Project: Predictive Analytics Capstone

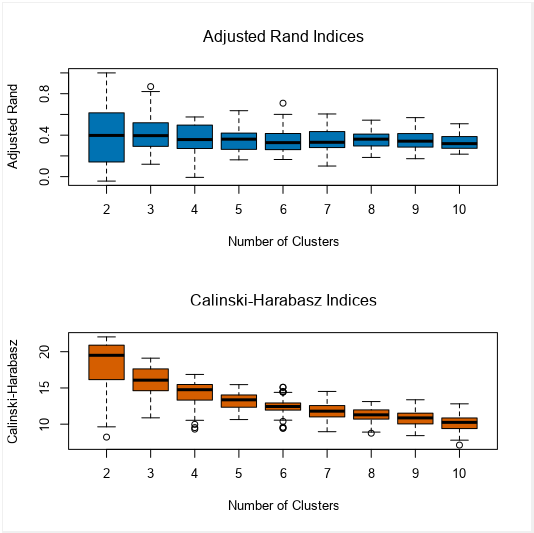
Complete each section. When you are ready, save your file as a PDF document and submit it here: <https://coco.udacity.com/nanodegrees/nd008/locale/en-us/versions/1.0.0/parts/7271/project>

## Task 1: Determine Store Formats for Existing Stores

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

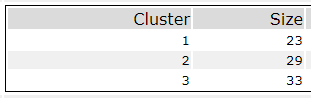
It has been determined to implement 3 store formats.

Using the K-Centroid Cluster Diagnosis reports for the K-Means method, has been observed that with 3 Clusters the AR and CH indices have the higher median and smaller variation at the same time per the whisker plot below.



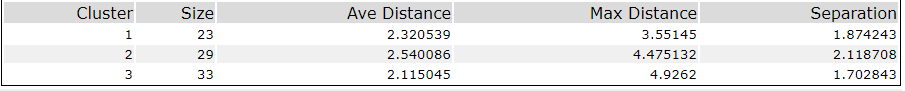
1. How many stores fall into each store format?

Using the K-Centroid Cluster Analysis for K-Means, has been obtained the next report indicating the next number of stores per cluster id.

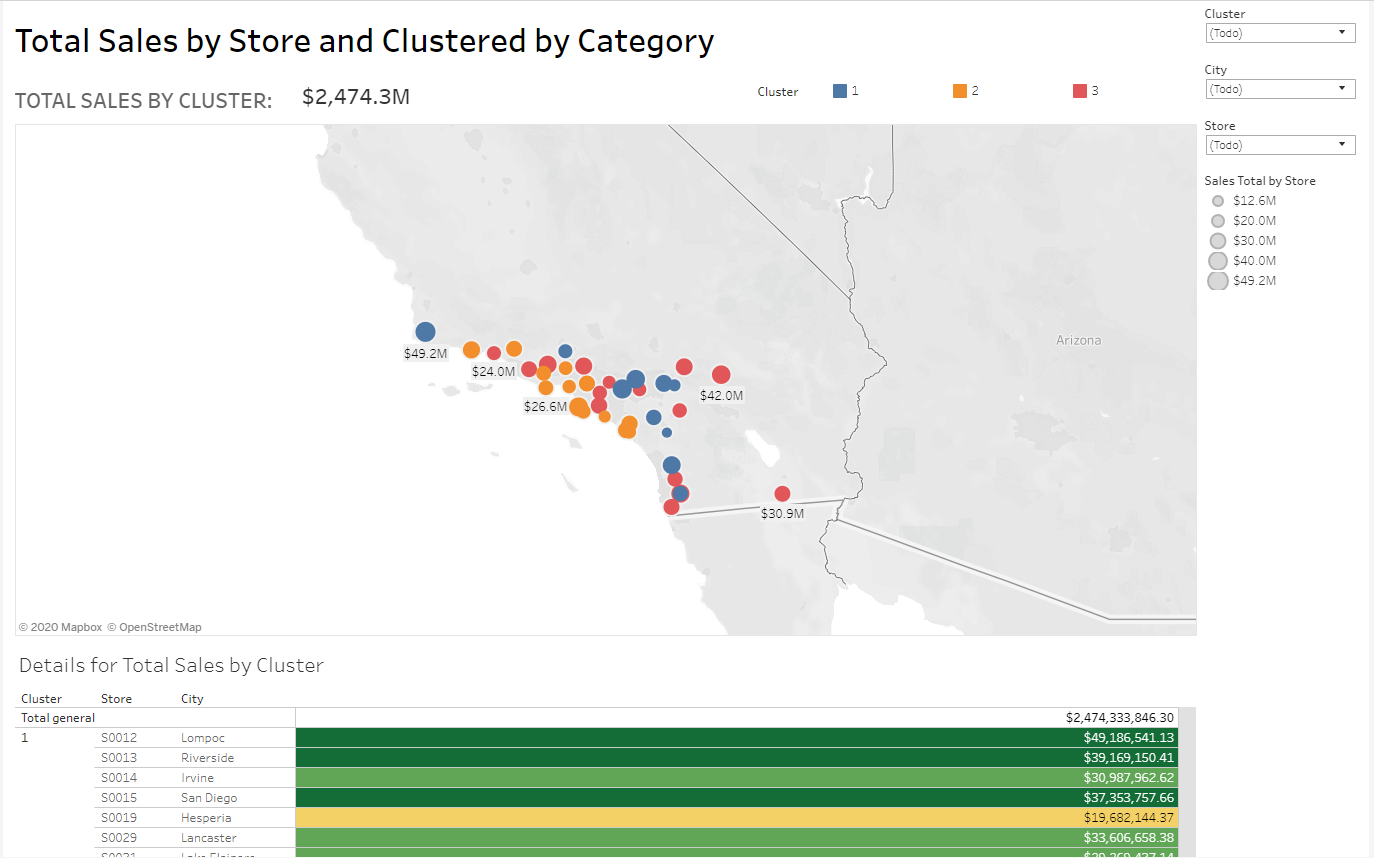


1. Based on the results of the clustering model, what is one way that the clusters differ from one another?

This difference for every cluster is given by the Distance and Separation. And, when comparing cluster values for same category, by the higher and lower values.



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.



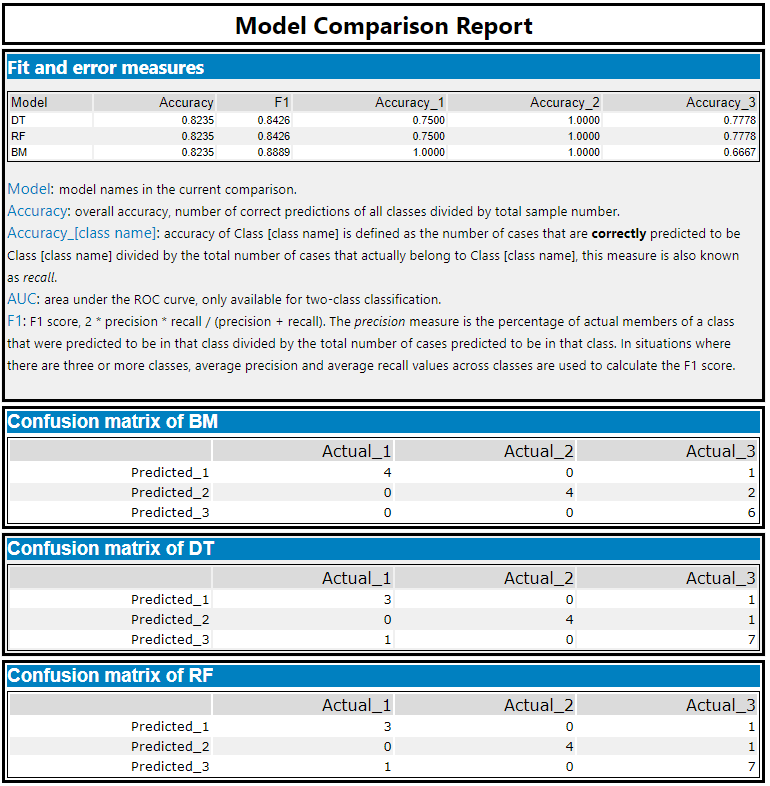
You can see the tableau dashboard using the next link:

<https://public.tableau.com/profile/felix.hernandez8665#!/vizhome/Task1_Clustered_Stores/DashboardStoreSalesClustered?publish=yes>

## 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? (Remember to Use a 20% validation sample with Random Seed = 3 to test differences in models.)

It has been used the suggested configuration for sampling and random seeds, and comparing the accuracy level for the 3 models is the same, but also considering the F1 Score the best performer is Boosted Model (BM) as per the table and confusion matrix below.



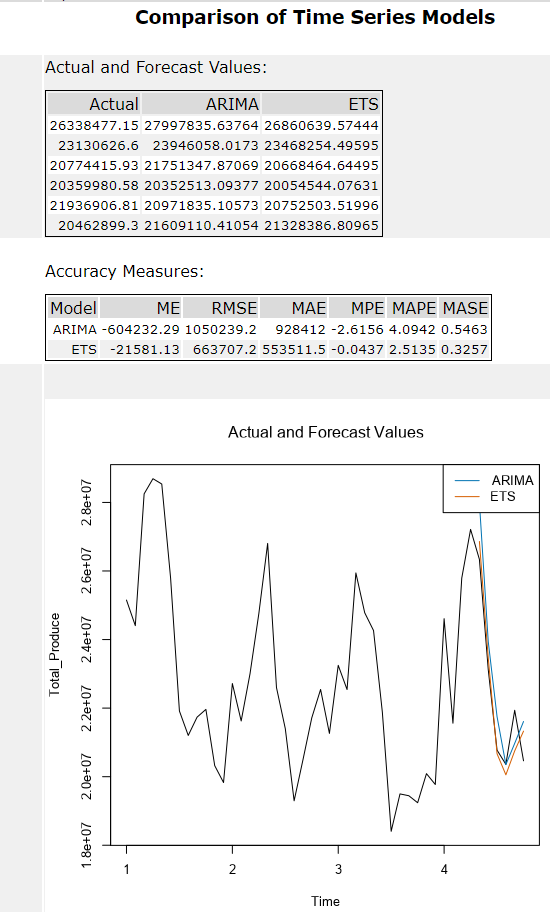
1. What format do each of the 10 new stores fall into? Please fill in the table below.

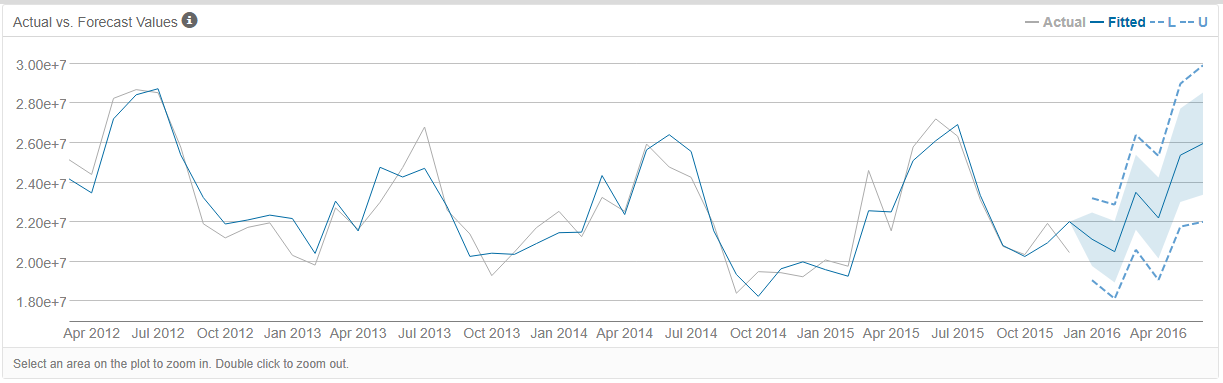
|  |  |
| --- | --- |
| Store Number | Segment |
| S0086 | 3 |
| S0087 | 2 |
| S0088 | 1 |
| S0089 | 2 |
| S0090 | 2 |
| S0091 | 1 |
| S0092 | 2 |
| S0093 | 1 |
| S0094 | 2 |
| S0095 | 2 |

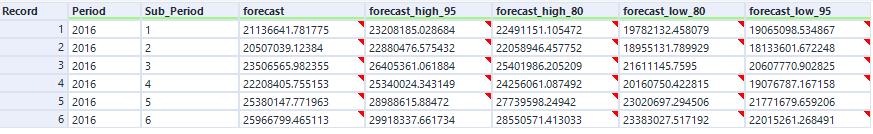
## Task 3: Predicting Produce Sales

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?

The ETS (M, N, M) Model has given a more accurate forecast per the report generated for Time Series Comparison, mainly considering the lower values for RMSE and MASE, and a nearest results compared against the Hold Out validation for 6 months.

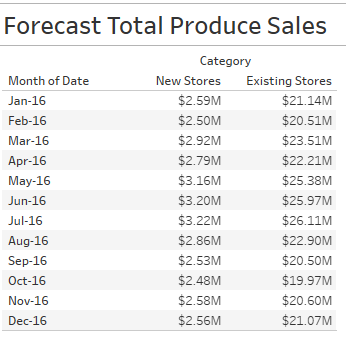




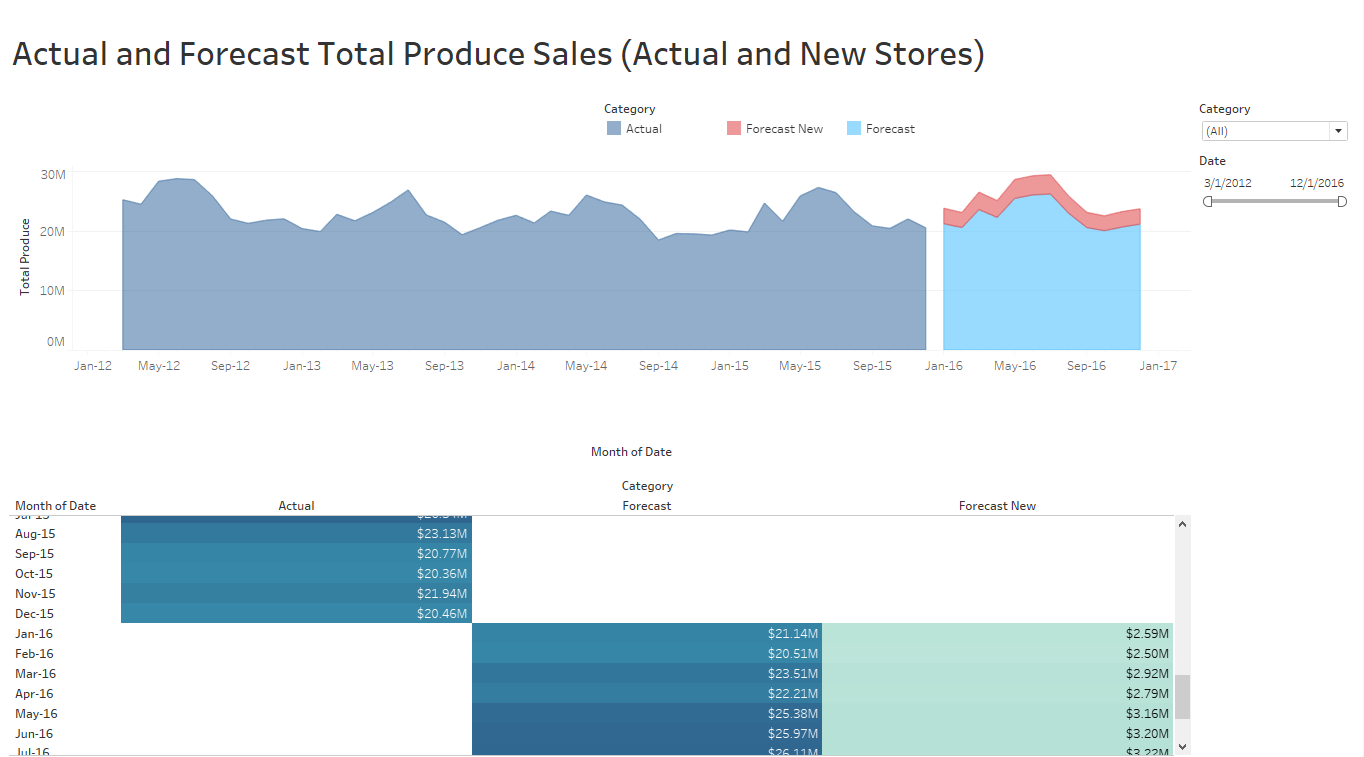


1. Please provide a table of your forecasts for existing and new stores. Also, provide visualization of your forecasts that includes historical data, existing stores forecasts, and new stores forecasts.

The table with forecast for new stores and existing stores can be seen below



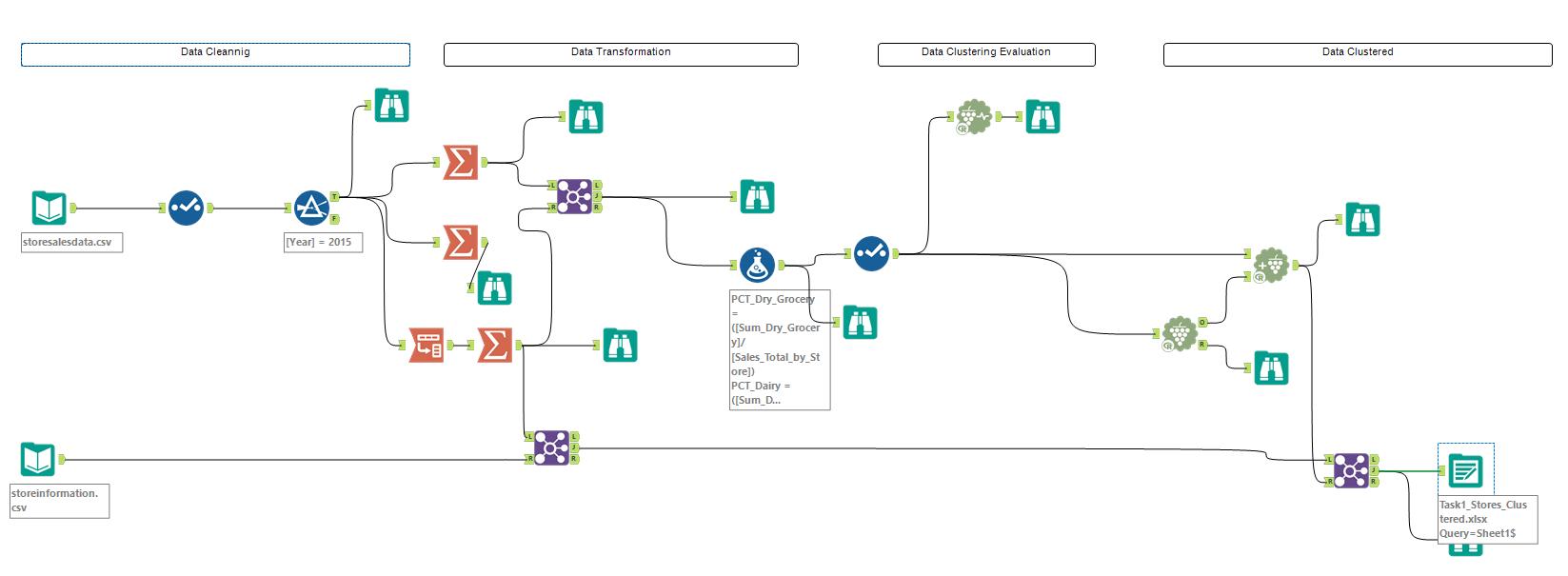
The tableau dashboard can be seen below.



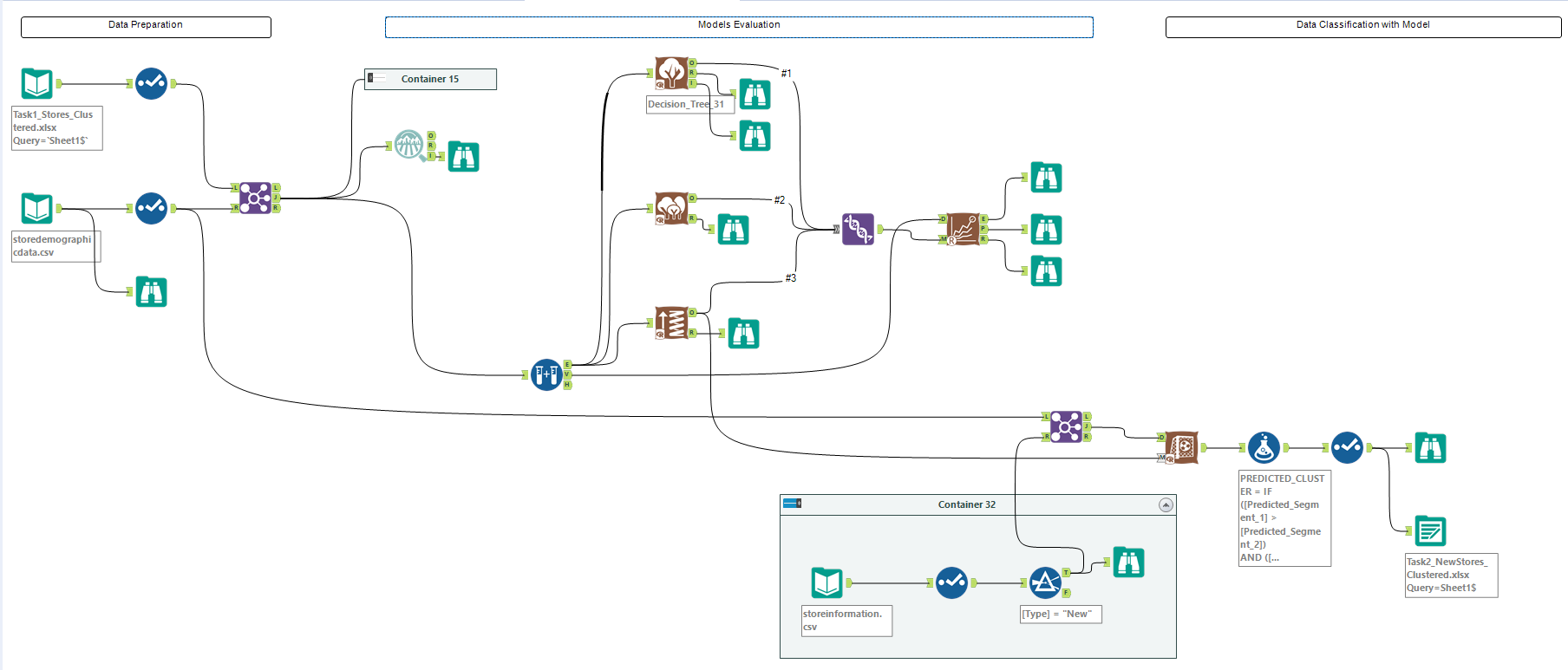
You can see the tableau dashboard using the next link:

<https://public.tableau.com/profile/felix.hernandez8665#!/vizhome/Task3_TotalSales_for_Actual_and_New_Stores_with_Forecast/TotalProduceSalesActualandNewStores?publish=yes>

Task 1 Alteryx Workflow:

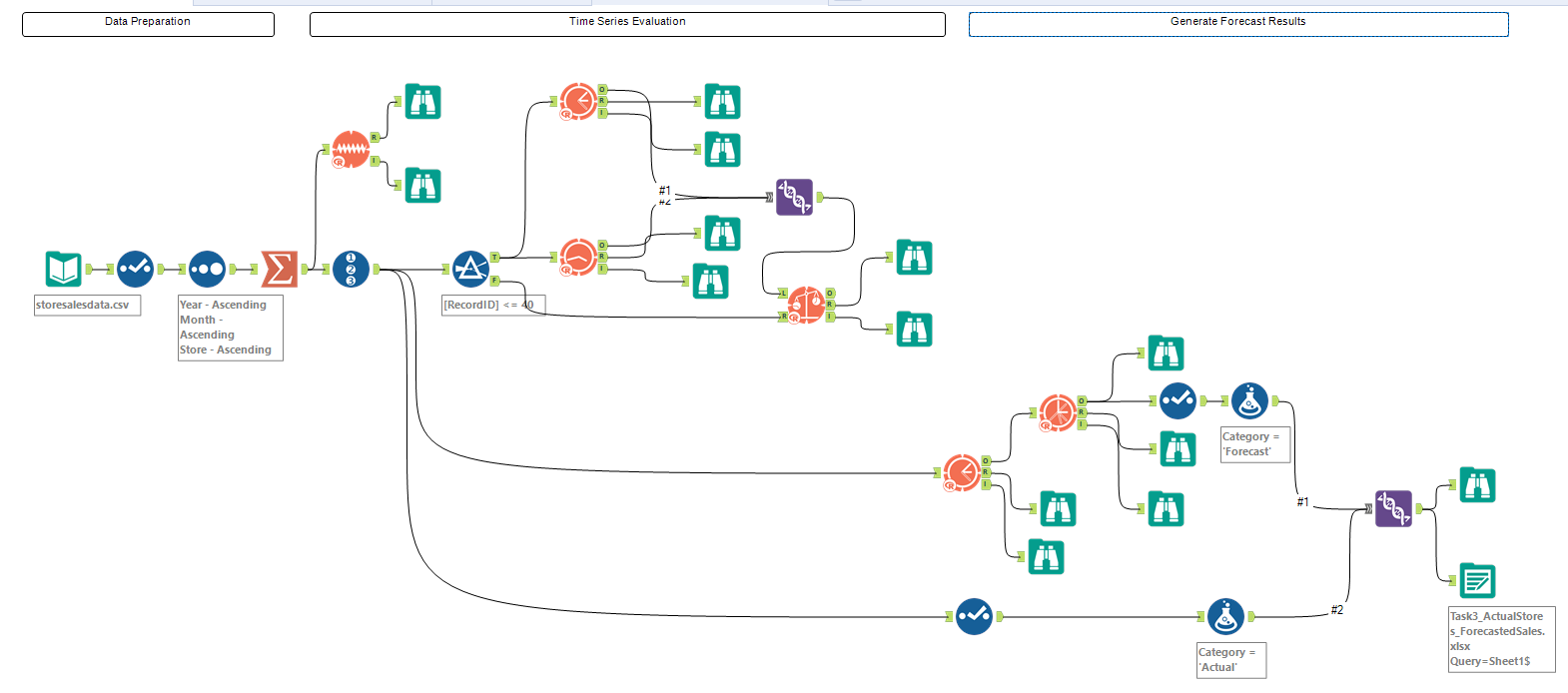


Task 2 Alteryx Workflow:



Task 3 Alteryx Workflows:

Existing Stores:



New Stores:

