 items to 50 items. Cluster 3 and 4, were further broken down into sub-clusters. Cluster 3 formed 4 sub-clusters and Cluster 4 formed 2 sub-clusters. The result was visibly clearer in terms of the sub-cluster patterns.

## Clustering Results II



Based on the results, we can conclude that Cluster 1 & 2 (Total 10 items) experienced very sharp increases in prices whereas a greater proportion of items saw moderate price increase (Cluster 3 – 50 items) while the rest remained the same or experience some declines too (Cluster 4 – 31 Items). Both Cluster 3 and Cluster 4 exhibited differing patterns across sub-clusters and fluctuations fell within different ranges.



**Individual items within the clusters highlights two main types of similarities:**

- (1) time series were clustered together because of **similarity in time**,
- (2) time series with **similarity in shape**, with time lag

The time series clustering results reveals the usefulness of the Dynamic Time Warp (DTW) method in time series data mining. The results can be seen especially from sub-clusters 3.1, 3.2, and 4.1. This finding provides strong justification for using DTW method for time series data mining.

## Panel Data Model

Using Cluster 2 as a sample for running the panel data regression model, we run the model using the Panel Models Data Task in SAS/ETS® module. We run the analysis using Linear, Fixed effects, One-way model. In a Fixed effects model, the relationship between unobserved variables and the independent variables are “fixed” into a constant over time, controlling for heterogeneity.

Variable	Estimate	Standard Error	t Value	Pr >  t
Intercept	-54.7036	41.5411	-1.32	0.1887
SGDperUSD	47.00489	19.6069	2.40	<b>0.0170</b>
RMBperSGD	1.287203	1.4170	0.91	0.3642
Average SORA	7.04645	1.1500	6.13	<b>&lt;.0001</b>
Crude Oil Price	0.10796	0.0574	1.88	0.0606
Gold Price	-0.06674	0.2843	-0.23	0.8146
Gold Price Index	1.133693	3.5428	0.32	0.7491
Imports SG	0.000198	0.000473	0.42	0.6755
Exports SG	0.000206	0.000448	0.46	0.6467

The R-square of the model is 0.7131, which indicates goodness of fit. The regression analysis result shows that p-values of **SGDperUSD** (exchange rate pair of SG\$ and US\$) and **Average SORA** are statistically significant to reject the null hypothesis. There is sufficient evidence to conclude that there is a correlation between SGDperUSD and Average SORA on the CPI.

## Conclusion & Future work

**The devil is in the details** –The CPI is a commonly used metric to measure consumer price inflation, but there is a lack of research and awareness on the items *within* the CPI. This project shines the light on the items and their price movements over time.

**Dynamic Time Warp method on Time Series Clustering** –. From the time series clustering analysis, we confirm that dynamic time warp method is useful with its strength to cluster similarity across two time series both in *time* and in *shape*.

**CPI as independent variable** – More research is needed on how the CPI affects other socio-economic metrics like poverty levels, median incomes, GDP per capita etc. Future works can focus on the reverse effects since this project examines on the factors contributing to the CPI.

**Fixed effects** – Further research needs to be conducted on panel data regression modelling techniques. Depending on the variables, special methods could be used to control and test model.