1. Data Exploration and Description

In order to successfully plan marketing actions, it is important to have a deep understanding of customer characteristics and the drivers behind customer decisions. Failing to understand customer dynamics leads to poor business performance. In this assignment the objective is to understand which customer segments are most likely to make a booking on Airbnb and effectively target those segments with appropriate promotional actions.

a) Customer Characteristics

After analyzing the basic metrics from the data provided by Alaska Airlines the following insights are extracted:

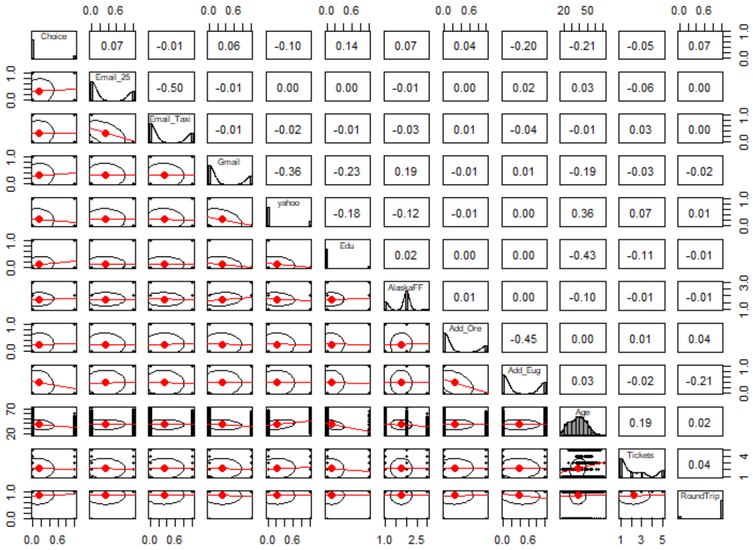
- Age span of potential customers is between 18-71 years old. Average age is 38 years old. Distribution is somewhat right-skewed (customers tend to be young)
- Most customers live in Eugene (37.97%) or out of state (36.93%).
- Most customers use Gmail (30.93%) followed by yahoo (22.37%).
- Most customers travel alone (52.03%)
- Most customers book ticket with return (86.90%)
- Most customers are frequent flyer members (66.40%)

Table 1: Summary of Customer Characteristics

Characteristic	Variable	Min	Max	Mean	Distribution
Age	Age	18.00	71.00	38.04	0-20: 95 (3.16%) 20-40: 1650 (55.00%) 40-60: 1211 (40.36%) 60-80: 400 (14.67%)
Address	Add_Eug Add_Ore	0	1	0.25 0.38	Oregon: 25.1% Eugene: 37.97% Out of state: 36.93%
Email Type	Gmail yahoo Edu	0	1	0.31 0.22 0.11	Gmail: 30.93% Yahoo: 22.37% Edu: 10.53% Other: 36.17%
Number of Tickets	Tickets	1	5	1.99	1:1561 (52.03%) 2: 405 (13.50%) 3: 572 (19.06%) 4: 428 (14.27%) 5: 34 (1.13%)
Ticket Type	Return	0	1	0.87	Return: 86.90% No Return: 13.11%
Frequent Flyer Status	AlaskaFF	1	3	1.74	1: 892 (29.73%) 2: 1992 (66.40 %) 3: 116 (3.87%)

b) Relationship between target variable and customer characteristics

Figure 1: Correlation Matrix & Plot



Edu & Gmail email domains seem to have *positive* relationship with target variable (0.14 & 0.06 correlation respectively). So clients with Gmail or Edu are more likely make a booking. Yahoo email domain has *negative* relationship with target variable (-0.10). This could be interpreted as Yahoo email owners being less likely to make a booking. Regarding customer address, correlation is *negative* (-0.20) for customers living in Eugen and slightly *positive* (0.03 for those with address in Oregon). Regarding age & choice: *Negative* relationship (-0.20) between age and booking.

2. Impact of Targeting on Choice

In order to better understand whether targeting of customers has impact on choice and pin down the most important drivers for choice, it is necessary to model the relationship between target variable and independent variables.

From Figure 1 it is concluded that no single variable seems able to strongly explain Choice. However, we notice significant correlation (>0.01 with all other variables in the dataset). So no variable is to be excluded from next steps.

In order to better plan and estimate the relationships between the variables it is necessary to define the exact mathematical model including all interactions among the variables.

Model definition

The model to be used is logistic regression response model (logit). We are using a logit model as it is the most suitable for binary (Yes/No or 1/0) classification, which is the target variable in this case (Choice: booking or not). The logit model has the benefit of directly modeling the likelihood of a customer making a booking, which in turn will be used to estimate acquisition costs and optimize the promotional actions of the firm.

- Model input: Promotion (*controlled*) variables: "Email_25", "Email_Taxi" Customer (*uncontrolled*) variables : "Gmail", "yahoo", "Edu", "AlaskaFF", "Add_Ore", "Add_Eug", "Age", "Tickets", "RoundTrip"
- Model output: Choice
- Model type: Logit, binomial regression
- Model objectives: Optimize probability P(Choice=1)

Model specification and functional form

Formal definition of the logistic regression model:

$$P(Choice=1)=\frac{\exp(u)}{1+\exp(u)}$$

with u defined as:

$$u=\beta_0+\beta_1*Email25+\beta_2*EmailTaxi+\beta_3*Gmail+\beta_4*yahoo+\beta_5*Edu$$
$$+\beta_6*AlaskaFf+\beta_7*AddOre+\beta_8*AddEug+\beta_9*Age+\beta_{10}*Tickets+\beta_{11}*RoundTrip$$

3. Model Estimation & Key Results

Estimating the logistic regression model's coefficients will help understand which variables affect the most the target variable Choice. This will in turn help to decide which promotional activities are the most appropriate and which customer segments are the best to target. The result of estimating the previously defined logistic model in R is shown in Table 2 (using 3 decimal points):

Table 2: Estimation of logistic regression model coefficients

Coefficients:	Estimate	Std.	Error	z value	Pr(>|z|)
(Intercept)	-0.048	0.308	-0.156	0.876	
df$Email_25	0.589	0.135	4.347	1.38E-05	***
df$Email_Taxi	0.208	0.140	1.487	0.137	
df$Gmail	0.290	0.137	2.118	0.034	*
df$yahoo	-0.171	0.178	-0.958	0.337	
df$Edu	0.521	0.184	2.82	0.004	**
df$AlaskaFF2	0.322	0.131	2.463	0.013	*
df$AlaskaFF3	0.456	0.287	1.586	0.112	
df$Add_Ore	-0.357	0.125	-2.841	0.004	**
df$Add_Eug	-1.629	0.150	-10.818	<0.2E-8	***
df$Age	-0.049	0.007	-7.013	2.33E-12	***
df$Tickets	-0.031	0.052	-0.602	0.547	

Observing the values for the coefficients the *statistically significant predictors* can be determined: Email_25, Add_Eug, Age, Edu, Add_Ore, Gmail and AlaskaFF. The exponents of the coefficients can be interpreted as odds ratios, which will be used to calculate booking probabilities and acquisition costs.

Table 3: Odds ratios

(Intercept)	Email 25	Email Taxi	Gmail	yahoo	Edu	Alaska FF2	Alaska FF3	Add Ore	Add Eug	Age	Tickets
0.95	1.80	1.23	1.34	0.84	1.68	1.38	1.58	0.70	0.20	0.95	0.97

According to the above we can identify the most *significant drivers of choice*:

- Email25: customer that receives email with $25 offer on Airbnb booking increases their chance of booking by 80.4% (1.80 odds ratio), also customers that receive EmailTaxi offer (odds ratio 1.23)
- Email domain Gmail or Edu & AlaskaFF = 2 or 3

b) Importance & nature of email targeting

As seen in Table 3, *email targeting matters*, and quite significantly as such. Customers that receive the offer for $25 discount on the next Airbnb booking are more *likely* to actually book by 80.89%. Those that receive the taxi offer get a lower increase of 23.29% to book. *Causality* cannot be claimed, as association is not causation. To prove causality, further *experiments* would be needed (for example After-Before testing).

c) Best customer email domain and interpretation

According to the analysis, customers with *email domain "Edu" are more likely to* book on Airbnb. (odds ratio of 1.68), followed by Gmail (odds ratio = 1.34). One possible reason for this is that customer email domain can be related to age, which is related with choice, as seen in Figure 1 (corr. = 0.21). In Figure 1 it is also observed that age and Edu & Gmail domains are related (-0.43 and 0.36 correlation respectively).

4. Action Plan to increase bookings

a) Choice of customer segment

From the analysis so far it can be concluded that a sensible strategy to increase the number of booking to Airbnb is to target travelers to Eugene that use *Gmail or Edu* email domain (odds ratios 1.34 & 1.68 resp.) and are frequent flyers or VIP frequent flyers (AlaskaFF=2 or AlaskaFF=3, odds ratios 1.38 & 1.58 resp.) as those customer segments have *highest increase in probability to book.* Those customer segments should be targeted with email offer for discount $25 or taxi offer $25. Resulting customer segments are shown in Table 4:

Table 4: Customer Segments

Promotion	AlaskaFF	Email Domain
Email25	2	Gmail
Email25	2	Edu
Email25	3	Gmail
Email25	3	Edu
Email_Taxi	2	Gmail
Email_Taxi	2	Edu
Email_Taxi	3	Gmail

Email_Taxi	3	Edu

Based on this customer segmentation the campaign costs and subsequent acquisition costs per customer will be calculated.

b) Campaign costs and acquisition costs per customer & segment

According to the existing customer segmentation, the acquisition costs are calculated, shown in detail in Table 5:

Table 5: Acquisition Costs per Customer Segment

Promotion	AlaskaFF Status	Email	Targeted Customers	Acquisition Probability	Acquired Customers	Total Costs	Acq. Cost per Cust.
Email25	2	Gmail	226	0.31	70	$10750	$154
Email25	2	Edu	75	0.36	27	$9675	$358
Email25	3	Gmail	20	0.34	7	$9175	$1311
Email25	3	Edu	4	0.39	2	$9050	$4525
Email_Taxi	2	Gmail	217	0.24	52	$10300	$198
Email_Taxi	2	Edu	64	0.28	18	$9450	$525
Email_Taxi	3	Gmail	21	0.26	5	$9125	$1825
Email_Taxi	3	Edu	6	0.31	2	$9050	$4525

It can be concluded that for *smaller customer segments* the acquisition cost per customer can be too *high*. Thus, it would be better to focus targeting efforts on *bigger customer segments,* for example, customers that use Gmail and have frequent flyer status = 2, which have acquisition cost per customer under $200.

In addition, the acquisition cost per customer drops the more customers actually make a booking. Thus, the best course of action is to target customer segments that are *most likely to make a booking* with the Email25 or EmailTaxi promotions. Even though Email25 has a higher probability of leading to a booking than EmailTaxi, EmailTaxi it is still worth pursuing for bigger customer segments.

Acquisition costs were calculated as follows:

- Nr.targeted customers: subset of dataset for each segment
- Fixed Costs: $9,000 (amount paid by Airbnb to Alaska Airlines)
- Variable Costs: $25 * Nr. Acq. Customers
- Total Cost = Fixed Costs + Variable Costs
- $Acquisition\,Costs = \frac{\text{¿}\,Costs + Variable\,Costs}{Number\ of\ Acquired\ Customers}$
- $Number\ of\ Acquired\ Customers = Acquisition\ prob. * Nr.\ targeted\ customers$