

# Shopify

Jessica Chan, Sai Haran, Kate Larsen,  
Adit Venkataraman, Chaoran Wei



# Agenda

- Shopify Business Model
  - Platform Overview
  - Customer Lifecycle
- Customer Conversion Project
  - Overview and Business Value
  - Dataset
  - Methodology
  - Modeling
  - Applications



# What is Shopify?



# Factoids

- Based out of Ottawa, Canada
- E-commerce Platform
- Software for a “DIY Online Store”



**400,000** **\$34 Billion**

ACTIVE SHOPIFY STORES

WORTH OF SALES



# Birth of a Fictitious Store

“I have an Idea”



Fill in some Information



Store is Created

Start your free 14-day trial of Shopify

Email address  
canadiens\_gear@gmail.com ✓

Password  
\*\*\*\*\* ✓

Your store name  
Canadiens\_Hockey\_Gear ✓

Create your store

Tell us a little about yourself

• •

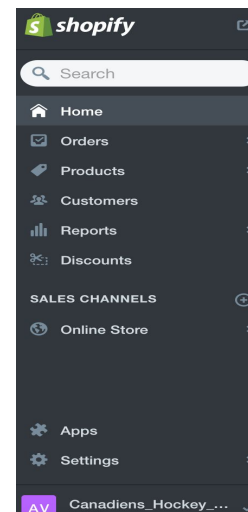
ARE YOU ALREADY SELLING?  
I'm selling, just not online ✓

WHEN WOULD YOU LIKE TO LAUNCH YOUR STORE?  
I'll be ready in a few weeks ✓

HOW MUCH REVENUE DOES YOUR BUSINESS CURRENTLY MAKE IN A YEAR?  
\$250,000 to \$1,000,000 ✓

ARE YOU SETTING UP A STORE FOR A CLIENT?  
☒ Yes, I'm designing/developing a store for a client

Next

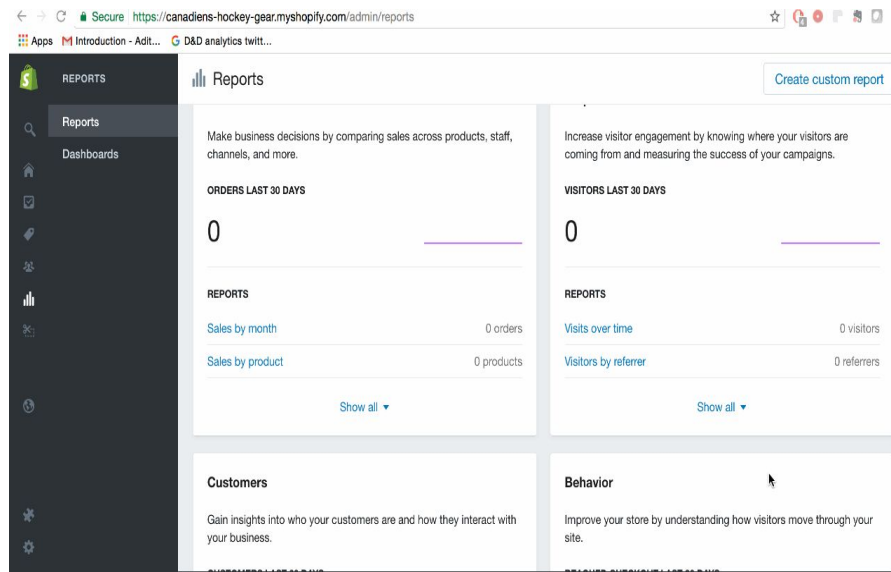
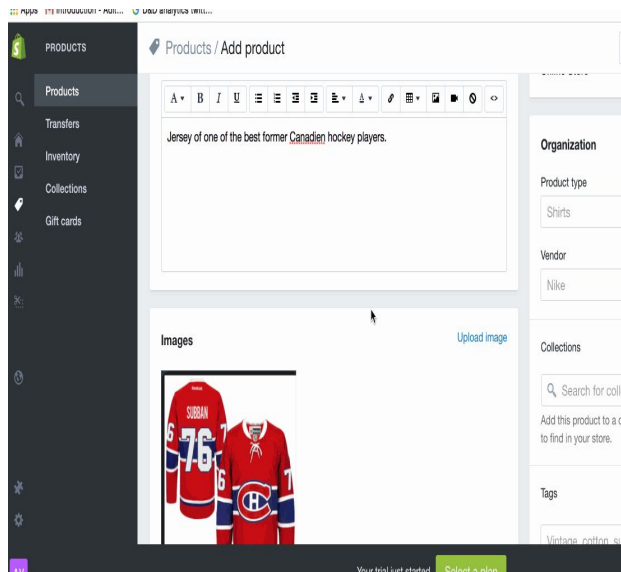


# Sample Store: Product Analysis

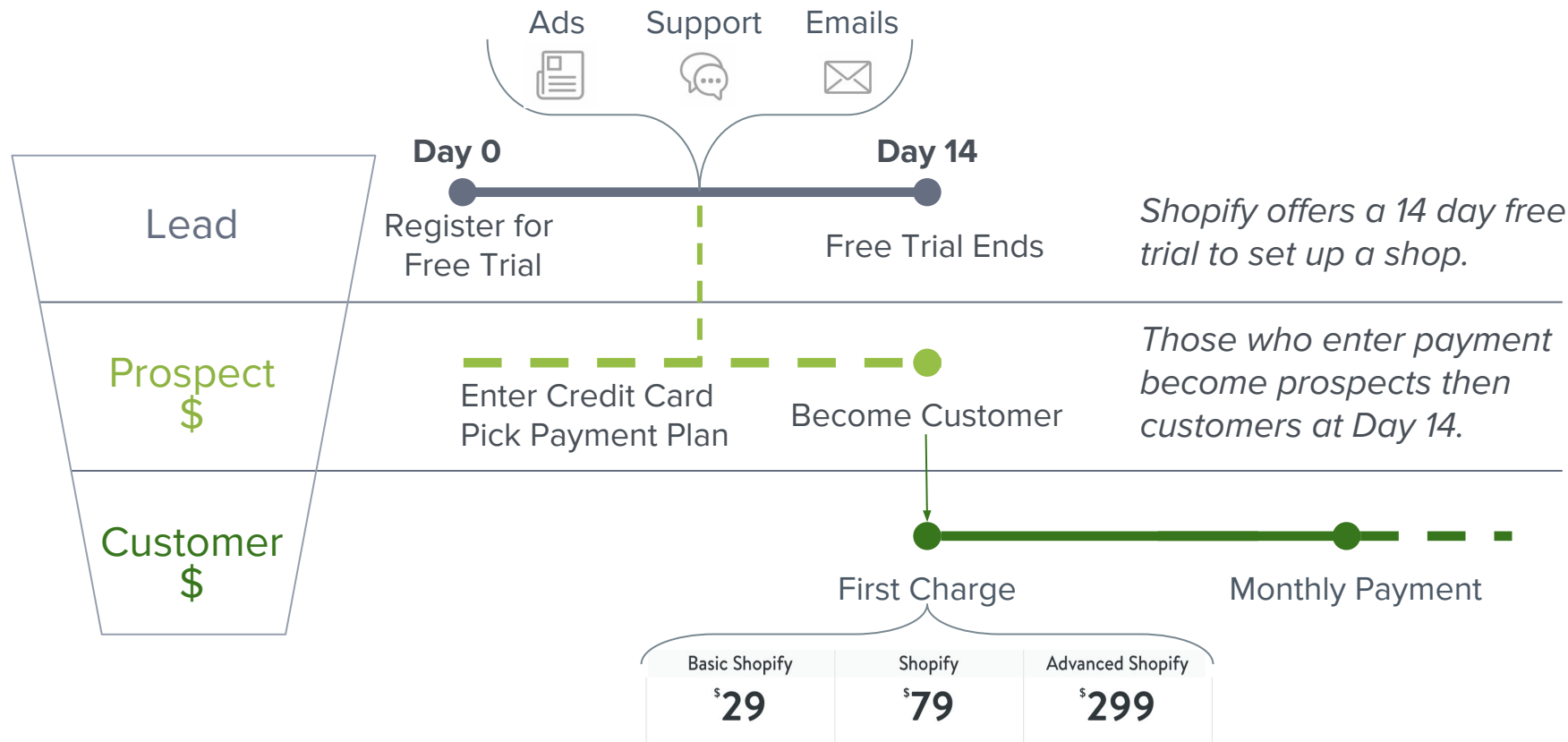
## Add Products



## View Product Reports



# Customer Lifecycle





# Project Overview





# Customer Conversion

## Goal

- Predict lead to customer conversion
- Discover key variables used to predict conversion

## Approach

- Develop 14 different models for each day in the free trial (Day 1, Day 2, ...)

## Benefits

- Reduce customer acquisition cost
- Effectively target customers

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4
5	6	7	8	9	10	11

$P(\text{customer on day 14} \mid \text{shop is } N \text{ days old})$



# Deliverables

	1) Predictive Model	2) Exploratory Web App
Details & Purpose	<ul style="list-style-type: none"><li>- Input: Merchant information from free trial</li><li>- Output: Probability of conversion to customer</li></ul>	<ul style="list-style-type: none"><li>- Demonstrate how important variables influence conversion</li><li>- Visualize model efficacy for each day of the free trial</li></ul>
Who Benefits	Marketing, Engineering	Marketing



# Dataset

- Shop information
- User information
- Web traffic data
- Sales data
- Marketing
- Support



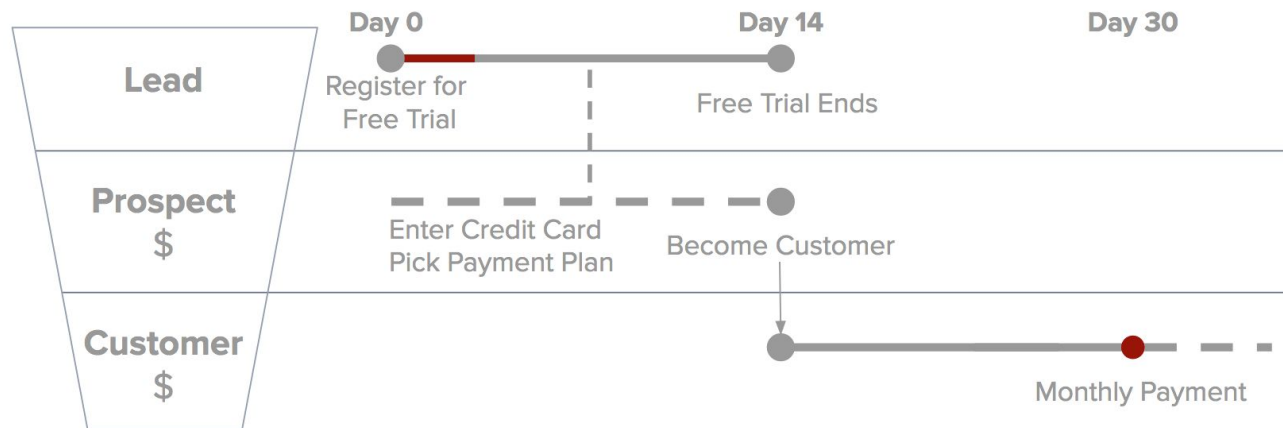
# Obstacles

Obstacle	Approach
Size of data	Random sampling
Additional data	Python data pipeline
Feature transformation	Binning, log transformation
Data Leakage	Time-boxing



# Data Leakage

Using data generated after the response variable (after lead period) to predict response variable (lead-to-customer conversion rate)



## Examples

- Use sales data after customer as predictor
- Email domain changes



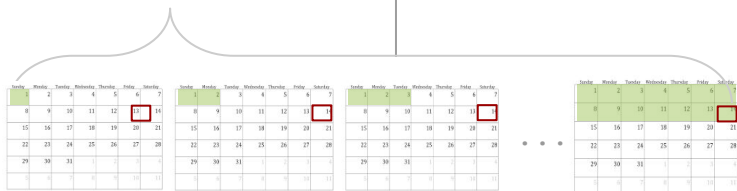
# Modeling



# Methodology

## Data Extraction

- 14 data sets, one for each day



## Train

- 14 models, one for each day

## Modeling

- Logistic Regression
- Boosted Tree
- Random Forest

## Measure Accuracy

- Classification accuracy
- Recall
- F1 score

## Web App

- Visualize importance measures and model performance

Goal:  $P(\text{customer on day 14} \mid \text{shop is } N \text{ days old})$



# Model Selection

- Random Forest
  - Feature importance interpretability
  - Account for imbalanced dataset
  - Speed in iteration
- Results
  - Accuracy, F1, Recall per day
  - Very good!

Day	Accuracy	F1 Score	Recall
1	$a_1$	$f_1$	$r_1$
2	$a_2$	$f_2$	$r_2$
3	$a_3$	$f_3$	$r_3$
...	...	...	...
14	$a_{14}$	$f_{14}$	$r_{14}$

	$p'$ (Predicted)	$n'$ (Predicted)
$p$ (Actual)	True Positive	False Negative
$n$ (Actual)	False Positive	True Negative





# Web App

## Technology

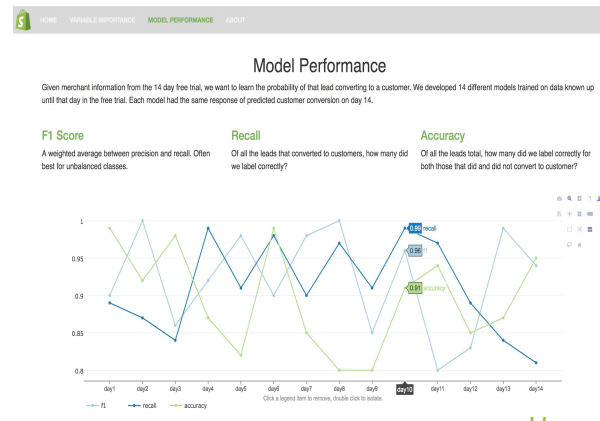
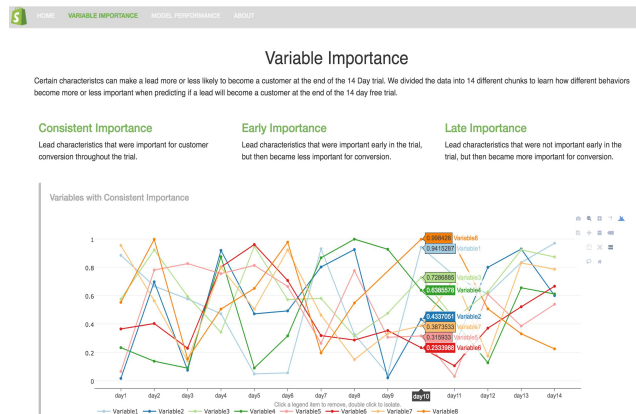
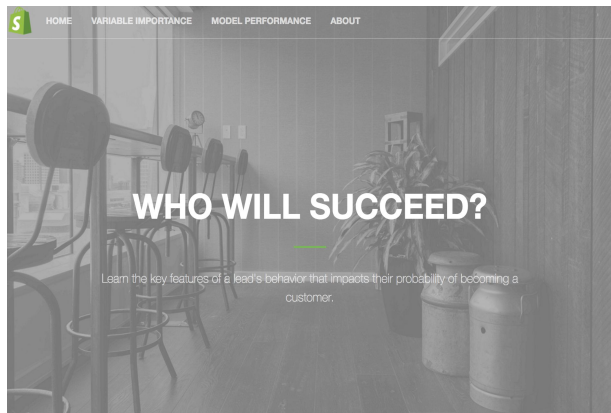
- Built in Python
- Flask app
- Interactive Plotly visuals
- Bootstrap framework

## Variable Importance

- Trace how merchant behavior becomes more or less important during the 14 day free trial

## Model Performance

- Visualize how model performance changes throughout the trial



# Applications

## 1) Identify at-risk merchants



Support queue optimization



Merchant segmentation

## 2) Leverage impact variables



Decrease CAC  
(Customer Acquisition Cost)



Nudge merchants in free trial



Questions?

