Demand Forecasting for a Luxury Fashion Retailer

BCG Team: Arun Ravindran & Anton van Pamel
MIT Mentor: Robert Freund
Location: Boston, Massachusetts

Project Overview

Project Importance

- Luxury retailers make little revenue from ready-to-wear clothes.
- Approximately 90% of revenue comes from handbags, shoes, accessories, and fragrances.
- Gross margins for handbags are often the highest across all departments, so an accurate demand forecast is crucial.

Project Scope

- Our project was forecasting demand for women's handbags in their European stores.
- Specifically, we created a model to predict demand for handbags that are part of the new seasonal collection, meaning they have no historical sales.

Below is the criteria on which we filtered the data.

- Women's handbags
- Full-price items
- European stores
- Retail locations
- Gross sales
- New Collection

Project Timeline

January - February	→ Exploratory Data Analysis → Initial Demand Forecasting Models
March - April	→ Feature Engineering → Constructing Panel Data
May - June	→ Pairwise Comparison Research → Efficient Algorithm Implementation
July - August	→ Sophisticated Demand Forecasting Models → Summer Capstone Showcase

Data Overview

Raw Data

We were given four datasets:

Stores | Products | Transactions | Inventory

- We merged these four datasets and then filtered the data to reflect the scope of our project.
- Our merged dataframe had approximately 45,000 rows. Within that dataframe, there are over 1,300 unique stock keeping units (SKUs) and about 120 unique store locations.

Store Clusters

The client provided us with five store clusters, labeled A through E. We analyzed each of these clusters and created a short description for each.

- **A** **Flagship Store:** highest amount of stock and sales with the most expensive purchases
- **B** **Large City Locations**: comparable sell-through rate to the flagship store, but with fewer overall sales and less expensive sales
- **C** **Resort Locations:** low stock and low sell-through rate, but high prices and located in resort towns
- **D** **Traditional City Locations:** do not carry the highest price-point, but has a high sell-through rate and high revenue to square footage ratio
- **E** **Low Volume Stores:** assortment of low volume store with the overall lowest price point (includes airport locations)

Data Processing

Clustering Data using *k*-Prototypes

- We applied clustering to our data using *k*-prototypes, which integrates *k*-means and *k*-modes algorithms to cluster both continuous and categorical variables.
- We selected the number of clusters by validating on the model's overall performance.
- These clusters helped us build new features, such as historical sales and stock-made by cluster.

Reducing Proportion of Null Values

- We imputed missing values in the dataframe using analytical expertise and ETL techniques.
- Using our analytical expertise, for example, we inspected the data and replaced null values with zero for binary features.
- We used ETL techniques to create an aggregated feature, and by merging datasets on this aggregated feature, the number of null values was dramatically reduced.

Dummifying Data and Deleting a Degree of Freedom

- We dummified the categorical variables and, when doing so, we deleted the extra degree of freedom.
- This approach decreased the complexity of the data and increased the performance of our model.

For example, the variable `Ornaments` has four levels: `Pearls`, `Studs`, `Swarovski`, and `None`. If we reduced these dummified features to `Pearls on handbag`, `Studs on handbag`, and `Swarovski on handbag`, all of the information can be captured.

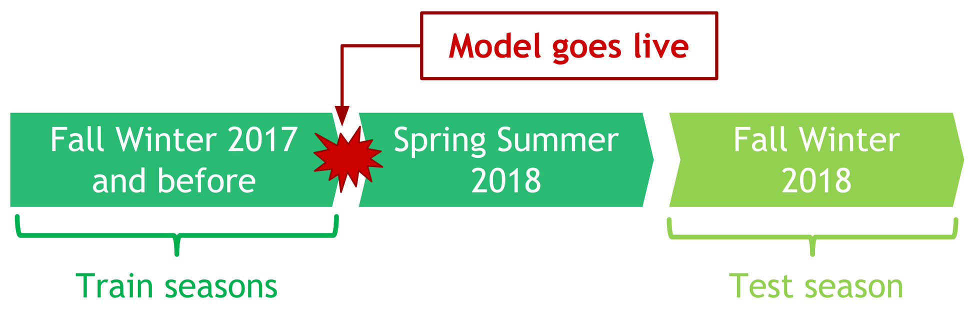
Project Challenges

We faced two main challenges:

- Forecasting is done in advance to allow for manufacturing.
- There are few similarities between the train and test set.

Manufacturing Timeline

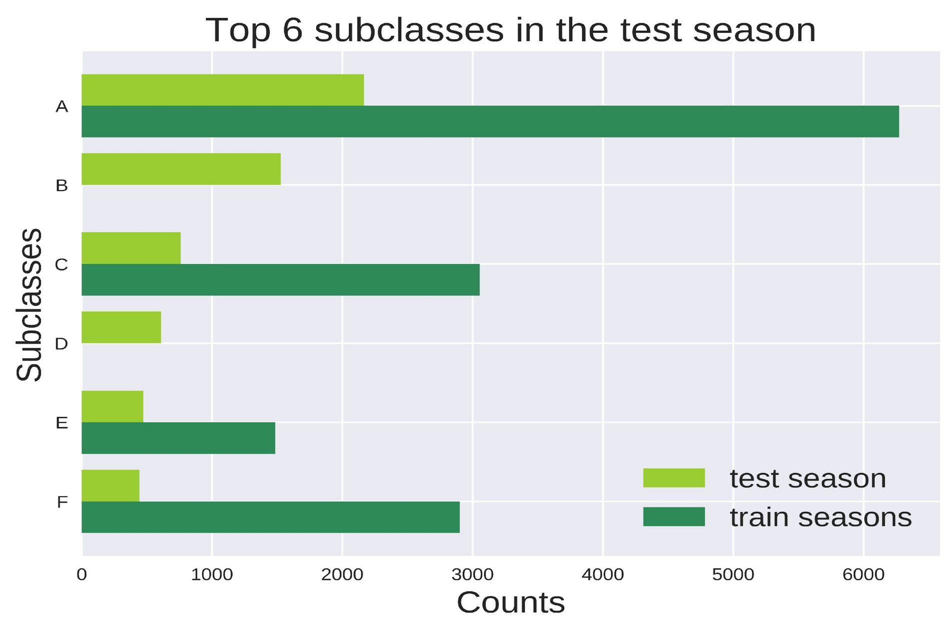
- Demand forecasting must be done six months before the season begins because of the manufacturing timeline.
- We predicted handbag demand for the Fall-Winter 2018 season.
- We trained our modeling using data from the Fall-Winter 2017 season and earlier.
- However, to allow for manufacturing and shipping, our model has to goes live at the beginning of the Spring-Summer 2018 season in order to predict demand for Fall-Winter 2018.



Dissimilarities between Train and Test

- Significant discrepancies exist between the train and test sets, which makes accurate predictions difficult for the test set.
- We inspected the number of SKUs made per subclass for the two datasets to asses the dissimilarity.

Below is a plot showing that some subclasses are prevalent in the test set but absent from the train set.



Feature Engineering

Historical Features

We created historical features by lagging the last two seasons of data. Because these SKUs are part of the new collection, however, we have no historical sales for the SKUs so we lagged on product category (i.e. type of material, color of bag, etc.).

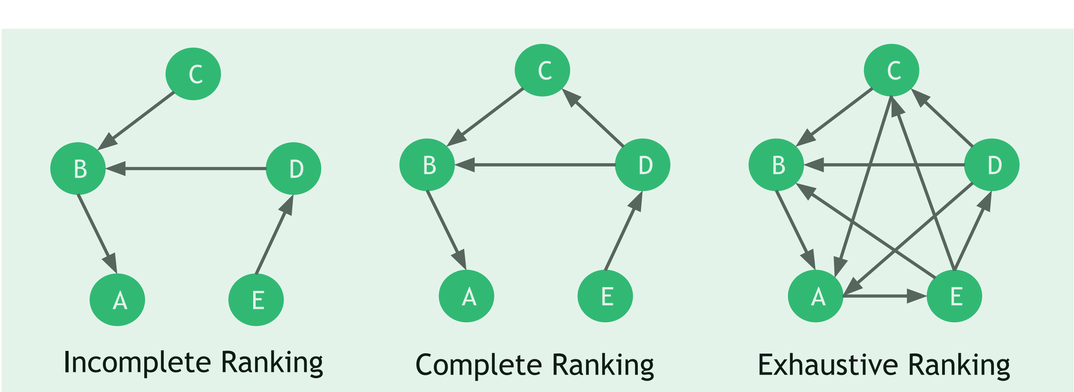
- Sales for that category for season-1
- Sales for that category for season-2
- Stock-made for that category for season-1
- Stock-made for that category for season-2
- Sell-through rate for that category for season-1
- Sell-through rate for that category for season-2

Product and Store Features

- Aggregated style and color features were created to decrease dissimilarity between the train and test sets.
- Consider a granular five-digit color code for a green bag, where the first three digits indicate that it is green, the next digit indicates the brightness of the shade, and the final digit signifies the exact hue of green.
- By reducing this feature to an aggregated three-digit code, we are able to find more similarities between the train and test set.

SKU Popularity: A Bayesian Approach

- We created the SKU popularity feature and included it in the model before its training, like in a Bayesian framework.
- To create this feature, store managers will perform pairwise comparisons of SKUs, allowing us to include human intelligence to our machine learning model.
- It is too time consuming to compare every pair of SKUs to obtain a global ranking. Therefore, we use an adaptive ranking algorithm to select the next pair to compare in order to minimize the total number of comparisons needed.



- In our algorithm, we use directed graphs: each node represents a SKU and an edge is added between two nodes when those SKUs have been compared.
- Our ranking can be obtained if all nodes are connected in our directed graph, as shown above.

Model and Results

Model Selection: Random Forest

- We tried three models: Elastic Net, CART, and Random Forest.
- We evaluated these three models on mean absolute error (MAE) and mean absolute percentage error (MAPE).
- We selected the Random Forest model not only because it has the best performance, but also because it is interpretable.

	Elastic Net	CART	Random Forest
MAE	6.08	5.64	**4.74**
MAPE	138%	147%	**130%**

Feature Importance

- Below are the most significant features in our model.
- We created all of these top features except for SKU price.
- This emphasized to us the importance of feature engineering to extract important signals from the data to feed into the model.

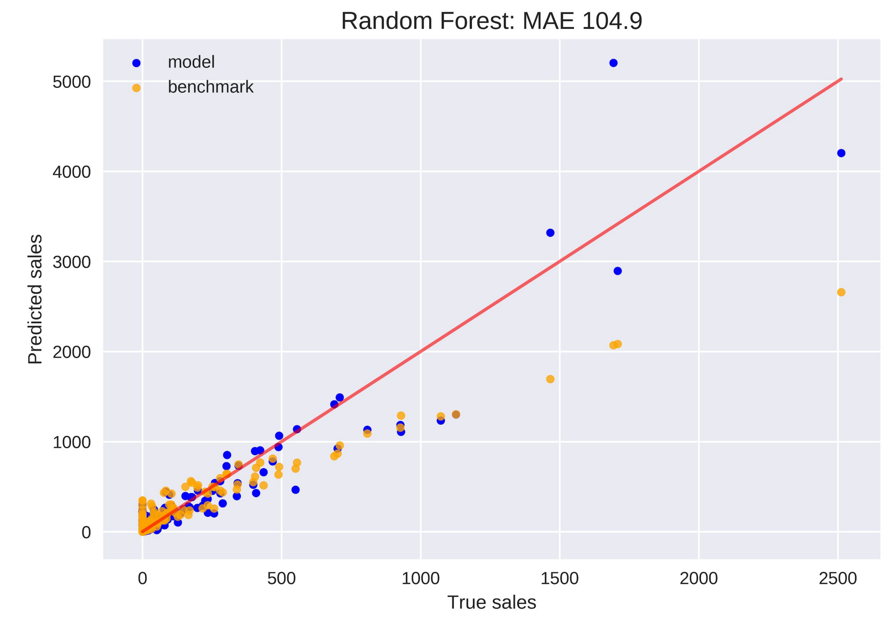
Rank of Importance	Feature
1	SKU popularity
2	Historical sales
3	Store popularity
4	Number of competing SKUs
5	SKU price
6	SKU launch month

Performance Compared to Benchmark

- In order to convince the client of the validity of our model, we compared its performance to the benchmark, which is the amount of stock made by the client per SKU for each season.
- We assessed the performance of our model and the benchmark using MAE and price MAE, which is MAE weighted by SKU price.

	Benchmark	Model
MAE	126.8	104.9
Priced MAE	204k €	168k €

- The client wants to make twice as much stock as they expect to sell to buffer for supply-chain logistics and ensure that stores are fully stocked. Therefore the objective of our model, which is the red line below, is to predict *2 x sales*.
- In the figure below, we plot our model's predictions in blue and the benchmark's predictions in orange.



Recommendation: Potential Demand

- Sales are a proxy for demand since stock-outs could have caused fewer sales to occur.
- We trained a Random Forest model to predict sales for which there were no stock-outs and then predicted sales for weeks in which there were stock-outs.
- Below is the plot for the Fall-Winter 2016 season, for which we predict that demand is 10.01% higher than sales. The MAE of this model is 0.461 and MAPE is 27.3%.

- Demand and sales are equal at the beginning of the season because no stock-outs have yet occurred, so the client is meeting all demand. Later in the season, however, there are many instances in which SKUs are out of stock.

Impact

Our Capstone project resulted in a better performing forecasting model in comparison to the client's model. This superior performance is the result of our data insights, feature engineering, and model selection. Ultimately, better forecasting improves the organizational and business performance, resulting in the following benefits:

- **Fewer missed sales:** accurately forecasting demand will ensure that inventory is in the right place at the right time.
- **Lower working capital:** the client can operate with less inventory because of confidence in demand projections.
- **Less waste:** the client is more likely to sell stock at full-price, without having to discount it because it is no longer part of the new season's collection.
- **Improved customer service:** with a deeper understanding of customer demand and unique store selling behaviors, the client can effectively deploy inventory to provide higher sell-through rates, improved on-time availability, and fewer stock-outs.