MIT MANAGEMENT
BUSINESS ANALYTICS

macy's

What WEAR When?

Demand Forecasting for Future Fashion Products

Alexander Birch
Master of Business Analytics

Vincent Lao
Master of Business Analytics

Mohammad Fazel
Faculty Advisor

Sandra Han
SVP, Merchandise Planning & Allocation

Iris Singhania
VP, Enterprise Data Science

Jingjing Ni
Principal Data Scientist

Problem Statement

"Currently, Macy's, Inc. uses a manual process for deciding product allocations to their stores across the U.S. Our goal is to make this process more data-informed -- to forecast the demand for new fashion products at the product-color-store-month level in order to inform the Merchandise Planning and Allocation team at Macy's, Inc."

44M+ Annually — Active Customers

4.3B+ Annually — Women's Apparel Revenue

$24B+ Annually — Total Revenue

Forecast Window

Historical Data — Today — First Sale — Six Month Forecast

Six Month Production Gap

Data Sources

- Product Descriptions, Product Sales
- Store Popularity, Demographics
- Time, COVID-19, Weather, Consumer Price Index

Data Summary

450 Stores + macys.com
4 Years of Data
20K products
30 Store/Time Features
86 million observations

Methodology

Input: *"lavender top, long slv ruched floral style blouse spring, weekend retreat, drawstring front elasticized details sleeves Kit & Sky, pattern floral, age group adult, print, occasion spring, casual, sleeve length sleeve, cropped, neckline square, peasant, color purple group..."*

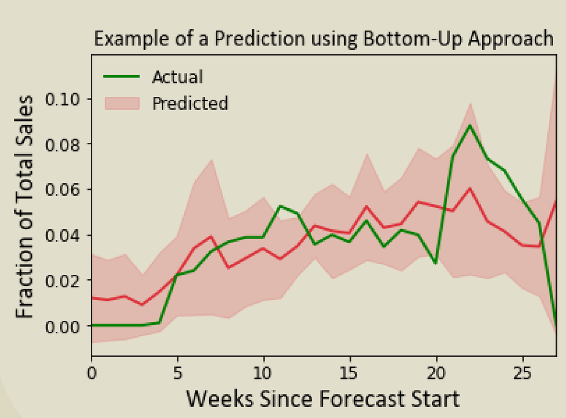
Natural Language Processing (NLP)

- We use **sentence embedding** to turn product descriptions into numerical vectors
- ***SWIVEL*** was chosen over *USE* or *BERT* for increased scalability without sacrificing performance

Product-Trend Matching

Bottom-Up

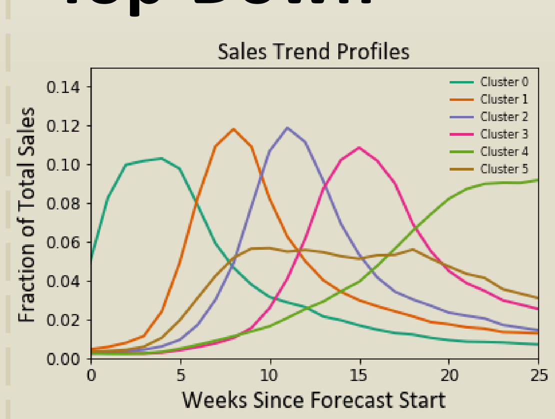
K-Nearest Neighbors (KNN) matches new products to similar products from the past.



We use the **six nearest neighbors**' sales to estimate sales for the new product.

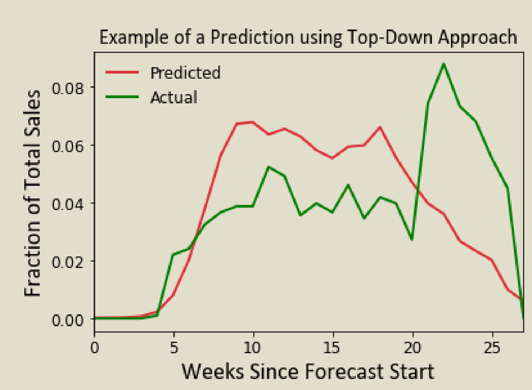
Six was found to be the optimum number of neighbors through the elbow method.

Top-Down



We use Weighted K-means Clustering to group all products into **six distinct profiles** based only on their sales trends. Six was found through a voting process between four metrics.

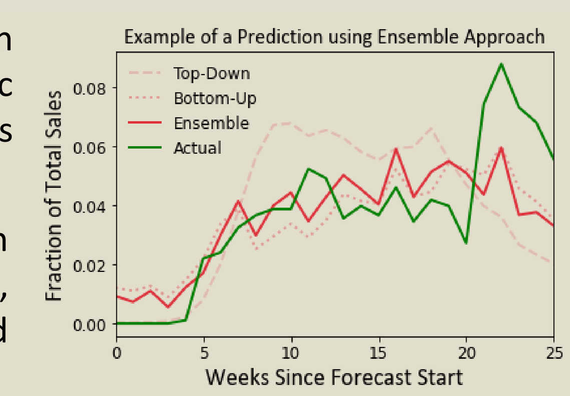
Then, we use Random Forest to **classify** products into one of these profiles utilizing their unique NLP embeddings.



Ensemble

The **Bottom-Up** approach captures the idiosyncratic trends from the products style and attributes.

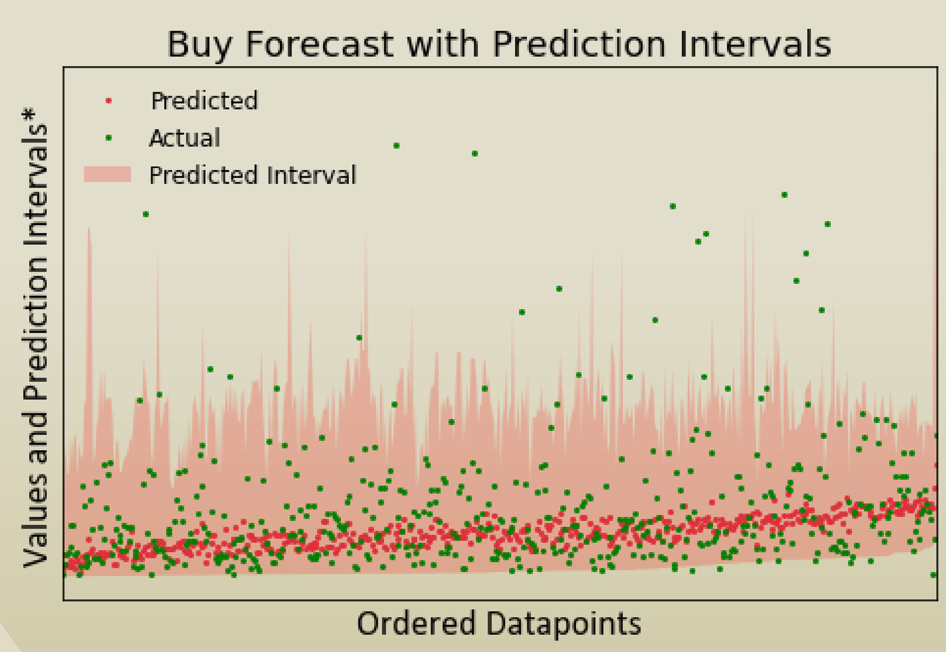
The **Top-Down** approach captures general trends, which are highly correlated with markdown schedules.



Combining the two approaches results in a model that outperforms either individually.

The models are combined using weighted mean and 'Savitzky–Golay' filtering. The weights are found by **cross-validation**.

Location-Level Buy Forecast



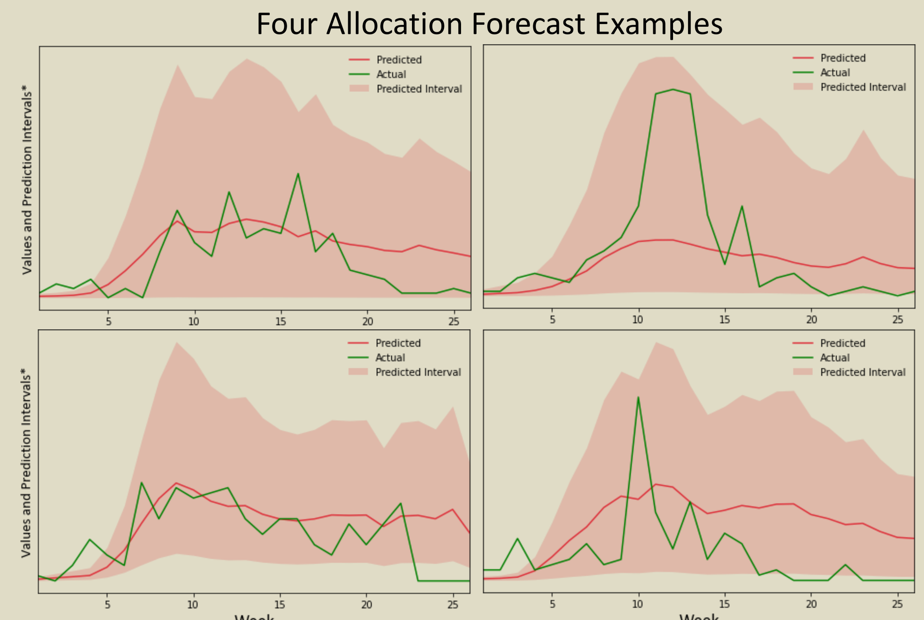
Using Random Forest, we predict the demand for a product by location, and create **prediction intervals** based on the 5th and 95th percentile tree predictions.

Since our model often under-forecasts, we can tune the interval width to tighten the range and improve usability. We separate out a second training set to tune the prediction interval percentiles in order to prevent overfitting.

Aggregating up our location-level buy forecasts, we obtain an overall **omni-channel buy forecast** for each product.

*Data censored at request of Macy's

Output: **Allocation Forecast**



By combining the location-level buy forecast and the ensemble trend forecast, we obtain a **product-store-monthly demand forecast**.

Prediction intervals provide more insight for merchandise planners than a single point prediction, as they specify a range of values for which planners can have more confidence.

From there, planners can override the prediction at their discretion if they believe the product will fall outside of the normal behavior.

Results

We implement a model pipeline that takes a product description as input and produces a product-level demand forecast across 450 stores and for *macys.com*.

The Product-Trend Matching and the .COM Buy Forecast perform well. However, the store-level buy forecast performance suffers due to data skew and low granularity. These errors propagate through the store-level allocation forecast.

Product-Trend Matching

	Bottom-Up	Top-Down	Ensemble
Spearman Correlation	0.537	0.620	**0.638**
Mean Absolute Error	0.027	0.025	**0.024**
First Peak Capture Percentage	35%	22%	**39%**

Forecasting

	Buy			Allocation		
	Store	.COM	Omni-channel	Store	.COM	Omni-channel
Root Mean Squared Error	7.23	337.35	1144.75	1.38	50.83	148.57
Prediction Interval Capture Percentage	39%	91%	79%	2%	49%	48%

Business Impact

- Innovated novel methodology for demand forecasting, pioneering a new class of forecasting at Macy's
- Models can be reinterpreted for different use cases:
  1. Top-Down Method → product popularity estimator
  2. NLP+KNN → product matching tool for recommendation systems

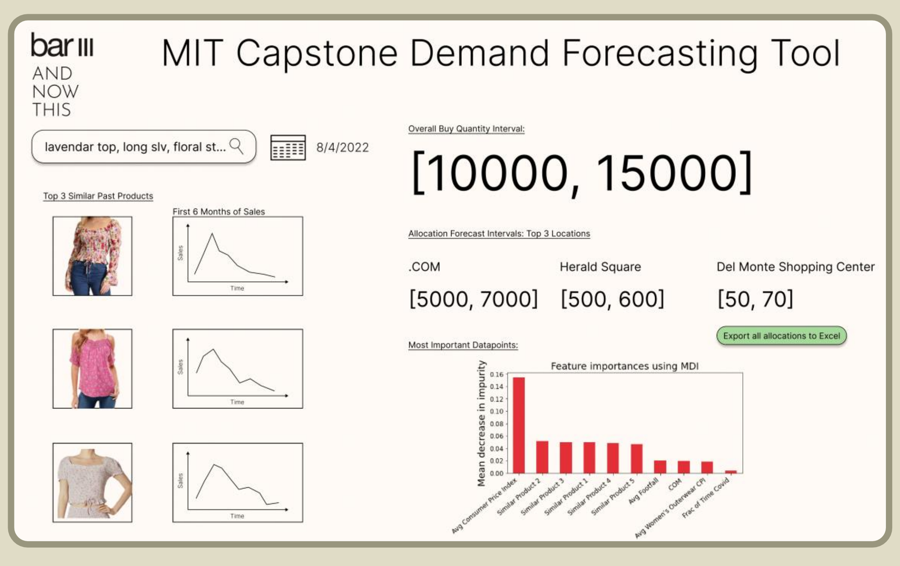
Future Considerations

Incorporate New Data Sources
- Predicted Fashion Trends
- Stockouts

Consider New Scopes of Products
- Aggregate allocation forecast across regions or climates
- Partition products based on their intended life cycle

New Methodologies
- Time-series Approaches
- Tensor Completion

Dashboard Prototype



By prototyping a dashboard that would be used by our stakeholder, we offer an interface where they can interact with our model and tangibly see results, all without ever needing to dive into the data science.

We handpicked some metrics and visualizations to give the user a general understanding of what data the model relies on to make decisions, with **feature importance** being especially useful to provide interpretability.