# Executive summary on

# **MARKETING**

Alexandre Olivier, Anchal Jaiswal, Asier Sarasua Amundarain, Ignacio Rupérez Larrea, Jina Kim, Nicolas Clemens Linsenmaier, Rahul Singh

Group 3, O1 Section, Master in Business Analytics and Big Data Program, IE School of Human Sciences and Technology

Machine Learning I

#### I. Objective

- 1. Identify clusters in the online travel market customer base, and segment the market based on demographic and ticket purchase information
- 2. Design and deploy segment oriented marketing campaign, including increasing cross-selling of online rental cars and hotels

#### II. Descriptive statistics of the dataset

#### Dataset overview (Annexure Table 1)

- Original dataset: 1999 observations, 12 variables: 1 unique identifier (id), 3 date columns (departuredate, returndate, purchasedate), 5 Boolean columns (children, car, hotel, insurance, gender), 1 continuous column (age), 1 discrete column (seats), and 1 categorical column (customer type)
- Final dataset: 1991 observations, 20 variables: 8 new variables created, 3 variables transformed

#### Key explanatory data analysis findings

#### Dates:

- The standard holiday months have the highest number of passengers departing. 45% of the people travel on summer holidays (July and August), closely followed by 16% on Christmas holidays (December). (Annexure Figure 1)
- The most popular departure-day of the week is Friday, where nearly 21% of the flights depart from the airport. Friday is most popular during the regular months (not in summer or Christmas holidays). During the standard holiday months, however, Friday no longer becomes the predominant day of week in terms of departure date. (Annexure Figure 2, Figure 4)
- The number of purchases increase during June, July, and December. High purchases in the former two months are not surprising as there is always some lag time between ticket purchase and actual departure. The surge in December is interesting as this may indicate that people tend to buy tickets in rush for Christmas holidays than they do so for the summer holidays. (Annexure Figure 5)
- Most people usually take a week or two to make decisions, but there are people who
  make very quick decisions in less than a week. (Annexure Figure 6)
- 3 days long trip is the most common; this may be due to weekend travels or non-leisure travels that don't last long. Longer holidays go on in a weekly interval, where 1 week is the most common, followed by 2 weeks. (Annexure Figure 7)
- While short trips lasting less than 3 days are also common during standard holidays, it is clear that the duration of trips usually last much longer than those during regular months, where the maximum duration is 11 days. (Annexure Figure 8)
- In general, people who go on weekend trips are younger than those who do not, although there are outliers. (Annexure Figure 10)

## Age:

- The early 20s (until 25) are the most common customers (37%). (Annexure Figure 11)
- The outliers of this sample of customers as per 1.5 IQR are those over the age of 72. There were 8 outliers, who were dropped from subsequent analysis as they tend not to be a major customer of the company, but still may distort the segmentation result during clustering analysis. (Annexure Figure 12)
- The average age of customers in general increases with the increase in number of seats bought, though there are some variances. This is likely because middle aged people with a family tend to buy tickets for all family members, while young singles can buy a ticket for one. (Annexure Figure 13)

#### Ticket purchase information:

 Customers travelling in pairs purchased more seats (43%) in comparison to others. Those travelling in big groups (over 4 people) tend to be a really small in number. Only 13% of customers travel alone, assuming that groups of people for the same trip did not buy tickets separately. (Annexure Figure 14)

■ 79% of customers do not travel with children.

#### Up-selling information:

- 80% of the customers either do not record rental information or do not rent cars. The rate drops to 65% for hotel room rentals. This may be because while not all travels require car rentals, most, if not all, require some sort of accommodation. (Annexure Figure 15, Figure 16)
- The probability of renting a car or hotel at the time of airline ticket purchase is not affected whether or not the trip involves children. This indicates that there are other reasons as to why some people rent car and hotel with the flight ticket.

### III. Technical design of the study

#### **Choice of clustering variables**

- Select variables that a customer considers when buying a flight ticket: cars, insurance, hotels are not the primary decision factors while choosing to purchase a flight ticket, so was left for profiling variables
- Avoid highly correlated variables: age, duration, and purchase lag are highly correlated with one another. Car and hotel are correlated.
- Use discriminatory variables: gender was dropped because it was not a discriminatory variable in any of the clusters formed. Standard holiday is an important variable, but because most people went on trips during these months, it was not a discriminatory variable among clusters.

#### Data transformation for modeling

- 1. Standardize continuous variables to z scale
- Calculate Gower distance for all observation values: this dataset has a mix of categorical
  and numeric data, meaning there isn't a single metric that can handle all variables. Gower
  distance method uses the most appropriate distance calculation method for each type of
  variable, then standardizes all points to scale from 0 to 1

## Choice of clustering algorithm

Partitioning Around the Medioid (PAM): the basic principle of PAM is similar to k-means, where clusters are formed around the closest center. The main difference is that (a) the center, or the medioid, is an actual point in the dataset, and (b) unlike k-means, PAM doesn't need the original data matrix to perform clustering, which is essential, because this analysis standardizes the data matrix to a distance matrix using Gower distance. This method also allows the clusters to be more robust to outliers, as the medioids are not arbitrary points in the dimensional space.

## **Clustering model process**

- 1. Select clustering variables: different variations tested based on choice of clustering variables
- 2. Calculate silhouette score for selection of k values, then select k value based on the score. Note that the final k selected may not produce the highest silhouette score, as clusters were manually checked to ensure each cluster made sense.
- 3. Create cluster using PAM algorithm
- 4. Iterate steps 1-3 until appropriate cluster is found

#### Final clustering analysis result

Clustering variable: Z\_age, travel\_companion, weekend\_trip

Number of clusters: 4Silhouette score: 0.72

#### IV. Customer segmentation based on cluster output

	Comfort duo	Young family	Young and free	Big family retreat
Cluster size	871 (43.75%)	714 (35.86%)	252 (12.66%)	154 (7.73%)
Age	Mean: 32.27	Mean: 34.23	Mean: 29.68	Mean: 38.58
	Median: 29	Median: 35	Median: 26	Median: 41
Travel	Couple	Small family	Solo	>4 large group
companion		(3 people)		
Weekend trip?	Υ	Υ	N	N
Priority	1	3	4	2

#### **Segmentation summary**

- 1. Comfort duo: This group consisted of mainly those who travelled in pairs, consisted of second lowest age range among consumers (late 20s and early 30s), and went on standard holidays (more often than the big family and young family clusters regarding average time spent on vacations). This group took equally long time to purchase a ticket as compared to cluster one (but with a high variance). This group had the most number of cars and hotels purchased, and the second highest in terms of percentage, indicating that this group intends to make its travel as hassle free as possible during the initial purchase stage.
- 2. Big family retreat: This identified cluster reserved more than four seats, enjoyed long vacations and consisted of people with higher age in comparison to other groups. 95% of this cluster is accompanied by a child in the trip. This cluster included people who took more than 20 days to purchase the tickets and preferred going on vacations during standard holidays (namely Summer and Christmas break). This group also tends to get prepared beforehand, with almost half of the cluster booking cars and hotels at the time of purchasing the ticket.
- 3. **Young family:** This cluster consisted of three people, with 26% of them accompanying a child. This group by far goes on the most weekend trips out of all the clusters, and also enjoy standard holiday periods, meaning when given the opportunity, they will travel.
- 4. **Young and free:** This cluster is a unique group of young and solo travelers that tries to avoid the crowd. This group is the only group with majority of the people avoiding standard holiday months, and even weekend travels, both of which are guaranteed to have lots of people travelling at the same time. They make almost instantaneous decisions when purchasing tickets, and do not tend to stay too long for travels.

#### V. Marketing strategic lines

### **Priority 1: Comfort duos**

**Reason**: This is the largest cluster among customers, taking almost half of the market. This group is also at a prime age group around late 20s and early 30s, where customers are not only young and active, but also have higher purchasing power than any other group without much burden.

#### Marketing strategy:

- Romantic package offers for young couples: if the traveling duo is a couple, honeymoon travel package can be sold including attractive tour offers and special rates. If the ticket is for overseas marriage, a coupon code for family members can be offered too. Similarly, advertising campaign can be tailored around romantic and festive days such as Christmas or Valentine's Day. Online social media will be used as a marketing channel keeping in mind the most active social websites among young people: Instagram, Facebook for tailor made campaigns.
- Focus on advertising luxury trips and hotels to young professionals where many times couples have dual income and no kids. Here channels can include professional networks

such as LinkedIn but also media outlets focussed on professionals, such as The Economist or Manager Magazine.

#### **Priority 2: Big family retreat**

**Reason**: despite being the smallest segment, this group is large in actual travel group size, as at least 4 people travel together. This group also rents car and hotel, meaning that that the size of the car and the number of rooms rented are likely to yield higher revenues if more customers from this cluster were to cross-purchase hotels and cars with the flight tickets.

#### Marketing strategy:

- Early bird bulk sales: offer steep discounts for flight tickets and hotel (or rental car) package based on number of tickets bought. Since these groups at least travel with 4 people or more, discounts based on more head counts will be attractive
- Large group insurance price promotion: offer insurance price discount based on headcount for groups of more than 4 or 5 people
- Deals with family friendly accommodations: promotions to hotels that can take in accommodate large number of people or places with various family-friendly activities can attract these big groups that need to accompany children in their travels

#### **Priority 3: Young family**

**Reason**: this group's size is relatively big in the market, but it is not as big as nor lucrative as are the Comfort duos, which have high purchasing power. However, as in the case of Big family retreats, there is a possibility of packaged sales for families, as the group has the most number of customers accompanying a child in travels. Furthermore, this group can be offered special weekend deals, as they enjoy weekend travels too.

#### Marketing strategy:

- The airline can deploy a strategy to appeal to children and teens in order to attract parents. For this, associating with key events for children in cities with major transport hubs can provide a good business opportunity to market to family demographics. For example, the company may associate with an open gym or arena night for a local school that gives parents confidence that their children are in a safe and fun environment in order to build a trendy brand image among kids.
- Offer package deals for example for a long weekend to Disney-Land or similar attractions
  which are fun for the entire family and yet affordable. TV advertisement in family channels
  will reach the right target and directly create conversation among the family.

#### **Priority 4: Young and free**

**Reason**: despite being a unique group, this group's relative size is small, and being solo travelers, it is hard to yield huge revenue from individual customers, unlike the 'Big family retreat' group.

## Marketing strategy:

- Price promotion on odd departure dates: offer price promotions on routes departing on weekdays and during less busy months. The airline could also provide best information regarding travel destination via email for solo travelers as 'access to best information' has been ranked as a priority among solo travelers.
- Price promotion on last minute deals: offer price promotions on last minute routes that departure within 2 days. This would allow the airline to sell empty seats to solo travelers who already make instantaneous decisions. Time sales on social media such as twitter or Facebook will attract the right crowd.
- Create a "Travel Picture Challenge" for example on Instagram. The young and free traveller should share her most authentic travel picture from a remote and unique experience. The winner will recieve a travel voucher for the airline, yet the reach achieved through the challenge will create publicity among the target group.

# <u>ANNEXURE</u>

Table 1. List of feature creation and transformation

Feature name	Type of variable	Feature explanation	Creation / Transformation method
dep_month	Categorical (12 values)	Month of departure	Parsed month from departuredate
dep_dow	Categorical (7 values)	Day of week of depature, where Sunday = 1	·
Duration	Continuous	Duration of trip	returndate – departuredate
standard_holiday	Binary categorical	Departure month lies in holiday months, including summer and Christmas	dep_month == 7    dep_month == 8    dep_month == 12
weekend_trip	Binary categorical	Trip is a short weekend trip	<pre>(dep_dow == 5 &amp;&amp; duration &lt;= 3)    (dep_dow == 6 &amp;&amp; duration &lt;= 2)    (dep_dow == 7 &amp;&amp; duration &lt;= 1)</pre>
purchase_lag	Continuous	Time taken for customer to buy and actually go on a trip	departuredate - purchasedate
up_purcased_carhotel	Binary categorical	Either rented a car or hotel at time of ticket purchase	car + hotel > 0
travel_companion	Categorical (4 values)	Number of people travelling together	<pre>if(seats == 1, "solo", if(seats == 2, "couple", if(seats == 3,     "small_family",     "large_group")))</pre>
Z_age	Continuous	Standardized age values	
Z_duration	Continuous	Standardized duration values	
Z_purchase_lag	Continuous	Standardized purchase lag time values	

## **Figures: Exploratory Analysis**

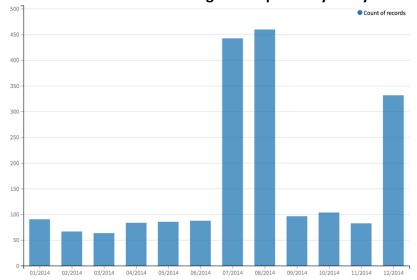


Figure 1. Count of departure date by month

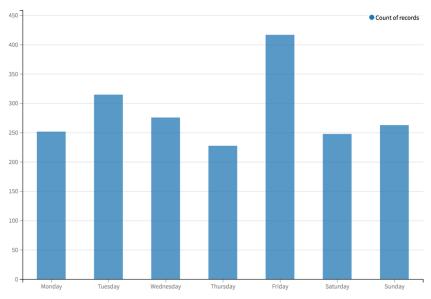


Figure 2. Count of departure date by day of week

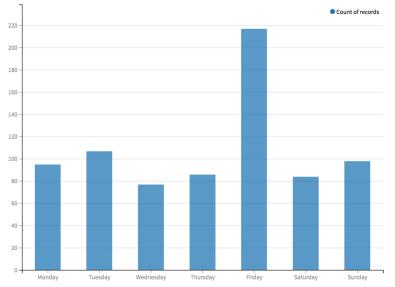


Figure 3. Count of departure date by day of week during regular months (exclude July, August, December)

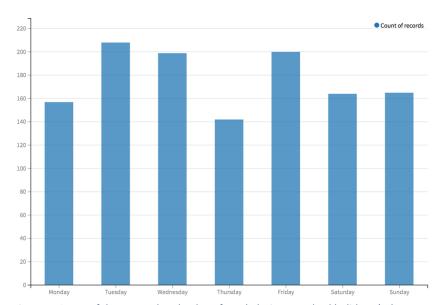


Figure 4. Count of departure date by day of week during standard holidays (July, August, December)

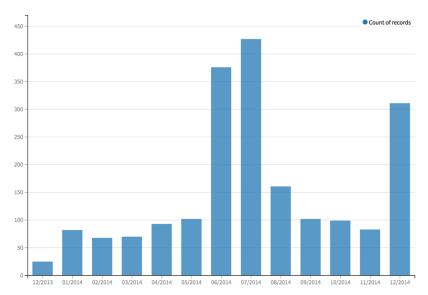


Figure 5. Count of purchase date by month

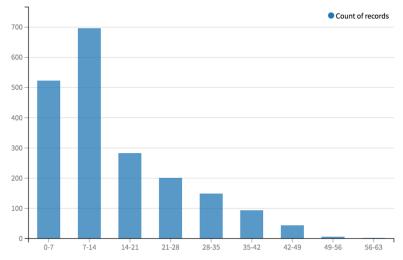


Figure 6. Distribution of purchage lag in days

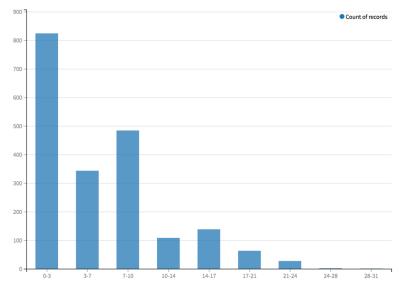


Figure 7. Distribution of duration of trips in days

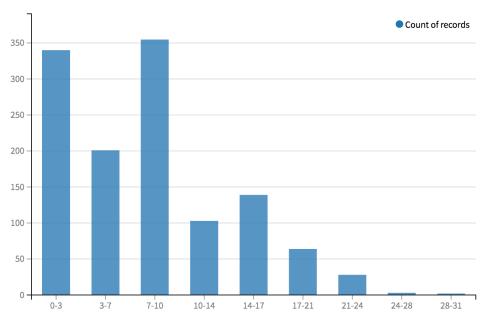


Figure 8. Distribution of duration of trips during standard holidays

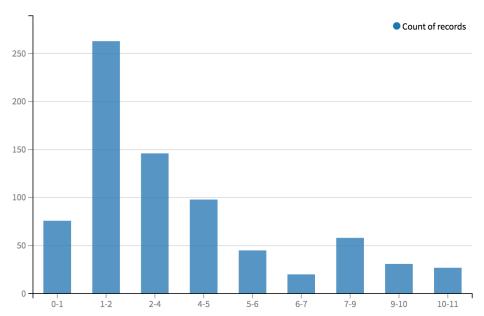


Figure 9. Distribution of duration of trips during regular days

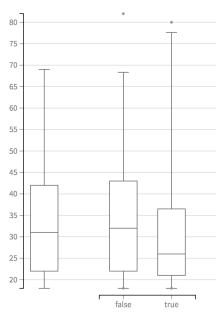
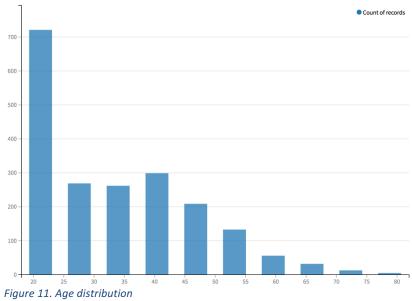


Figure 10. Distribution of age by weekend trip



STATISTIC	CS	TOP VALUES			OUTLIERS	(0 low, 8 high)
Min	18	18	175	8.8 %	73	
Max	82	20	117	5.9 %	74	
Mean	33.310	21	104	5.2 %	76	
Median	31	22	104	5.2 %	78	
StdDev	12.913	23	96	4.8 %	80	
Mode	18	19	80	4.0 %	81	
Distinct	62	45	71	3.6 %		
IQR	20	More values and actions				
Sum	66587					

Figure 12. Age statistics summary

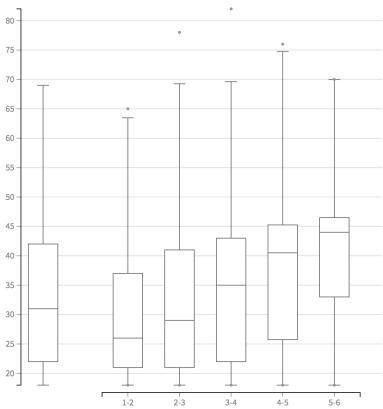


Figure 13. Distribution of age by number of seats purchased

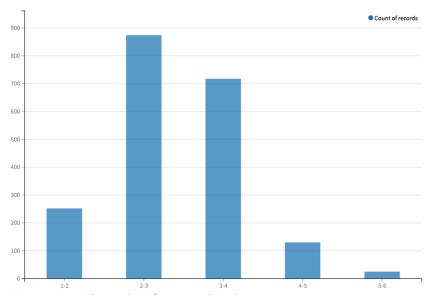


Figure 14. Count by number of seats purchased

Run: In DSS		Count by car and children	7
	0	1	
0	80%	20%	
1	77%	23%	

Figure 15. Percentage of car rentals (column) given the trip involves a child (row)

Run: In DSS	Count by hotel and children		
	0	1	
0	65%	35%	
1	64%	36%	

Figure 16. Percentage of hotel rentals (column) given the trip involves a child (row)

## **Graphs – cluster vs profiling variable analysis**

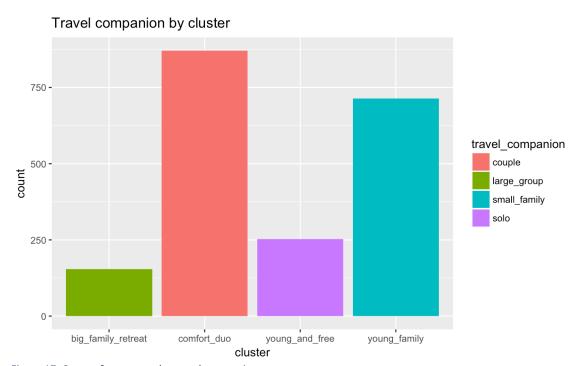


Figure 17. Count of customers by travel companion

## Density of age distribution by cluster

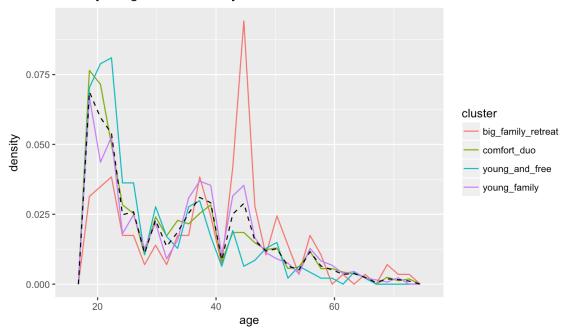


Figure 18. Density of age distribution by cluster

## Distribution of age by cluster

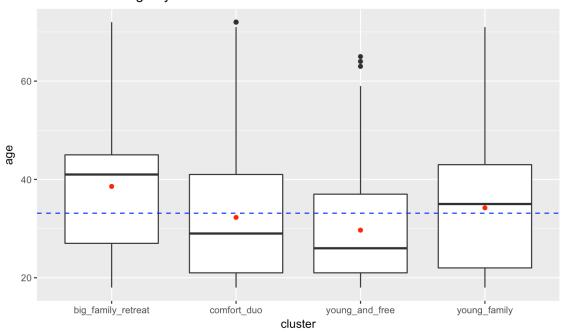


Figure 19. Distribution of age by cluster

## Density of purchase lag distribution by cluster

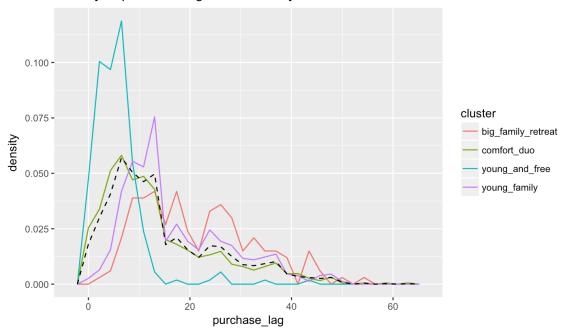


Figure 20. Density of purchase lag distribution by cluster

## Distribution of purchase lag by cluster

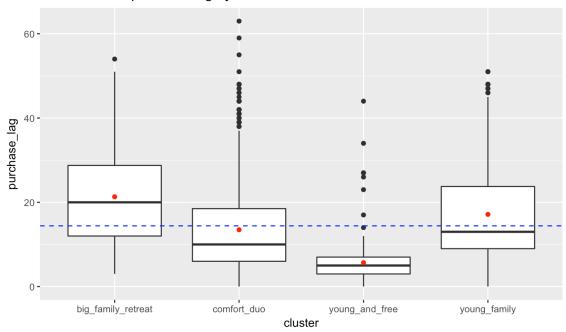


Figure 21. Distrubution of purchase lag by cluster

# 

20

duration

30

Figure 22. Density of trip duration distribution by cluster

10

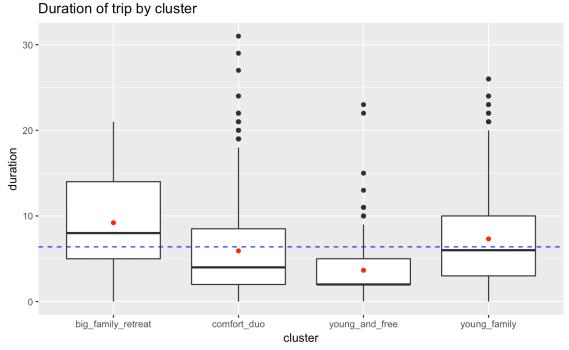


Figure 23. Density of trip duration distribution by cluster

## Proportion of people going to standard holiday by cluster

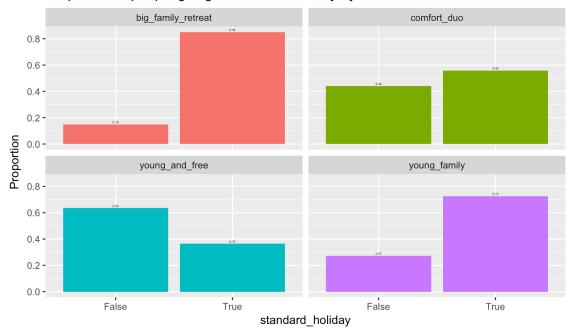


Figure 24. Proportion of people going to standard holiday by cluster

## Number of people going on standard holiday months by cluster

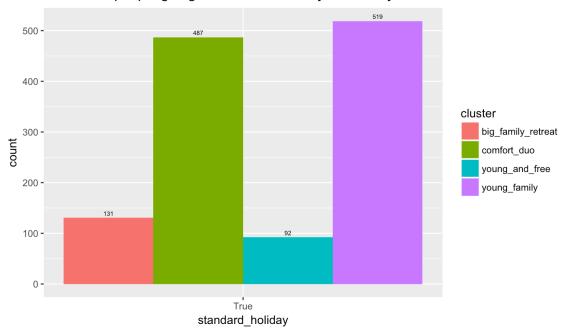


Figure 25. Count of people who go on standard holiday months by cluster

## Proportion of people going on weekend trip by cluster

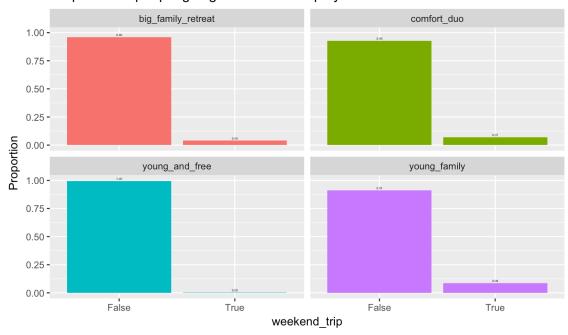


Figure 26. Proportion of people going on weekend trip by cluster

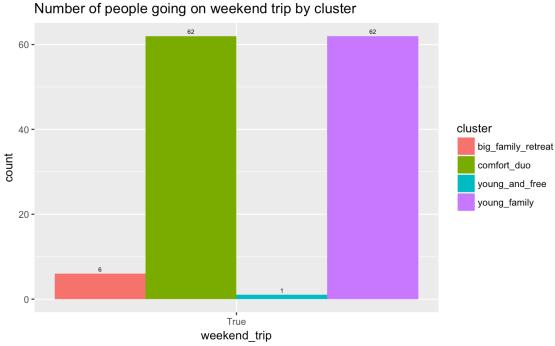


Figure 27. Count of people going on weekend trip by cluster

## Departure month and day of week proportion for 'Big family retreat' cluster

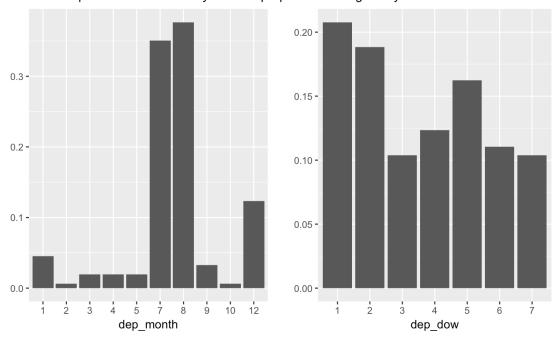


Figure 28. Departure month and day of week proportion for 'Big family retreat' cluster

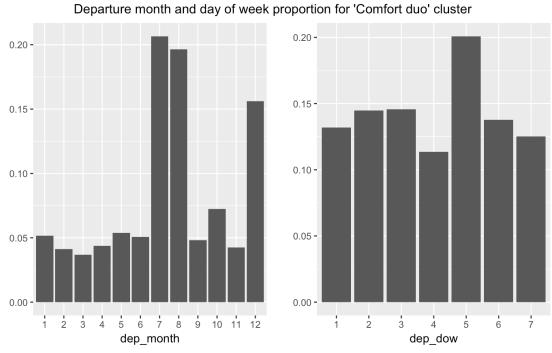


Figure 29. Departure month and day of week proportion for 'Comfort duo' cluster

## Departure month and day of week proportion for 'Young and free' cluster

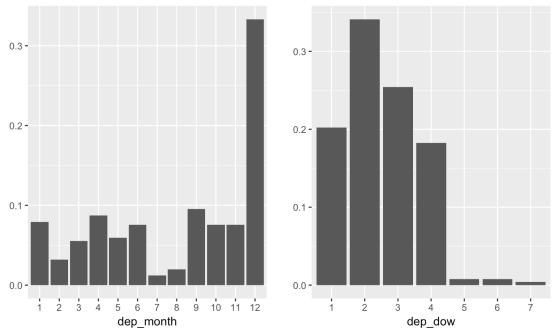


Figure 30. Departure month and day of week proportion for 'Young and free' cluster

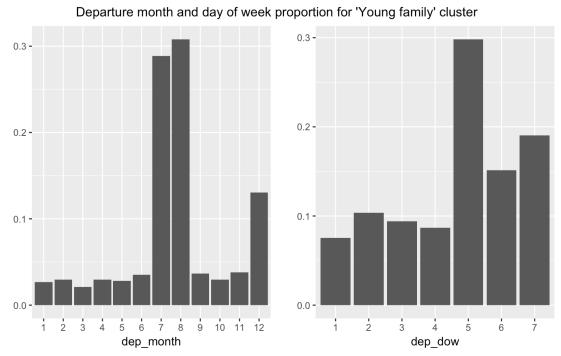


Figure 31. Departure month and day of week proportion for 'Young family' cluster

## Proportion of people accompanying children by cluster

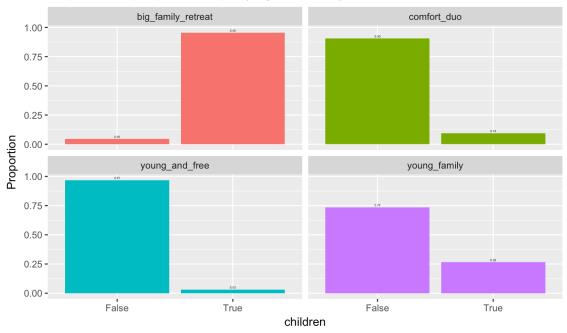


Figure 32. Proportion of people accompanying children in the trip by cluster

## Number of people accompanying children by cluster

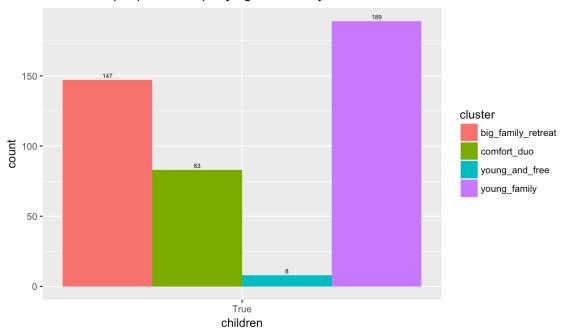


Figure 33. Number of people accompanying children in the trip by cluster

## Proportion of people renting car by cluster

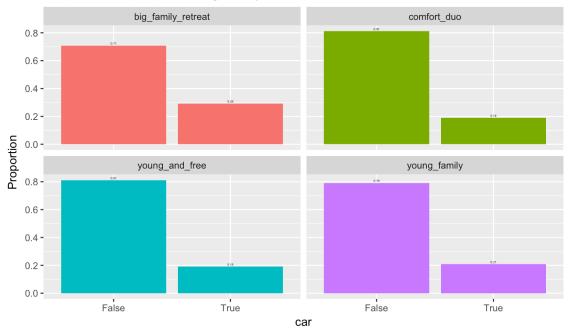


Figure 34. Proportion of people renting car by cluster

## Number of people renting car by cluster

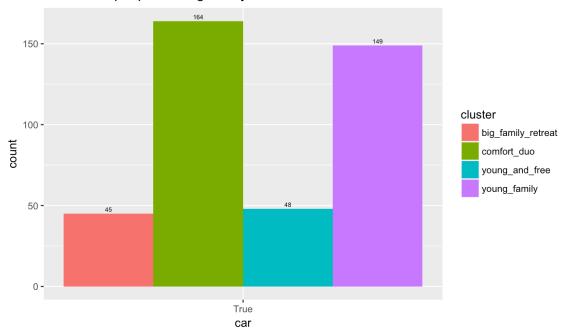


Figure 35. Number of people renting car by cluster

## Proportion of people renting hotel by cluster

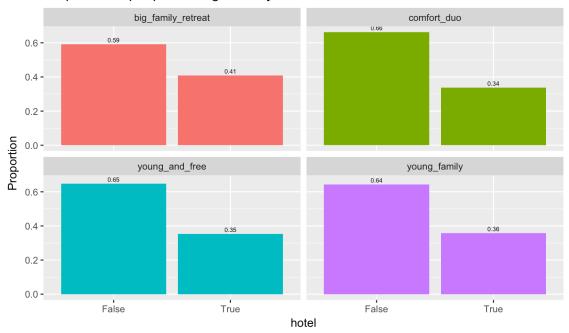


Figure 36. Proportion of people renting hotel by cluster

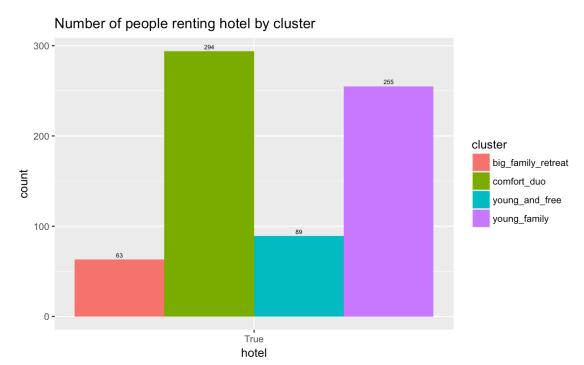


Figure 37. Number of people renting hotel by cluster

## Proportion of people with insurance by cluster

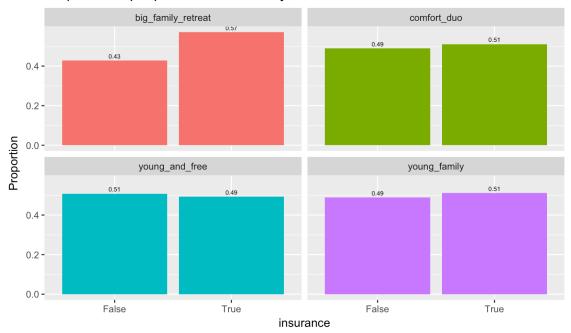


Figure 38. Proportion of people with insurance by cluster

## Number of people with insurance by cluster

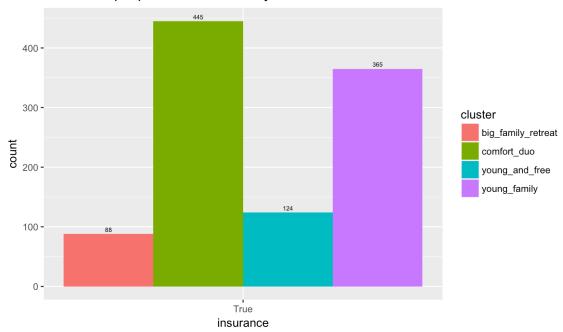


Figure 39. Number of people with insurance by cluster

## Proportion of people for each customer type by cluster

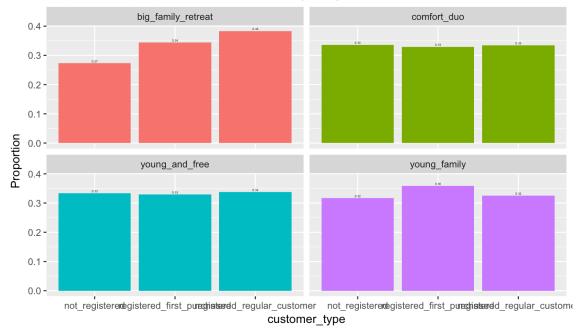


Figure 40. Proportion of people for each customer type by cluster

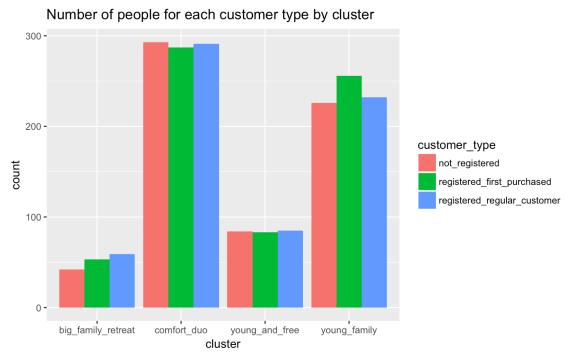


Figure 41. Number of people for each customer type by cluster

## Gender proportion by cluster

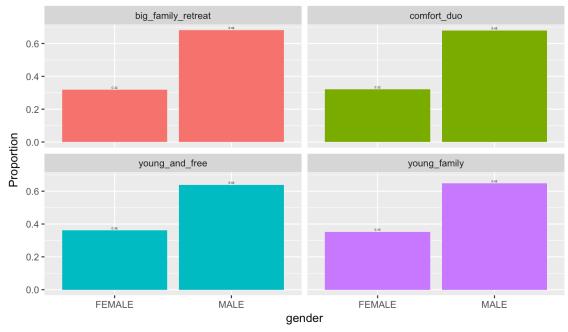


Figure 42. Proportion of gender by cluster

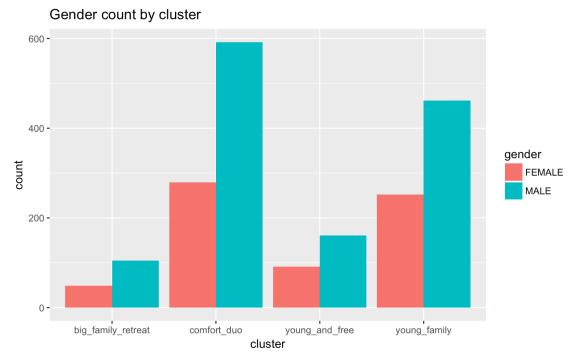


Figure 43. Number of each gender by cluster