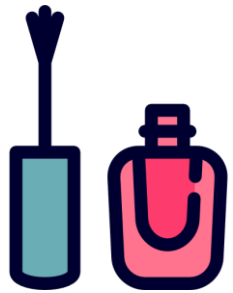
PROFILING & ANALYSING CUSTOMER BEHAVIOUR – USING CLUSTER & ASSOCIATION ANALYSIS

Team V
Andrew Lim
Sarah Chow

SPONSOR BACKGROUND

- A local startup launched in May 2015
- Bridge the gap between customers & beauty professionals

- **Customers** can book beauty services at their convenience
- **Beauty professionals** grow their brand & customer base



PROJECT OBJECTIVES

Customers

1. To determine the customer segmentation (different groups of customers) from the current booking patterns
2. To understand customers' behaviour. When was the last time a customer used the app? How frequent does a customer use the app? How much does a customer spend on average and in total?
3. To understand how customers select the type of services within each booking

PROJECT OBJECTIVES

Beauty Professionals

1. To determine the type of services that are commonly published
2. To determine the different price points for the different types of services

METHODOLOGY

jmp®
Statistical Discovery.™ From SAS.

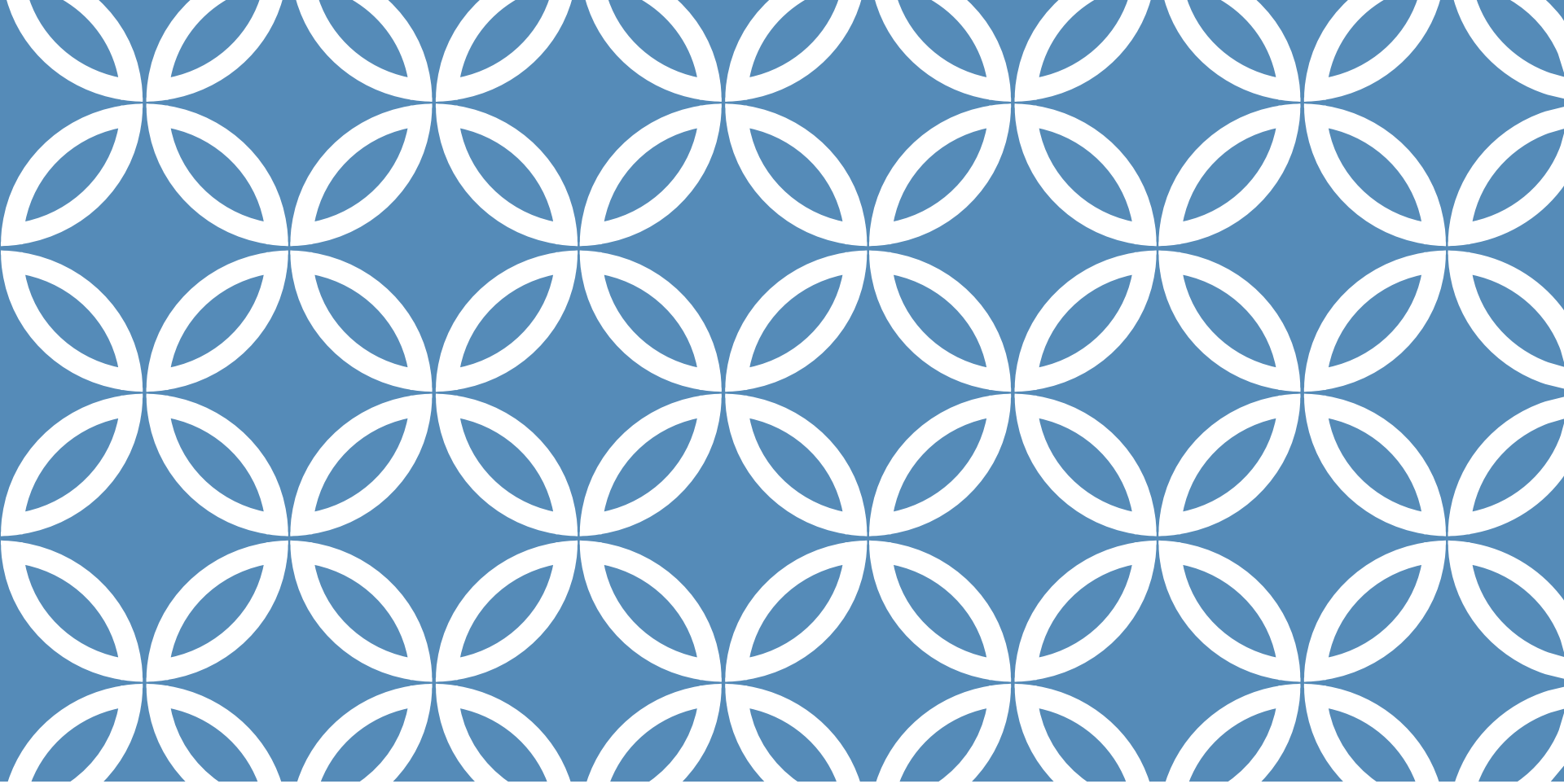
DATA COLLECTION

- Access to their MongoDB database
- Date range used from **Jan 2015 to Dec 2016**
- Narrowed down from 59 to 7 relevant data tables

DATA CLEANING

1. Filtering out test data
   - Removed by specific columns (test_at, is_test, deleted_at)
   - Removed test data across data tables
   - e.g. remove admin users AND bookings made by them

2. Duplicate & missing values
   - e.g. missing monetary values within bookings data

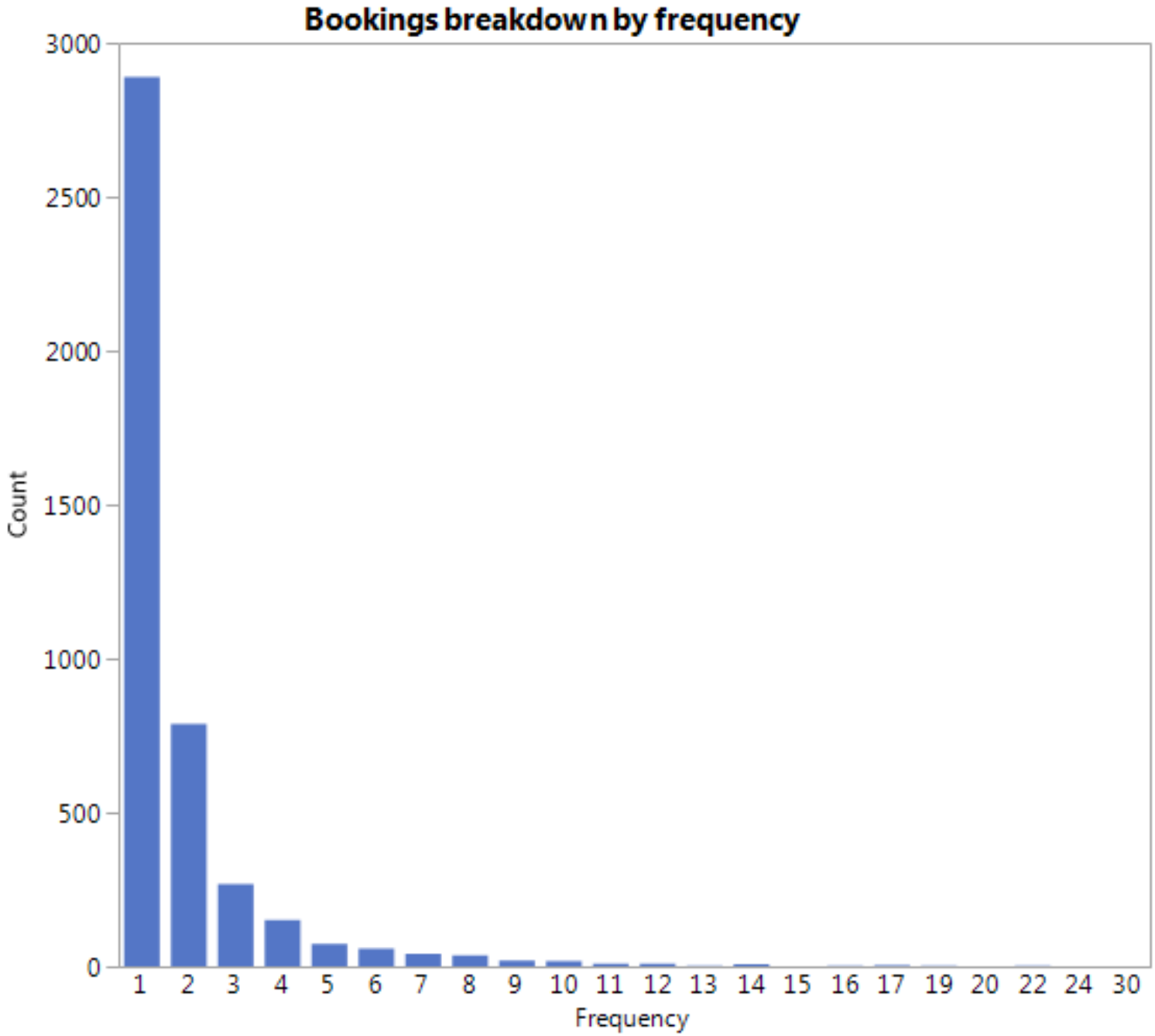


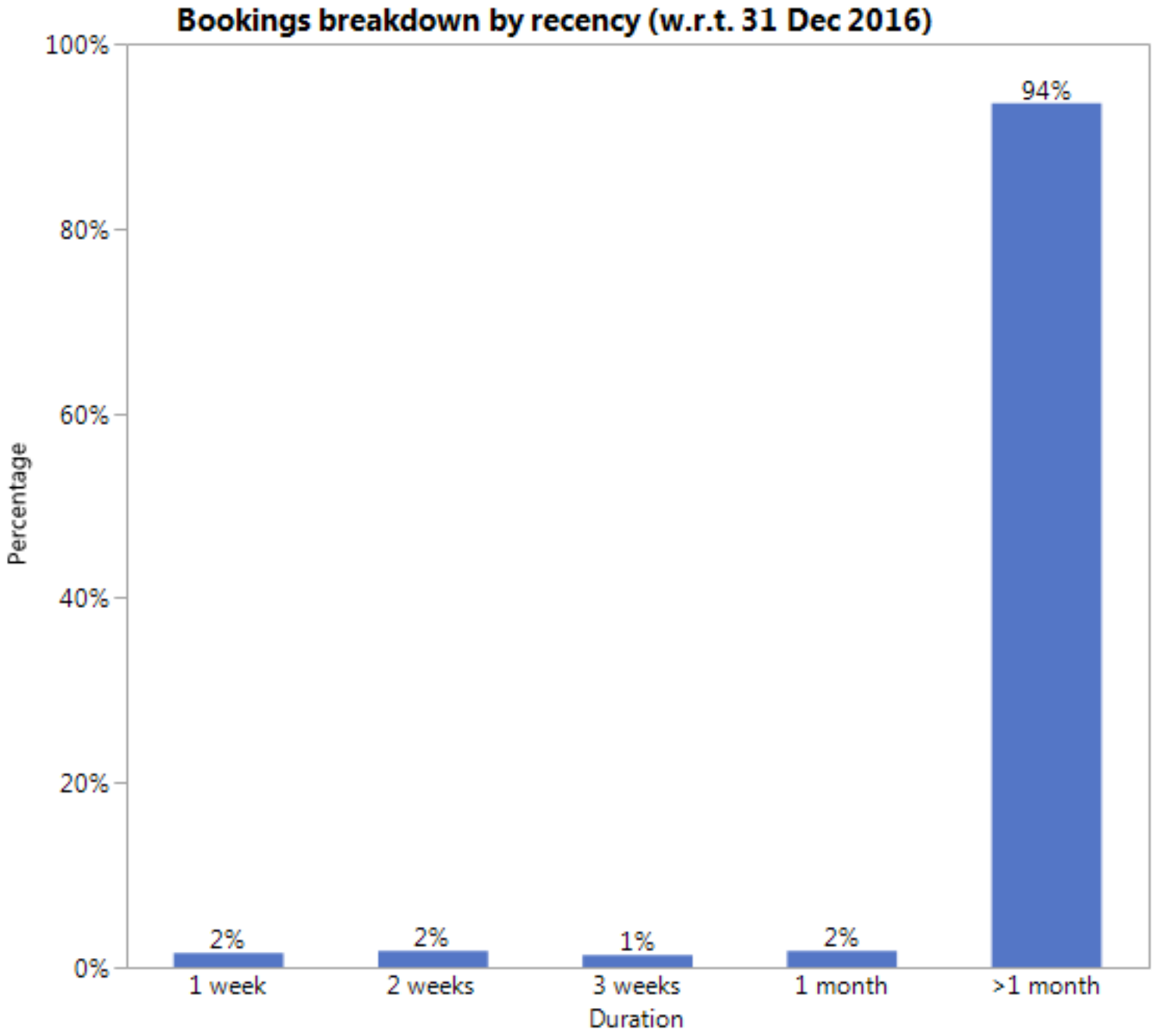
EXPLORATORY DATA ANALYSIS

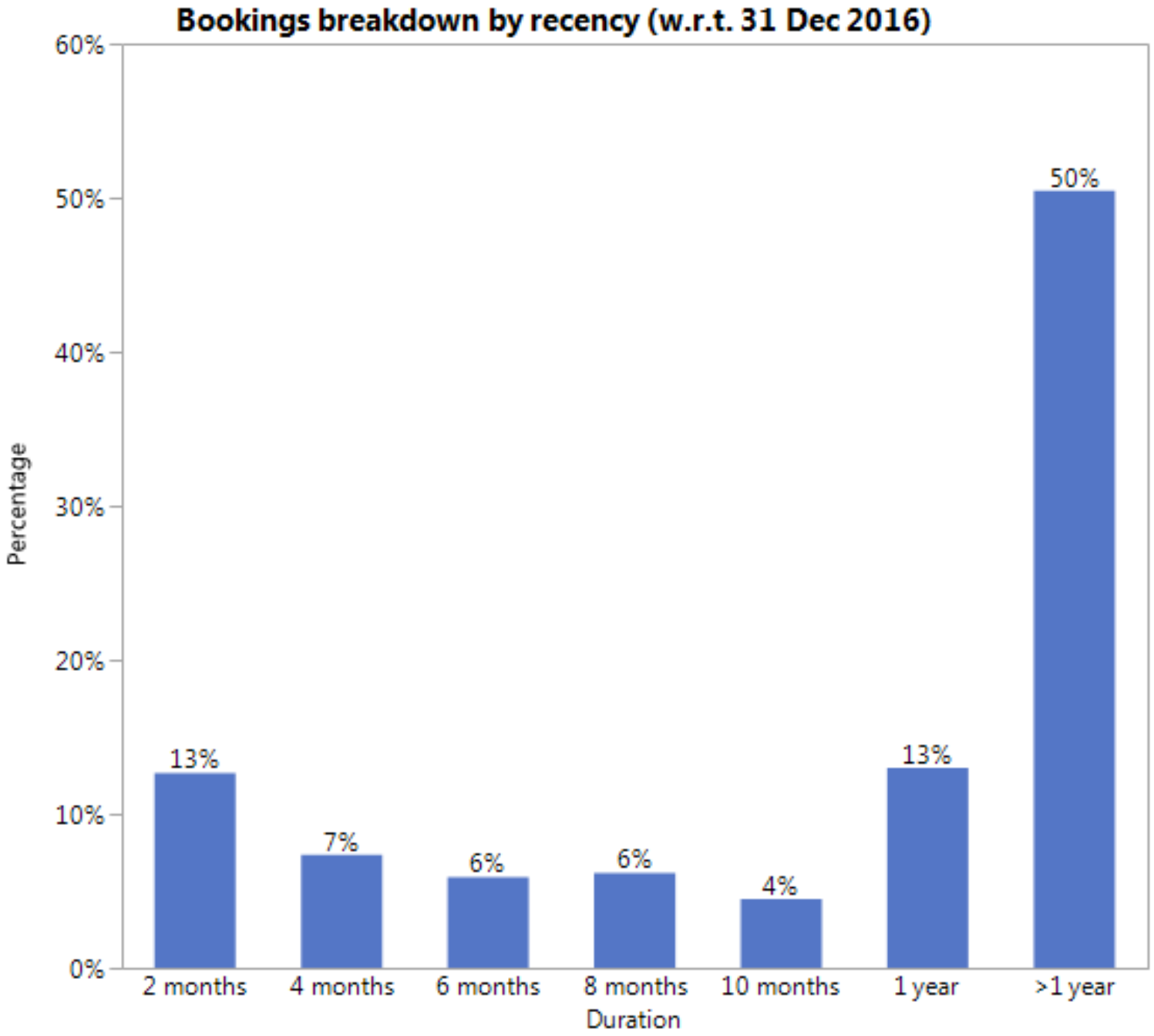
Recency
Frequency
Monetary

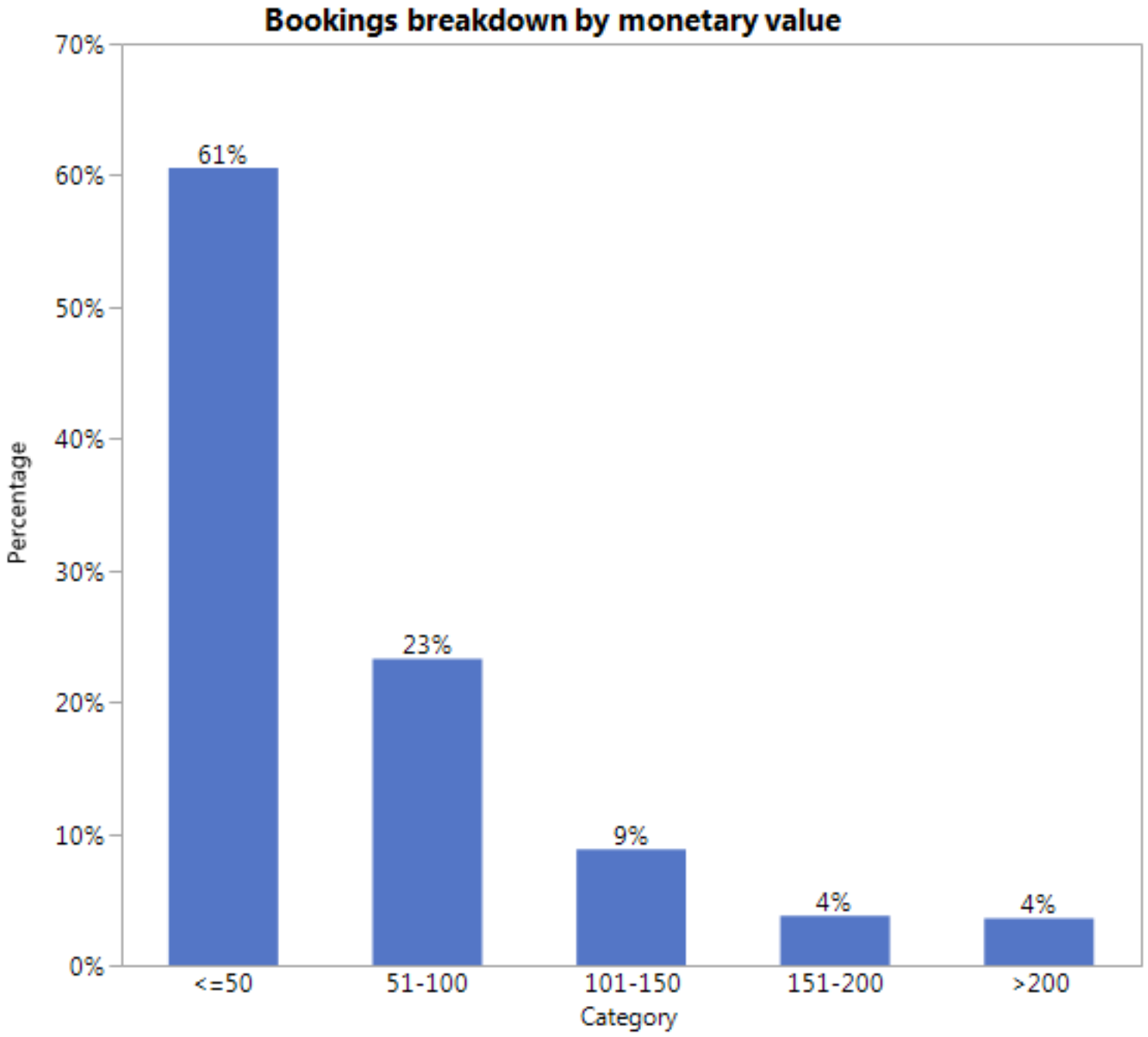
EXPLORATORY DATA ANALYSIS (EDA)

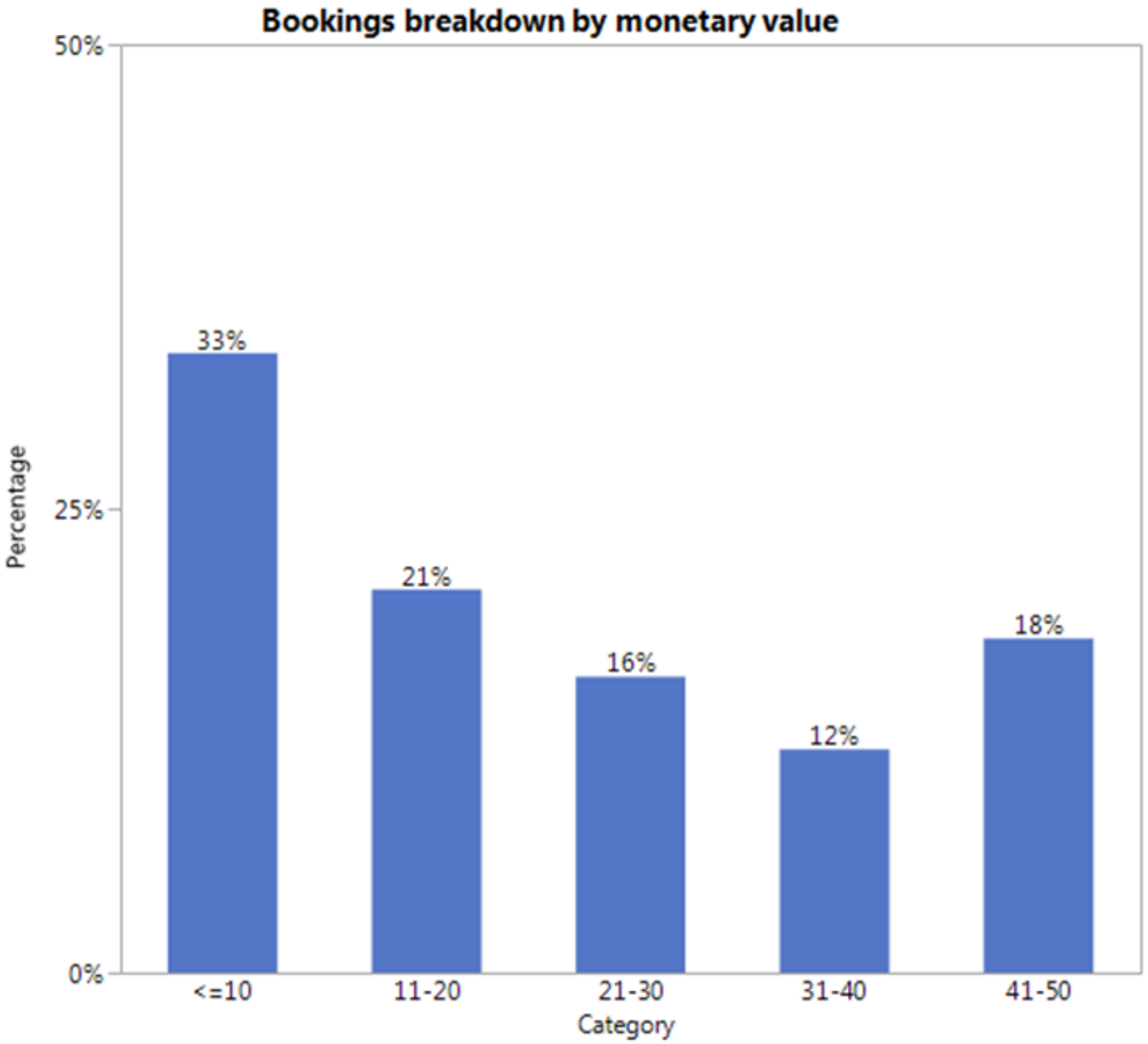
- Users breakdown by type
- Customers breakdown by age, gender
- Professionals breakdown by age, gender
- Bookings breakdown by type
- Bookings breakdown by status
- Bookings breakdown by year, month, day
- Bookings breakdown by category
- **Bookings breakdown by frequency, recency, monetary amount**
- Bookings breakdown by duration from sign up to 1st online booking
- Bookings breakdown by service count
- Services breakdown by price, category


Bookings breakdown by frequency
3000
2500
2000
1500
1000
500
0
Count
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
20
22
24
30
Frequency

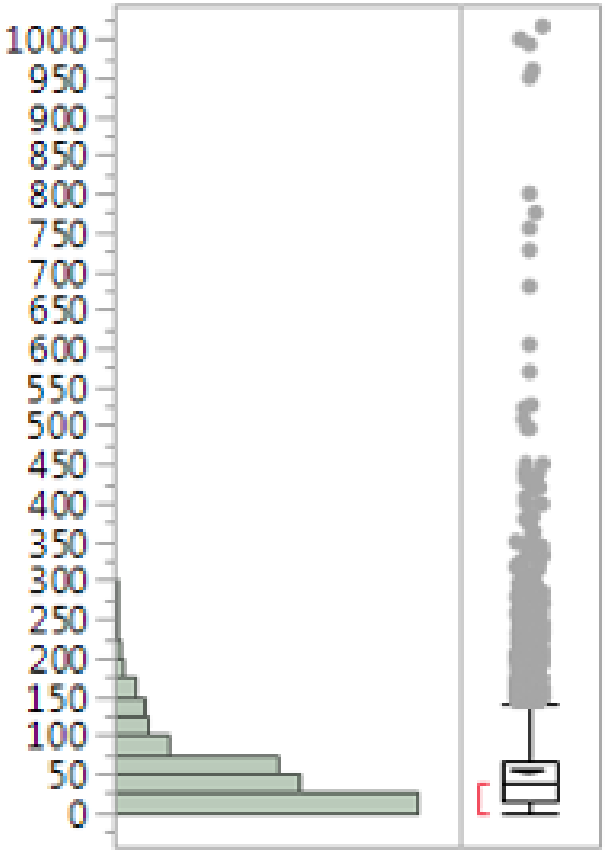

Bookings breakdown by recency (w.r.t. 31 Dec 2016)
100%
80%
60%
40%
20%
0%
Percentage
94%
2%
2%
1%
2%
1 week
2 weeks
3 weeks
1 month
>1 month
Duration


Bookings breakdown by recency (w.r.t. 31 Dec 2016)
60%
50%
40%
30%
20%
10%
0%
Percentage
13%
7%
6%
6%
4%
13%
50%
2 months
4 months
6 months
8 months
10 months
1 year
>1 year
Duration


Bookings breakdown by monetary value
70%
60%
50%
40%
30%
20%
10%
0%
Percentage
61%
23%
9%
4%
4%
<=50
51-100
101-150
151-200
>200
Category


Bookings breakdown by monetary value
50%
25%
0%
Percentage
33%
21%
16%
12%
18%
<=10
11-20
21-30
31-40
41-50
Category

BOOKINGS (NAILS)

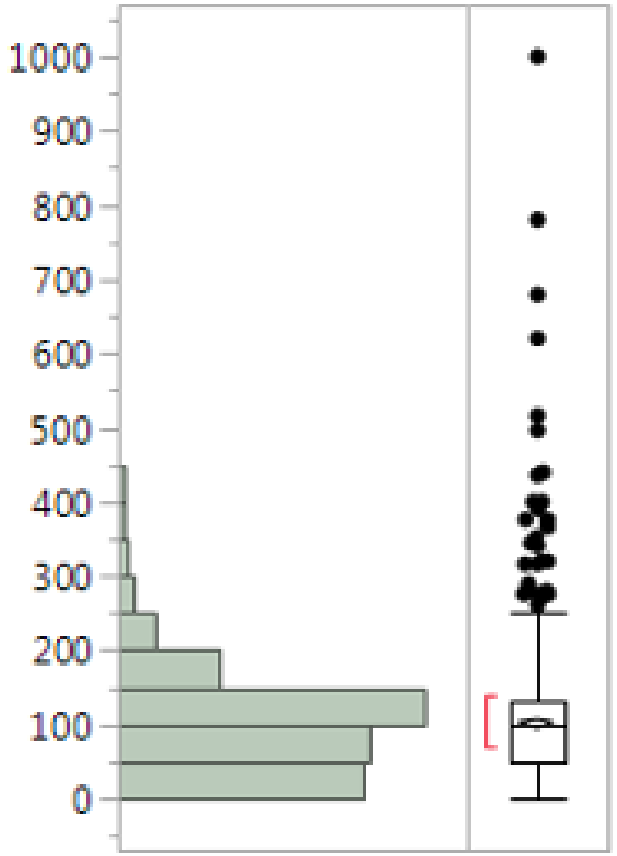


100.0%	maximum		1016
99.5%		392.74975	
	97.5%		210
	90.0%		127.15
	75.0%	quartile	66.485
	50.0%	median	37.84
	25.0%	quartile	15
	10.0%		0
	2.5%		0
	0.5%		0
0.0%	minimum		0

niche

mass

BOOKINGS (MAKEUP)

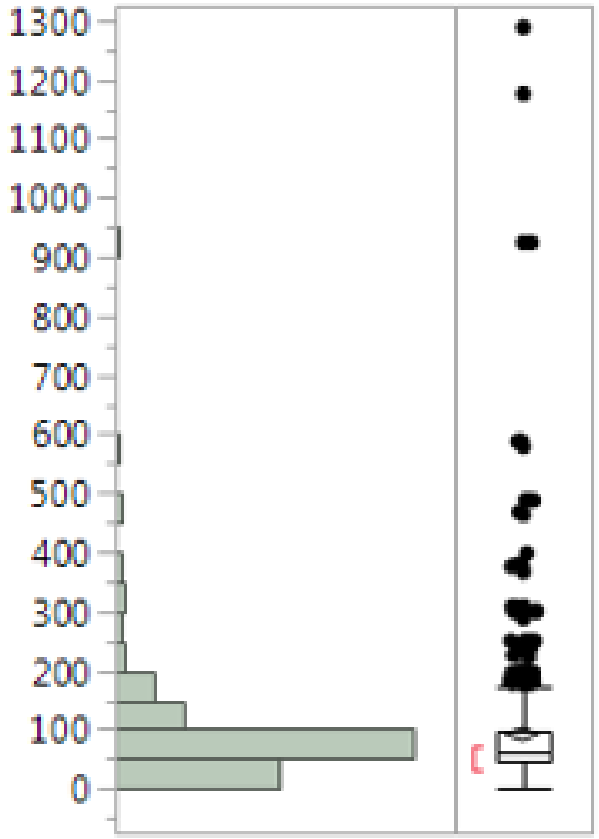


100.0%	maximum		1000
	99.5%		660.415
	97.5%		342.275
	90.0%		180
	75.0%	quartile	130
	50.0%	median	98.985
	25.0%	quartile	50
	10.0%		0
	2.5%		0
	0.5%		0
0.0%	minimum		0

niche

mass

BOOKINGS (BROWS)

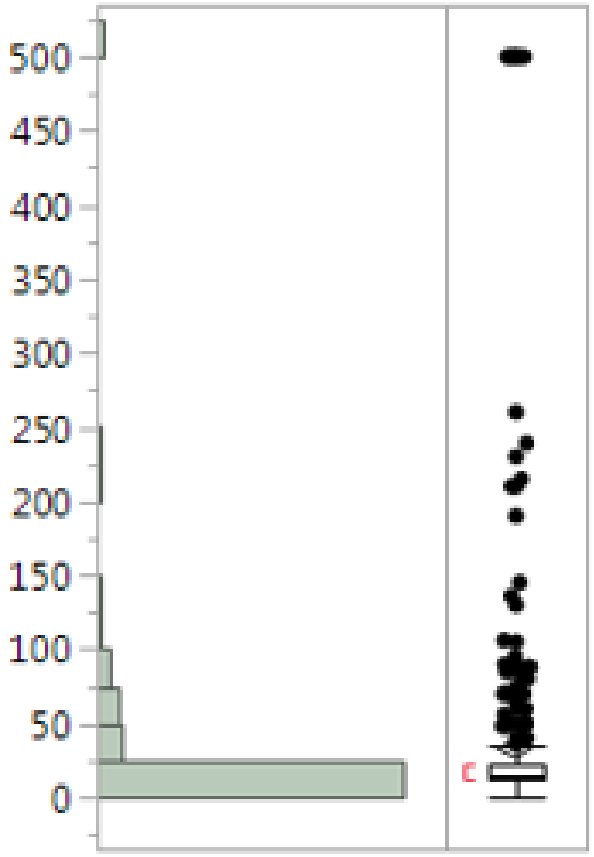


100.0%	maximum		1288
	99.5%		1003.38
	97.5%		390.35
	90.0%		178.6
	75.0%	quartile	98.6
	50.0%	median	60
	25.0%	quartile	46.74
	10.0%		20
	2.5%		0
	0.5%		0
0.0%	minimum		0

niche

mass

BOOKINGS (HAIR STYLING)

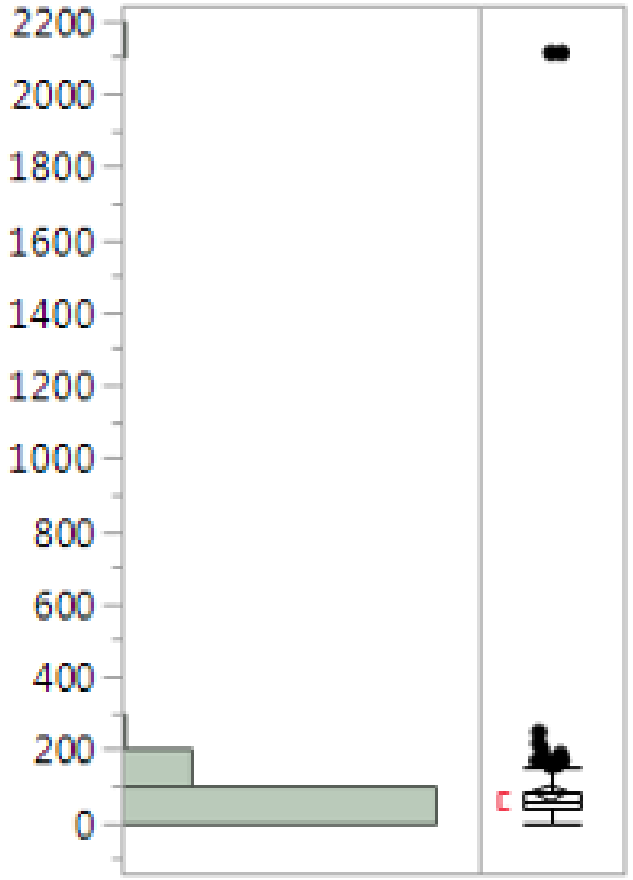


100.0%	maximum		500
	99.5%		500
	97.5%		220.625
	90.0%		69
	75.0%	quartile	22
	50.0%	median	13.5
	25.0%	quartile	13
	10.0%		10
	2.5%		0
	0.5%		0
0.0%	minimum		0

niche

mass

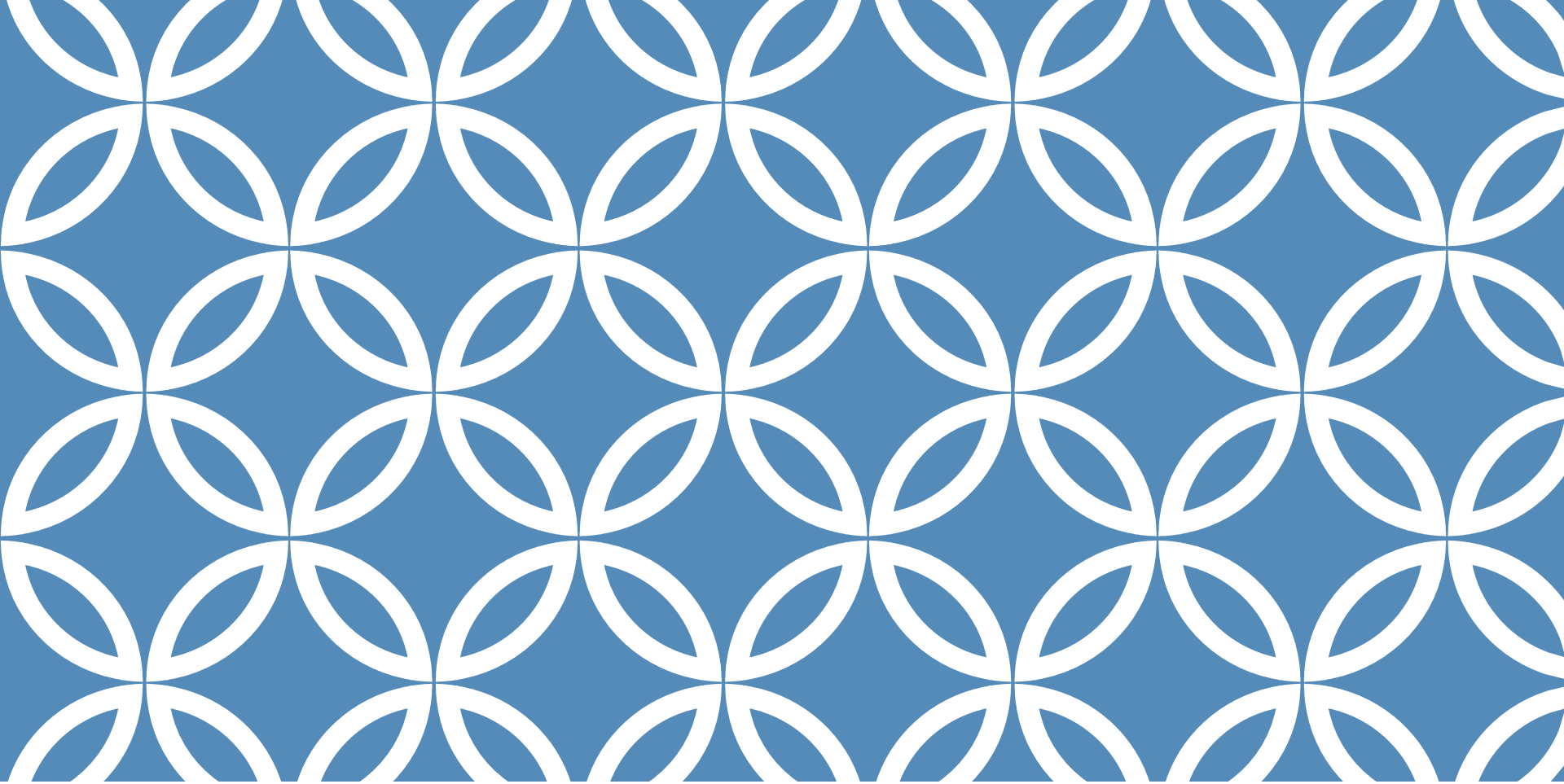
BOOKINGS (FACIAL)



100.0%	maximum		2112
	99.5%		2112
	97.5%		169.8
	90.0%		128
	75.0%	quartile	85
	50.0%	median	58
	25.0%	quartile	40
	10.0%		19.8
	2.5%		0
	0.5%		0
0.0%	minimum		0

niche

mass



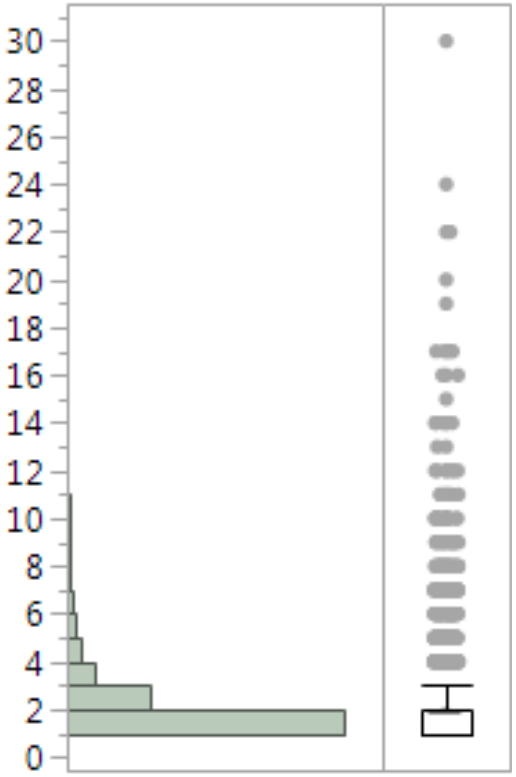
CLUSTER ANALYSIS

Customer Segmentation

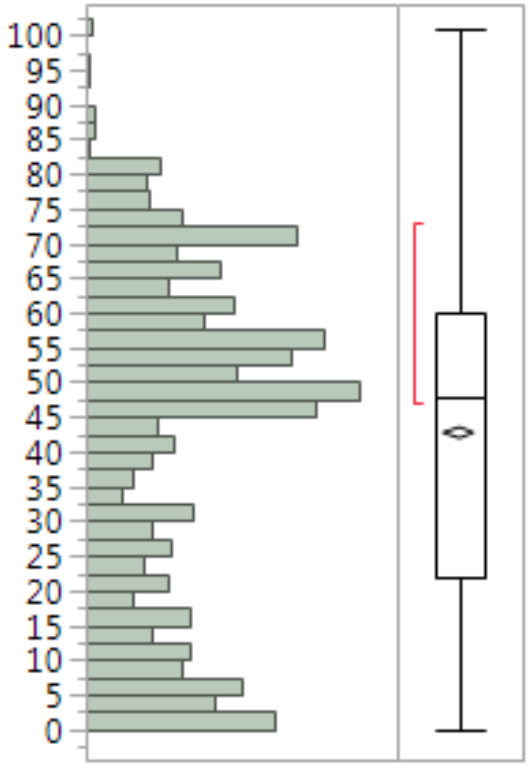
CLUSTER ANALYSIS

- Aims to discover any potential customer segmentation
- K-means clustering
- Looked at customers who have made **at least one** booking
- Clustering variables
  - Booking recency
  - Booking frequency
  - Total monetary value

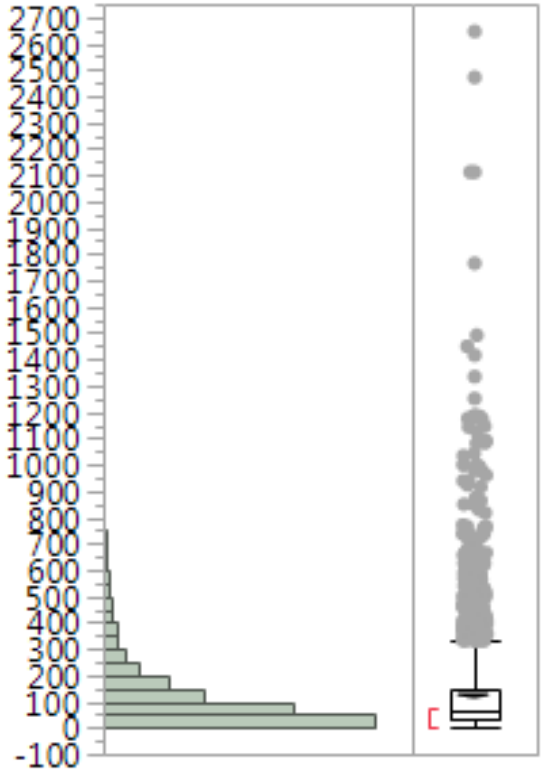
CLUSTER ANALYSIS



booking_frequency

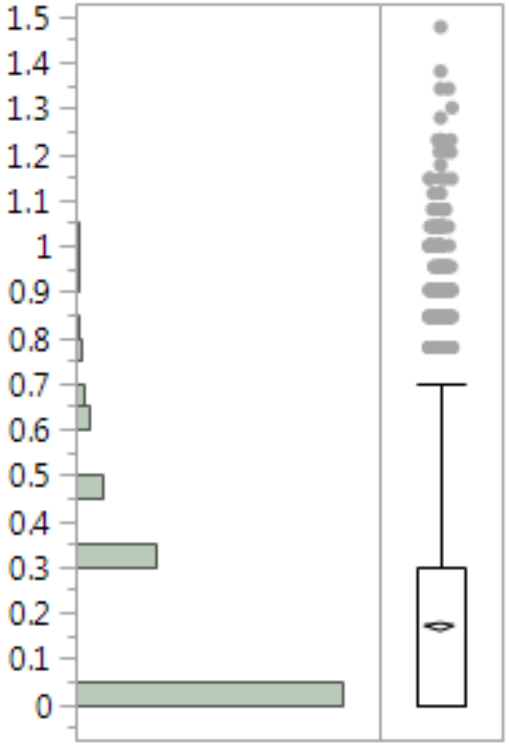


booking_recency

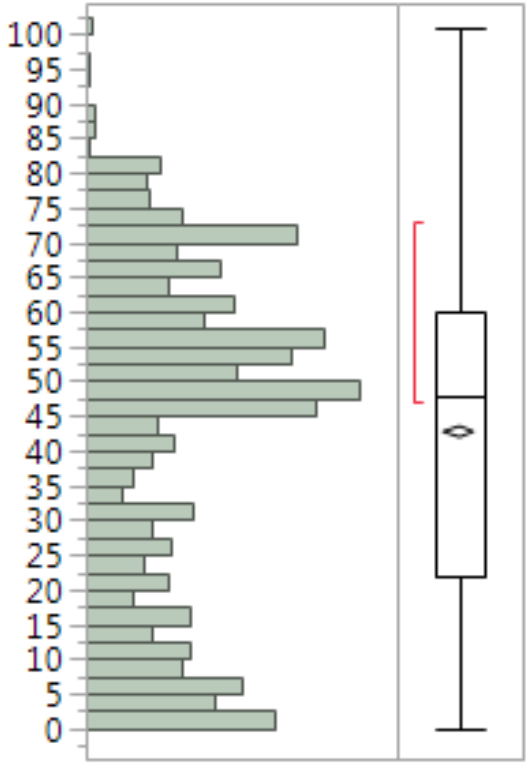


booking_monetary_total

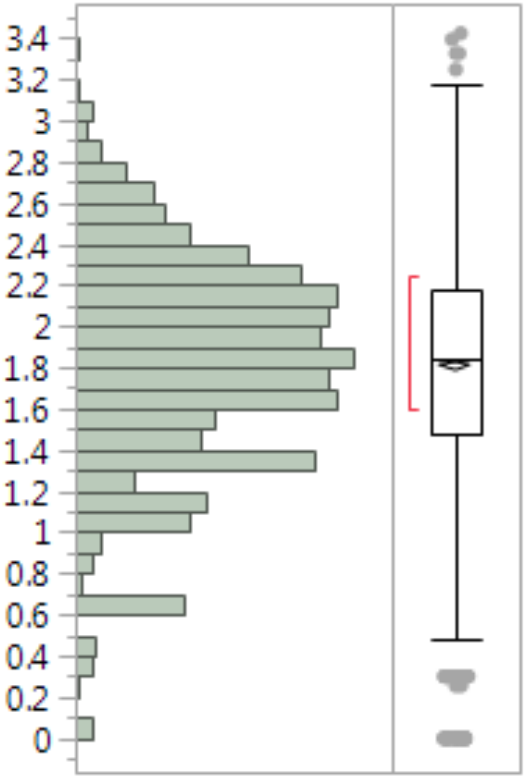
CLUSTER ANALYSIS



booking_frequency (log10)

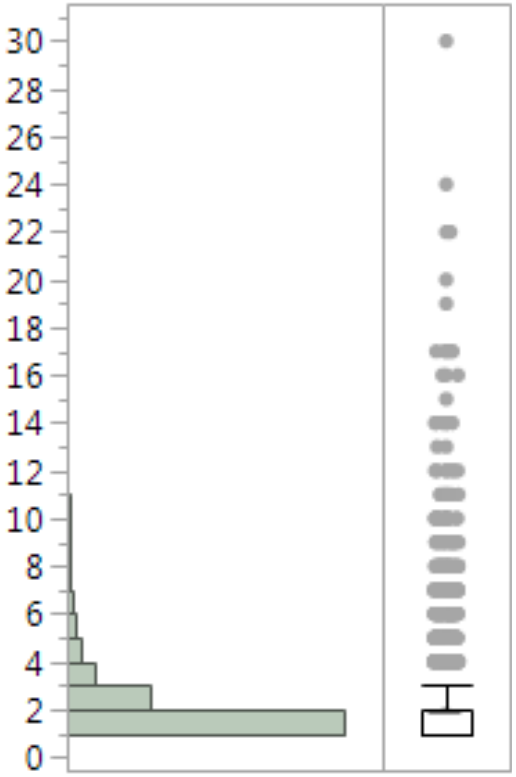


booking_recency

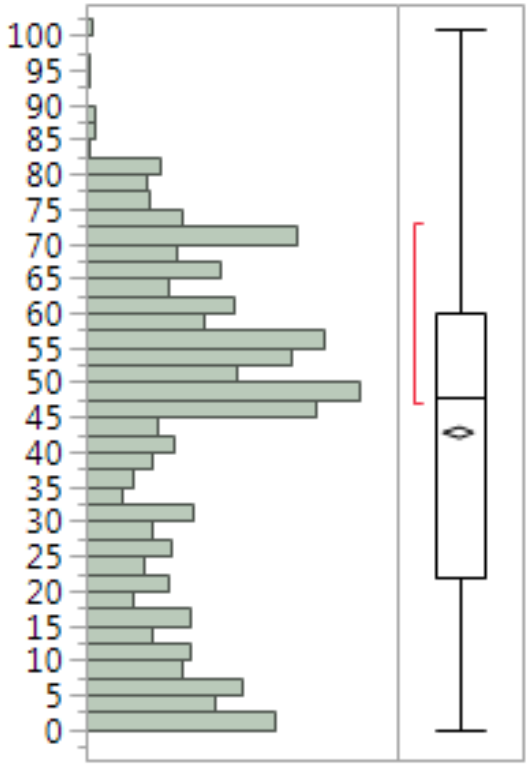


booking_monetary_total (log10)

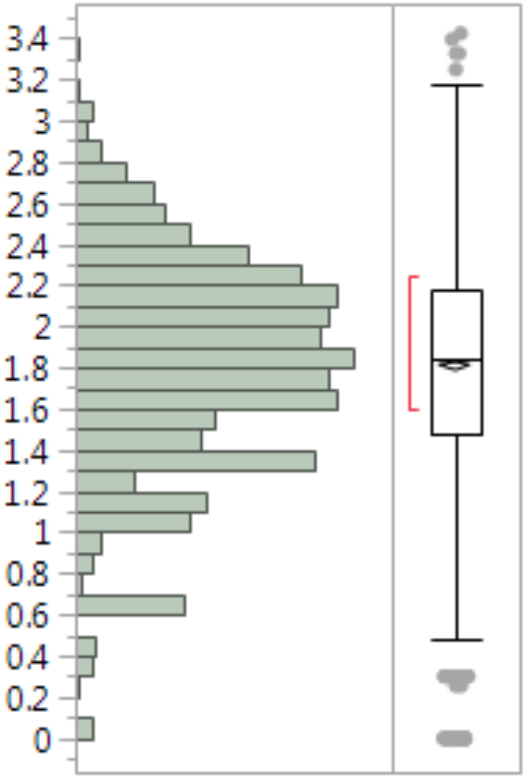
CLUSTER ANALYSIS



booking_frequency



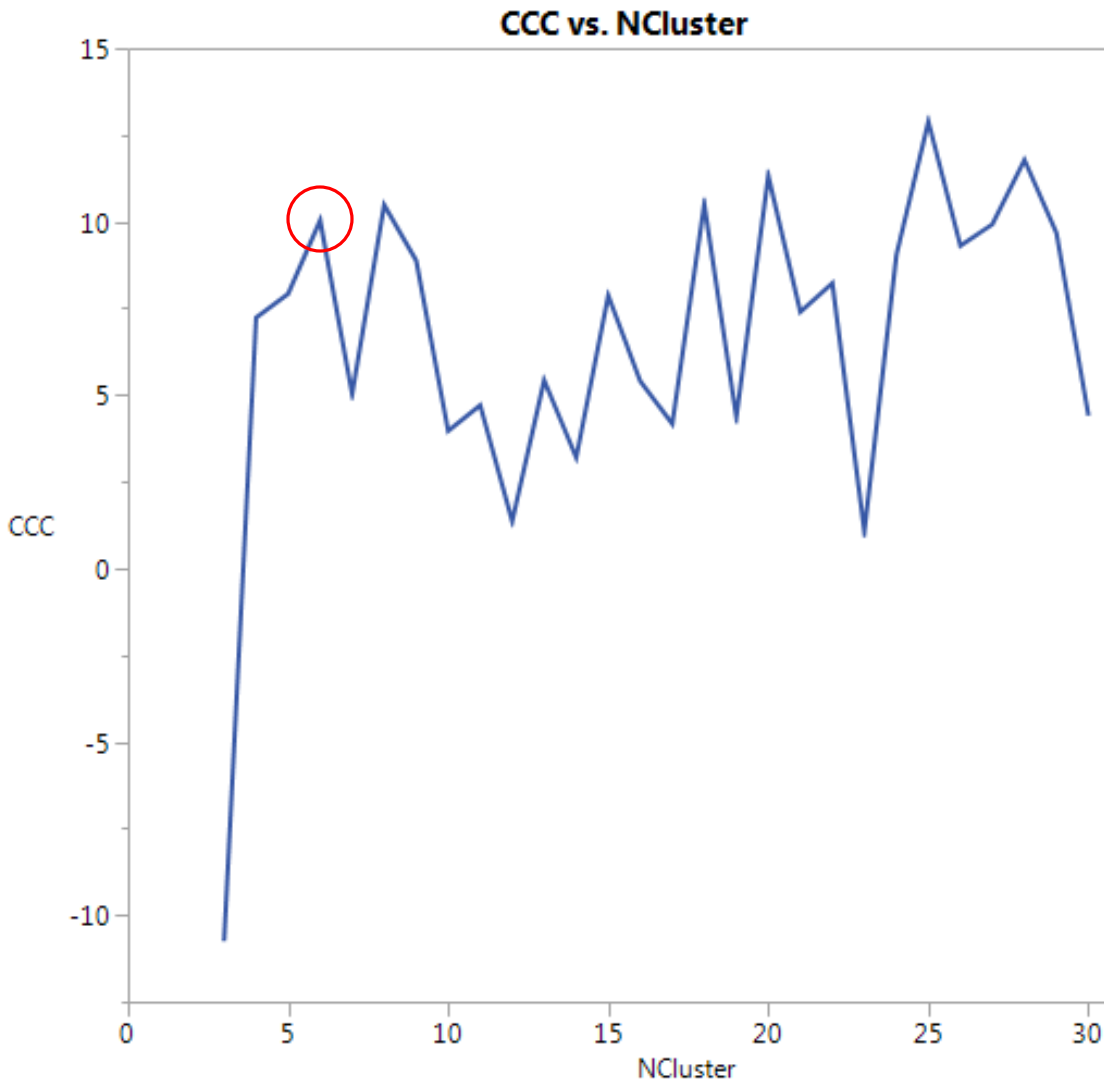
booking_recency



booking_monetary_total (log10)

K-means clustering

Cluster range 3 to 30



Cluster Comparison

Method	NCluster	CCC	Best
K-Means Clustering	3	-10.75	
K-Means Clustering	4	7.23595	
K-Means Clustering	5	7.92032	
K-Means Clustering	6	10.0372	
K-Means Clustering	7	5.04086	
K-Means Clustering	8	10.4839	Optimal CCC
K-Means Clustering	9	8.88005	
K-Means Clustering	10	3.96878	
K-Means Clustering	11	4.70582	
K-Means Clustering	12	1.36393	
K-Means Clustering	13	5.42802	
K-Means Clustering	14	3.20073	
K-Means Clustering	15	7.87067	
K-Means Clustering	16	5.40176	
K-Means Clustering	17	4.17096	
K-Means Clustering	18	10.5065	
K-Means Clustering	19	4.36813	
K-Means Clustering	20	11.326	
K-Means Clustering	21	7.40043	
K-Means Clustering	22	8.22926	
K-Means Clustering	23	1.07648	
K-Means Clustering	24	9.03791	
K-Means Clustering	25	12.8891	
K-Means Clustering	26	9.30894	
K-Means Clustering	27	9.92589	
K-Means Clustering	28	11.7841	
K-Means Clustering	29	9.67832	
K-Means Clustering	30	4.40883	

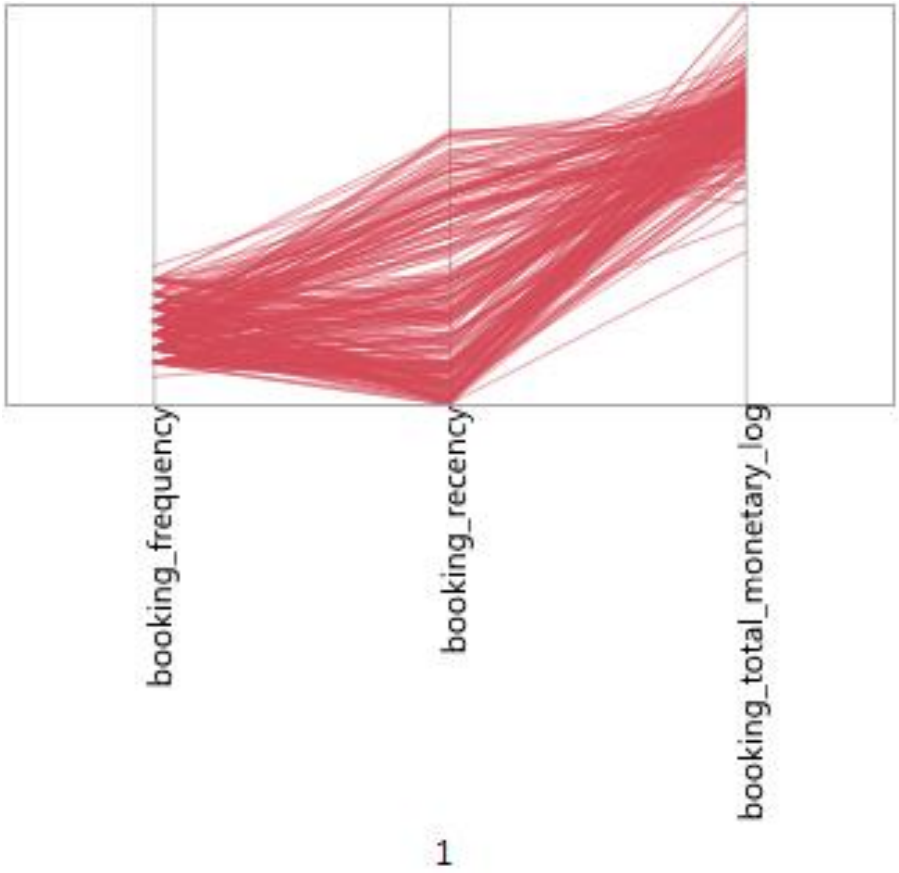
No. of clusters = 6

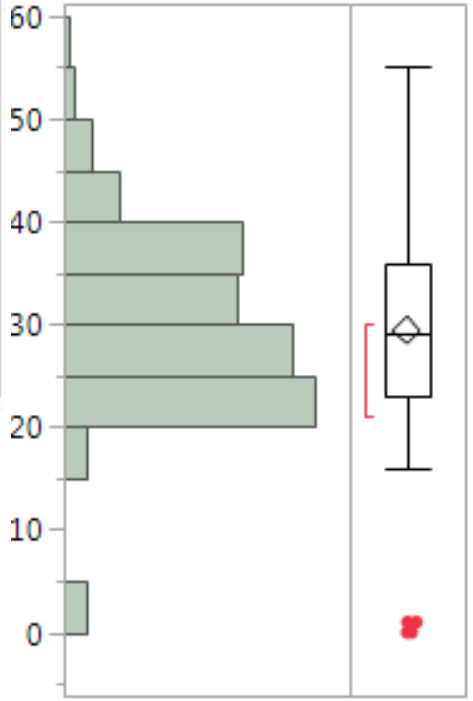
Cluster Summary

Cluster	Count	Step	Criterion
1	258	55	0
2	620		
3	42		
4	1121		
5	793		
6	1020		

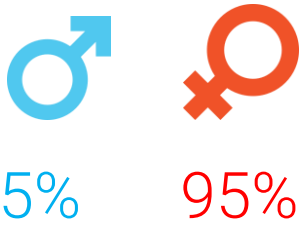
Cluster Means

Cluster	booking_frequency	booking_recency	booking_total_monetary_log
1	6.29844961	27.0232558	2.47834188
2	1.1516129	68.4483871	0.97168195
3	14.6190476	16.1904762	2.74301569
4	1.29884032	56.8412132	1.73241173
5	1.85119798	47.0063052	2.30626783
6	1.46666667	12.7460784	1.81295839



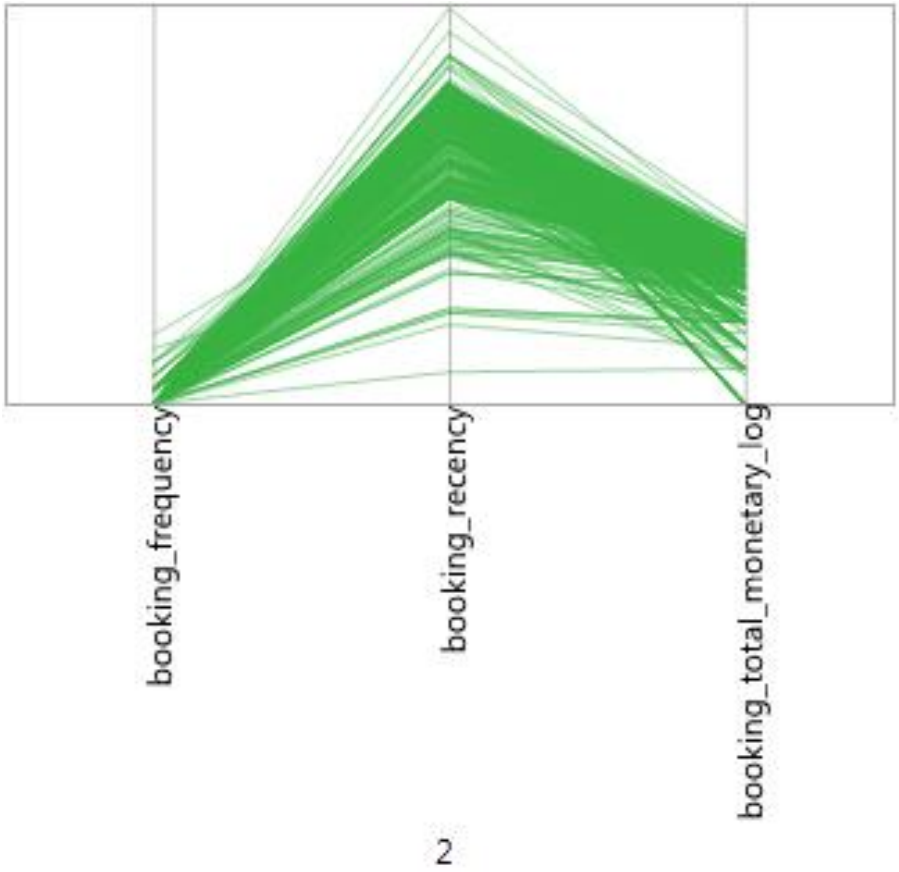


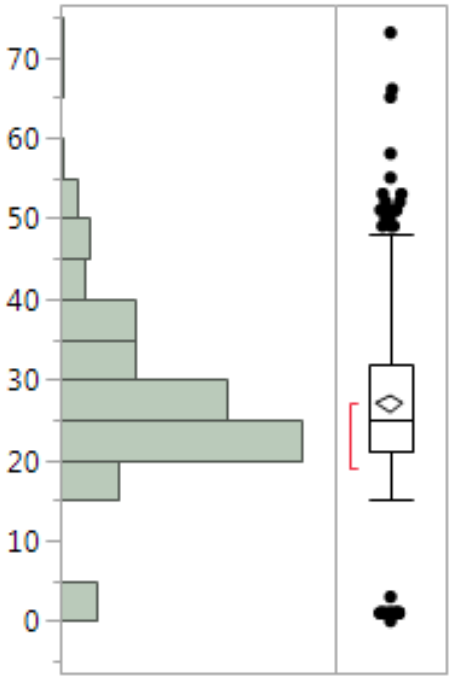
Quantiles		
100.0%	maximum	55
99.5%		55
97.5%		47
90.0%		40.4
75.0%	quartile	36
50.0%	median	29
25.0%	quartile	23
10.0%		21
2.5%		7
0.5%		0
0.0%	minimum	0



- Moderate frequency
- Widespread recency
- High monetary value

- 258 customers
- Age around 23 to 36
- Average total monetary spending of $125





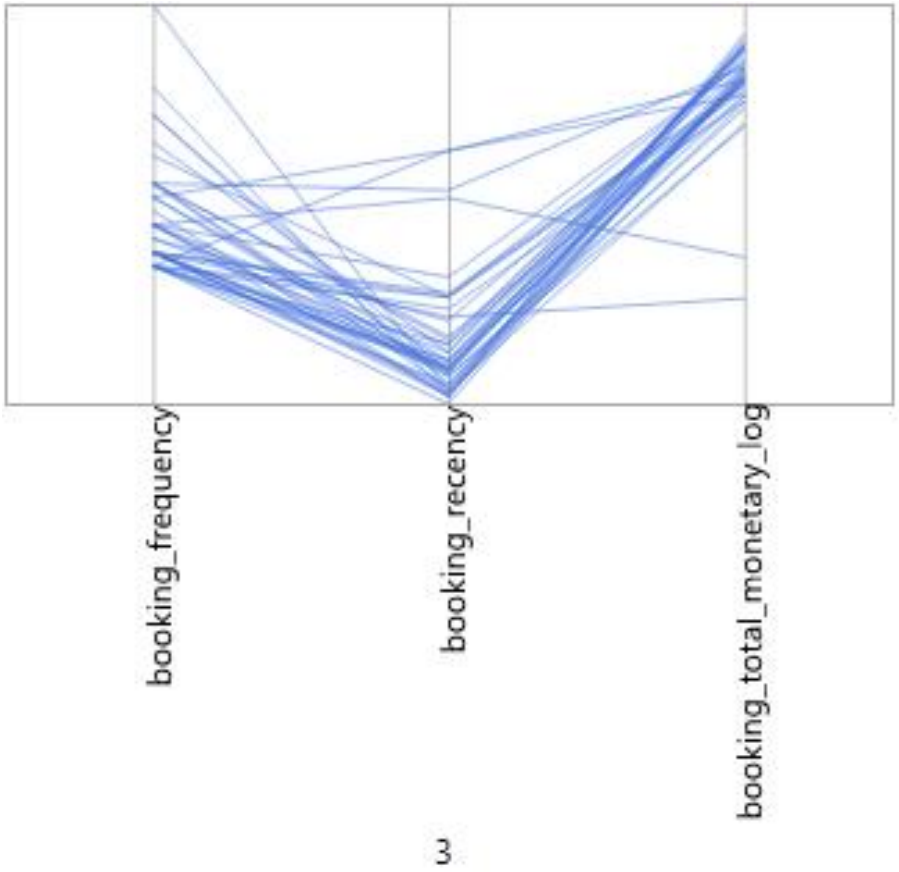
Quantiles		
100.0%	maximum	73
99.5%		65.82
97.5%		51
90.0%		40.4
75.0%	quartile	32
50.0%	median	25
25.0%	quartile	21
10.0%		18
2.5%		1
0.5%		1
0.0%	minimum	0

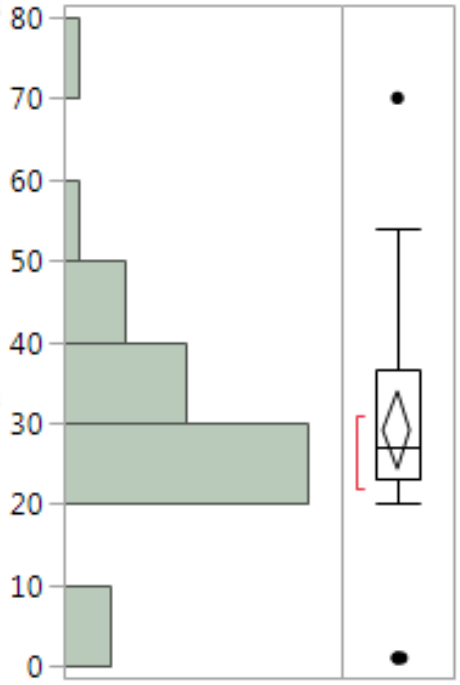
♂ 6%

♀ 94%

- Low frequency
- High recency (haven't booked in a year)
- Low monetary value

- 620 customers
- Age around 21 to 32
- Average total monetary spending of ~ $12



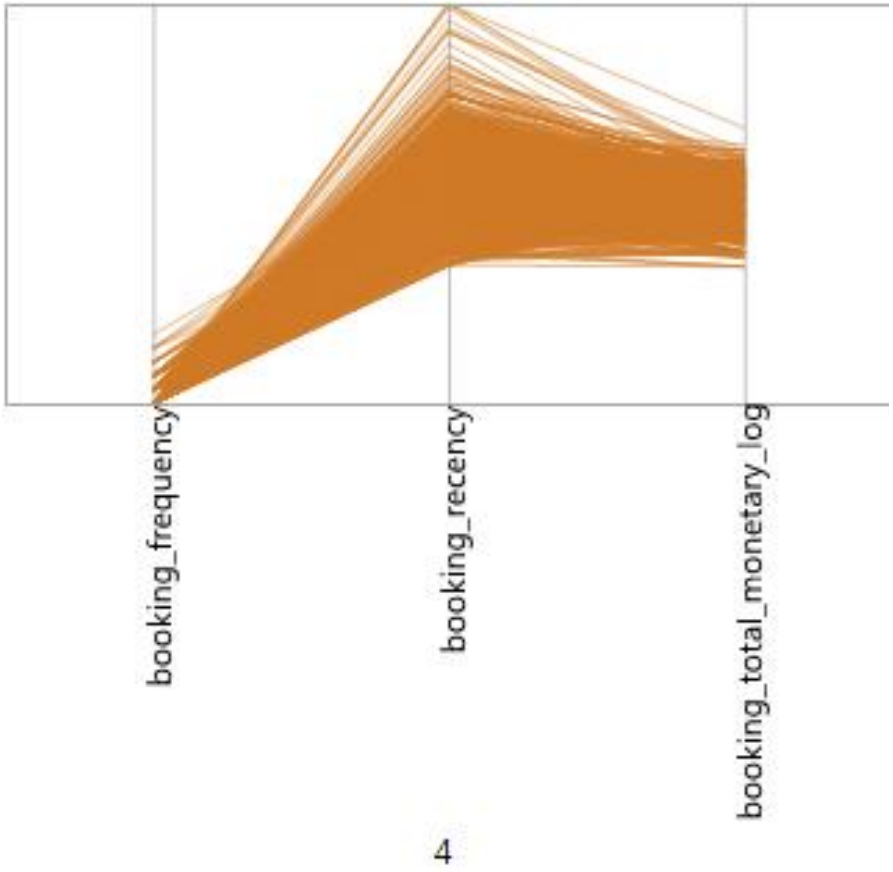


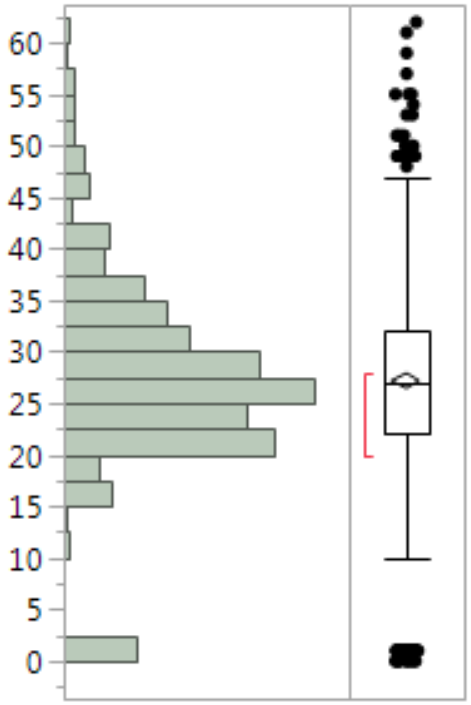
Quantiles		
100.0%	maximum	70
99.5%		70
97.5%		70
90.0%		43.6
75.0%	quartile	36.5
50.0%	median	27
25.0%	quartile	23
10.0%		8.6
2.5%		1
0.5%		1
0.0%	minimum	1

3% 97%

- High frequency
- Low recency (<4months)
- High monetary value

- 42 customers
- Age around 23 to 36
- Average total monetary spending of ~ $700





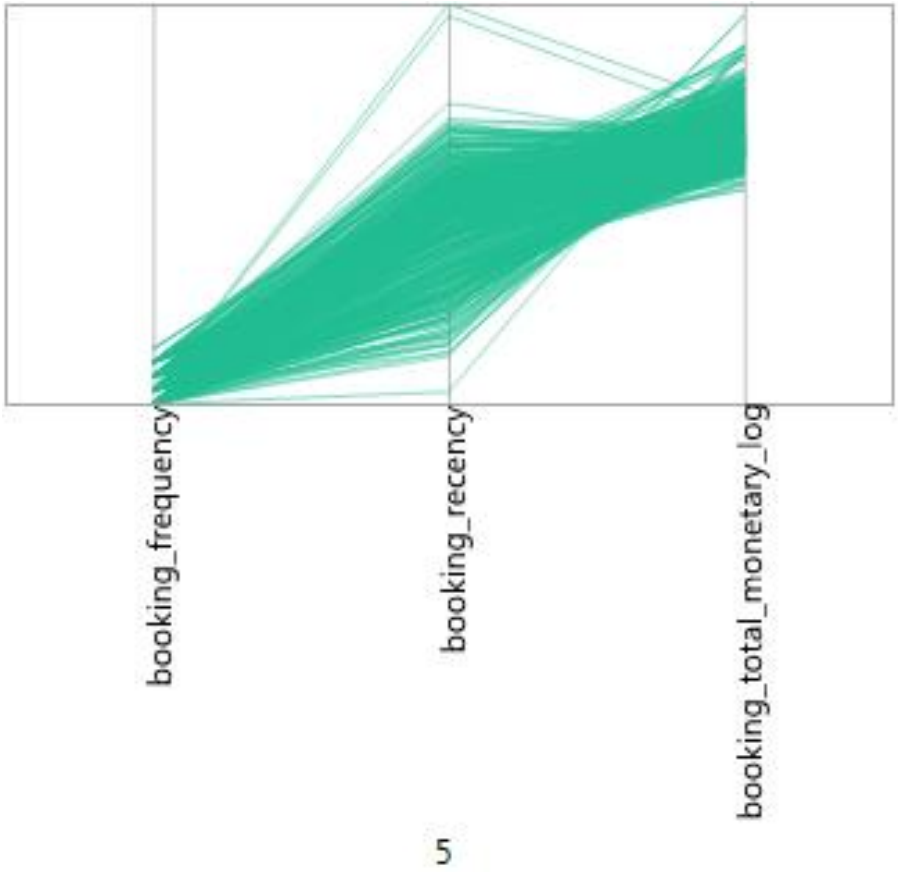
Quantiles		
100.0%	maximum	62
99.5%		57.17
97.5%		49
90.0%		38
75.0%	quartile	32
50.0%	median	27
25.0%	quartile	22
10.0%		19
2.5%		1
0.5%		0
0.0%	minimum	0

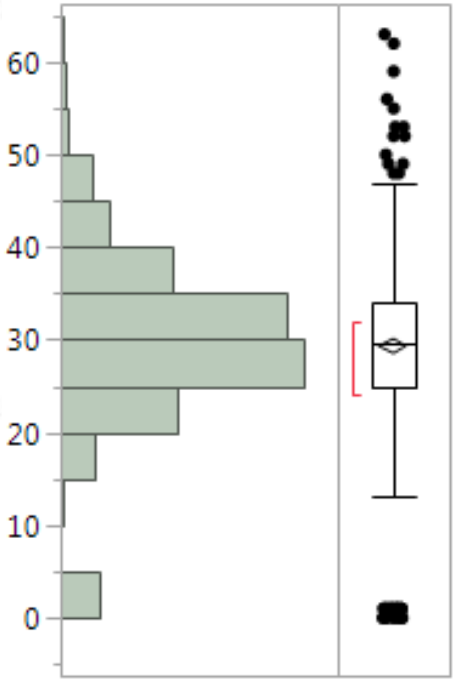
♂ 3%

♀ 97%

- Low frequency
- High recency (haven't booked in a year)
- Medium monetary value

- 1121 customers
- Age around 22 to 32
- Average total monetary spending of ~ $60



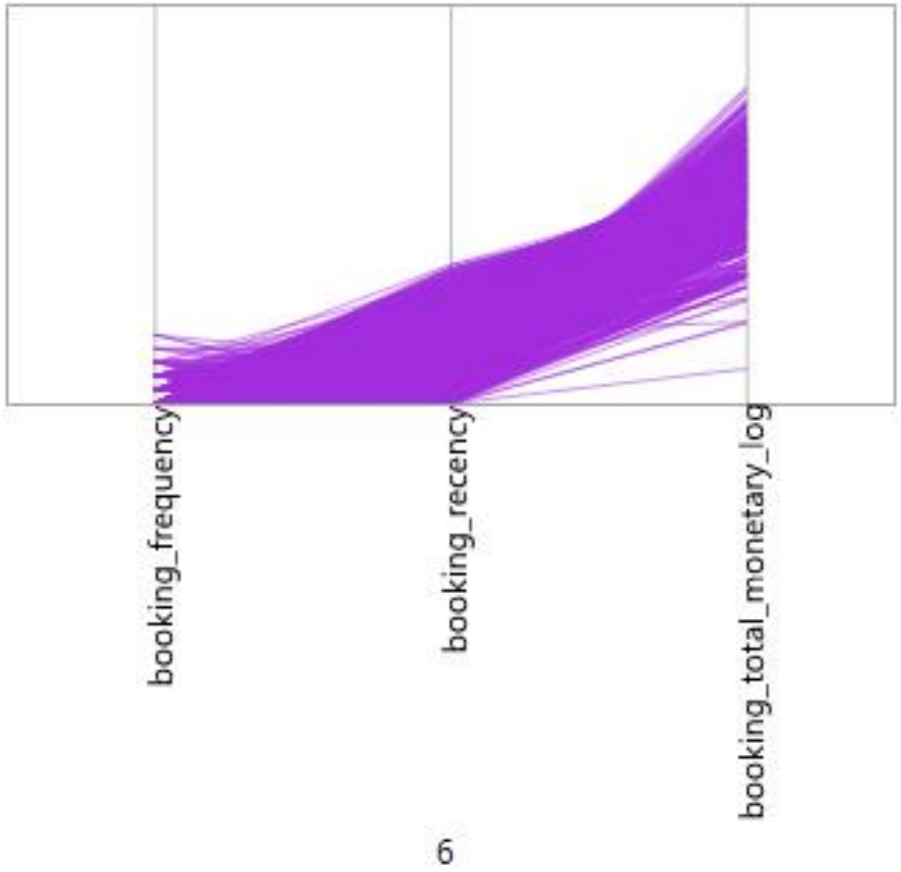


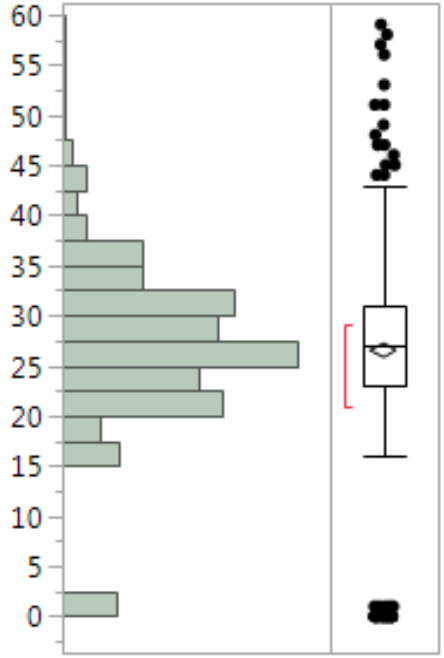
Quantiles		
100.0%	maximum	63
99.5%		59
97.5%		48
90.0%		40
75.0%	quartile	34
50.0%	median	29.5
25.0%	quartile	25
10.0%		20.3
2.5%		1
0.5%		0
0.0%	minimum	0

3% 97%

- Low frequency
- Moderate recency
- High monetary value

- 793 customers
- Age around 25 to 34
- Average total monetary spending of ~ $238





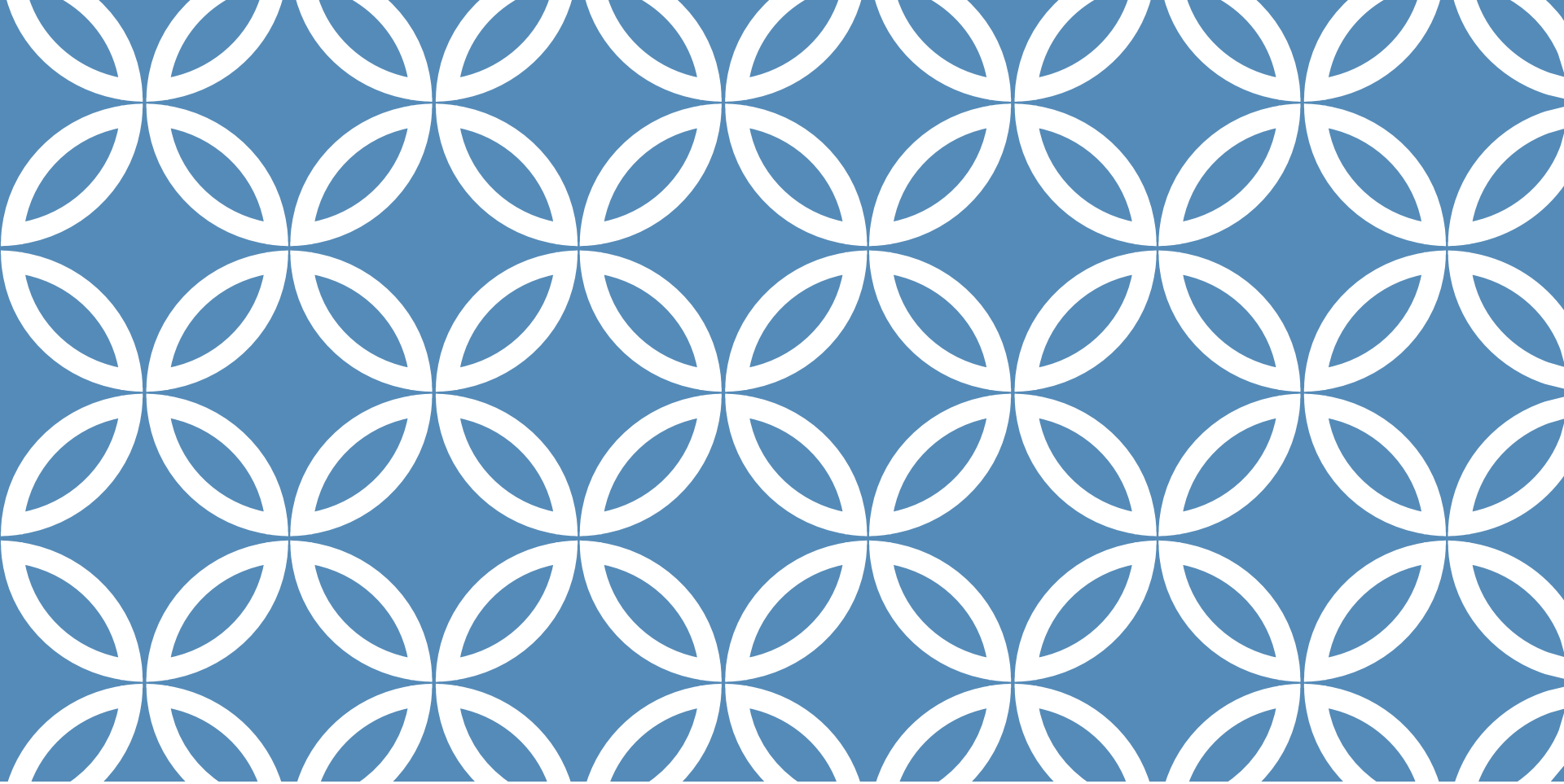
Quantiles		
100.0%	maximum	59
99.5%		56.425
97.5%		43
90.0%		35
75.0%	quartile	31
50.0%	median	27
25.0%	quartile	23
10.0%		18
2.5%		0
0.5%		0
0.0%	minimum	0

♂ 12%

♀ 88%

- Low frequency
- Low recency
- Medium monetary value

- 1020 customers
- Age around 23 to 31
- Average total monetary spending of ~ $89



ASSOCIATION ANALYSIS

Market Basket Analysis

ASSOCIATION ANALYSIS

Item – each item is a category of a service (i.e. nails, makeup etc.) that the customer includes in a booking.

ID – the ID of the transaction that the item belongs to (i.e. booking ID)

Support – proportion of transactions in which an item set appears

Confidence – proportion of transactions that contain the consequent item set, given that the condition item set is in the transaction

Lift – ratio of an association rule's confidence to its expected confidence

ASSOCIATION ANALYSIS

service_count	service_ids.0	service_ids.1
1	570a8572d65405706d6aee56	
2	570a8573d65405706d6af2ed	570a8572d65405706d6aee75
2	570a8572d65405706d6aee4a	570a8572d65405706d6aee4b

booking_id	service_count	label	service_id	category_1
570a8716d65405706d6b7956	1	service_ids.0	570a8574d65405706d6afa92	Nails
570a8716d65405706d6b7964	2	service_ids.0	570a8574d65405706d6afb43	Nails
570a8716d65405706d6b7964	2	service_ids.1	570a8573d65405706d6af480	Nails
570a8716d65405706d6b7965	1	service_ids.0	570a8572d65405706d6aef94	Makeup
570a8716d65405706d6b7966	1	service_ids.0	570a8572d65405706d6aef94	Makeup
570a8716d65405706d6b7967	1	service_ids.0	570a8573d65405706d6af7d8	Makeup
570a8716d65405706d6b7968	1	service_ids.0	570a8573d65405706d6af7d8	Makeup
570a8716d65405706d6b796a	1	service_ids.0	570a8573d65405706d6af7d8	Makeup

ASSOCIATION ANALYSIS

Including bookings with service count = 1

Minimum Support	0.001
Minimum Confidence	0.001
Minimum Lift	0.001
Maximum Antecedents	3
Maximum Rule Size	99

Frequent Item Sets

Item Set	Support	N Items
{Nails}	75%	1
{Makeup}	8%	1
{Brow & Lash}	7%	1
{Hair Styling}	5%	1
{Facial}	4%	1
{Hair Removal}	2%	1
{Brow & Lash, Nails}	1%	2
{Massage}	1%	1
{Hair Removal, Nails}	0%	2
{Brow & Lash, Facial}	0%	2
{Facial, Hair Removal}	0%	2
{Body Therapy}	0%	1
{Brow & Lash, Hair Removal}	0%	2
{Facial, Nails}	0%	2

Rules

Rule			
Condition	Consequent	Confidence	Lift
Hair Removal	Nails	23%	0.303
Brow & Lash	Nails	12%	0.156
Hair Removal	Facial	8%	2.169
Facial	Brow & Lash	6%	0.878
Hair Removal	Brow & Lash	6%	0.788
Facial	Hair Removal	5%	2.169
Brow & Lash	Facial	3%	0.878
Facial	Nails	3%	0.04
Brow & Lash	Hair Removal	2%	0.788
Nails	Brow & Lash	1%	0.156
Nails	Hair Removal	1%	0.303
Nails	Facial	0%	0.04

ASSOCIATION ANALYSIS

Excluding bookings with service count = 1

Minimum Support	0.01
Minimum Confidence	0.01
Minimum Lift	0.01
Maximum Antecedents	3
Maximum Rule Size	99

Frequent Item Sets

Item Set	Support	N Items
{Nails}	92%	1
{Brow & Lash}	7%	1
{Hair Removal}	5%	1
{Brow & Lash, Nails}	3%	2
{Facial}	2%	1
{Hair Removal, Nails}	2%	2
{Makeup}	1%	1

Rules

Rule			
Condition	Consequent	Confidence	Lift
Brow & Lash	Nails	49%	0.533
Hair Removal	Nails	42%	0.453
Nails	Brow & Lash	4%	0.533
Nails	Hair Removal	2%	0.453

ASSOCIATION ANALYSIS

booking_id	service_count	label	service_id	category_1
570a86fed65405706d6b757f	3	service_ids.0	570a8572d65405706d6af11f	Nails
570a86fed65405706d6b757f	3	service_ids.1	570a8572d65405706d6af120	Nails
570a86fed65405706d6b757f	3	service_ids.2	570a8572d65405706d6af124	Nails
570a86ffd65405706d6b7596	1	service_ids.0	570a8572d65405706d6af09b	Nails
570a86ffd65405706d6b759a	1	service_ids.0	570a8573d65405706d6af36d	Nails
570a86ffd65405706d6b75b1	3	service_ids.0	570a8572d65405706d6af071	Hair Styling
570a86ffd65405706d6b75b1	3	service_ids.1	570a8572d65405706d6af073	Hair Styling
570a86ffd65405706d6b75b1	3	service_ids.2	570a8572d65405706d6af06c	Hair Styling

booking_id	service_count	label	service_id	category_1	category_2
570a86fed65405706d6b757f	3	service_ids.0	570a8572d65405706d6af11f	Nails	Gel
570a86fed65405706d6b757f	3	service_ids.1	570a8572d65405706d6af120	Nails	Nail Art
570a86fed65405706d6b757f	3	service_ids.2	570a8572d65405706d6af124	Nails	Nail Art
570a86ffd65405706d6b7596	1	service_ids.0	570a8572d65405706d6af09b	Nails	Gel
570a86ffd65405706d6b759a	1	service_ids.0	570a8573d65405706d6af36d	Nails	Gel
570a86ffd65405706d6b75b1	3	service_ids.0	570a8572d65405706d6af071	Hair Styling	Colour
570a86ffd65405706d6b75b1	3	service_ids.1	570a8572d65405706d6af073	Hair Styling	Colour
570a86ffd65405706d6b75b1	3	service_ids.2	570a8572d65405706d6af06c	Hair Styling	Cut

ASSOCIATION ANALYSIS

Minimum Support	0.01
Minimum Confidence	0.01
Minimum Lift	0.01
Maximum Antecedents	3
Maximum Rule Size	99

Including bookings with service count = 1

Frequent Item Sets

Item Set	Support	N Items
{Classic}	40%	1
{Express}	24%	1
{Nail Art}	14%	1
{Removal}	9%	1
{Classic, Nail Art}	7%	2
{Lash}	6%	1
{Extension}	5%	1
{Nail Spa}	4%	1
{Classic, Removal}	4%	2
{Dinner & Dance}	4%	1
{Cut}	3%	1
{Express, Nail Art}	3%	2
{Express, Removal}	2%	2
{Classic, Express}	2%	2
{Nail Art, Removal}	2%	2
{Waxing}	2%	1
{Brow}	2%	1
{Others}	1%	1
{Classic, Nail Art, Removal}	1%	3
{Extension, Nail Art}	1%	2
{Classic, Nail Spa}	1%	2

Rules

Rule			
Condition	Consequent	Confidence	Lift
Nail Art, Removal	Classic	61%	1.524
Removal	Classic	48%	1.197
Nail Art	Classic	48%	1.195
Classic, Removal	Nail Art	31%	2.242
Nail Spa	Classic	29%	0.716
Extension	Nail Art	25%	1.841
Removal	Express	25%	1.056
Removal	Nail Art	24%	1.761
Nail Art	Express	22%	0.913
Classic, Nail Art	Removal	19%	2.246
Classic	Nail Art	17%	1.195
Nail Art	Removal	15%	1.761
Removal	Classic, Nail Art	15%	2.246
Express	Nail Art	13%	0.913
Classic	Removal	10%	1.197
Nail Art	Classic, Removal	9%	2.242
Express	Removal	9%	1.056
Nail Art	Extension	9%	1.841
Express	Classic	9%	0.224
Classic	Express	5%	0.224
Classic	Nail Art, Removal	3%	1.524
Classic	Nail Spa	3%	0.716

ASSOCIATION ANALYSIS

Excluding bookings with service count = 1

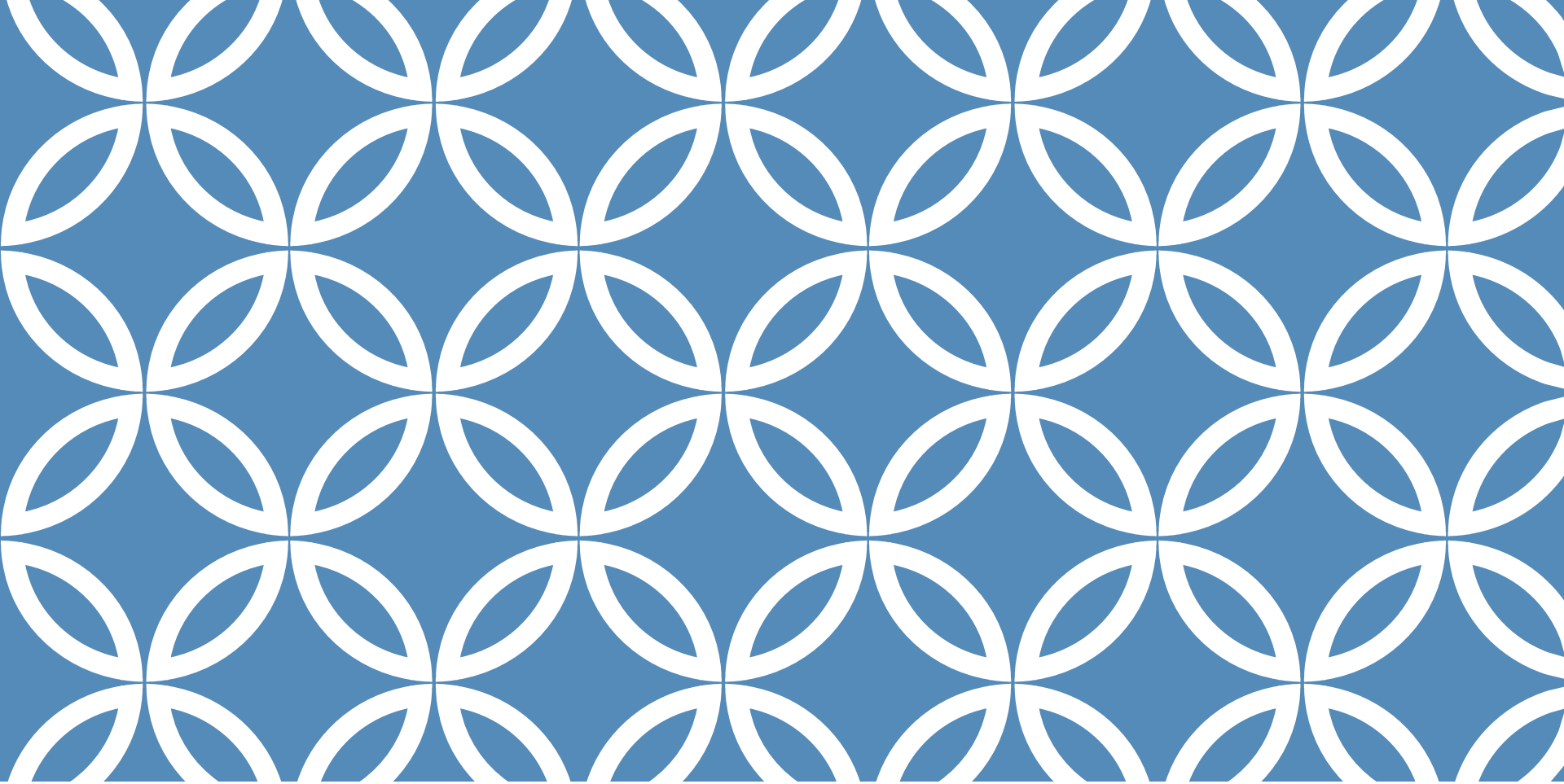
Minimum Support	0.1
Minimum Confidence	0.1
Minimum Lift	0.1
Maximum Antecedents	3
Maximum Rule Size	99

Frequent Item Sets

Item Set	Support	N Items
{Classic}	60%	1
{Nail Art}	42%	1
{Express}	29%	1
{Removal}	26%	1
{Classic, Nail Art}	26%	2
{Classic, Removal}	16%	2
{Express, Nail Art}	12%	2

Rules

Rule			
Condition	Consequent	Confidence	Lift
Removal	Classic	61%	1.002
Nail Art	Classic	60%	0.998
Classic	Nail Art	42%	0.998
Express	Nail Art	40%	0.934
Nail Art	Express	27%	0.934
Classic	Removal	26%	1.002



RECOMMENDATIONS

RECOMMENDATIONS

Acquire	Enhance	Retain
• Who are the customers • Exploratory Data Analysis • Cluster Analysis	• Enhance profitability of existing customers • Cross-selling • Association Analysis	

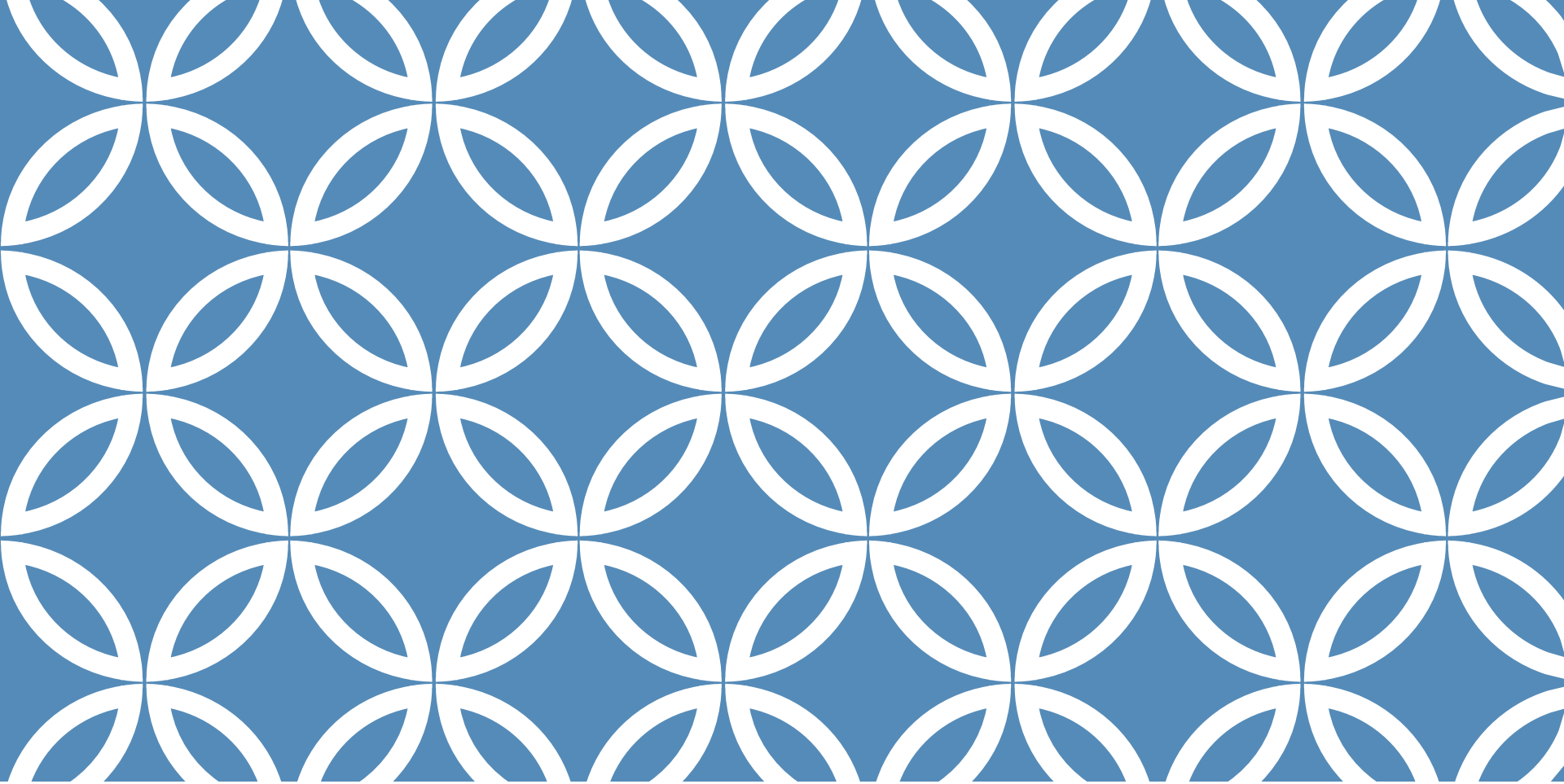
RECOMMENDATIONS

1. Targeted marketing at different groups of customers

2. Keeping track of RFM attributes for each customer

3. Have campaigns that are tailored to the association between 2 or more services in a booking

    e.g. 20% off classic service if there is removal service



THANK YOU