FINAL PROJECT:

PREDICTIVE ANALYSIS ON CANADIAN DRUG SHORTAGES

BY: ANGELO REYES

TOPICS FOR DISCUSSION

01. Introduction

02. Motivation for this Project

03. Dataset

04. Exploratory Data Analysis

05. Model Building

06. Results

07. Conclusion

MOTIVATION FOR THIS PROJECT

This project aims to address the significant issue of drug shortages in Canada, where over 20,573 unique drugs were reported unavailable as of November 2023, according to Drug Shortages Canada. As a Pharmacy technician, I experience firsthand the challenges of managing medication supplies and finding alternatives, which critically impact patient well-being. By analyzing factors such as supply chain dynamics, manufacturing disruptions, regulatory changes, and demand fluctuations, this project will utilize machine learning techniques to predict drug shortages. The goal is to iteratively train and test supervised learning models to develop a robust solution that helps mitigate this ongoing problem.

Shortage reports	**20573**
Actual shortage:	**1859 (9%)**
Anticipated shortage:	**104 (1%)**
Avoided shortage:	**578 (3%)**
Resolved:	**18032 (88%)**

DATASET

Our dataset is derived from drugshortagescanada.ca, the official site for reporting and accessing data on drug shortages and discontinuations in Canada. This comprehensive database is maintained by market authorization holders and is publicly accessible.

Report ID

Drug Identificaion Number

Report Type

Brand Name

Company Name

Common Name

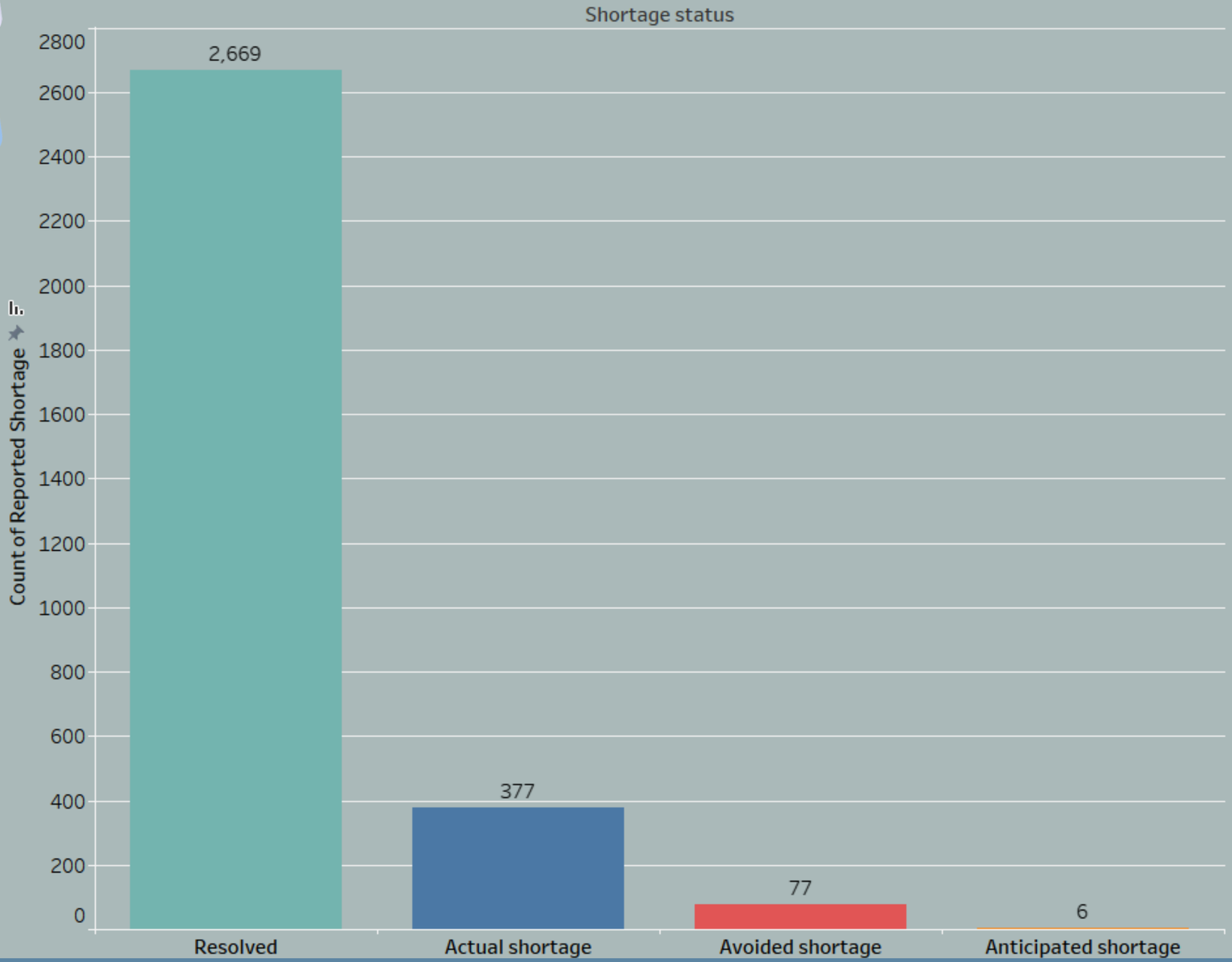
Shortage Status

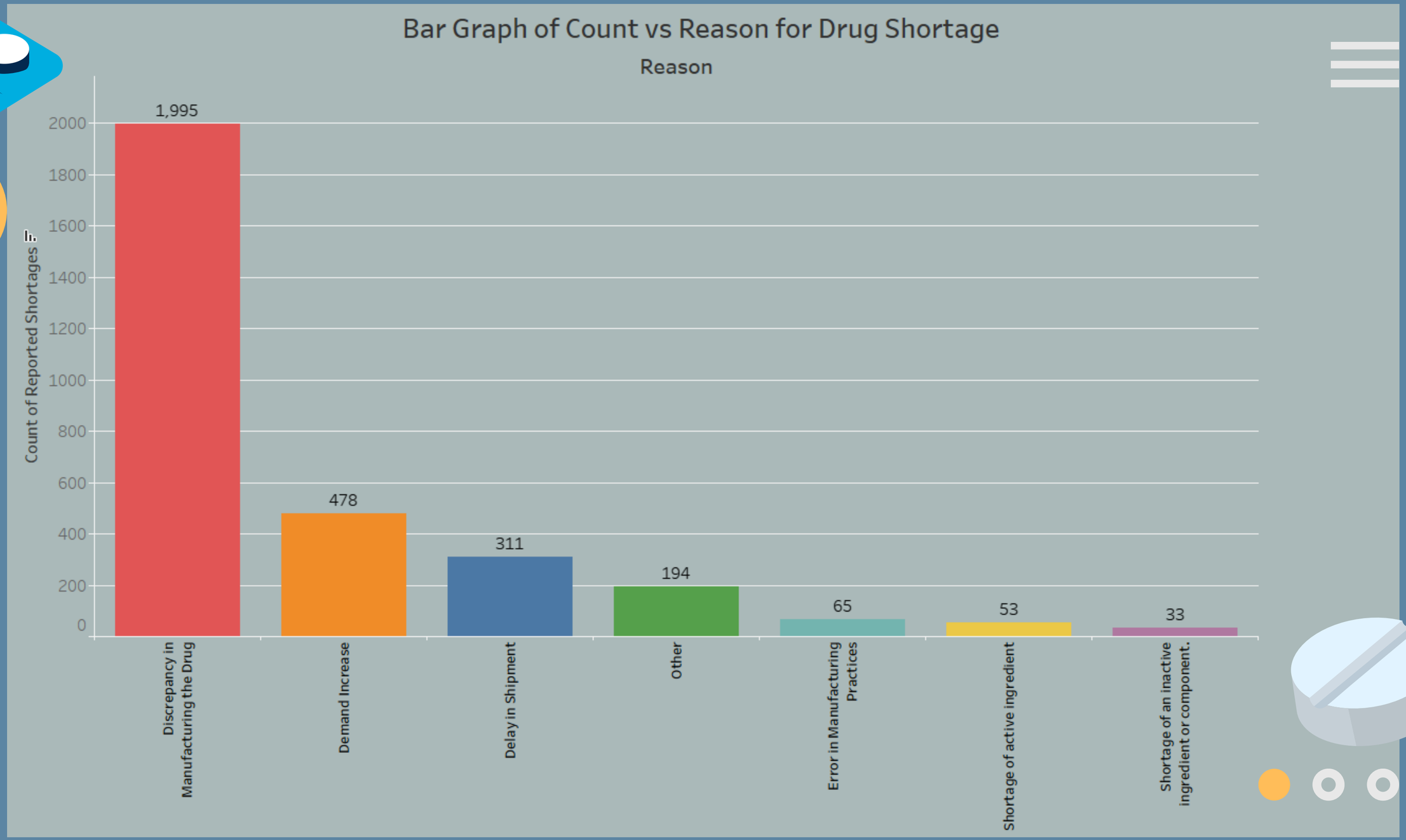
Actual Start Date

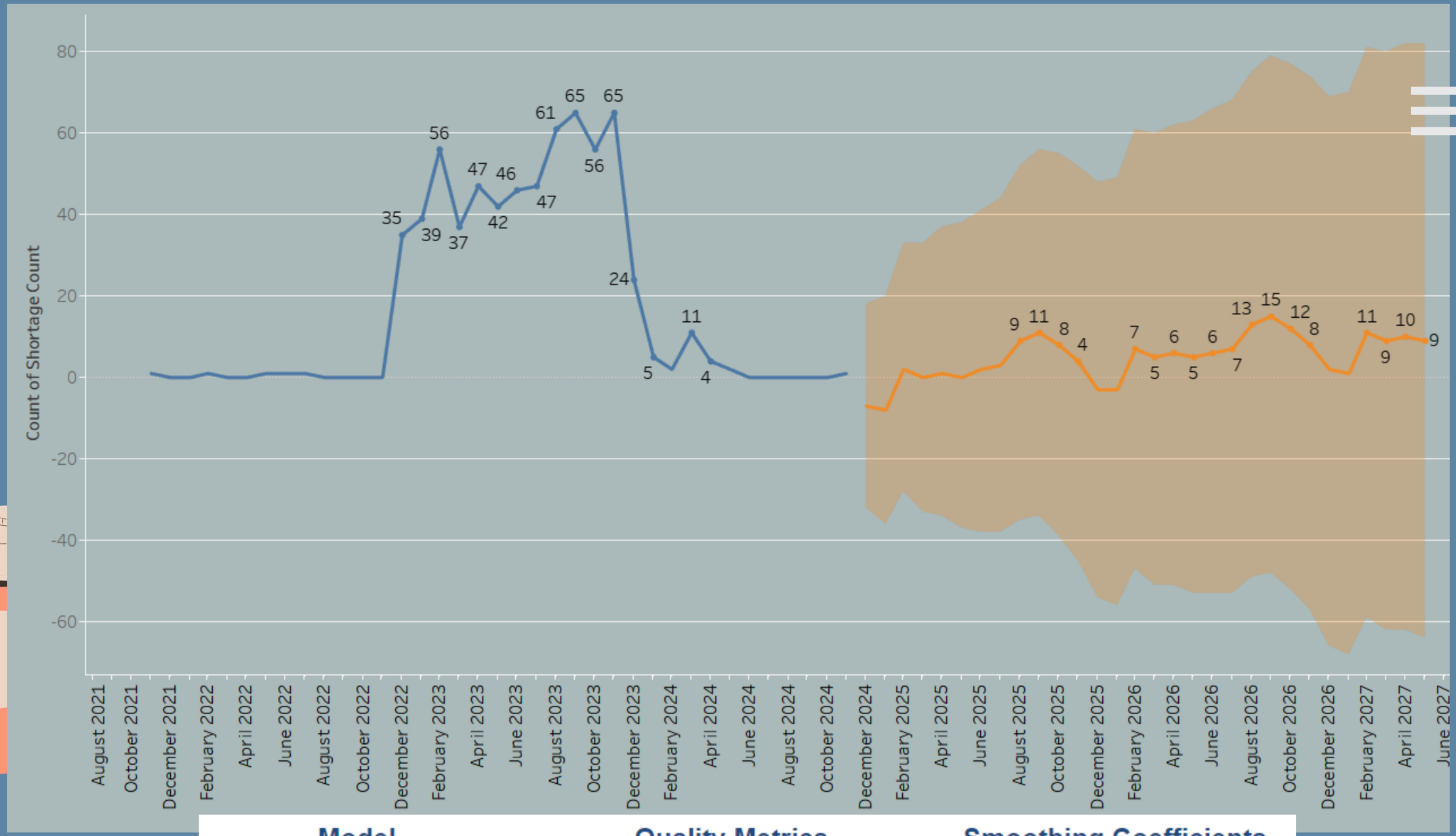
Reason

EXPLORATORY DATA ANALYSIS

Bar Graph of Reported Shortage vs Shortage Status




Bar Graph of Count vs Reason for Drug Shortage
Reason
Count of Reported Shortages
2000
1800
1600
1400
1200
1000
800
600
400
200
0
1,995
478
311
194
65
53
33
Discrepancy in Manufacturing the Drug
Demand Increase
Delay in Shipment
Other
Error in Manufacturing Practices
Shortage of active ingredient
Shortage of an inactive ingredient or component.



Model			Quality Metrics					Smoothing Coefficients		
Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Additive	None	Additive	15	11	0.23	204.7%	232	0.500	0.000	0.339

MODEL BUILDING

PREPROCESSING

In this step we identified and handled any missing values. Furthermore, I used one hot encoding to convert categorical values into features that will be used by my ML algorithm.

FEATURE ENGINEERING

Extracted features from date columns such as day of the week, month, and year and created seasonal trend features. Since my data contained imbalances, the use of SMOTE was applied.

MODEL TRAINING

Ensemble learning methods was used such as, Random Forest Classifier and Gradient Boosting Classifier. Then was compared with one another to see which model performed better.

HYPERPARAMETER TUNING

The choice of a tuning method was Randomized Search Cross Validation from Scikit-learn. The choice of using this compared to Grid Search was due to it being generally faster.

MODEL BUILDING

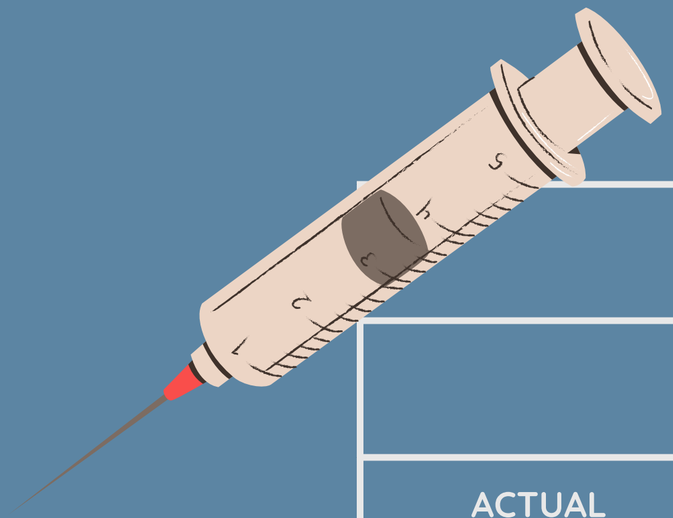
MODEL TESTING

After tuning the hyperparameters and selecting the best model configuration,. application of this finalized model was tested on unseen data.

MODEL EVALUATION

Evaluated the model's performance using several key metrics that we will discuss in the upcoming slides.

MODEL METRICS



RANDOM FOREST MODEL METRICS:

	PRECISION	RECALL	F1-SCORE	SUPPORT
ACTUAL SHORTAGE	0.93	0.64	0.76	89
ANTICIPATED SHORTAGE	0.00	0.00	0.00	1
AVOIDED SHORTAGE	0.67	0.22	0.33	18
RESOLVED	0.94	0.99	0.96	675
ACCURACY: 0.9349, PRECISION: 0.9294, RECALL: 0.9349, F1-SCORE: 0.9257				

MODEL EVALUATION

- Both models shows promise in managing and predicting drug shortages, which is critical for the pharmacy industry to ensure reliable patient care and efficient drug distribution.

- Overall since the recall scores of Random Forest Model is slightly higher. This became the model of choice.
  - Since Recall scores considers to be the ability of the model to find all actual drug shortages. This became the basis of what we were looking for in the model.
-
- Accuracy, Precision and F1-Scores were around 93% indicating on a somewhat promising model performance.

CHALLENGES

API limitations. Opted to just export dataset from website directly.

Class imbalances. Had to try to balance the minority classes.

Loading hyperparameter tuning techniques.

FUTURE GOALS

Use different classification models to compare with the model chosen.

Build a predictive model that would accurately anticipate drug or active ingredients susceptible to shortages.

Create and launch an interactive app for business integration.

THANK YOU!

(403) 397-5104

angelo.reyes10@outlook.com

www.linkedin.com/in/angelo-reyes-02988413a/