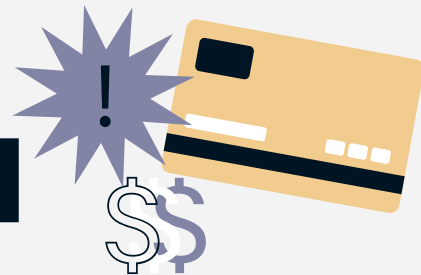




Predicting H&M Customer Value



Team 72

Roopam Chakrabarty, Natania Christopher,
Helen Cunningham, and Janine Mis



CUSTOMER LOYALTY

- Digital advertising has gained market share
- Direct-to-customer advertising has gained in popularity

↑ BRANDS
↓ SWITCHING
COSTS



Cost for Business



Value to Customer

- Determine the factors that truly drive purchase behavior.
- Identify consumers likely to purchase more in the future.





OBJECTIVE

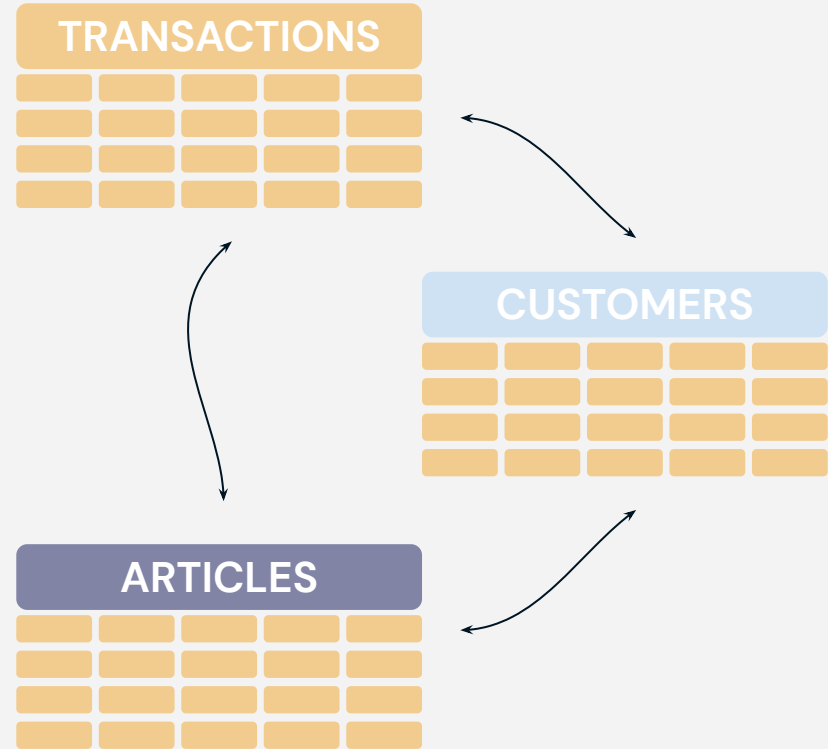
How Much Will A Current Customer Spend At H&M In The Future Based On Their Attributes?

- Can we identify the probability that a customer will be “loyal” (make high amounts of transactions) to the brand?
- Which attributes are most heavily correlated with spending behavior?















THE DATA

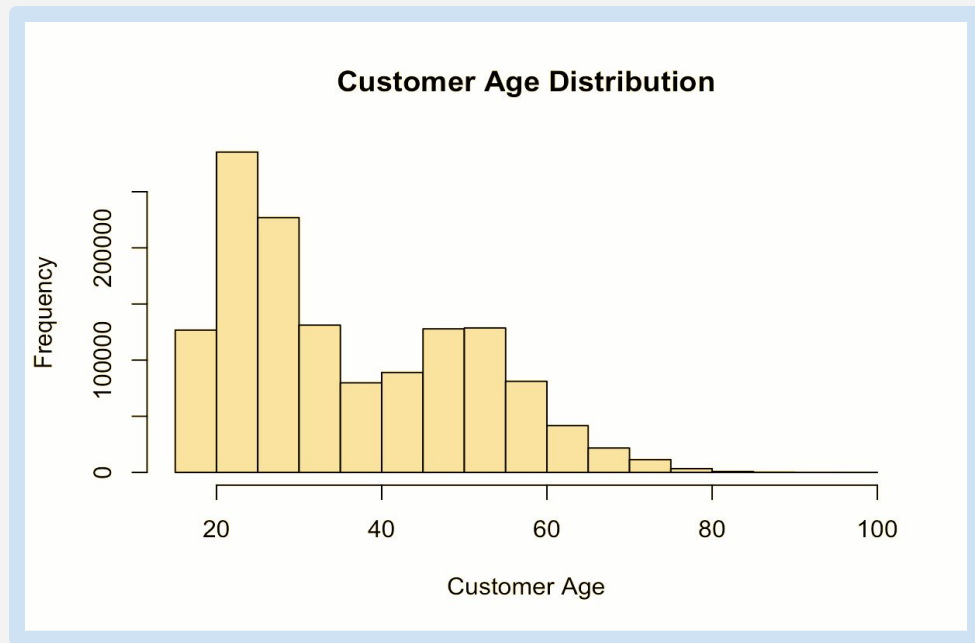
- H&M Data sourced from Kaggle
- Segmented data into two time periods:
 - **T1:** September 2018 through February 2019
 - **T2:** September 2019 through February 2020



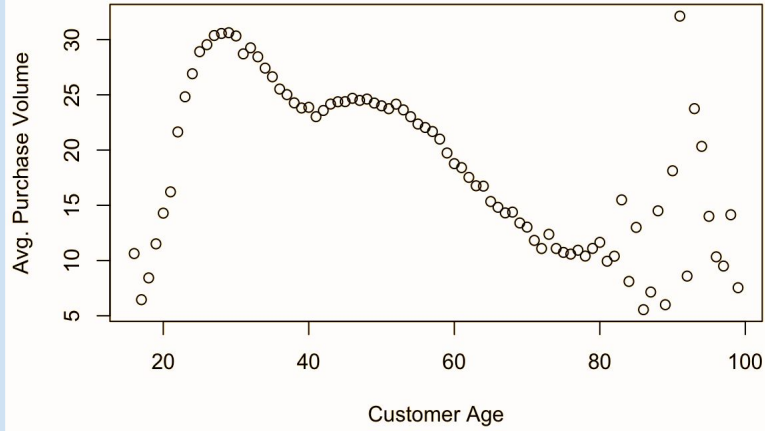
- Hypothesis: Previous purchasing behavior will be a strong predictor of future purchasing behavior. Additionally, club members would be likely to have a higher \$ amount of purchases.

 Receives Newsletter	 Age	 # Transactions Per Month
 Newsletter Frequency	 Postal Code	 Online vs. In-Store
 Club Member Status	 Clothing Category	 \$ Spent In T1
 Club Member Stage	 # Items Purchased	 \$ Spent In T2

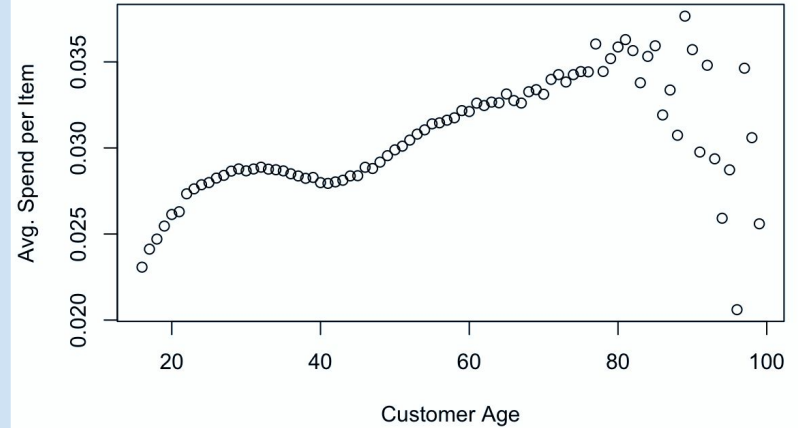
EXPLORATORY ANALYSIS



Average Purchase Volume by Customer Age



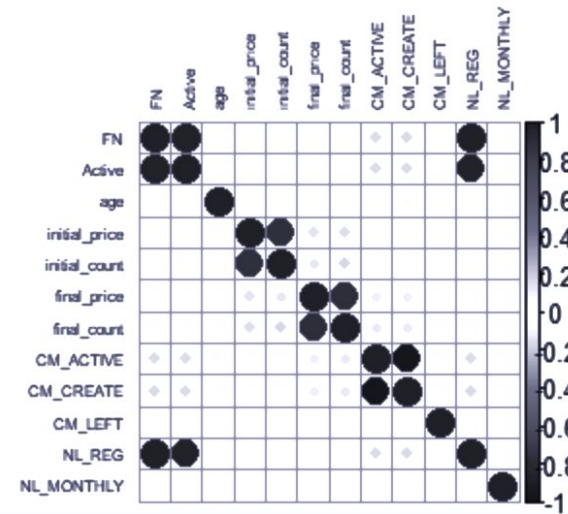
Average Spend per Item by Customer Age



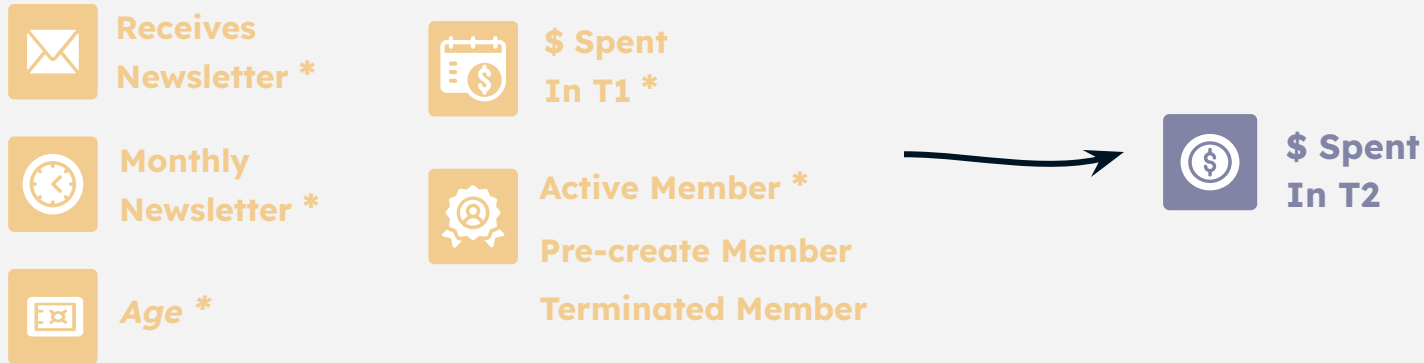
Item prices are masked but can be compared directionally.

The chart displays the volume of transactions in millions, categorized by purchase channel (In Store and Online) over a 24-month period. The Y-axis represents Transactions (Millions) from 0.0 to 2.0. The X-axis represents YearMon from Jul 2018 to Jul 2020. The 'In Store' channel is represented by dark blue bars, and the 'Online' channel is represented by light blue bars. The total transaction volume shows a general upward trend, peaking in mid-2019 and then declining slightly in early 2020.

YearMon	In Store (Millions)	Online (Millions)	Total (Millions)
Jul 2018	0.20	0.40	0.60
Aug 2018	0.40	0.95	1.35
Sep 2018	0.35	0.90	1.25
Oct 2018	0.45	0.70	1.15
Nov 2018	0.35	0.90	1.25
Dec 2018	0.35	0.75	1.10
Jan 2019	0.40	0.90	1.30
Feb 2019	0.35	0.75	1.10
Mar 2019	0.35	0.90	1.25
Apr 2019	0.45	1.00	1.45
May 2019	0.45	1.05	1.50
Jun 2019	0.55	1.35	1.90
Jul 2019	0.55	1.20	1.75
Aug 2019	0.40	0.85	1.25
Sep 2019	0.35	0.85	1.20
Oct 2019	0.35	0.75	1.10
Nov 2019	0.35	0.85	1.20
Dec 2019	0.40	0.65	1.05
Jan 2020	0.35	0.75	1.10
Feb 2020	0.30	0.65	0.95
Mar 2020	0.15	0.90	1.05
Apr 2020	0.40	1.30	1.70
May 2020	0.20	1.15	1.35
Jun 2020	0.40	0.80	1.20
Jul 2020	0.25	0.55	0.80

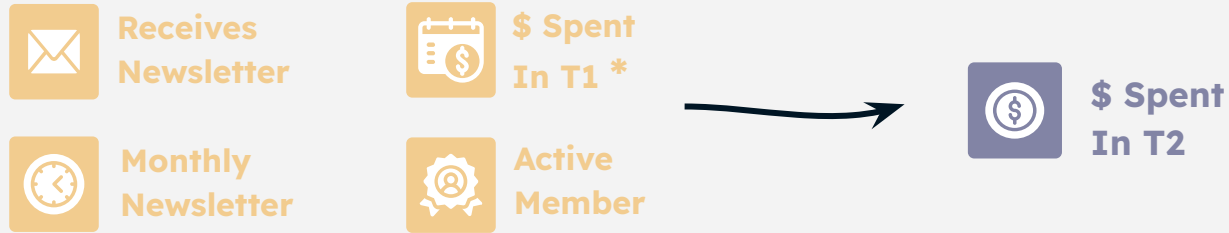


LINEAR REGRESSION #1



- Significant predictors indicated with an [*]
 - The coefficient for **Age** is more than 100x smaller than the other coefficients in the model, so it can be discarded.
-

LINEAR REGRESSION #2



- * The coefficient for **\$ Spent in T1** was the largest.
 - \$1 increase in customer spend in T1 → \$0.60 increase in customer spend in T2.
 - R^2 value: 0.3697
-

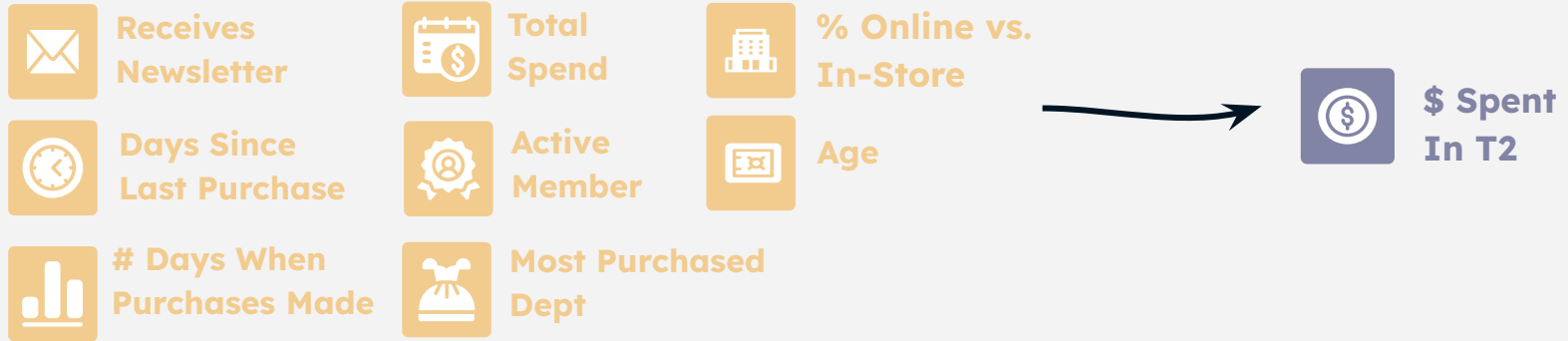
TESTING & VALIDATION

- Divided data into Train (60%), Test (20%) and Validation (20%) sets.
- Created 7 different regression models based on the significant variables and a Random Forest Model.

ERROR METRICS FOR TOP 2 REGRESSIONS & RANDOM FOREST

Model	MSE	RMSE	MAE	MAPE
Model 6	0.0178	0.1334	0.1028	0.9558
Model 7	0.0178	0.1334	0.1028	0.9555
Random Forest	0.0166	0.1290	0.0978	0.9696

ADDITIONAL MODELING



- Step-wise regression was used to select the factors for this model.
- Adj. R-squared was 0.3991.

MSE	RMSE	MAE	MAPE
0.25	0.5	0.29	5.82

MODEL'S MOST VALUABLE CUSTOMERS

- Younger
 - Subscribe to the newsletter
 - Active club members
 - Had high pre-period spend
 - Made purchases on numerous days
 - Made purchases most recently
 - Purchase online
 - Frequently buy ladies' accessories
-

LOGISTIC REGRESSION



Newsletter
Reader Status



Age



T1 Quantity
Purchased



\$ Spent
In T1



Club Member
Status

0.5 Cutoff



**Loyalty: Customers
who purchased 6 or
more items in T2**

73%
ACCURACY

98.5%
SPECIFICITY

5.6%
SENSITIVITY

CONCLUSION



- Club member status, newsletter status, past purchase behavior, age, and purchase channel are strong predictors of future sales.
- Our models can be used to decide how to focus efforts & investments to optimize ROI.

**LOGISTIC
REGRESSION**



Focus Advertising &
Maximize Ad Revenue

**LINEAR
REGRESSION**



Identify High-Value Customers &
Steer Product Development

References

1. Huang, Jess, and Thomas O'Toole. "Customer Loyalty: The New Generation." *McKinsey & Company*, McKinsey & Company, 14 Aug. 2020, <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/customer-loyalty-the-new-generation>.
2. Katz, Elie Y. "Council Post: How to Build Customer Loyalty." *Forbes*, Forbes Magazine, 10 June 2022, <https://www.forbes.com/sites/forbesbusinesscouncil/2022/06/09/how-to-build-customer-loyalty/?sh=78bee284b0ae>.
3. Polkes, Aliza. "The State of Brand Loyalty 2021: Global Consumer Survey." *Yotpo*, 28 Dec. 2022, <https://www.yotpo.com/blog/the-state-of-brand-loyalty-2021-global-consumer-survey/>.

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

H&M Logo: brandslogos.com/fashion/hm-logo-black-and-white/



