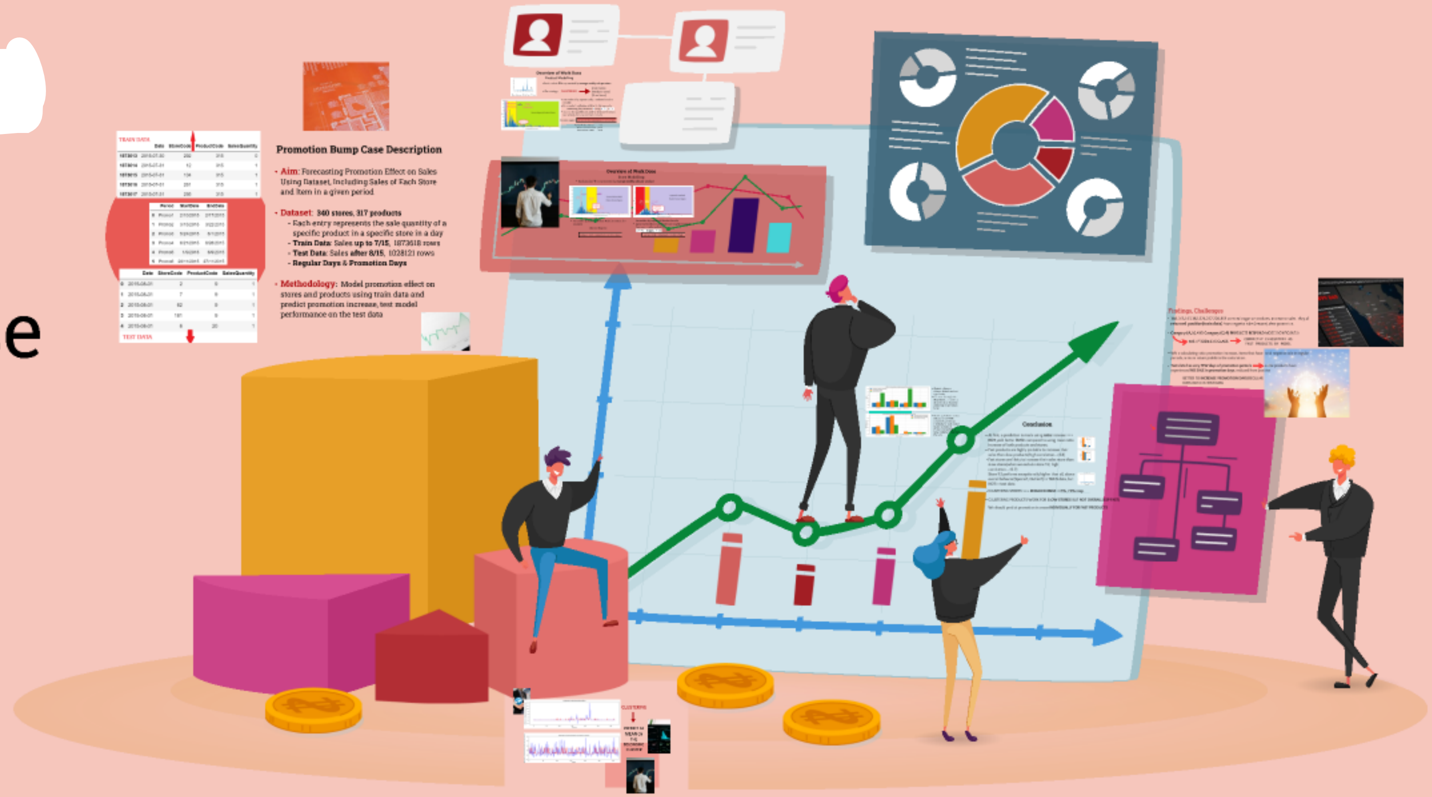
8/12/2024

Promotion Bump Case

Ender Erkaya



Promotion Bump Case Description

- **Aim**: Forecasting Promotion Effect on Sales Using Dataset, Including Sales of Each Store and Item in a given period

- **Dataset**: **340 stores, 317 products**
  - Each entry represents the sale quantity of a specific product in a specific store in a day
  - **Train Data**: Sales **up to 7/15**, 1873618 rows
  - **Test Data**: Sales **after 8/15**, 1028121 rows
  - **Regular Days** & **Promotion Days**

- **Methodology**: Model promotion effect on stores and products using train data and predict promotion increase, test model performance on the test data

TRAIN DATA

	Date	StoreCode	ProductCode	SalesQuantity
1873613	2015-07-30	292	315	0
1873614	2015-07-31	12	315	1
1873615	2015-07-31	104	315	1
1873616	2015-07-31	261	315	1
1873617	2015-07-31	295	315	1

	Period	StartDate	EndDate
0	Promo1	2/10/2015	2/17/2015
1	Promo2	3/15/2015	3/22/2015
2	Promo3	5/24/2015	6/1/2015
3	Promo4	6/21/2015	6/28/2015
4	Promo5	1/9/2015	6/9/2015
5	Promo6	20/11/2015	27/11/2015

	Date	StoreCode	ProductCode	SalesQuantity
0	2015-08-01	2	9	1
1	2015-08-01	7	9	1
2	2015-08-01	62	9	1
3	2015-08-01	181	9	1
4	2015-08-01	6	20	1

TEST DATA

Overview of Work Done

Product Modelling

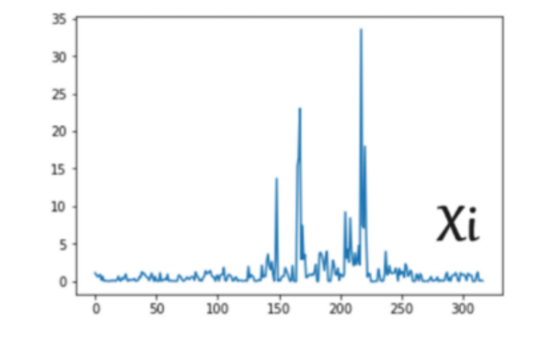
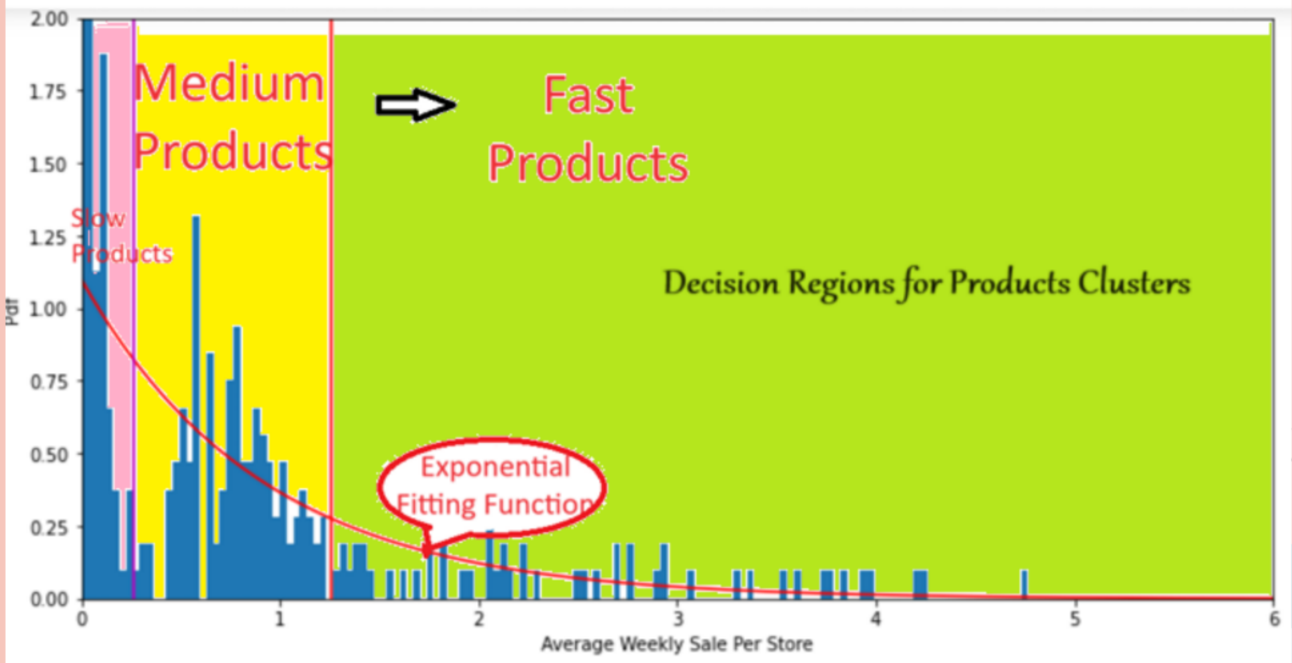


Figure 1: Average Weekly Sales Per Stores

- Each product **Xi** is represented by **average weekly sale per store**
- Our strategy: **CLUSTERING** → {Fast Items} {Medium Items} {Slow Items}



- X is modeled by exponentially distributed random variable.
- Exponential distribution is fitted to histogram by
  - estimating the parameter $\hat{\lambda}$ using $\hat{\mu}_{robust}$, $\hat{\sigma}_{robust}$
- 1st and 3rd quartiles are used for decision threshold respectively for Slow and Fast Products

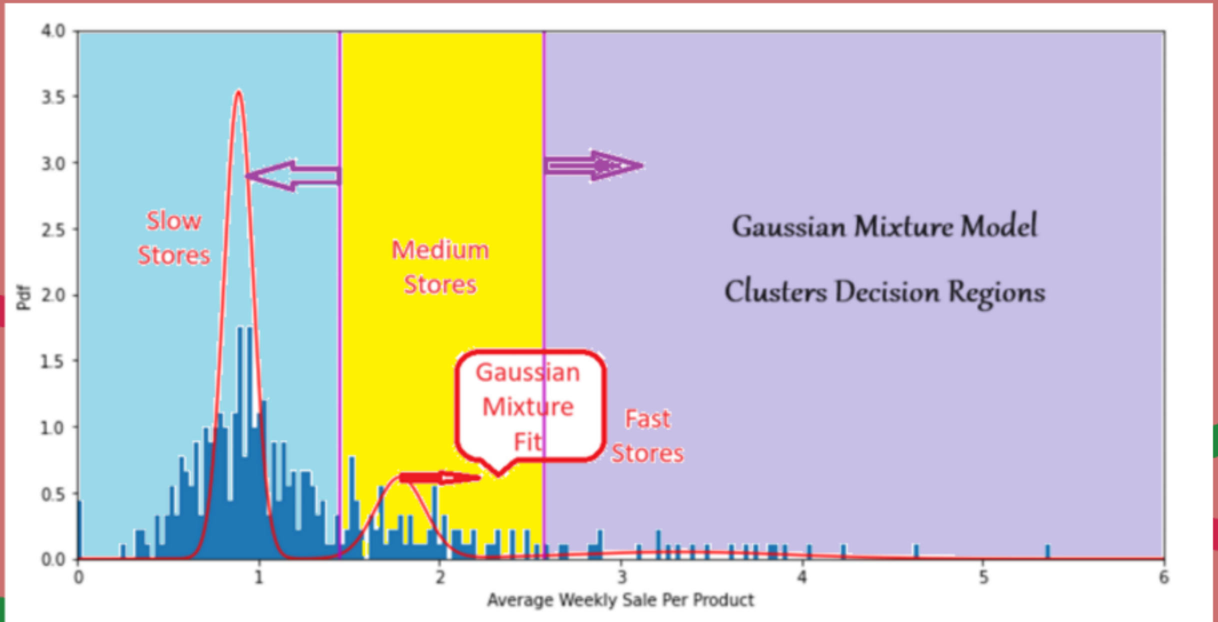
Decision Region: xslow < 0.26 < xmedium < 1.26 < xfast

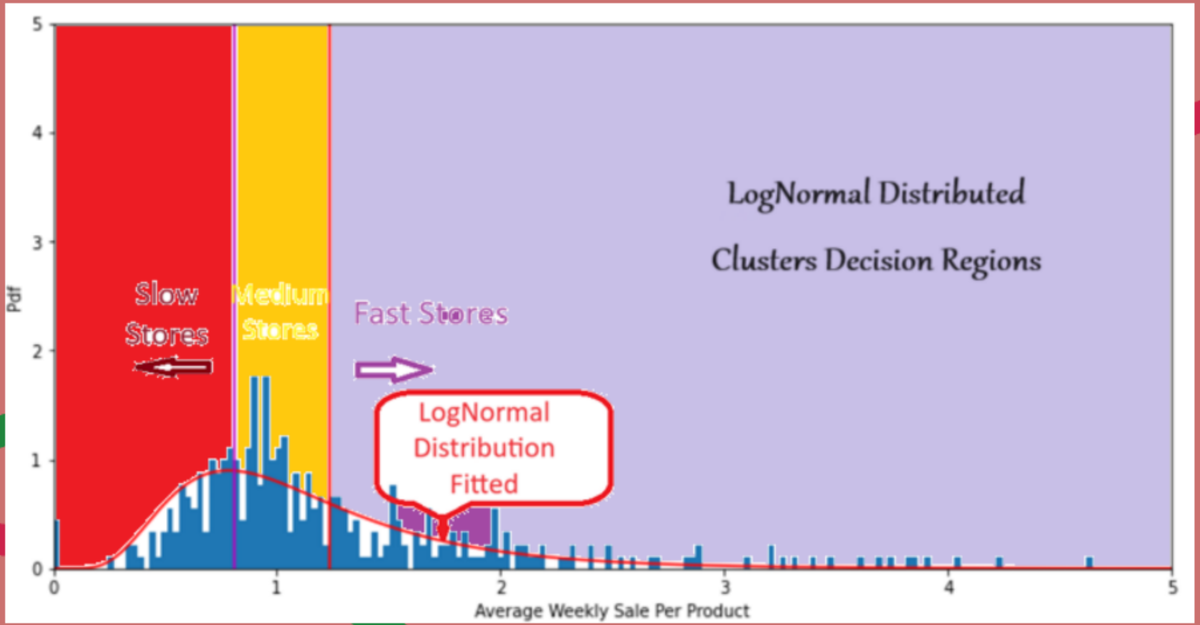
- |Fast Product Cluster| = 63
- |Medium Product Cluster| =126
- |Slow Product Cluster| =128

Overview of Work Done

Store Modelling

- Each product **Yi** is represented by **average weekly sale per product**





- *Y is modeled by Gaussian Mixture Model, ie mixture of 3 Gaussians*

Decision Regions:

xslow < 1.45 < xmedium < 2.58 < xfast

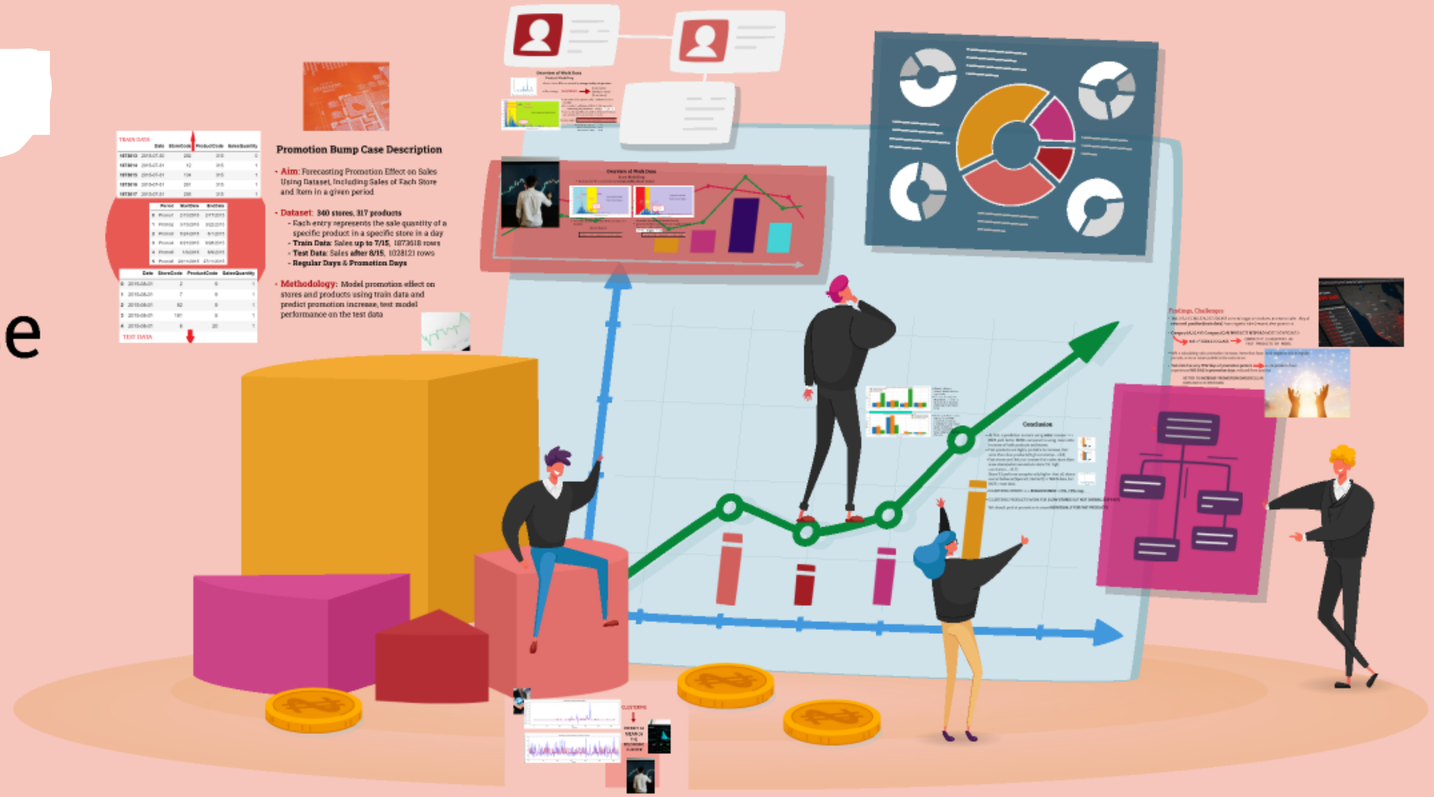
- *Y is modeled by LogNormal Random Variable*
- *Lognormal distribution parameters are roughly estimated from mean, median and std as*
- $\mu = 7e - 3$ and $\sigma = 0.5$

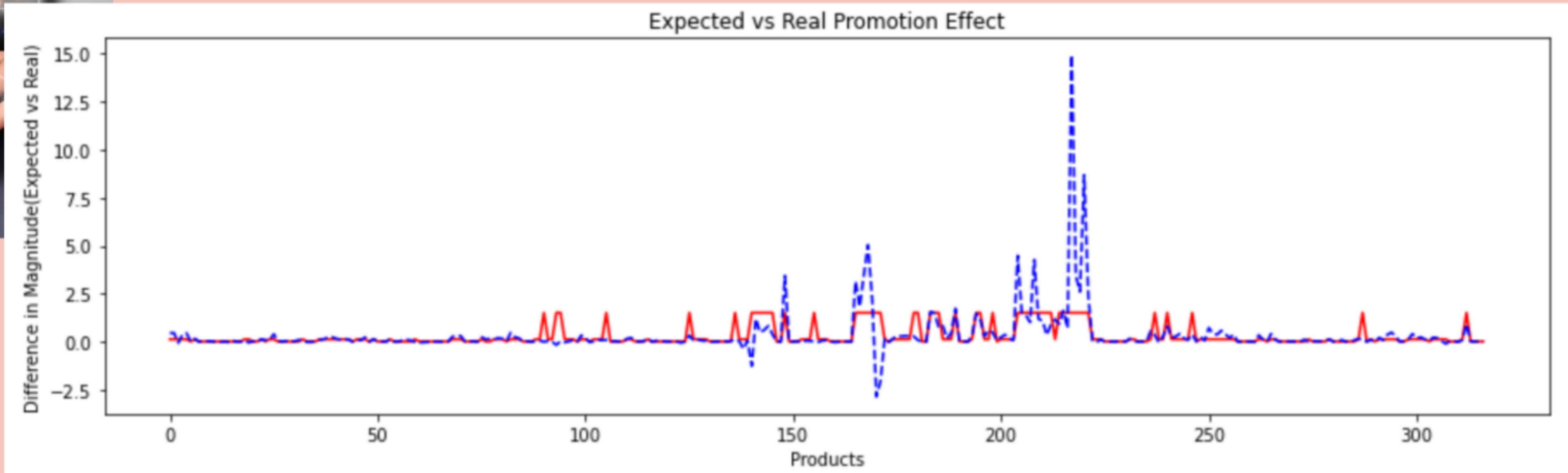
xslow < 0.8 < xmedium < 1.24 < xfast

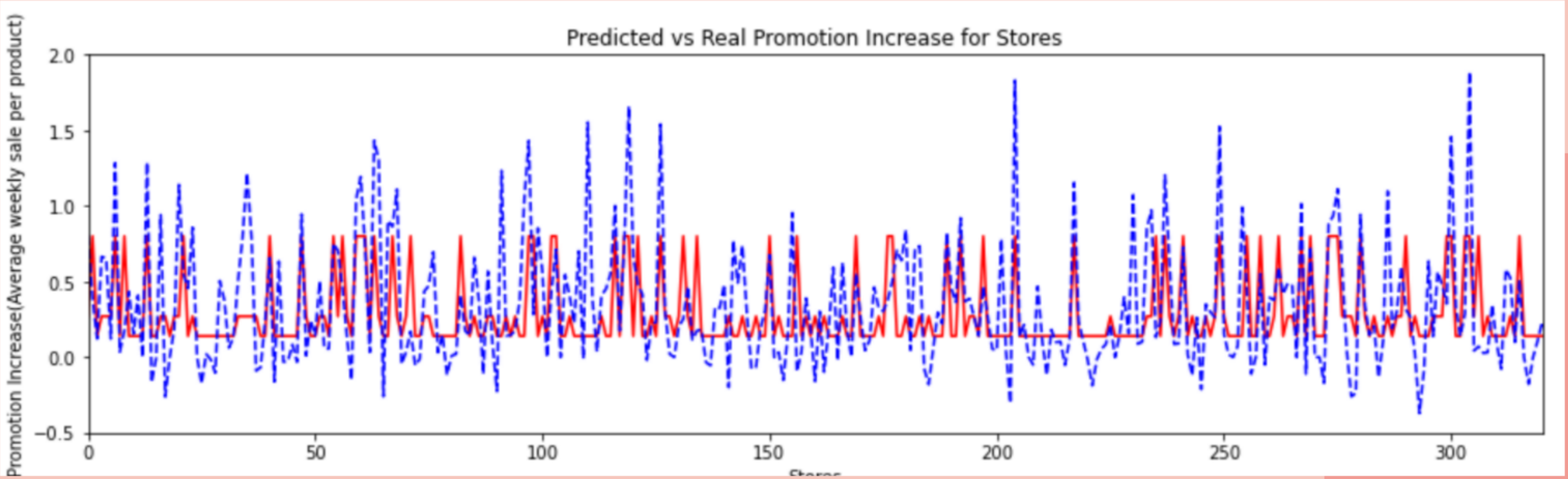
8/12/2024

Promotion Bump Case

Ender Erkaya

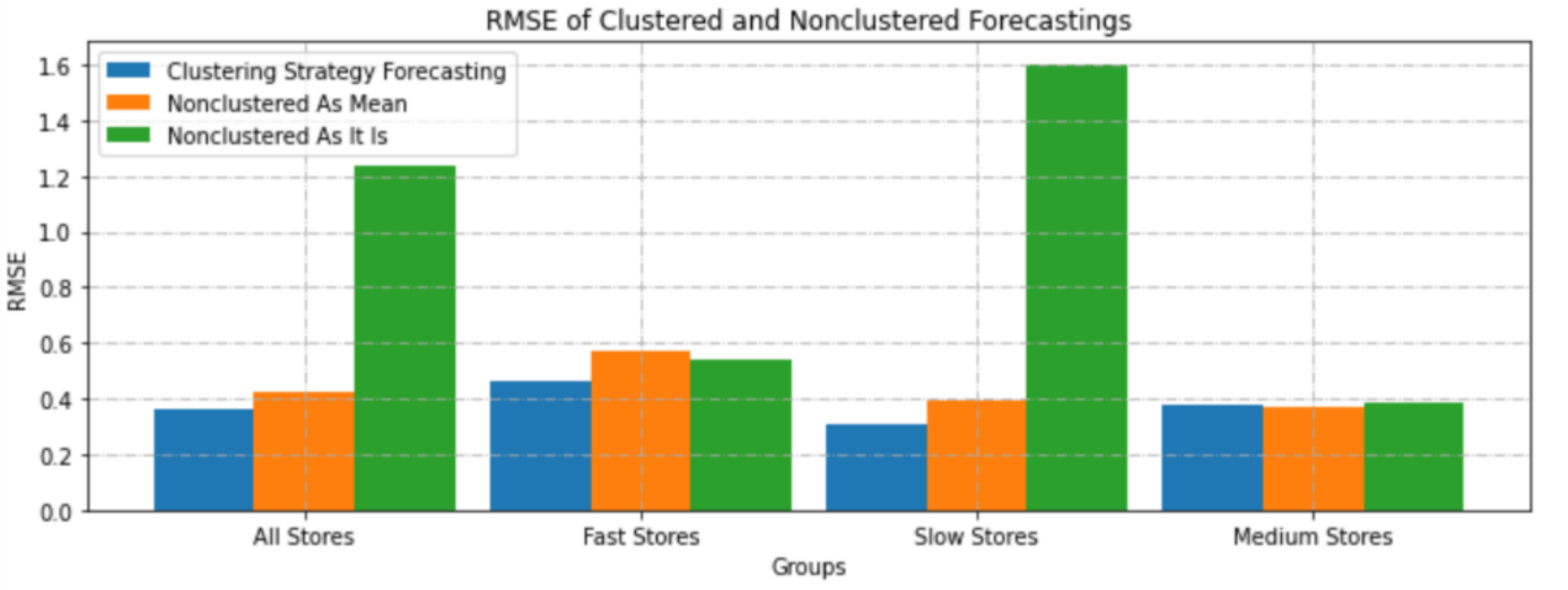




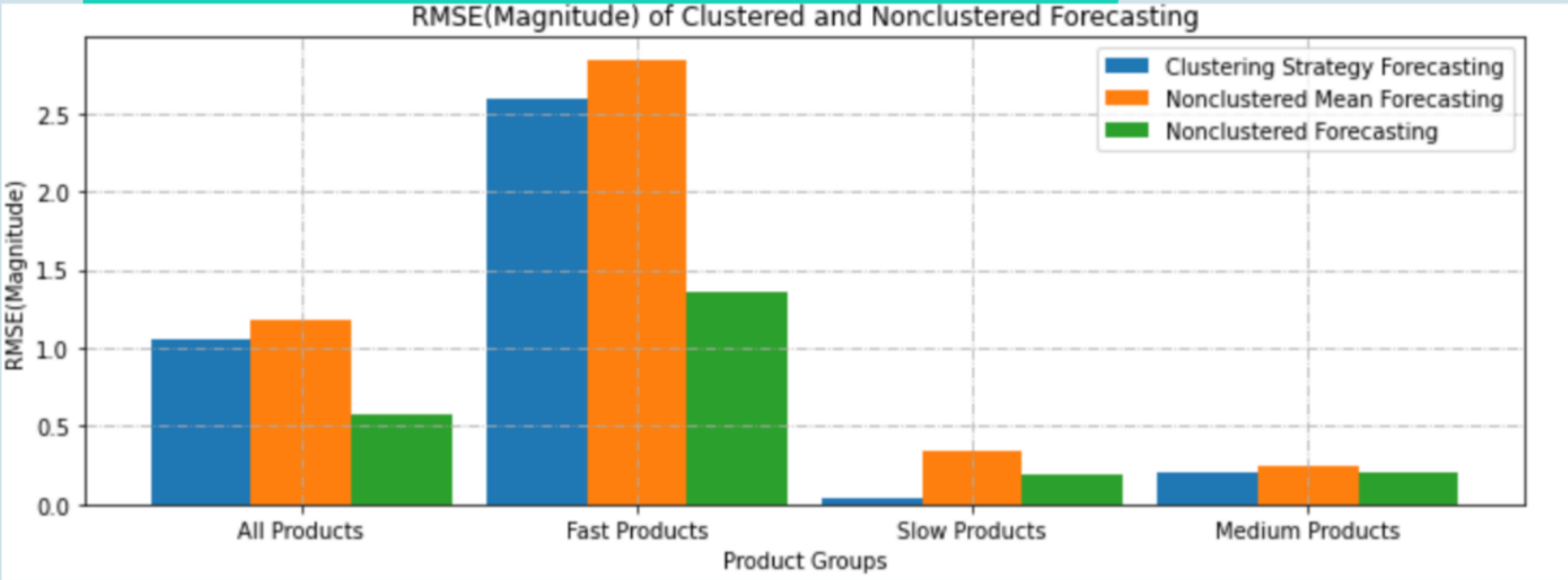


CLUSTERING

PREDICT AS MEAN OF THE BELONGING CLUSTER

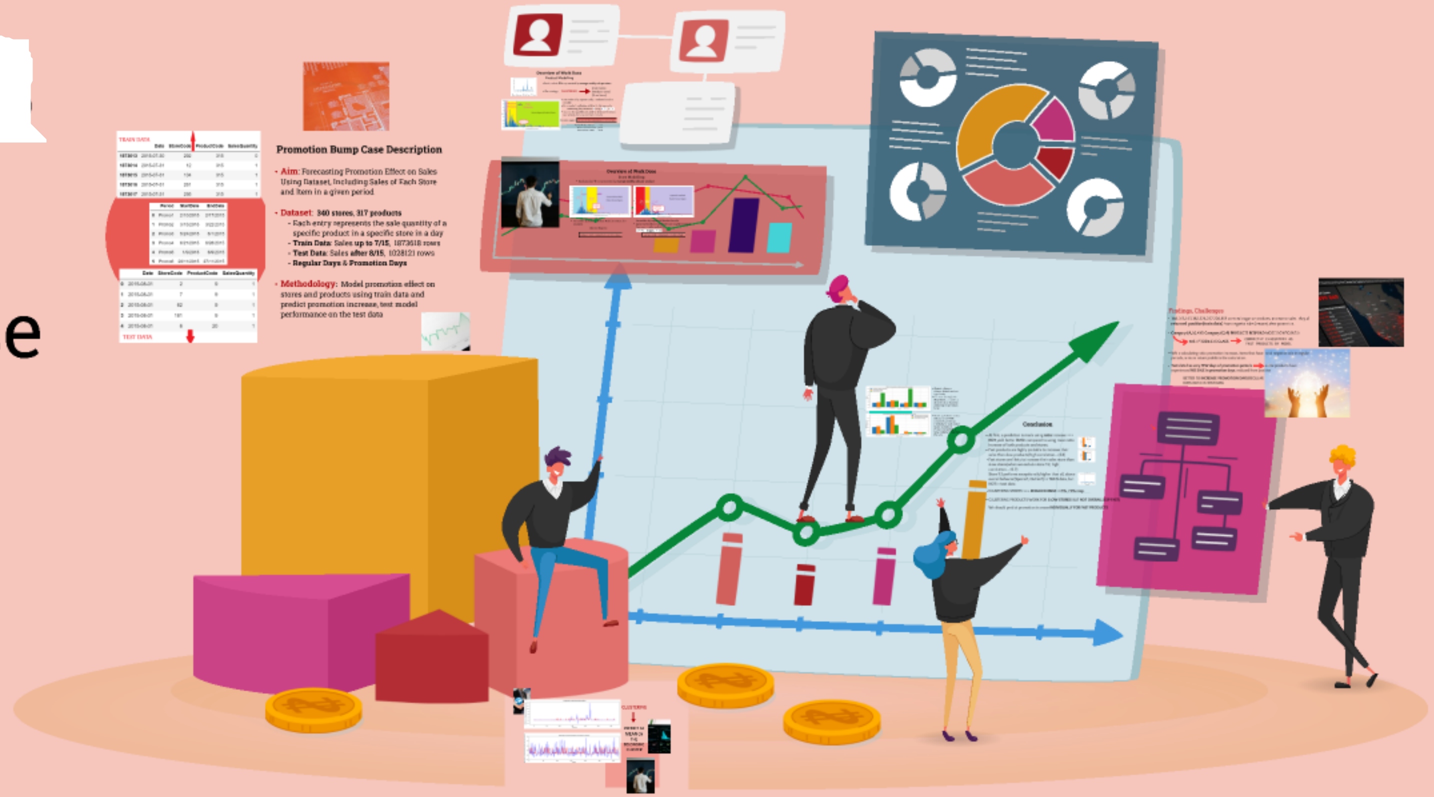


- Clustering Stores is efficient, ie RMSE reduced significantly.
- It is more meaningful for **Slow Stores.** It is better to determine Slow Stores for estimating the promotion bump.



- Clustering Products has **not reduced** overall RMSE compared to predicting individually for each product
- Clustering Model of Products has **worked** for **Slow Products, but not Fast Products**

8/12/2024

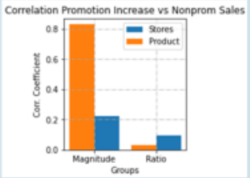


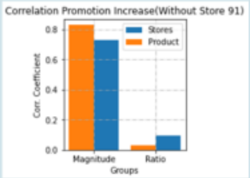
Promotion Bump Case

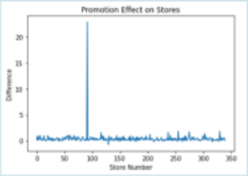
Ender Erkaya

Conclusion

- At first, a prediction is made using **ratio** increase ==> **NOT** yield better RMSE compared to using mean ratio increase of both products and stores.
- Fast products are highly probable to increase their sales than slow products(high correlation~>0.8)
- Fast stores are likely to increase their sales more than slow stores(when we exclude store 92, high correlation ~>0.7)
- Store 92 performs exceptionally higher than all, above overall behavior(Special?, Outlier?) in TRAIN data, but NOT in test data

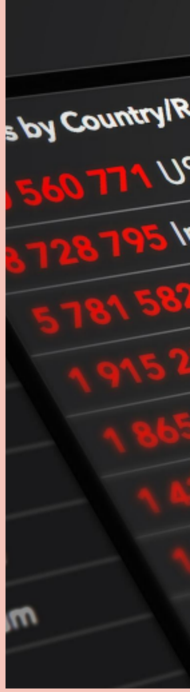






- CLUSTERING STORES ==> **REDUCED RMSE : 15%, 70% resp.**
- CLUSTERING PRODUCTS WORK FOR **SLOW STORES** BUT **NOT OVERALL(ESP FAST)**
- We should predict promotion increase **INDIVIDUALLY FOR FAST PRODUCTS**

Findings, Challenges



- 160,163,165,182,226,227,228,309 are total negative products, ie returns>sales. They all **returned positive(train data)** from negative sales(returns) after promotion.

- **Category (A,5)** AND **Category (G,4)** PRODUCTS RESPOND MORE SIGNIFICANTLY

  → FAST PRODUCTS CLASS → CORRECTLY CLASSIFIED AS FAST PRODUCTS BY MODEL

- While calculating ratio promotion increase, items that have total negative sale in regular periods, ie more return prohibits the calculation.

- **Test data has very FEW days of promotion periods** → some products have experienced **NO SALE in promotion days**, reduced from positive

  BETTER TO **INCREASE PROMOTION DAYS**/REGULAR DAYS RATIO IN TEST DATA

8/12/2024

Promotion Bump Case

Ender Erkaya