General Dilemmas of Fingerhut Design and Suggestions

Charles Barnes, Christopher Fu, Zach Godick, and Han Qiang

Content of Presentation

1. **Overview**
   Project Introduction
2. **Question #1**
   Related Work
3. **Question #2**
   Related Work
4. **Question #3**
   Related Work
5. **Conclusion**
   Summary + Next Steps

Guiding Questions

- How optimal is Fingerhut's ideal path for customer journeys?
- How can we maximize the customer conversion rate by predicting customer behaviors?
- How can we maximize a customer achieving the "Order shipped" milestone using their interaction with intermediate steps?

Question #1 - Data Clean Up and Methods

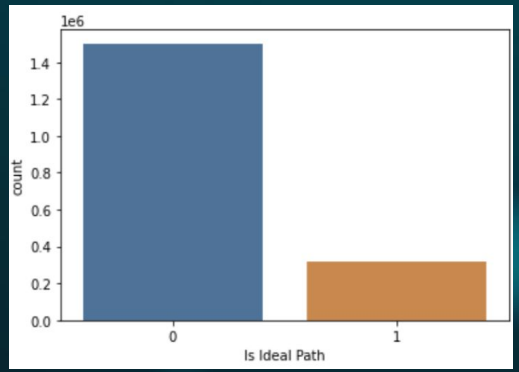
1. Data Cleaning
   - Table merge main dataset with Event Definitions csv file to get stages
   - Assign journey # based on when customer restarts (journey_steps_until_end = 1)
   - Group by customer and aggregate stages into list

	customer_id	journey	stage	Path
0	-2147483541	1	['Apply for Credit', 'First Purchase', 'First ...	[12, 5, 4, 11, 5, 6, 1, 4, 11, 4, 11, 5, 6, 5,...
1	-2147481037	1	['Apply for Credit', 'misc', 'Discover', 'Firs...	[12, 1, 2, 4, 11, 6, 5, 6, 6, 5, 5, 4, 4]
2	-2147474335	1	['Discover', 'Apply for Credit', 'Apply for Cr...	[2, 12, 19, 19, 19, 19, 19, 19, 19, 19, 19, 4,...
3	-2147474305	1	['First Purchase', 'Apply for Credit', 'Apply ...	[6, 19, 19, 19, 19, 19, 19, 19, 19, 1, 4, 19, ...
4	-2147472167	1	['Apply for Credit', 'Apply for Credit', 'Appl...	[12, 19, 3, 19, 19, 19, 19, 19, 19, 19, 19, 19...
...	...	...	...	...
1818995	2147480051	1	['Prospecting', 'Discover', 'Discover', 'Apply...	[21, 2, 22, 12, 24, 1, 21, 1, 1, 21, 1, 1, 1, 1]
1818996	2147480182	1	['Apply for Credit', 'Discover', 'misc', 'Firs...	[12, 2, 24, 4, 4, 4, 4, 4, 4, 24, 24, 24, 24, 24]
1818997	2147480920	1	['Discover', 'Apply for Credit', 'Apply for Cr...	[2, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 3,...
1818998	2147482120	1	['Apply for Credit', 'Apply for Credit', 'Appl...	[12, 19, 19, 2]
1818999	2147483471	1	['Apply for Credit', 'First Purchase', 'First ...	[12, 4, 4, 4, 1, 21, 1, 1, 21, 1, 1, 1, 1]

Stages	Features
Apply for Credit	1,2,3
First Purchase	4,5
Misc	6
Prospecting	7
Discovery	8
Downpayment	9
Credit Account	10
Order Shipped	11,12

3. Feature Engineering
   - Vectorize string input of stage lists
   - 80/20 Train-Test split
   - Oversampling
   - PCA (3rd model only)

2. Labeling
   - IDEAL PATH: Apply for credit > Make a first purchase > Make the down payment > Order Ships
   - Scoring based on if all stages appear and are in that order
   - 1/0 for binary classification



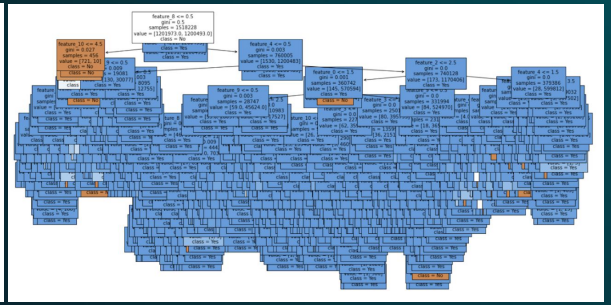
Question #1 - Results and Main Insights

Logistic Regression

```
Classification Report with L2 Regularization:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    300175
           1       0.99      1.00      0.99     63625

    accuracy                           1.00    363800
   macro avg       0.99      1.00      1.00    363800
weighted avg       1.00      1.00      1.00    363800

Confusion Matrix with L2 Regularization:
[[299523    652]
 [     0  63625]]
```

- Overfitting even with hyper-parameter tuning
- Not informative on its own

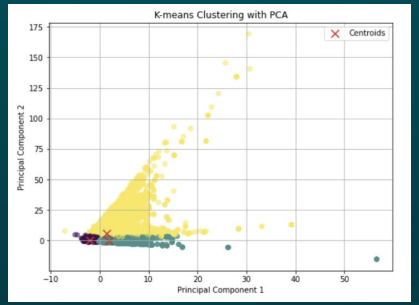
Random Forest



- Most Important Features: First Purchase, Down Payment, and Order Shipped
- Apply for Credit valued lower

Feature Importances	Value
Feature 1	0.1340
Feature 2	0.0001
Feature 3	0.0357
Feature 4	0.0000
Feature 5	0.2503
Feature 6	0.0249
Feature 7	0.0003
Feature 8	0.0014
Feature 9	0.2891
Feature 10	0.0032
Feature 11	0.0138
Feature 12	0.2472

K-Means Classifier



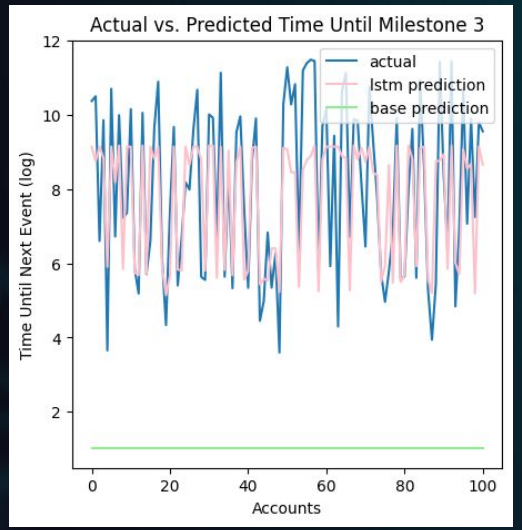
- K-Means Clustering + Logistic Regression
- Prospecting had largest variance
- Uneven density of data

Features	PCA 1	PCA 2
Feature 1	0.45549	-0.11045
Feature 2	0.09983	0.55710
Feature 3	0.11997	0.55105
Feature 4	-0.00433	0.00362
Feature 5	0.41625	-0.10800
Feature 6	0.25416	0.08294
Feature 7	0.09983	0.55710
Feature 8	-0.03740	0.06918
Feature 9	0.46086	-0.12794
Feature 10	-0.17186	0.02161
Feature 11	0.25416	0.08294
Feature 12	0.46086	-0.12794

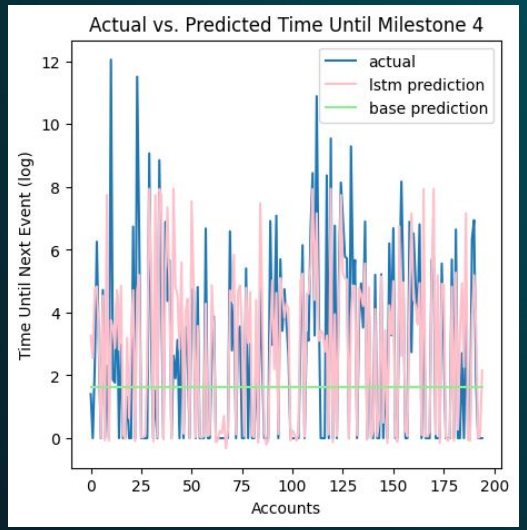
Question #2 - Data Clean Up and Methods

- Select ed_id corresponds to milestones:
  - Milestone 1 - 12; Milestone 2 - 7; Milestone 3 - 29; Milestone 4 - 8; Milestone 5 - 27
- Feature engineering:
  - calculate duration between the current and the previous milestones and creates time_elapsed variable
- Data manipulation:
  - Identify and address outliers in time_elapsed
  - Applied log transformation to time_elapsed
- Model:
  - Average base model as benchmark
  - LSTM model

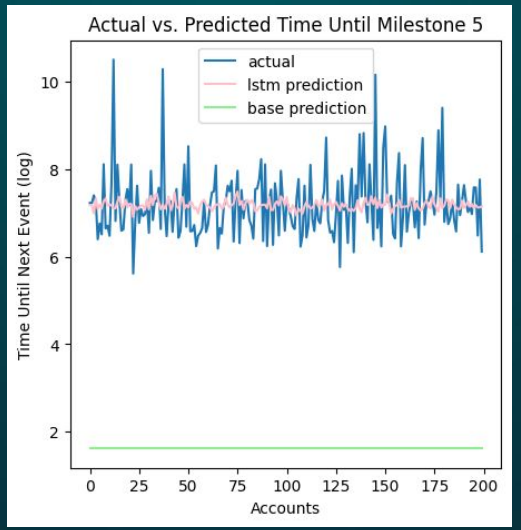
Question #2 - Results and Main Insights



LSTM Model Loss: 2.34
Base Model Loss: 54.38



LSTM Model Loss: 3.78
Base Model Loss: 9.89



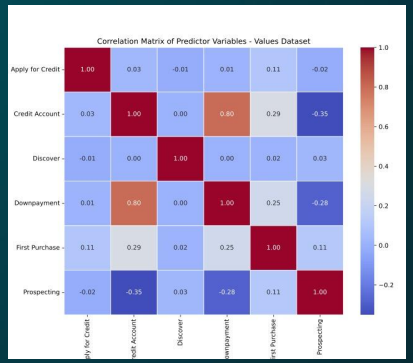
LSTM Model Loss: 0.55
Base Model Loss: 31.72

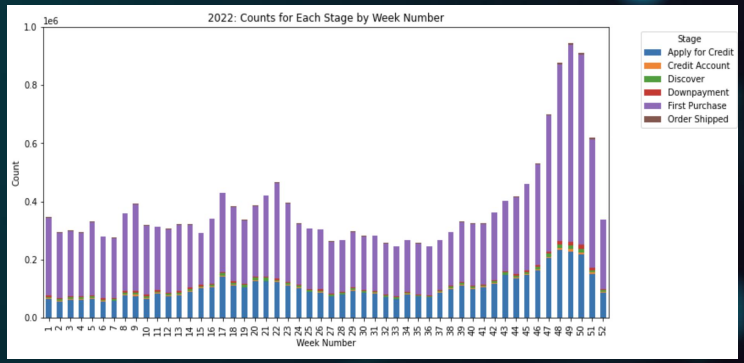
Question #3 - Data Clean Up and Methods

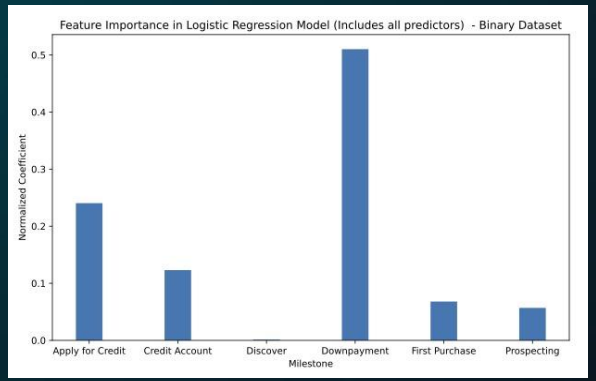
- Journey Definition - 60 days, grouping them by unique customer ID and account ID pairs.
- Data manipulation - filtering, table pivoting, merges, grouping, data type conversions, NA value removal, etc.
- Feature Engineering - Create our new variables from timestamp data and from the existence of stages and milestones in each journey.
- Dataset Creation - Binary and Numerical Variables for the existence of milestone events during each journey
- Model creation and training - Logistic Regression, Random Forest Decision Trees

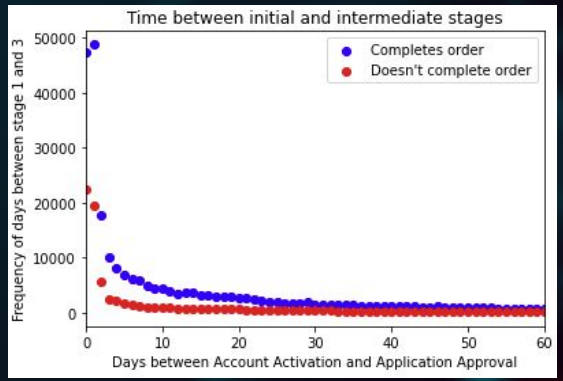
Question #3 - Visualizations

- Correlation Matrix of Predictor Variables
- Week over Week analysis of Journey Stages during 2022
- Feature Importance in Logistic Regression Prediction Model
- Frequency chart for Predicting the influence of intermediate steps on journey completion









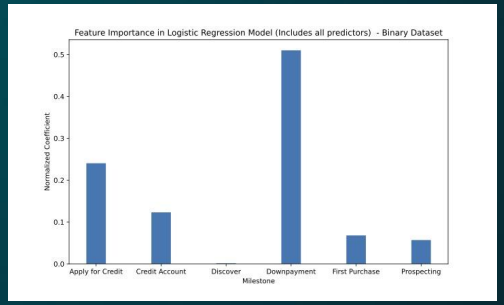
Question #3 - Results and Main Insights

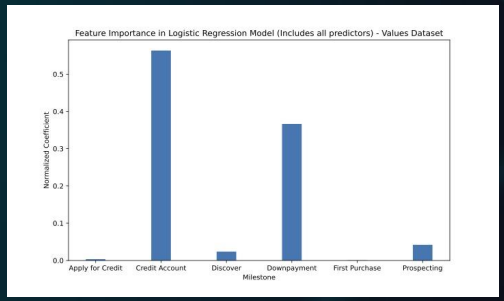
Time between Stage 1 (Application Approval) and Stage 3 (Credit Activation)

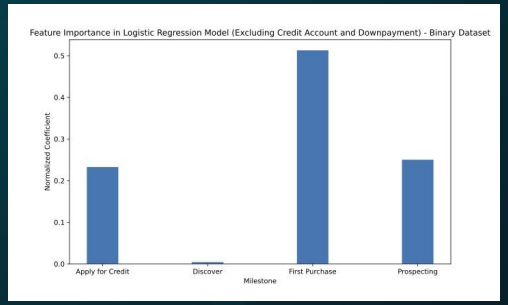
	Order Shipped	Not Shipped
# Journeys	249718	78417
Mean	11.146537	8.0734
St. Dev	14.846087	13.367
IQR	1 - 17 days	0 - 10 days

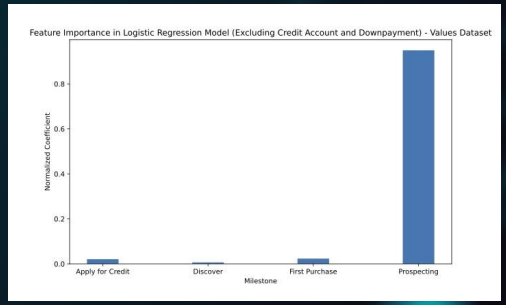
Testing Accuracy

Baseline	0.7644719
Logistic Regression	0.7651449
Random Forest	0.7677470









Logistic Regression Models Illustrate that “Downpayment” was the main predictor of “Order Shipped,” followed by “Credit Account” and “Apply for Credit”

Summary and Next Steps

- "First Purchase," "Downpayment," and "Orders Shipped" were the greatest contributing factors for an ideal path
  - "Prospecting" may be important considering the magnitude of variance it had on the PCA components
  - The frontend design of the website could have played a role in dictating where customers would click to interact with the website
  - Next steps: expanding into multi-class classification of ideal journey paths and degrees of ideal
- LSTM model is a viable method for predicting the time until the next milestone, potentially aiding in enhancing the customer conversion rate.
  - Next steps: Adopting this approach or similar methods to study customer behaviors while addressing the limitations concerning model accuracy / extreme cases; consider more accounts as well as hyperparameter tuning
- Used *event_timestamp* variable to determine hotspots in the customer interaction cycle
  - Next steps: Maximize the percentage of customers that complete their credit rehabilitation journey and graduate to the next program "Fetti"
- Logistic Regression models suggest that "Downpayment" and "Credit Account" are highly correlated and predictive of "Order Shipped"
  - When considering whether certain milestones had been achieved (binary), and excluding "Downpayment" and "Credit Account," "First Purchase" had the highest feature importance
  - When considering occurrences of various milestones (values), and excluding "Downpayment" and "Credit Account," "Prospecting" had the highest feature importance
  - Next steps: addressing data imbalance with more predictors and more data

Questions?