ABOUT EVITE

EVITE + WCAI DATATHON AND FINDINGS

QUESTIONS
EVITE BY THE NUMBERS

The most trusted online invitation service, bringing people together face-to-face to celebrate their most important life moments

100+MM
ANNUAL USERS

32MM
registered users

200+MM
Invitations Sent Every Year

2OK
Invitations Sent Every Hour

3 BILLION
Unique Face-to-Face Connections Enabled

$21 BILLION
Spent on Parties By Evite Users

$12MM
How does Evite Work?

Hosts organize events
How does Evite Work?

Hosts organize events

Some of the guests will become hosts
How does Evite Work?

Hosts organize events

Some of the guests will become hosts

Leading to more events with more guests
How does Evite Work?

Hosts organize events

Some of the guests will become hosts

Leading to more events with more guests

Some hosts will become premium users
Revenue = $1 \times (# \text{ free events}) + $10 \times (# \text{ premium events})\ *

More Hosts

More Events

More Guests

How to drive guest to host conversion?
THE CHALLENGE:
Develop a ML/data science approach to benefit Evite
Business Motivation

Goals:
Identify free event hosts who will later host premium events

Identify guests who are likely to later become hosts

Identify events where guests are likely to later become hosts
Business Motivation

Goals:

Identify free event hosts who will later host premium events

Identify guests who are likely to later become hosts

Identify events where guests are likely to later become hosts
## Provided Data

### Events table
- event_id
- user_id (hosts only)
- event_type
- event_category
- seasonal
- event_date
- sent_on
- created_date
- zip
- premium_vs_free
- guests
- rsvp_undecided
- rsvp_maybe
- rsvp_no
- rsvp_yes

### Guests table
- user_id
- event_id
- event_type
- event_category
- Correspondence
- event_category_group
- Seasonal
- invitee_type
- rsvp_response
- event_date
- sent_on

---

# unique events – 525300
# unique users – 8222685
User/event comb. – 11080614
Stages of the Approach

Data Cleaning
- Incorrect data type.
- Missing/unreal values.

Feature Engineering
- Time-based features.
- Location-based features.
- Event-based data aggregation.

Model Training
- Split training/testing data.
- Optimise training parameters.
- Evaluate model.

Training model to predict the probability of the guests of an event later hosting an event through Evite themselves.
Data Cleaning

**Date**
Incorrect years: 8018, 2050
Filter by year < 2020
Sent date > Event date
Filter by Sent date < Event date

**Zip code**
Different formats/missing values:
90815, L5B3e2, 98840-1504, etc.
Fill missing values with 00000
Filter by r"^[d0-5]" (keeps both 5 digit and 9 digit zips)
Sent date > Event date
Filter by Sent date < Event date

**# of guests**
Wrong data types:
Some text instead of number
Remove

**Constant features**
Remove
Feature Engineering

New features

- Event date – Sent date
- Event date – Create date
- Frequency of hosting the events
- Total number of hosted events
- RSVP-Yes/Total
- Average household income (based on zip code)
- Population density (based on zip code)
Training Feature

Identify events where guests turned into hosts

Events table
>> Group by user_id
>> Filter by earliest date
user_id
event_id
hosting_date

Gives earliest date of hosting the event

Guests table
>> Group by user_id
>> Filter by earliest date
user_id
event_id
attending_date

Gives earliest date of attending any the event

>> merge on user_id
>> if attending_date < hosting_date # if user attended an event before hosting
>> Guest2host = True
>> else
>> Guest2host = False

Merge back to the events table
The model is able to predict G2H conversion with a useful level of accuracy.

Top Features
- Zip code
- Median income
- RSVP/Total
- Event date – Sent Date

ROC Curve
- Accuracy = 73%
- Metric = ‘RMSD’
- Method = LightGBM
- Classification = Binary
Implementation

- Using probability output, calculate expected revenue for each event

- For each event’s $P(\text{guest} \rightarrow \text{host})$,
  
  ```python
  if (expected revenue > external advertisement revenue):
      display Evite advertisement
  else:
      display external advertisement
  ```

- The content of Evite ads determined through recommendation engine
  
  - Similar events organized in the area.
  - Seasonal factors.
  - Network analysis.
Closing Thoughts
THANK YOU