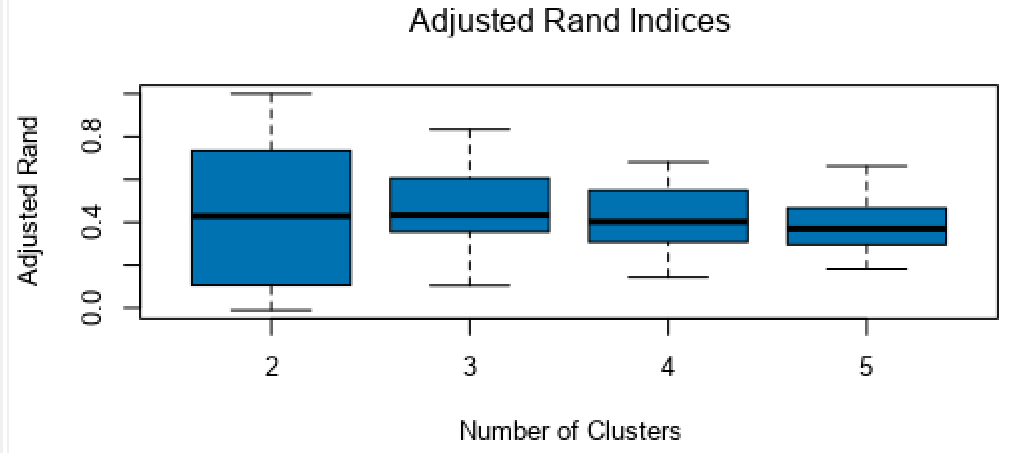
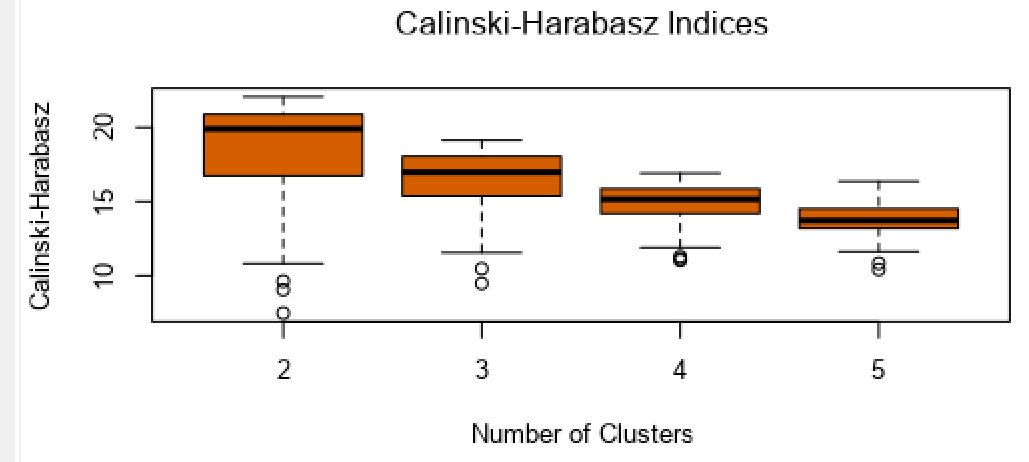
Project: Predictive Analytics Capstone

## Task 1: Determine Store Formats for Existing Stores

1. What is the optimal number of store formats? How did you arrive at that number?

By performing cluster analysis methods using k-centroids diagnostic tool with number of clusters ranging from 2-5. We were able to see the below reports.

**K-Means Cluster Assessment Report**

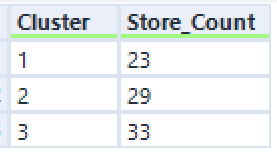
 

Based on the K-means report’s Adjusted Rand and Calinski-Harabasz indices, number of

clusters 3 has the highest median and considerably less variation. Hence the

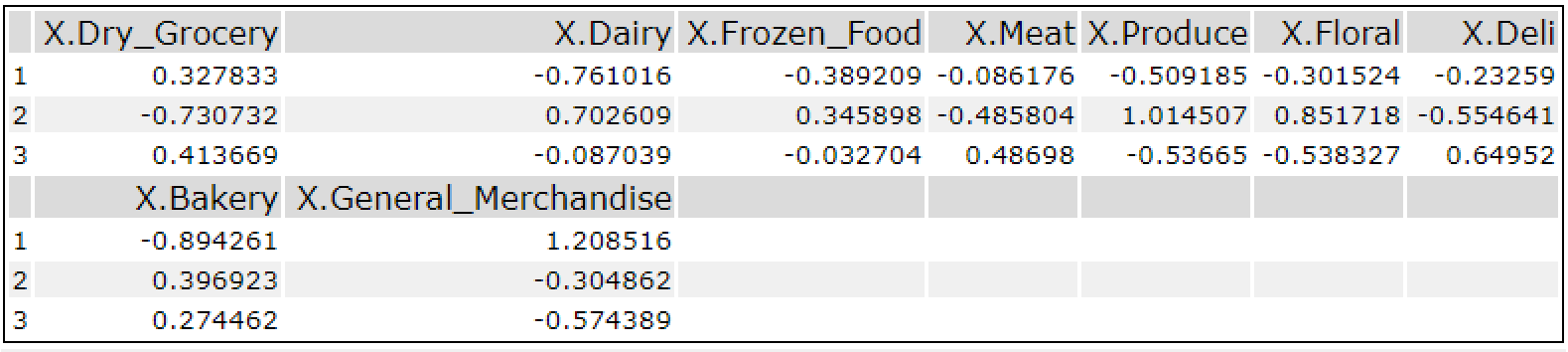
optimal number of store segments is 3.

1. How many stores fall into each store format?



1. Based on the results of the clustering model, what is one way that the clusters differ from one another?

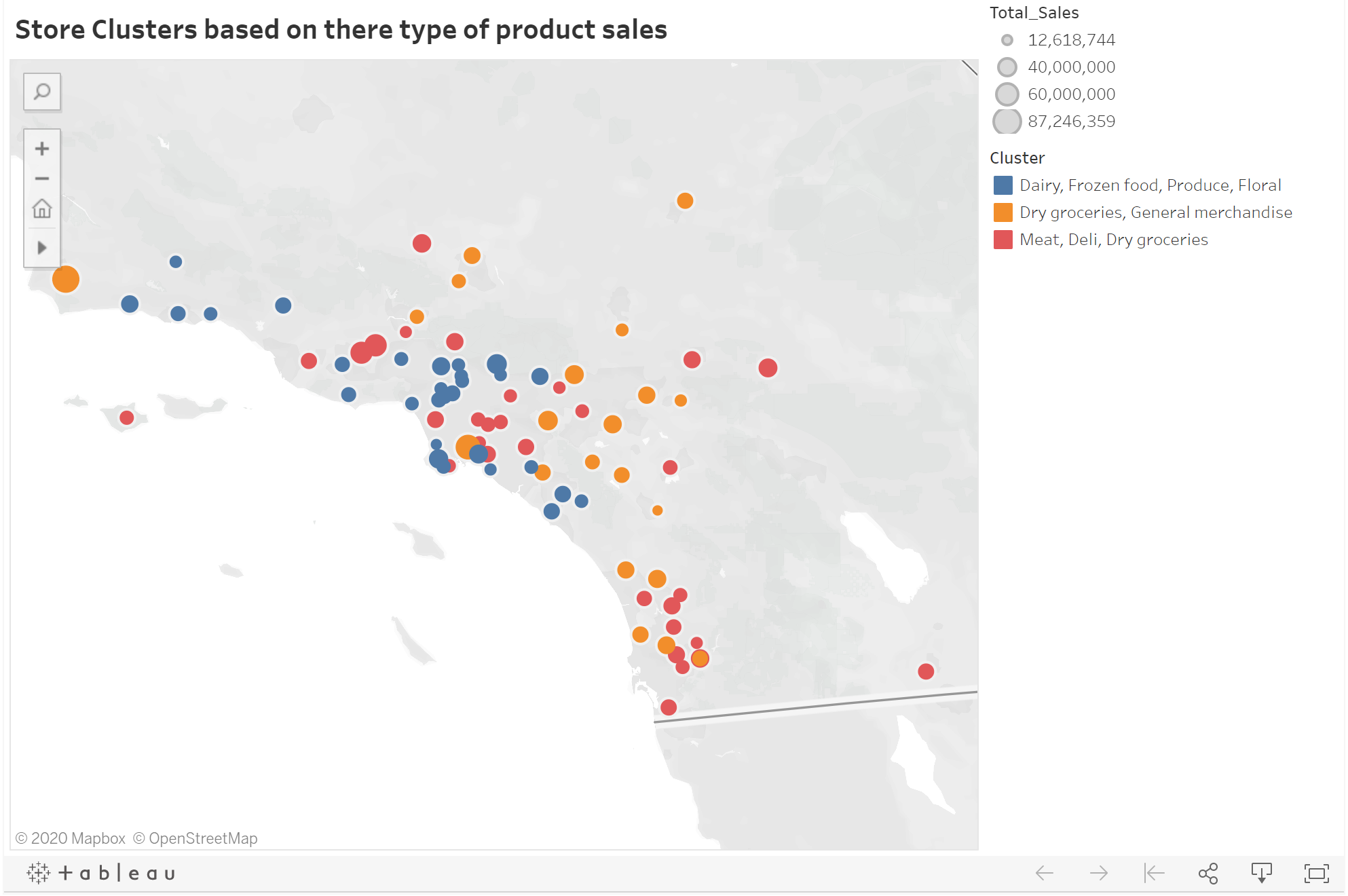
Based on the clusters report



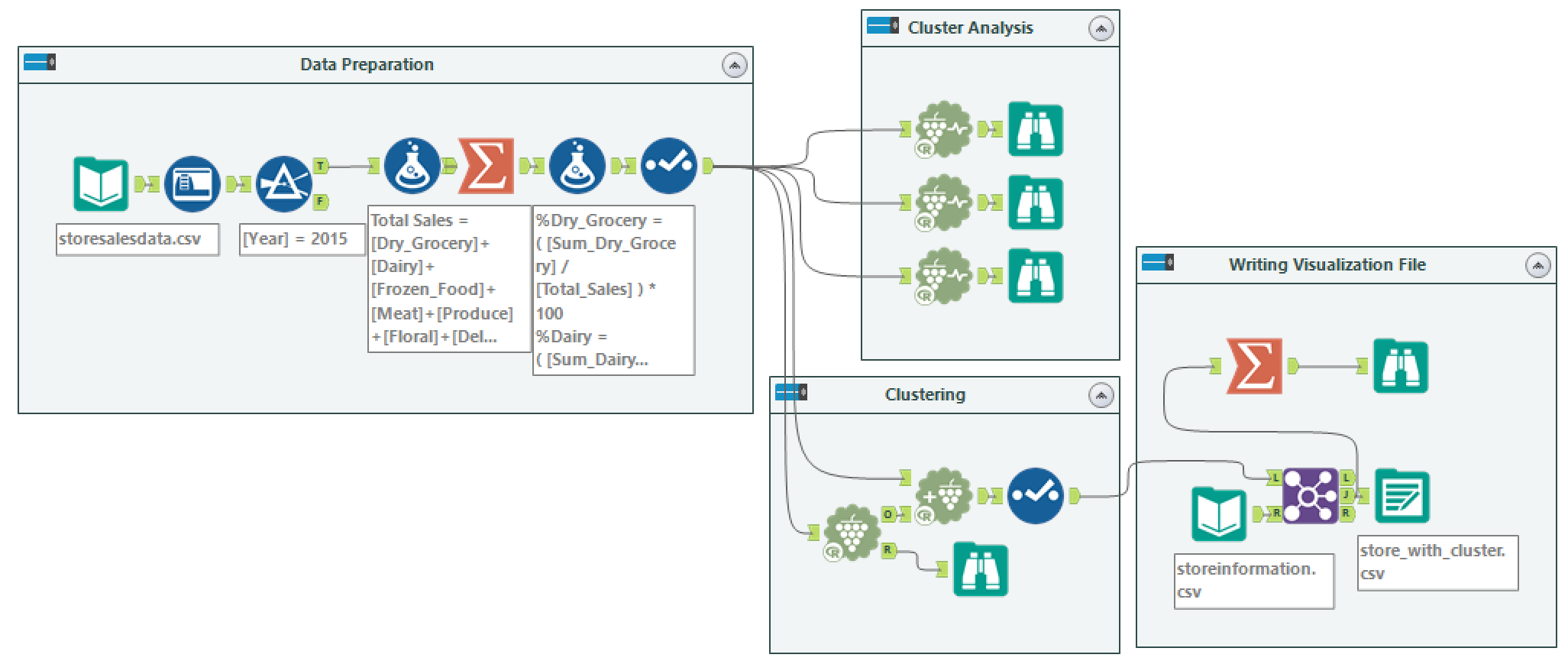
It is clear that,

* **Segment-1** mostly sold more dry groceries and general merchandise which has longer or no expiry period.
* **Segment-2** mostly sold more Dairy, Frozen food, Produce(farm-produced crops) and Floral which are FMCG(Fast Moving Consumable Food) products with a shorter expiry period compared to segment-1 products.
* **Segment-2** mostly sold more meat, deli and dry groceries which are meat and processed food.

1. Please provide a Tableau visualization (saved as a Tableau Public file) that shows the location of the stores, uses color to show cluster, and size to show total sales.



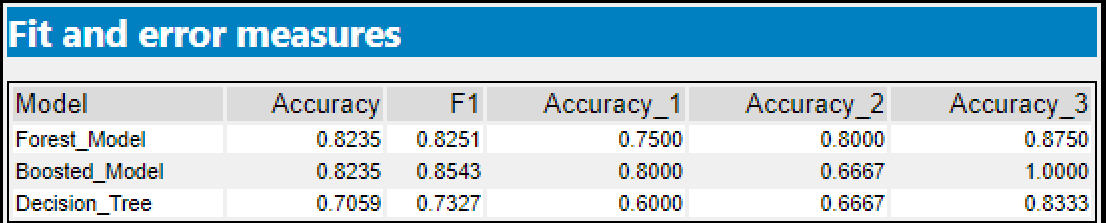
<https://public.tableau.com/views/StoreClustersbasedontheretypeofproductsales/Sheet1?:display_count=y&publish=yes&:origin=viz_share_link>



## Task 2: Formats for New Stores

1. What methodology did you use to predict the best store format for the new stores? Why did you choose that methodology?

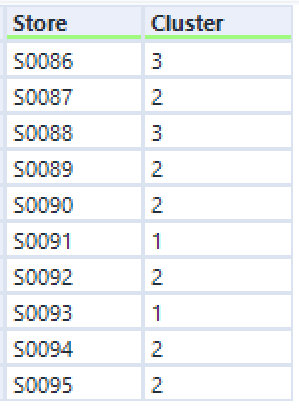
Based on the model performance report

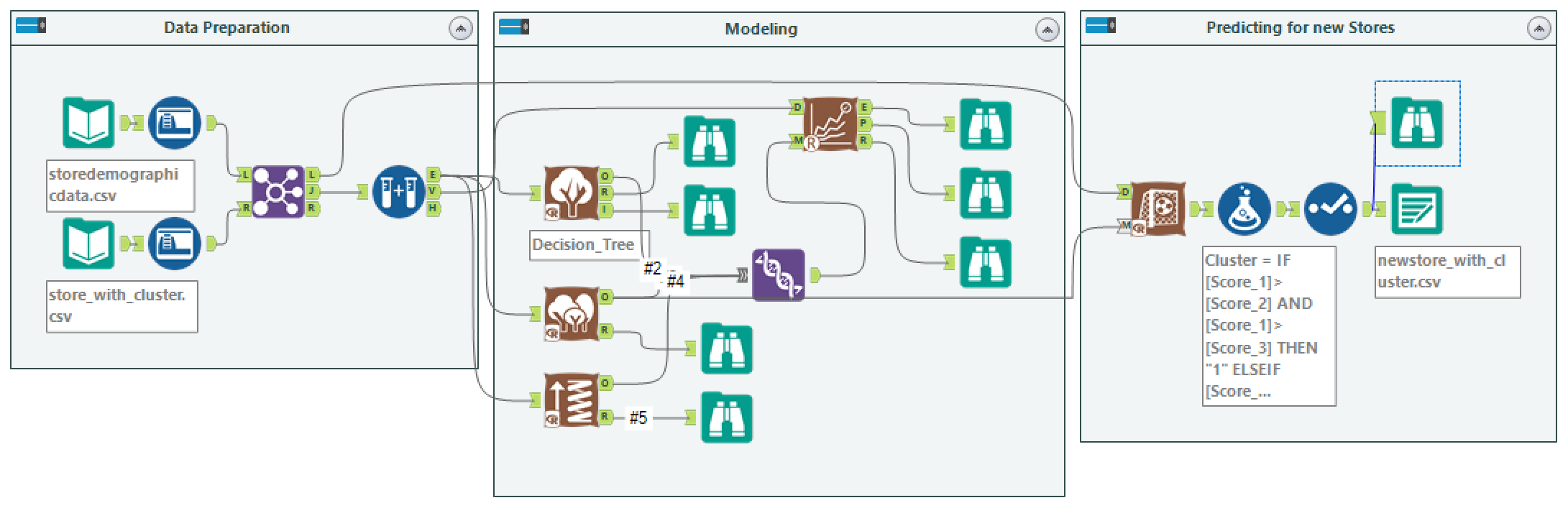


Forest model shows the highest accuracy and very less biasness between categories. Hence selecting Forest Model for predicting segment for new stores.

1. What format do each of the 10 new stores fall into? Please fill in the table below.

Segment interchangeably named as cluster.

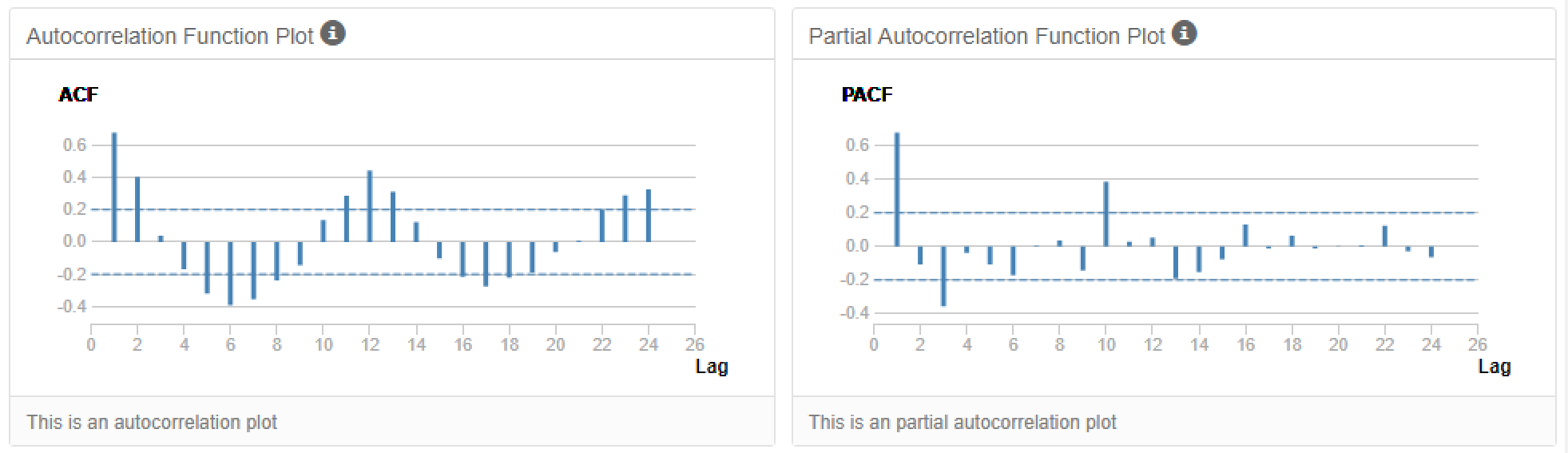
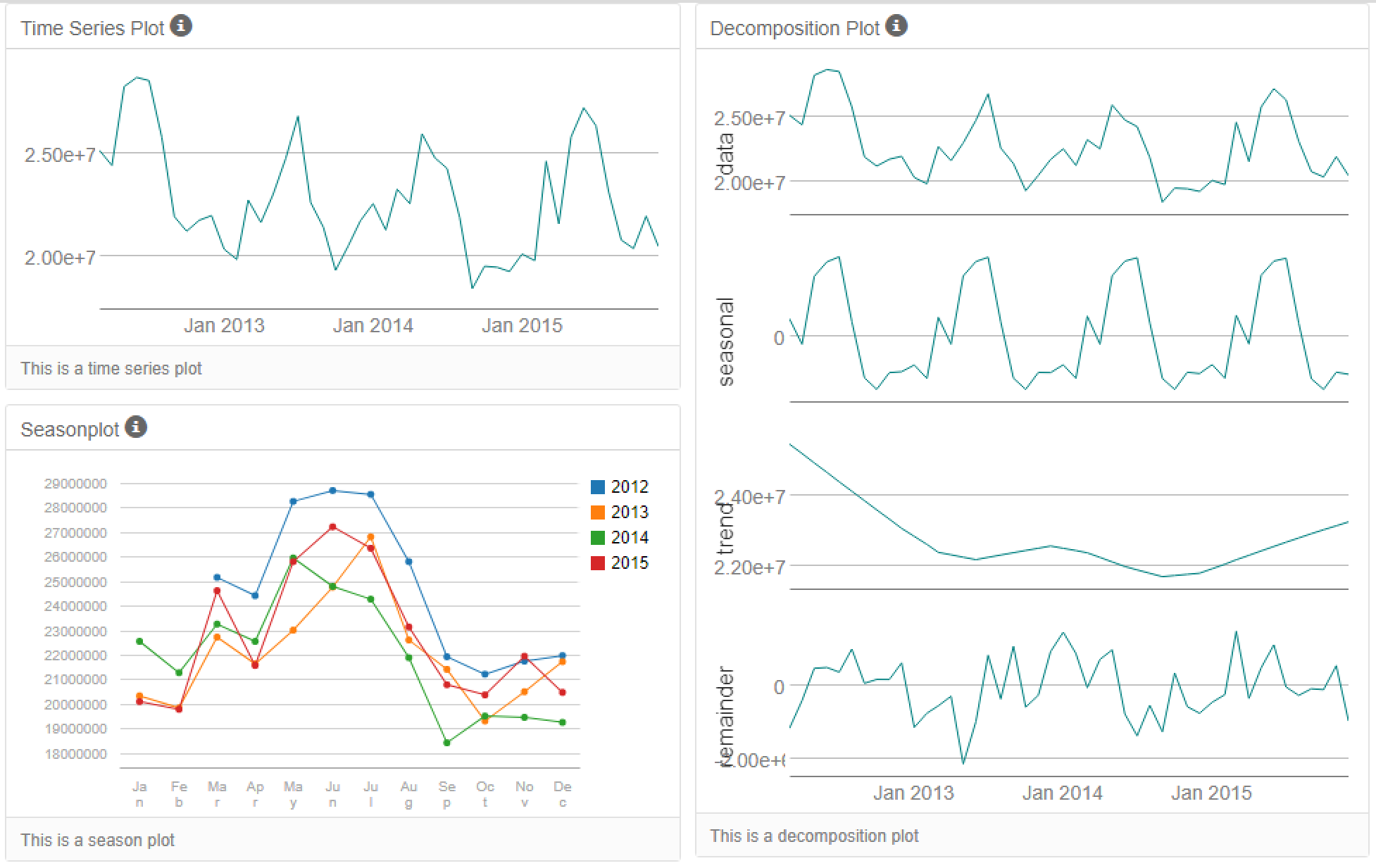




## Task 3: Predicting Produce Sales

1. What type of ETS or ARIMA model did you use for each forecast? Use ETS(a,m,n) or ARIMA(ar, i, ma) notation. How did you come to that decision?

By referring to the TS Plot report



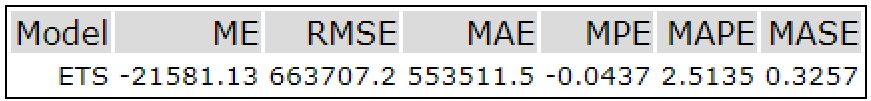
There is an existent seasonality no prominent trend and error element has a slight

increase.

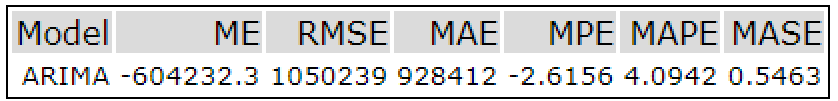
Holding 6 months of data for validation as suggested.

Below are the reports after validating ETS and ARIMA models on the hold out data.

* **ETS**

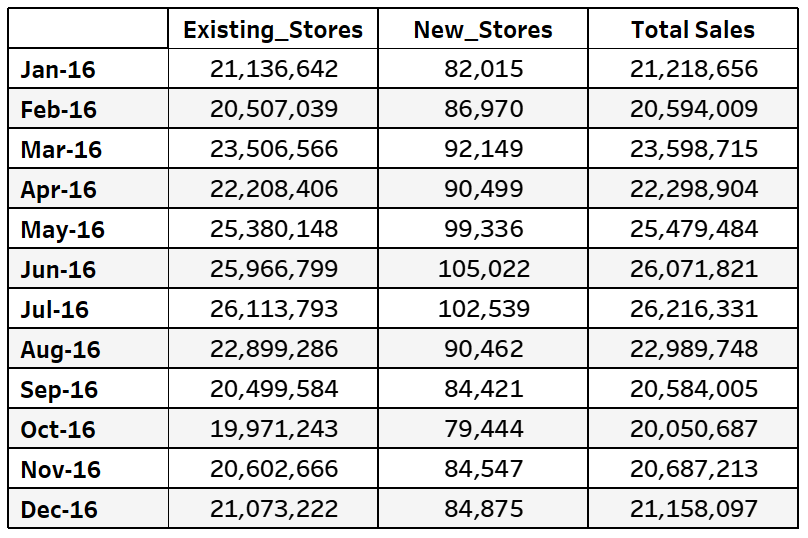


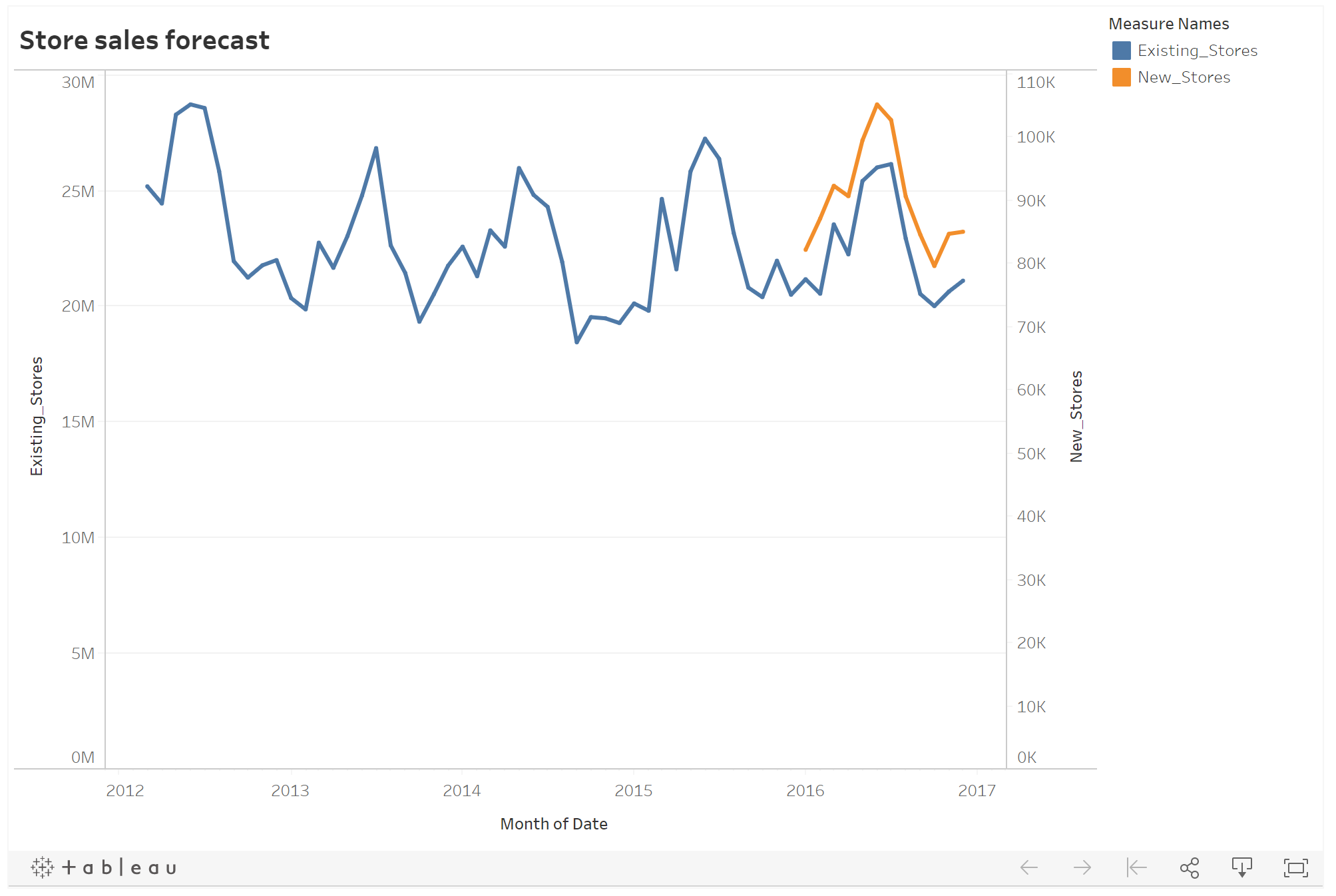
* **ARIMA**



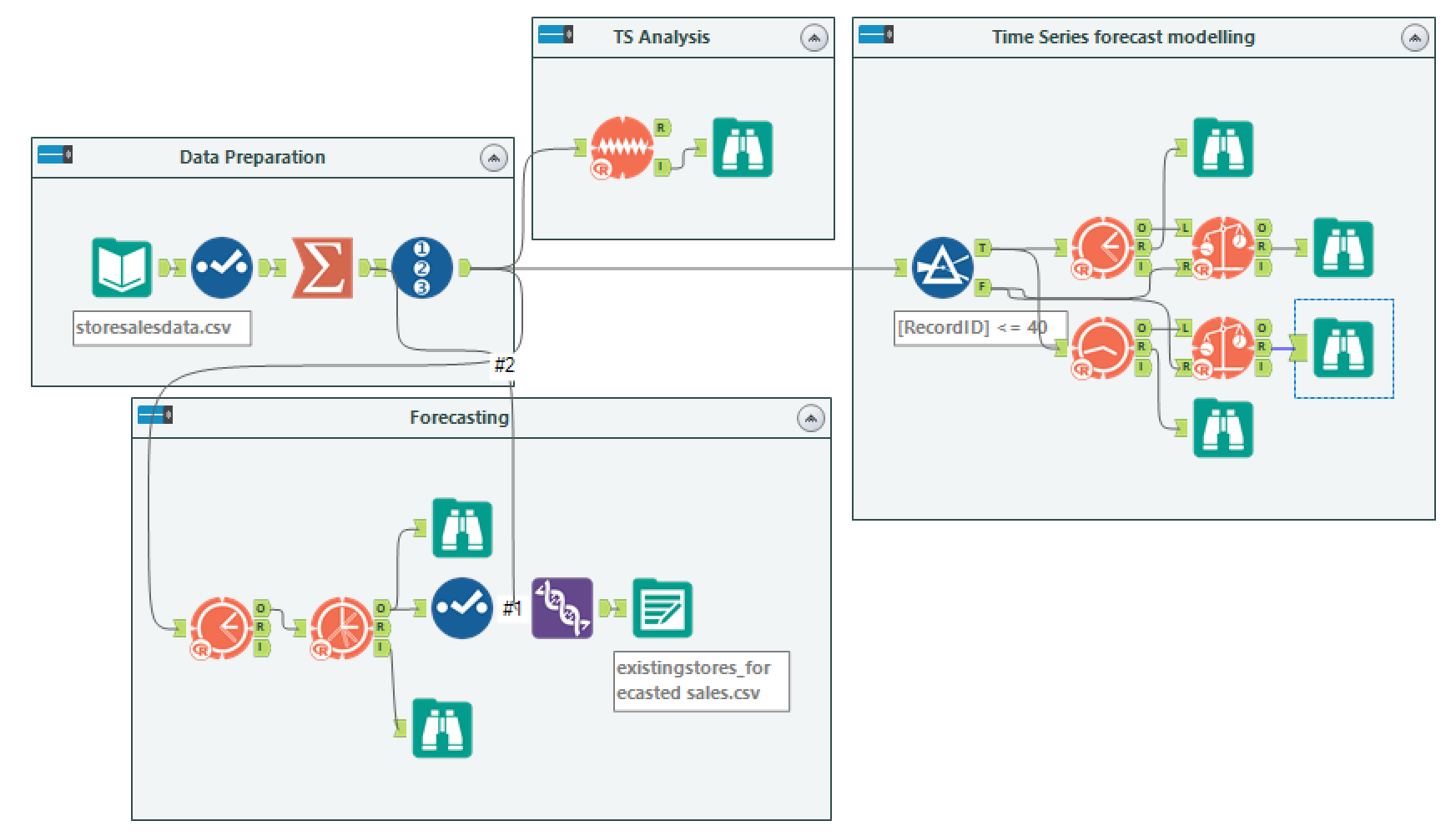
ETS model has considerably low RMSE and other error ratings hence choosing ETS model for forecasting.

2. Please provide a table of your forecasts for existing and new stores. Also, provide visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts.

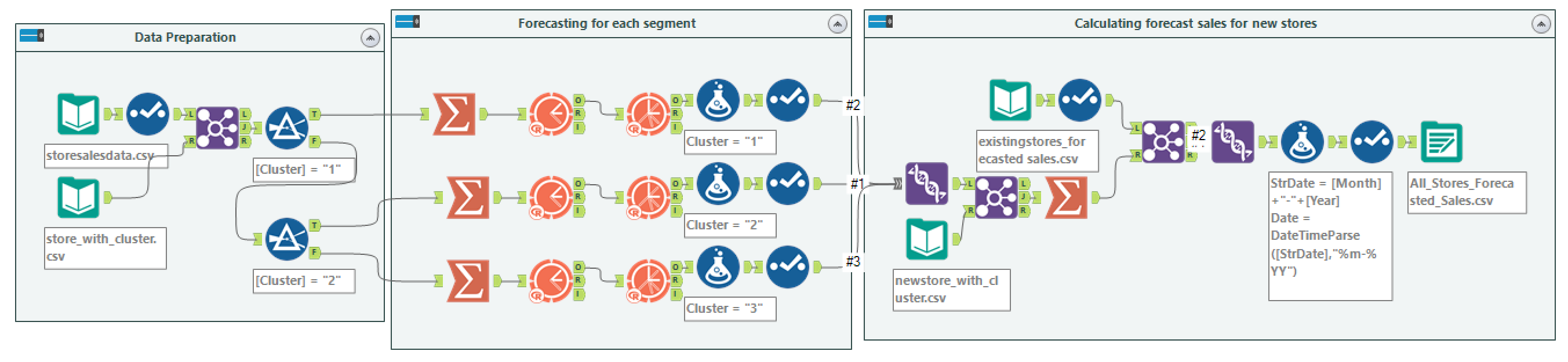




<https://public.tableau.com/views/Storesalesforecast/Sheet1?:display_count=y&publish=yes&:origin=viz_share_link>



**Total produce sales forecast for exiting stores**



**Produce sales forecast for new stores**