Wayfair Case Study: Using Data to bring Customers Home

Authors: Devanshi Gariba, Alefiya Naseem

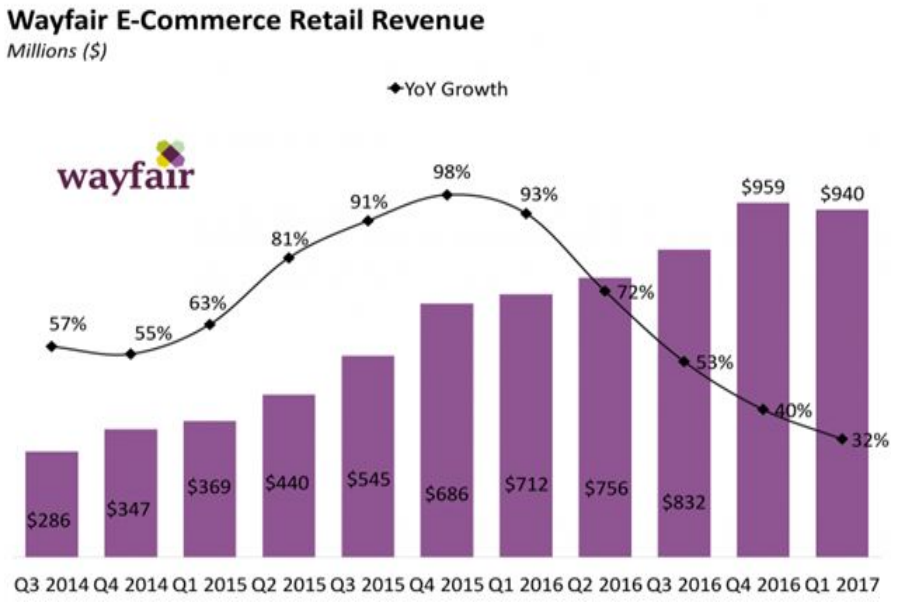
Project Overview

- Wayfair is a large e-commerce retailer
- Large B2B (Business-to-Business) division that sells to business customers
- Leverage data science to provide them excellent service
    - Predict customer needs
    - Predict purchasing patterns

Goals

- Use of Wayfair's B2B customer interaction dataset
- Work on the following problems:
    - Whether a B2B customer will purchase or not in the next 30 days
    - How much a B2B customer will spend in the next 30 days
    - Deploy as a web app

Our aim is to not just take care of high retention and revenue prediction accuracies but also to drive insights from these predictions, evaluate and provide reasoning for which models seem to work best for our use-case.



Data

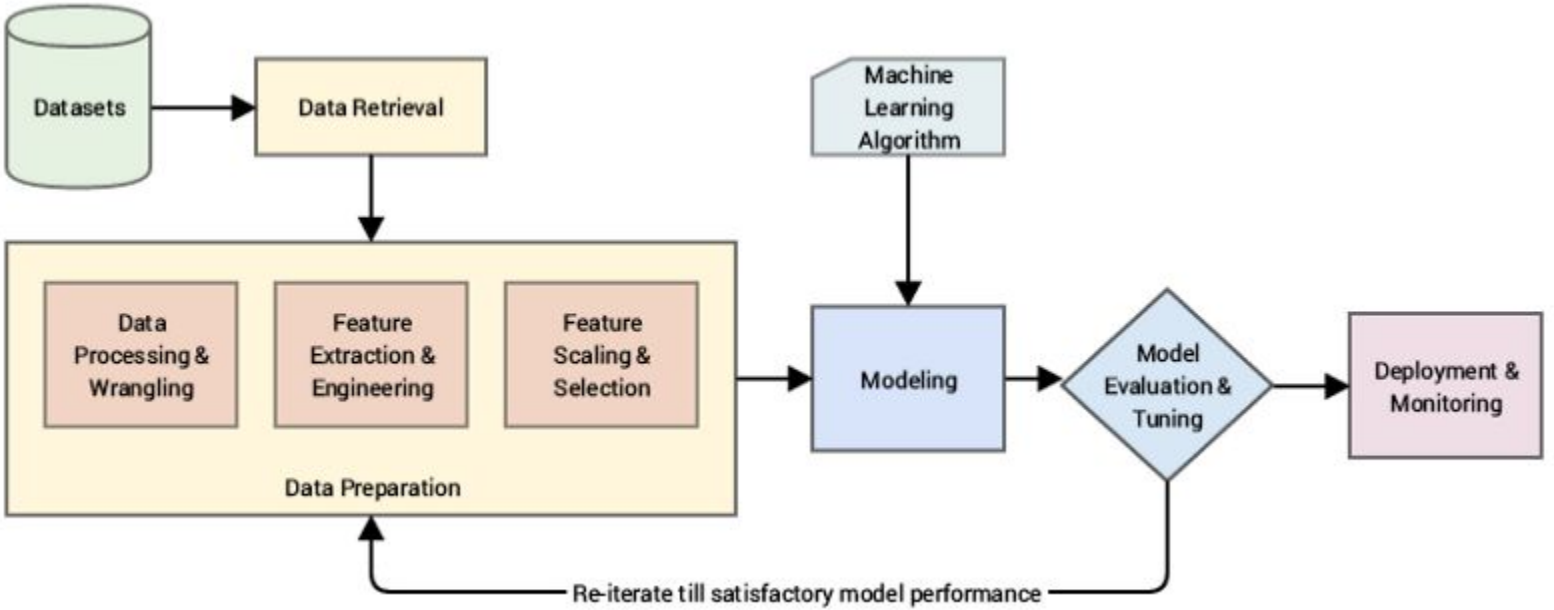
Training data : This data includes 181 features. These features can be divided into the following feature groups.

- **Two outcome variables:**
  - convert_30 (boolean)
  - revenue_30 (numeric)
- **Customer:** basic customer information like status, type, role, team, etc.
- **Enrollment questions:** number of employees, number of purchases and total cost of purchases per year
- **Order:** order information like number of orders, size, influence, etc. over a time frame

Data

- **Satisfaction:** customer satisfaction(customer issue) information over a time frame
- **Visit:** number of visit to the site or favorites list over a time frame
- **Search:** number of search terms over a time frame
- **SKU(Stock Keeping Unit):** SKU’s and their average price viewed over a time frame
- **Task:** task information like introduction, cadence, reassignment and others over a time frame
- **Call:** call information over a time frame
- **Email-BAM:** Information about emails exchanged between customer and sales rep
- **Email-Wayfair:** Email subscription information

Machine Learning Pipeline



Research Questions

1) What are the most appropriate methods for tidying the dataset? How to handle missing data, categorical data?
    a) How do we make sure our results are not biased?
    b) Will missing data imputation improve our prediction or yield high-performing feature-sets?
    c) Does missing data have any effect on feature-subset selection?
2) Dataset is highly unbalanced. What are the appropriate sampling techniques to deal with class imbalance since there's a huge amount of missing values too?
    a) Identify and implement appropriate resampling techniques to deal with class imbalance.
3) How to develop an intuition from the dataset given the high dimensionality of the problem?
    a) Are we able to produce interesting visuals demonstrating the relationship between the target variable and the associated features?
    b) Perform exploratory Data Analysis

Research Questions

4) Can we use data adaptive machine learning algorithms to predict whether a B2B customer will purchase or not in the next 30 days.

a. There are an overwhelming number of features in the dataset. What are the most important features that help in the prediction of the conversion.

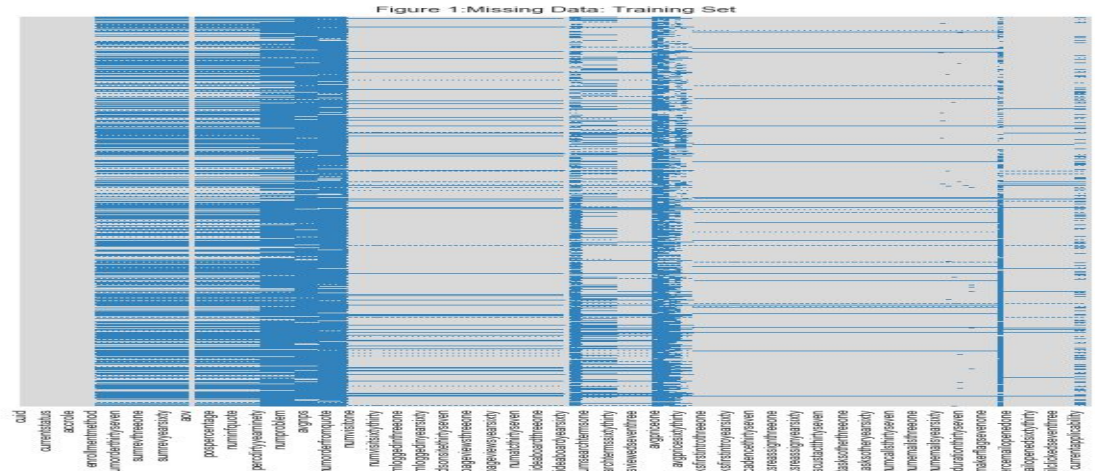
b. What techniques can be useful in determining the right feature-set?

c. Determine the optimal number of features in predicting whether a customer will purchase or not. Infer different feature sets to see if there is a logical explanation and interpretation of why some features are important over others and see if we can personalize efforts on customers that don't convert.

d. What modelling strategies best help in predicting the score?

Data Cleaning, Missing Values

- There were 181 features categorised into 12 parts.
- Once every feature was set to its required datatype we took a look at missing values.



Data Cleaning, Missing Values

- The heatmap shows many missing values, more precisely, there were continuous features which had at least 70% missing data.
- Normally, we would omit such features entirely as it’s difficult to make reliable imputations based on just 30% of your data but these were intuitively important features like number of previous orders, average overall value, emails opened etc.
- So we made the following assumptions for imputation.

Data Cleaning, Missing Values

- Most of the features that had missing values were intuitively correlated to the outcome of conversion. Take for example emails opened to conversion rate(outcome variable).
- Looking at the remaining 30% of the data and subsetting by class we saw that most customers who did not convert within 30 days barely had any activity prior to the time period we were looking at and vice-versa.
- So we subsetted the missing data by class, imputed 0 for observations that did not convert and imputed median for customers who did convert.
- This seemed like a safer assumption than median imputing everything.

Handling Categorical Data

- All the categorical variables did not have more than 5 levels at best.
- So dummy encoding all categorical variables was preferred as it did not blow up the dimensions much.

Normalizing and Standardizing Data

- There weren’t necessarily any numerically large features, but they were either proportions or integers. So some kind of scaling was needed.
- Min-Max scaling was used on all the features to avoid some errors in the chi-squared tests which can’t work for negative values.
- Lasso Regularization needed standardizing variables since it applies

Handling Class Imbalance

- This is a highly imbalanced dataset with just about 3000 customers converting out of a possible 28,000.
- We used SMOTE(Synthetic Minority Oversampling Technique) sampling to oversample the minority class instead of resampling existing observations.
- Note that it was crucial to oversample just the train data after splitting into train-test sets, and not oversample any test data to avoid model bias on synthetic test samples.
- Lastly, besides accuracy, recall is used as a metric to evaluate models.

Feature Selection Techniques

1. **Filter Method:** Using Pearson's Correlation Coefficient, the filtering here is done using correlation matrix and it is done using Pearson correlation. All the relevant features whose correlation with the target (more than 0.2) were extracted. Correlation between the features was checked and the highly correlated features were dropped to remove redundancy.
2. **Wrapper Method**: The Recursive Feature Elimination (RFE) method works by recursively removing attributes and building a model on those attributes that remain. It uses accuracy metric to rank the feature according to their importance. The RFE method takes the model to be used and the number of required features as input. It then gives the ranking of all the variables, 1 being most important. It also gives its support, True being relevant feature and False being irrelevant feature.

3. **Chi-square Tests**

Compute chi-squared stats between each non-negative feature and class. This score can be used to select the n_features features with the highest values for the test chi-squared statistic from X.

Dataset contains a lot of negative values and hence it was already processed to create a normalized dataset to perform Chi-square tests..

4. **Embedded Method: Lasso Regularization**

If the feature is irrelevant, lasso penalizes it's coefficient and makes it 0. Hence the features with coefficient = 0 are removed and the rest are taken.

Interpretation and Exploration of Feature Subsets

1. **Number of Visits**: Number of online visits in the past 1 day , Number of online visits in the past 1-3 days , Number of online visits in the past 30-7 days ,Number of online visits in the past 60-30 days.

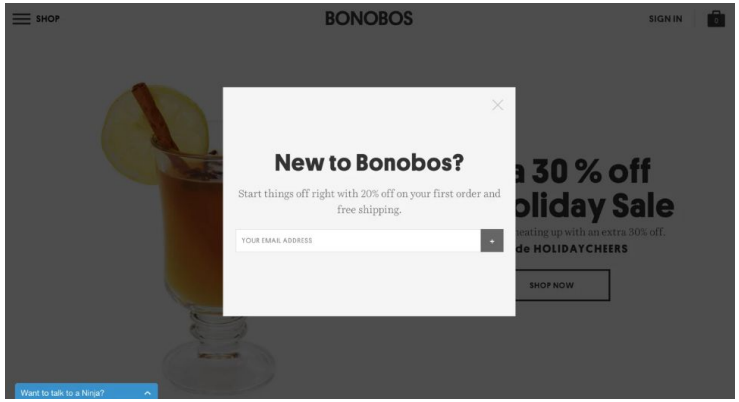
   This metric brings in the powerful idea of user segmentation - New users v/s Old users and it can be used in determining and designing new strategies around new and old users.

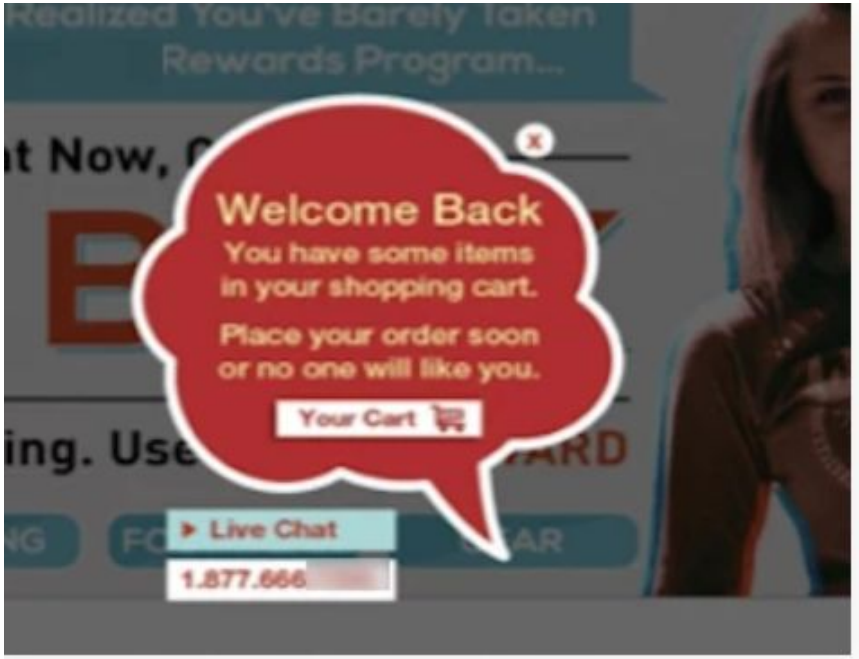
   Possible strategies for New visitors?

   a. When there is a large number of new visitors, one of the possible strategies could be to add them to their marketing list, send them promotional emails or give them a generous discount on their first purchase.

b. Possible strategies for old visitors?

When there are a large number of return visitors, it might be a good idea to tailor a personalized experience for them which is more consistent with their needs. Personalizing efforts to remind them of their previous purchases, product views, or interactions in general can help in boosting sales. Thus, this metric definitely does help in predicting purchase and if used wisely can help in channelizing efforts to convert the non-buying ones into the buying ones. The concept of a returning customer being easier to convert than a new customer can be validated by this set of relevant features as given out by the filter method.


SHOP
BONOBOS
SIGN IN
New to Bonobos?
Start things off right with 20% off on your first order and free shipping.
YOUR EMAIL ADDRESS
SHOP NOW
Want to talk to a Ninja?

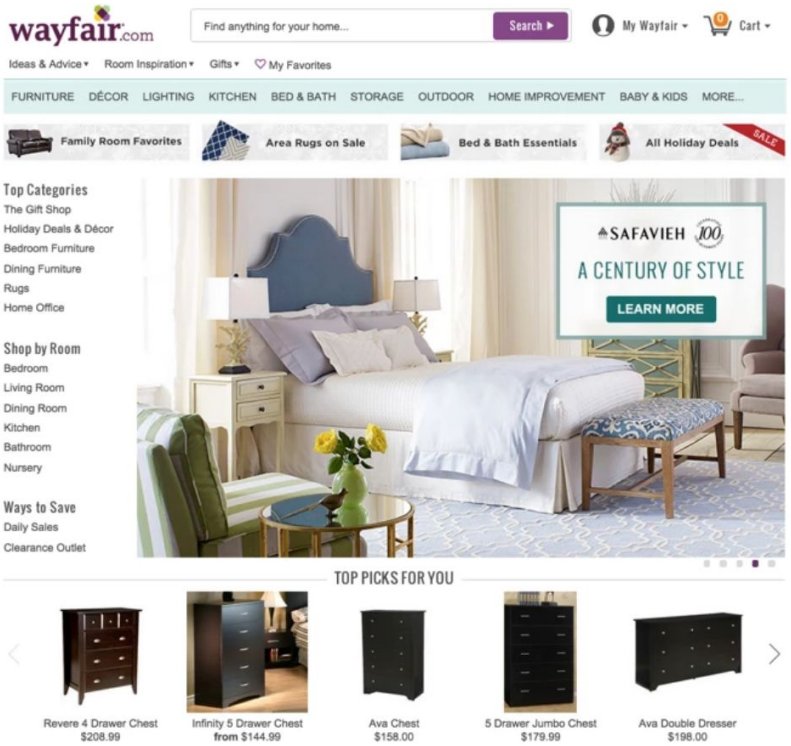

Rewards Program...
Welcome Back
You have some items in your shopping cart.
Place your order soon or no one will like you.
Your Cart
Live Chat
1.877.666

2. **Number of search terms in the past 30-7 days.**

What the customers typically want and what pages aren't meeting their needs.

a. If the number of search terms are low for a number of different products~ it probably means that search terms could be more specific and most likely coming from existing or returning customers who came back, knowing what they want.

b..If the number of search terms are general and appear larger in number, then it may mean customers can't find what they are looking for. Another important perspective from this metric could be to test the Search Results Page. Are the customers able to find their items of interest on Page 1? It becomes very important to show the most relevant results first. Showing related products or recommended products to the returning customers can definitely pass the personalization test.For eg. showing Top picks.


wayfair.com
Find anything for your home...
Search
My Wayfair
Cart
Ideas & Advice
Room Inspiration
Gifts
My Favorites
FURNITURE
DÉCOR
LIGHTING
KITCHEN
BED & BATH
STORAGE
OUTDOOR
HOME IMPROVEMENT
BABY & KIDS
MORE...
Family Room Favorites
Area Rugs on Sale
Bed & Bath Essentials
All Holiday Deals
SALE
Top Categories
The Gift Shop
Holiday Deals & Décor
Bedroom Furniture
Dining Furniture
Rugs
Home Office
Shop by Room
Bedroom
Living Room
Dining Room
Kitchen
Bathroom
Nursery
Ways to Save
Daily Sales
Clearance Outlet
SAFAVIEH
A CENTURY OF STYLE
LEARN MORE
TOP PICKS FOR YOU
Revere 4 Drawer Chest
$208.99
Infinity 5 Drawer Chest
from $144.99
Ava Chest
$158.00
5 Drawer Jumbo Chest
$179.99
Ava Double Dresser
$198.00

3. **Days since last order** and **Days since last visit**.

These metrics could specifically be important in the context when customers visit the website, come back more times, depending on the price point and also on why the brand exists, and then maybe close the deal. These metrics also enhance the understanding of “how long” it takes for customers to buy from the website and is that behavior different across different segments of the website customers. If exploited the information wisely, it can definitely help optimize the marketing campaigns, promotions, other efforts to boost conversion.

4. **Number of Add To Carts in the past 30-7 Days/ Number of Add to Favourite Lists**

While it is an intuitive one, there are many factors around this feature that can lead upto the actual conversion - Is there a guest check out option to enable faster processing (even if it’s a bulk order)? Is there any shipping estimate provided when the product is added to the cart? Now that the customer has already made up his mind but is only a few clicks away - Are there any personalized efforts involved in reminding him to complete the purchase?

5. **Number of Seconds on the website**

It inarguably is another very important metric - How long is the website able to interest the user? Is the website search efficient in producing relevant results quickly? Again, is there any personalization to keep the user engaged? How is the product recommendation working. If the number is really low - we can definitely predict a higher bounce rate - The bounce rate is the rate at which new visitors visit the site and immediately click away without doing anything.

Common problems may include poor website design, low usability, or higher load times that could potentially affect the conversion rates.

6. **Percentage of orders in the past 30 days that were a 'dirty' (problematic) order**

E-satisfaction is definitely an important factor and lesser the number of problematic order, the greater the satisfaction. While this metric can be a direct metric by itself, there are a range of factors around this metric which can still help if the percentage of dirty orders is high. Customer obsession can be of the greatest value - How is the company tackling dirty order issues? how efficient and fast is the dirty-order and re-order processing? How does the company make sure it is not losing on its valuable customers?

Is there any financial or product based incentive, especially because the case study concerns B2B customers and the orders associated with such customers are the bulk orders.

7. **Percentage of Emails Opened from Wayfair** and **Number of Emails between Customer and Sales Representative**

Email marketing is one of the most common forms of product marketing - often times A top priority for email marketing is to increase the subscriber engagement, which in turn has the power to increase sales, average revenue per customer and in general keep the customers informed about the new launches, on-going promotions and recommendations around their past purchases and in some way anticipate their future needs.

Modelling Strategies

1. **Logistic Regression:**

   We fit a Logistic Regression model on all the feature subsets for predicting whether a customer will purchase in the next 30 days or not. While it is a simple linear classification model, it did not work really well on all of our feature-subsets. Model was fit on 5 different feature sets (starting with all features and working with feature-subsets extracted via techniques described above)) and the corresponding accuracies were 0.77, 0.75 , 0.75, 0.78, 0.73

Modelling Strategies

2. **Decision Trees:**

Logistic Prediction accuracies weren't great, decision trees have better prediction powers. We fit decision trees on all the feature subsets. It was quite surprising to see the entire dataset underperforming in comparison to the other feature subsets because of the tendency of decision trees to oftentimes overfit.

This modelling strategy yielded accuracies of 64.20, 66.263,63.18,76.53 and 65.49 and recall scores of 0.33, 0.33,0.31, 0.46 and 0.33.

Again, this method did not yield predictive accuracies or good precision/recall measures.

Modelling Strategies

3. **Random Forests:**

While it was easy to identify the star feature-subset via decision trees, the predictive scores were still not great.

Random Forests performed quite well with great predictive accuracies and AUC scores

```
Feature_subsets = [all_features,Feature_Selectionset1,Feature_Selectionset2,chi_feature,FeatureSelectionset4]

for i in Feature_subsets:
    Random_Forest(i)
```

```
0.9908568913492006 0.8903306078919303 0.8030183818786798 203
0.9707371878183115 0.8807323142552436 0.7479456155570938 13
0.9786412864290794 0.8777106292214717 0.7639856518869154 21
0.9873880366734099 0.8905083540703875 0.7876082418933261 30
0.9787651591796784 0.8871311766797014 0.7603088494363923 12
```

Modelling Strategies

4. **Gradient Boosting:** Similarly, gradient boosting after performing a hyperparameter tuning sweep, performed just as good as random forests.

Future Work

1. Phase-1 was heavily focussed on data cleaning and wrangling, handling class imbalance and most importantly on identification of important feature subsets and interpreting the actionable features. A lot of classification modelling strategies have been discussed, we will continue to work on modifying the strategies with hyperparameter tuning and discussing more strategies.
2. Second Phase of the project focuses on the ***regression problem***: If a customer is predicted to most likely purchase in the next 30 days - How much a B2B customer will spend in the next 30 days?
3. Deploy the machine-learning pipeline as a web-app which helps in prediction and also suggest actions and strategies as discussed in the previous section based on the feature values.

We would love to know your thoughts :)