1. About Page

On this page, a short description of the functions of each tab in the application is shown. Links to this user guide and to the development page (Project site) are also provided.
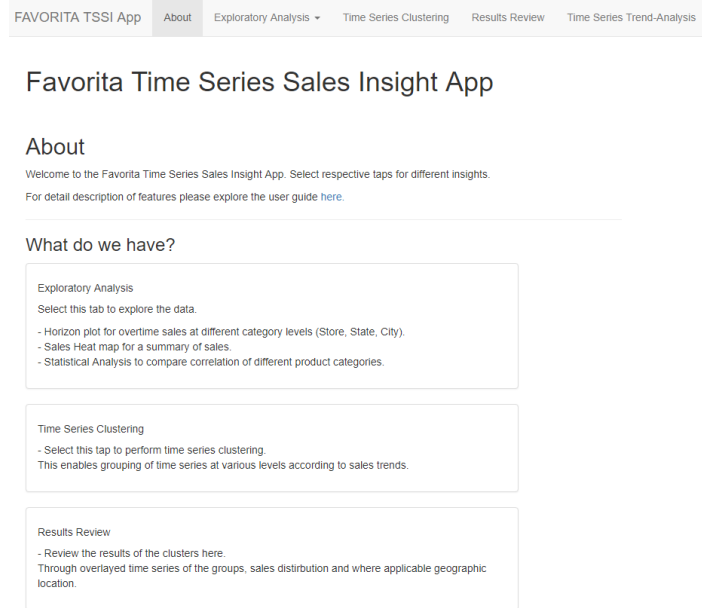
2. Exploratory Analysis

Functions under this tab provide ways to explore an overview of the data, identify time-series trends and correlation between products. There are three ways to explore the data: horizon plot, regional sales heat map and statistical analysis.

2.1 Horizon Plot
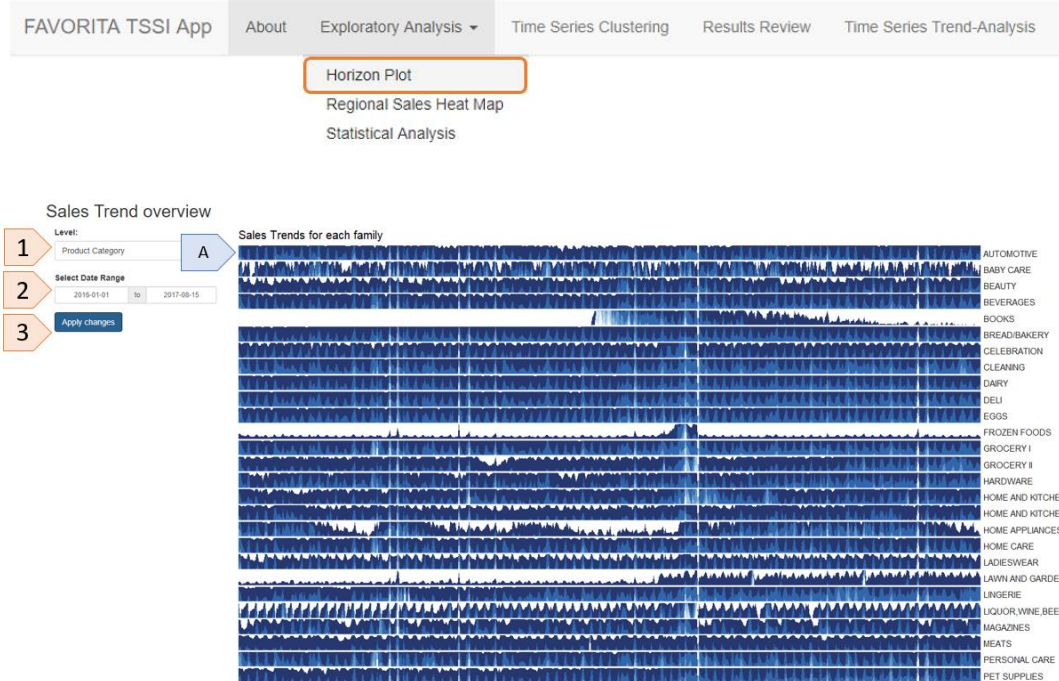
Graphics

[A] The horizon plot gives a quick overview to compare time series trends of different categorical variables. Lighter shades represent peaks in the horizontal shaded areas and darker colors represent relatively lower values.

Options

[1] Select the categorical variable for comparison. Currently data can be compared at State, City, Store Number, Store Type, or product category level.

[2] Select the time frame for comparison. Choose a time frame according to expected seasonality trends. Certain product categories have quarterly trends, and a period of 1-2 years could be selected for comparison. When Store level sales trends could be viewed according to longer periods to observe their overtime sales. (Online version of application is limited to data 1 year data in 2016).

[3] Click on Apply changes to update the graph after selection.

2.2 Regional Sales Heatmap
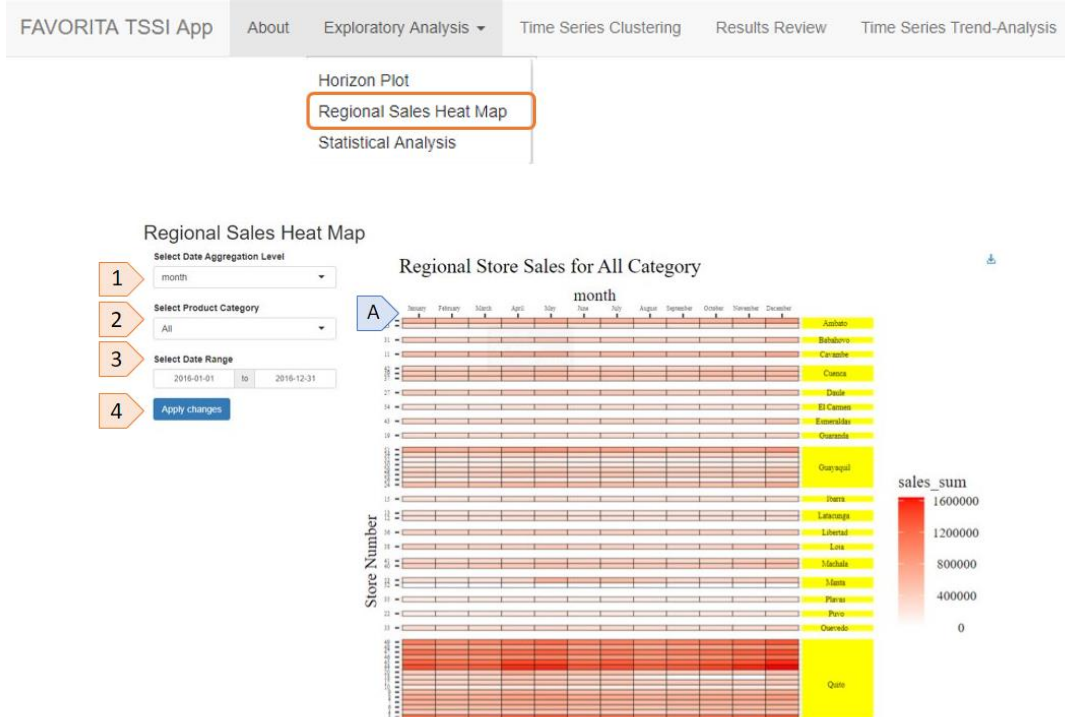
Graphics

[A] The regional sales heat map provides a summary of sales at the store and city level. Users could choose to view the sum of sales at a monthly or day of the week level for all products or a certain product category. Giving users another view of the time series data. Hovering over a cell would show details of the store, city, and aggregated sales.
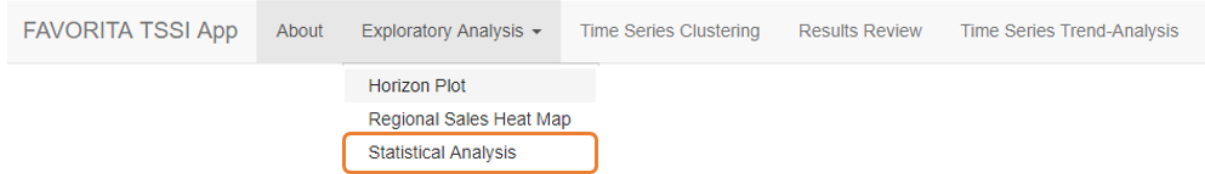
Options

[1] Select the Date aggregation level. Sales could be grouped by their monthly or day of the week sales.

[2] Product category to view sales. Users could choose either all products or a specific category.

[3] Select date range to summarize sales.

[4] Click on Apply changes after settings are made.

2.3 Statistical Analysis – Scatter Plot and Correlation Matrix

This tab allows the user to compare correlations between different product categories in terms of sales and perform a statistical test to determine strength of correlations.
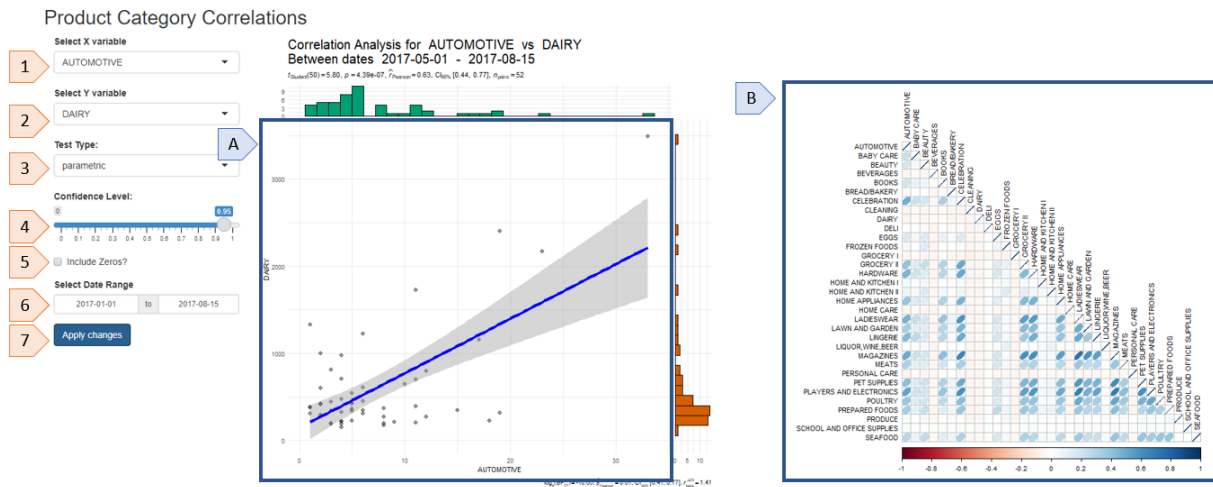
Graphics:

[A] A scatter plot is used to visually show correlations between two product categories in terms of sales. Marginal frequency distribution plot is used to indicate whether categories chosen are normally distributed or not before choosing the type of statistical test for verification.

[B] A correlation matrix is provided as a guide to provide suggestions which pairs of categories are highly correlated.

Options:

[1] Select product category to be placed on the X-Axis.

[2] Select product category to be placed on the Y-Axis.

[3] Select type of statistical test method to assess correlation between the two product categories. Options currently available are parametric, non-parametric, robust and bayes.

[4] Select the confidence level required for the test.

[5] Optional check box, to remove zero values from the original data or not, as purchases may not always occur on the same data.
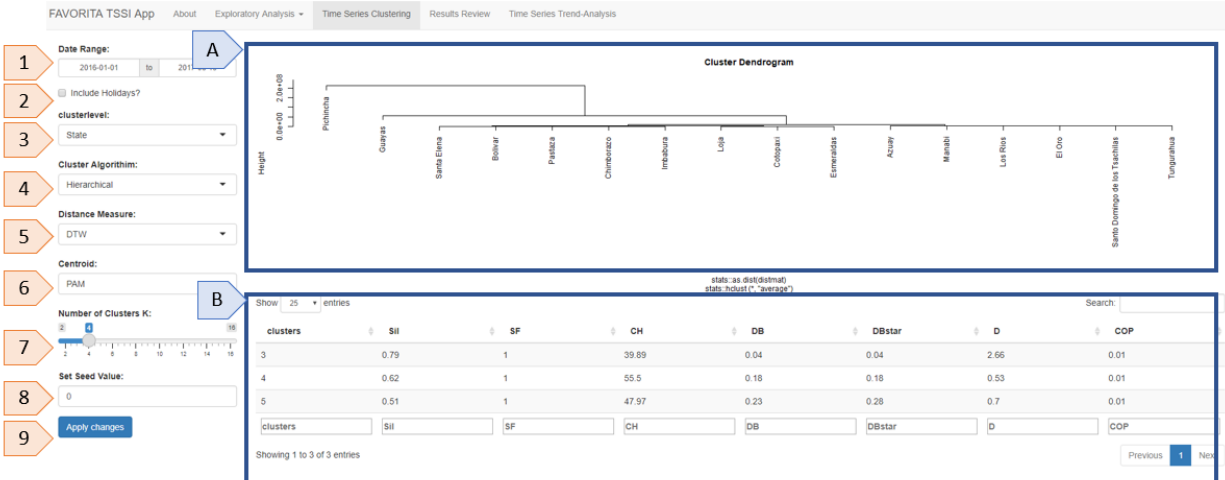
[6] Select date range of data to test from. Note that for online version of application date range greater than 1-month may cause the application to crash.

[7] Click on Apply changes after settings is made.

3. Time Series Clustering

This tab allows users to perform clustering of the time series data based on earlier derived observations from exploratory analysis.

Graphics:

[A] Provides a visual representation of the results from clustering the time series. Depending on the cluster method chosen different graphs will be shown. For partitional clusters, each group is shown in a separate facet plot with their time-series data together. For Hierarchical clustering, a dendrogram is shown instead.
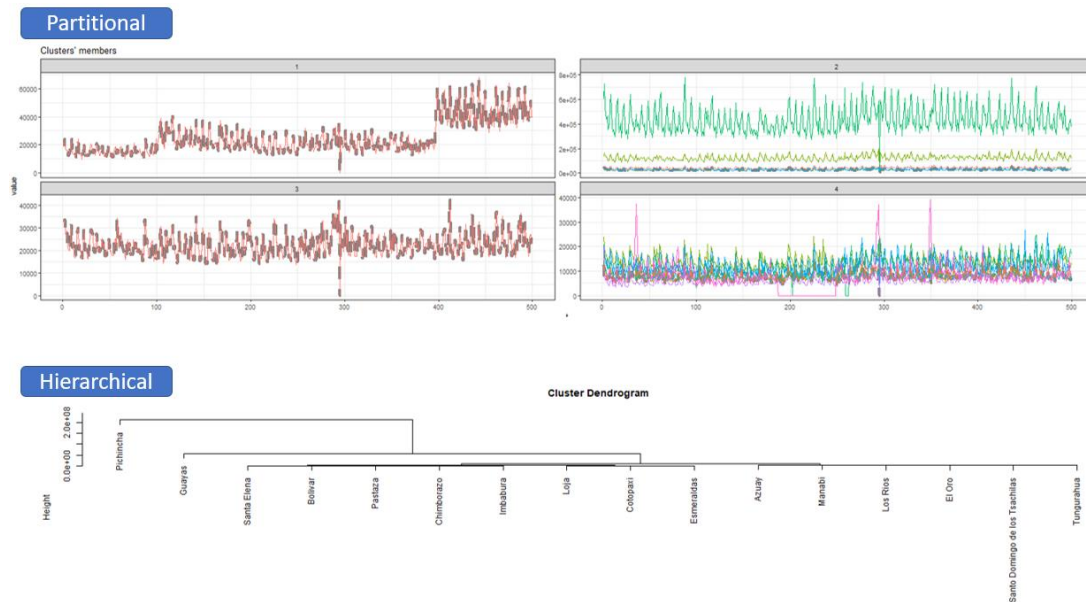
[B] A table summarizing the various cluster validation indicators is shown here for varying number for clusters from the option chosen by the user. As clustering is an unsupervised technique, these indicators

serve as gauge how well the cluster members conform to the group. Users could then adjust the model and rerun the analysis.

Options:
[1] Select date range of data to be used for time-series clustering.

[2] Select this option to include data from days where sales occur on holidays. Holidays contribute to a spike in sales for certain days and products. Uncheck to remove these data points before clustering.

[3] Select the categorical level to apply time-series clustering. Currently State, City, Store, Store type and product category data could be grouped according to their time-series sales data.

[4] Select the clustering algorithm to be applied. Currently partitional and hierarchical options are available.

[5] Select the Distance measure used to separate the time series data. Currently 'DTW', 'SBD' and 'GAK' options are available.

[6] Select the Centroid calculation method in this option. Currently 'PAM' and 'DBA' are available.

Note that this option is not applicable to hierarchical clustering and is not applied when clustering via hierarchical method.

[7] Select the number of clusters for the group. A larger number of groups may allow the data to be better classified, however too many groups would invalidate the need for clustering. Note not to select more clusters than the number of categorical levels in the data.

[8] Enter a random number to set seed. This is to allow consistent results to be reproduced if running the analysis multiple times.

[9] Click on Apply changes after settings are made.

4. **Results Review**

This tab gives users an overview of the results from time-series clustering generated in Time Series Clustering Tab. Providing additional views of the data for users to spot commonalities in the attributes that have been grouped together by time-series clustering.
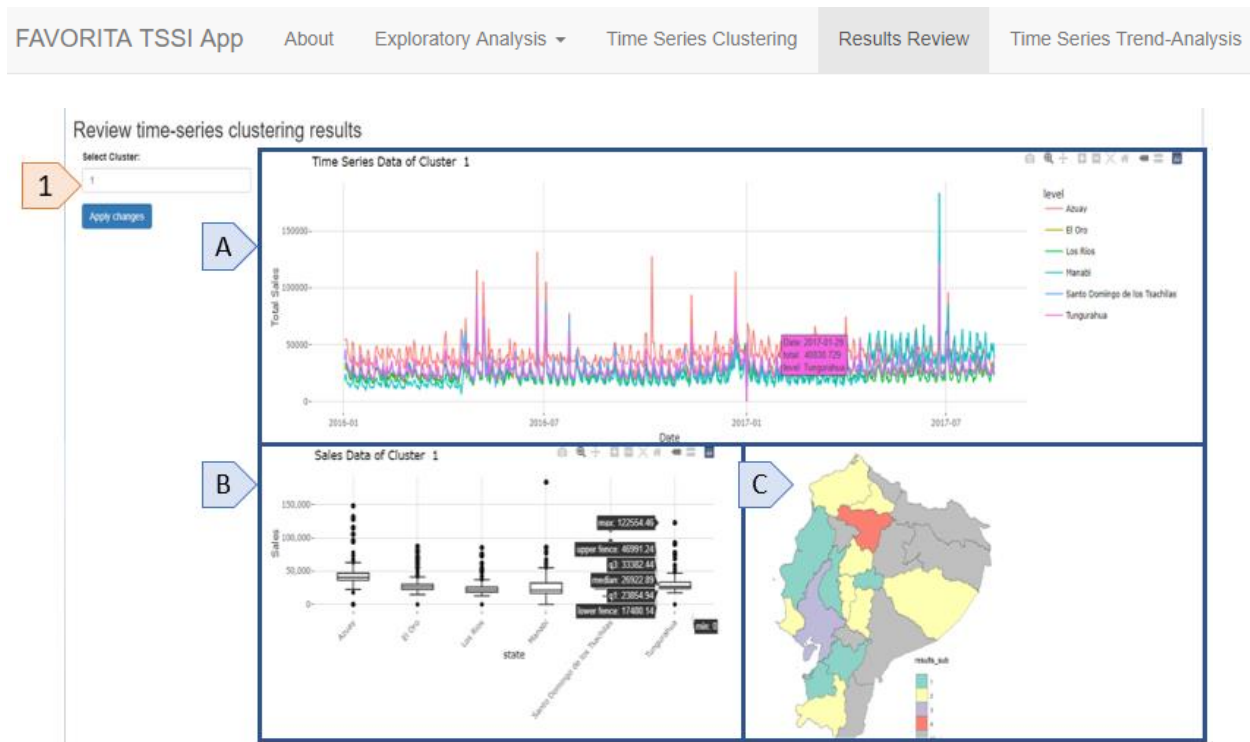

Graphics

[A] Line chart of data grouped together in time-series clustering presented in a larger view. Hovering over lines in the chart will bring up tool tips to show details of the sales, date, and categorical data label. Note that by clicking on each the legend labels, the series could be removed from the view. This does not affect other graphs.

[B] Box plot indicating summary of sales for each of the categorical labels. Hovering the mouse at the side of the median line within the box plot will bring up summary values that describe the box plot.

[C] This view summarizes the grouping categorical labels. For geographic labels such as state and city a map is shown to indicate the location of each group.

For other labels such as product category, store type and store number, a tree map is displayed to indicate the composition of each group. Size and shade of the cells represent the number of sales with more sales indicated by larger cells and lighter color.
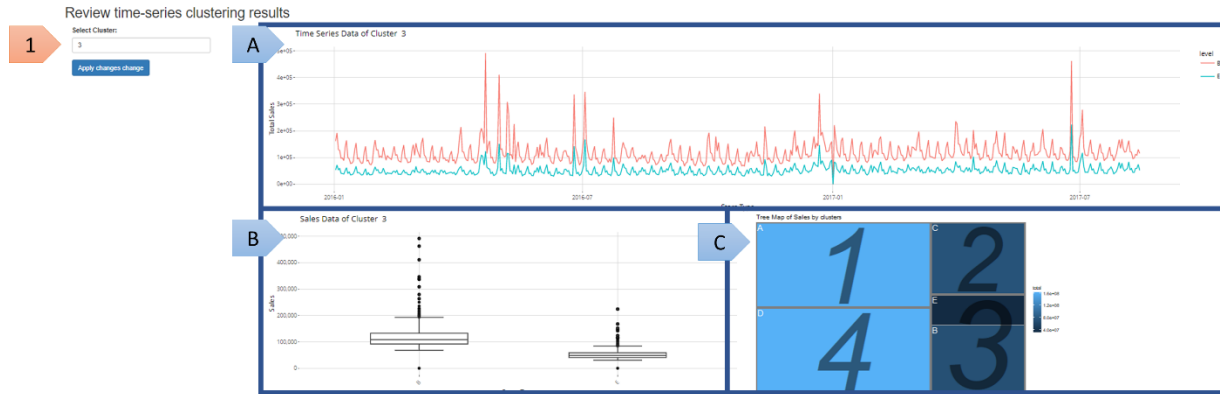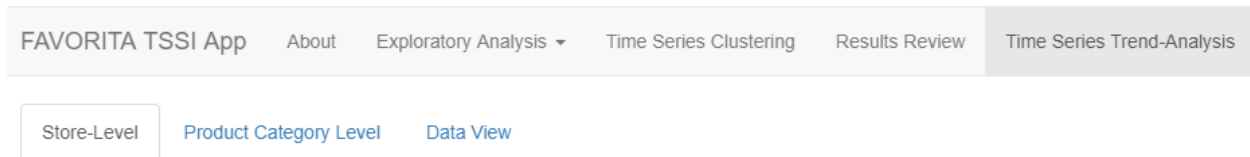
Options:

[1] Toggle from this option the different clusters to view. Note that data displayed are results from time-series clustering tab. To change to a different set of data or category label, the model would have to change in Time-series clustering tab.

[2] Click on Apply Changes to update table.

5. **Time Series Trend Analysis**

This tab allows users to revisit the time-series data and overlay different categorical levels to investigate the trends leading to grouping of between series. Line plots and corresponding day of the week cyclic plot are presented. Shown in separate tabs users could investigate the Store-level similarities and product category similarities.

5.1 Store – Level

In this tab users could choose data from which store to overlay for comparison.
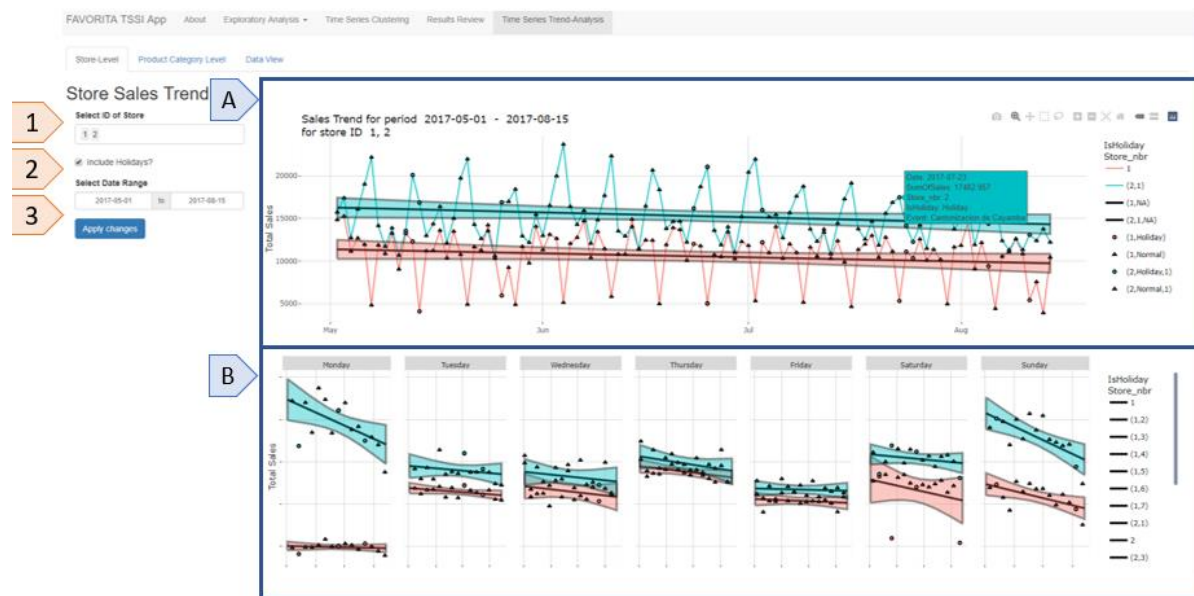
Graphics:
[A] Overlay line charts from chosen data. Hovering over the points will bring up a tool tips giving details of the data such as Date, total amount of sales, store number, whether data point corresponds to a holiday and which holiday it is.

[B] Cyclic day of the week plot to show sales trends on days of the week.

Options
[1] Enter the store number for charting. Clicking in the window will bring up a drop down of suggestions for selection. Note that multiple different stores could be selected.

[2] Uncheck this option to include data from holidays or uncheck to remove them as holidays may contribute to a single spike in sales.

[3] Select the date range of data. Note that too long a time would result in a cluttered graph. This view is better for investigation of shorter time periods.

[4] Click on Apply Changes to update table.

5.2 Product Category Level

Similar to the store level tab this page allows users to overlay product category sales data.

Graphics
[A] Overlay line charts from chosen data. Hovering over the points will bring up tool tips giving details of the data such as Date, total amount of sales, the product category, whether data point corresponds to a holiday and which holiday it is.

[B] Cyclic day of the week plot to show sales trends on specific days of the week.

Options
[1] Enter the product category for charting. Clicking in the window will bring up a drop down of suggestions for selection. Note that the selected product category can be multiple.

[2] Uncheck this option to include data from holidays or uncheck to remove them as holidays may contribute to a single spike in sales.

[3] Select the date range of data. Note that too long a time would result in a cluttered graph. This view is better for investigation of shorter time periods.

[4] Click on Apply Changes to update the graph.

5.3 Data View

This tab assists the user as a reference to the data for selecting corresponding stores related to a specific city or state.

Graphics:

[A] At the top half of the page, this tables gives a summary of the data and number of entries in the selected date period. Users may also use this table to see which state or city a store belongs to.

[B] At the bottom half detailed row entry of each data point for reference.

Date	store_nbr	DayOfWeek	WeekOfYear	Month	city	state	StoreType	cluster	family	sales	onpromotion	IsHoliday	type of holidays_e	description
2016-01-02	1	7	1	1	Quito	Pichincha	D	13	MAGAZINES	1.000	0		Normal	
2016-01-02	1	7	1	1	Quito	Pichincha	D	13	MEATS	222.967	0		Normal	

3

Showing 1 to 25 of 666,468 entries

Options:

[1] Select the date range of data for reference. Click on Apply Change to update the table.

[2] & [3] Enter the values for different columns to check, for example, store number, city, state etc.