amazon

Amazon Ads Dog Food Project

EMORY GOIZUETA BUSINESS SCHOOL

Master of Science in Business Analytics MSBA

Our Team

amazon

Sherraina Song

Rick Wang

Chris Chou

Pandora Shou

Content

Pet Food Market Background

70%+ Households

in the United States now own a pet

43% Purchases

of pet supplies are online transactions

57% Pet Owners

in the US purchase through Amazon more than other online shops

Our Client

amazon

<u>Amazon Ads</u>

Our client's goal is to reinvent advertising, helping businesses to build brands, push creativity, and drive performance for millions of customers every day

Optimize Ad Channels for Targeted Customers

Best Customers

Identify the most responsive customer segments for specific brand

Best Ads

Predict the best type of ads for the segment and increase purchase possibility

5K Transaction & Demographic Data

amazon

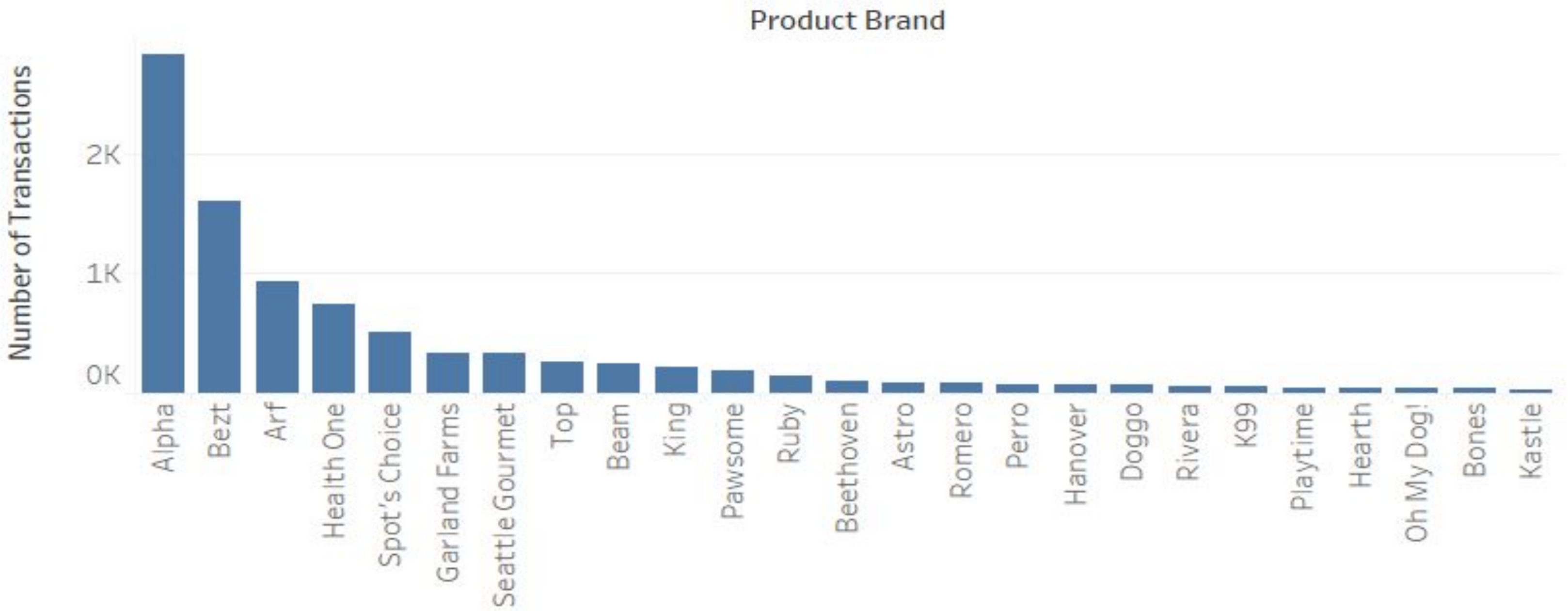
Data come from customers' survey who have purchased dog food products.

Column	Description
sale_id	ID of the sale
sale_date	Date of the sale
ad_exp	Ad experience of the sale
product_id	ID of the product
product_brand	Brand of the product
product_name	Name of the product
price	Unit price of the product
qty	Quantity of the product purchased

Column	Description
customer_id	ID of customer who purchased the product
gender	Gender of customer who purchased the product
city	City where customer resides
st	State where customer resides
zip	Zip code where customer resides
lat	Latitude of customer's residence
lng	Longitude of customer's residence
marital	Marital status of customer
education	Highest education level of customer
income	Income bracket of customer
age	Age range of customer
prime	Amazon Prime status of customer (1 or 0)

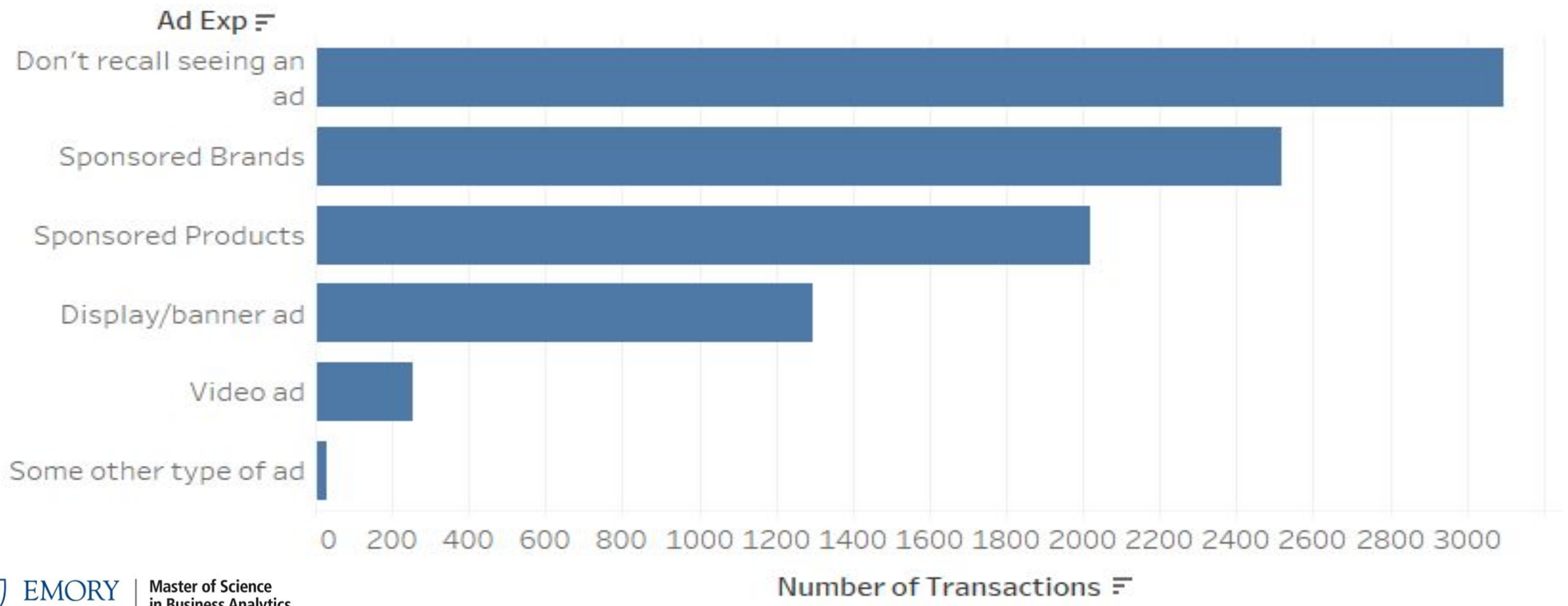
Top 5 Brands Account for 70% Total Sales

- The size of data for smaller brands is too small for model building
- We will focus on the top 5 brands for model building, smaller brands group as “other”



⅓ of the Customers don't Recall Seeing Ads

- ⅓ of the transactions are not under the influence of ads



Data Preparation

- Remove unnecessary attributes
- Create brand price index (average price of all products: 43.5)
- Create category column for each brand
- Create dummy variable

product_brand	
Perro	242.996173
Playtime	199.257311
K99	168.691506

brand price index example

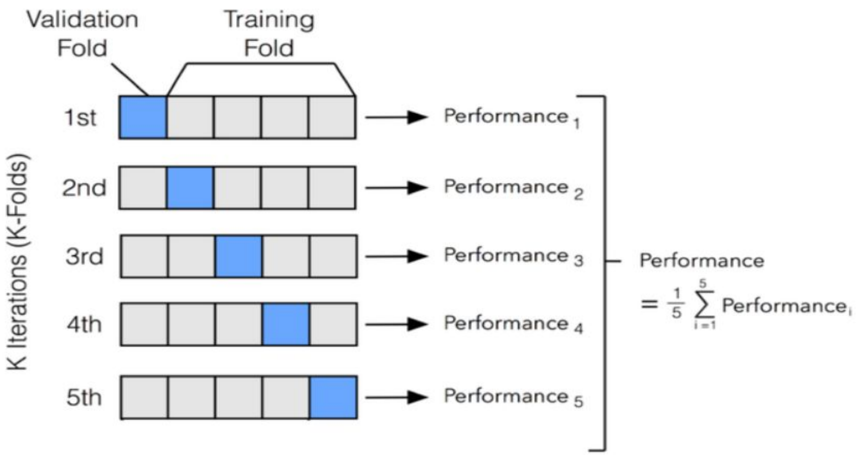
category	
Dried dog food	79
Wet dog food	9
Veterinary dog food	2
Dehydrated dog food	1
Freeze-dried dog food	1

category example

See Appendix for more information

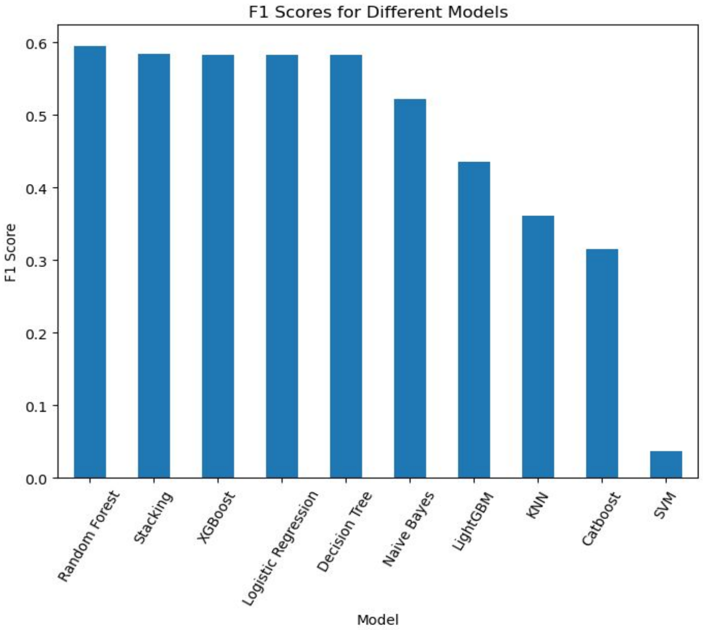
Model Building

- Split data into two parts, 70% for training; 30% for testing
- Set “purchase” as target variable, other columns as attribute
- Utilize 10-fold cross validation to avoid overfitting



Model performance - score

- Evaluate models based on **Accuracy, Precision, Recall, and F1 score**
- Tree-based models generate the best performance among all models

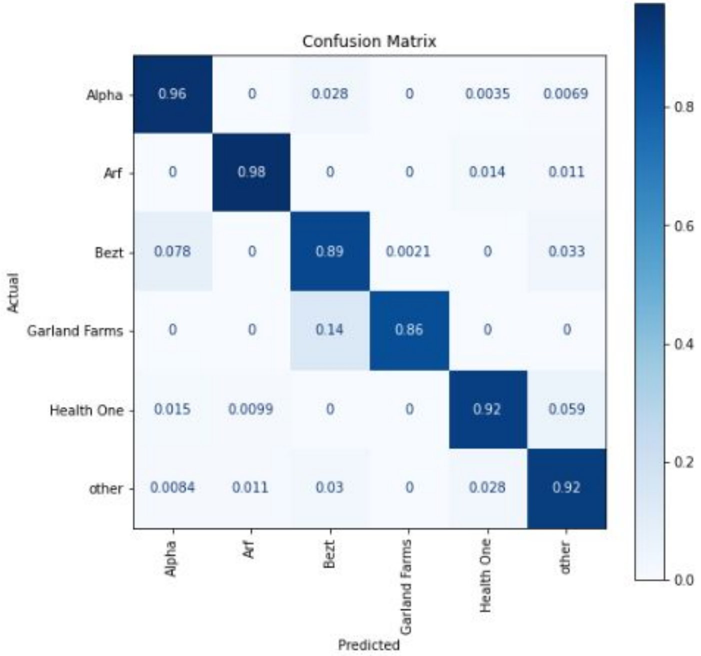


Feature importance:

- Price_index
- Date(Day, Month)
- Category of product(Dried, Wet)
- Brand

Multiple Classification Model

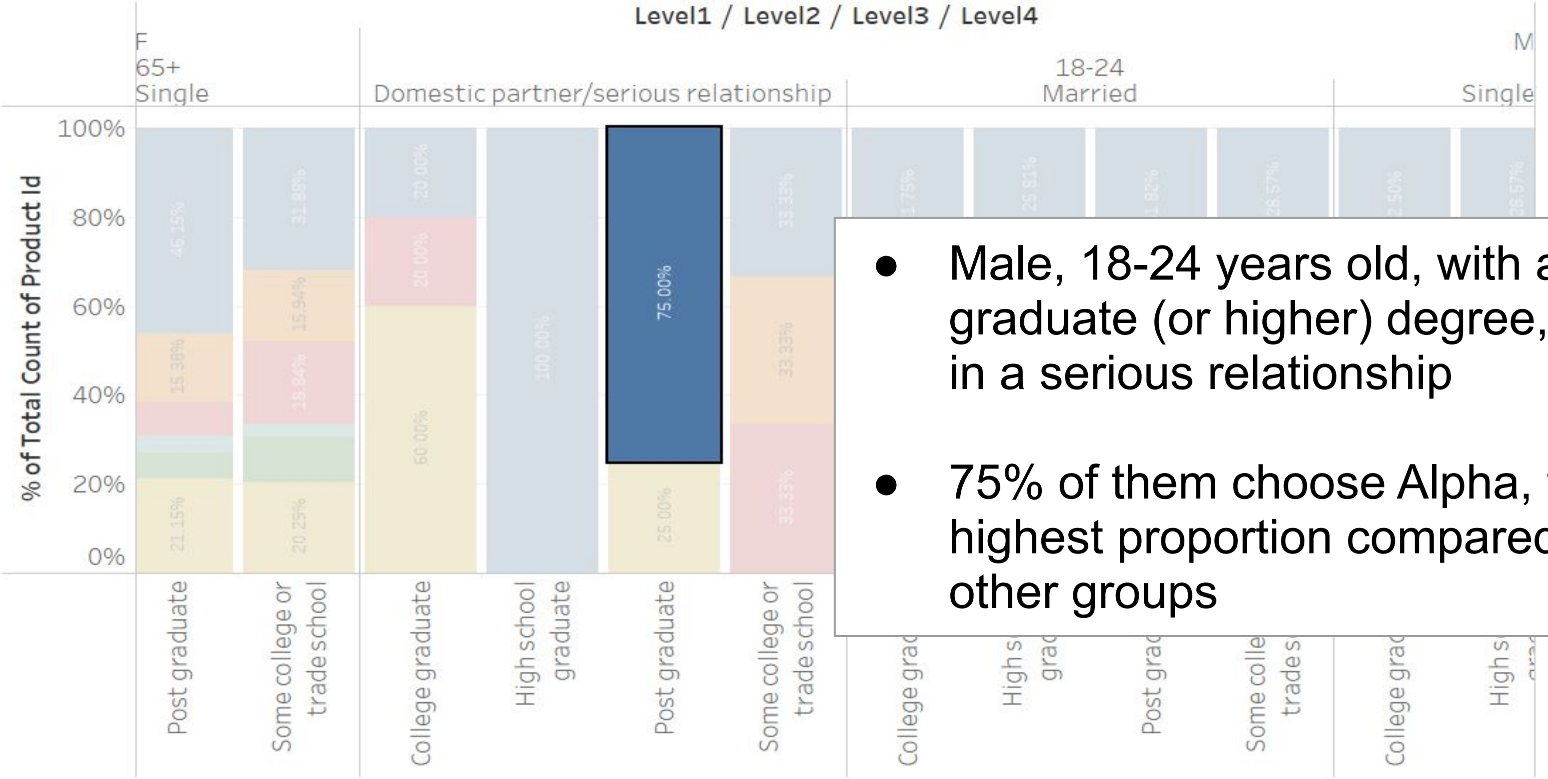
Decision Tree Multi-classification



Feature importance:

- Price
- Category of product(Dried, Wet)
- Date(Day, Month)
- Accuracy score: 0.93

Insights for Target Customers: Alpha



- Male, 18-24 years old, with a graduate (or higher) degree, & in a serious relationship
- 75% of them choose Alpha, the highest proportion compared to other groups

See detailed visualization through Tableau link

Insights for Target Customers: Alpha

Male, 18-24 years old, with a graduate (or higher) degree, & in a serious relationship

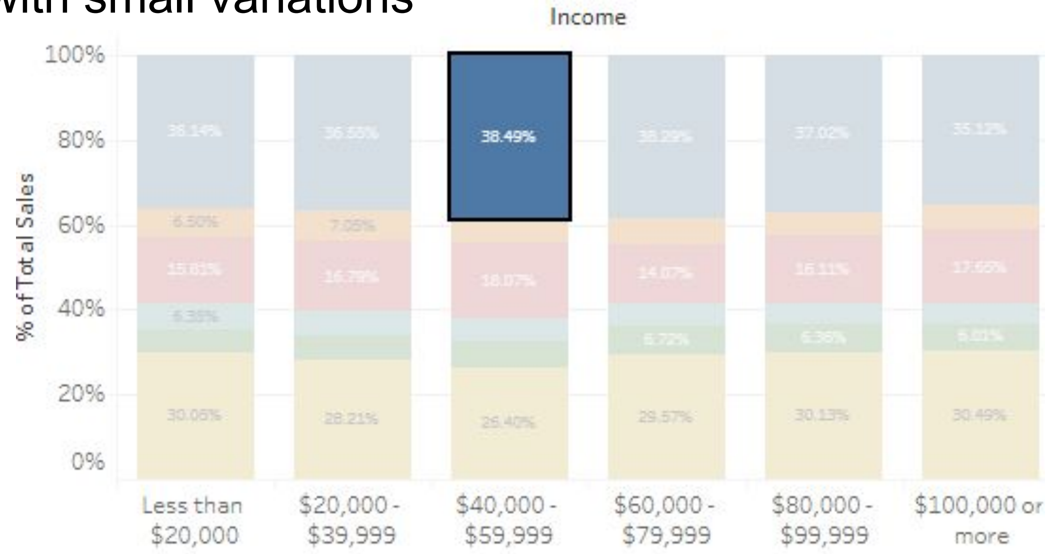
They are most likely from South regions of USA

Insights for Pricing Strategy - Alpha

- **Middle-high price range:** Alpha's prices are middle-high among top 5 brands
- **Small price variation**: most-purchased products are similarly priced

Insights for Pricing Strategy - Alpha

- **Average income customers**: 38.49% customers who make $40K - $60K choose Alpha, the highest proportion compared to other income groups
- **Overall,** more average income customers choose Alpha, whose prices are middle-high with small variations



Insights for More Brands

Explore the Tableau Dashboard to develop insights for other top brands

Exploration Tips:

1. Want to only look at age groups? Increase or decrease demographic criterias by editing levels on the right!
2. Wonder where your best customers are? Click on a bar in "Demographic Distribution for Each Brand" to find that customer group's regions on the map
3. Want to only look at one brand? Try hover over the dashboard; for the map, select the brand in the filter below the map to find all customers' regions

Recommendations

amazon

Target Customers

Find the most valuable customer segment for each brand & which US region are they from, and target this brand at these customers

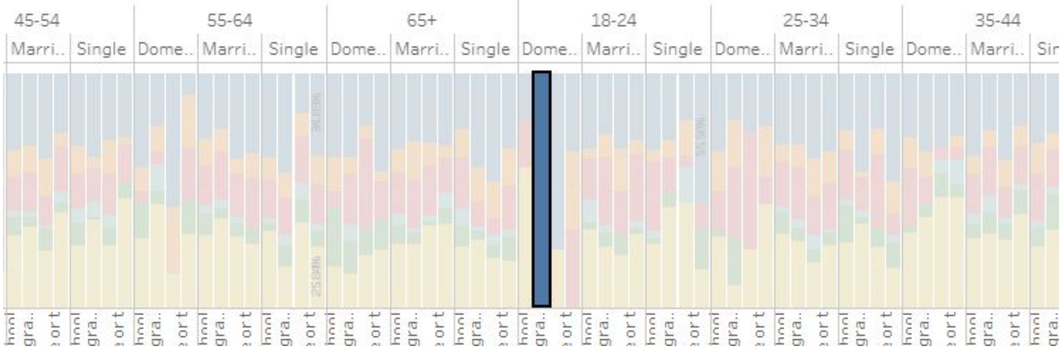
Pricing Strategies

Show brands their price range statistics and the income levels of their customers to help brands design pricing strategies

(e.g. how much discount if customers purchase through ads)

Limitations

1. **Small data size:** Due to the small data size, the distribution of customers in each segment are likely biased.
   a. For example, some customer segments have 100% people buying one brand, but mostly this customer segment only has one person
   b. We can avoid this problem in real life if we have enough data for each customer segment



APPENDIX

- Data preparation
- Model performance

Remove unnecessary attributes

- **Sale ID, Product ID:**
Unique identifiers are not useful in the prediction

- **Product Name:**
Textual variables are not useful after NLP

- **Quantity:**
Not replicable when we expand the dataset later

- **Zip, Lat, Lng:**
Not useful because is highly related to “City” and “State”

Create brand price index

Create brand price index

(Brand average price / average price of all brands) * 100

```
product_brand
Perro              242.996173
Playtime           199.257311
K99                168.691506
```

gender	city	st	marital	education	income	age	prime	category	purchase	price_index	sale_year	sale_month	sale_day
M	Shreveport	LA	Single	2	3	4	True	Dried dog food	True	80.566583	2022	1	1
M	Columbus	OH	Married	1	1	4	True	Dried dog food	True	80.566583	2022	1	1

Cleaned dataset

Data Prep - Category

amazon

Create category column

1. Prioritize keywords and phrases for each product categories
2. Assign brand main category by most popular category of a brand

```
def  categorize_product_description(product_name):
        if  "freeze"  in  product_name  and  "dried"  in  product_name:
                return  "Freeze-dried  dog  food"
        elif  "dehydrated"  in  product_name:
                return  "Dehydrated  dog  food"
        elif  "wet"  in  product_name:
                return  "Wet  dog  food"
        elif  "diet"  in  product_name:
                return  "Veterinary  dog  food"
        else:
                return  "Dried  dog  food"
```

```
category
Dried dog food           0.858696
Wet dog food             0.097826
Veterinary dog food      0.021739
Dehydrated dog food      0.010870
Freeze-dried dog food    0.010870
```

```
category
Dried dog food           79
Wet dog food              9
Veterinary dog food       2
Dehydrated dog food       1
Freeze-dried dog food     1
```

Data Prep - Dummy Variables

Ordered Categorical Variables

Indexed “income”, “education”, “age”

```
#change other attributes has order to categorical value
df_modify.income[df_modify.income == 'Less than $20,000'] = 0
df_modify.income[df_modify.income == '$20,000 - $39,999'] = 1
df_modify.income[df_modify.income == '$40,000 - $59,999'] = 2
df_modify.income[df_modify.income == '$60,000 - $79,999'] = 3
df_modify.income[df_modify.income == '$80,000 - $99,999'] = 4
df_modify.income[df_modify.income == '$100,000 or more'] = 5

df_modify.education[df_modify.education == 'High school graduate'] = 0
df_modify.education[df_modify.education == 'Some college or trade school'] = 1
df_modify.education[df_modify.education == 'College graduate'] = 2
df_modify.education[df_modify.education == 'Post graduate'] = 3

df_modify.age[df_modify.age == '18-24'] = 0
df_modify.age[df_modify.age == '25-34'] = 1
df_modify.age[df_modify.age == '35-44'] = 2
df_modify.age[df_modify.age == '45-54'] = 3
df_modify.age[df_modify.age == '55-64'] = 4
df_modify.age[df_modify.age == '65+'] = 5
```

education	income	age
College graduate	$60,000 - $79,999	55-64
Some college or trade school	$20,000 - $39,999	55-64

education	income	age
2	3	4
1	1	4

Data Prep - Dummy Variables

Index Ad_exp

Index with 0,1,2 since we want to **regroup** this categorical variable

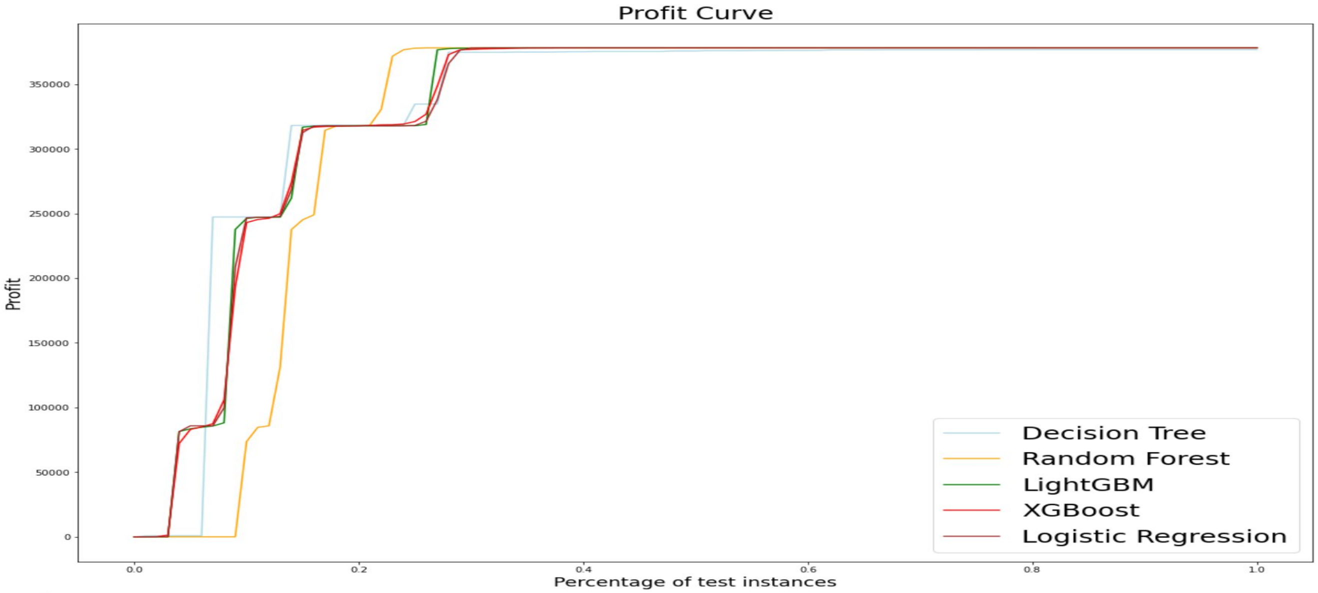
```
df_modify.ad_exp[df_modify.ad_exp == 'Sponsored Brands'] = 0
df_modify.ad_exp[df_modify.ad_exp == 'Some other type of ad'] = 0
df_modify.ad_exp[df_modify.ad_exp == 'Sponsored Products'] = 0
df_modify.ad_exp[df_modify.ad_exp == 'Display/banner ad'] = 1
df_modify.ad_exp[df_modify.ad_exp == 'Video ad'] = 1
df_modify.ad_exp[df_modify.ad_exp == "Don't recall seeing an ad"] = 2
```

Regular dummy variable

Get dummy variables from “product brand”, “gender”, “city”, “st”, “marital”, and “category”, since they are unordered

Model performance - profit curve

- Average Sales per transaction: **$39.02**
- Average cost per click of Ad: **$0.97**
- Average conversion rate: **10%**



Model performance - expected value

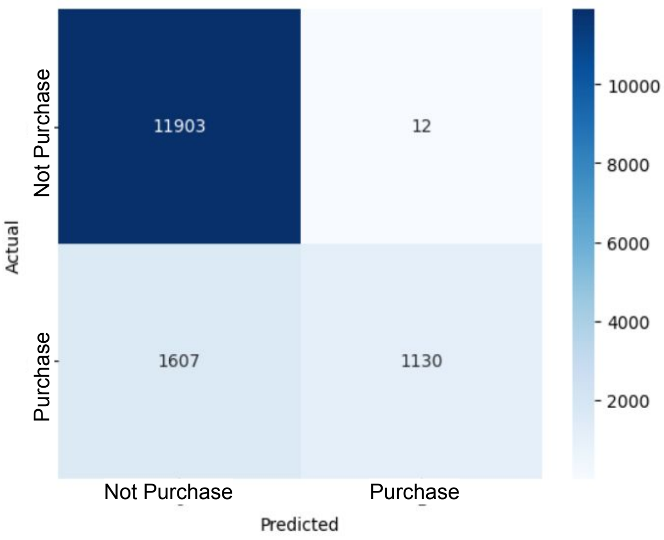
Cost-Benefit Matrix		
	Actual Purchase	Actual Not Purchase
Predicted Purchase	39.02	-9.7
Predicted Not Purchase	0	0

Confusion Matrix (for Random Forest optimized parameters)		
	Actual Purchase	Actual Not Purchase
Predicted Purchase	1130 (7.71%)	12 (0.08%)
Predicted Not Purchase	1607 (10.97%)	11903 (81.24%)

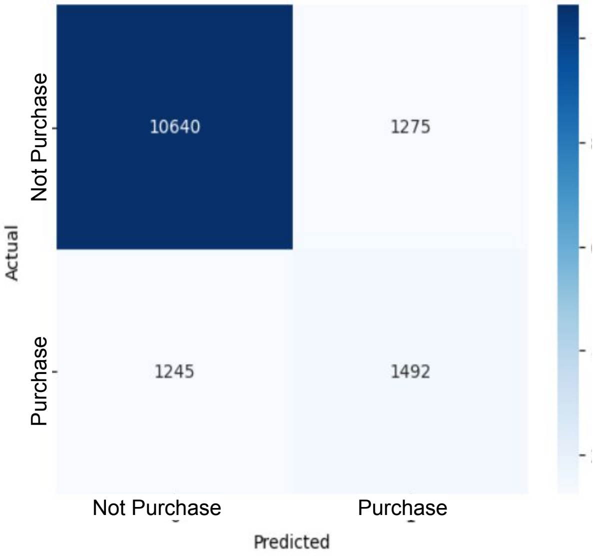
Expected Value(per ad):
($39.02)*(7.71%)+($-9.7)*(0.08%)+0+0=$3.00

Model performance - confusion matrix

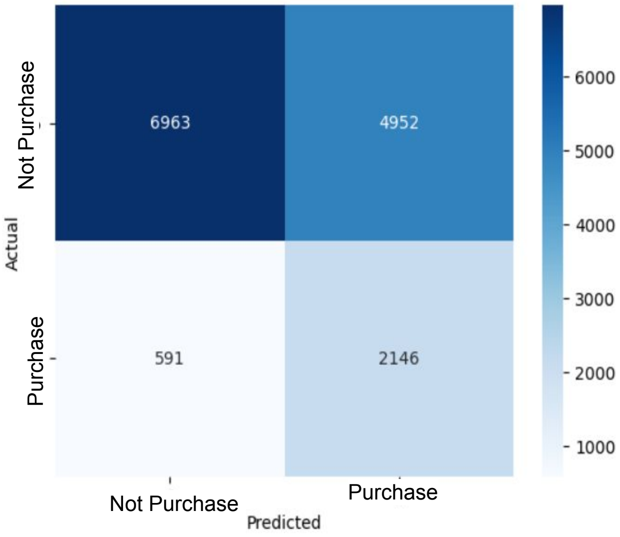
- Utilize machine learning to predict the possibility of purchase
- Build confusion matrix by different class-weight



Class weight {0:1, 1:1}



Class weight {0:1, 1:3}



Class weight {0:1, 1:10}