**Capstone Project Overview**

The capstone project has three main tasks, each of which requires you to use a skills you developed during the nanodegree program. One you complete all three tasks, please submit a PDF and Tableau Public file with your project.

**Tips**

* **Split up your Alteryx workflows:** There will be using multiple data sources and complex tools, which can slow down software. Splitting up the workflows makes the process more manageable.
* **Map out your work:** Before you dive into your analysis, think about the steps and plan ahead. This will reduce the amount of unnecessary work.
* **Post to Slack and/or Forums for help:** Other students and forum mentors can help you if you get stuck. When posting, make sure you include enough specificity so others can help.

## The Busine

## ss Scenario: Store Format

Your company currently has 85 grocery stores and is planning to open 10 new stores at the beginning of the year. Currently, all stores use the same store format for selling their products. Up until now, the company has treated all stores the same, shipping the same amount of product to each store. This is beginning to cause problems as stores are suffering from product surpluses in some product categories and shortages in others. You've been asked to provide analytical support to make decisions about store formats and inventory planning.



### Task 1: Determining Store Format

To remedy the product surplus and shortages, the company wants to introduce different store formats. Each store format will have a different product selection in order to better match local demand. The actually building sizes will not change, just the production selection and internal layouts. The terms formats and segments will be used interchangeably throughout this project. You’ve been asked to:

* Determine the optimal number of store formats, based on sales data. HINT: Calculate the average sales (or percentage sales) per category per store for clustering.
* Segment the 85 current stores into the different store formats.
* Use the StoreSalesData.csv and StoreInformation.csv files.

### Task 1 Submission

1. What is the optimal number of store formats? How did you arrive at that number?
2. How many stores fall into each store format?
3. Based on the results of the clustering model, what is one way that the clusters differ from one another?
4. Please provide a map created in Tableau that shows the location of the stores, uses color to show cluster, and size to show total sales. Make sure to include a legend! Feel free to simple copy and paste the map into the submission template.

NEXT

## The Business Scenario: New Stores

The grocery store chain is has 10 new stores opening up at the beginning of the year. The company wants to determine which store format each of the new stores should have. However, we don’t have sales data for these new stores yet, so we’ll have to determine the format using each of the new store’s demographic data.



Pretty sweet grocery store, right?

### Task 2: Store Format for New Stores

You’ve been asked to:

* Develop a model that predicts which segment a store falls into based on the demographic and socioeconomic characteristics of the population that resides within a five minute drive around each new store.
* Use the model to predict the best store format for the each of 10 new stores.
* Use the StoreDemographicData.csv file.

### Task 2 Submission

* What methodology did you use to predict the best store format for the new stores? Why did you choose that methodology?
* What are the three most important variables that help explain the relationship between demographics indicators and store formats? Please include a visualization.
* What format do each of the 10 new stores fall into? Please provide a data table.

## The Business Scenario: Forecasting

Fresh produce has a short life span, and due to increasing costs, the company wants to have an accurate monthly sales forecast.



### Task 3: Forecasting Produce Sales

You’ve been asked to prepare a monthly forecast for produce sales for the year of 2015 for both existing and new stores. To do so, follow the steps below.

**Step 1:** To forecast sales for existing stores you should aggregate sales across all stores by month and produce a forecast.

**Step 2:** To forecast sales for new stores:

* Forecast produce sales for the average store (rather than the aggregate) for each segment.
* Multiply the average store sales forecast by the number of new stores in that segment.
* For example, if the forecasted average store sales for segment 1 for March is 10,000, and there are 4 new stores in segment 1, the forecast for the new stores in segment 1 would be 40,000.
* Sum the new stores sales forecasts for each of the segments to get the forecast for all new stores.

**Step 3:** Sum the forecasts of the existing and new stores together for the total produce sales forecast.

### Task 3 Submission

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?
2. Please provide a Tableau Dashboard (saved as a Tableau Public file) that includes a table and a plot of the three monthly forecasts; one for existing, one for new, and one for all stores. Please name the tab in the Tableau file "Task 3".

### Review

Use the [**project rubric**](https://review.udacity.com/#!/projects/252/rubric) to review your project. If you are happy with your submission, then you're ready to submit your project. If you see room for improvement, keep working to improve your project.

### Submission Template

Use the submission template at the bottom of this section to submit your project. After filling it out, save it as a PDF and submit the PDF in the next section.

#### Data

* StoreSalesData.csv - This file contains sales by product category for all existing stores for 2012, 2013, and 2014.
* StoreInformation.csv - This file contains location data for each of the stores.
* StoreDemographicData.csv - This file contains demographic data for the areas surrounding each of the existing stores and locations for new stores.

#### Supporting Materials

**[StoreSalesData](https://d17h27t6h515a5.cloudfront.net/topher/2016/November/582bff6a_storesalesdata/storesalesdata.csv" \t "_blank)**

**[StoreInformation](https://d17h27t6h515a5.cloudfront.net/topher/2016/November/582bff5f_storeinformation/storeinformation.csv" \t "_blank)**

**[StoreDemographicData](https://d17h27t6h515a5.cloudfront.net/topher/2016/November/582bff9e_storedemographicdata/storedemographicdata.csv" \t "_blank)**

**[P8 Submission Template](https://d17h27t6h515a5.cloudfront.net/topher/2016/November/582c0282_p8-submission-template/p8-submission-template.docx" \t "_blank)**

NEXT