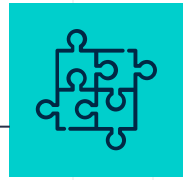
Optimizing Customer Acquisition Strategies Using Predictive Analytics

Group: Sophia Mei, Alex Ng, Ying Qing, Laurie Ye

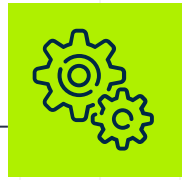
Agenda



01

Business & Data

Understanding of the business, the data, and the problem presented



02

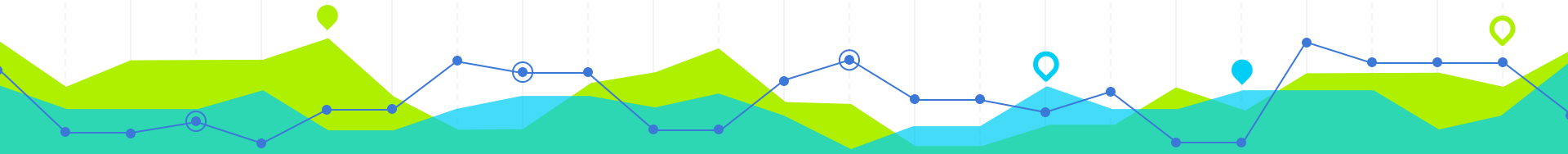
Modeling Process

Create an efficient model to help answer the business problem

03

Evaluation & Deployment

Evaluate model effectiveness and deploy the model in real life



Business Problem 1

Motivation

Fewer
advertisements
than Walmart and
Sprout retail stores

Less profitable
than 7 Eleven and
Circle K Franchises

Business Problem: Cost-Saving of Customer Acquisition

Problem Statement

Convenient Food Mart is **facing increasing competition** from other grocery retailers and franchises, resulting in high costs to acquire new customers.

According to recent market research, Convenient Food Mart has **much higher customer acquisition costs** than industry benchmark. This is due to a lack of precision in their customer targeting efforts, resulting in wasted marketing spend and inefficient acquisition channels.

Solution Approach

By analyzing the Convenient Food Mart customer dataset and R as a tool, we can identify **demographic traits or behavior** that are common among customers who are likely to be profitable for Convenient Food Mart.

We can identify **key factors** that influence profitability and use this information to target similar customers with higher precision, reducing the cost of acquisition.

Data Preparation 2

Summary of Variables

Variables	Description	Type
Marital Status	Married or Single (M or S)	Binary
Gender	Female or Male (F or M)	Binary
Total Children	Number of children customers have	Continuons
Education	Partial High, High School, Partial College, Bachelors, Graduate Degree	Categorical
Member Card	Normal, Bronze, Silver, Golden	Categorical
Occupation	Clerical, Management, Manual, Professional, Skilled Manual	Categorical
Houseowner	Customer is owner or not	Categorical
Average Cars at Home	Number of cars at customer's home	Continuous
Average Yearly Income	Range of yearly income	Categorical
Children at Home	Number of children at home	Continuous
Media Type	Different types of media used	Categorical

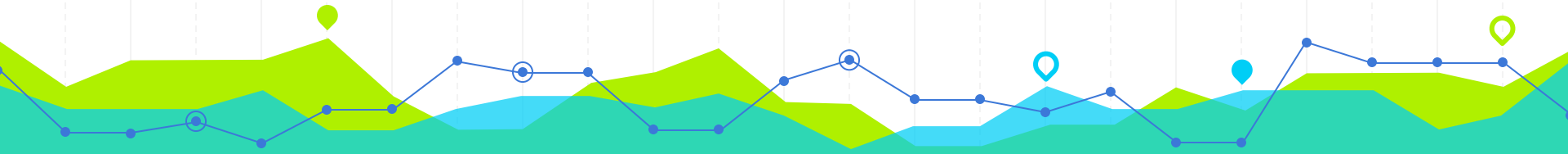
Dummies Variables

Media Types	Average Yearly Income	Occupation	Education	Member Card
- Bulk Mail - Cash Register Handout - Daily Paper - Daily Paper, Radio - Daily Paper, Radio,TV - In-Store Coupon - Product Attachment, Radio - Street Handout - Sunday Paper - Sunday Paper,Radio - Sunday Paper, Radio, TV - TV	- $10K - $30K - $30K - $50K - $50K - $70K - $70K - $90K - $90K - $110K - $110K - $130K - $130K - $150K - $150K +	- Clerical - Management - Manual - Professional - Skilled Manual	- Bachelors Degree - Graduate Degree - High School Degree - Partial College - Partial High School	- Bronze - Golden - Normal - Silver

Final Table

```
 [1] "sales_country"                  "marital_status"                    "gender"
 [4] "total_children"                 "houseowner"                        "avg_cars_at.home.approx."
 [7] "num_children_at_home"           "cost"                              "media_typeBulk Mail"
[10] "media_typeCash Register Handout" "media_typeDaily Paper"             "media_typeDaily Paper, Radio"
[13] "media_typeDaily Paper, Radio, TV" "media_typeIn-Store Coupon"        "media_typeProduct Attachment"
[16] "media_typeRadio"                "media_typeStreet Handout"          "media_typeSunday Paper"
[19] "media_typeSunday Paper, Radio"  "media_typeSunday Paper, Radio, TV" "media_typeTV"
[22] "avg..yearly_income$10K - $30K"  "avg..yearly_income$30K - $50K"     "avg..yearly_income$50K - $70K"
[25] "avg..yearly_income$70K - $90K"  "avg..yearly_income$90K - $110K"    "avg..yearly_income$110K - $130K"
[28] "avg..yearly_income$130K - $150K" "avg..yearly_income$150K +"        "occupationClerical"
[31] "occupationManagement"           "occupationManual"                  "occupationProfessional"
[34] "occupationSkilled Manual"       "educationBachelors Degree"         "educationGraduate Degree"
[37] "educationHigh School Degree"    "educationPartial College"          "educationPartial High School"
[40] "member_cardBronze"              "member_cardGolden"                 "member_cardNormal"
[43] "member_cardSilver"
```

Dependent Variables:
Cost of acquiring one customer



Data Cleaning 3

Dropping Variables

- Given the multitude of individual, distinct, and unique promotion methods, it is not necessary to include promotion name as a predictor variable in our modeling approach.

- Since 'Avg_cars_at.home.approx..1' and 'Avg_cars_at.home.approx.' contain duplicate information, we can drop the 'Avg_cars_at.home.approx..1' column.

```
colSums(is.na(model_data))
sum(is.na(model_data))
```

```
                sales_country                  marital_status                          gender
                            0                               0                               0
              total_children                     houseowner                avg_cars_at.home.approx.
                            0                               0                               0
        num_children_at_home                           cost                  media_typeBulk Mail
                            0                               0                               0
media_typeCash Register Handout            media_typeDaily Paper       media_typeDaily Paper, Radio
                            0                               0                               0
media_typeDaily Paper, Radio, TV          media_typeIn-Store Coupon       media_typeProduct Attachment
                            0                               0                               0
           media_typeRadio                media_typeStreet Handout          media_typeSunday Paper
                            0                               0                               0
media_typeSunday Paper, Radio media_typeSunday Paper, Radio, TV                   media_typeTV
                            0                               0                               0
avg..yearly_income$10K - $30K     avg..yearly_income$30K - $50K     avg..yearly_income$50K - $70K
                            0                               0                               0
avg..yearly_income$70K - $90K     avg..yearly_income$90K - $110K    avg..yearly_income$110K - $130K
                            0                               0                               0
avg..yearly_income$130K - $150K        avg..yearly_income$150K +              occupationClerical
                            0                               0                               0
      occupationManagement                  occupationManual          occupationProfessional
                            0                               0                               0
occupationSkilled Manual          educationBachelors Degree       educationGraduate Degree
                            0                               0                               0
educationHigh School Degree           educationPartial College      educationPartial High School
                            0                               0                               0
          member_cardBronze                  member_cardGolden                member_cardNormal
                            0                               0                               0
          member_cardSilver
                            0
```

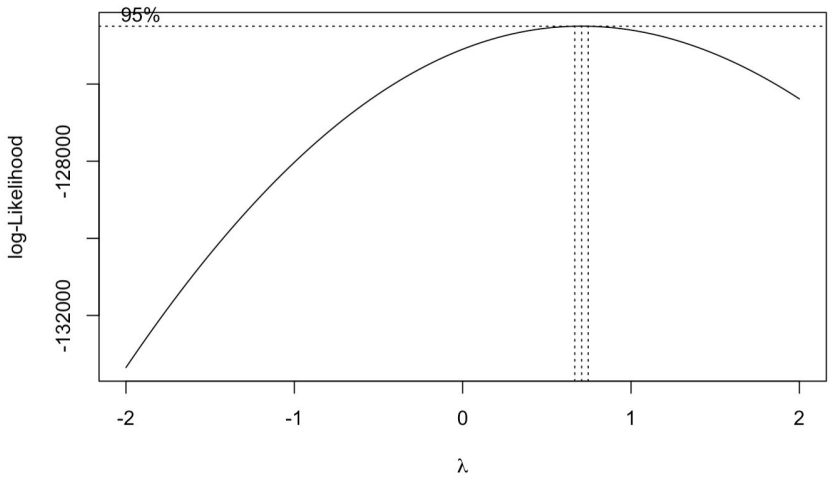
Linear Regression 4

Correlation

	media_typeStreet Handout	media_typeSunday Paper	media_typeSunday Paper, Radio
marital_status	0.0034759075	0.0001526972	-0.008325913
gender	-0.0093782565	-0.0071956493	-0.011917652
total_children	0.0116658725	-0.0227189113	-0.006334334
houseowner	0.0169774160	-0.0006968848	0.002778026
avg_cars_at.home.approx.	0.0147487887	-0.0294619188	-0.012072989
num_children_at_home	0.0053456678	-0.0068229740	-0.004330210
cost	0.0695081792	-0.0466006706	0.052623528
media_typeBulk Mail	-0.0697097995	-0.0705701884	-0.072624508
media_typeCash Register Handout	-0.0668744310	-0.0676998245	-0.069670587
media_typeDaily Paper	-0.0925560217	-0.0936983886	-0.096425977
media_typeDaily Paper, Radio	-0.1022045621	-0.1034660155	-0.106477942
media_typeDaily Paper, Radio, TV	-0.0884467043	-0.0895383523	-0.092144841
media_typeIn-Store Coupon	-0.0726200358	-0.0735163440	-0.075656427
media_typeProduct Attachment	-0.0853175288	-0.0863705551	-0.088884828
media_typeRadio	-0.0728634961	-0.0737628093	-0.075910067
media_typeStreet Handout	1.0000000000	-0.0709135488	-0.072977864
media_typeSunday Paper	-0.0709135488	1.0000000000	-0.073878588
media_typeSunday Paper, Radio	-0.0729778637	-0.0738785884	1.000000000
media_typeSunday Paper, Radio, TV	-0.0653678577	-0.0661746564	-0.068101021

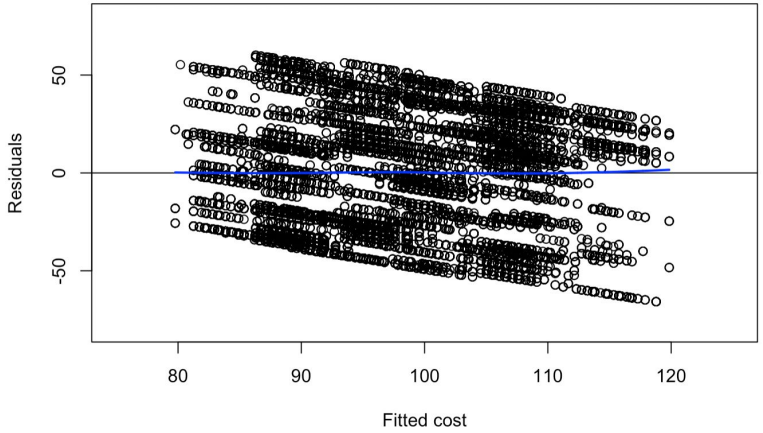
Variables Selection - Backward Stepwise & Box-Cox Transformation

```
Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                        94.4173     0.5241 180.143  < 2e-16 ***
avg_cars_at.home.approx.            1.0634     0.1547   6.873 6.43e-12 ***
`media_typeBulk Mail`              13.7350     0.7582  18.115  < 2e-16 ***
`media_typeCash Register Handout`  19.2161     0.7861  24.444  < 2e-16 ***
`media_typeDaily Paper`            -6.8909     0.6393 -10.779  < 2e-16 ***
`media_typeDaily Paper, Radio, TV` -6.7887     0.6550 -10.365  < 2e-16 ***
`media_typeIn-Store Coupon`         4.4423     0.7396   6.007 1.91e-09 ***
`media_typeProduct Attachment`     11.5112     0.6734  17.095  < 2e-16 ***
media_typeRadio                     4.3185     0.7401   5.835 5.44e-09 ***
`media_typeStreet Handout`         11.0303     0.7581  14.551  < 2e-16 ***
`media_typeSunday Paper`           -2.8219     0.7564  -3.731 0.000191 ***
`media_typeSunday Paper, Radio`     8.2154     0.7405  11.094  < 2e-16 ***
`media_typeSunday Paper, Radio, TV` -12.2605     0.7994 -15.336  < 2e-16 ***
occupationProfessional             -1.4880     0.3735  -3.984 6.79e-05 ***
`educationHigh School Degree`      -1.9647     0.3934  -4.994 5.95e-07 ***
`educationPartial College`          1.9575     0.5998   3.264 0.001101 **
```

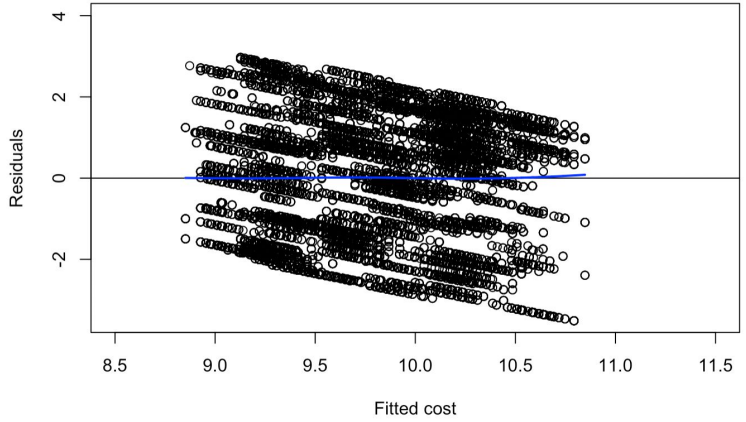


Linear Model & Result

Linear Model (Before Box-Cox)



Linear Model (After Box-Cox)



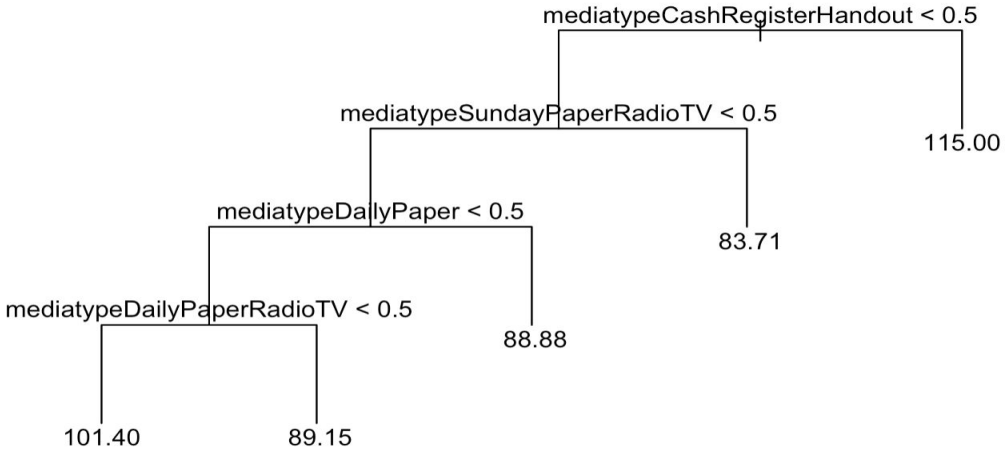
W/O BC	ME	RMSE	MAE	MPE	MAPE
Test Set	-0.8955734	29.31947	25.44998	11.66243	30.50074

With BC	ME	RMSE	MAE	MPE	MAPE
Test Set	-4.814887e-13	29.10552	25.12472	-10.42456	29.71302

Regression Tree 5

Regression Tree

Min.	1st Qu.	Median	Mean	3rd. Qu	Max.
-61.970	-30.880	4.359	0.000	24.860	57.260



Regression Tree Result

	ME	RMSE	MAE	MPE	MAPE
Test Set	-0.9381577	29.63132	25.96069	-11.96705	31.12783

Regression Tree Result After Random Forest

	ME	RMSE	MAE	MPE	MAPE
Test Set	-0.1825283	16.18703	9.935356	-3.987051	11.75767

Comparison 6

Model Comparison Analysis

Linear Regression

- Provides formulas and coefficients to understand the relationship between media costs and target variables
- Includes weights on non-media_type variables (e.g. `avg_cars_at.home.approx`, `occupation`, `education`)

	ME	RMSE	MAE	MPE	MAPE
Test Set	-4.814887e-13	29.1055	25.1247	-10.425	29.7130

✅ Random Forest

- Lower RMSE, MAE, MPE, and MAPE
- Capture non-linearity in the data
- Ease of interpretation by non-data background stakeholders

	ME	RMSE	MAE	MPE	MAPE
Test Set	-0.1825	16.1870	9.93535	-3.9870	11.7576

Conclusion 7

Deployment

Conclusion

Media outlets (e.g., TV, radio, print advertising) have different levels of effectiveness in attracting new customers.

Sunday Paper, TV, Radio are the most effective marketing strategies in seizing customers. Convenient Food Mart may carefully consider their media strategy when trying to minimize customer acquisition costs.

Limitation

We did not include other dependent variables such as store data or promotion type that may potentially influence marketing costs.

Discussion

Our model can be applied during the promotion planning stage to estimate potential media costs incurred and the segment of customers attracted.

By further combining costs and past revenue data, the ROI of different media promotions could be derived.

Future approach

Pair Customer Lifetime Value with CAC for different customers.

Segment customer demographic groups to uncover further purchasing pattern.

THANKS!

Any questions?