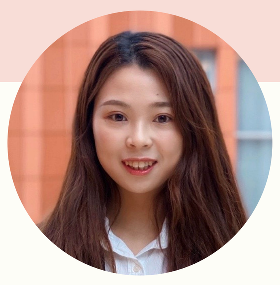
True Sales Potential

Unleashing the untapped opportunity with Big Data



Tico Han　　Sydney Mo

Faculty Advisor: Dr. Arnold I. Barnett
Company Representative: Estefanía Elizondo Gómez

Project Charter

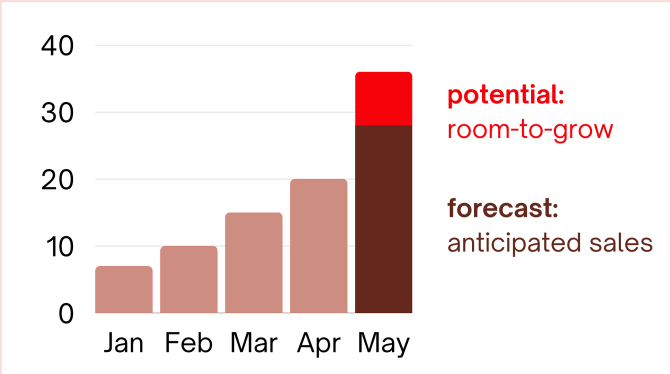
Mar	Apr	May	Jun	Jul	Aug
On-boarding Stake-holder Meeting Data Exploration		Code Outline 1st Revision: Accuracy	1st On-site Visit 2nd Revision: Methodology	2nd On-site Visit 3rd Revision: Validation	Hand-over

Problem

With its wide range of customers through different trade channels, geographical locations, and product portfolios, CCSWB aims to create tailored service experiences for its customers by understanding their characteristics. Therefore, **sales potential** is introduced to measure the unrecognized room-to-grow of a customer.

How much more can we sell?

It is a win-win solution for both CCSWB and its customers because it drives extra sales for both sides.



Objective

- Identify **internal & external factors** that can affect sales potential
- Calculate a **monthly, customer-level** true sales potential **by product category**
- Provide guidance to **optimize sales and resource allocation**

Data

The data we use can be divided into three parts:

Scope: 2018 ~ 2021 **Dallas Fort-Worth (DFW)** Region
- **Internal:** Sales, Customer Characteristics, Product Information
- **External:** District Demographics (i.e. population, median income)
- **Nielsen:** Beverage Category Market Share

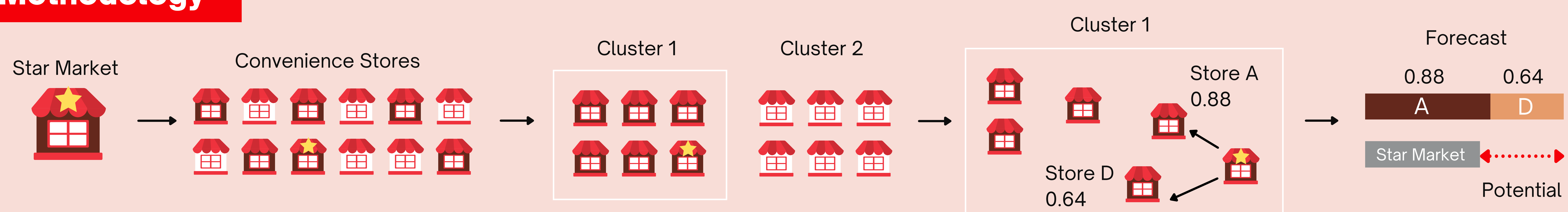
Features extracted from the data:

Sales-related:
Number of categories sold
Last purchase date/quantity
Last 30/90 days total sales
Local market share

Time series:
Year/Month/Quarter/Weekday
Seasonality (Holiday)

Non-sales-related:
Sub Trade Channel
Sales Office
Business Type
Geographic
Demographic (education, gender, race, income)

Methodology



1. Segmentation:
The data are first filtered to **25 trade channels** that CCSWB already defined.

2. Clustering:
For each trade channel, we use **k-means clustering** to group similar customers together based on features.

3. K-Nearest Neighbors:
We find the **closest "neighbors"** of a target customer based on the difference of features within its assigned cluster.

4. Potential Calculation:
The potential is the **gap of forecast** from the target customer to the **weighted average** of its "neighbors".

Prediction

To accurately estimate potential, we need an accurate forecast of sales volume. During data exploration, we observed very distinct purchasing habits across channels and beverage categories. Therefore, our ensemble model tackles this problem by taking consideration of sales propensity:

V: Volume forecast = **unit** of sales volume given purchase
D: Propensity forecast = **whether** customer will make purchase

Volume Forecast		Propensity Forecast		Forecast
V | (D = 1)	x	D = {0,1}	=	E (V)

Prescription

Based on the calculated potential and the prediction of purchase decisions, our model segments customers into 4 tiers to instruct visitation:

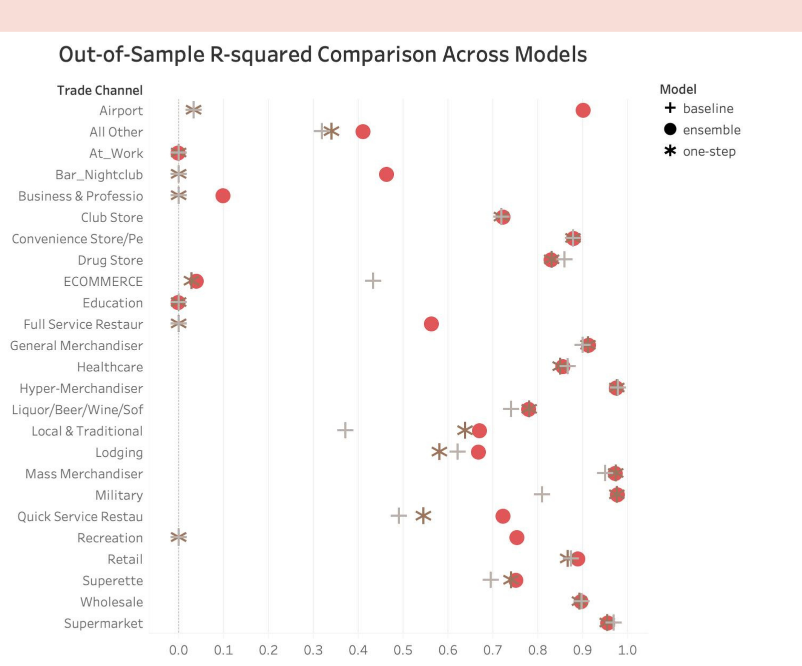
Priority of Visit
1. Target Customer: (Potential is **high** & **will** purchase)
2. Maintain Relation: (Potential is low & **will** purchase)
3. Rising Star: (Potential is **high** & will not purchase)
4. Limited Opportunity: (Potential is low & will not purchase)

Rising Star	Target Customer
Limited Opportunity	Maintain Relation

Impacted Customers	Correctly Identify	Out-of-Sample R-squared	Out-of-Sample Area Under Curve
13.7K	**99%** sales opportunity	**+37.9%** than baseline	**+29.2%** than baseline

Result

Our result is very promising. We compared the OSR2 across all three models. It turned out that ensemble model outperforms the other two(or equivalent) in all trade channels.



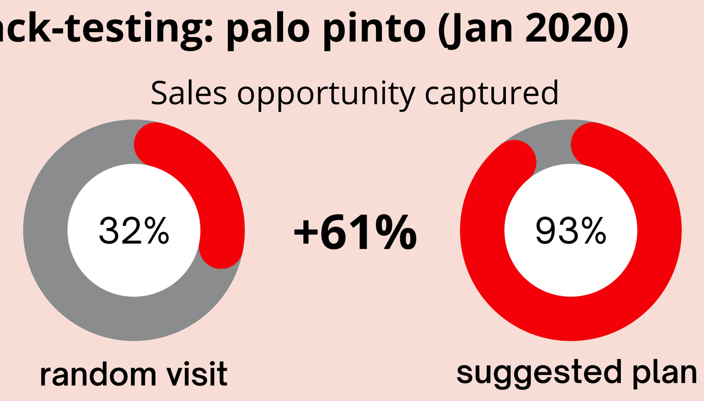
Baseline:
V = V of last month

One-step Model:
Linear Regression / Random Forest / XGBoost — choose **Best OSR2 in validation**

Ensemble Model:
Linear Regression / Random Forest / XGBoost — choose **Best OSR2 (volume)**
X
Linear Regression / Random Forest / Gradient Boosting / XGBoost — choose **Best OSR2 (propensity)**

Validation

1. Back-testing: palo pinto (Jan 2020)



2. On-Shelf-Availability (OSA) vs. Potential

OSA = % of items that are in-stock on the shelf

Assumption
- low **OSA** => high **potential**
(Demand not met, we can sell much more!)
- high **OSA** => low **potential**
(We are selling as much as we can.)

Result
- Correlation can be proved in several channels.
- Limitations: many other factors also affect OSA.

Future Steps

- Flexibility for add-on constraints
- Optimize resource, consider profitability
- Pilot launch / Proof of Value
- Automation / Runtime Efficiency
- Internal CRM System Integration