DEMAND FORECASTING

& Inventory Management

Presented by Emma Le Bars and Ivan Chertov

A Data-Driven Approach to Optimize Warehouse Space and Stocking

Agenda

- Business Context & Objectives
- Data Exploration
- Prediction Models
- Forecasting & Results
- Inventory Rebalancing Strategy
- Key Insights & Takeaways

Business Context

Objective

Accurate demand forecasts

Business Importance

Effective stock management

Primary Goal

Develope a forecasting model with machine learning

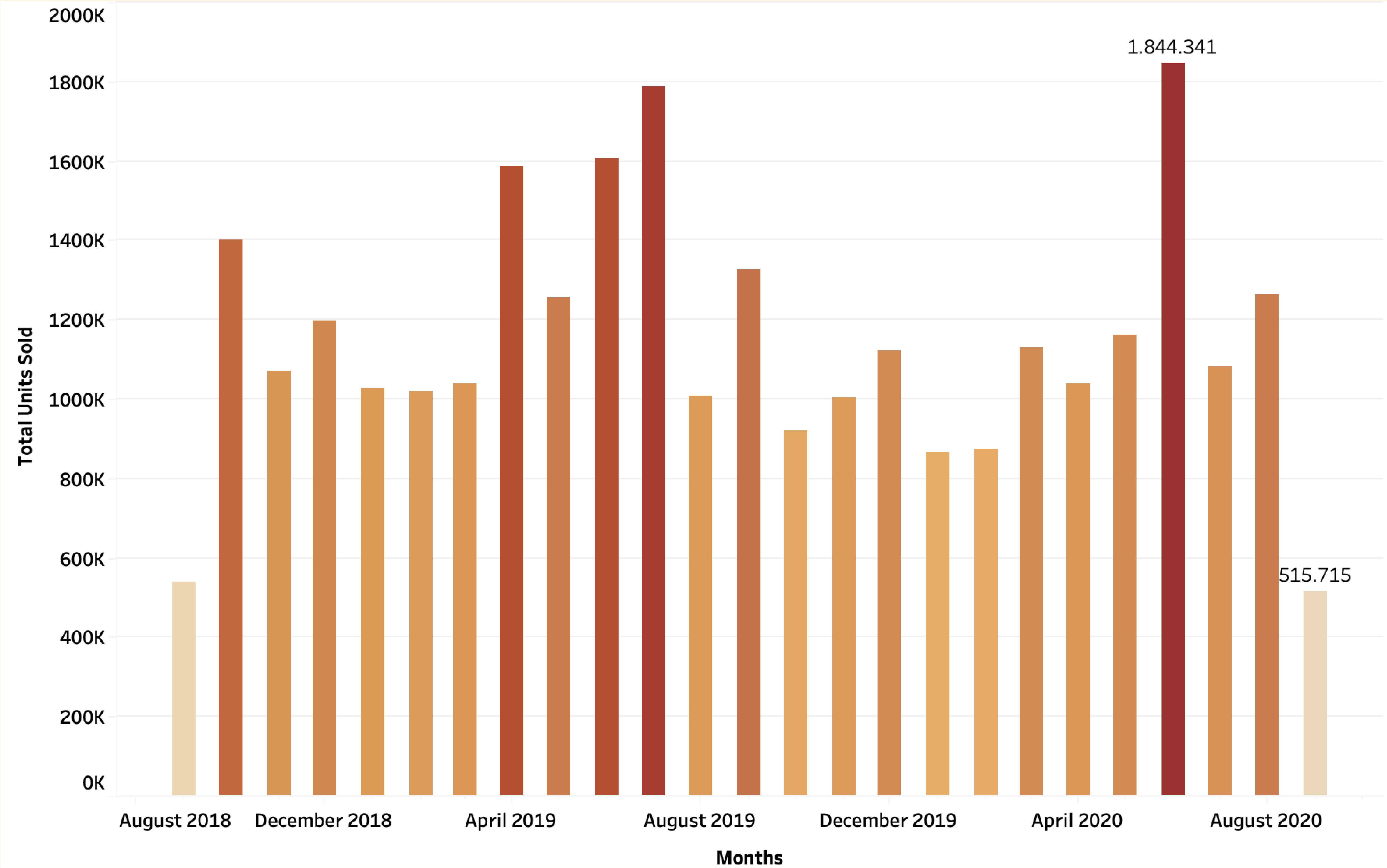
Data Exploration

Dataset:

- Transactions data from H&M's online sales
- September 2018 to September 2020
- Data on over 30 million transactions
- Data on 104.000 articles

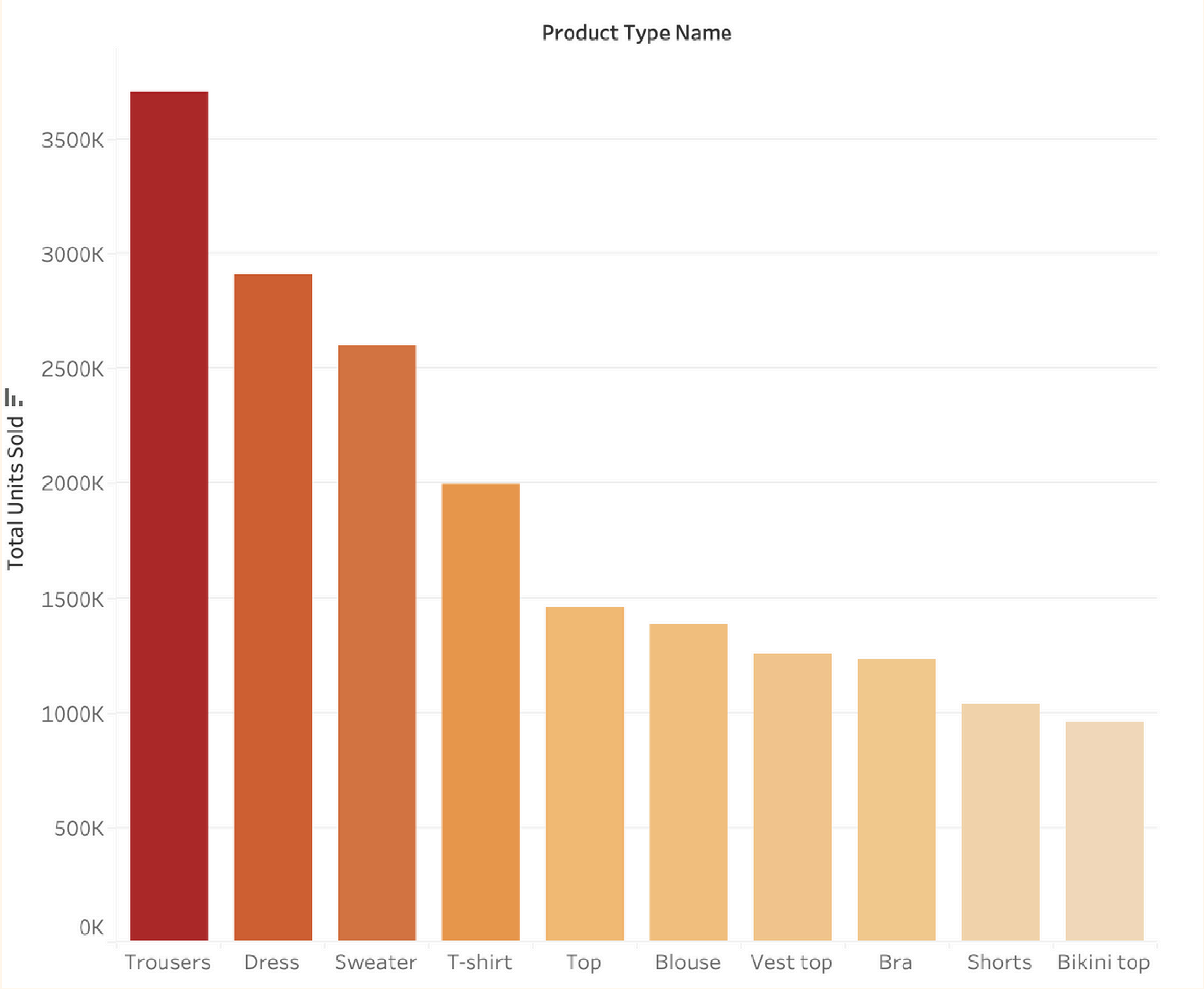
Key Attributes: Product type, color, weekly sales (total units sold), average price, and customer counts.

Sales over time



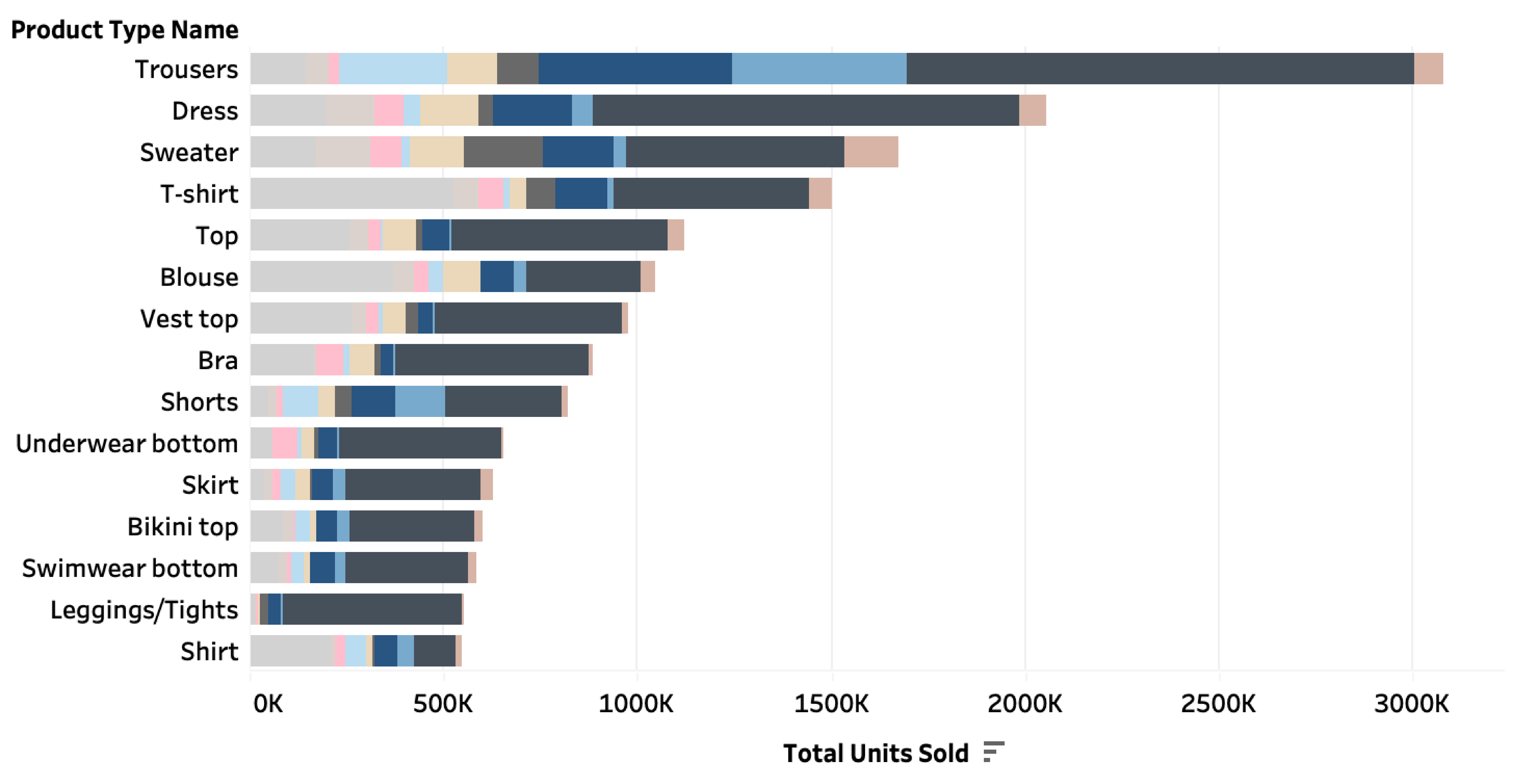
Top Sellers

Dress, Sweaters and Trousers are the most sold categories.

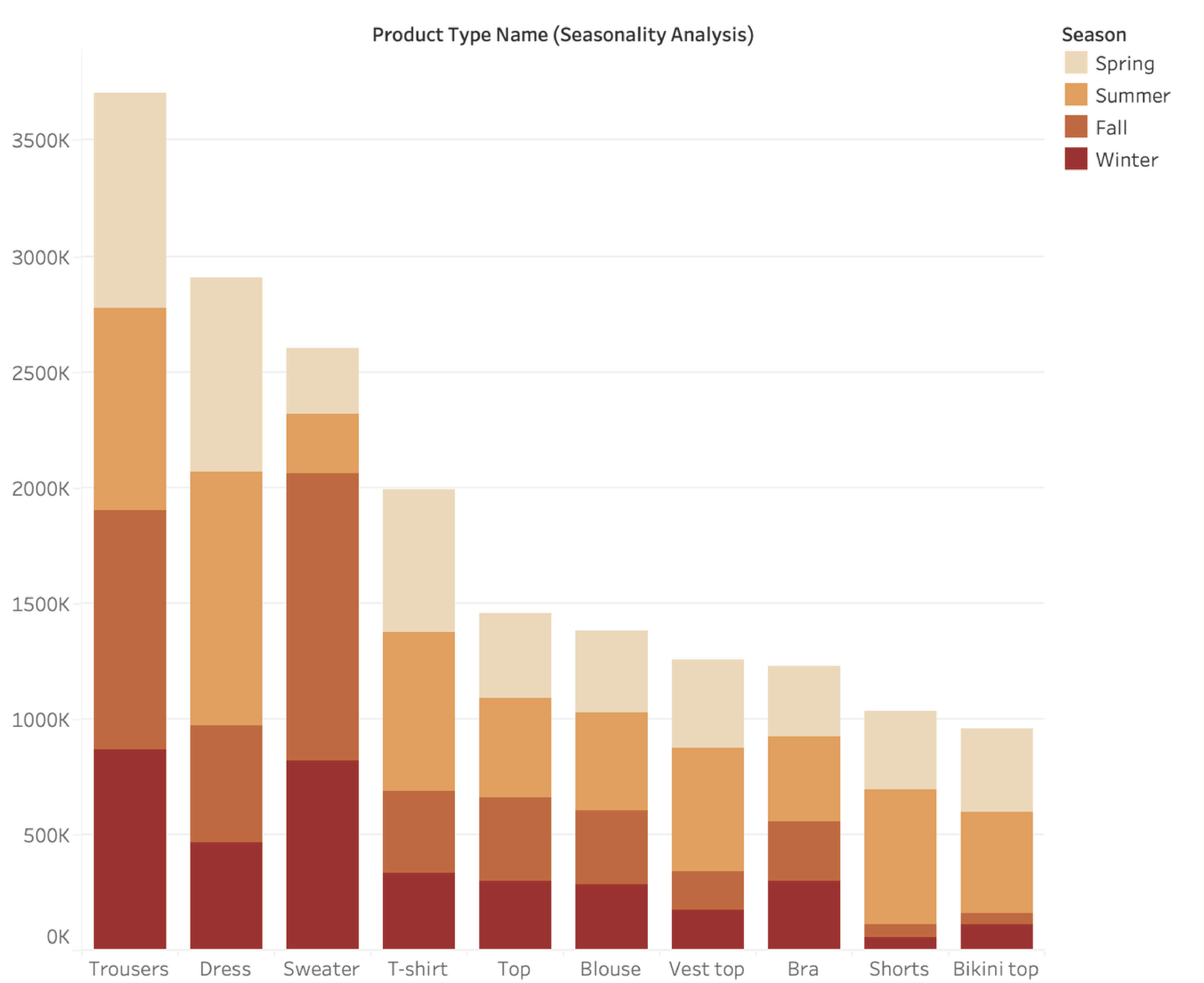


Sales based on color

The most sold colors are black and blue!

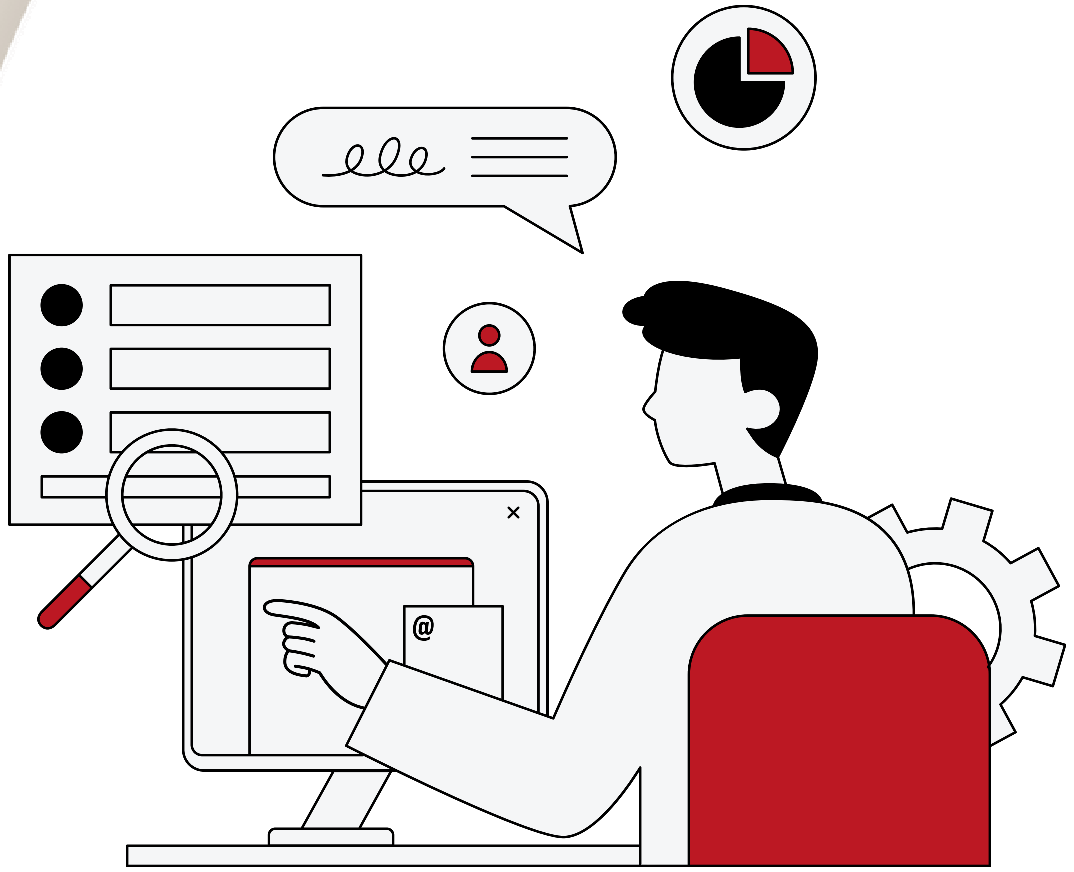


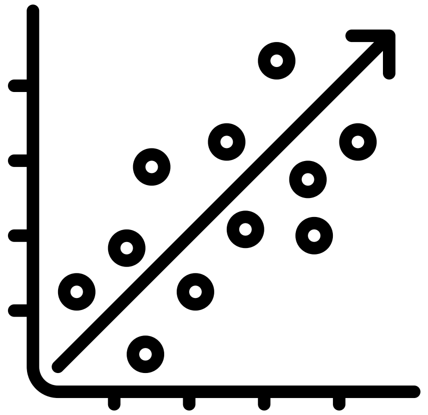
Seasonality

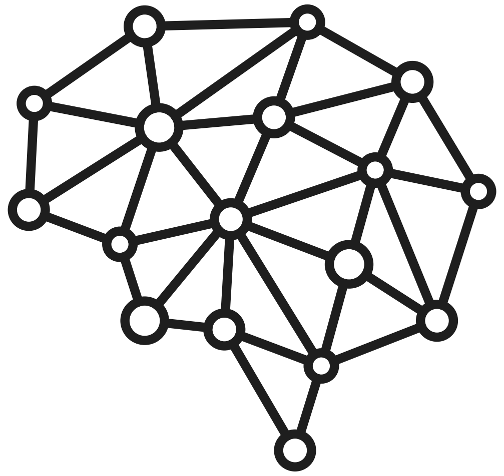


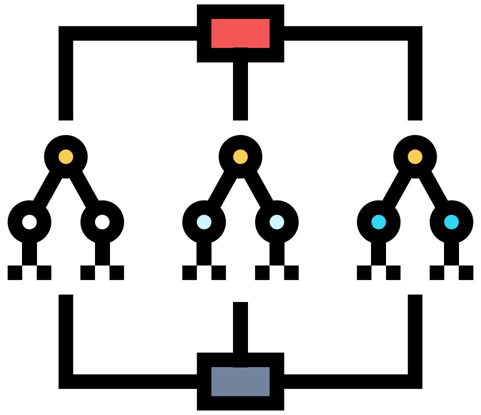
Prediction Model

Creating a forecast!



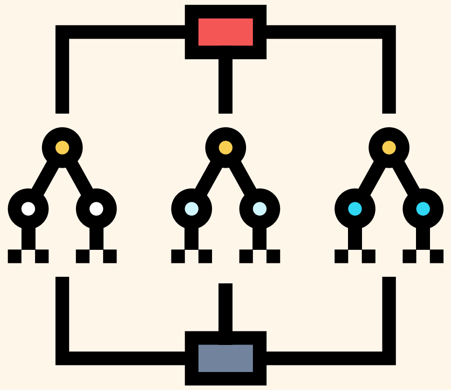






Exploration

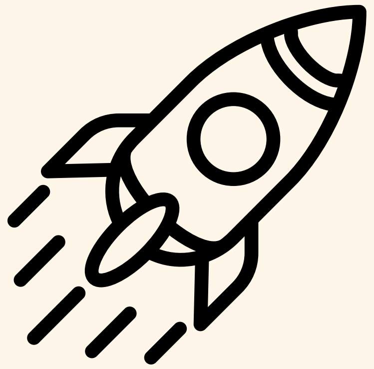
01



Random Forest

Too inacurate

02



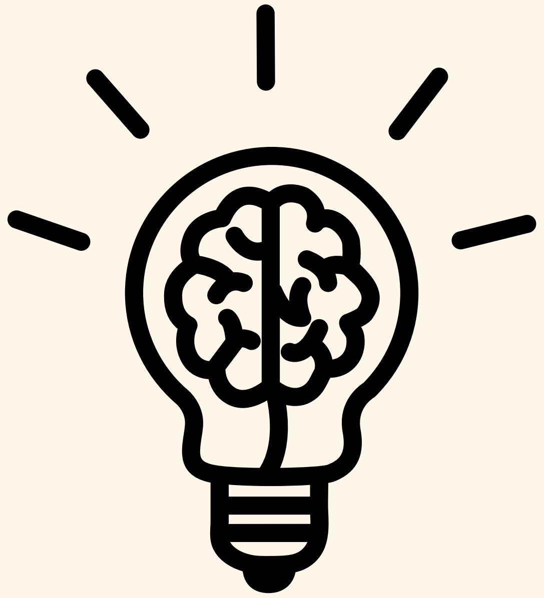
XGBoost

CPU heavy

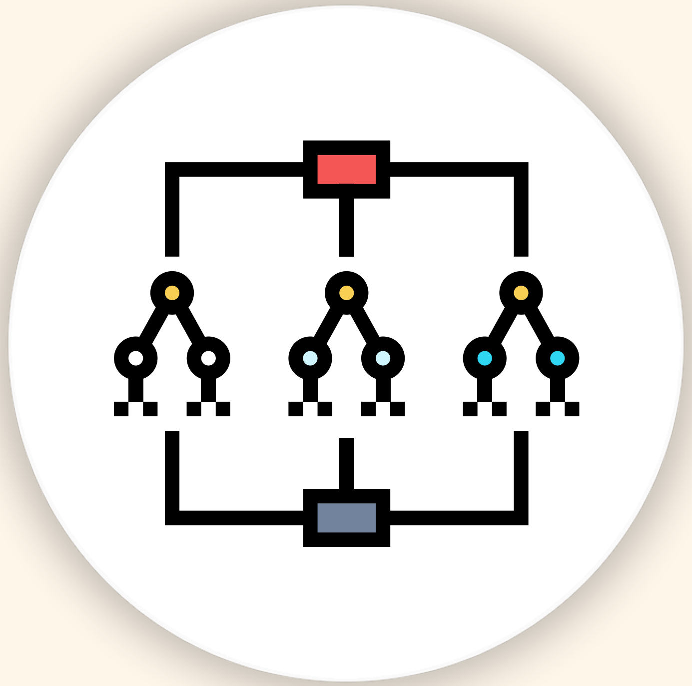
03

Random Forest, XGBoost & PyTorch

Too CPU heavy

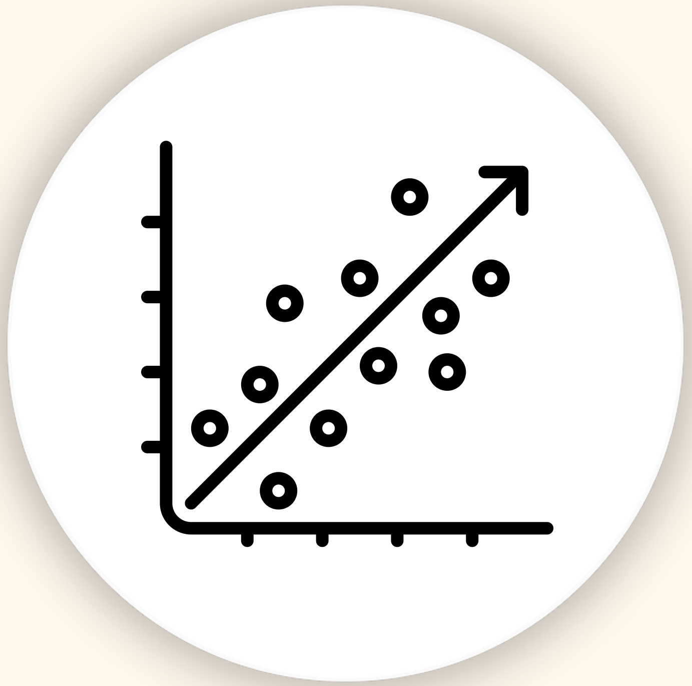


Ensemble Model



Random Forest Regressor

Captures complex interactions in the data



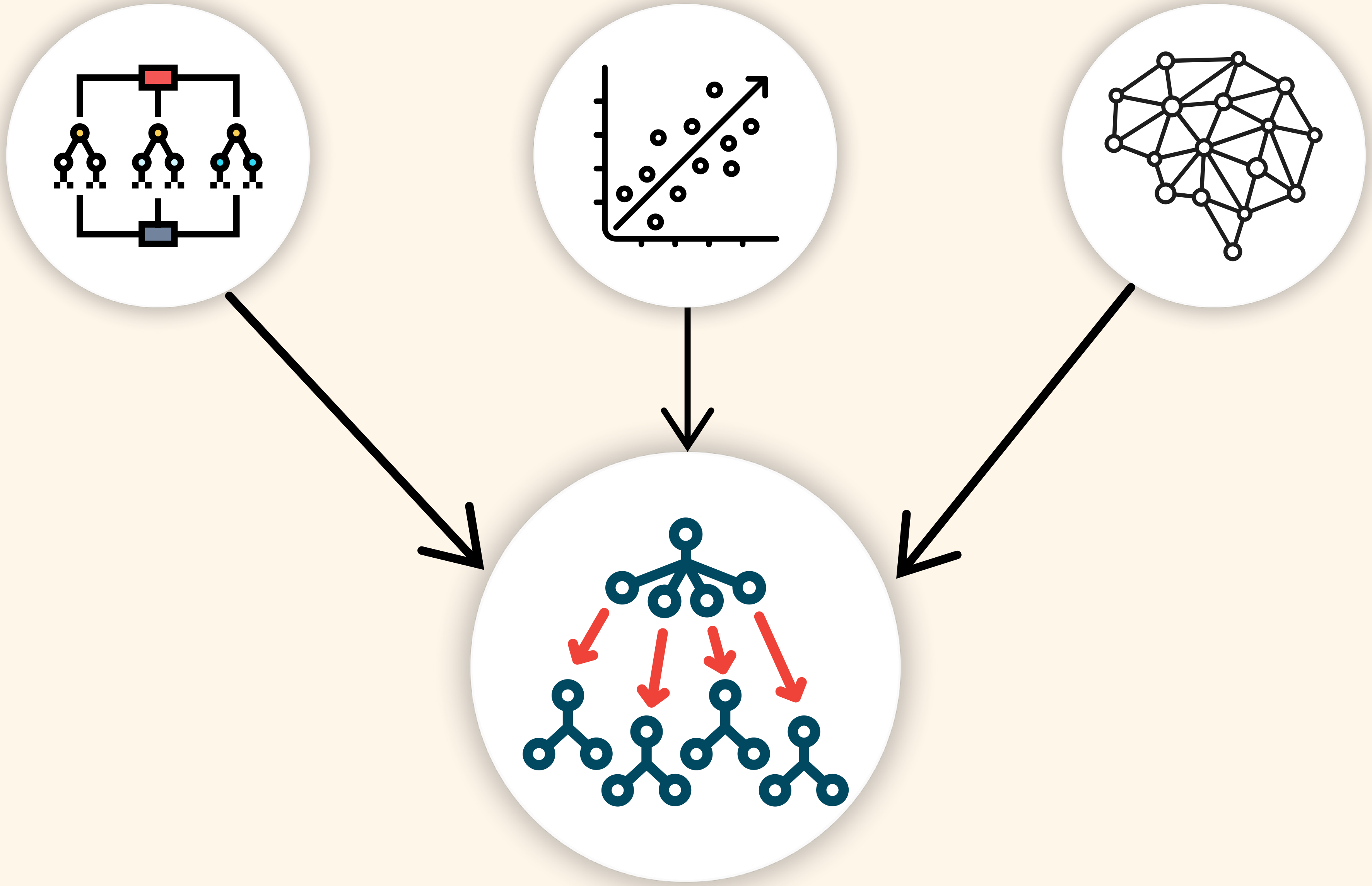
Linear Regression

Provides a baseline trend analysis

MLP Regressor (Neural Network)

Handles complex non-linear patterns

Meta-Model



Random Forest Regressor

Data Splitting for Model Training and Testing

Training Set
Data before
Jan 2020

Validation Set
Jan 2020

Test Set
Feb 2020

Forecasting
March 2020

No Covid
Data

Model Training Results

February 2020

Predicted Sales
881 000

Actual Sales
875 000

MAE 24

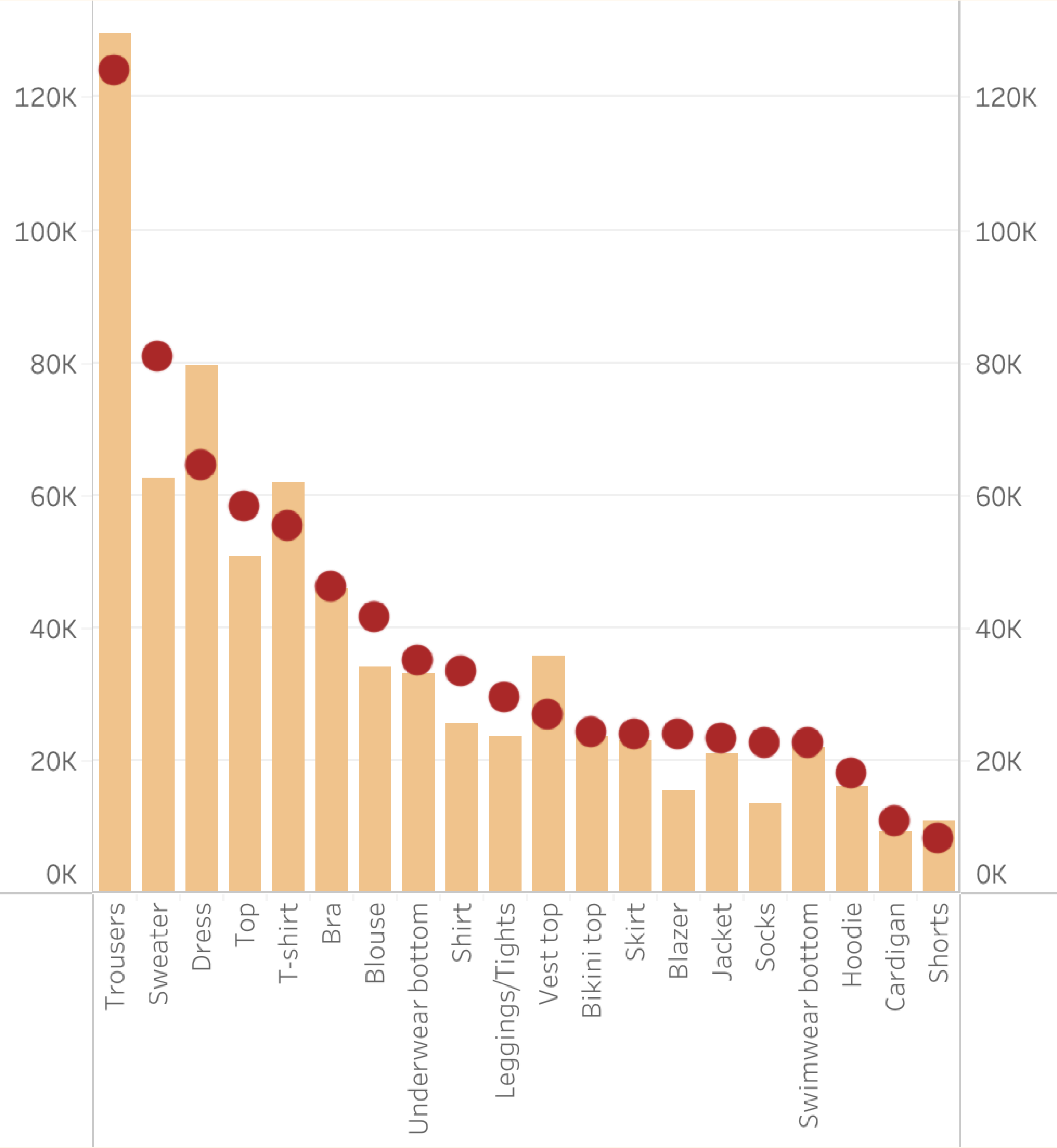
RMSE 76

Overall Positive Variance
0.68%

Forecating Results

Forecast vs. Actual (March)

- Red dots: Model's predicted sales.
- Orange bars: Actual sales data for March.
- Close alignment highlights forecast accuracy.
- Effective tool for inventory optimization and planning.
- High accuracy in high-volume products; opportunities to refine for lower-volume items.



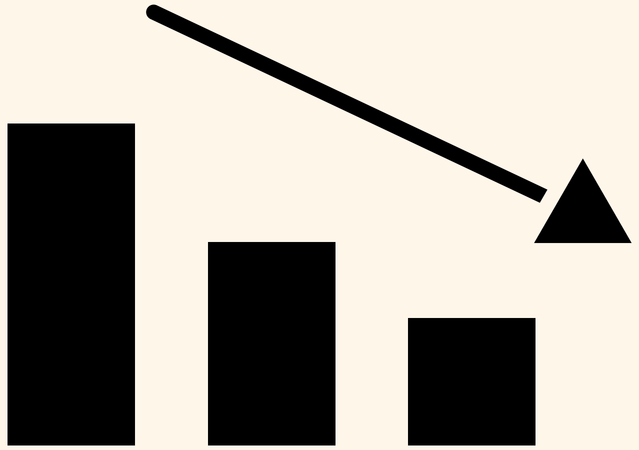
Inventory Rebalancing Strategy

Objective

- Free warehouse space by removing low-demand items
- prioritizing high-demand products

Rebalancing Mechanism

Threshold



Non-Peak Items

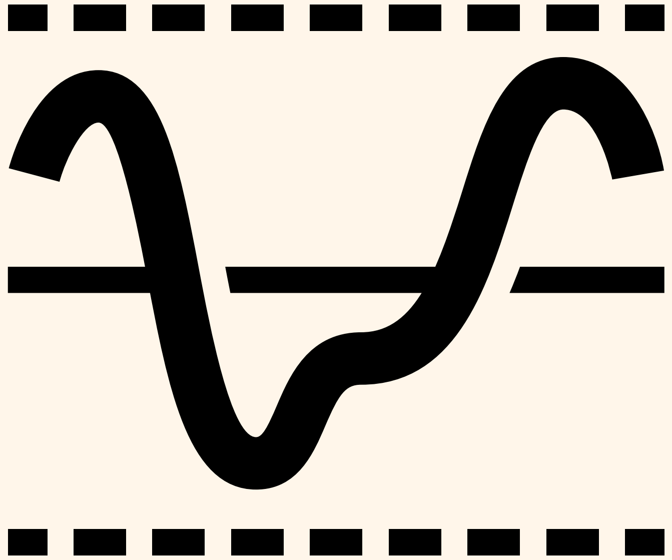
quick rebalancing

50% avg. demand

Peak Items

retain stock

100% avg. demand



Dynamic

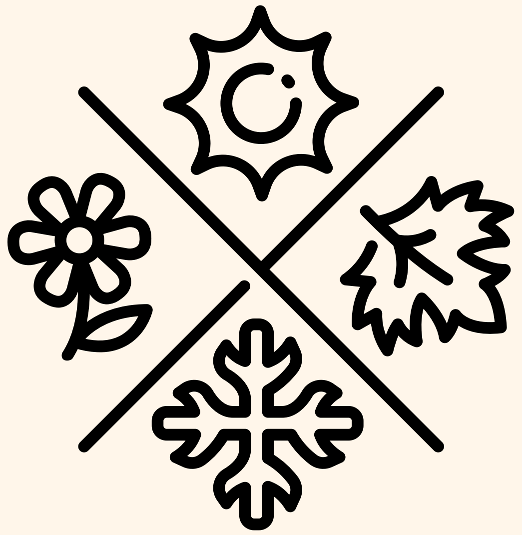
rolling demand

over a 4-week window

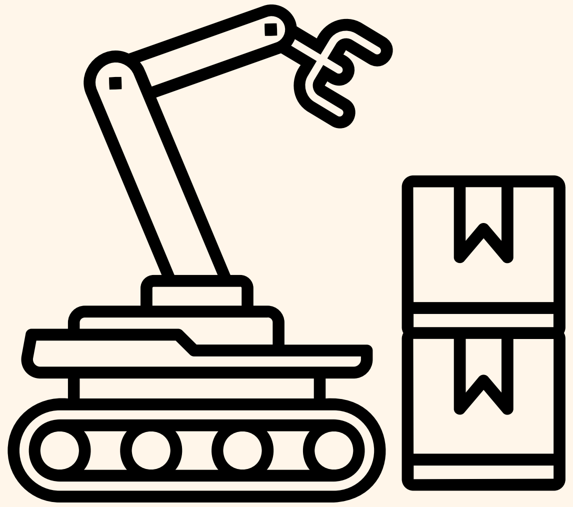
Rebalance Dashboard

- **Swimwear Top**:
  - First 3 Weeks - No stock increase needed
  - Sales expected to rise in the final week
- **Moccasins**:
  - Flagged all four weeks

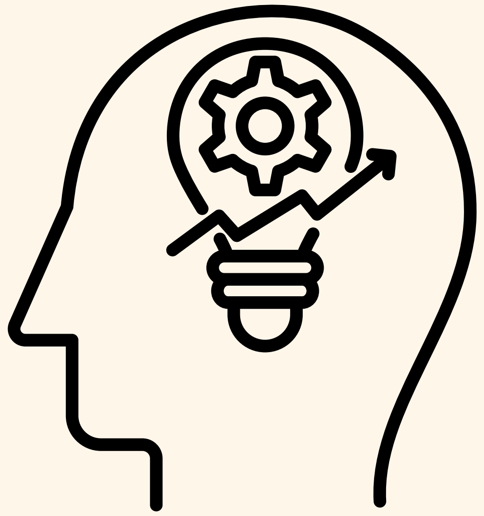
Key Insights



Seasonal Demand Sensitivity



Space Optimazation



Refinements & Future Enhancements

Seasonal Demand Sensitivity

- is_peak_season flag
- highly valuable in capturing demand fluctuations
- model responded dynamically to peak and off-peak seasons.

Space Optimization

- based on forecasts
- removing low-demand items
- create space for high-priority items aligning with H&M's stock management goals.

Next Steps

01 Further tuning of peak season thresholds and exploration of additional external factors, like marketing campaigns.

02 Integration of economic and promotional data to improve forecast precision.

Bonus

Example for future tuning

Model Training Single Articles

Black Trousers

average size of the errors

2.5

$\sqrt{\text{RMSE}}$

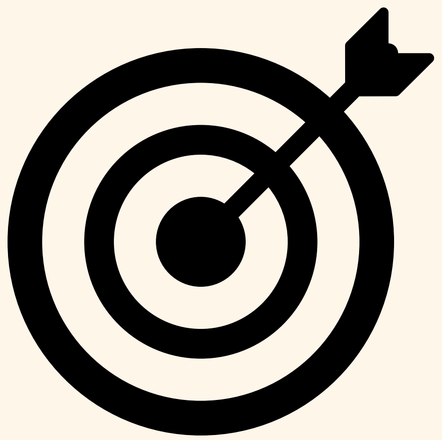
9.2

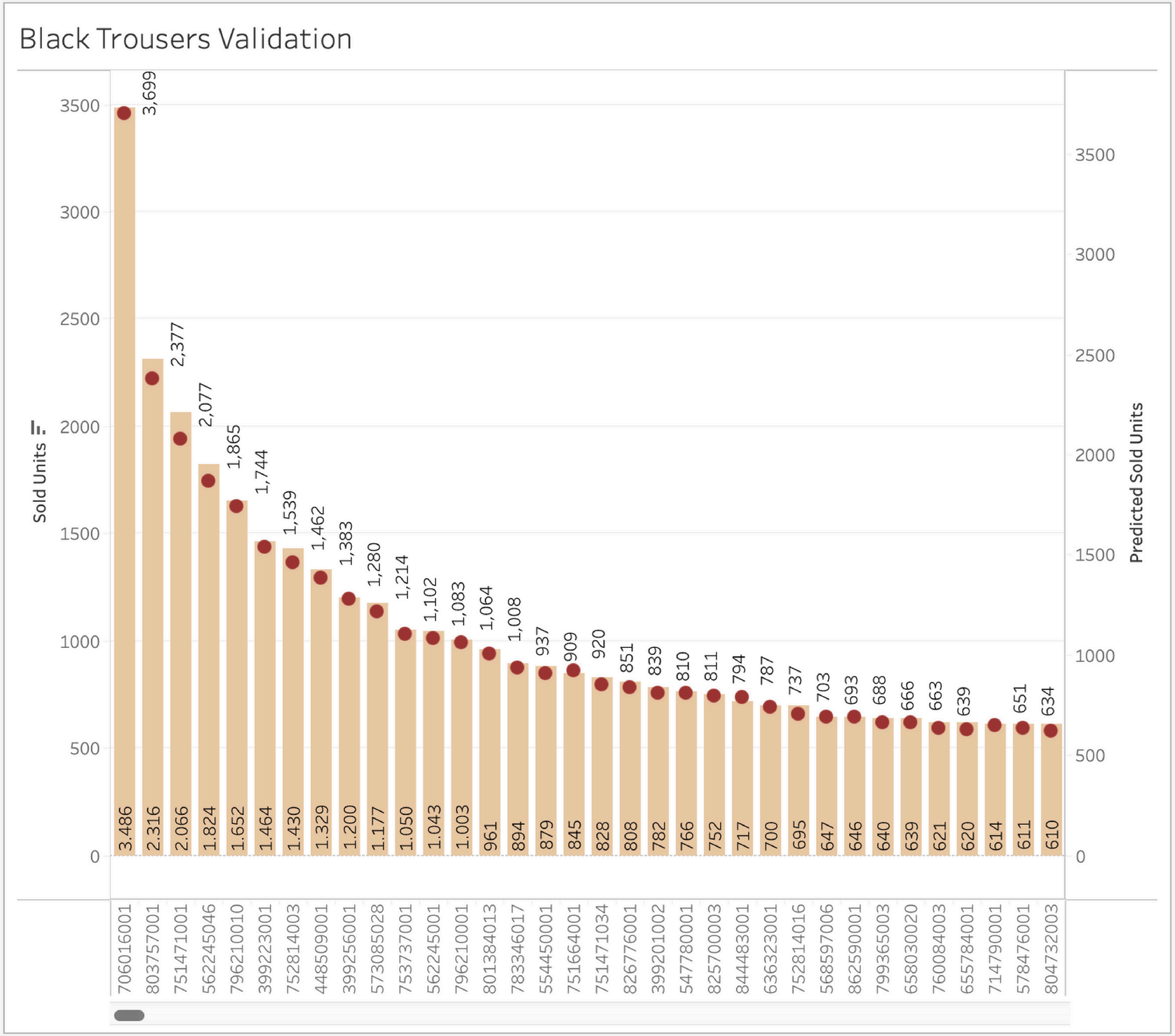
Predicted Sales
8442

Actual Sales
8712

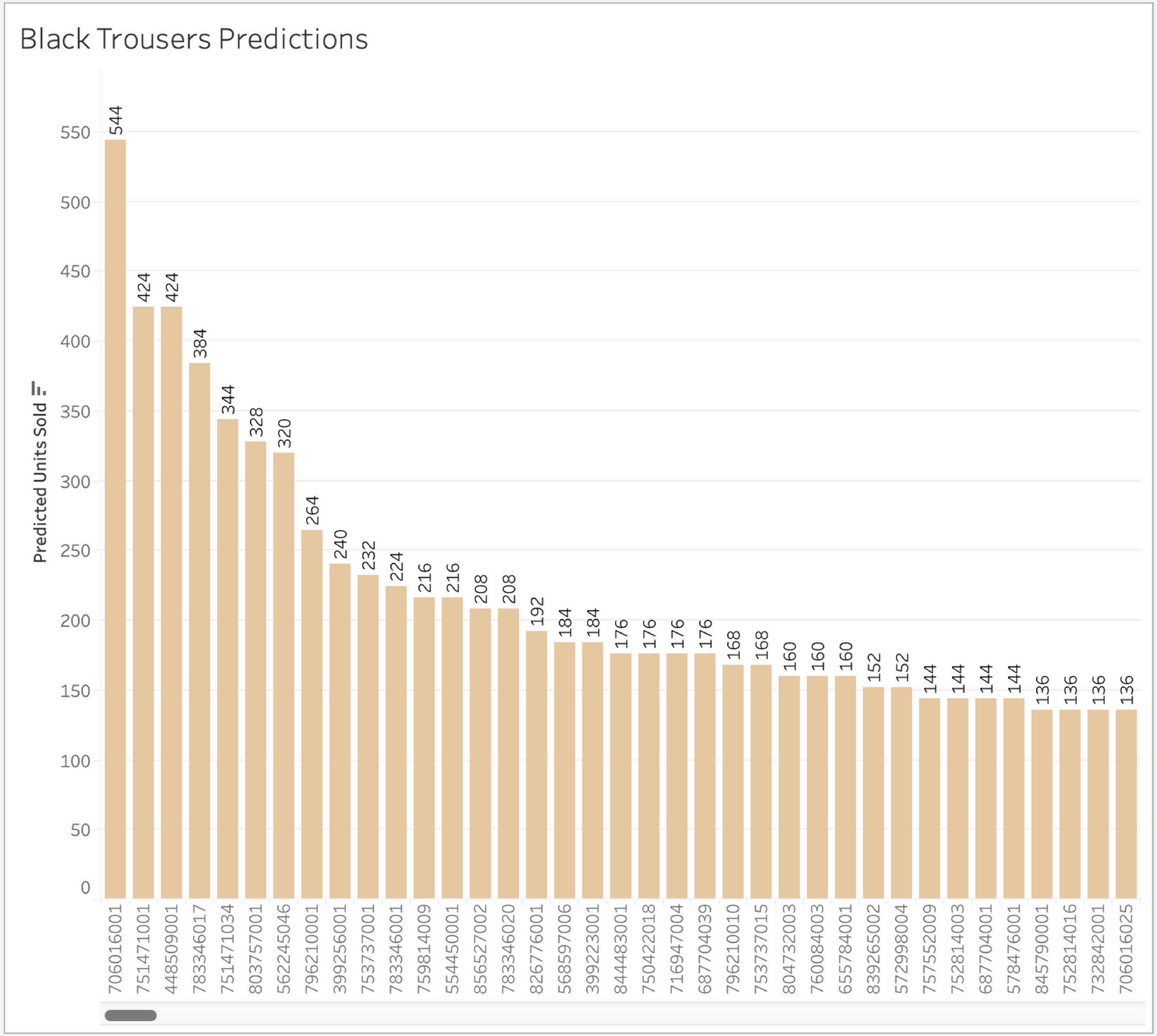
Overall Variance
3.2%

Model Training

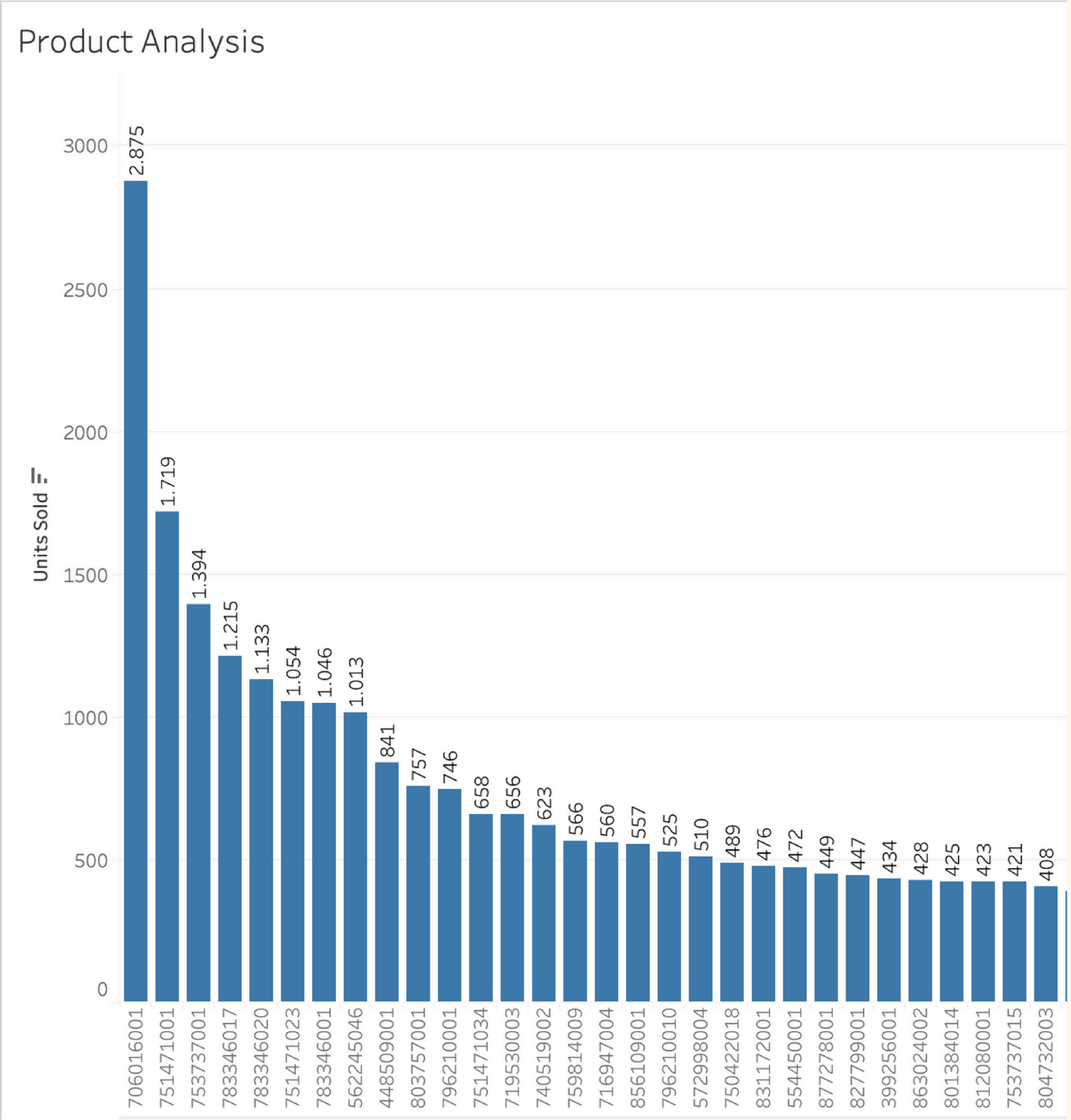




Forecast


Black Trousers Predictions
Predicted Units Sold
0
50
100
150
200
250
300
350
400
450
500
550
544
424
424
384
344
328
320
264
240
232
224
216
216
208
208
192
184
184
176
176
176
176
168
168
160
160
160
152
152
144
144
144
144
136
136
136
136
706016001
751471001
448509001
783346017
751471034
803757001
562245046
796210001
399256001
753737001
783346001
759814009
554450001
856527002
783346020
826776001
568597006
399223001
844483001
750422018
716947004
687704039
796210010
753737015
804732003
760084003
655784001
839265002
572998004
757552009
752814003
687704001
578476001
845790001
752814016
732842001
706016025

Actual Numbers



Thank you!

Find out more at :

https://github.com/Barstov-Industries/Capstone

Appendix

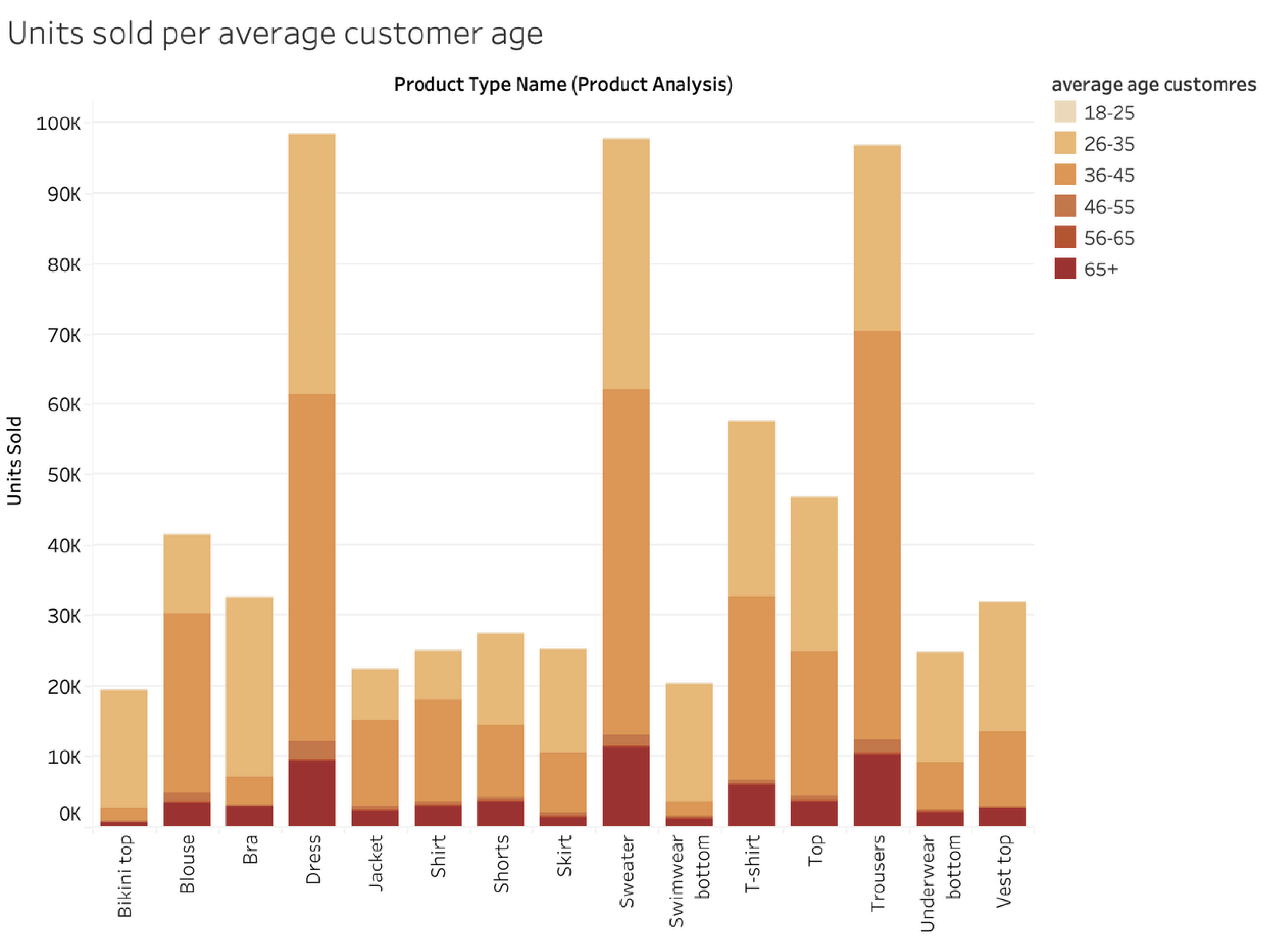
Demographics

Core customers group:

- 26 to 35

Second most presend group:

- 36 to 45

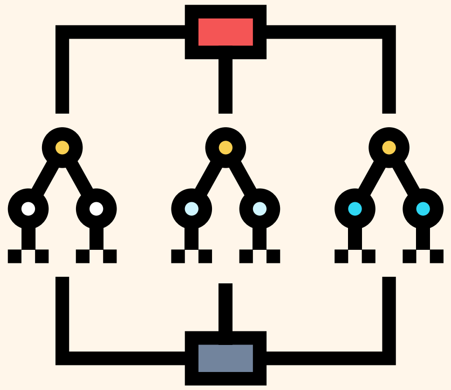


Sales based on color

Total Units sold by products and colors

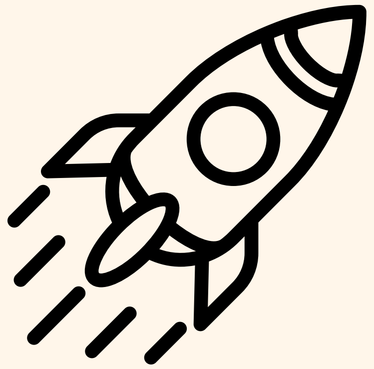
	Product Type Name (Product Analysis)									
Colour Grou..	Blouse	Bra	Dress	Shorts	Skirt	Sweater	T-shirt	Top	Trousers	Vest top
Beige	0,40%	0,20%	0,46%	0,14%	0,20%	1,68%	0,35%	0,42%	0,62%	0,13%
Black	2,44%	2,85%	9,36%	1,89%	2,53%	4,83%	3,03%	4,02%	7,69%	2,78%
Blue	0,22%	0,04%	0,47%	0,67%	0,21%	0,23%	0,14%	0,06%	3,04%	0,11%
Dark Blue	0,55%	0,24%	1,51%	0,88%	0,38%	1,64%	1,06%	0,64%	3,44%	0,39%
Grey	0,01%	0,16%	0,30%	0,23%	0,09%	1,76%	0,65%	0,23%	0,70%	0,16%
Light Beige	0,93%	0,24%	1,36%	0,37%	0,27%	1,20%	0,31%	0,78%	0,86%	0,29%
Light Pink	0,27%	0,53%	0,67%	0,15%	0,19%	0,94%	0,63%	0,28%	0,18%	0,24%
Off White	0,38%	0,04%	1,08%	0,17%	0,21%	1,42%	0,39%	0,39%	0,53%	0,25%
Red	0,26%	0,41%	0,92%	0,05%	0,13%	0,81%	0,34%	0,30%	0,28%	0,20%
White	2,20%	1,11%	1,66%	0,46%	0,18%	1,78%	3,16%	1,82%	0,76%	1,36%

First Steps

01 

Random Forest

RMSE 400+

02 

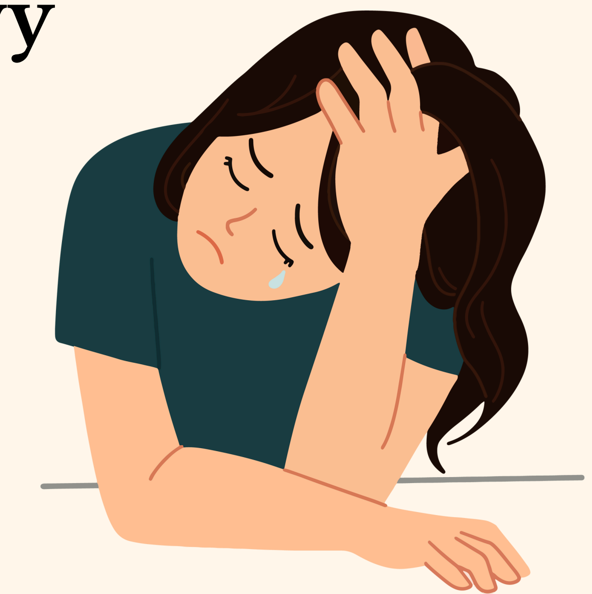
XGBoost

Never Ending Story

03 

Random Forest, XGBoost & PyTorch

Too CPU heavy



Feature Engineering

Lagged Sales

Seasonality

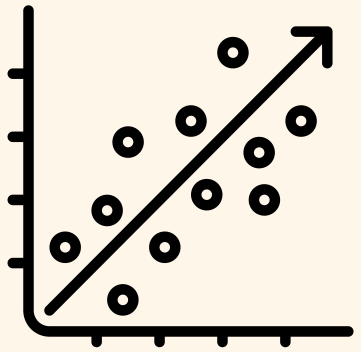
Price Scaling

```
# Define feature columns
feature_columns = ['product_type_no', 'colour_group_code', 'average_price',
                   'lag_units_sold_1week', 'lag_units_sold_2weeks', 'month', 'is_peak_season']
```

Model Training Results

February 2020

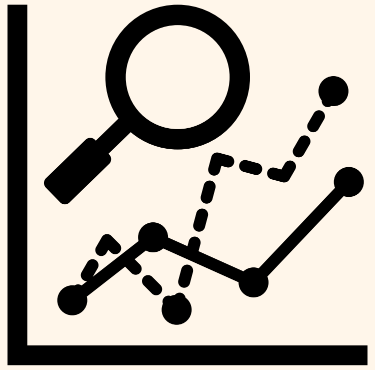
average size of the errors



24

RMSE

76



Predicted Sales
881 000

Actual Sales
875 000

Overall Positive Variance
0.68%