



# H&M RECOMMENDATION SYSTEM



Aditya Parameswaran, Prashanthi Ramachandran, Shashidhar Pai, Shipra Priyadarshini

# Roadmap

- I Introduction
- II Motivation
- III Data
- IV The TrainDy Recommender<sup>TM</sup>
- V Results
- VI Demo
- VII Conclusion & Discussions

# Introduction

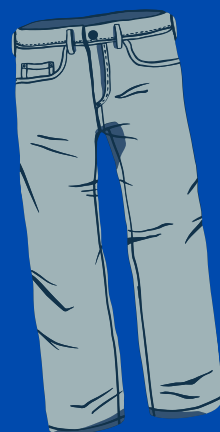
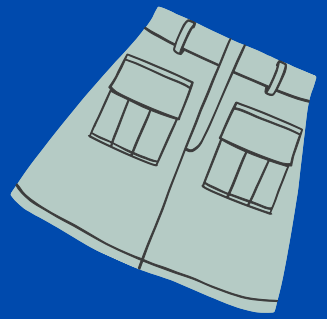
- H&M Recommender System
- Part of the Kaggle Competition
- Explored different techniques such as collaborative and content-based filtering
- ML models: k-means clustering and conditional VAEs



# Motivation

- H&M Group: 53 online markets and approximately 4,850 stores.
- Improving Very difficult for customers to find what they want or are interested in quickly
- Customer engagement for business
- Reduces returns, thereby reducing emissions from transportation





# Data

- Purchase history, (anonymized) customer demographic, and product data
  - articles.csv
  - customers.csv
  - transactions.csv
- EDA and preprocessed

**1.3 million**

CUSTOMERS

**31 million**

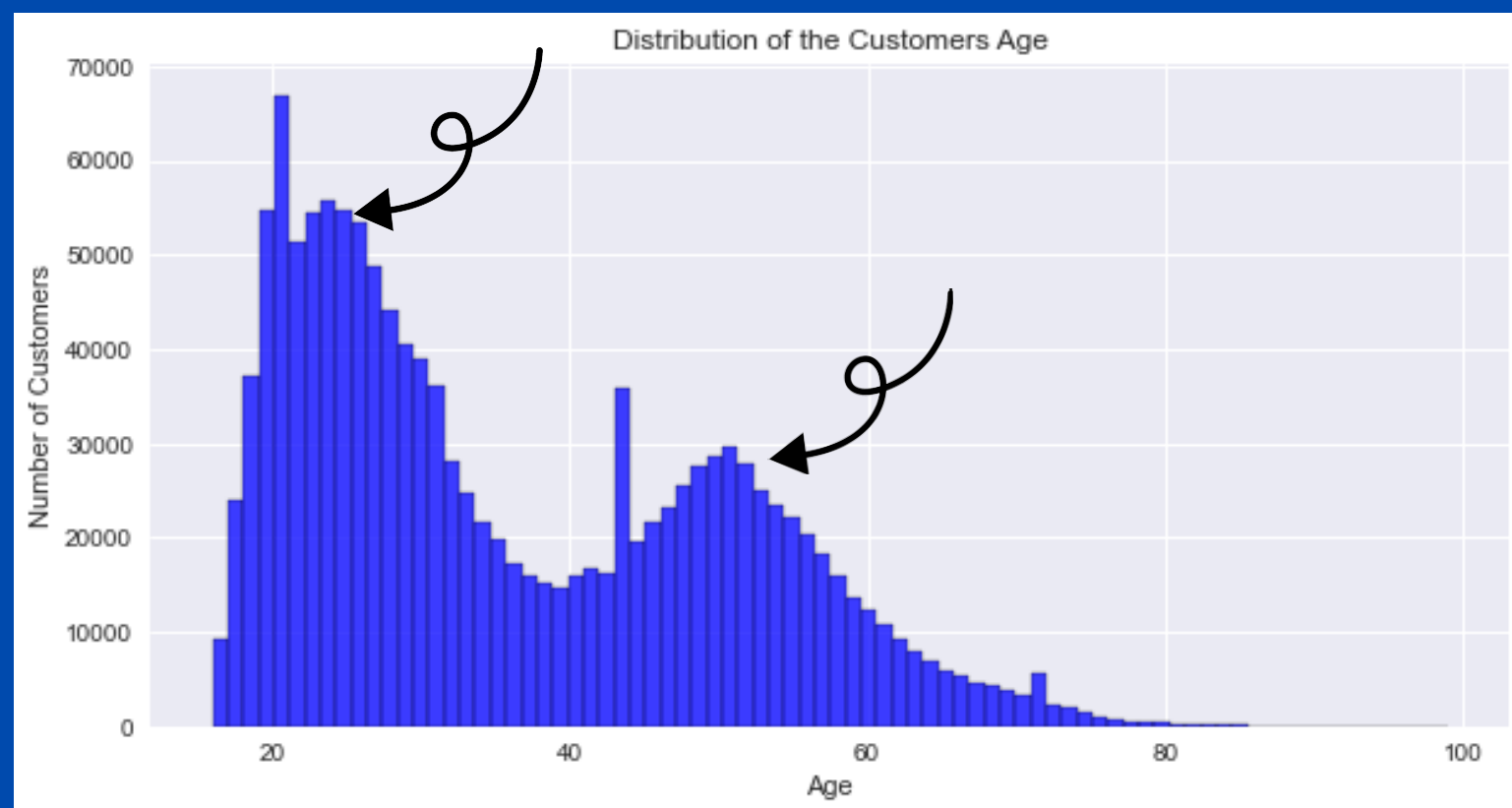
TRANSACTIONS

**300K**

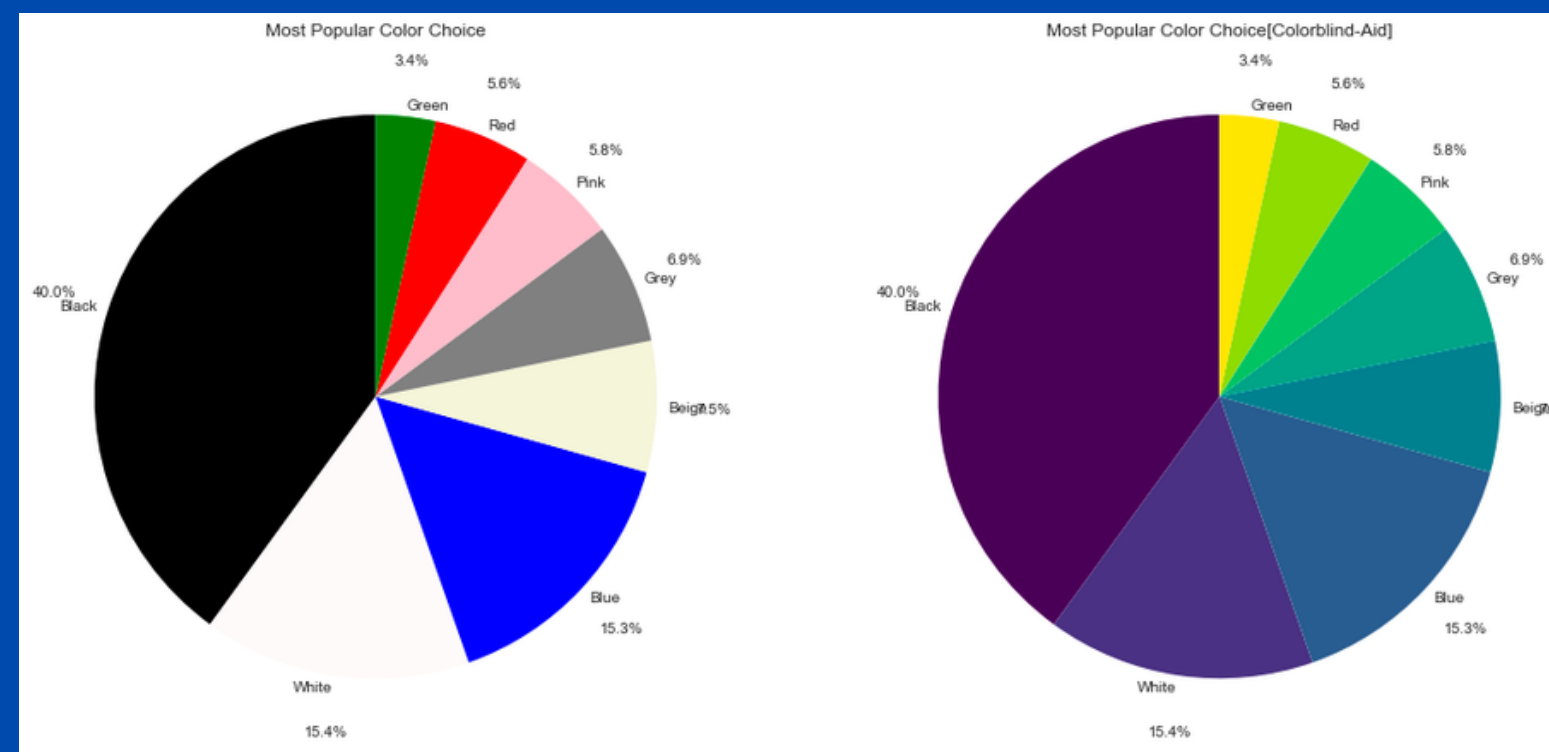
ARTICLES



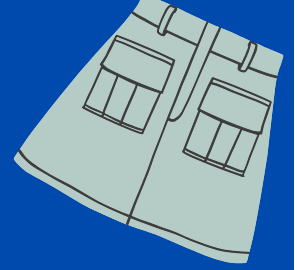
# EDA and Viz. (I)



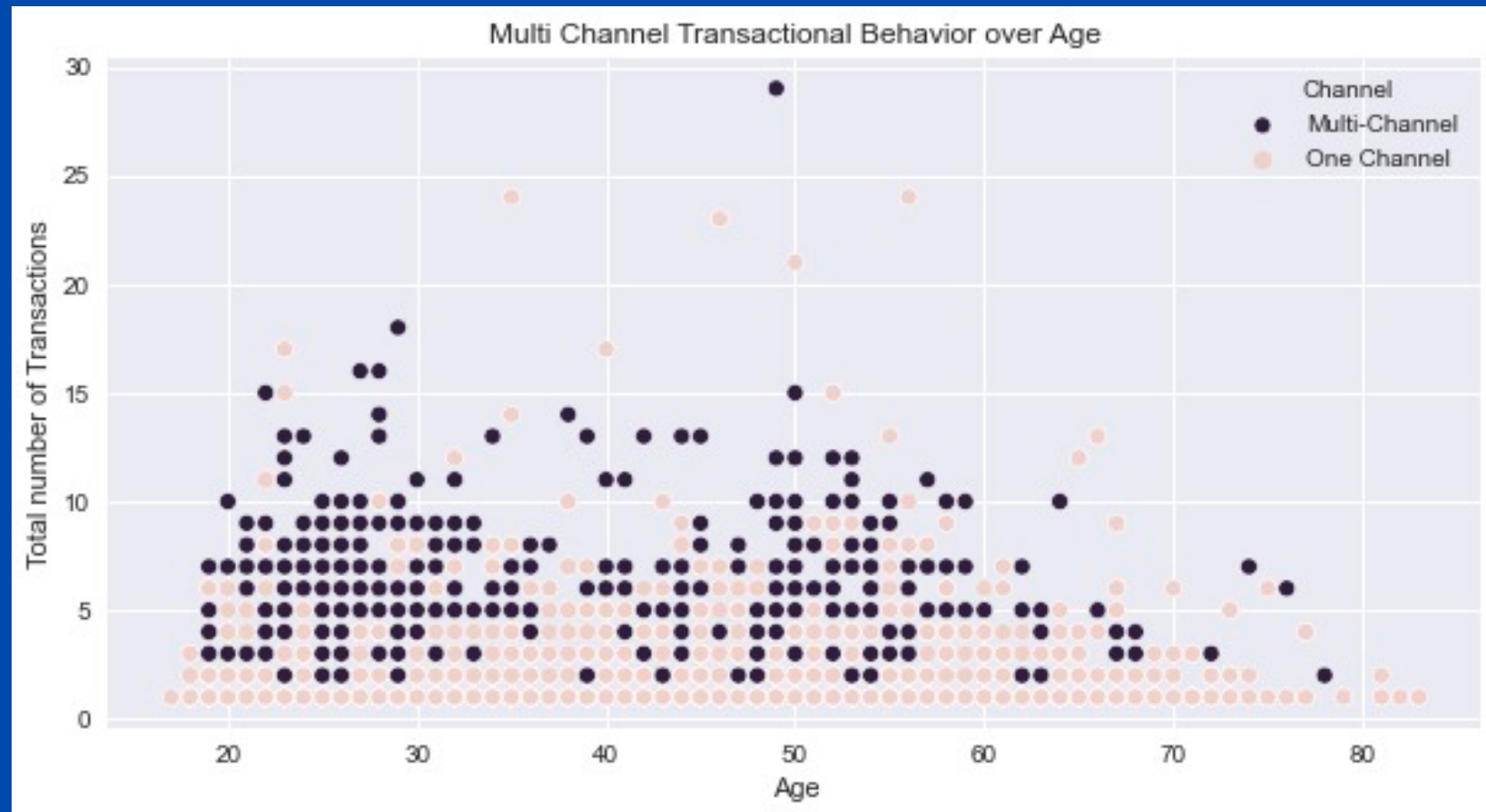
Age Distribution of customers



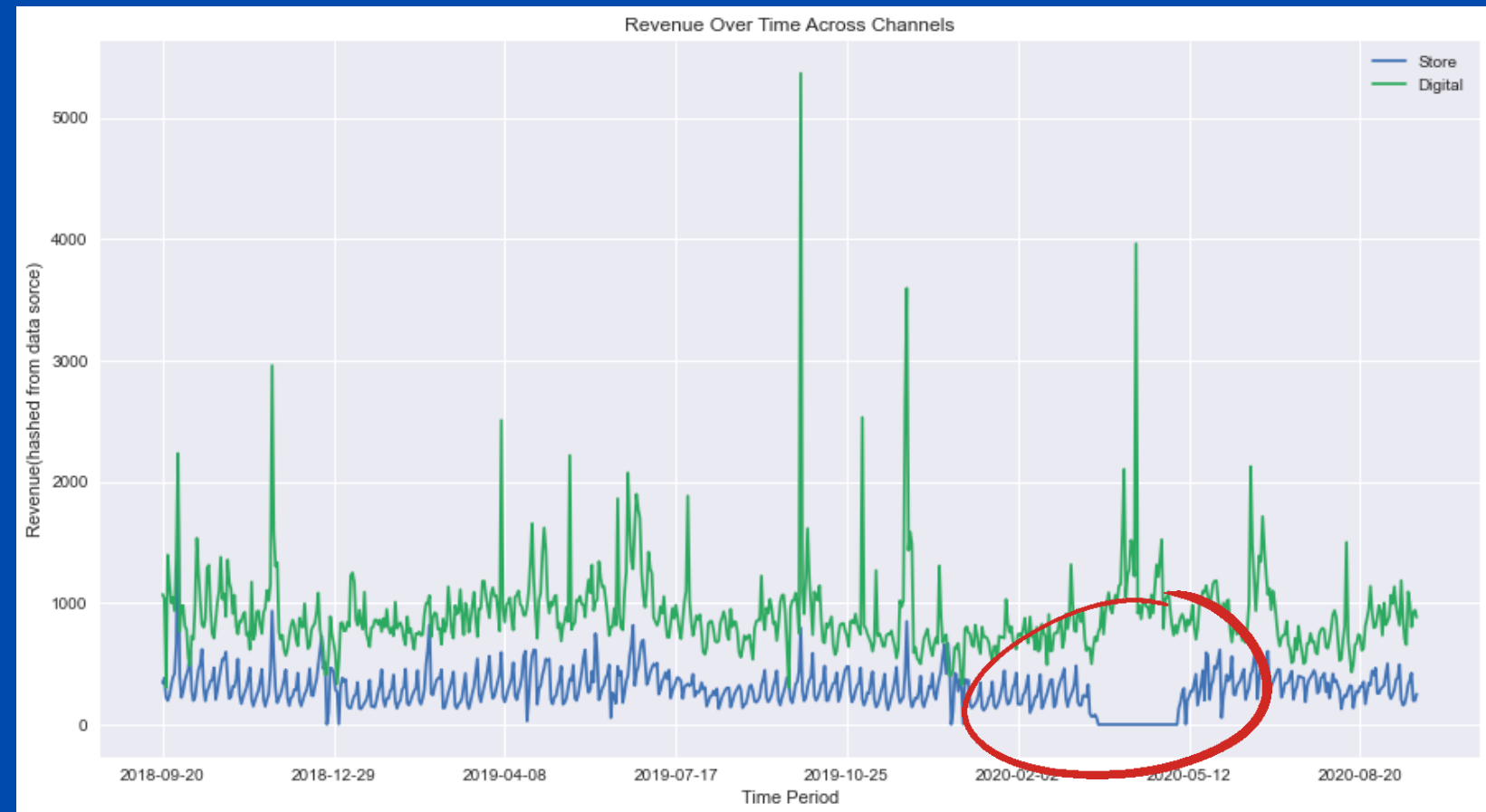
Most Popular Color choices



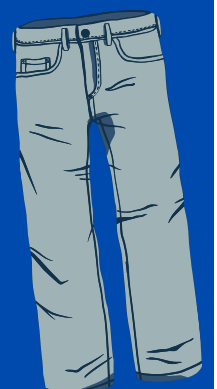
# EDA and Viz. (II)



Behaviour of customers who purchase through one channel vs those that purchase through both channels



Revenue Over Time Across Channels





# Hypothesis I

- Hypothesis:

We want to know whether the average day gap between transactions (ADGT) for digital customers is different from that of store customers.

- Null hypothesis ( $H_0$ ):

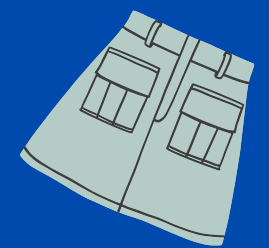
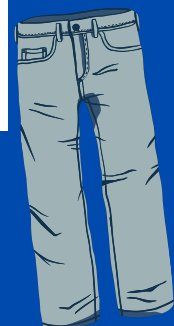
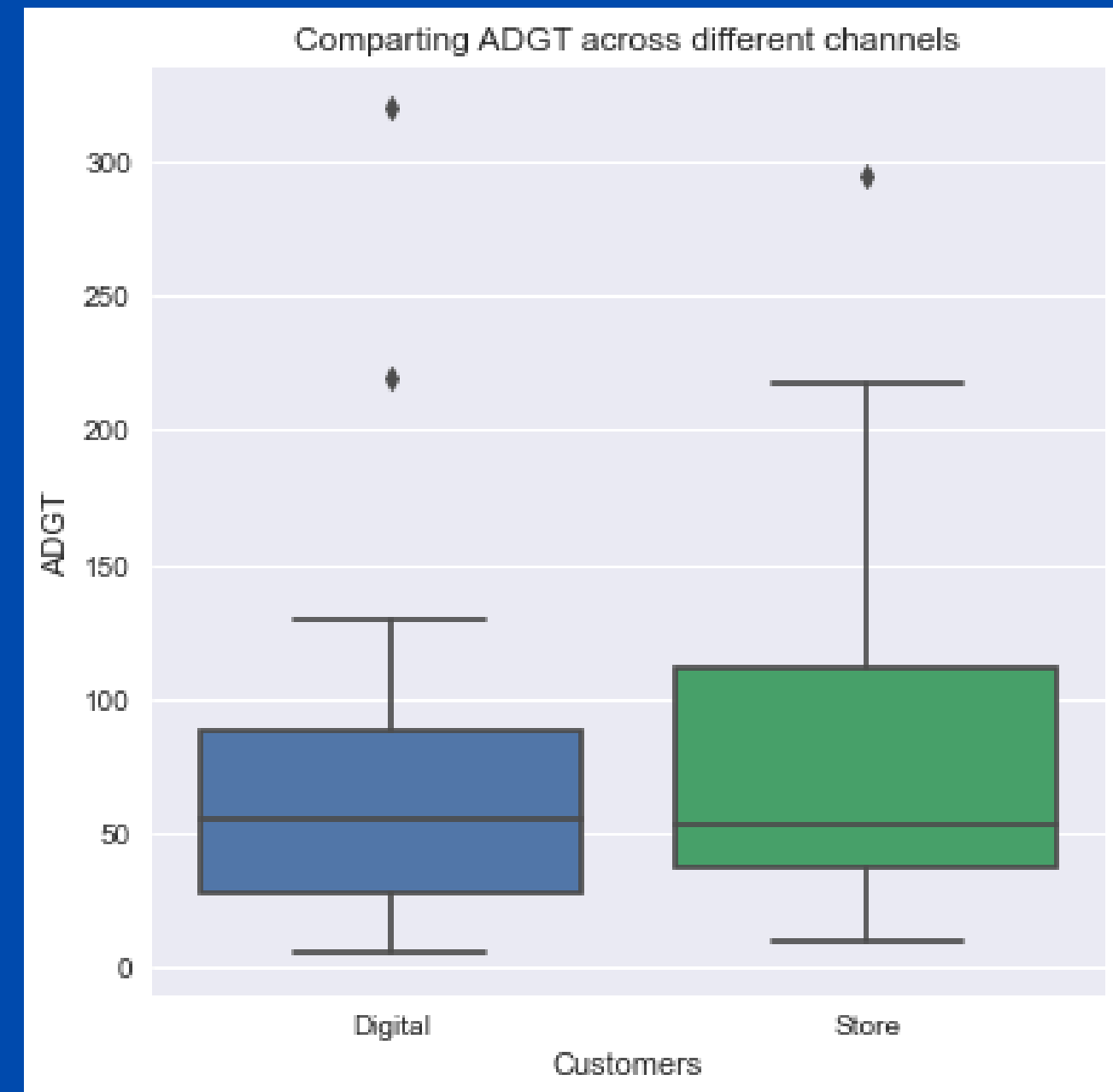
The means of the ADGTs for digital and store customers are similar to each other.

- Alternate hypothesis, ( $H_a$ ):

The means of the ADGTs for digital and store customers are different from each other.

- Two-sample T-test
- p-value: 0.07; since  $0.07 \gg 0.001$ , we accept the null hypothesis

$$\text{adgt} = \frac{(\text{date of last purchase} - \text{date of first purchase})}{\text{total number of purchases in this period}}$$

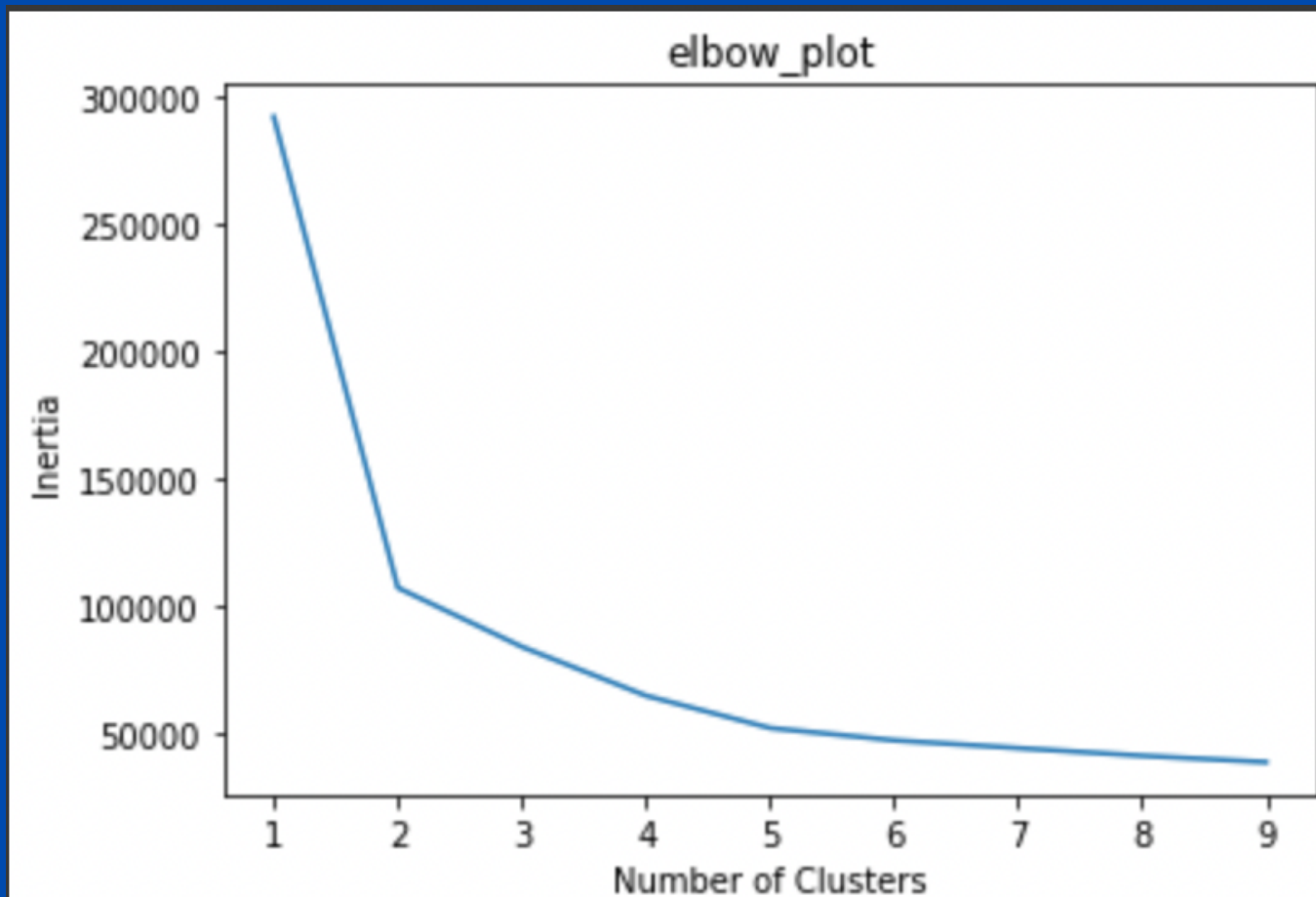




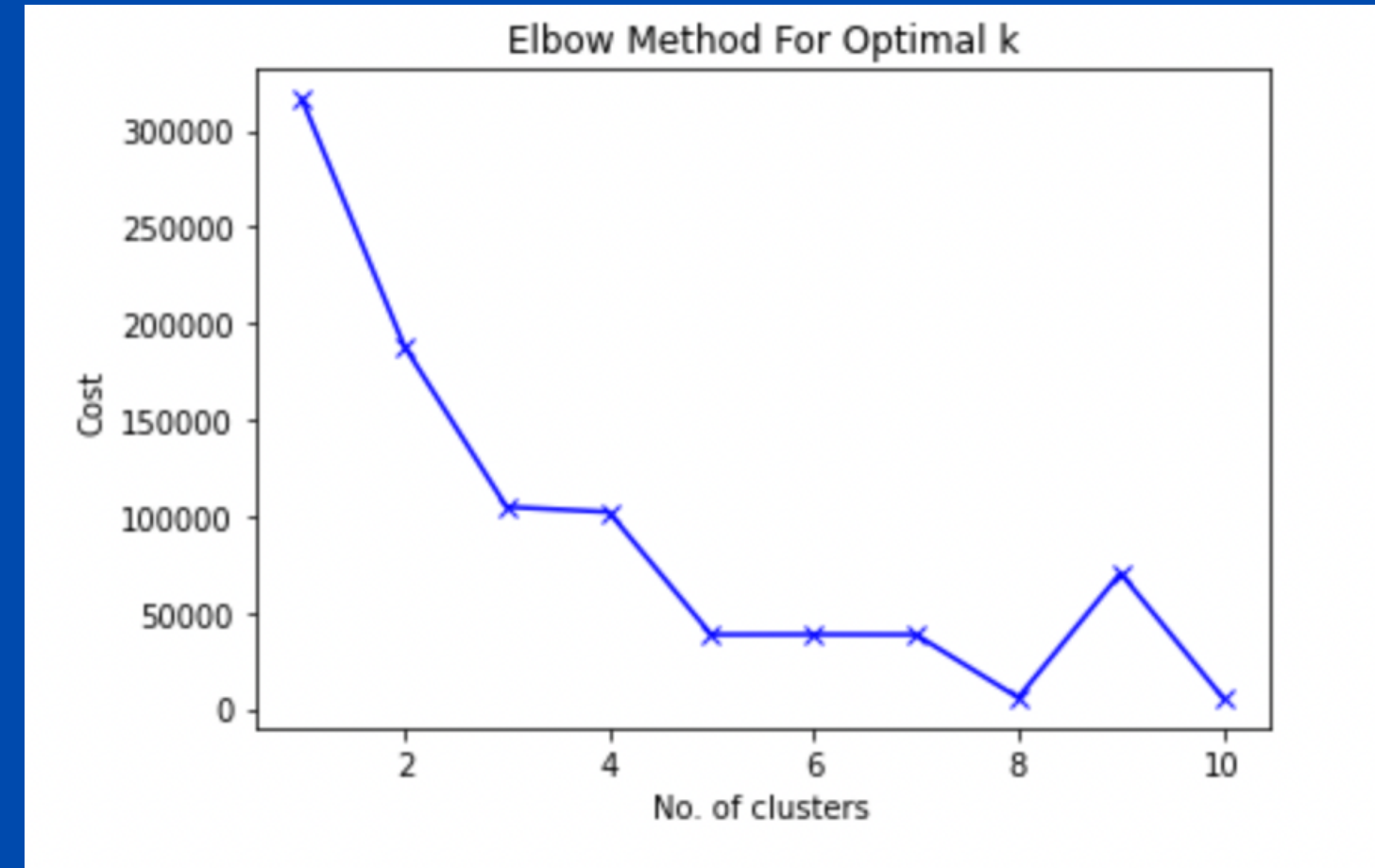
# The TrainDy Recommender<sup>TM</sup> (I)

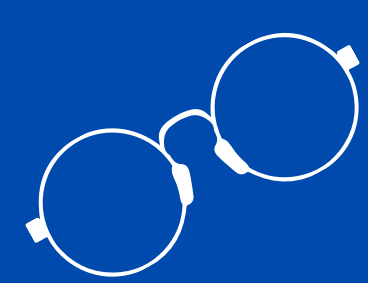
- User Clustering
  - k-means: on a combination of user demographic + purchase history
  - k-modes
  - elbow graphs

## k-means



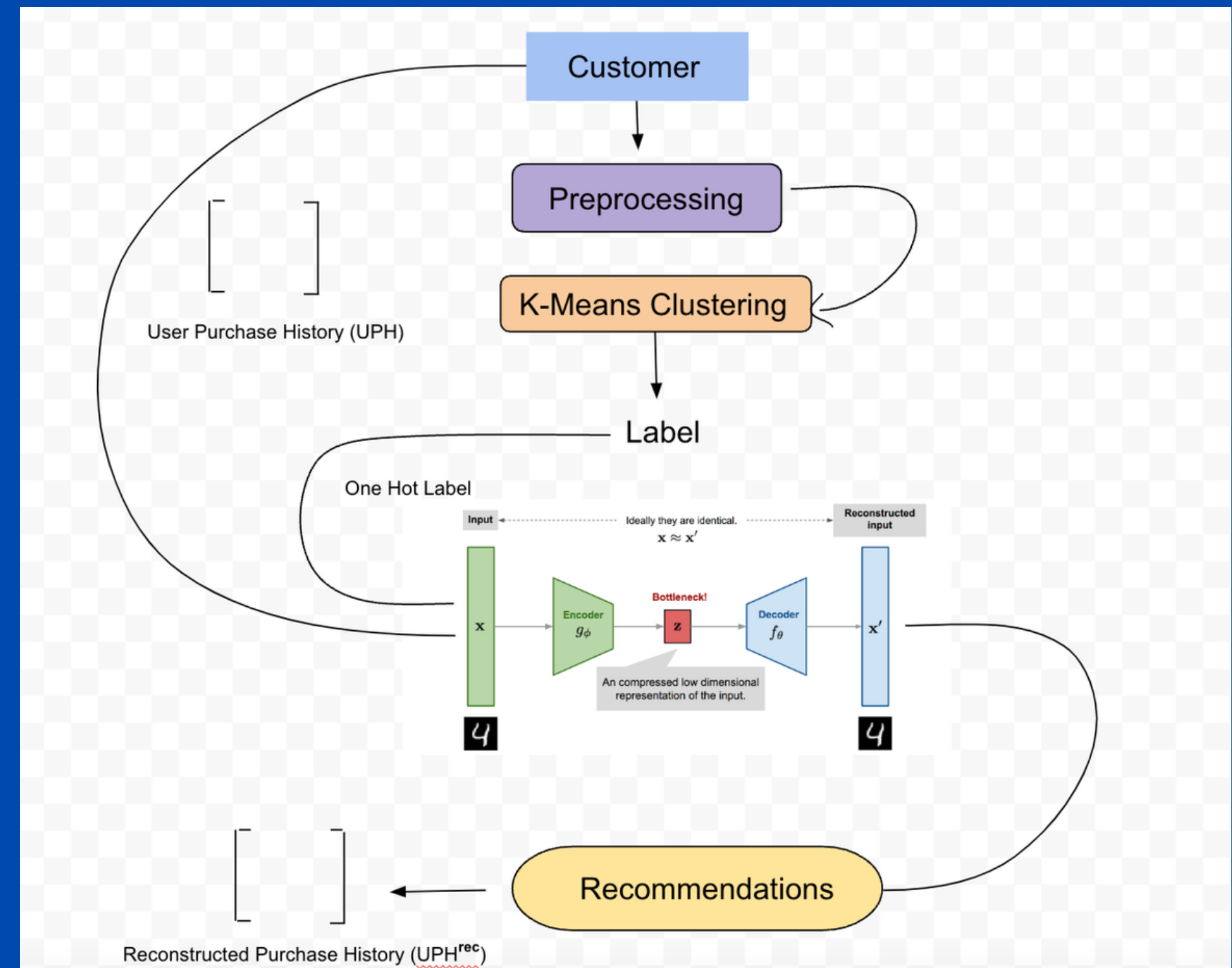
## k-modes





# The TrainDy Recommender<sup>TM</sup> (II)

- Prediction
  - User Purchase History Matrix
  - CVAE using purchase history and user clustering
- Success metric
  - Recall<sub>30</sub>

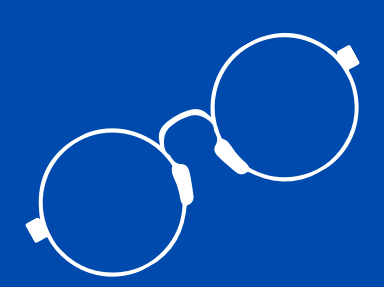




Demo

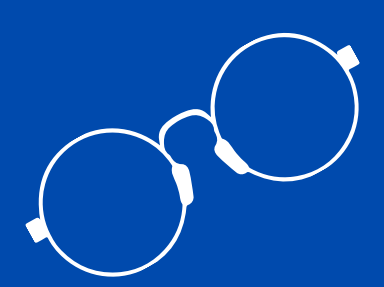


# Results



# Conclusion

- 
- Future work



# THANK YOU!

- We would like to thank Prof. Lorenzo De Stefani and the TAs, especially our mentor TA, Thien Nguyen for their guidance and timely help concerning the project.