American Express Campus Challenge 2023
Modeling Problem Statement

Recommendations to drive **incremental activation** on relevant merchants for each customer

What's in it for Participants?

Use AI to save marketing costs! $

Home Depot saved an estimated $200 million in marketing costs by using **AI** by **personalizing** products on its website

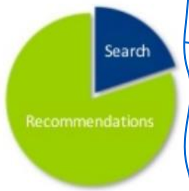
Kroger saved an estimated $100 million in marketing costs by using **data science** to **personalize** its customer loyalty offers

Build your own Recommender System

„The company reported a **29%** sales **increase to $12.83 billion** [...] Amazon has **integrated recommendations into nearly every part of the purchasing process from product discovery to checkout.**"

amazon

„Our **recommender system** [...] **in total influences choice for about 80% of hours streamed at Netflix**. The remaining 20% comes from search [...]"



NETFLIX

Uplift - the Holy Grail to drive marketing efficiency!

It's time to stop wasting marketing spend and start using uplift models

Context

Merchant Recommender: Overview

Merchant Recommender is a capability to **connect** Amex **Merchants** (shops accepting Amex cards) & Amex Credit Card **Customers**

This capability recommends* **personalized** merchants to Credit Card Customers, to help Customers **discover** merchants near to them to shop, increase **spend on merchants** & benefit Amex enterprise by driving more **revenue**

Customer visits the channel

Recommender Algorithm

Personalized Merchants for that Customer

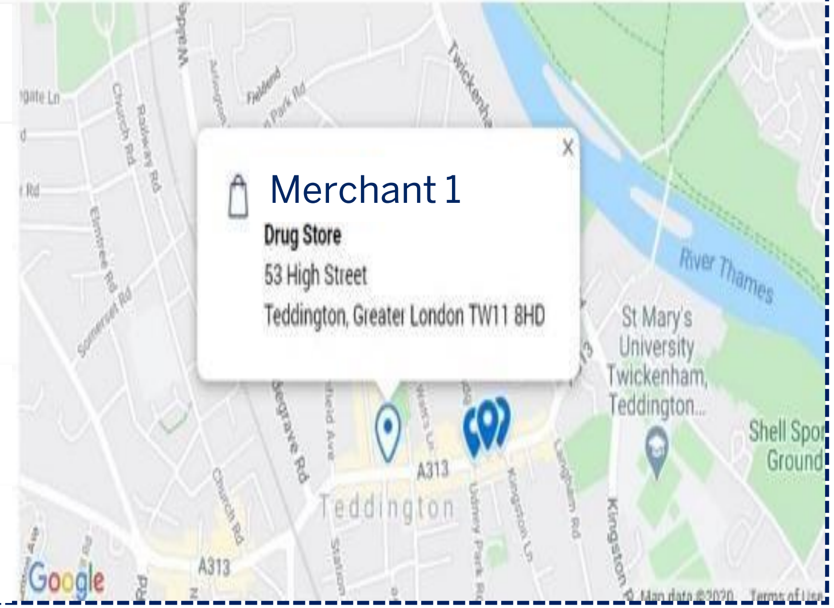
Merchant 1

Merchant 2

Merchant 3

Merchant 4

Merchant 5

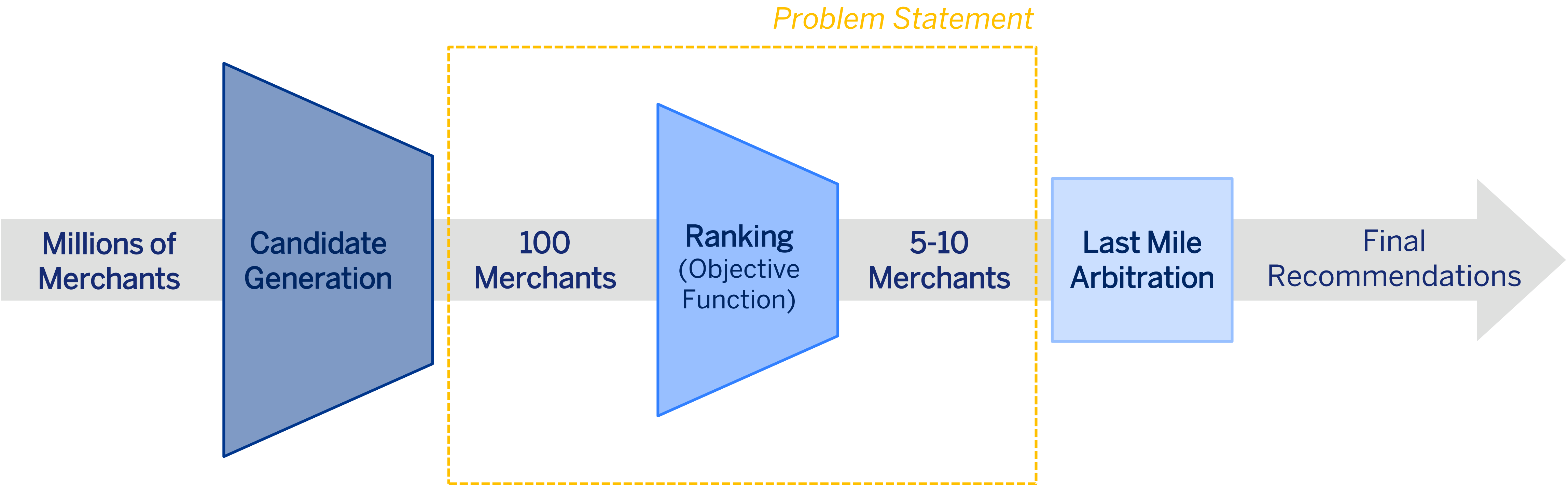


Sample logged in experience of Merchant recommender capability on Amex Website

*Recommendation means ***showing*** an Amex Merchant to an Amex Credit Card Customer on a marketing channel

Merchant Recommender: Architecture

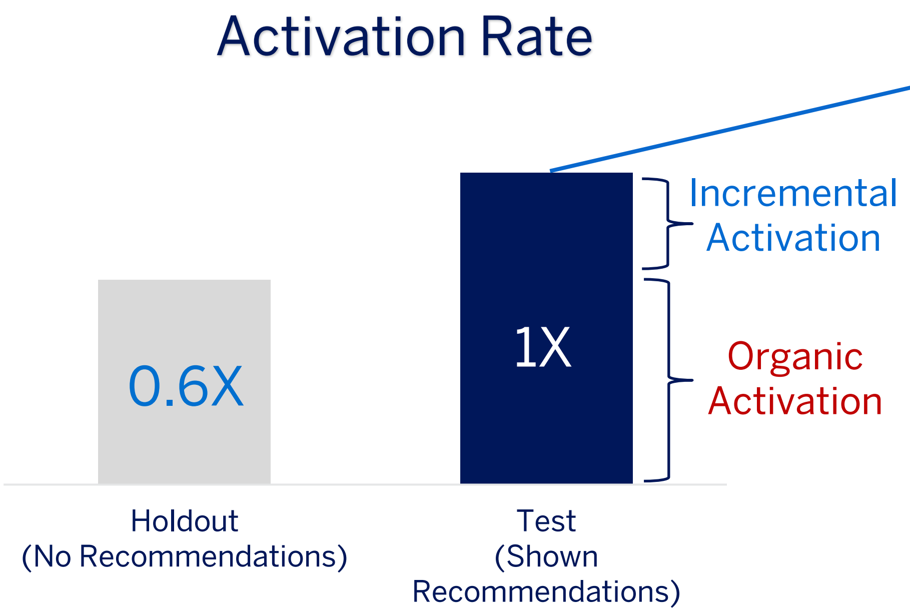
Merchant Recommender architecture is designed as per industry best practices, and comprises of 3 modules -



Motivation

To Drive Incremental Activations from each Customer

By activating 'net-new' merchants that the Customer would not have discovered otherwise, unless recommended



Response Algorithm tries to maximize only Test Activation Rate

Objective Function

P (Y = 1 | Recommendation)

Uplift Algorithm tries to maximize the **difference** of Test & Holdout activation Rate (Incremental activations)

Objective Function

P (Y = 1 | Recommendation) –
P (Y = 1 | no Recommendation)

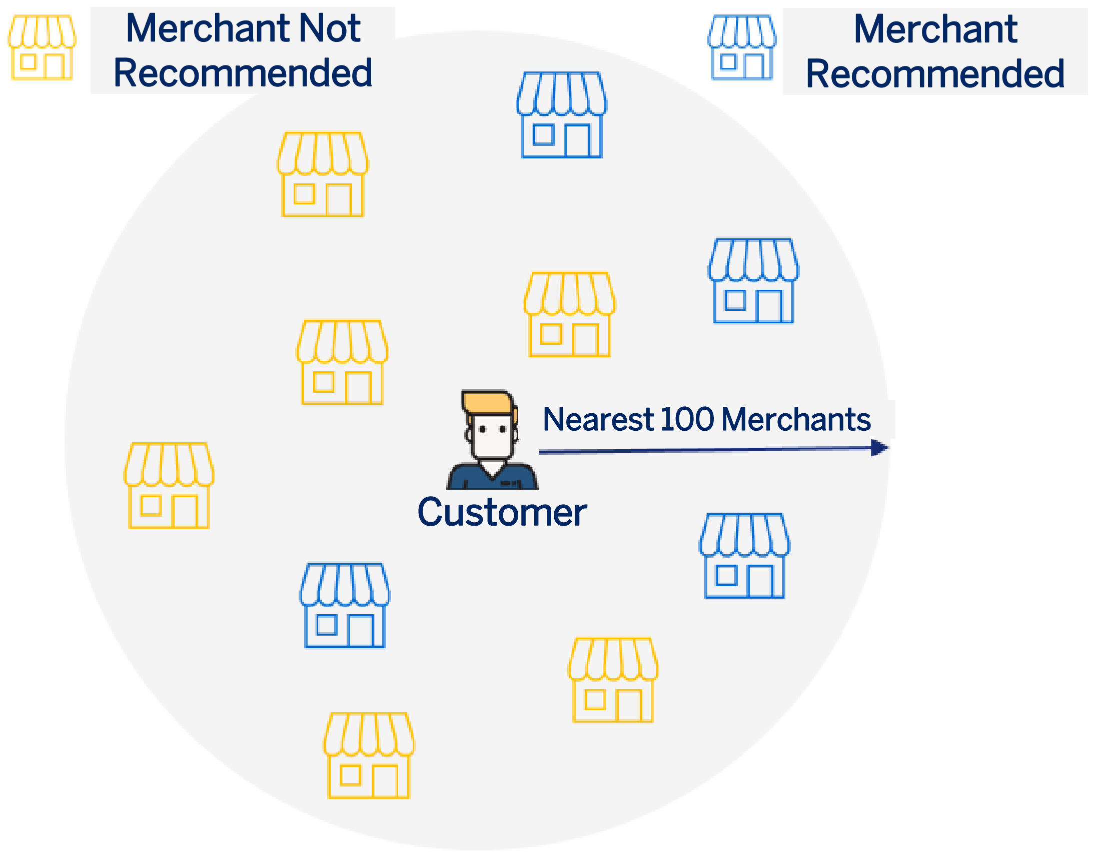
#Recommendations – Unique Customer x Merchant

#Activations - First transaction (irrespective of $ amount) on Merchant within 30 days post seeing the recommendation

Activation Rate - #Activations / #Recommendations

Problem Statement

Problem Statement



Illustration

- ➢ A Customer logs into Amex website
- ➢ Out of the nearest 100 merchants to the Customer, we **randomly** recommend some merchants and don't recommend remaining merchants
- ✓ **Goal:** Our primary goal is to **maximize Incremental** (Recommended vs Not-Recommended) **Activation Rate** on Merchants for **each** Customer.

Objective Function (to maximize)

P [(Y = 1 | Recommendation) –
(Y = 1 | no Recommendation)]

Evaluation Criteria

A good solution should be able to rank order well with Incremental Activation Rate

Illustration

Rank within Customer	Activation Rate (%) Recommended	Not-Recommended	Incremental
1 - 10	0.55%	0.21%	0.33%
11 - 20	0.20%	0.11%	0.09%
21 - 30	0.14%	0.09%	0.05%
31 - 40	0.10%	0.08%	0.02%
41 - 50	0.08%	0.07%	0.01%
51 - 60	0.08%	0.07%	0.01%
61 - 70	0.08%	0.09%	-0.01%
71 - 80	0.09%	0.11%	-0.02%
81 - 90	0.12%	0.19%	-0.06%
91 - 100	0.19%	0.35%	-0.16%

This table is created at dataset level post collating ranks of all Customers

Steps:

- Data is scored at Customer x Merchant level
- Ranks are created **within each Customer** basis **predicted** score
- **Incremental Activation rate** is calculated at each decile using Recommended Activation Rate **minus** Not-Recommended Activation Rate

Evaluation Metric:

- ❑ Incremental Activation Rate on Top 10 merchants by prediction for each Customer

*Evaluation code to be provided to participants

Data

Illustration

Sample of the Modeling Dataset

Primary Key Columns

Customer *(masked)*	Merchant *(masked)*	*ind_recommended*	*activation*	X Features (67)
1	1	1	0	....
1	2	1	1	....
...	...	...	...	....
1	6	0	1	....
1	7	0	0	....
...	...	...	...	....

Variable Name	Description
Customer (Masked)	Masked Customer Identifier
Merchant (Masked)	Masked Merchant Identifier
ind_recommended	1 if Merchant was recommended to Customer, 0 otherwise.
activation	1 if Customer transacts at Merchant within 30 days post recommendation, 0 otherwise. This cannot be used as independent feature directly or indirectly.
X Features (67)	67 Ready-to-use Independent Features are provided (details on next slide)

*ind_recommended & activation are only going to be provided for Training data

Details about Independent Predictors - X Features

Feature Type	# Features
Numerical	66
Categorical (Merchant_Profile_01)	1
	67

- For some of the features, raw feature descriptions have also been provided. For others, it's same as the feature names.
- None of the features have been normalized or imputed by any value

Feature Category	# Features	Feature Names
Customer Digital activity	22	Customer_Digital_activity_1 ... Customer_Digital_activity_22
Customer Spend	14	Customer_Spend_1 ... Customer_Spend_14
Merchant Spend	11	Merchant_Spend_1 ... Merchant_Spend_11
Distance	5	Distance_1 ... Distance_5
Customer Industry Spend	5	Customer_Industry_Spend_1 ... Customer_Industry_Spend_5
Customer Profile	4	Customer_Profile_1 ... Customer_Profile_4
Customer Merchant	3	Customer_Merchant_1 ... Customer_Merchant_3
Merchant Profile	3	Merchant_Profile_1 ... Merchant_Profile_3
	67	

Datasets

Data	Column Details	# Rows	# Columns	Location/Name
Train Data	Customer, Merchant, ind_recommended, activation + 67 X Features	12,229,978	71	*To be downloaded from the unstop website (available to all registered candidates)*
Round 1 Submission Data	Customer, Merchant + 67 X Features	8,496,466	69	
Round 2 Submission Data	Customer, Merchant + 67 X Features	4,108,134	69	*To be downloaded from the website (available to R1 shortlisted candidates)*

- All 3 datasets have been created on different time periods and are unique at Customer x Merchant level
- All 3 datasets have no overlap with respect to Customer, Merchant or Customer x Merchant
- Datasets are delimited by ‘,’
- Please note that *ind_recommended* & *activation* columns will not be provided with Round 1 & 2 Submission data

Sample of a scored Round 1/2 submission data

Primary Key Columns

Customer *(masked)*	Merchant *(masked)*	Predicted Score
1	1	0.5
1	2	0.01
...	...	....
1	6	0.2
1	7	0.09
...	...	....

Predicted Score:
P [(Y = 1 | Recommendation) –
(Y = 1 | no Recommendation)]

- Score to be calculated at each row i.e., unique combination of **Customer x Merchant**
- This predicted score will be used for evaluation
- Scored submission data unique Customer x Merchant counts should match with Submission Data Round 1 & Submission Data Round 2 counts mentioned in previous slide

*Sample Round 1/2 Submission file can be downloaded from unstop website

Stages of Competition

Round 1 Submission

(scored submission data)

Round 2 Submission

(scored submission data)

Final Round

(Deck Submission & Presentation)

Round 1 Guidelines

- Participants will train their model using labeled Training Data

- They are free to split the training data into Train & Out-of-sample/In-time data into whatever ratio they deem fit. They are also free to use any sampling technique on Train data*

- An evaluation custom code (in python) to calculate "Incremental Activation Rate (at dataset level) on Top 10 merchants (by prediction) for each Customer" would be provided to participants to run on their training/out-of-sample data scores to mimic the exact evaluation process that will run on their submitted Round 1 & Round 2 data. Directions to run the code have been mentioned in the beginning of the code. This can't be used on scored Round 1/2 submission data as it doesn't have 'activation' column.

- Once they have the best model according to them, they can score the Round 1 submission data and upload it

- Scored Round 1 submission data should be a csv & delimited by "," with 1st row as the header with column names as *customer, merchant & predicted_score*. None of these 3 columns can be missing/null/na etc.

- Scored Round 1 submission data should have only 3 columns and unique Customer x Merchant count should be 8,496,466

- Top Teams with Highest "Incremental Activation Rate on Top 10 merchants by prediction for each Customer (at dataset level)" will be shortlisted for Round 2

*Train data has already been sampled once before sharing, hence will contain < 100 merchants per Customer

Round 2 Guidelines

- Only Participants shortlisted from Round 1 will be shared Round 2 submission data
- Once they have the best model according to them, they can score the Round 2 submission data and upload it
- Scored Round 2 submission data should be "," delimited and have 1st row as the header with column names as *customer, merchant & predicted_score.* . None of these 3 columns can be missing/null/na etc.
- Scored Round 2 submission data should have only 3 columns and unique Customer x Merchant count should be 4,108,134
- Top Teams with Highest "Incremental Activation Rate on Top 10 merchants by prediction for each Customer (collated at dataset level)" will be shortlisted for Deck Submission & Presentation Round

Final Round Guidelines

- Top teams from Round 2 will be asked to share details of the codes used to arrive at the model
- Teams will also create a presentation detailing their approach including (but not limited to) sampling technique, Feature Innovation, Intuitiveness, Selection, Incremental Trends, Algorithm/Modelling Framework used, Presentation, QnA etc. They will be asked to present the same to a panel.

Top 3 teams will be selected as winners based on Round 1 & Round 2 scores, as well as scores from the presentation

All the Best!!!