Dynamic Order Routing

Connor Underwood, Ethan Empaces, Sania Kapoor, Stephen Kruse, Bhavya Venkataraghavan

Purpose Statement

Finish Line, a subsidiary of JD Sports, is a fashion retailer that specializes in providing athleisure and footwear across the world. This project aims to build a model that predicts expected delivery time of customer orders in various scenarios to assist JD Finish Line in optimizing their order routing within the U.S.

Objectives

- Compute the most advantageous locations within the business to fulfill an order
- Be cost efficient to both the business and the customer
- Be Dynamic: provide order routing options in the case of natural disasters
- Leverage Learnings to develop a demand model for specific products and their expected ROI in various scenarios

Key Business Questions

- What is the best way to fulfill an order?
- How do we decide which store to route from without breaking size runs?
- What products tend to sell the best at specific locations?

Tools

- RStudio & Power BI - data visualization
- PySpark (Python API) - parallel processing framework for big data queries, machine learning. Over 1 TB of data!

Exploratory Data Analysis

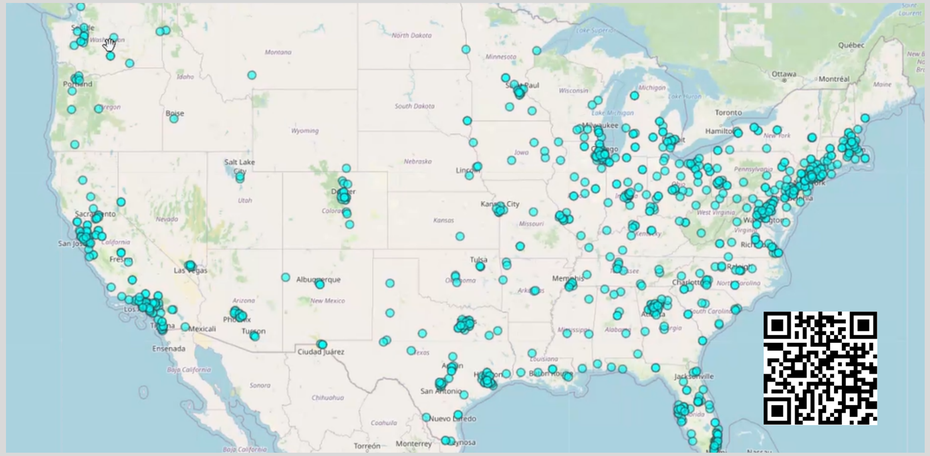
What products tend to sell the best at specific locations?

Table 1

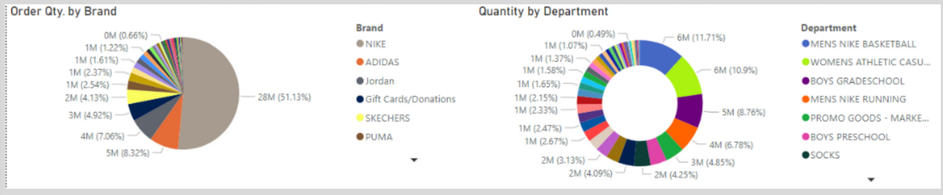
OCCURRENCES	NAME	SIZE
2618	Men's Nike Sportswear Tech Fleece Taped Full-Zip Hoodie	MED
2298	Unisex Crocs Classic Clog Shoes (Men's Sizing)	7.0
2194	Unisex Crocs Classic Clog Shoes (Men's Sizing)	6.0
2190	Men's Nike Sportswear Tech Fleece Taped Full-Zip Hoodie	MED
2014	adidas Yeezy Slide Sandals	10.0

Table 1: Extracted information about a sample store's top selling items as well as their respective sizes and sales (occurrences).

Interactive RStudio map of every U.S. JD Finish Line location that displays the findings from Table 1 but for each unique store.



Pie charts of order quantity by brand and department, respectively, that help us conceptualize which brands and product departments sell the best.



Post EDA Goal: Implement an algorithm to select the optimal store(s) to ship a given order from

Model Implementation

How do we decide which store to route an order from?

Preprocessing Phase

- Dirty Data! - Was Not Initially Machine Learning Compatible
- One Hot Encoding: technique to represent categorical features in a numeric format, for machine learning capabilities
  - *Encoded shipping carriers & U.S. states within the data
- Feature Integration: added other important features not already present within the data - lead time, days late (of order), order distance (in miles)

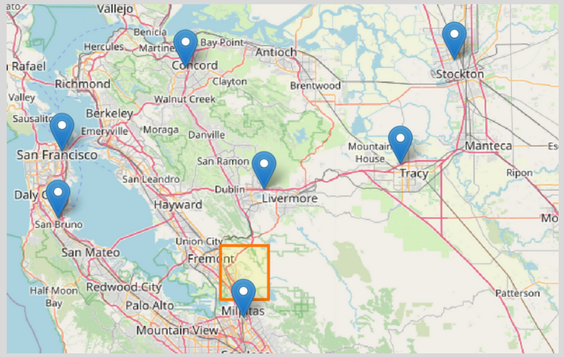
Model Brainstorming

- Performed a K-means clustering algorithm on our model data
- Trimmed Mean Approach: Prioritize based on lowest average lead time

Coordinates in 3D-space that indicates the center of each cluster

```
Cluster 0 centroid: [0.9329375  0.62333237 1.66468112]
Cluster 1 centroid: [1.3184352  2.77474681 3.29118706]
Cluster 2 centroid: [2.68763497 0.61660386 1.89345745]
```

Interactive map output of the Top 10 stores for Zip Code 94539 sorted by lead time



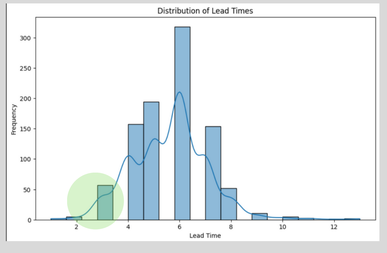
Future Goals

Current Capabilities

- Given any U.S. zip code, the model can output optimal stores based on user-defined metrics (lead time, distance, etc.)
- List the lead time history of any store-zip code pair

Future Capabilities

- Apply weights on certain metrics for increased accuracy
- Incorporate extra metrics such as product information



Conclusion

- Utilizing the provided data set, extracted insights regarding top selling items, their corresponding sizes, and sales frequencies for each U.S. store.
- Identified optimal fulfillment locations based on implemented machine learning techniques such as K- means clustering, including factors such as proximity to customers and available inventory

Acknowledgements

Thank you to Paice Fuller, Briana Pedersen, Matthew Market, 2024 JD Finish Line Team, & Data Mine Staff