Table of Contents

01

Focus Areas

Guiding questions

02

Data Preparation

Data cleaning, data manipulation, etc.

03

Modeling

Markov chains, classification models

04

Recommendations

Examination of findings

01
Focus Areas

Guiding Questions

Customer Behavior

What do typical customer behavior patterns look like? Are certain actions on Fingerhut's platform more or less likely to lead to another action?
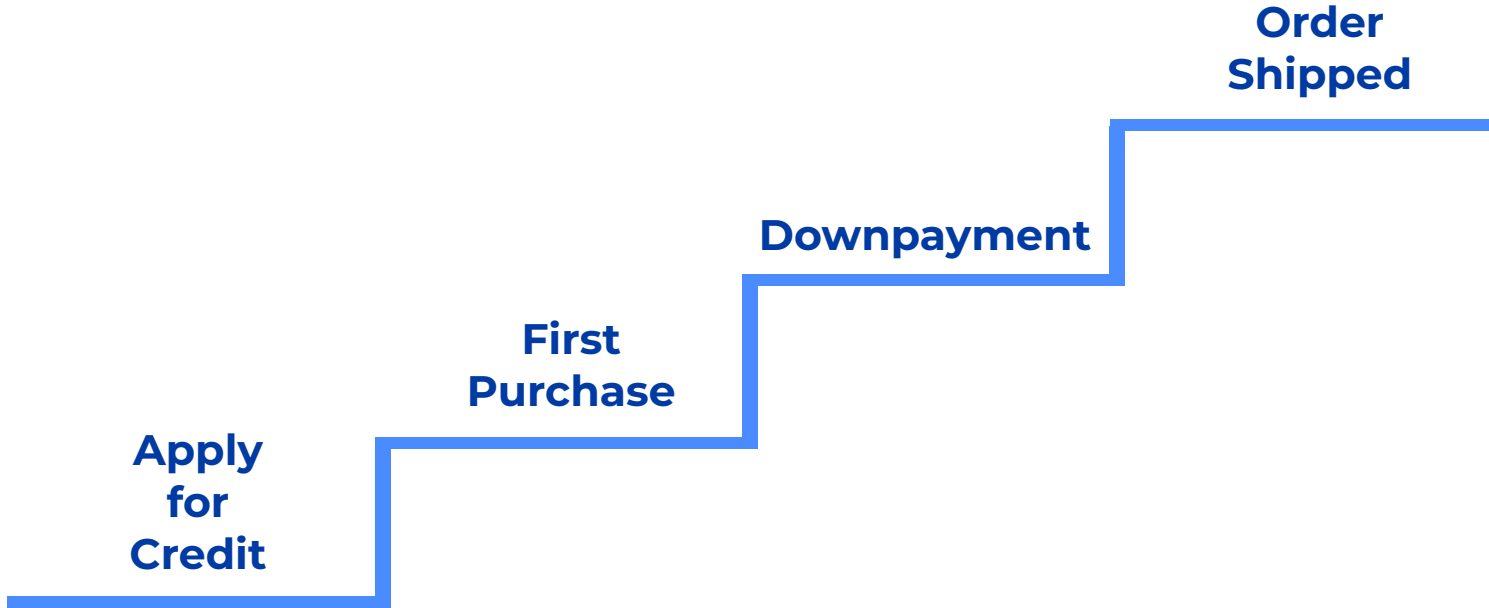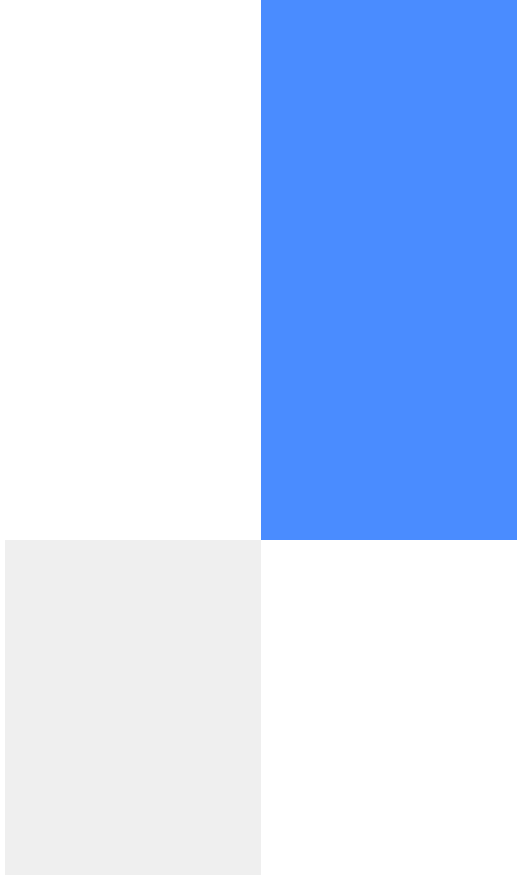
Ideal Journeys

Which customer behaviors or features make them more likely to follow the "ideal journey" as outlined by Fingerhut?

Promotions

Which promotional materials launched by Fingerhut are more likely to result in initial purchases by customers?
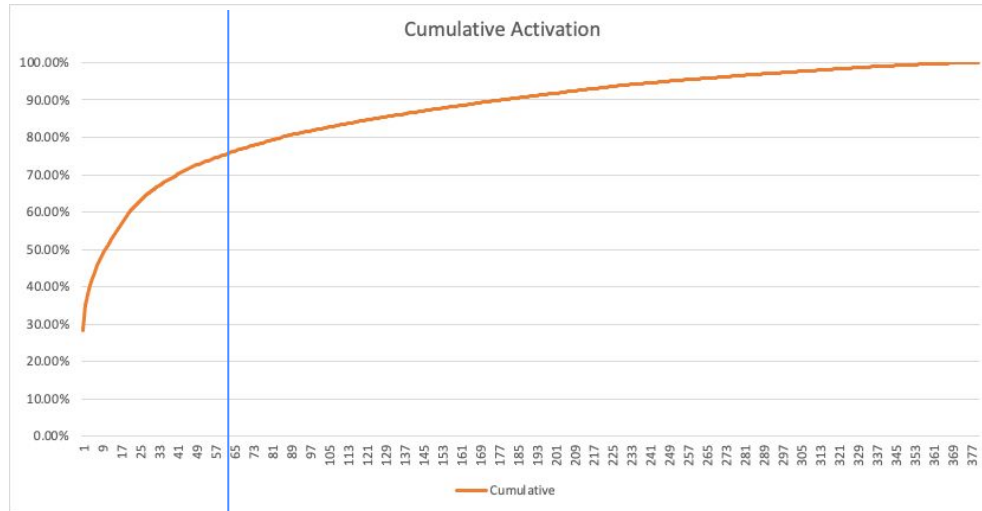
02

Data Preparation

Data Cleaning

- **Removed duplicates**
  - Reindexed journey_steps_until_end
- **Removed possibly incomplete customers**
  - Customers who activated account in the last 60 days, but not yet placed an order

Wide-Format Data

	customer_id	account_id	ed_id	event_name	event_timestamp	journey_steps_until_end
0	-2147206560	2082689427	[12, 1, 19, 5, 11, 3, 4, 6, 4]	[application_web_approved, promotion_created, ...	[2023-05-02 20:20:18+00:00, 2023-05-02 20:58:1...	[1, 2, 3, 4, 5, 6, 7, 8, 9]
1	-2145360520	1467252181	[19, 19, 19, 19, 19, 19, 3, 19, 12, 4, 4, ...	[application_web_view, application_web_view, a...	[2022-01-08 00:40:01+00:00, 2022-01-08 00:40:0...	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
2	-2144991170	-2081722203	[19, 19, 19, 19, 19, 3, 19, 12, 4, 4, 4...	[application_web_view, application_web_view, a...	[2022-09-19 14:47:59+00:00, 2022-09-19 14:48:0...	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
3	-2142086624	-484534102	[2, 12, 4, 4, 4, 5, 6, 1, 1, 6, 24, 1, 21	[campaign_click, application_web_approved, bro...	[2021-07-07 06:00:00+00:00, 2021-07-07 18:58:0...	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...

ed_id	stage	event_timestamp	application	activation	promotion_exposure	place_order_web	place_order_phone	order
[2, 12, 1, 4, 4, 4, 11, 1, 5, 6, 1, 1, 4, 11, ...	[Discover, Apply for Credit, Promotion Created...	[2021-09-03T06:00:00.000Z, 2021-09-03T21:46:20...	1	1	1	1		0
[2, 4, 11, 4, 4, 11, 4, 11, 4, 5, 4, ...	[Discover, First Purchase, First Purchase, Fir...	[2021-02-20T06:00:00.000Z, 2021-02-20T22:23:39...	1	1	1	1		0
[19, 19, 19, 19, 19, 19, 3, 19, 12, 4, ...	[Apply for Credit, Apply for Credit, Apply for...	[2022-12-11T07:43:02.000Z, 2022-12-11T07:43:04...	1	0	1	0		0
[12, 2, 22]	[Apply for Credit, Discover, Discover]	[2023-07-02T13:43:31.000Z, 2023-07-02T19:43:31...	1	0	1	0		0

03
Modeling

Journey Analysis (Events)

Wanted to incorporate all events:

- Utilized a random sample of 83,284 unique customer ids with 3,246,341 corresponding journey steps
- Assumed customers cannot repeat the same event in succession

Results:

- Many transitions provided expected results but could be optimized further (i.e. 52% of the time browse_products was immediately followed by add_to_cart)

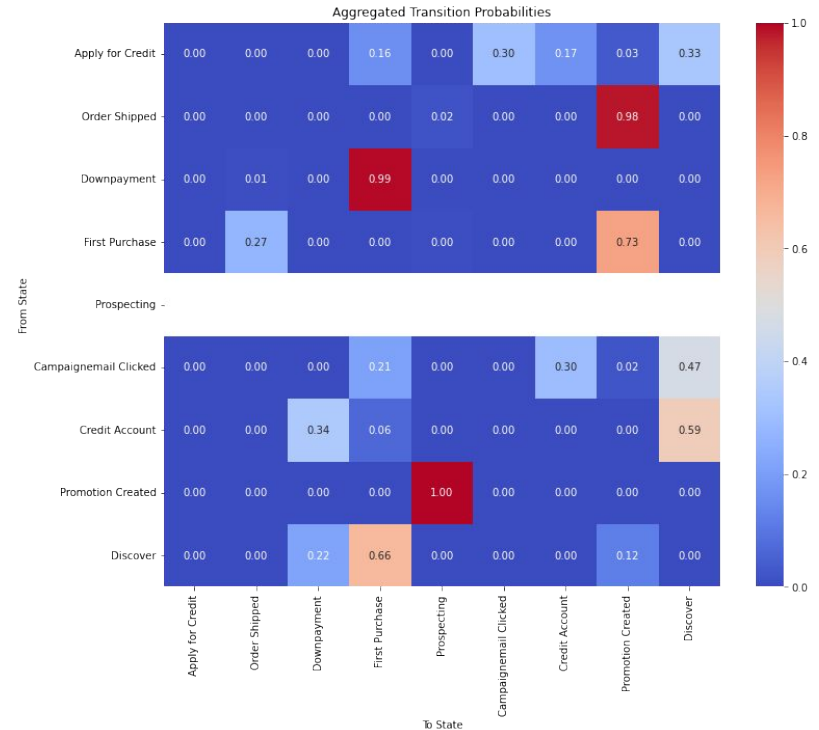Markov Chain Transition Matrix Heatmap

Journey Analysis (Stages)

Aggregated Events to Stages:

- Unique stages 'Promotion Created', 'Campaignemail Clicked' and 'Order Shipped' were created as they were not accounted for in the event_definitons aggregations
- Assumed customers cannot repeat the same stage in succession for continuity

Results:

- Prospecting stage irrelevant
- Discover is a significant step for a successful journey
 - optimize actions that push people at these junctions

Aggregated Transition Probabilities

From State \ To State	Apply for Credit	Order Shipped	Downpayment	First Purchase	Prospecting	Campaignemail Clicked	Credit Account	Promotion Created	Discover
Apply for Credit	0.00	0.00	0.00	0.16	0.00	0.30	0.17	0.03	0.33
Order Shipped	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.98	0.00
Downpayment	0.00	0.01	0.00	0.99	0.00	0.00	0.00	0.00	0.00
First Purchase	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.73	0.00
Prospecting									
Campaignemail Clicked	0.00	0.00	0.00	0.21	0.00	0.00	0.30	0.02	0.47
Credit Account	0.00	0.00	0.34	0.06	0.00	0.00	0.00	0.00	0.59
Promotion Created	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Discover	0.00	0.00	0.22	0.66	0.00	0.00	0.00	0.12	0.00

Classification Model: Ideal vs Non-ideal Journey
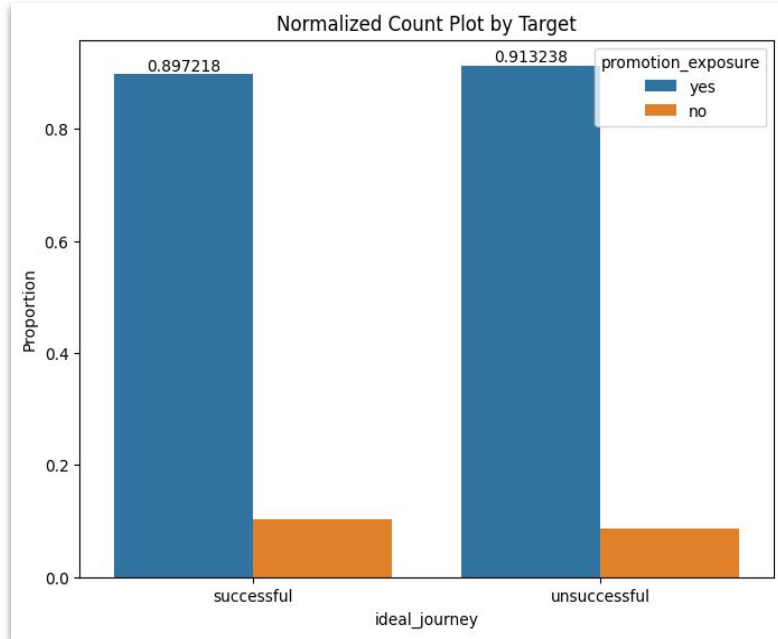
Wanted to focus on single journeys:

- Treat each unique customer_id and account_id combination as a single customer
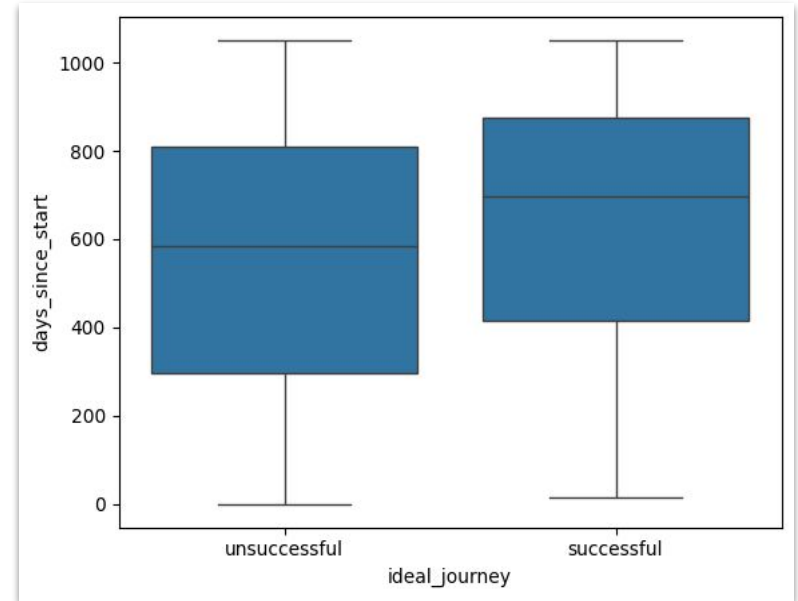- Removed customers with multiple activations

Feature engineering:

- Included only features that could be extracted from a "new" customer
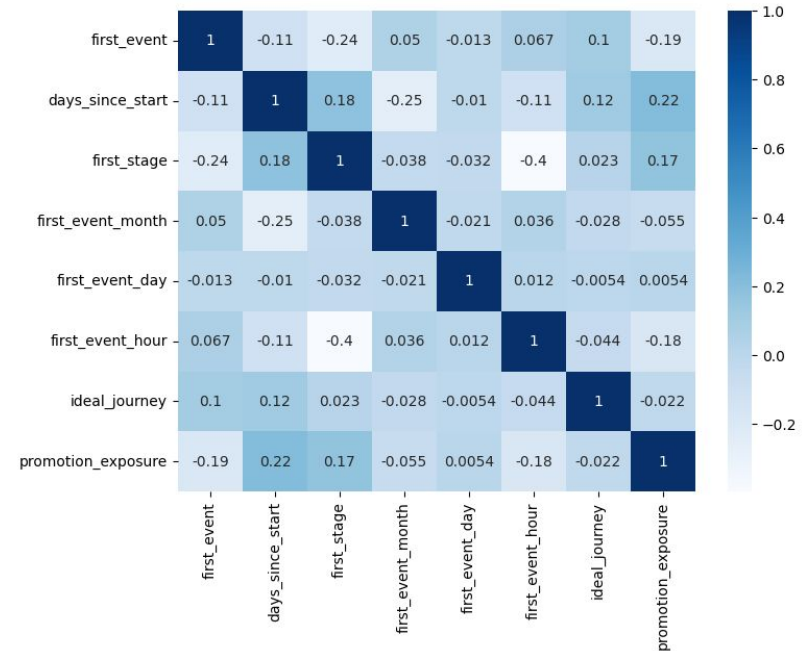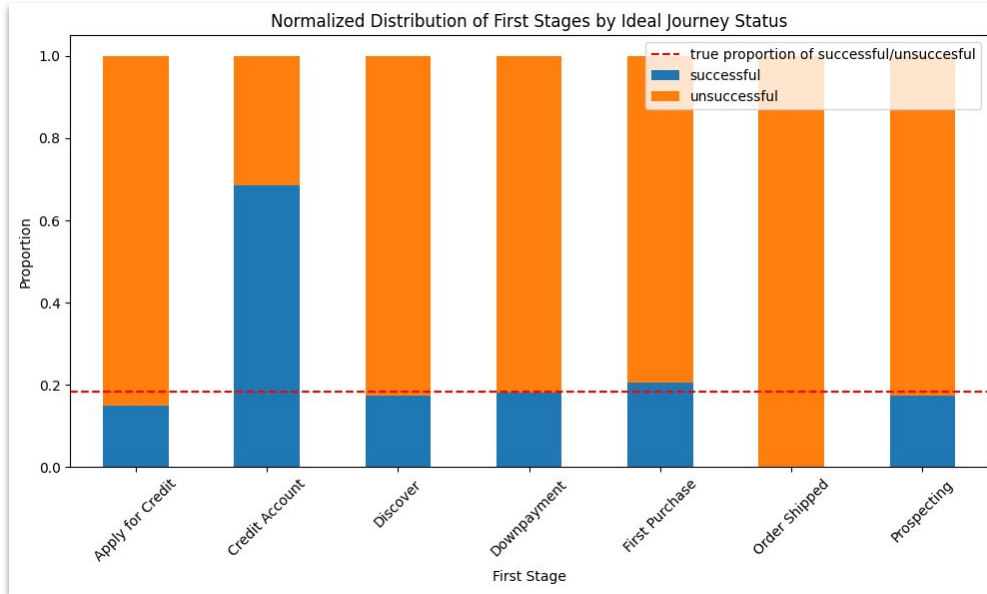
EDA of Feature-Engineered Data

Lower percentage of customers with successful ideal journeys were exposed to promotions
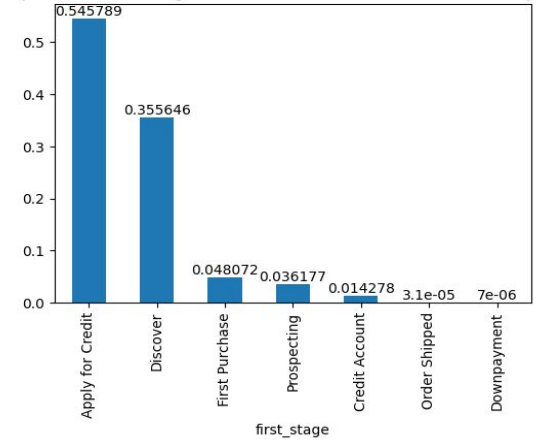

Successful journeys typically started earlier than unsuccessful journeys
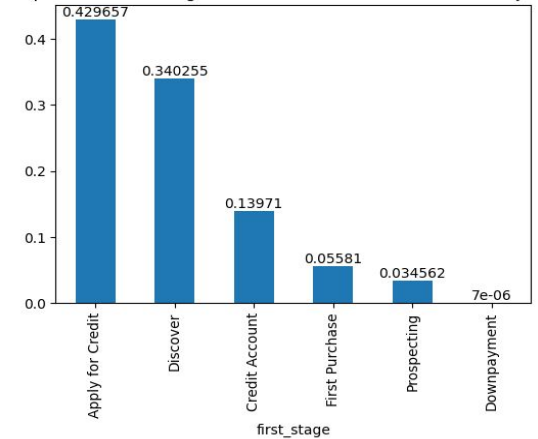
Examining First Stage
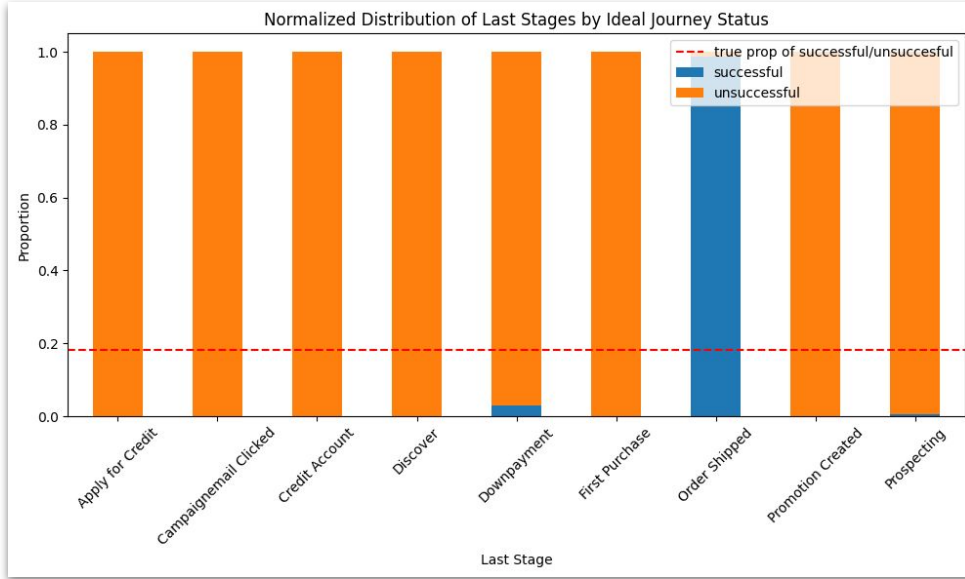

Most successful journeys start with 'Credit Account'

Examining Last Stage

Normalized Distribution of Last Stages by Ideal Journey Status
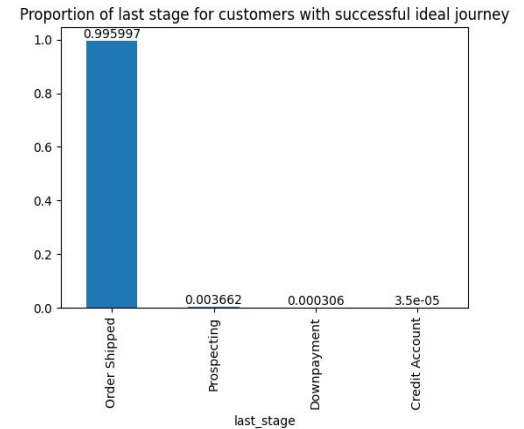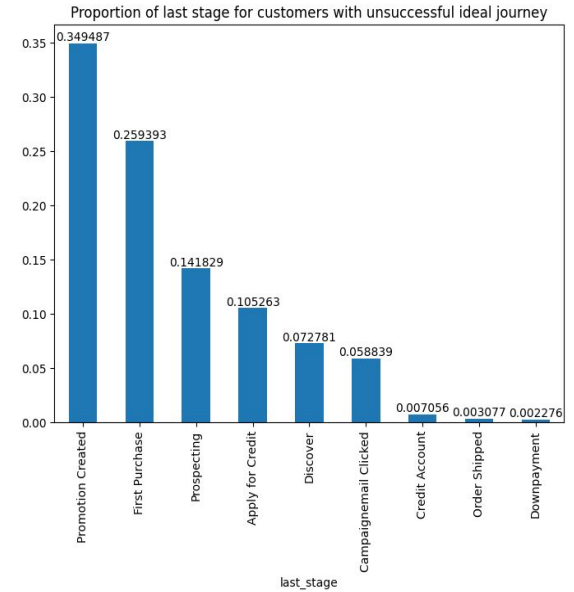

Proportion of last stage for customers with unsuccessful ideal journey

Proportion of last stage for customers with successful ideal journey

Most unsuccessful journeys end with an event in the promotion created stage, or first purchase stage
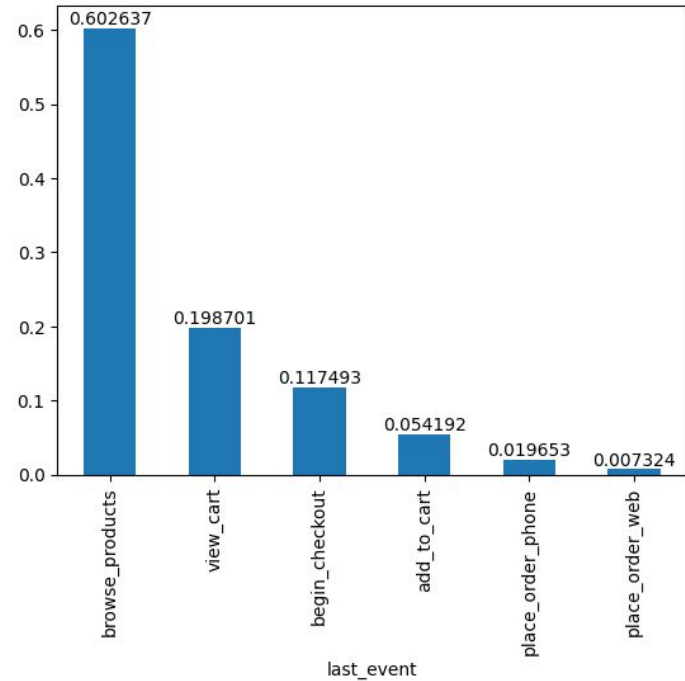
Examining Last Stage: First Purchase

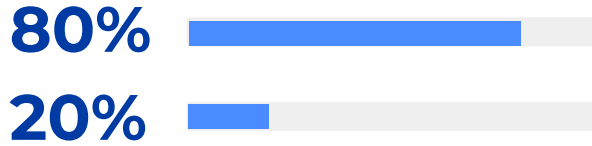**Of all the customers who ended with 'First Purchase' as their final stage:**

- Most stopped after browsing products or viewing their cart
- Very few made it to checking out

Distribution of final event for customers with First Purchase as final stage
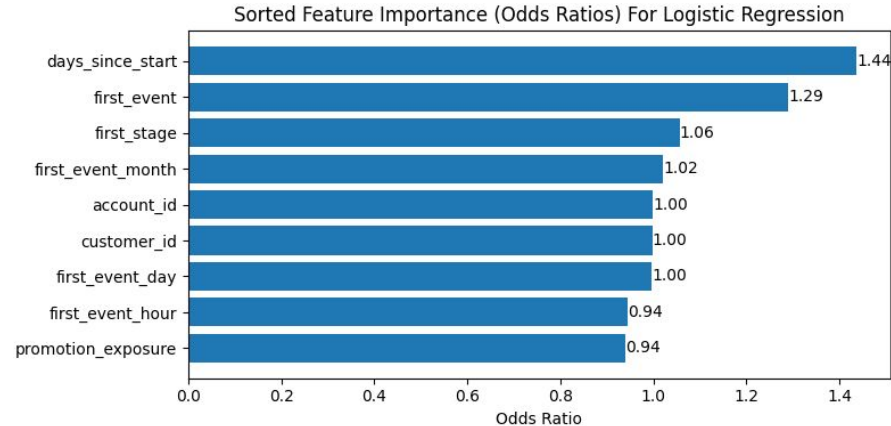
Logistic Regression Model

Ideal Journeys

80%

20%

Limitation: Dataset was extremely imbalanced - used class_weight = 'balanced' parameter

- uses the values of y to adjust weights inversely proportional to class frequencies

	precision	recall	f1-score	support
unsuccessful	0.86	0.57	0.68	465695
successful	0.23	0.58	0.33	104706
accuracy	-	-	**0.57**	570401
macro avg	0.55	0.58	0.51	570401
weighted avg	0.74	0.57	0.62	570401

Sorted Feature Importance (Odds Ratios) For Logistic Regression

Feature	Odds Ratio
days_since_start	1.44
first_event	1.29
first_stage	1.06
first_event_month	1.02
account_id	1.00
customer_id	1.00
first_event_day	1.00
first_event_hour	0.94
promotion_exposure	0.94

Odds Ratio

Odds ratio interpretation: for every 1 day increase in days_since_start, the odds of a successful journey increase by 1.44 times.

Random Forest Model

Results of best model after hyperparameter tuning + cross validation

	precision	recall	f1-score	support
unsuccessful	0.87	0.67	0.76	465695
successful	0.27	0.55	0.36	104706
accuracy	-	-	**0.65**	570401
macro avg	0.57	0.61	0.56	570401
weighted avg	0.76	0.65	0.68	570401

Promotion Exposure Events

ed_id	event_name
1	Promotion Created
2	campaign_click
9	customer_requested_catalog _digital
20	catalog_email_experian
21	catalog_mail
24	Campaignemail Clicked

Promotion & Activation Paths

Total

1,665,431

Promotion Exposure

1,523,909

Activation 372,432

No Activation 1,151,477

No Promotion Exposure

141,522

Activation (24%)
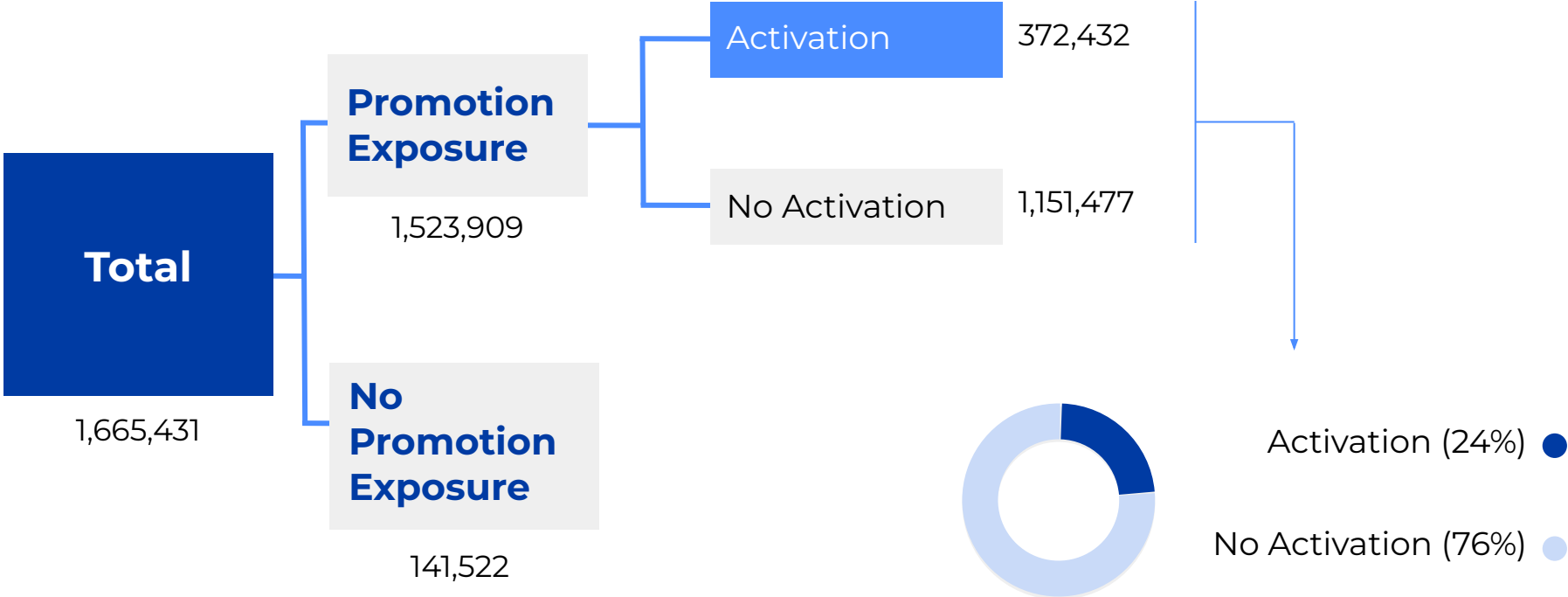
No Activation (76%)

Percentages by activation & purchases

21%
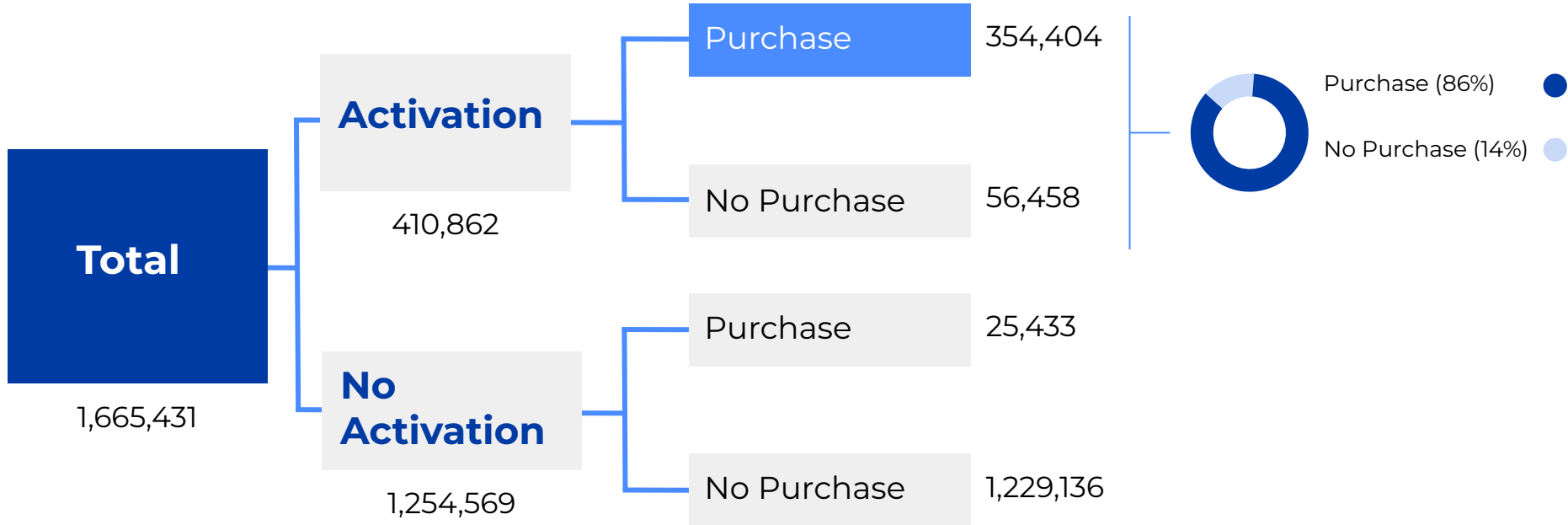
Total users with **activation** & **purchase**

3%

Total users with **activation** & **no purchase**

76%

Total users with **no activation** & **no purchase**

Activation & Purchase Paths

Total

1,665,431

Activation

410,862

Purchase — 354,404

No Purchase — 56,458

No Activation

1,254,569

Purchase — 25,433

No Purchase — 1,229,136

Purchase (86%)

No Purchase (14%)

Identifying Effect of Promotion Exposure on Activation Rates

w/ Logistic Regression

	precision	recall	f1-score	support
Activation	0.86	0.50	0.64	250911
No activation	0.33	0.75	0.46	82176
accuracy	-	-	0.57	333087
macro avg	0.60	0.63	0.55	333087
weighted avg	0.73	0.57	0.59	333087

Predictor (ed_id / event_name)		Coefficient	Odds Ratio
24	**Campaignemail Clicked**	**0.6607**	**1.9362**
1	**Promotion Created**	**0.3192**	**1.3760**
-	promotion_y_n	0.0378	1.03855
9	customer_requested_catalog_digital	0.0	1.0
2	campaign_click	-0.1431	0.8666
20	catalog_email_experian	-0.3926	0.6753
21	catalog_mail	-1.4094	0.2443

04

Recommendations
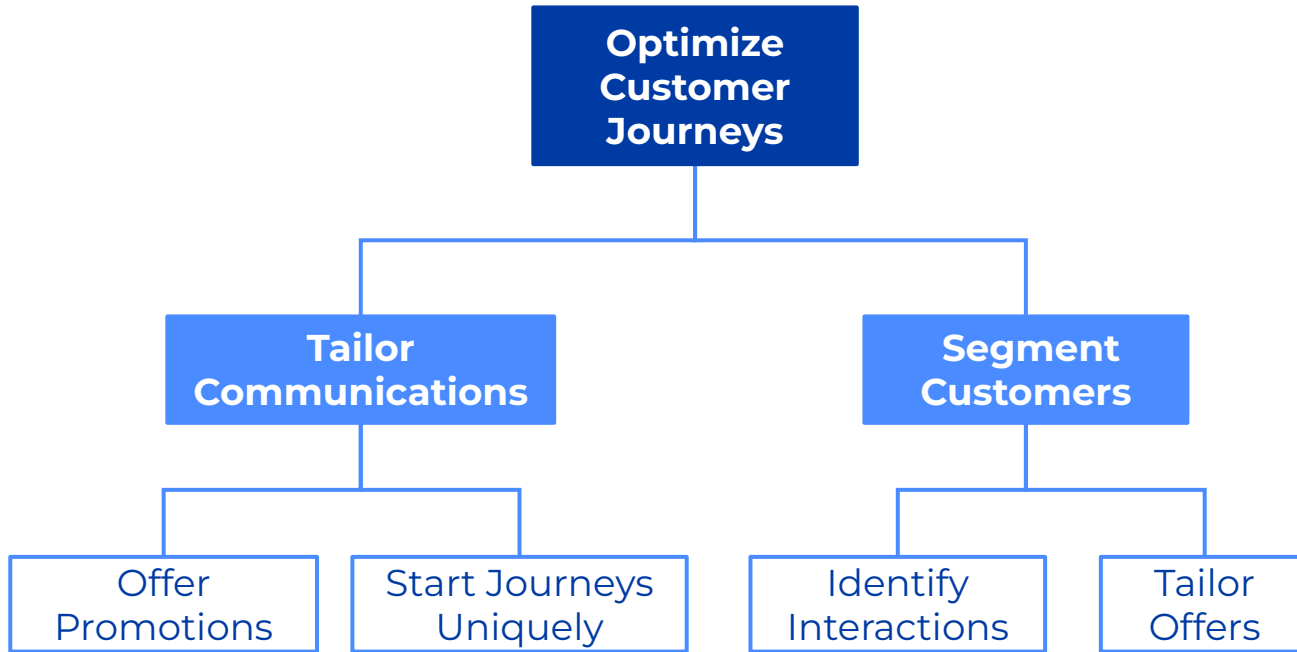
Strategy Map

Journeys	Learn causation behind crucial junctions	Tailor communication based on stage
Promotions	Keep more detailed record on promotion types	Experiment with more email campaigns
Learning & Growth	Expand Fingerhut inventory	Improve customer information

Questions?

Thank you!