MIT MANAGEMENT BUSINESS ANALYTICS

ARCACONTINENTAL Coca-Cola SOUTHWEST BEVERAGES

Prioritizing Customers Visits

Coca-Cola Southwest Beverages
Estefanía Elizondo Gómez
Raúl Arcadio Castro Ramírez
Alejandro Valdés Valdés

Luis Honsel
Killian Farrell

Faculty Advisor
Stephen C. Graves

Context

The **COVID pandemic** had a **strong impact** on Coca-Coca Southwest Beverages' **revenues**

Recapturing the lost revenue **as soon as possible** has become a **top business priority**

► How can **analytics** help us recapture revenue?

Data

Google COVID-19 Community Mobility Reports

Mobility reports show where users are currently spending their time, e.g. residential areas, supermarkets, etc.

United States Census Bureau

Census data contains labor, economic, and population data by county

Data Lake

Over 2 years of customer sales (tens of millions of rows)

Master data is all about Customer

Customer master data

Customer profitability

! Problem Statement

Optimize the allocation of business development visits to customers to boost Coca Cola's revenue recovery

The customers here are outlets where you drink Coca-Cola on-site (restaurants, bars, etc.)

Therefore, our task is to to

► **prioritize customers decide on who to contact first**

Visiting customers is key for boosting relationship and revenue growth

How:

Linear Regression model predicting revenues

Clustering

Gradient boosting model predicting order likelihood

Mixed Integer Optimization model

Automated ranking

When:

March-April: Data Acquisition & Stakeholder meetings

April-May: Data cleansing, first results with clustering

July: CEO meeting

August: Final Presentations

April: Project scope change due to COVID

June-July: Model developments, testing & integration

Upcoming: Roll-Out

Methodology

1 – Clustering Customers

Customer Types

airports, at-work, bars/clubs, education, restaurants, healthcare, lodging, quick-service restaurants, recreation, retail, wholesale

These are very heterogenic!

Challenge

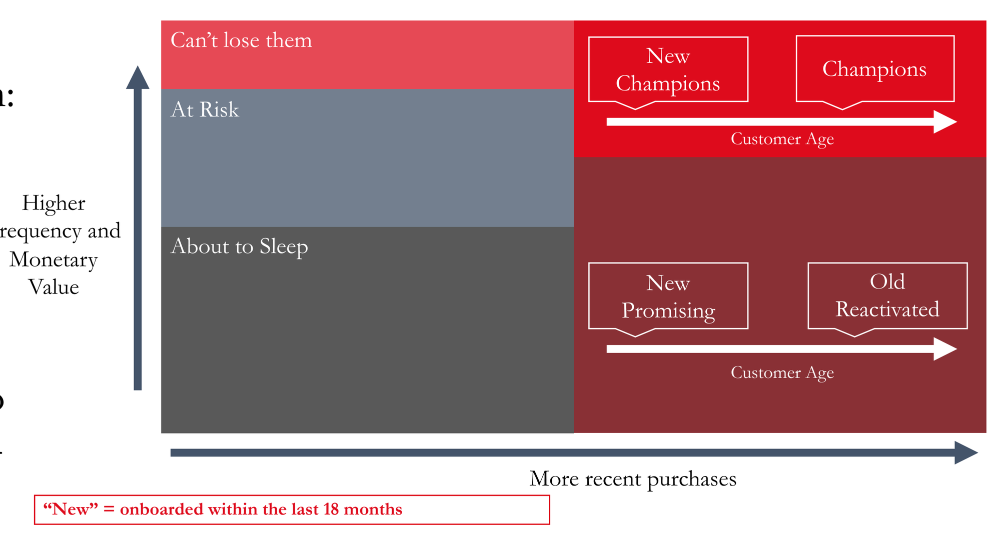
Create clusters that compare customers to relevant peers and are globally consistent

Solution: RFM Analysis with Tenure-Aspect

Per subtrade channel, each outlet gets a score based on:

- Recency – time since last purchase
- Frequency – total number of purchases
- Monetary – revenue generated

Customer ranked on a 1-4 scale in each category, overall weighted scores gives customers' priority. In addition, we incorporate the tenure of the customer to account for different approaches towards new vs long-lasting customers.



Breakout already revealed interesting results as to how customers are spread across clusters

Understandable business logic that gives results that are

- **Quick**
- **Interpretable**
- **Consistent**

Business already started using this data-based language to describe their customers in conversation!

2 – Identifying high Potential Customers

Predicting Future Revenues:

Using regularized (elastic net) linear regression, we predict next month's total revenue for each customer

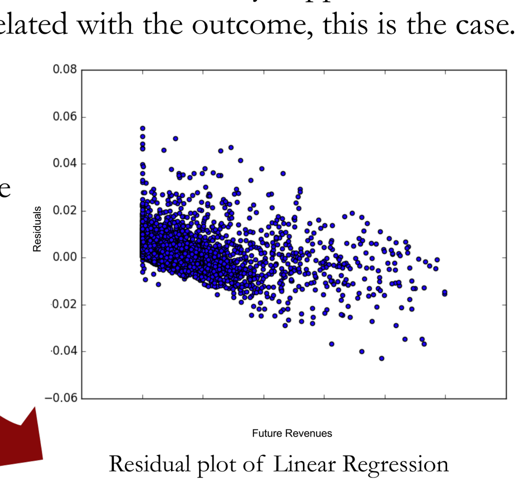
Target: Revenue in $ for next month

Predictors: Engineered revenue features, times between orders, customer-specific characteristics, economical data

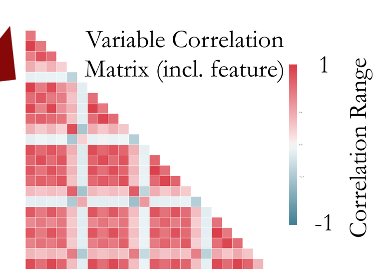
Why linear regression, it sounds so obvious?

Good question! Linear regression—out of all the tested models—performed the best. This usually happens when features are very correlated with the outcome, this is the case.

Residual plot shows very good performance on low to medium revenues



Residual plot of Linear Regression



Out of sample R^2: 0.85

Most important features:

- Location
- Historic revenue features

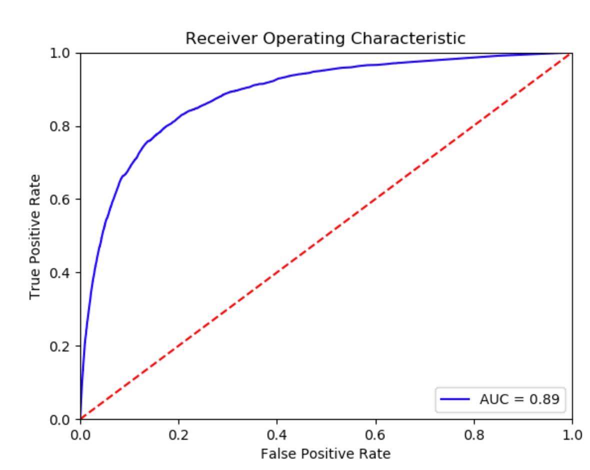
10-fold cross-validation to tune hyper-parameters

Order Likelihood:

Using Gradient-Boosted Trees, we determine whether a customer will order in the next month

Target: Will the customer order within the next month?

Predictors: Engineered revenue features, times between orders, customer-specific characteristics, economical data



ROC Curve

ROC AUC: 0.89

Most important features:

- Location
- Historic revenue features

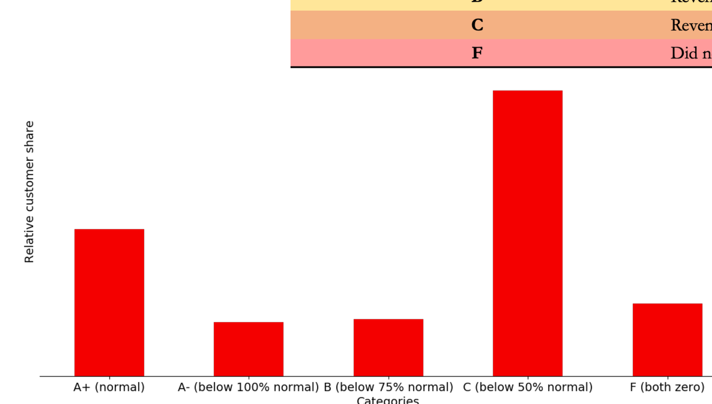
	Predicted Order	Predicted no Order
Order	54%	11%
No order	20%	15%

Revenue Recovery Grade:

Showing drop in revenue levels after COVID

This method is descriptive, it compares last years revenue to the current year (over a fixed period of weeks, e.g., week 6 to week 12). The relative change is mapped to a grade:

Revenue Recovery Grade	Explanation
A+	Revenue same or better vs previous year
A-	Revenue up to -25% vs previous year
B	Revenue up to -50% vs previous year
C	Revenue lower than -50% vs previous year
F	Did not buy in both time periods



While before COVID lockdowns the revenues were overwhelmingly at normal levels, we now see a significant portion of customers with "below-normal revenues"

Chart showing shares during some post-lockdown period

Proved to be a very valuable business metric
Already being implemented in a dashboard

3 – Optimizing Customer Ranking

Goal: Maximize customer value objective

Optimization Objective
Easy to adapt

Revenue Prediction · Order Likelihood · Outlet Profitability · Clustering · Revenue Recovery Grade · Local Accounts

Customer value objective: weighted combination of predicted revenue, order likelihood, revenue recovery, clustering results, profitability, and location parameters

Constraints: Remember, the subtrade channels are very heterogeneous! We constraint our solution space so that within the list of customers of visited, there is a balance of subtrade channels.

Output: Prioritized list of customers

134% more revenue visited

15% more profitable customers

21% more relevant customers

Backtesting

Backtesting performed on **half a year of data**, comparing **actual visit plans versus the optimization**

vs.

The optimization prioritizes more valuable customers (see above)

There is a statistically significant correlation between our prioritization and revenue growth

Top prioritized customers significantly less affected by COVID impact

During recovery time the prioritized customers show higher growth

Senior and operations stakeholders convinced by these initial results to run pilot of model to uncover causal links and recapture revenues

Impact

Optimization results integrated seamlessly into current planning tool

Our model is adaptable to different business context by re-weighting

Brick structure, you can add and remove components as you need

- **Pilot** test will be put in place **next weeks**
- Model will impact **more than 40k customers**
- **First** Advanced Analytics project at Coca-Cola Southwest Beverages