Predictive Analytics for Expedia

MKT 6342 Customer Insights and Data Visualization

Term Project Report

5th December,2016

Group 7

Prashant Shinde

Anjali Singh

Aditya Belpathak

Sandhya Ravindran

Contents

Executive Summary

While heaps of data exist in the travel sector, from how seasonality affects bookings to the types of travel packages that receive the highest conversion rates among consumers, leveraging some of the more unstructured streams into effective analytical modeling can be a challenge. Increasing profitability in this complex industry requires significant legwork on the part of businesses from being in the right place at the right time to using the predictive capabilities of the available data. No one has the ability to capture and analyze data from the future — however, there is a way to predict the future using data from the past. It's called predictive analytics and more and more businesses are doing it every day.

The dataset provided for analysis is from the "an Empirical analysis of the value of the complete information for eCRM models" which includes various site-centric and user-centric variables related to Expedia Website. The objective of our analysis is to prescribe the best predictive model to determine whether a particular user will make a booking through the Expedia website or not.

Thus, the main focus of the analysis was on creating a **recommender system** for the Expedia website to monitor the factors influencing the purchase of tickets from the website. Additionally, we will also be doing the **customer analytics** for Expedia. A detailed customer analytics seemed important as, the buying pattern for different segments based on the demographical factors such as age, gender, family size was observed to be different.

Based on the insights observed from the predictive model created and customer analytics we have also created some strategies which could be followed by Expedia to personalize the recommender system for each customer, thus gaining the loyalty of its customers and hence, increasing the overall tickets purchase. We have also created some strategies based on the various demographics segmentations, which will be useful for targeting the different created segments.

About Expedia

Expedia, Inc. (headquartered in Bellevue, WA, United States) owns and operates numerous international global online travel brands such as Expedia.com, Hotels.com, Hotwire.com, trivago, Egencia (formerly Expedia Corporate Travel), Venere.com, Egencia, Expedia Local Expert, Classic Vacations, Expedia CruiseShipCenters, Travelocity, Orbitz, and HomeAway. Expedia, Inc.'s companies operate more than 100 branded points of sale in more than 60 countries. Its clout extends over travel bookings for over 10,000 partners such as airlines and hotels, consumer brands, and high traffic websites through Expedia Affiliate Network; 80% of which is powered by their API.

The following infographic provides more details about Expedia:

Problem Statement

Expedia wants to take the proverbial rabbit hole out of purchase by providing personalized recommendations to their users. This is no small task for a site with hundreds of millions of visitors every month!

Currently, Expedia uses search parameters to adjust their booking recommendations, but there aren't enough customer specific data to personalize them for each user. Through this problem we are trying to contextualize customer data and predict the likelihood a user will make a certain booking.

The data set containing 3121 records includes 40 input variables out of which 6 are Binary, 5 are Nominal and rest are Interval. The variables provided in the dataset are as follows:

No.	Variable	Description
1	gender	"1"—Male, "0" – Female
2	age	Age of the user
3	income	Income of the user
4	edu	"0" – high school or less, "1"-- college, "2" – post college
5	hhsize	Size of house hold
6	child	"1" – have, "0" – not have
7	booklh	No. of bookings the user made at this site in the past
8	sesslh	No. of sessions to this site so far
9	minutelh	Time spent in this site so far in minutes
10	hpsesslh	Average hits per session to this site
11	mpsesslh	Average time spent per sessions to this site
12	booklc	Dummy variable, indicating if the user has booked at this site up to this point in the current session
13	httlc	No. of hits to this site up to this point in this session
14	minutelc	Time spent up to this point in this session
15	weekend	Indicating if this session occurs on weekend
16	bookgh	No. of past bookings of all sites so far
17	sespsite	Average sessions per site so far
18	sessgh	Total no. of sessions visited of all sites so far
19	minutegh	Total minutes of all sites
20	hpsessgh	Average hits per session
21	mpsessgh	Average minute per session
22	awareset	Total no. of unique shopping sites visited
23	basket	Average no. of shopping sites visited per session
24	single	Percentage of single-site sessions
25	booksh	Percentage of total bookings are to this site
26	hitsh	Percentage of total hits are to this site
27	sessh	Percentage of total sessions are to this site
28	minutesh	Percentage of total minutes are to this site
29	entrate	No. of sessions start with this site/total sessions of this site
30	peakrate	No. of sessions the user spend the most time within this site/total sessions of this site
31	exitrate	No. of sessions end with this site/total sessions of this site
32	SErate	No. of sessions coming from search engines/total sessions of this site
33	*bookgc*	Binary variable, indicating if this user has booked at any sites up to this point in the current session
34	*hitgc*	Total hits of all sites in the current session
35	*basketgc*	No. of shopping sites in this session
36	*minutegc*	Time spent of all sites in this session
37	*SEgc*	Indicating if this session uses search engines
38	*path*	Indicating if this site is an entry/peak
39	*hitshc*	Hits to this site/ hits to all sites in this session
40	*minutshc*	Minutes to this site/total minutes in this session
41	bookfut	Binary dependent variable, indicating if this user is going to book in the remainder of the session (after the clipping point)

We will use **Tableau, SAS Enterprise miner and SAS base** to get insights of the dataset. **The first part of the project is to create a stable Recommender system**, we will be performing the following steps:

1. **Basic data preprocessing**: Explore the statistical properties of the variables in the input data set. The results that are generated in this step will give you an idea of which variables are most useful in predicting the target response.
2. **Explorative analysis**: This will help us observe the influence of various demographic variables on the target response. This will be done using Tableau.
3. **Building Predictive models**: We will be creating 3 predictive models namely decision trees, Neural networks and regression models. We will then compare the above models we tried, and select a champion model. When evaluating the model performance, we will try to use confusion matrix as the main evaluation criterion
4. **Insights development and strategy creation**: We will develop strategy of targeting, acquiring and retaining customers from different segments.

   The second part of the project is to perform customer analytics.

   We will be performing the following steps for it:
5. **Factors influencing booking:** we will dig deeper into the factors that influence booking by customers using tableau.
6. **Touch Points and Pain points:** we will segment our customers into different segments according to their behavior and things that motivate them to purchase**.**
7. **Manage and Channel:** we will try to formulate different strategies to manage and channel our customers.

RECOMMENDER SYSTEM

Recommender systems are commonly defined as applications that web sites exploit to suggest products and provide consumers with information to facilitate their decision-making processes. They implicitly assume that we can map user needs and constraints, through appropriate recommendation algorithms, and convert them into product selections using knowledge compiled into the intelligent recommender. Knowledge is extracted from either domain expert (content- or knowledge-based approaches) or extensive logs of previous purchases (collaborative-based approaches). Furthermore, the interaction process, which turns needs into products, is presented to the user with a rationale that depends on the underlying recommendation technology and algorithms. For example, if the system funnels the behavior of other users in the recommendation, it explicitly shows reviews of the selected products or quotes from a similar user. Here we will try to create a recommender system for Expedia.

BASIC DATA PREPROCESSING

We could see that variable X 38 indicating whether the site is entry/peak has 1831 missing variables which is almost 59% of the total records and variable x32 indicating no of sessions coming from search engines/total session of Expedia site, has 1594 missing values which is almost 52% of the total records.

The other variables in the dataset also contains missing values but the missing percentage is less than 10 %.

Since both X32 and X38 has more than 50% missing values we decided to reject these two variables manually.

While checking the distribution of the input variables, we could skewness in many of the interval variables.

We have checked for collinearity among the variables and found out that below variables have really high correlation

x18 and x19 (0.89)

x36 and x14 (0.79)

x36 and x34 (0.79)

Data partition

We have used 70% of data for training the model and 30% for validation purpose. We have also set the random seed to 500 for random sampling of our input data based on hit and trial method.

Impute

Since Most of our input variables has missing values, we have used the Impute node to impute the missing values with default input method as “mean” for interval variable and “count” for class variables. Before using logistic regression and neural network, imputation is an important step since both the classifiers are sensitive to missing data.

Target selection

The main purpose of our project is to determine whether the user will make a booking through the website or not. The dataset defines variable “depend” as the target variable .But the target variable “depend “ only indicates if the particular user is going to book in the remainder of the session or not (after the clipping point)

One interesting observation that we found while analyzing the data was that variable x12 and the target variable “depend” is negatively correlated.

Variable 12 indicates if the user has booked at this site up to this point in the current session.

Since our main objective is to predict whether a user will book though the site or not, the actual “target” for us should be the combination of variable x12 and variable “depend”

Hence, the dependent variable was combined with x12 to create a new target variable “newtarget” which will correctly predict whether a user is going to book the ticket or not.

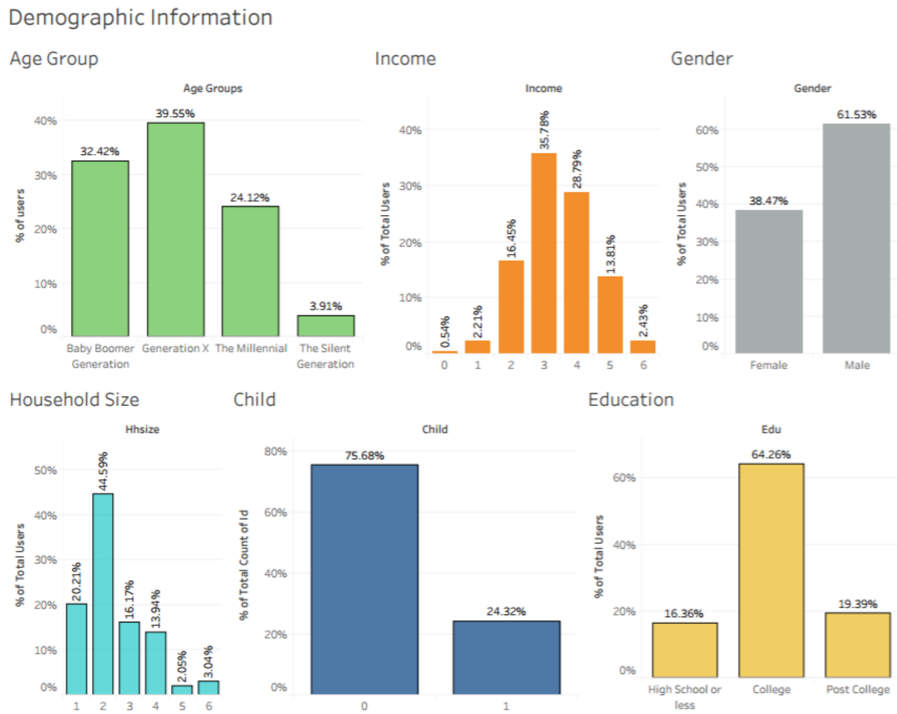
Code for combining x12 with depend to form new target

```
data aba.expedia_newtarget_notMI;
set aba.expedia;
if x12 = 1 or depend =1 then newtarget = 1;
else newtarget = 0;
run;
```

After that we have rejected variable “depend” and variable x12 from the new dataset with new target.

Exploratory Analysis

The entire agenda of exploratory analysis was to understand the distribution of overall demographic data. By using tableau, we created histograms for Age, Gender, income, household size etc. The dashboard for these datasets are as follows:



Predictive Models

Optimal Decision Tree

Decision tree models are conceptually easy to understand and they readily accommodate nonlinear associations between input variables and one or more target variables. They also handle missing values without the need for imputation.

Our First tree is an optimal decision tree with SAS default setting. The total no of leaves in this optimal decision tree is 14 i.e. the tree is pruned at 14.

Fit Statistics:

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
newtarget		_NOBS_	Sum of Frequencies	2183	938	.
newtarget		_MISC_	Misclassification Rate	0.137426	0.1258	.
newtarget		_MAX_	Maximum Absolute Error	0.926263	0.926263	.
newtarget		_SSE_	Sum of Squared Errors	495.1289	198.2822	.
newtarget		_ASE_	Average Squared Error	0.113406	0.105694	.
newtarget		_RASE_	Root Average Squared Error	0.336757	0.325106	.
newtarget		_DIV_	Divisor for ASE	4366	1876	.
newtarget		_DFT_	Total Degrees of Freedom	2183	.	.
newtarget		_APROF_	Average Profit for newtarget	-0.44709	-0.40618	.
newtarget		_PROF_	Total Profit for newtarget	-976	-381	.

The misclassification rate for the validation data in this optimal decision tree model is 0.1258

Interactive Decision tree

We have created the interactive decision tree based on the logworth value which is –log(pvalue) of the variables

The decision tree that we created based on logworth value has total 26 leafs.

Fit Statistics:

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
newtarget		_NOBS_	Sum of Frequencies	2183	938	.
newtarget		_MISC_	Misclassification Rate	0.124141	0.152452	.
newtarget		_MAX_	Maximum Absolute Error	0.979592	1	.
newtarget		_SSE_	Sum of Squared Errors	415.6782	204.0936	.
newtarget		_ASE_	Average Squared Error	0.095208	0.108792	.
newtarget		_RASE_	Root Average Squared Error	0.308558	0.329836	.
newtarget		_DIV_	Divisor for ASE	4366	1876	.
newtarget		_DFT_	Total Degrees of Freedom	2183	.	.
newtarget		_APROF_	Average Profit for newtarget	-0.35181	-0.43497	.
newtarget		_PROF_	Total Profit for newtarget	-768	-408	.

The validation misclassification rate for our interactive decision tree is 0.1524

Regression and Neural Network

We have used the imputation, variable selection and transform variable node to make the data suitable for logistic regression and neural network

Logistic Regression

Since our target variable is a binary variable we performed logistic regression. Initially we tried stepwise, forward selection and backward selection separately on our dataset and finally decided that for our dataset forward selection suits best.

For our analysis we have received the following results:

Summary of forward selection:

```
NOTE: No (additional) effects met the 0.05 significance level for entry into the model.
```

Summary of Forward Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Validation Misclassification Rate
1	IMP_x33	1	1	422.6050	<.0001	0.1908
2	IMP_x7	6	2	173.1946	<.0001	0.1887
3	LG10_IMP_x14	1	3	60.0292	<.0001	0.1738
4	LG10_IMP_x40	1	4	28.3676	<.0001	0.1599
5	IMP_x35	14	5	45.9745	<.0001	0.1546
6	LG10_IMP_x11	1	6	13.2493	0.0003	0.1610
7	LG10_IMP_x24	1	7	9.9502	0.0016	0.1578
8	LG10_IMP_x34	1	8	6.9339	0.0085	0.1588
9	LG10_IMP_x21	1	9	3.8846	0.0487	0.1588

```
The selected model, based on the misclassification rate for the validation data, is the model trained in Step 5. It consists of the following effects:

Intercept  IMP_x33  IMP_x35  IMP_x7  LG10_IMP_x14  LG10_IMP_x40
```

Fit Statistics:

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
newtarget		_AIC_	Akaike's Information Criterion	1894.087	.	.
newtarget		_ASE_	Average Squared Error	0.128932	0.12261	.
newtarget		_AVERR_	Average Error Function	0.422833	0.41242	.
newtarget		_DFE_	Degrees of Freedom for Error	2159	.	.
newtarget		_DFM_	Model Degrees of Freedom	24	.	.
newtarget		_DFT_	Total Degrees of Freedom	2183	.	.
newtarget		_DIV_	Divisor for ASE	4366	1876	.
newtarget		_ERR_	Error Function	1846.087	773.7007	.
newtarget		_FPE_	Final Prediction Error	0.131798	.	.
newtarget		_MAX_	Maximum Absolute Error	0.98508	0.998542	.
newtarget		_MSE_	Mean Square Error	0.130365	0.12261	.
newtarget		_NOBS_	Sum of Frequencies	2183	938	.
newtarget		_NW_	Number of Estimate Weights	24	.	.
newtarget		_RASE_	Root Average Sum of Squares	0.359071	0.350157	.
newtarget		_RFPE_	Root Final Prediction Error	0.363041	.	.
newtarget		_RMSE_	Root Mean Squared Error	0.361061	0.350157	.
newtarget		_SBC_	Schwarz's Bayesian Criterion	2030.61	.	.
newtarget		_SSE_	Sum of Squared Errors	562.9169	230.0162	.
newtarget		_SUMW_	Sum of Case Weights Times Freq	4366	1876	.
newtarget		_MISC_	Misclassification Rate	0.162162	0.154584	.
newtarget		_PROF_	Total Profit for newtarget	-1128	-434	.
newtarget		_APROF_	Average Profit for newtarget	-0.51672	-0.46269	.

The misclassification rate for our logistic regression model is 0.1545

Neural Network

SAS Enterprise Miner has two nodes that fit neural network model: the Neural Network node and the AutoNeural node. We have used the Neural Network node for our modelling. To converge the model fitting process, we have set maximum iteration to 150.

Result for Neural network:

Fit statistics:

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
newtarget		_DFT_	Total Degrees of Freedom	2183	.	.
newtarget		_DFE_	Degrees of Freedom for Error	2077	.	.
newtarget		_DFM_	Model Degrees of Freedom	106	.	.
newtarget		_NW_	Number of Estimated Weights	106	.	.
newtarget		_AIC_	Akaike's Information Criterion	1813.19	.	.
newtarget		_SBC_	Schwarz's Bayesian Criterion	2416.166	.	.
newtarget		_ASE_	Average Squared Error	0.112499	0.118873	.
newtarget		_MAX_	Maximum Absolute Error	0.990235	0.988818	.
newtarget		_DIV_	Divisor for ASE	4366	1876	.
newtarget		_NOBS_	Sum of Frequencies	2183	938	.
newtarget		_RASE_	Root Average Squared Error	0.335408	0.34478	.
newtarget		_SSE_	Sum of Squared Errors	491.1689	223.0066	.
newtarget		_SUMW_	Sum of Case Weights Times Freq	4366	1876	.
newtarget		_FPE_	Final Prediction Error	0.123981	.	.
newtarget		_MSE_	Mean Squared Error	0.11824	0.118873	.
newtarget		_RFPE_	Root Final Prediction Error	0.35211	.	.
newtarget		_RMSE_	Root Mean Squared Error	0.34386	0.34478	.
newtarget		_AVERR_	Average Error Function	0.366741	0.395359	.
newtarget		_ERR_	Error Function	1601.19	741.6928	.
newtarget		_MISC_	Misclassification Rate	0.148878	0.149254	.
newtarget		_WRONG_	Number of Wrong Classifications	325	140	.
newtarget		_PROF_	Total Profit for newtarget	-877	-439	.
newtarget		_APROF_	Average Profit for newtarget	-0.40174	-0.46802	.

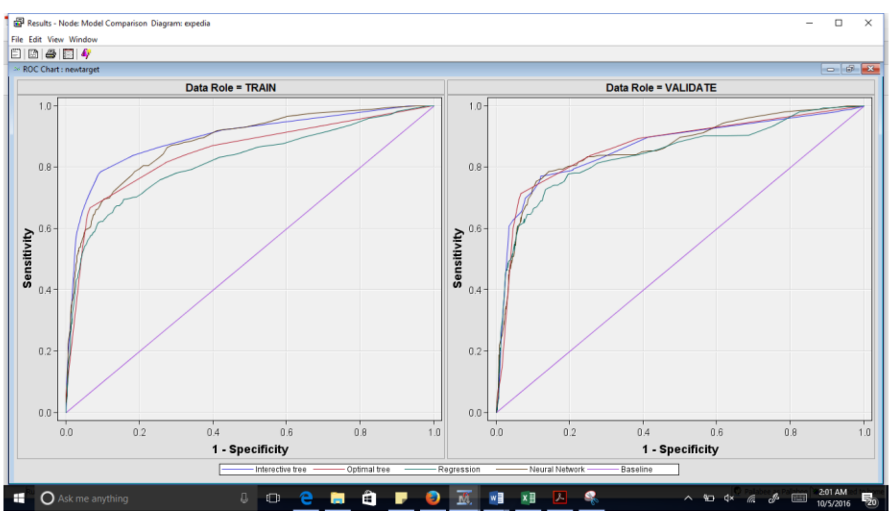
The validation misclassification rate for our neural network model is 0.1492

Model Comparison and Champion Model

Through model comparison, we aim to choose the champion model. We have compared the four model based on misclassification rate.

The validation ROC chart shows the comparison of performance by all the four models. The optimal tree model seemed to have higher performance than the other three models and this is mirrored in the fit statistics table and result output.

ROC Chart:



Fit Statistics
Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Tree	Optimal tree	0.12580	0.11341	0.13743	0.10569
	Neural	Neural Network	0.14925	0.11250	0.14888	0.11887
	Tree3	Interective tree	0.15245	0.09521	0.12414	0.10879
	Reg	Regression	0.15458	0.12893	0.16216	0.12261

On comparing the four models on the basis of misclassification rate, optimal tree model has lowest value for misclassification rate (0.1258) ensuring the results will be more accurate when compared to the other models.

Hence optimal tree is our champion model.

Improve Model Performance

Though through model comparison we have found Optimal tree as our champion model, there are still scope for improvement.

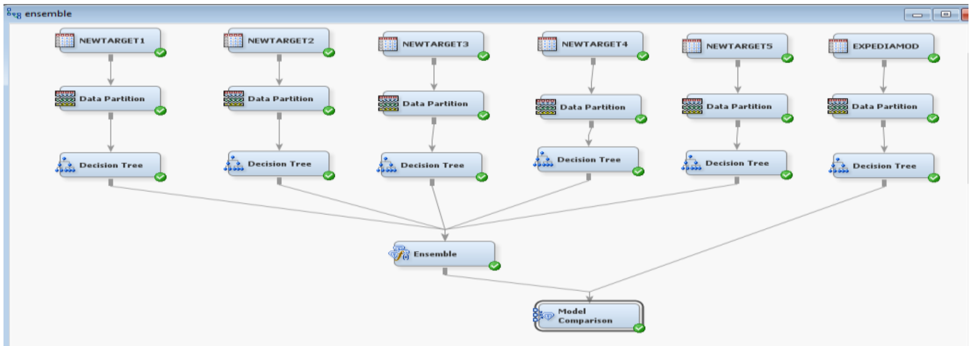
The first approach taken by us to improve the performance of our model is bagging (bootstrap aggregating) since bagging is less susceptible to noisy data and model overfitting.

The other approach which we have used for model performance is multiple imputation using MCMC method. Since Optimal tree was our champion model in initial analysis, we have applied decision tree for all the five data set and used the Ensemble node to create a new model. We have used "Average as the posterior probability function.

The output of the ensemble node and the output of the previous champion optimal tree model via compared using model comparison node.

Summary

Final diagram for preliminary predictive modelling



Insights from Recommender system

1. We realized that the best analytical model to be used by Expedia for booking prediction should be Optimal decision tree.
2. Out of all the 40 input variables used by the company, the variables which are mostly affecting the decision of purchase on the website are as follows:
    a. Size of household.
    b. No. of bookings the user made at the site in the past.
    c. Time spent in the site
    d. Number of sessions in this site so far
    e. Total hits of all sites in the current session
    f. Percentage of total hits are to this site
    g. Percentage of single-site sessions
    h. Age of the user
    i. Total number of unique shopping sites visited

So, we will be using these variables to focus on the customer analytics for the next section

Customer Analytics

The global travel industry comprises a wide variety of businesses, from hotels and inns to casino resorts, trains, buses, airplanes, cruise ships, tour operators and travel bookers, both online and physical. Partnerships between businesses are mutually beneficial and ensure a win-win situation for those involved.

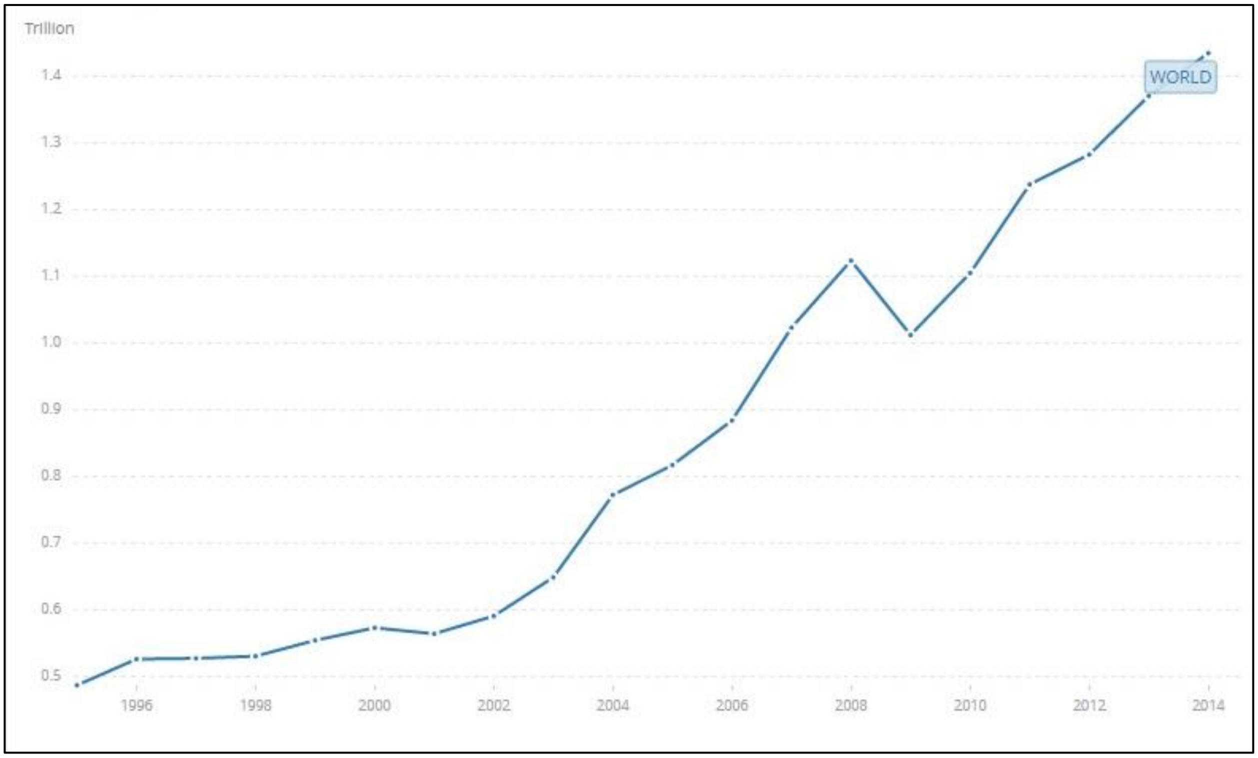
As per the United Nations and the World Travel & Tourism Council (WTTC), more than 1.3 billion tourists traveled the world during 2015. The WTTC discovered that the travel and tourism industry globally supported 108 million jobs on a direct basis in the year, and generated $2.2 trillion in direct global contribution to GDP (gross domestic product). The share is equivalent to the global mining industry, and significantly greater than chemicals and manufacturing (8.6%), agriculture (8.5%), education (8.4%), automotive manufacturing (7.0%) and banking (5.9%).

The sector's growth is projected to soar 4% year-on-year till 2026; which is appreciably greater than the forecasted global GDP growth rate. The WTTC notes that business travel accounts for only 25% of travel-related expenses globally, while leisure travel contributes the lion's share at 75%. Domestic travel accounts for more than 70% of the total expenditure.

To get a bite of this success story, companies look to service delivery providers to enhance their business processes through the use of customer analytics, performance optimization and by implementing progressive customer service solutions. By embedding analytics in every step of a customer's lifecycle, businesses have found that they are better at predicting a customer's likely behavior and thus can improve the balance sheet and drive shareholder value. So, as the next part of the project , we will be discussing customer analytics in detail

The Customers

As per World Bank data, the expenditure on International Tourism was US$ 1.434 trillion!



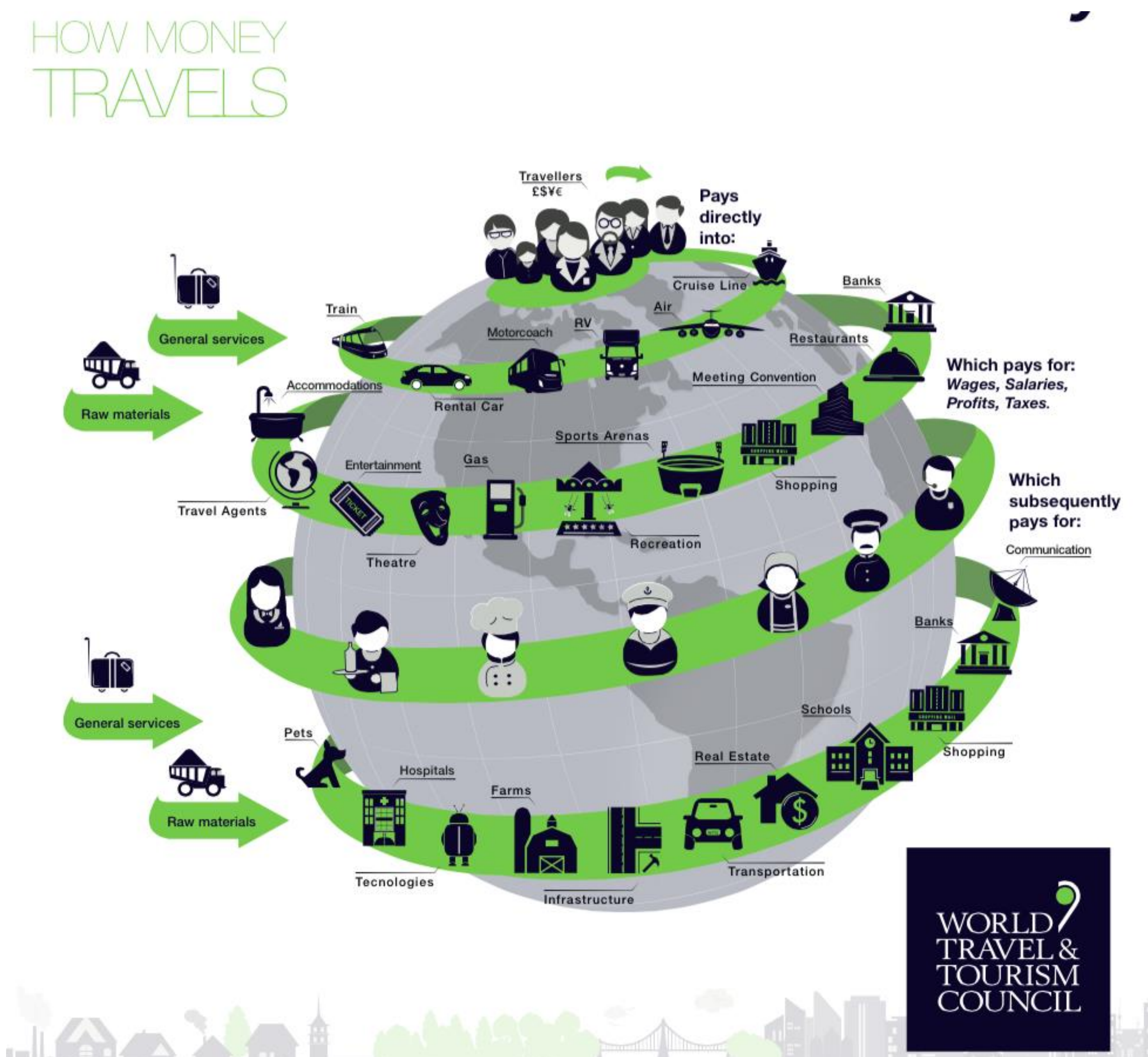
The Travel industry caters to two primary segments, business and leisure travel; though the motivation to travel could be varied (such as pleasure, relaxation, cultural studies, pilgrimages, trade, and so on).

The term ‘Business travel’ encompasses the individuals that are still working and being paid, but are away their regular workplace and home. Primary business tourism activities include meetings, and attending conferences and exhibitions. The Global Business Travel Organization forecast that U.S.-generated business travel spending touched almost US$300 billion in 2014.

The primary motivation of a leisure traveler is relaxation and vacationing. Comforting and welcoming environs are the key to sell to this segment. Various options are available to such customers, right from high-end to budget vacations; catering to all sections.

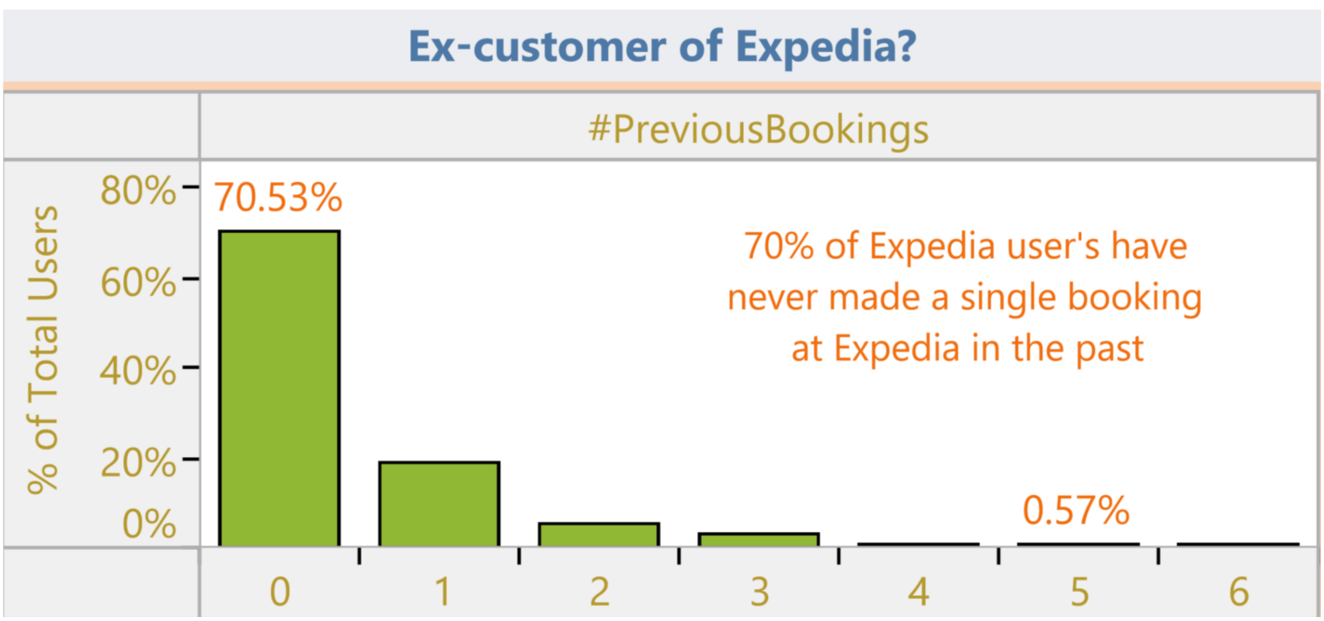
Over the past decade, online bookings have managed to snatch away nearly two thirds of travel-related bookings from traditional travel agents. The demography that still prefers the latter comprise the older, higher-spending travelers who prefer a personal service. The market is ripe for online travel companies to grow and prosper, by providing exemplary, hassle-free, and customized service; at the click of a button.

The following infographic details the journey of money:



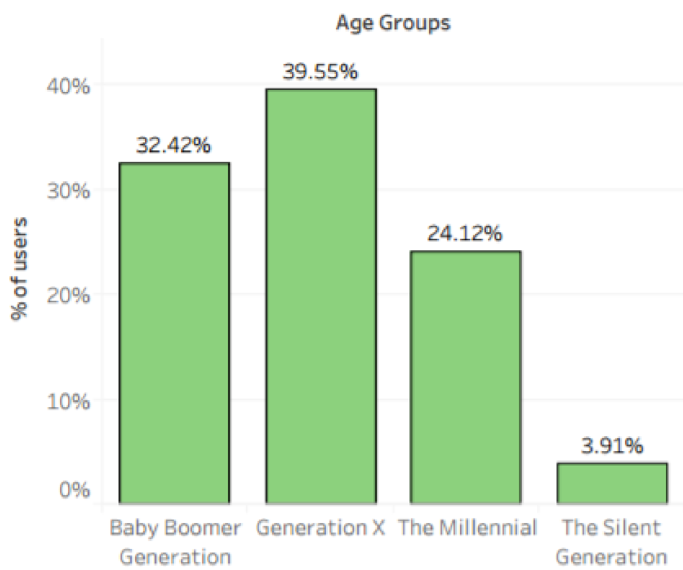
Factors influencing booking

Exploring the data lead us to very resourceful insights. From our data, we see that a staggering 70.53 % of Expedia's users has never made a single booking.



Let us dig deeper to identify what could be a possible cause.

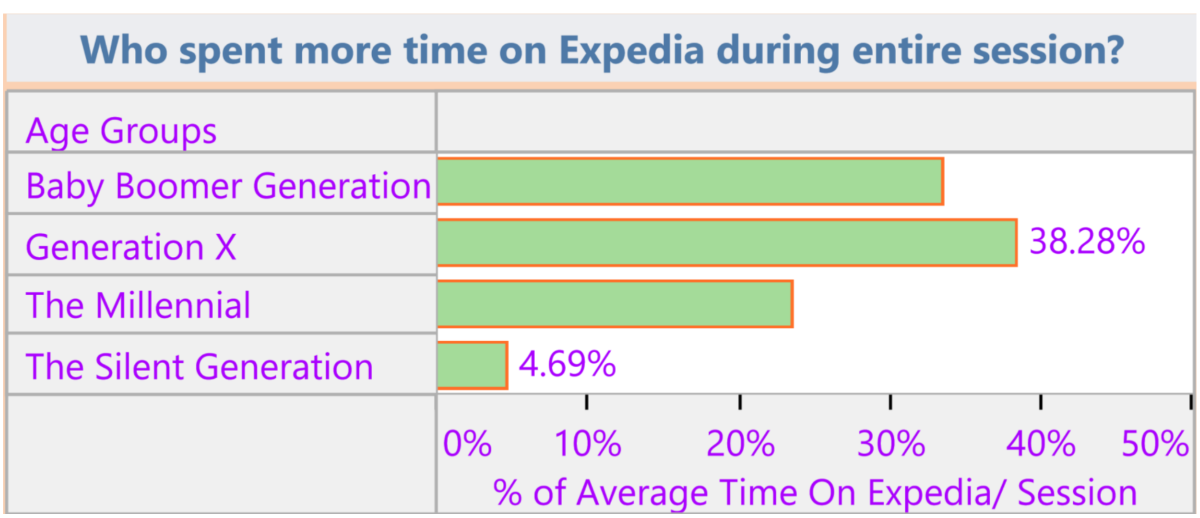
Age Group



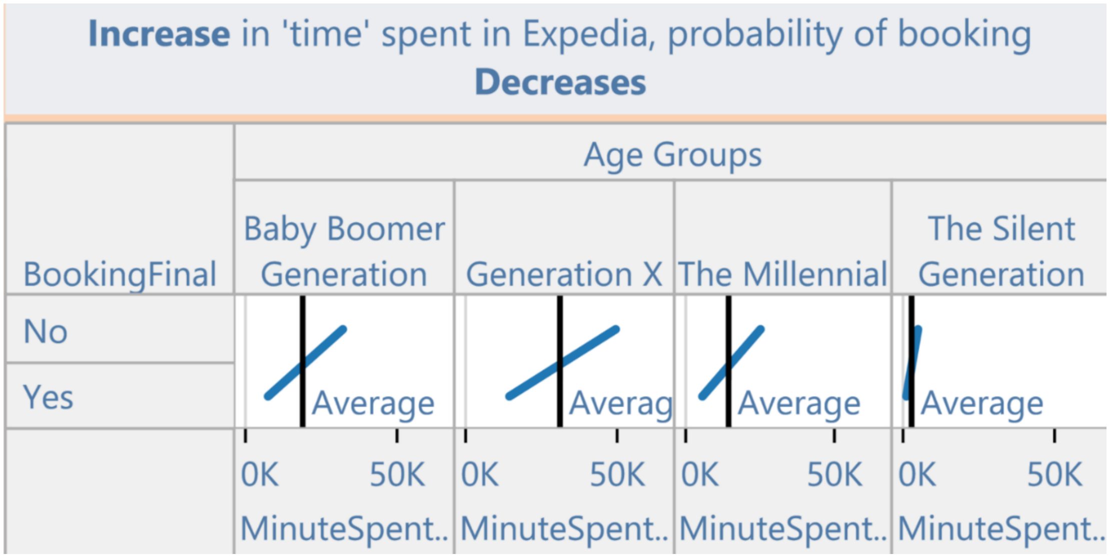
Our customers mainly belong to 3 demographics. Generation X accounting to 39.5% of users while Baby Boomer Generation and The Millenial account for 32.42% and 24.12 %.

It can be said that we are pretty much doing very well with the 3 demographics. We have done a great job in attracting all the 3 demographics. Why does this initial interest not convert to sales?

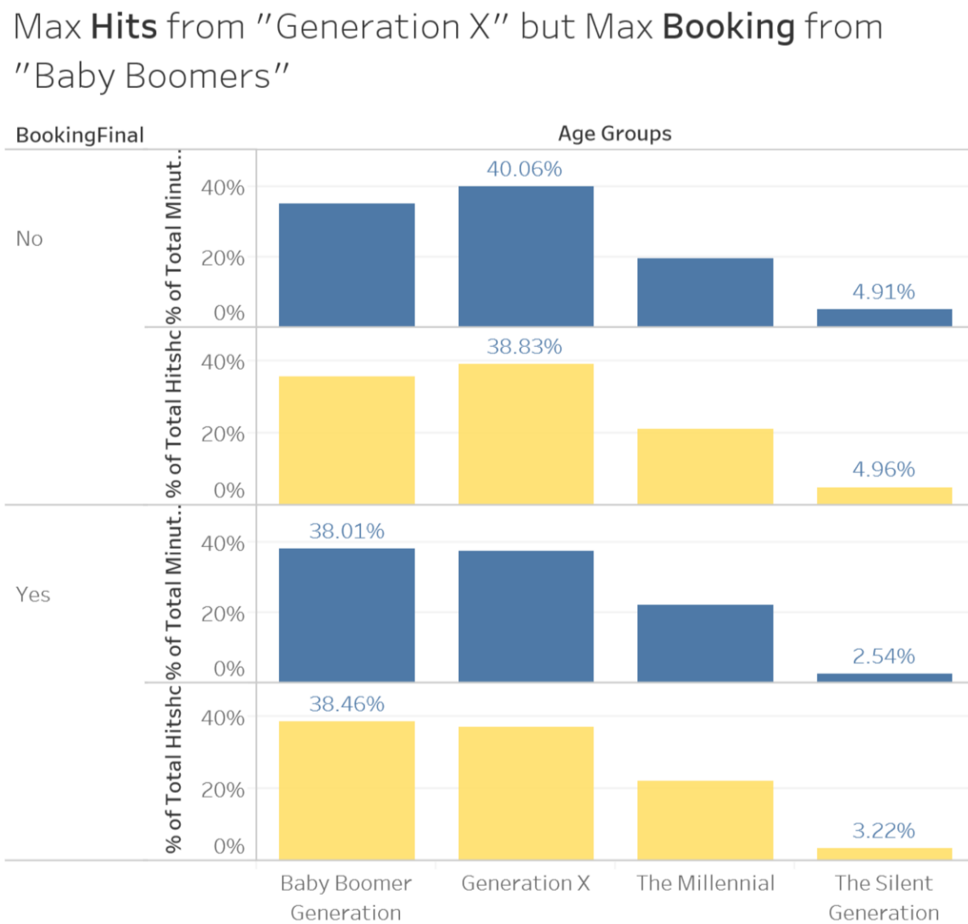
Let us do exploratory analysis to find more



They also happen to be spending maximum time on the website. This indicates that a lot of time is spent on browsing the options and choosing the right place.

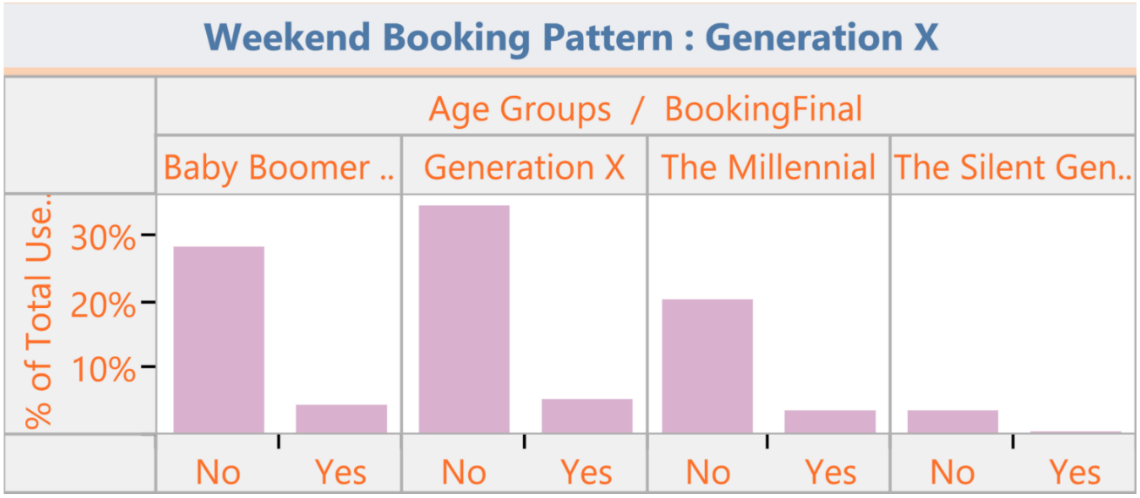


This data answers us and strongly supports our decision. To go ahead with 8 category search . This might reduce the overall time wasted in browsing every option.

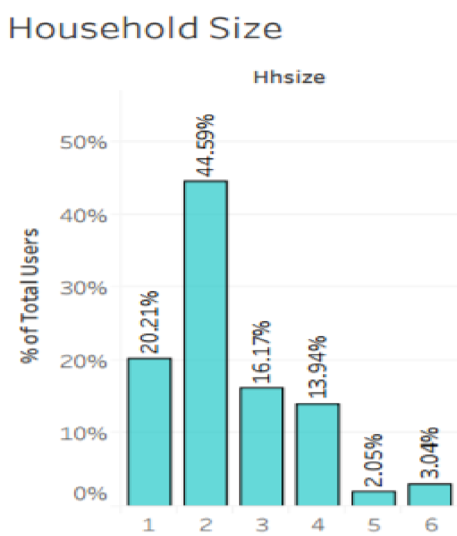


% of Total Minutshc and % of Total Hitshc for each Age Groups broken down by BookingFinal. The view is filtered on Age Groups, which keeps Baby Boomer Generation, Generation X, The Millennial and The Silent Generation.

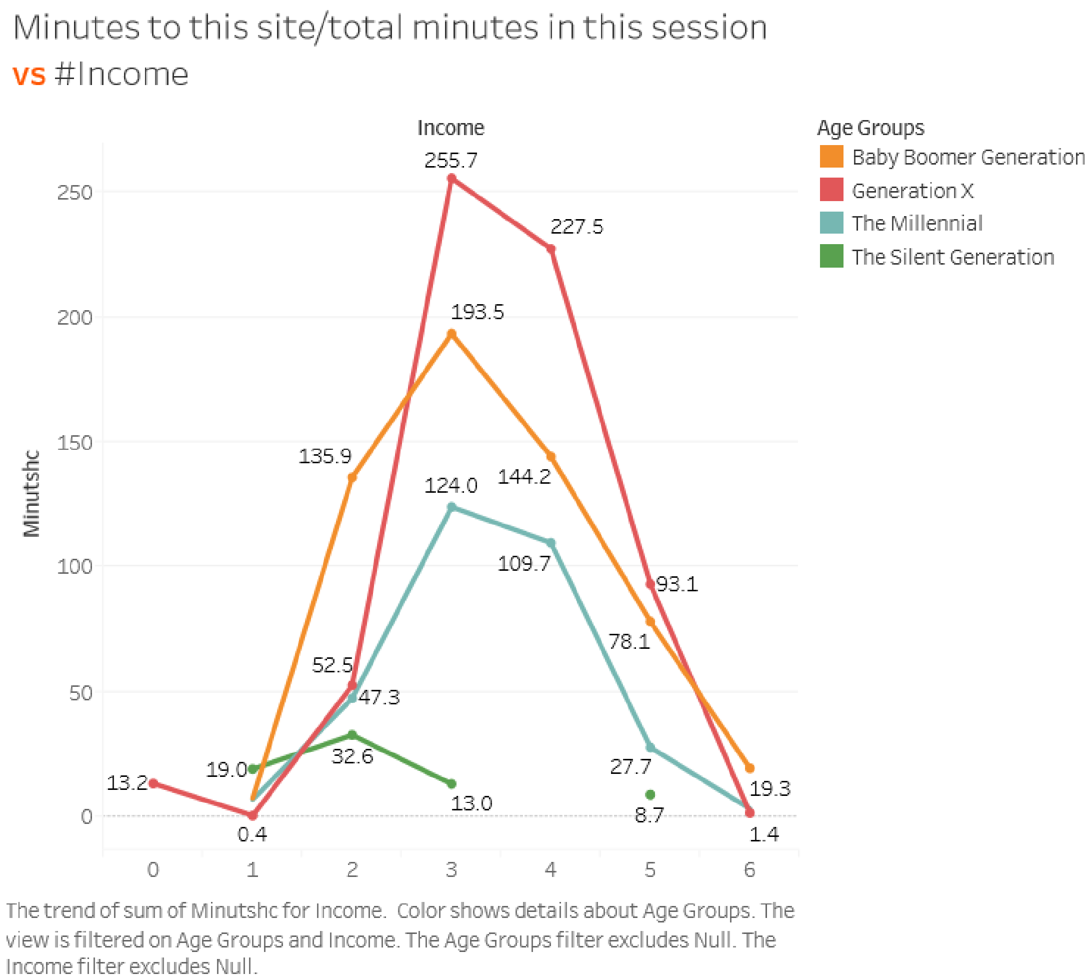
This data also echoes with the inferences made thus far. The time that a customer spends on the website decreases the chance of a booking. A model which can filter better choics for the customer can benefit expedia.The time saved in browsing through all the choices can lead an overall good customer experience and help boost the sales.



We can also infer that most customers don’t like to make a booking on weekends . This is symmetric with every demographic.



Our data also suggests that most bookings are done in the household of 2 people or couples.The next most bookings are from singles followed by families of 3 or 4



The above chart shows the distribution of the total time spent on the site against the income. From the chart we can observer that the generation x spend most time on the site in comparison to millennials and baby boomers. This generation comes in the mid income group.

Customer segmentation

Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unmet customer needs. Companies that identify underserved segments can then outperform the competition by developing uniquely appealing products and services. This prioritization can help companies develop marketing campaigns and pricing strategies to extract maximum value from both high- and low profit customers. A company can use Customer Segmentation as the principal basis for allocating resources to product development, marketing, service and delivery programs.

By using clustering we have segmented our customers bases into following personas:

1. **User persona 1 (richly retired)**

Couples

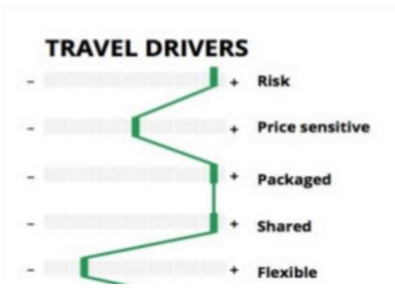
Baby boomers

Mid range income- high income

This persona typically consist of retired Baby boomer couples who fly to meet their grandkids.They are willing to spend.

Comfortable travel is a priority.

2. **User persona 2(traveller millenials)**



Students / millennial

Lower- middle income

This persona consist of Students / millennial who want to go new places. The experience of travelling to new place. Typically, they are constrained by budget. This type of a user is more flexible and adventure loving

3. **<u>User persona 3 (safe families)</u>**

Generation X

Family man

Mid range income – high range income

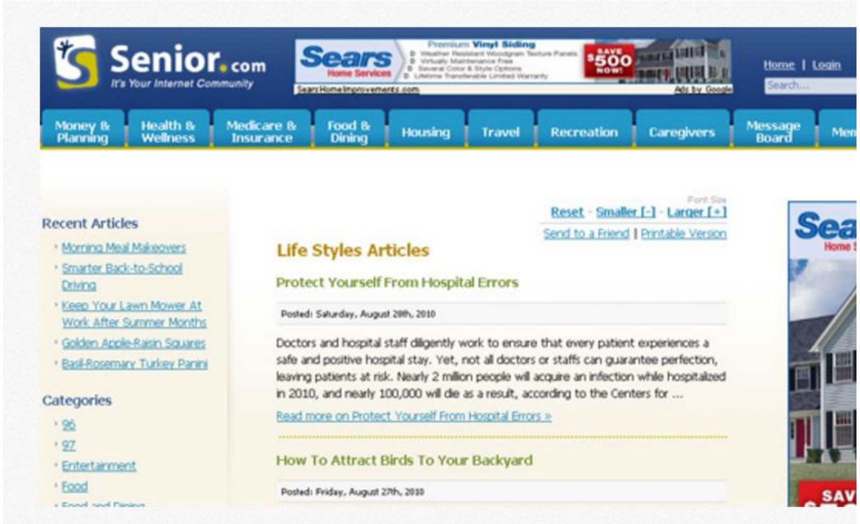
Is a working professional who wants a vacation with spouse and kids. Typically this user works 9am to 5pm, 5 days days a week. The vacation time he wants a safe and family friendly place. This user belongs to the Generation X demographic and typically belongs to a household of 3 or more people.

Marketing Strategy

We have created separate strategies for different segments.

<u>User persona 1 (Richly retired)</u>

This segmentaion consist mainly of couples who are rich and old. They are willing to spend. Message of reaching to them may empahise with festivals or reuniting with the Family. They can typically be reached by TV, Radio, or other media sources popular with baby boomers rather than internet content. Expedia can also place advertisements on some websites meant for the senior age group people. One example of such is senior.com.



Once when you have driven this persona to the website , we need to make them spend more time on our website. Some of the ways in which Expedia can make their website more baby boomer friendly is as follows:

- Make sure your font size is large enough for easy reading. Laptops and most monitors are relatively small, making reading difficult for many seniors. I suggest a minimum of 12 point.
- Keep the navigation as simple as possible.
- Age appropriate photos and graphics. If you are trying to get an older audience to buy, then make them feel comfortable and welcome on your site.
- Understand the market. Marketing to seniors is a sub-specialty. Most internet marketers tend to be younger and don't quite understand the senior market. It makes sense to add at least one person to your marketing team that has experience working with senior consumers.
- Create landing pages specifically targeted to the senior audience. If your online advertising is geared to attract a senior audience, then it makes sense to create a landing page or alternate home page specifically geared to that audience

User persona 2(traveller millennials)

Are Students / millennial who want to go new places. The experience of travelling to new place. Typically, they are constrained by budget. This type of a user is more flexible and adventure loving. This can be sold as a message in the advertising . This segment is highly active on social media and internet. They can be best reached by great content and through various online platforms:

- Start now. There's already a small contingency of Millennials coming into a modicum of wealth, and studies predict that more will join their ranks over the next 10 years. A universal truth about brands is that consumers feel loyal to them. Start building that loyalty online now so that Generation Y will already be attached to your brand once they can afford to buy it.
- Engage online. It's no secret that the Millennial generation is active online. They're not your core audience yet, so don't drop your core marketing tactics, but this is where the next generation of buyers is most active right now. You should have a strong presence online (especially in social media) and use this time to start listening. Find out what they like and dislike about your brand and start gathering insights on how you can strengthen your brand in their eyes.
- Be genuine. Generation Y believes they can see through phony marketing messages and have developed a love/hate relationship with certain over-the-top branding. Rather than using loud and overly promotional online messaging, identify what they like about a brand and develop a story around it that they will be excited to follow.
- Leverage peer recommendations. With social media becoming an increasingly important part of Millennials' daily routines, it's important to keep peer recommendations in mind. This generation has 24/7 access to their social circles and it's highly unlikely that they'll make a major purchase decision without consulting their friends and, or family first.
- Focus on experience. Hard selling is considerably less effective than creating a memorable brand experience. Static messages don't register, so figure out how to get Generation Y to interact with your brand. It could be online in the form of a game or a microsite, or maybe it's a live event or sponsorship.
- Show value. With the proliferation of online product reviews and peer recommendations, it's becoming more important than ever to highlight the value and quality of your product. Millennials do extensive research before buying a luxury product, so make sure your positioning is very clear on the reasons your product is the best.
- Offer exceptional customer service. Generation Y is hyper-connected and they expect their brands to be as well. Make sure you're available on multiple communication channels to answer any

questions they might have, without being pushy or trying to upsell. Upsell is seemingly less necessary with Millennials than with past generations because they are exceptionally loyal. According to a report by Edelman Digital, 70 percent will keep coming back to a product or service they like.

- Do good. One of the best ways to start resonating with Generation Y now is to engage in social good. They grew up with volunteering as a major part of their high school and college requirements and consequently, being socially conscious has become a part of their core belief system. Show that your brand cares about humanity and the environment and you'll start to build a positive image in the eyes of millennials

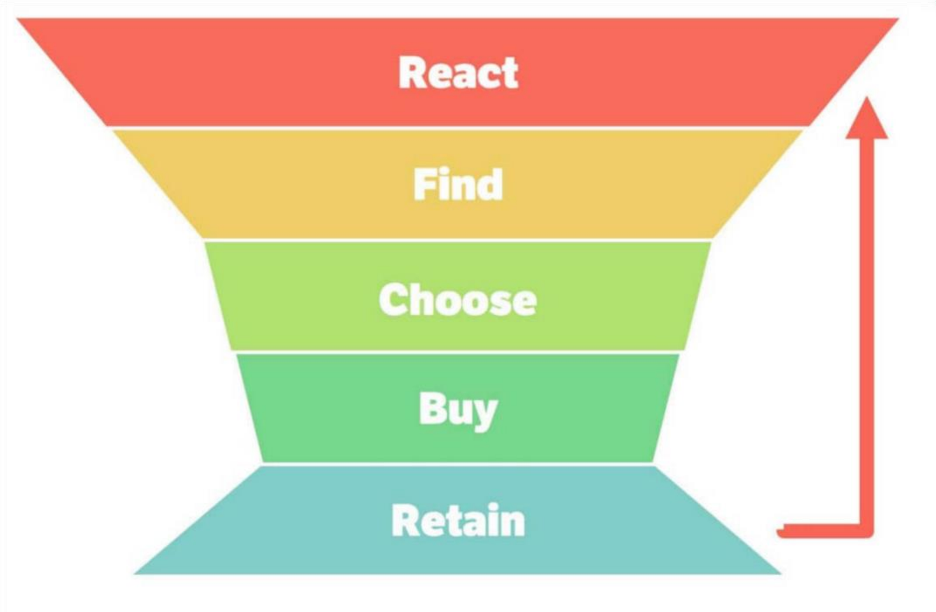
User persona 3(safe families)

Is a working professional who wants a vacation with spouse and kids. Typically this user works 9am to 5pm 5 days a week. The vacation time he wants a safe and family friendly place. This user belongs to generation X demographic and typically belongs to a household of 3 or more people. The message of time spent relaxing and with family resonates best with this user. This type of user can be typically reached on business platforms, online or through business newspapers.

- Target Gen Xers to build brand loyalty and grow market share. Though Generation X is far smaller than the Millennial cohort, Gen Xers report spending far more money with lower levels of brand attachment than Millennials spend. Gen Xers value reliable services when deciding to return to a previous booked hotel , and like Millennials, they place high importance on Internet connectivity.
- Be cautious about assuming that social media and hotel rankings will stimulate revenue. When it comes to selecting a hotel for the first time, social media and industry rankings are at the bottom of the list for all generations. Recommendations from friends have a strong influence on Baby Boomers and Generation Xers, while online guest reviews have the most influence with Millennials.
- Generation X seeks out child-friendly destinations. Since they lack time and crave relaxation, it's not surprising that Sun Belt locales are popular with Gen Xers. So, advertisements should be made keeping this in mind as well.

Conclusion

The following can be used to summarise our strategy with our customers



A customer should ideally buy the following ideas:

- **React**

  The customer should ideally like the website . As per the data, we are doing a great job with respect to the webpage visit. Useful content should be written and used across platforms to bring in users. The content should be sensitive to the needs of the user personas and written in a manner where it can be found easily. keywords based content writing can be done to achieve success. Also since Baby boomers and generation x are more into TV and Radio media. Creative advertisements can also reach them

- **Find**

  The customer should find what they want and the experience on the website should meet their need. Needs of each persona may vary but a common need could be how intuitive the website is. Another need could be whether the website is responsive or not based on what device they use to make a booking.

- **Choose**

  The customer should find choices he is interested in or maybe interested In. By providing relevant choices we can aim to provide the customer with a touchpoint . This would mean that we could get a loyal customer, enthusiastic word of mouth and they same touchpoint may work on another customer of the same user persona.

- **Buy**

  A booking or buying our service is a real indicator that the customer is won. The cart page or the billing page shouldn't be long or slow as that could be very frustrating.

- **Retain**

  The customer should book back and for this incentives and discouts might work. Keeping in touch with a past customer through mails and other active web presence is hence important.

Recommendation

- Generation X seems to be frugal (economic with regards to money spending) : Send discount coupons / offers to just that segment
- Probability of booking is sharply decreasing when a user browse 2 competitor (e.g. Airbnb) : Find out what competitors are offering better than Expedia
- Millennials seems to be 'hoppers' (not loyal) but potential revenue generators; Expedia needs to do targeted marketing to retain millennials or may be accommodate / address this segments special request / requirements (which competitors are doing !)
- 70% of the Expedia user's are new & the conversion rate of new users' is low; Expedia should welcome new customers with some special booking offers (e.g. package booking discount, etc.)

- Segment 3: Safe Families is our primary target, who will financially able to afford a luxurious vacation with their families
- Segment 1: Richly Retired is also our focus; we can provide customized services with a personal touch to such individuals (Customer-Care Executives to connect with them)
- Segment 2: Millennials has potential for the future, collecting their data and trends, and using that later to connect is ideal
- Monitor competitors' price fluctuation closely as changes in any changes in pricing by them affects how the customer perceives our brand; and also has an impact on market shares.
- Gain detailed demographic data of the customers to make personalised recommendations
- Analyse customer data and focus on ways to retain them.