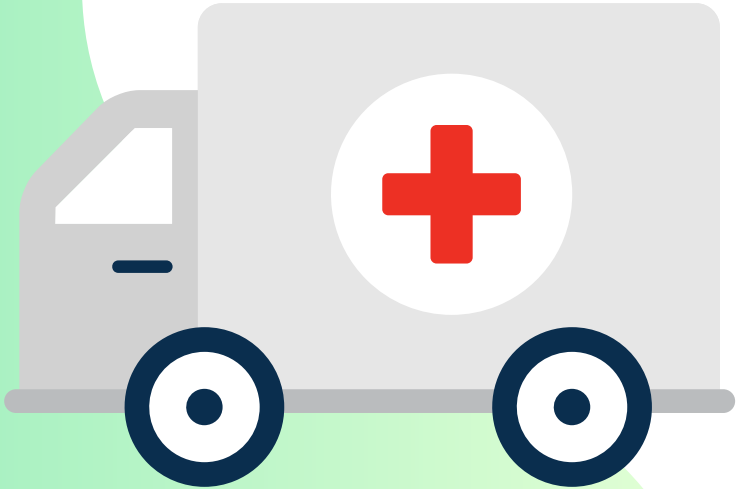
Kharagpur Data Science Hackathon

Team Name - Harvey Mike

Institute - IIT Kharagpur

Table of Contents



01

Introduction

Problem Statement

Task - 1 (Classification Type)

Objective: Forecast the HCP's optimal channel among the three promotional channels of Sales Rep calls, emails, and Sample Drops for the coming week (first week of Feb'20)

Constraint:
1. Limit on calls in last 12 months according to HCP segment
2. Only 25% HCPs can be reached through sample drop

Task - 2 (Regression Type)

Objective: Predict the Brand Rx (prescriptions) for the coming week (first week of Feb'20) generated through the optimized channel.

Constraint:
1. Overall predicted Brand Rx (across all HCPs) should not deviate more than 10% from the immediate previous week

HCP Data Description

Physician ID

Unique ID of the physician/HCP

Brand Rx

Prescriptions for the brand medicine written by the HCP during that week

Market Rx

Prescriptions for all Skin Disease market medications written that week by HCPs

Sales Rep Calls

Whether the HCP received a call from the salesperson during that week

Samples Dropped

Number of Brand drug samples received by the HCP in that week

Emails Delivered

Number of emails delivered to the HCP through campaigns in that week

Physician Segment

A segment assigned for every HCP based on their value and relevance (Low, Medium, High)

Physician Speciality

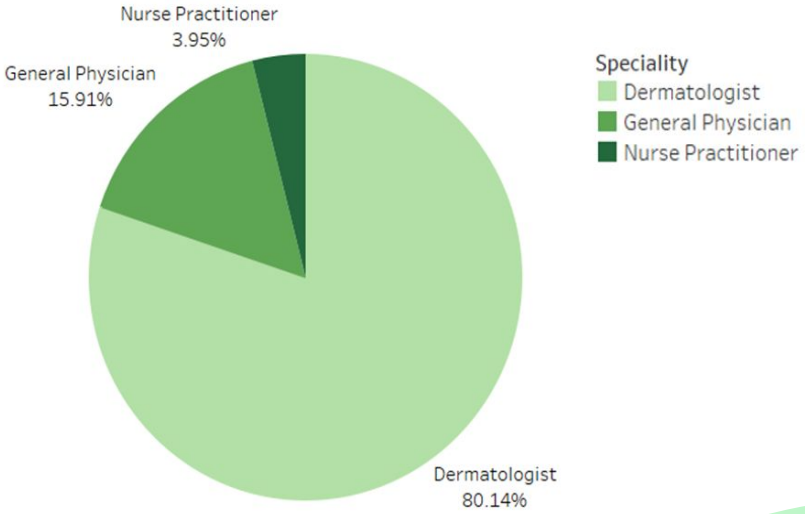
Medical speciality of the HCP (Dermatologist, General Physician, Nurse Practitioner)

02

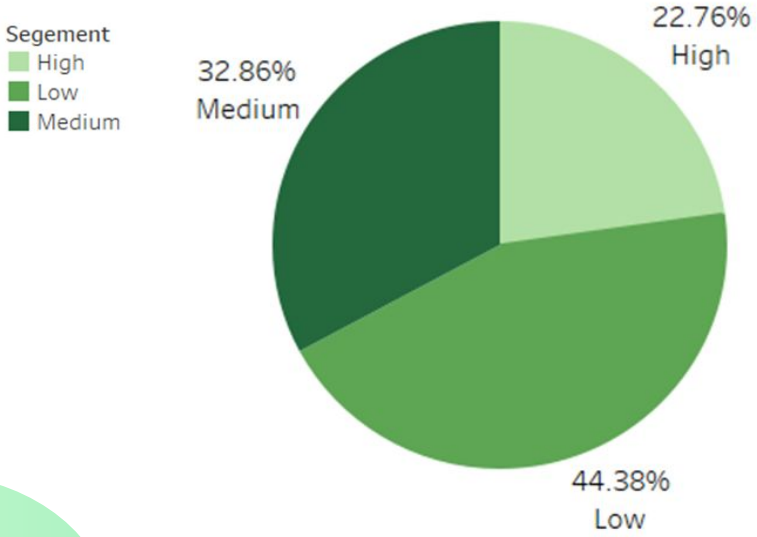
Exploratory Data Analysis (EDA)

Distribution of HCP Data

Speciality of HCP

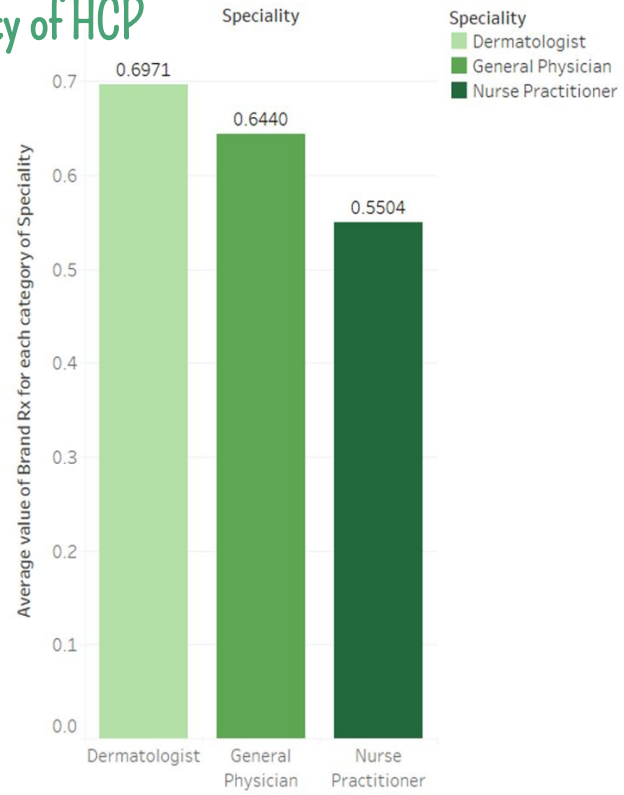


Segment of HCP

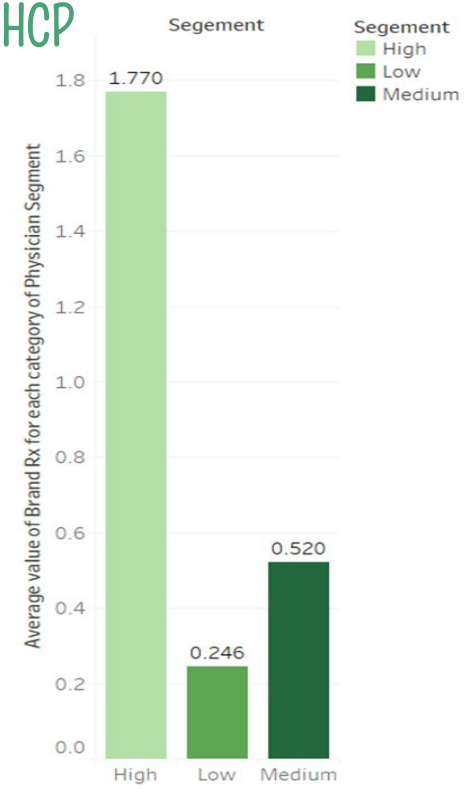


Influence on Brand Rx Value

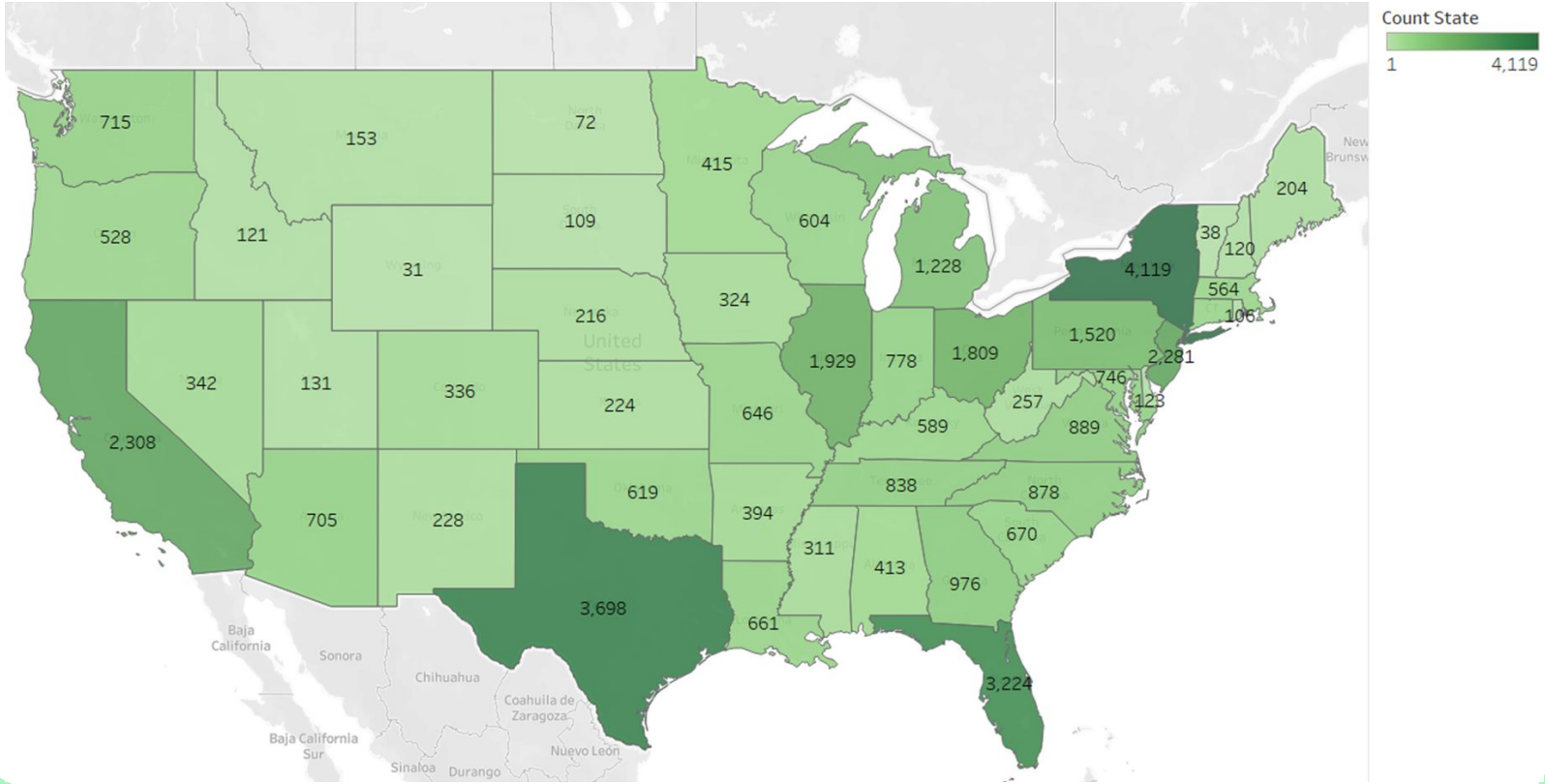
Speciality of HCP



Segment of HCP



Geographical Distribution

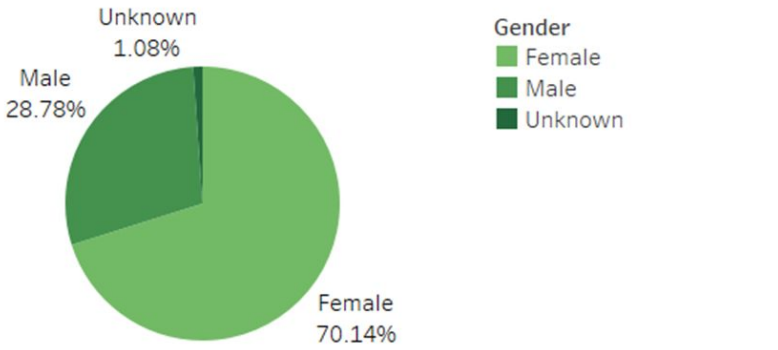


Data Observation

While observing data, we found **1324 HCPs** with either all the promotions as null or Brand Rx value as null.

This could be due to the following reasons:
1. Values are missing
2. Brand Rx is zero even after multiple attempts to reach the HCP
3. Company doesn't want to promote through that particular HCP

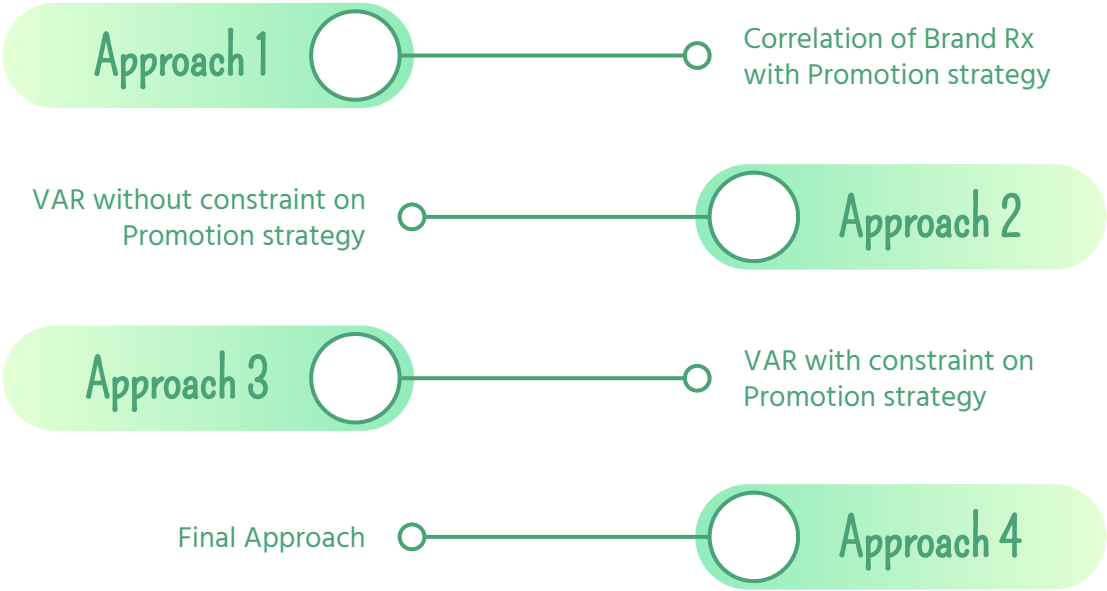
Females are found to be consuming this drug more often than Males.



03

Solution

Task - 1 (Classification Type)

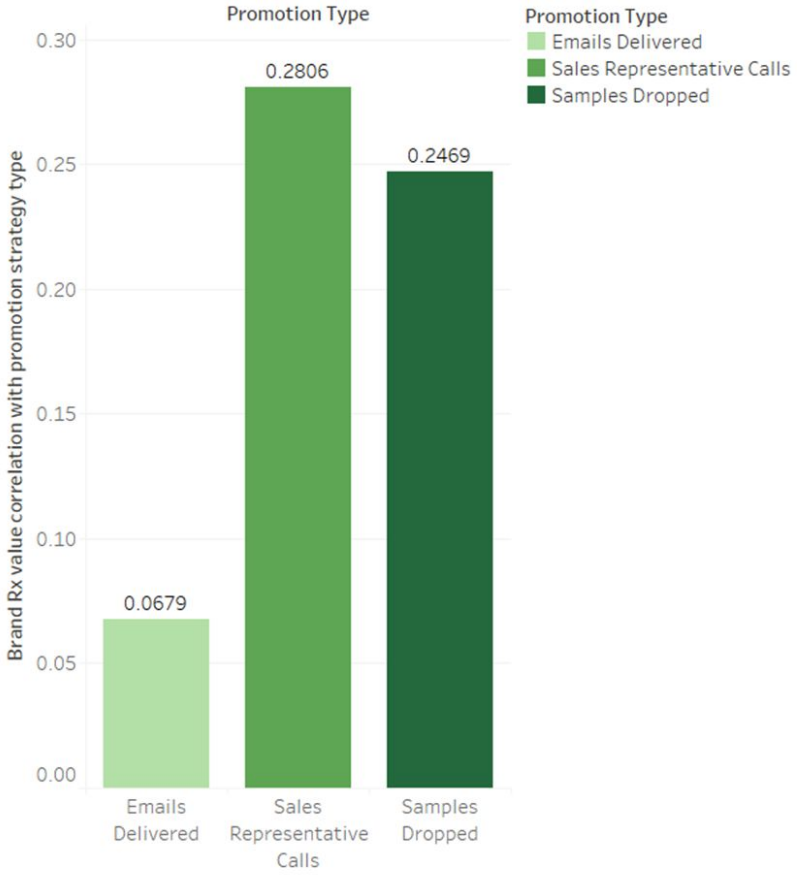


Approaches to Task - 1

A. Correlation of Brand Rx with Promotion strategy:

Sales representative calls has the highest impact, whereas emails barely impact the brand Rx.

For HCPs where either all Brand Rx values are zero or all promotion strategies are also zero. We believe calls is best option due to high correlation.

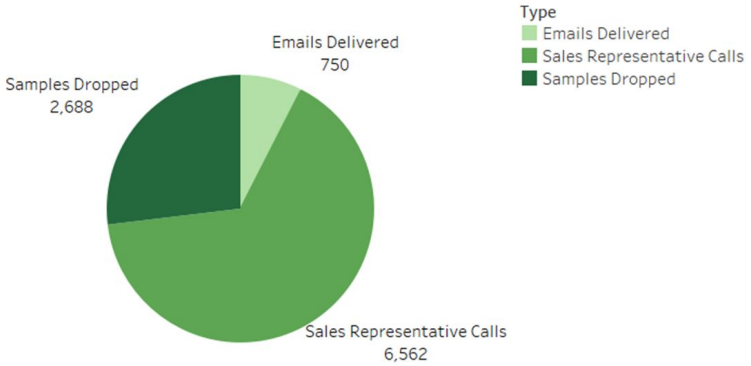


Approaches to Task - 1

B. VAR without constraint on Promotion strategy:

The number of calls, samples dropped and emails was predicted using Vector Autoregression (VAR) for each HCP.

This was done to see how model performs without constraints.



C. VAR with constraint on Promotion strategy:

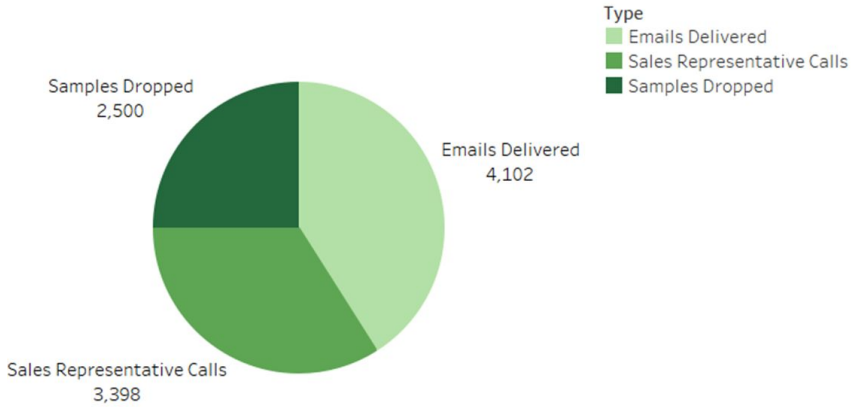
Constraints were sequentially resolved, starting with calls followed by samples dropped, and if the condition on either of those were violated, we predicted the promotion strategy as email.

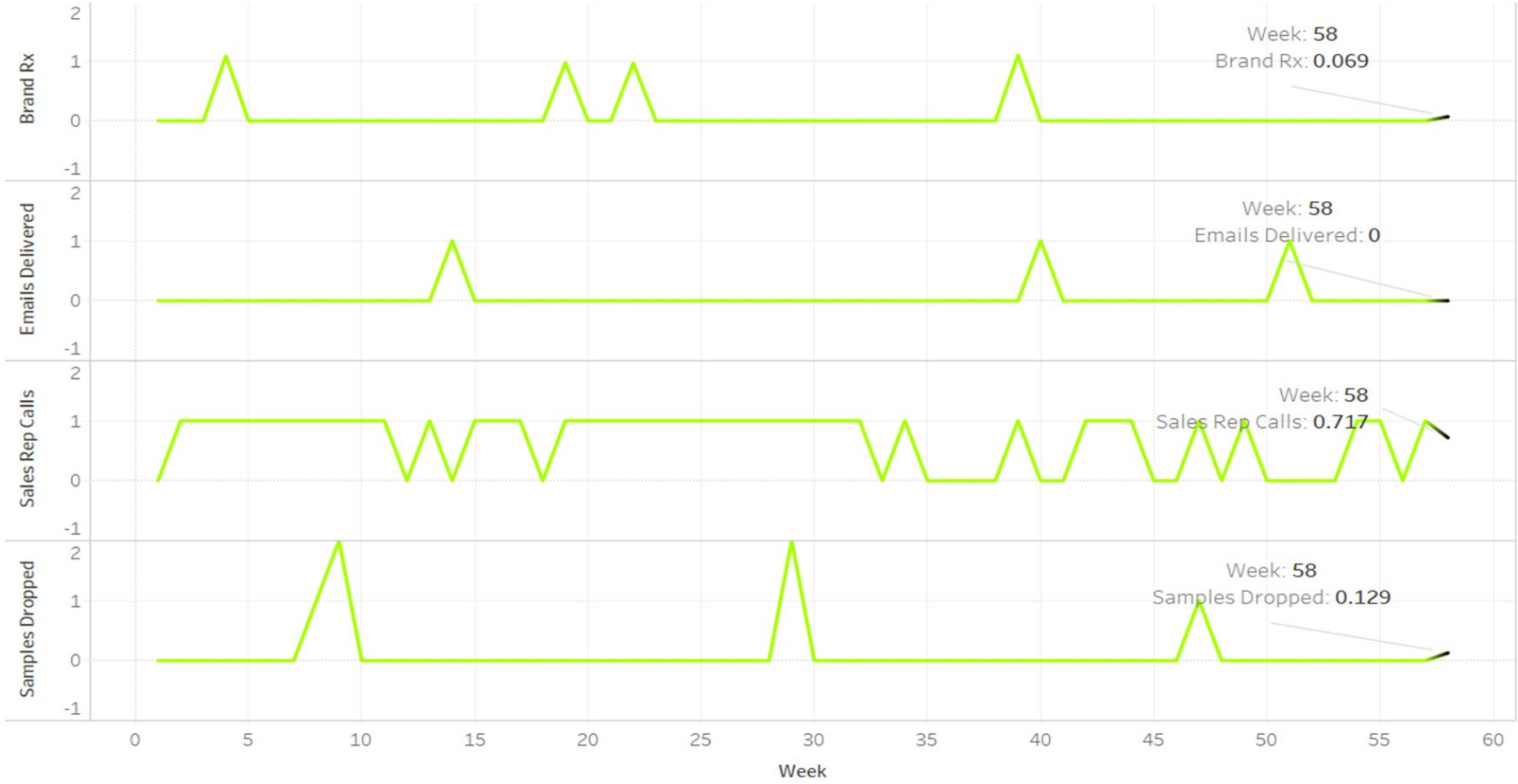
Approaches to Task - 1

D. Final Approach:

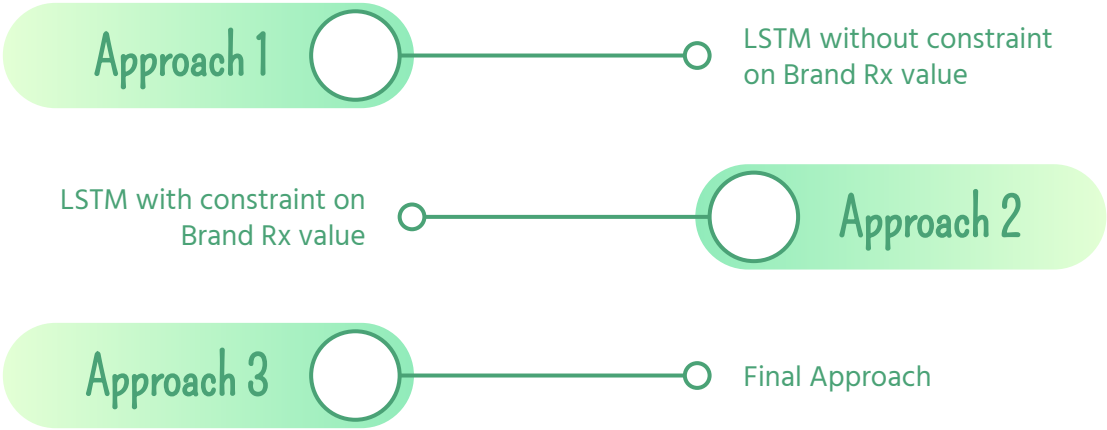
Building upon previous approach, we found that sample drop already violated the constraint before reaching the last HCP.

The final promotion strategy was decided based on the sorting scheme we designed.




Brand Rx
Week: 58
Brand Rx: 0.069
Emails Delivered
Week: 58
Emails Delivered: 0
Sales Rep Calls
Week: 58
Sales Rep Calls: 0.717
Samples Dropped
Week: 58
Samples Dropped: 0.129
2
1
0
-1
0
5
10
15
20
25
30
35
40
45
50
55
60
Week

Task - 2 (Regression Type)



Approaches to Task - 2

A. LSTM without constraint on Brand Rx value:

For all HCPs, a deep learning model (LSTM) was applied to predict the Brand Rx value.

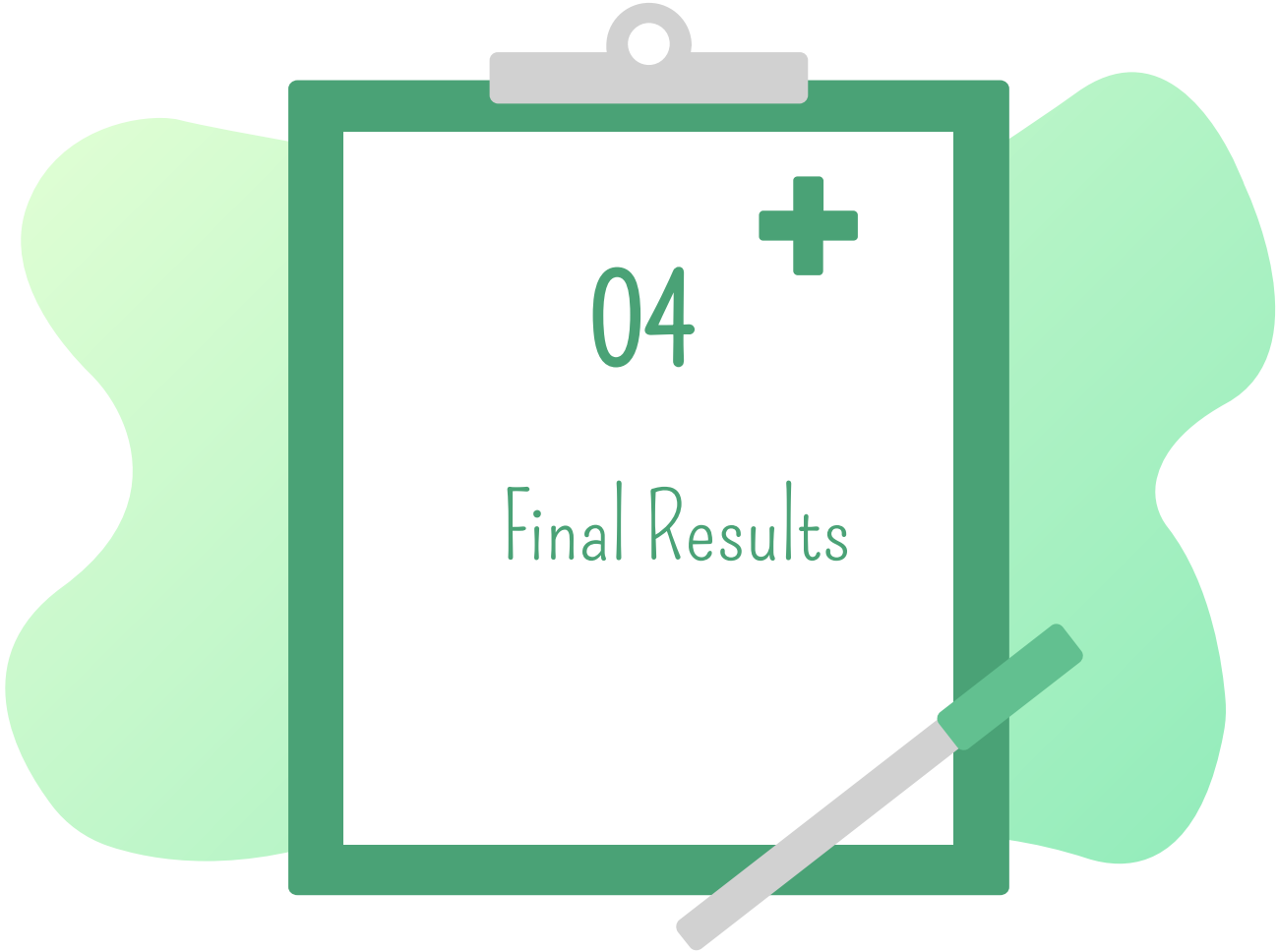
The 10% constraint was not satisfied.

B. LSTM with constraint on Brand Rx value:

Building on the previous approach, the predictions obtained were clipped to fall within the constraint.

C. Final Approach:

Our final approach was just a naive prediction of average of Brand Rx in the previous year.


04
Final Results

Summary

Task - 1:

The predicted distribution of promotion strategy is as follows:

Sales representative calls are **3398** and satisfy constraint based on HCP segment

Sample drops are **2500** and satisfy constraint of maximum 25% in a week

Emails are **4102**

Task - 2:

The sum of Brand Rx value for the last week of Jan'20 is **6376.952**

The sum of predicted Brand Rx value for the first week of Feb'20 is **6861.917**

The deviation is **7.6%** which satisfies the constraint of maximum of 10% variation from previous week

.

Thank You

Harsh Sukheja

AKA Mike

Jeevesh Mahajan

AKA Harvey

Manu Sehgal

AKA Louis