Analytics Practicum – Team V

Interim Presentation

23rd Feb 2017

Agenda

- Project Overview
- Data Collection
- Data Cleaning
- Exploration Data Analysis
- Methodologies
- Scope of Work
- Work Plan

Sponsor Background

Vanitee

- Officially launched in May 2015
- Bridge the gap between customers & beauty professionals

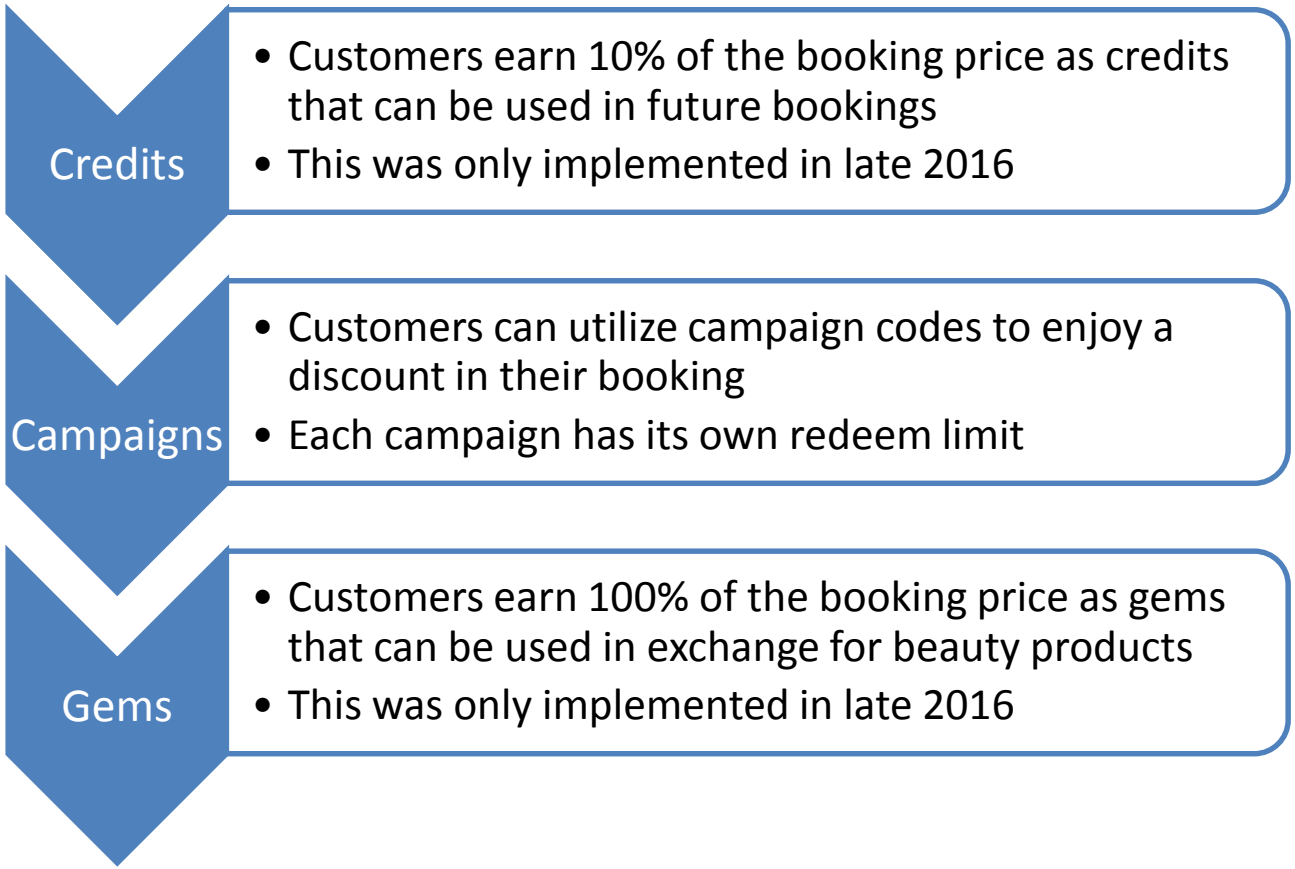
- Customers can
  - Book beauty services at their own convenience
- Beauty professionals can
  - Grow their brand & customer base
  - Showcase what they do best

Project Motivation

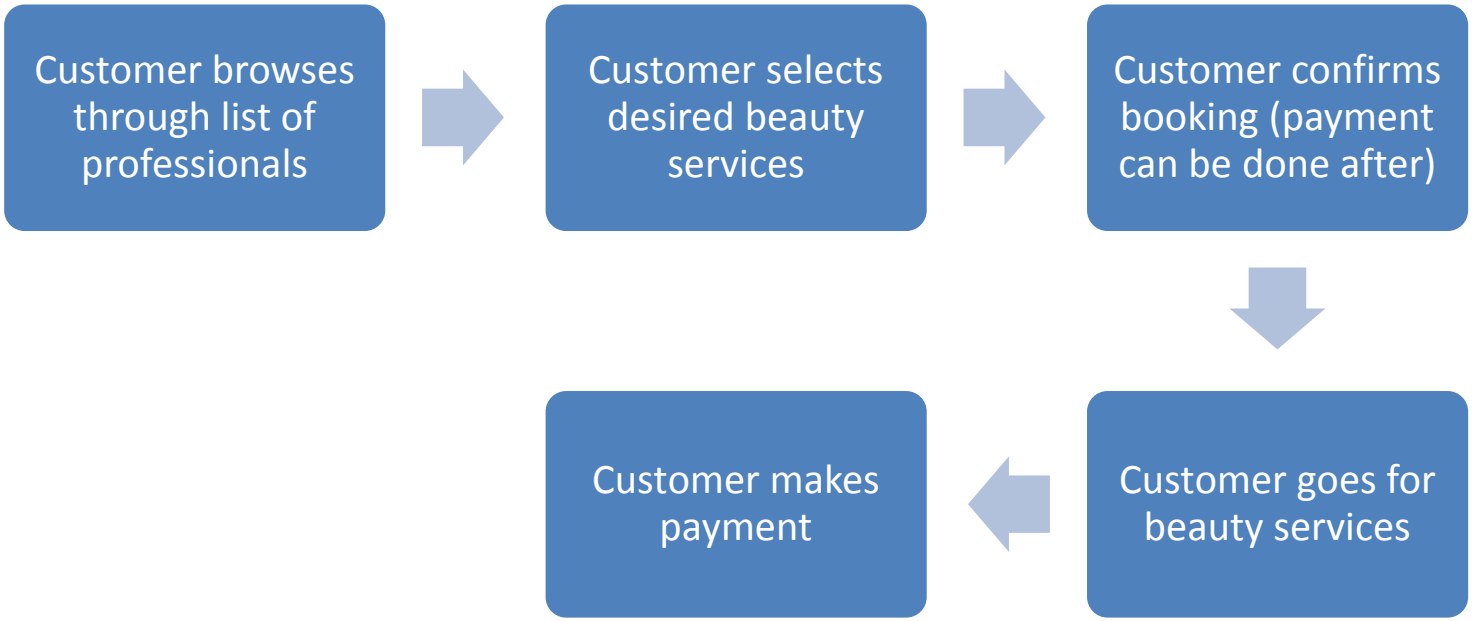
- To compete with brick-and-mortar shops, it is vital to retain existing customers

- To evaluate the effectiveness of their current loyalty program

Current Loyalty Program

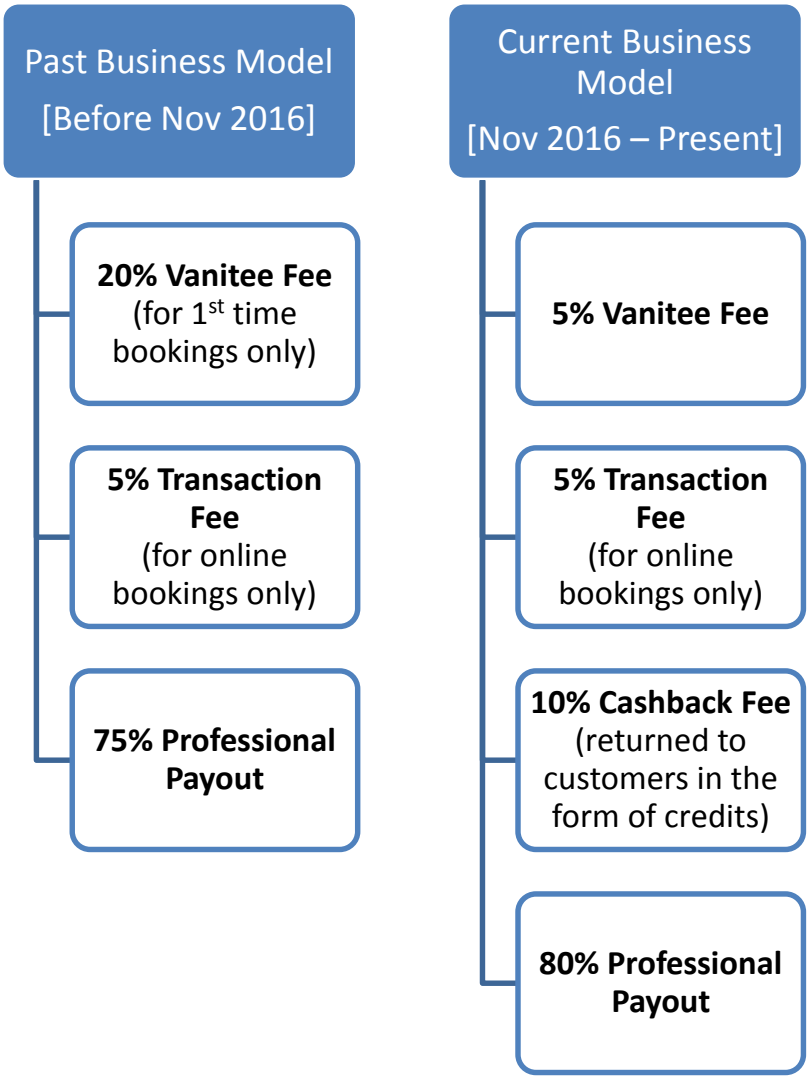


Current Process Flow



- Credits can be used here to reduce the price of booking
- Campaign codes can be used here to reduce the price of booking
- Credits (10%) & Gems (100%) are earned here

Business Model



Project Objectives

Customers

- To determine the customer segmentation from current booking patterns
  - Which customers are stagnant vs active?
- To understand customers' behavior
  - RFM analysis
- To evaluate the effectiveness of using campaign codes to ensure repeat bookings
- To understand how customers use credits

Project Objectives

Customers

- To determine the Customer Lifetime Value by campaign
  - Which campaigns do customers react to more
- To determine the type of services that generate the most revenue

Project Objectives

Professionals

- To determine if there is any correlation on what makes beauty professionals more attractive to customers
  - Higher chat response rate
  - Greater variety of services
  - Less expensive services compared to other professionals

Data Collection

- Access to Vanitee's MongoDB database
- Data range used from Jan 2015 to Dec 2016

- Narrowed down from 59 to 7 tables
  - Bookings
  - Campaigns
  - Categories
  - Users
  - Customers
  - Professionals
  - Services

New Columns – Bookings

New Columns Created	Description
created_at_sgt	The date & time that the booking was created at in SGT
created_at_year	The year that the booking was created
created_at_month	The month that the booking was created
created_at_day	The day that the booking was created
is_master	Indicates if the booking is a master booking or not
is_online	Indicates if the booking is an online or manual booking
service_count	The number of services selected in the booking
vanitee_fee	The profit earned by Vanitee (based on the new formula mentioned above)
use_campaign	Indicates if the booking utilized any campaign codes
category_name	The category that the booking falls under

Data Collection – Challenges

- Filtering out test data
  - Narrowed the initial range from Aug 2014 – Dec 2016 to Jan 2015 – Dec 2016
  - Removed by specific columns (test_at, is_test, deleted_at)
  - Removed across tables
    - Removed admin users → removed bookings made by them
    - Removed non-published professionals → removed bookings that involved them

Data Collection – Challenges

- Complex table structure
  - Tables with column master_id → created a new column, is_master to indicate if the record is a master record
  - Non-master records are excluded from analysis & formulation of graphs

Data Cleaning – Issues

- Duplicate values

Vanitee X Kelynn Nai	vaniteekelynn vtkelynn2016 vtkelynn	0/1	0/1	0	04 May 2016 6:17:48 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:47 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:47 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:46 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:47 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:46 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:46 PM

Data Cleaning – Issues

- Duplicate values
  - Assumed that campaigns with campaign codes are true campaigns
  - Used is_published column to filter out active campaigns

Data Cleaning – Issues

- Missing values
  - Missing monetary values in Bookings data
    - final_price
    - total_price
    - discount_amount
  - Missing start_at values in Campaigns data
    - Missing start_at → use created_at
  - Missing price_vanitee_transaction_fee values in Bookings data

Data Cleaning – Issues

- Changes in business model
  - Hard to calculate profit per booking

$$profit = final\ price - payout - transaction\ fee - cashback$$

Booking – No discount

Customer payment

Payment Method	Credit Card
Total Price	$70.00
Payment Transaction Id	5xes6gzg
Waived	$0.00
Discount Amount	$0.00
Professional Discount	$0.00
Vanitee Discount	$0.00
Credits Used	$0.00
Amount Charged	**$70.00**

Payout Status	Paid (585b85c(
Payout Amount	$56.00
Payout Transaction Id	SGGP161
Payout Created	2016-12-2
Currency	SGD
Chargeable Fee (to professional)	$70.00
Vanitee Fee	$3.50
Transaction Fee	$3.50
Customer Cashback	$7.00

- Profit = $70 - $56 - $3.5 - $7 = $3.5

Booking – Prof discount

Customer payment

Payment Method	Credit Card
Total Price	$45.00
Payment Transaction Id	9d2989ge
Waived	$0.00
Discount Amount	$20.00
Professional Discount	$20.00
Vanitee Discount	$0.00
Credits Used	$0.00
Amount Charged	**$25.00**

Payout Status	Paid (588075
Payout Amount	$20.00
Payout Transaction Id	SGGP17
Payout Created	2017-01
Currency	SGD
Chargeable Fee (to professional)	$25.00
Vanitee Fee	$1.25
Transaction Fee	$1.25
Customer Cashback	$2.50

- Profit = $25 - $20 - $1.25 - $2.5 = $1.25

Booking – Vanitee discount

Customer payment

Payment Method	Credit Card
Total Price	$60.00
Payment Transaction Id	hzx94gqd
Waived	$0.00
Discount Amount	$10.00
Professional Discount	$0.00
Vanitee Discount	$10.00
Credits Used	$0.00
Amount Charged	**$50.00**

Payout Status	Paid (586b565a547a
Payout Amount	$48.00
Payout Transaction Id	SGGP17010301
Payout Created	2017-01-03 15:4
Currency	SGD
Chargeable Fee (to professional)	$60.00
Vanitee Fee	$3.00
Transaction Fee	$3.00
Customer Cashback	$6.00

- Profit = $50 - $48 - $3 - $6 = -$7

Exploratory Data Analysis

- Data
  - 1 Jan 2015 to 31 Dec 2016
  - Removed testing data (is_test, deleted_at etc.)
- Data tables
  - Bookings
  - Users, Customers, Professionals
  - Services, Categories
  - Campaigns
- Bookings
  - Master + online + checkout

EDA – Booking

- Bookings breakdown by type
- Bookings breakdown by status
- Bookings breakdown by frequency
- Bookings breakdown by year
- Bookings breakdown by month
- Bookings breakdown by month & year
- Bookings breakdown by day
- Bookings breakdown by recency (w.r.t. 31 Dec 2016)
- Bookings breakdown by monetary amount
- Bookings breakdown by duration from sign up to 1st online booking
- Bookings breakdown by service count
- Bookings breakdown by category
- Bookings breakdown by campaign usage
- Bookings breakdown by credit usage

EDA – Customer & Professional

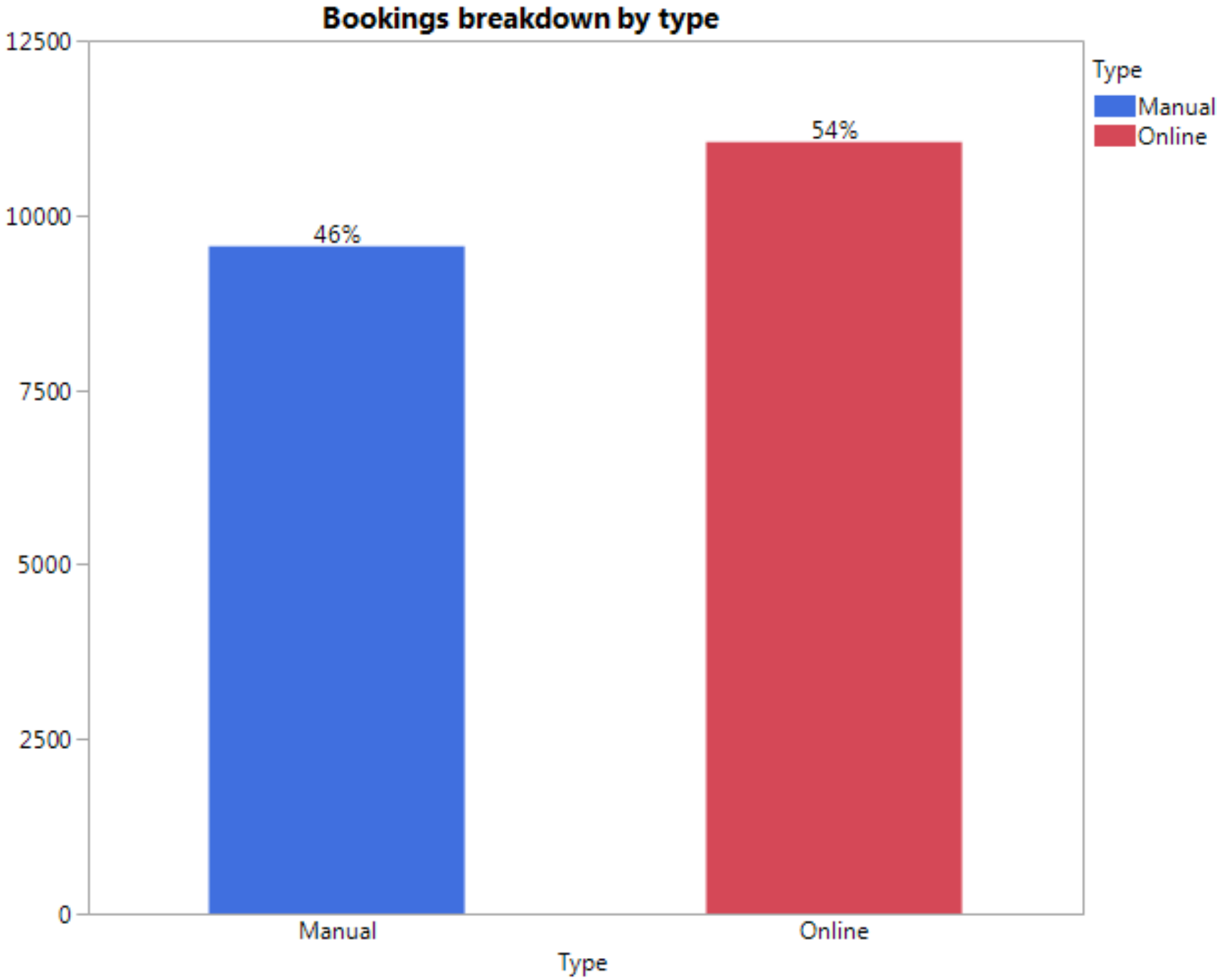
- Users breakdown by type
- Customers breakdown by age
- Customers breakdown by gender
- Professionals breakdown by age
- Professionals breakdown by gender

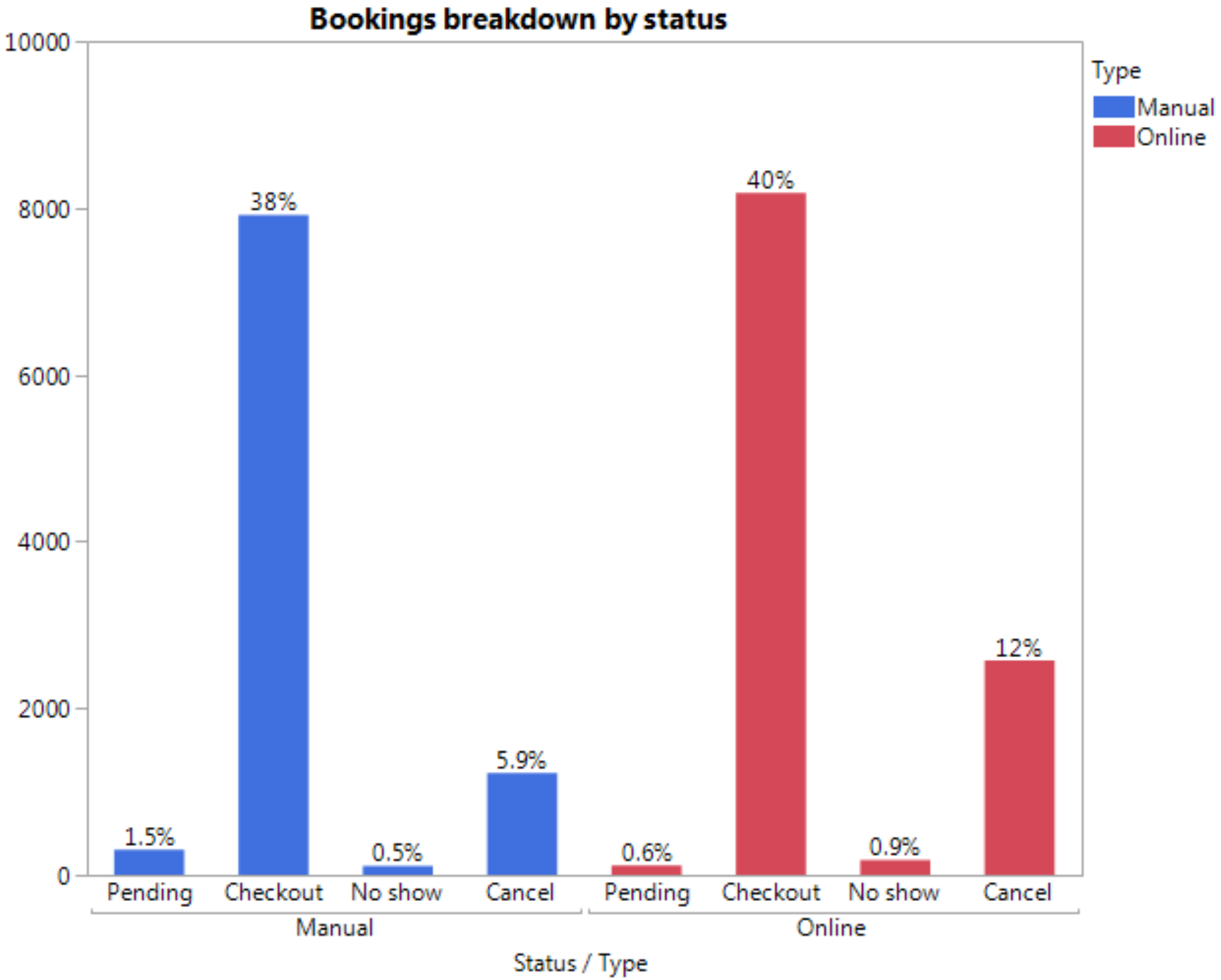
EDA – Service

- Services breakdown by price
- Services breakdown by professional
- Services breakdown by category

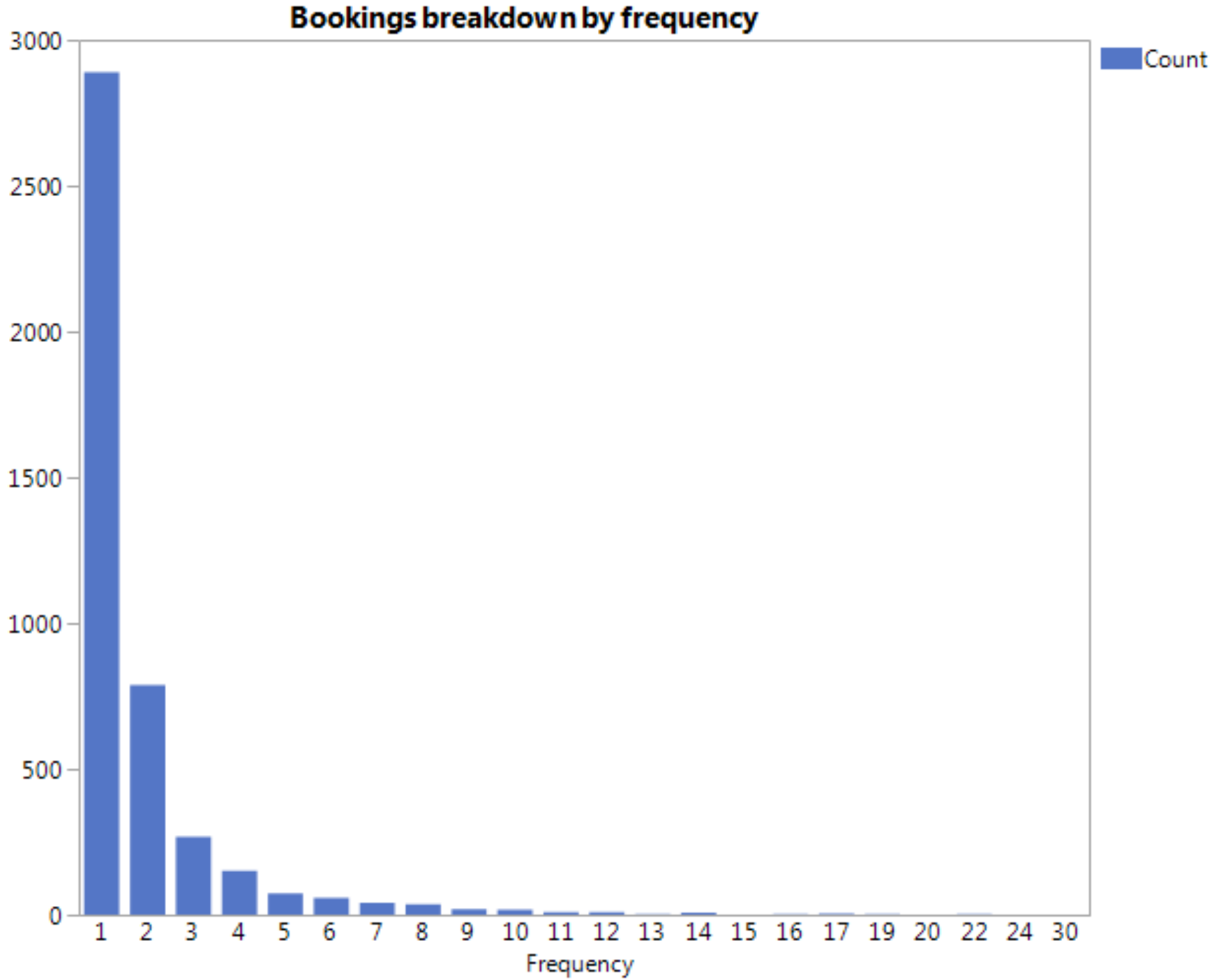
EDA – Campaign

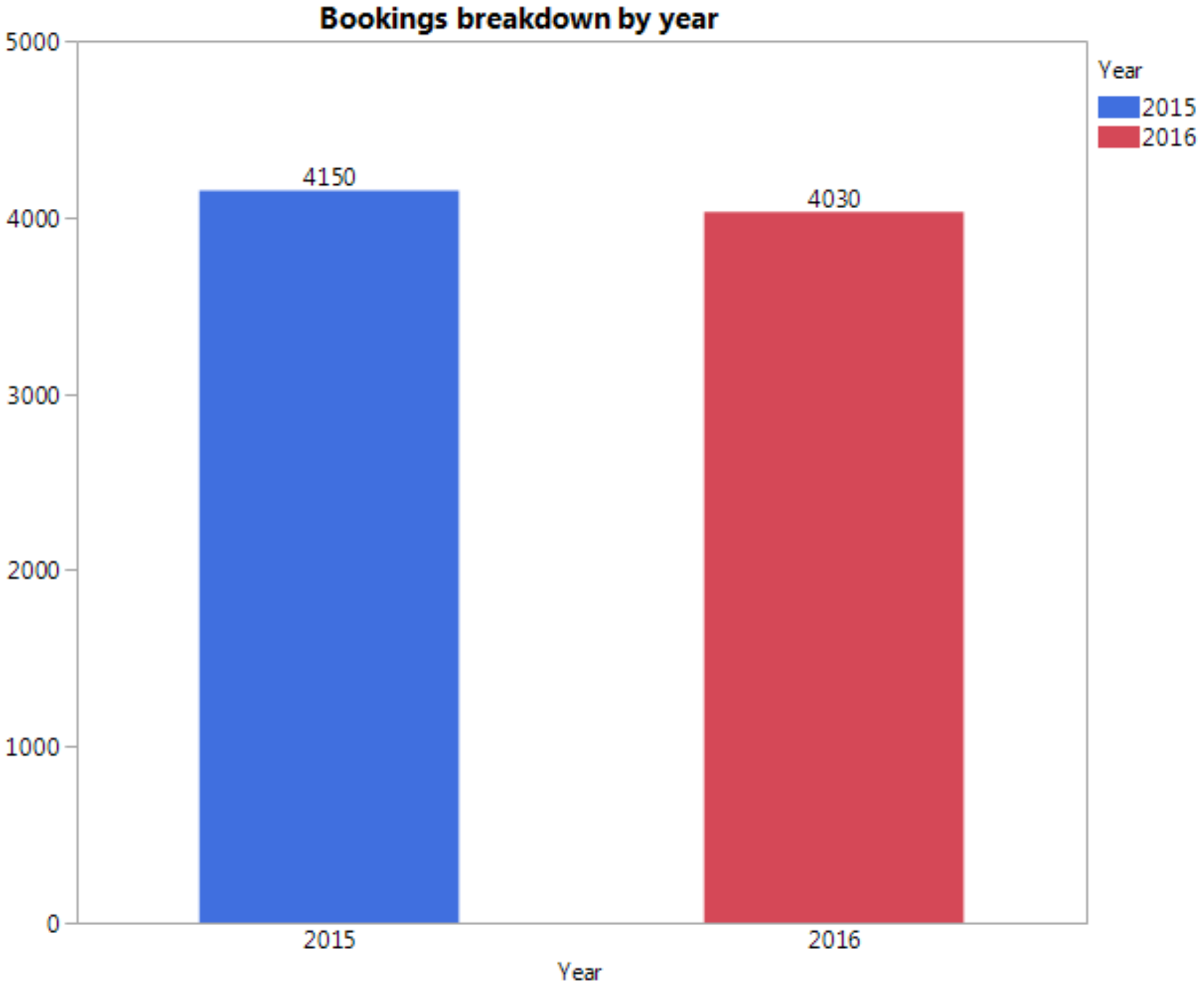
- Campaigns breakdown by duration
- Campaigns breakdown by type
- Campaigns breakdown by discount type
- Campaigns breakdown by discount amount
- Campaigns breakdown by usage
- Campaigns breakdown by duration from start to 1st usage

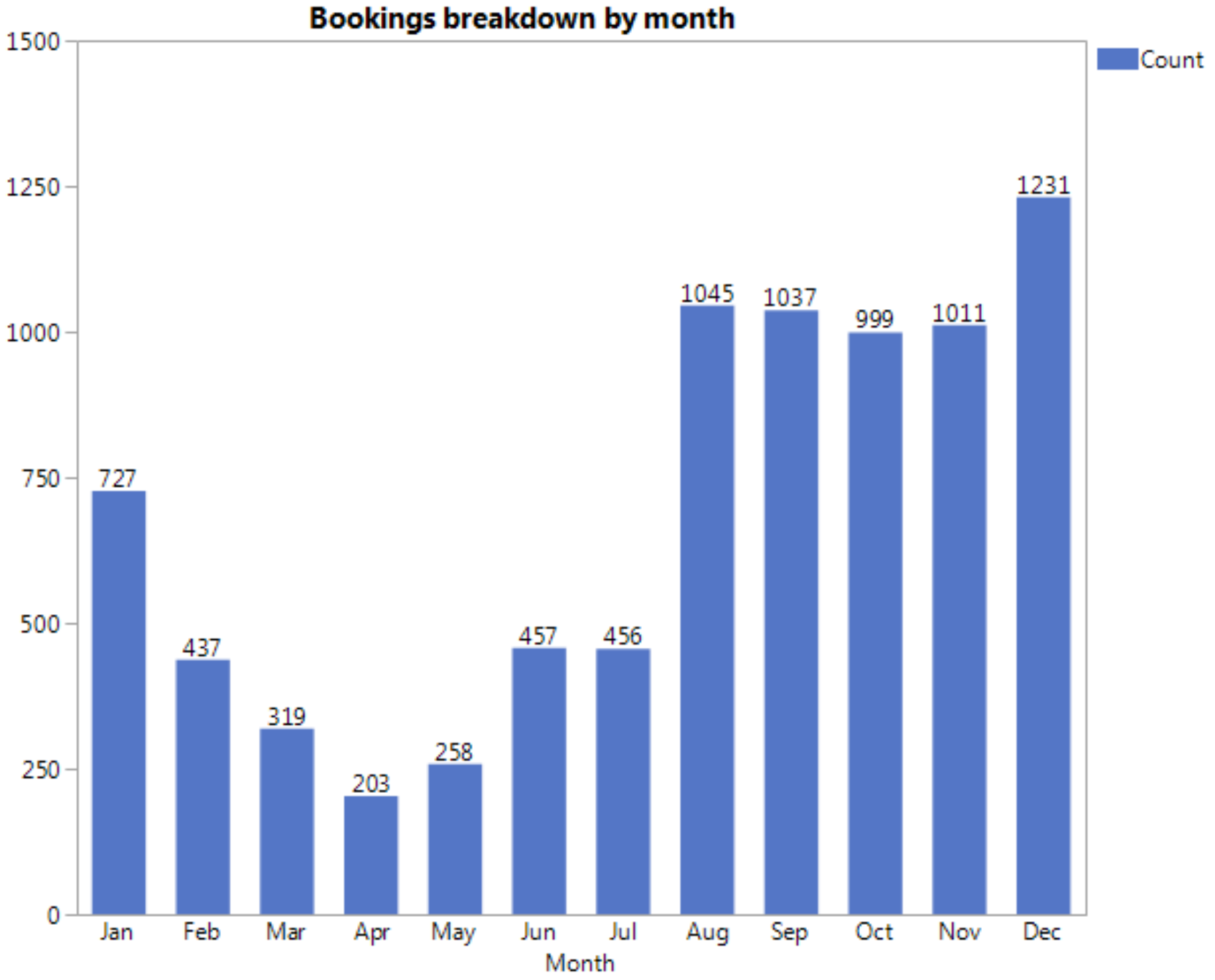

Bookings breakdown by type
12500
10000
7500
5000
2500
0
46%
54%
Manual
Online
Type
Type
Manual
Online

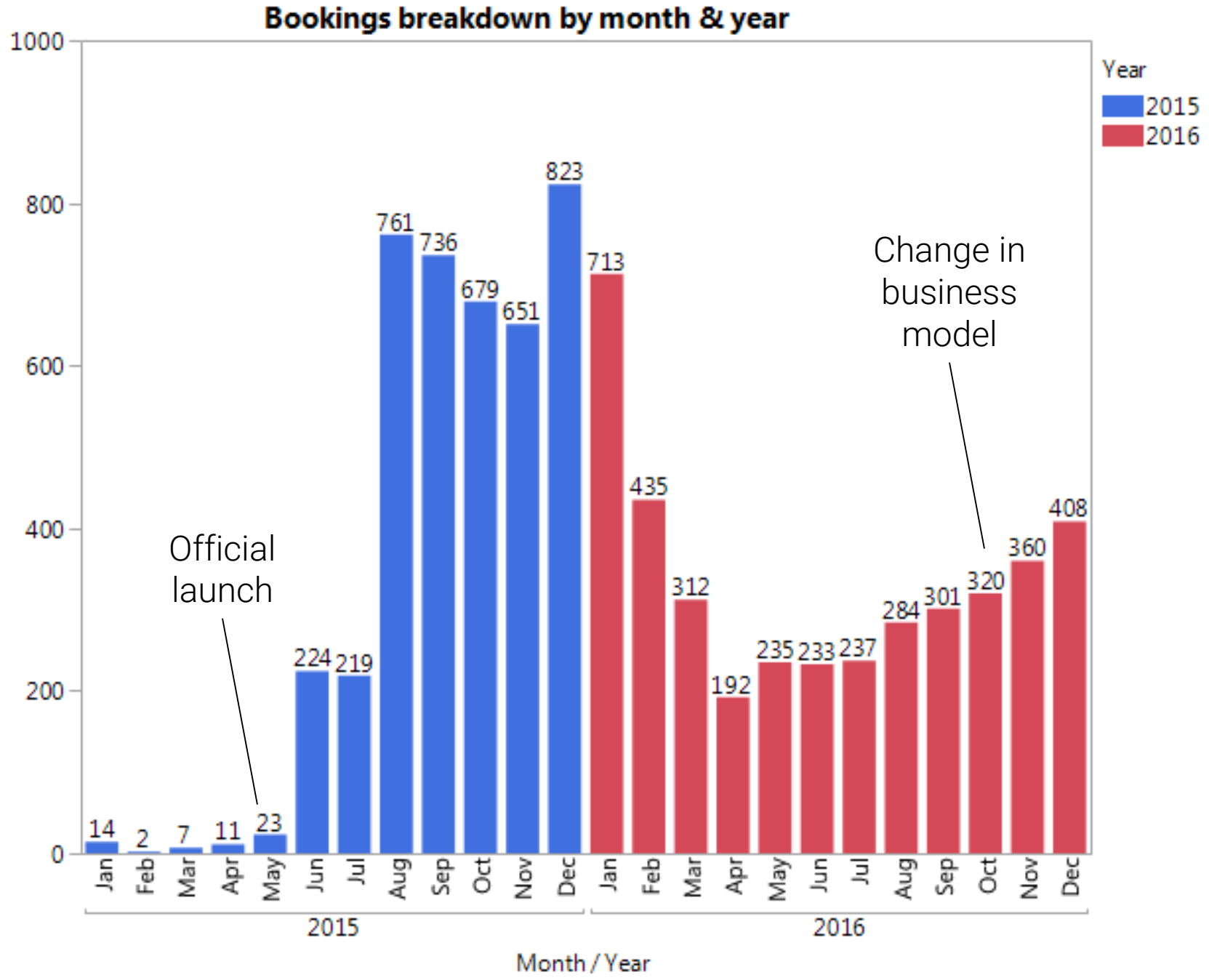

Bookings breakdown by status
10000
8000
6000
4000
2000
0
Type
Manual
Online
38%
40%
12%
5.9%
1.5%
0.5%
0.6%
0.9%
Pending
Checkout
No show
Cancel
Pending
Checkout
No show
Cancel
Manual
Online
Status / Type

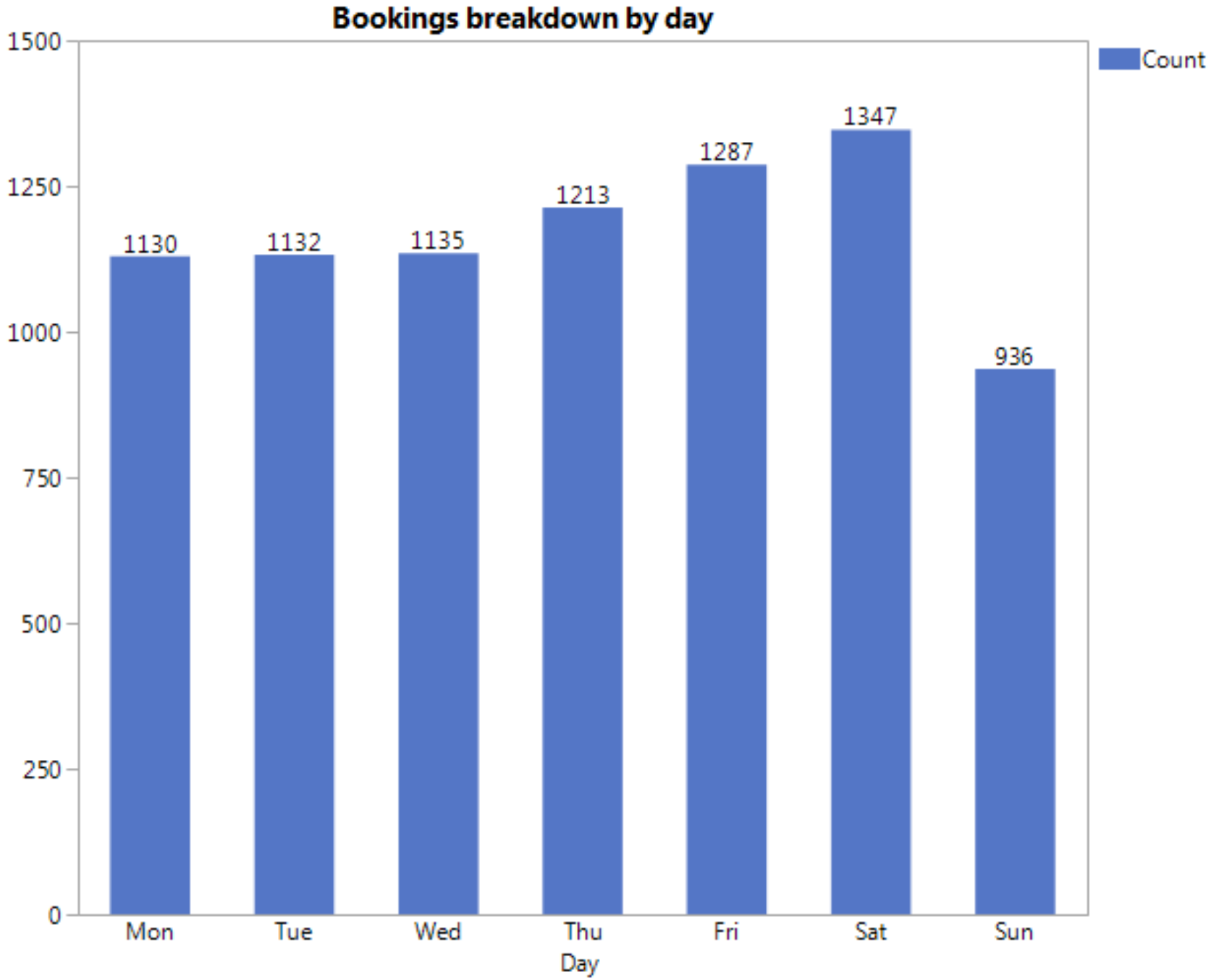
Online bookings (check out)

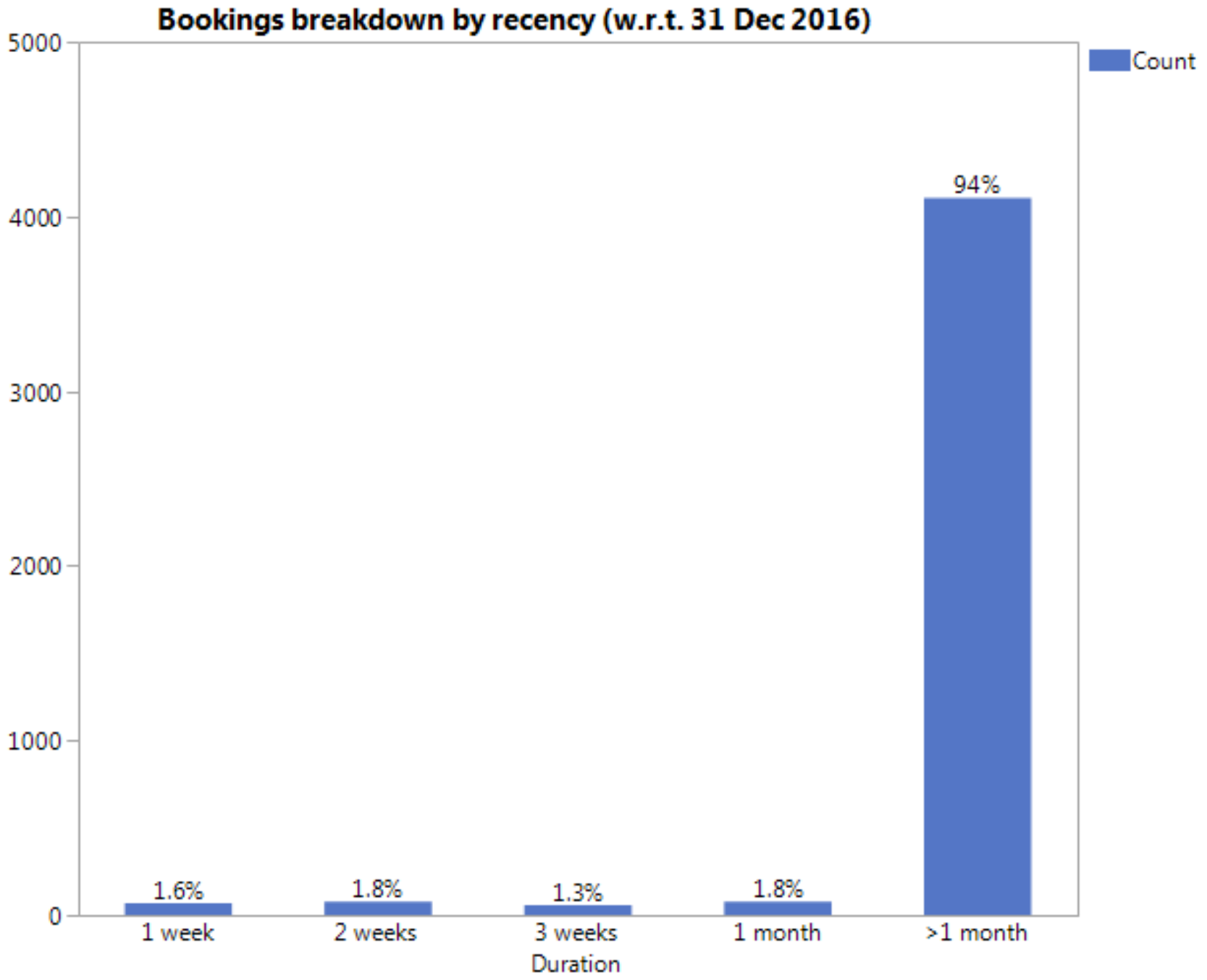

Bookings breakdown by frequency
Count
3000
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
20
22
24
30
Frequency

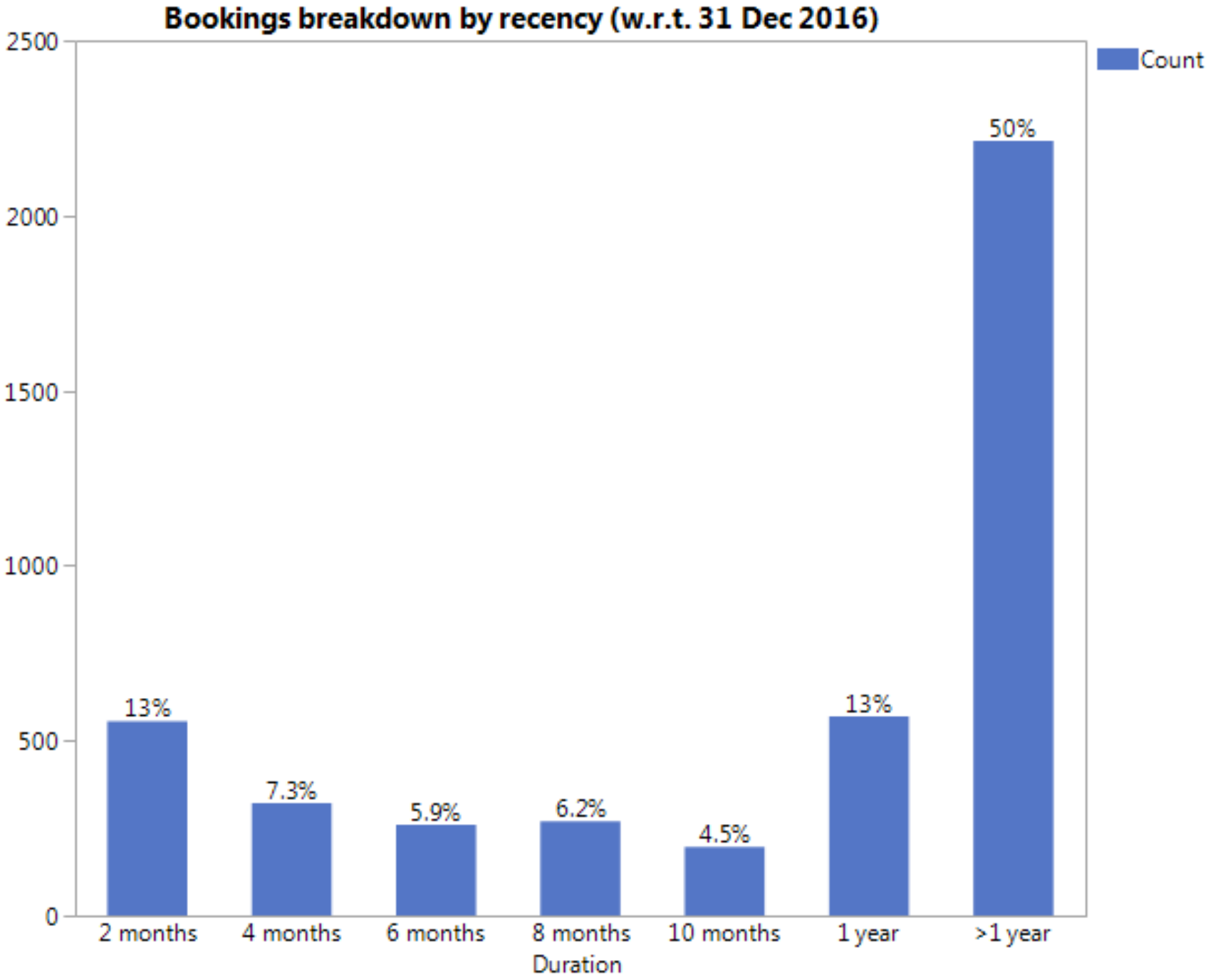

Bookings breakdown by year
5000
4000
3000
2000
1000
0
4150
4030
2015
2016
Year
Year
2015
2016

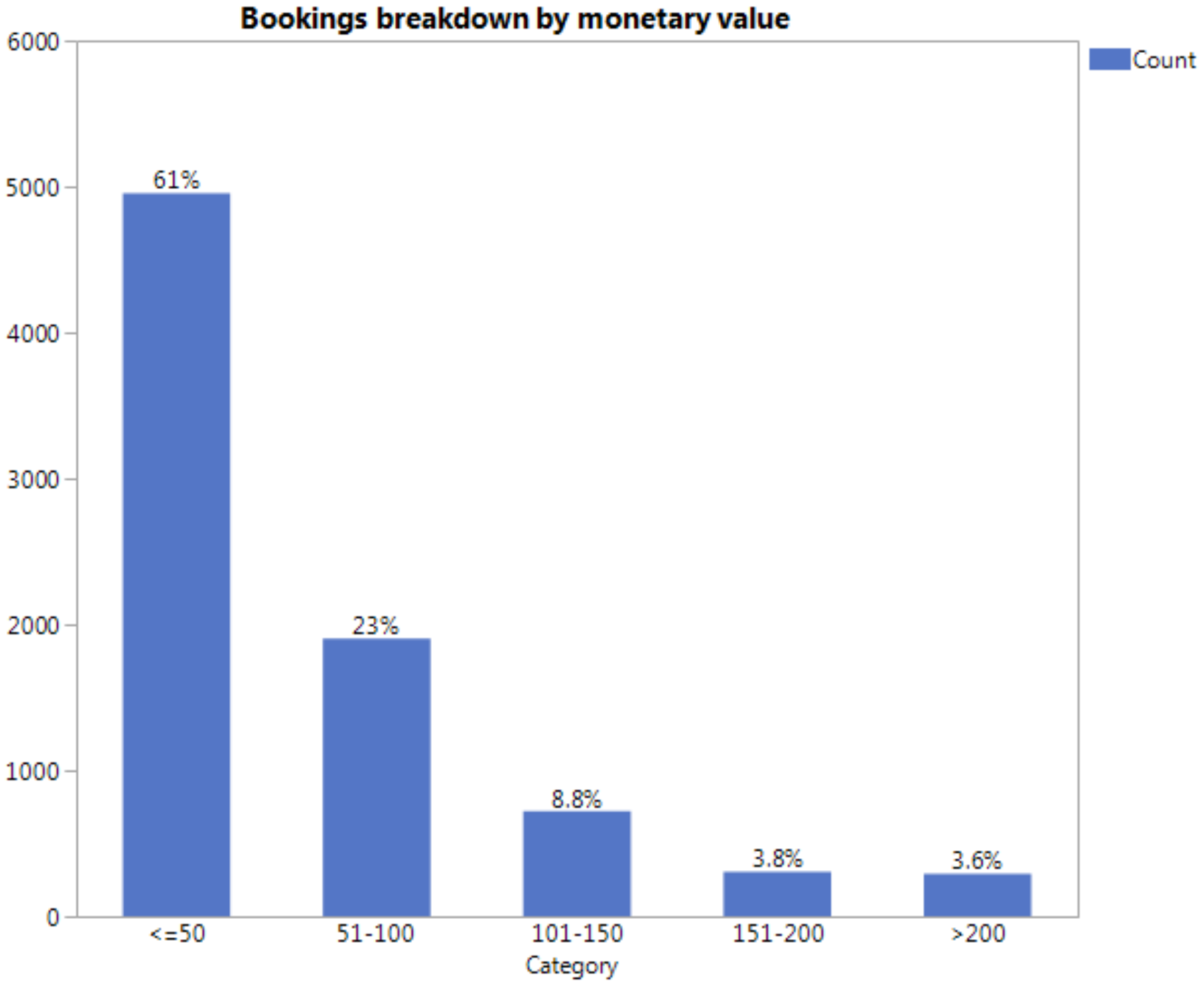

Bookings breakdown by month
1500
1250
1000
750
500
250
0
Count
727
437
319
203
258
457
456
1045
1037
999
1011
1231
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

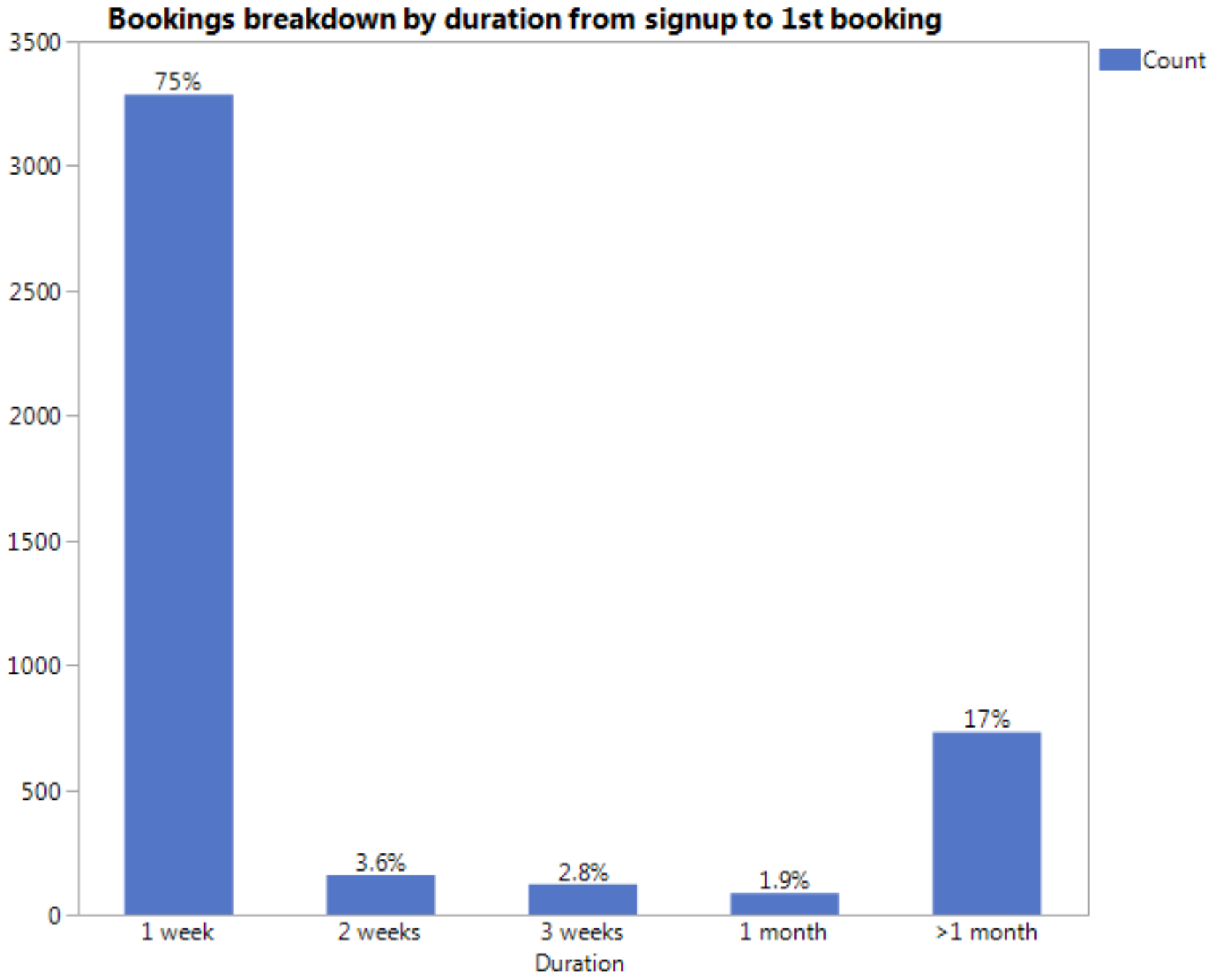

Bookings breakdown by month & year
1000
800
600
400
200
0
Year
2015
2016
Official launch
Change in business model
14
2
7
11
23
224
219
761
736
679
651
823
713
435
312
192
235
233
237
284
301
320
360
408
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2015
2016
Month / Year

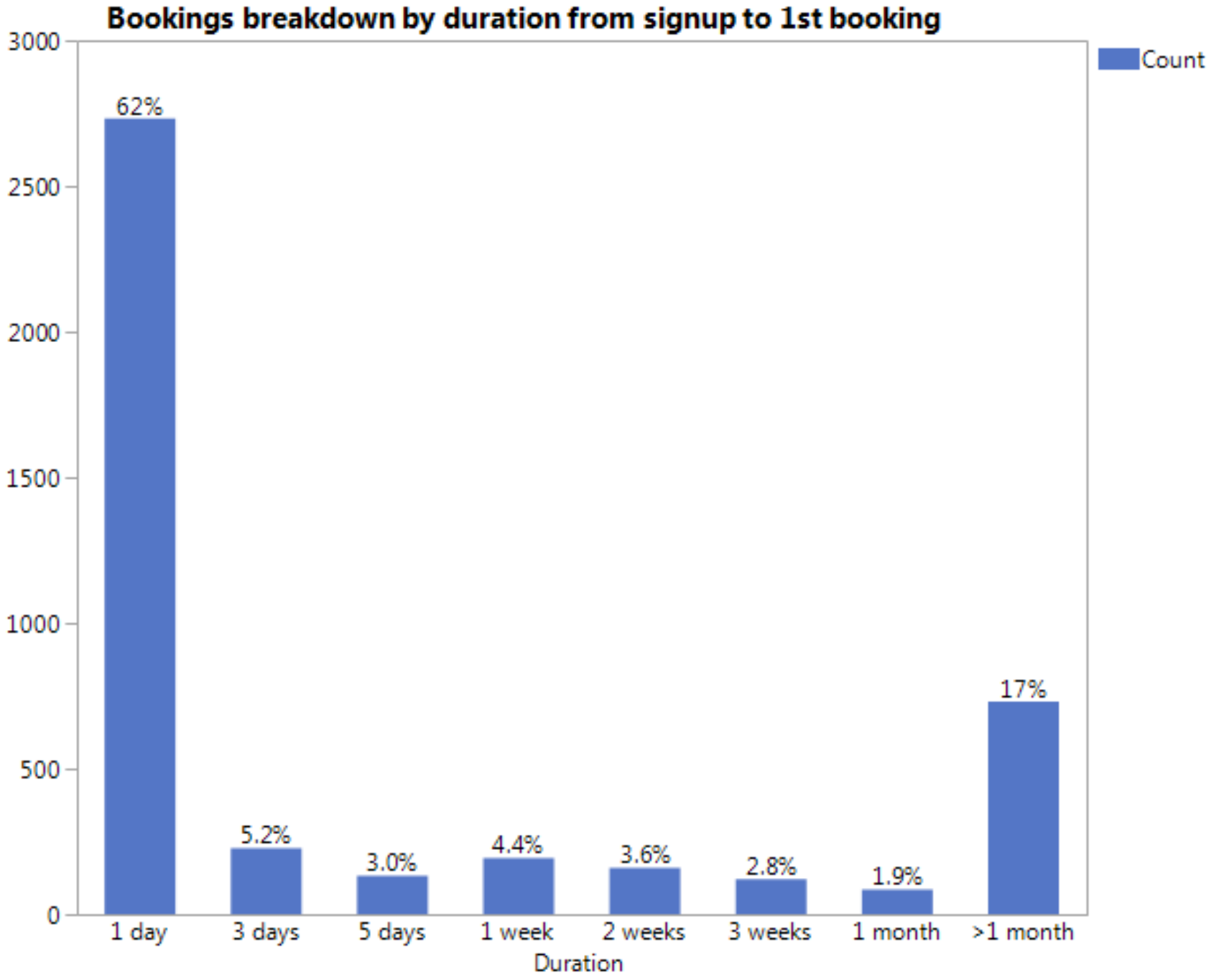

Bookings breakdown by day
1500
1250
1000
750
500
250
0
1130
1132
1135
1213
1287
1347
936
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day
Count

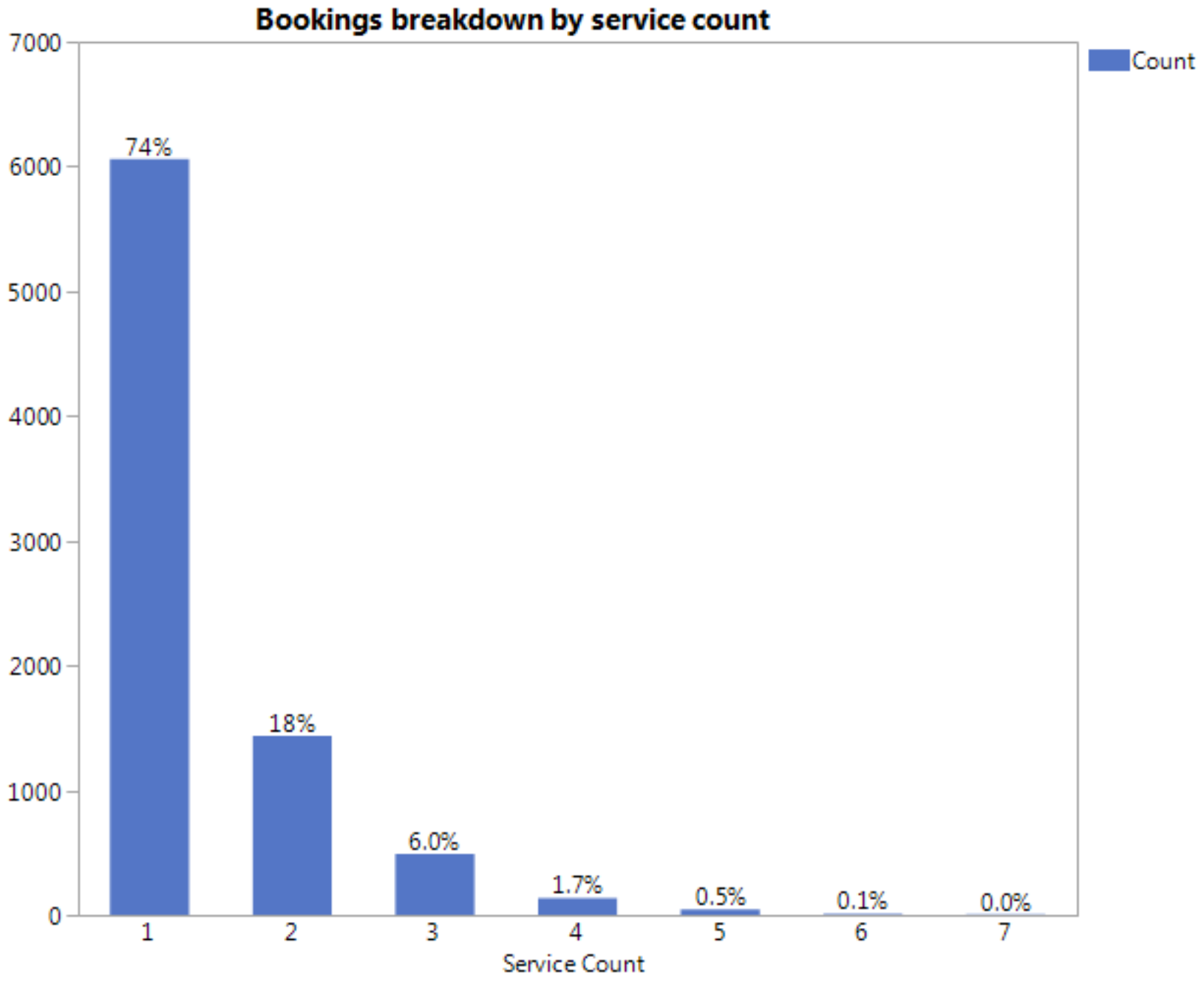

Bookings breakdown by recency (w.r.t. 31 Dec 2016)
Count
5000
4000
3000
2000
1000
0
94%
1.6%
1.8%
1.3%
1.8%
1 week
2 weeks
3 weeks
1 month
>1 month
Duration

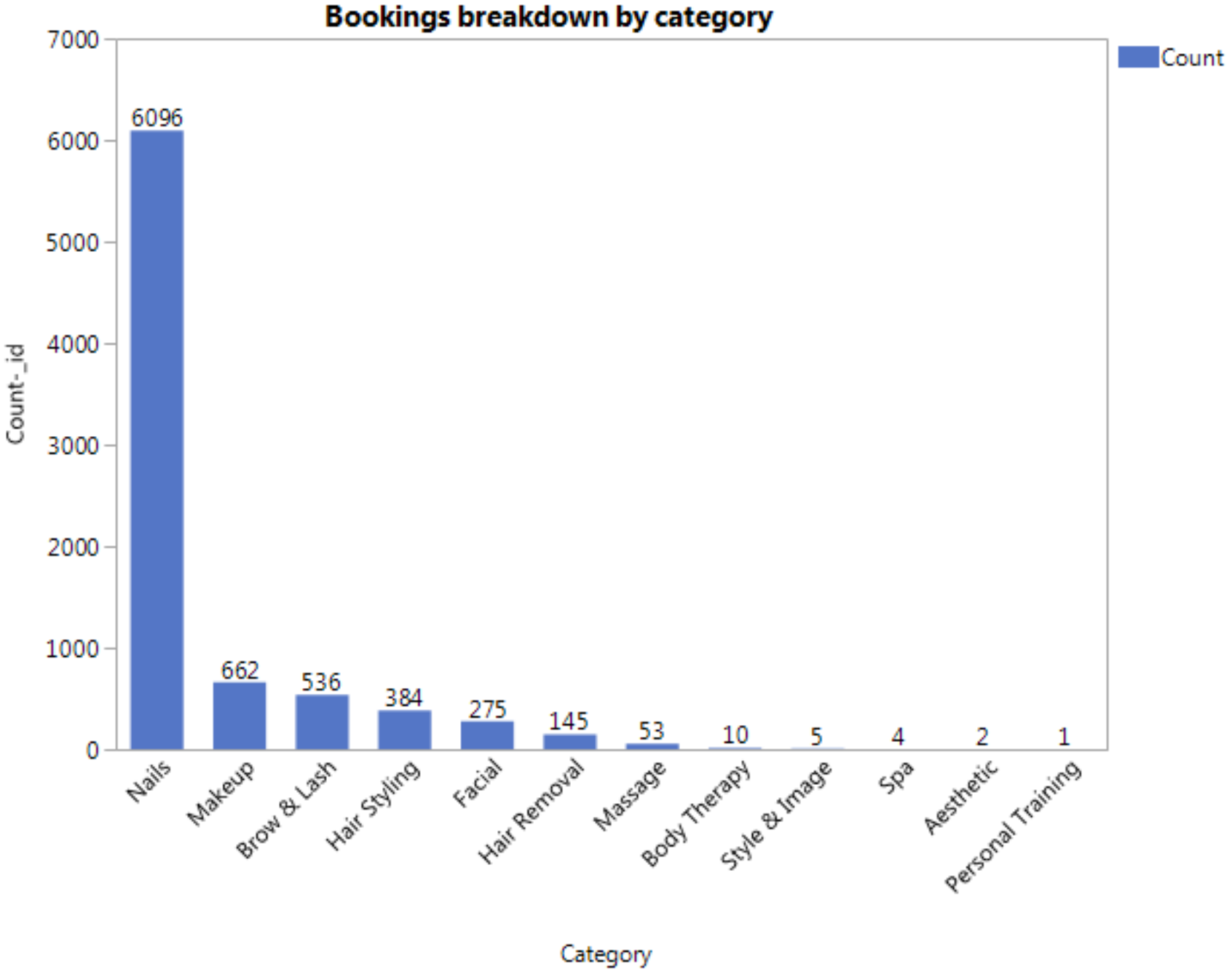

Bookings breakdown by recency (w.r.t. 31 Dec 2016)
Count
2500
2000
1500
1000
500
0
13%
7.3%
5.9%
6.2%
4.5%
13%
50%
2 months
4 months
6 months
8 months
10 months
1 year
>1 year
Duration

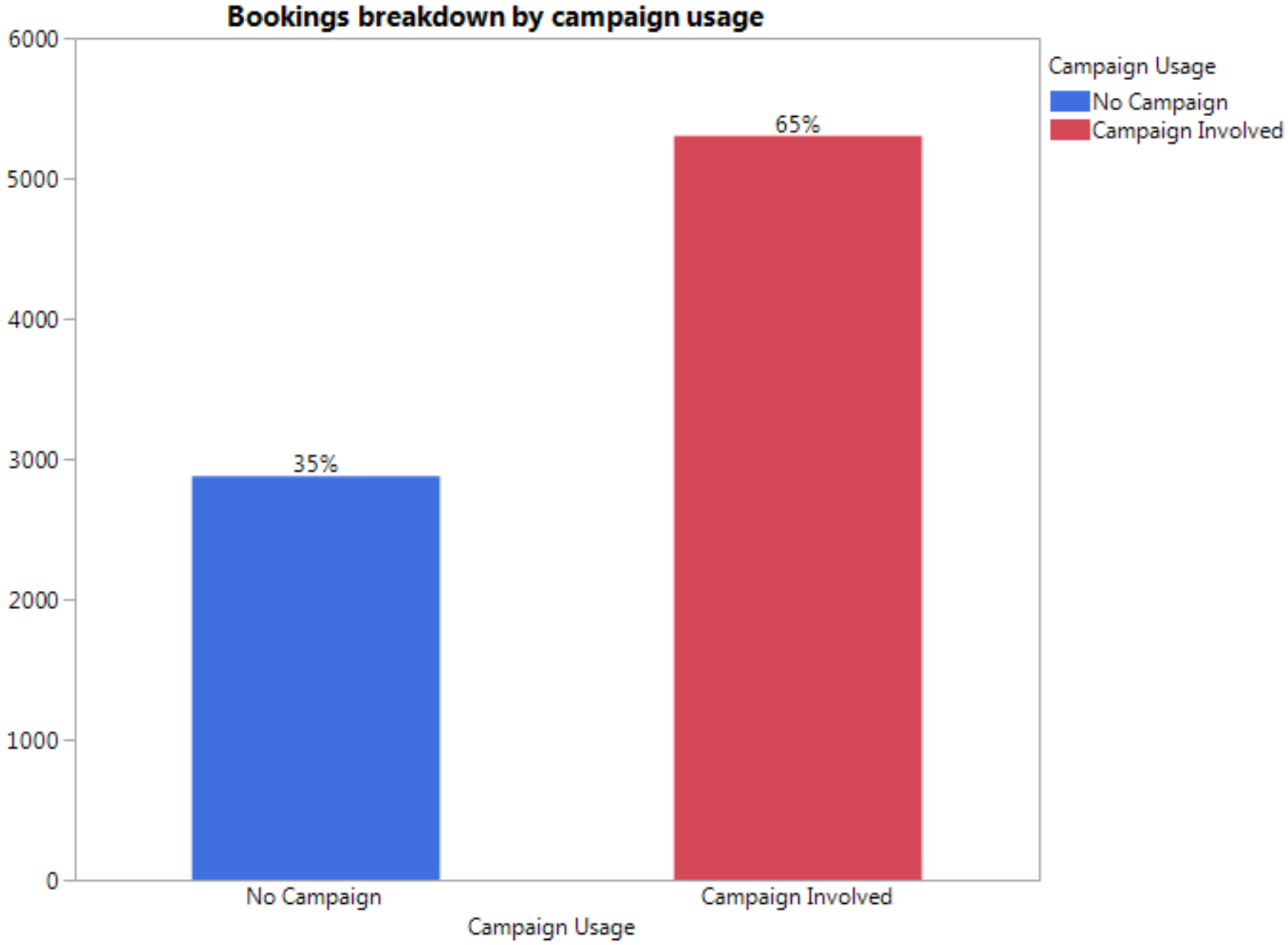

Bookings breakdown by monetary value
Count
6000
5000
4000
3000
2000
1000
0
61%
23%
8.8%
3.8%
3.6%
<=50
51-100
101-150
151-200
>200
Category

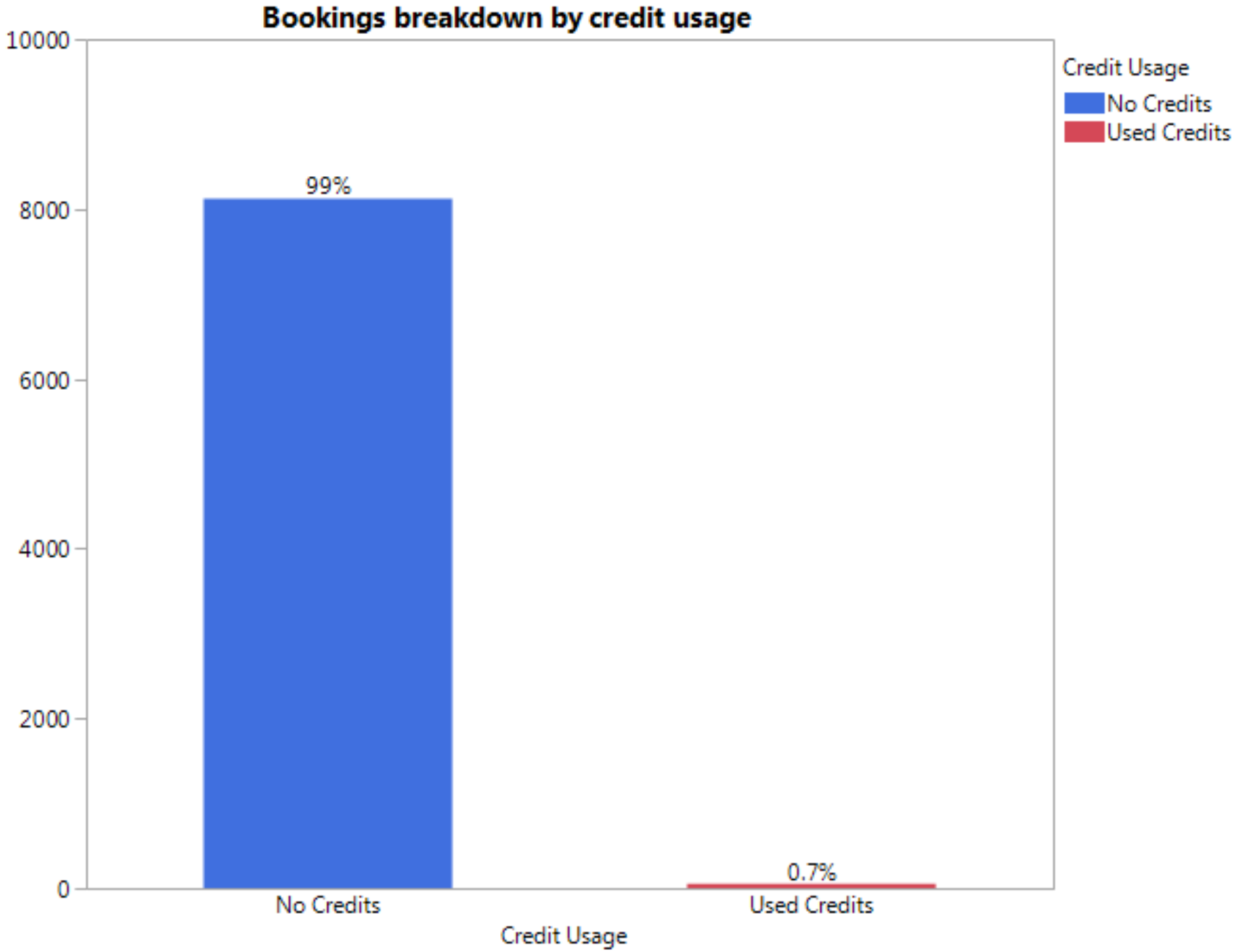

Bookings breakdown by duration from signup to 1st booking
Count
3500
3000
2500
2000
1500
1000
500
0
75%
3.6%
2.8%
1.9%
17%
1 week
2 weeks
3 weeks
1 month
>1 month
Duration


Bookings breakdown by duration from signup to 1st booking
Count
3000
2500
2000
1500
1000
500
0
62%
5.2%
3.0%
4.4%
3.6%
2.8%
1.9%
17%
1 day
3 days
5 days
1 week
2 weeks
3 weeks
1 month
>1 month
Duration


Bookings breakdown by service count
Count
7000
6000
5000
4000
3000
2000
1000
0
74%
18%
6.0%
1.7%
0.5%
0.1%
0.0%
1
2
3
4
5
6
7
Service Count


Bookings breakdown by category
Count
Count-_id
7000
6000
5000
4000
3000
2000
1000
0
6096
662
536
384
275
145
53
10
5
4
2
1
Nails
Makeup
Brow & Lash
Hair Styling
Facial
Hair Removal
Massage
Body Therapy
Style & Image
Spa
Aesthetic
Personal Training
Category


Bookings breakdown by campaign usage
6000
5000
4000
3000
2000
1000
0
35%
65%
No Campaign
Campaign Involved
Campaign Usage
Campaign Usage
No Campaign
Campaign Involved


Bookings breakdown by credit usage
10000
8000
6000
4000
2000
0
99%
0.7%
No Credits
Used Credits
Credit Usage
Credit Usage
No Credits
Used Credits

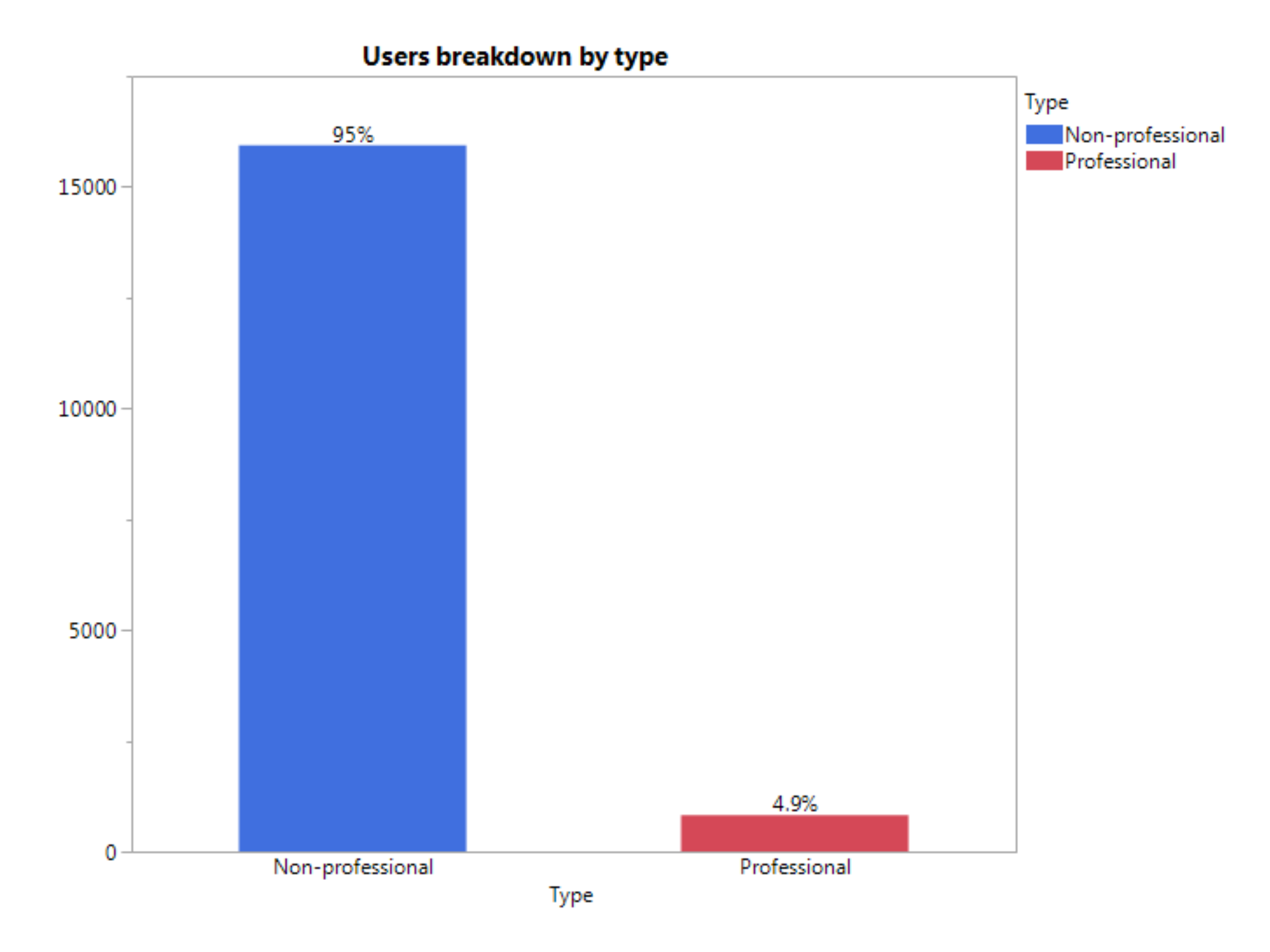
Bookings - Profit

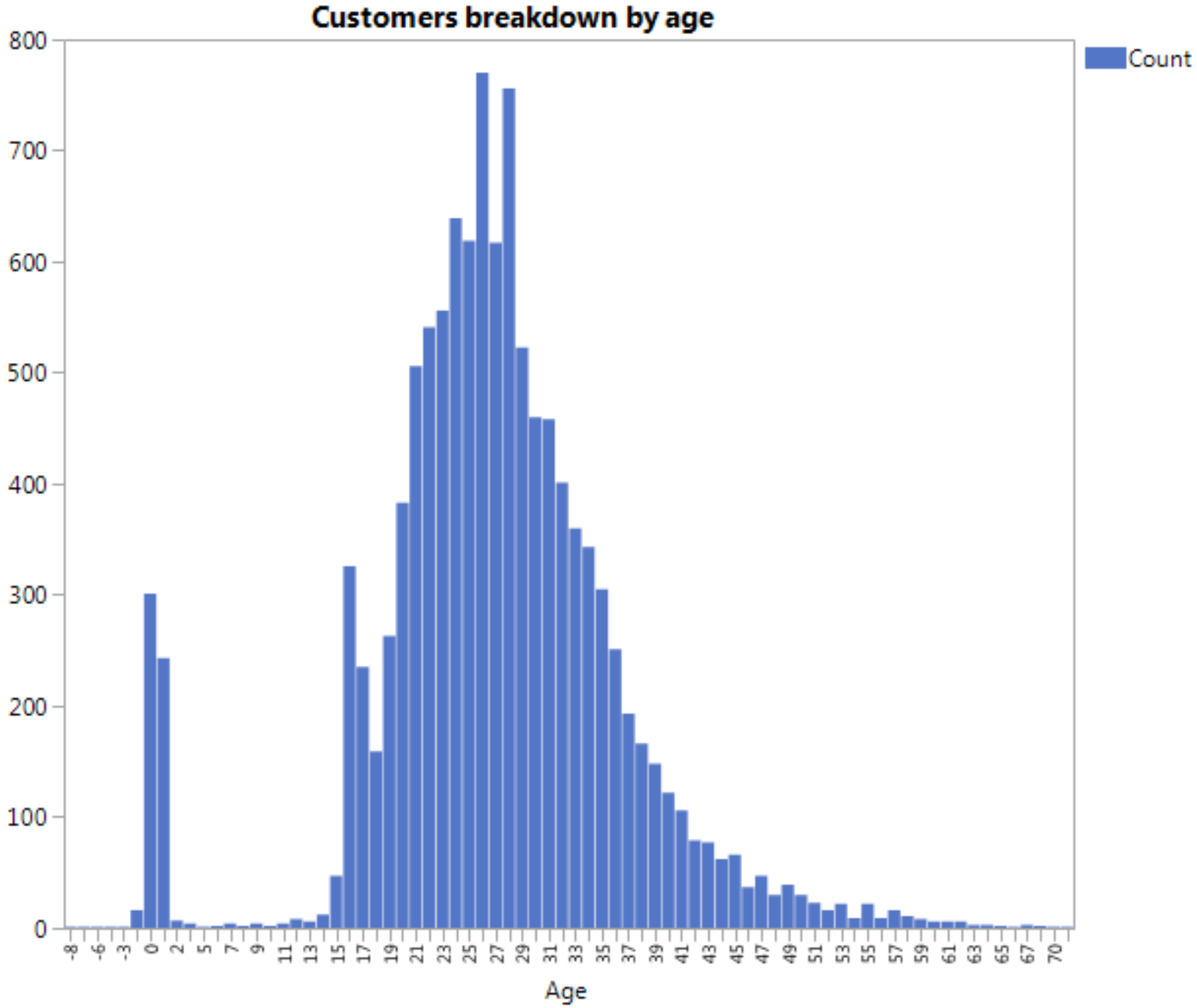
$$profit = final\ price\ - payout\ - transaction\ fee\ - cashback$$

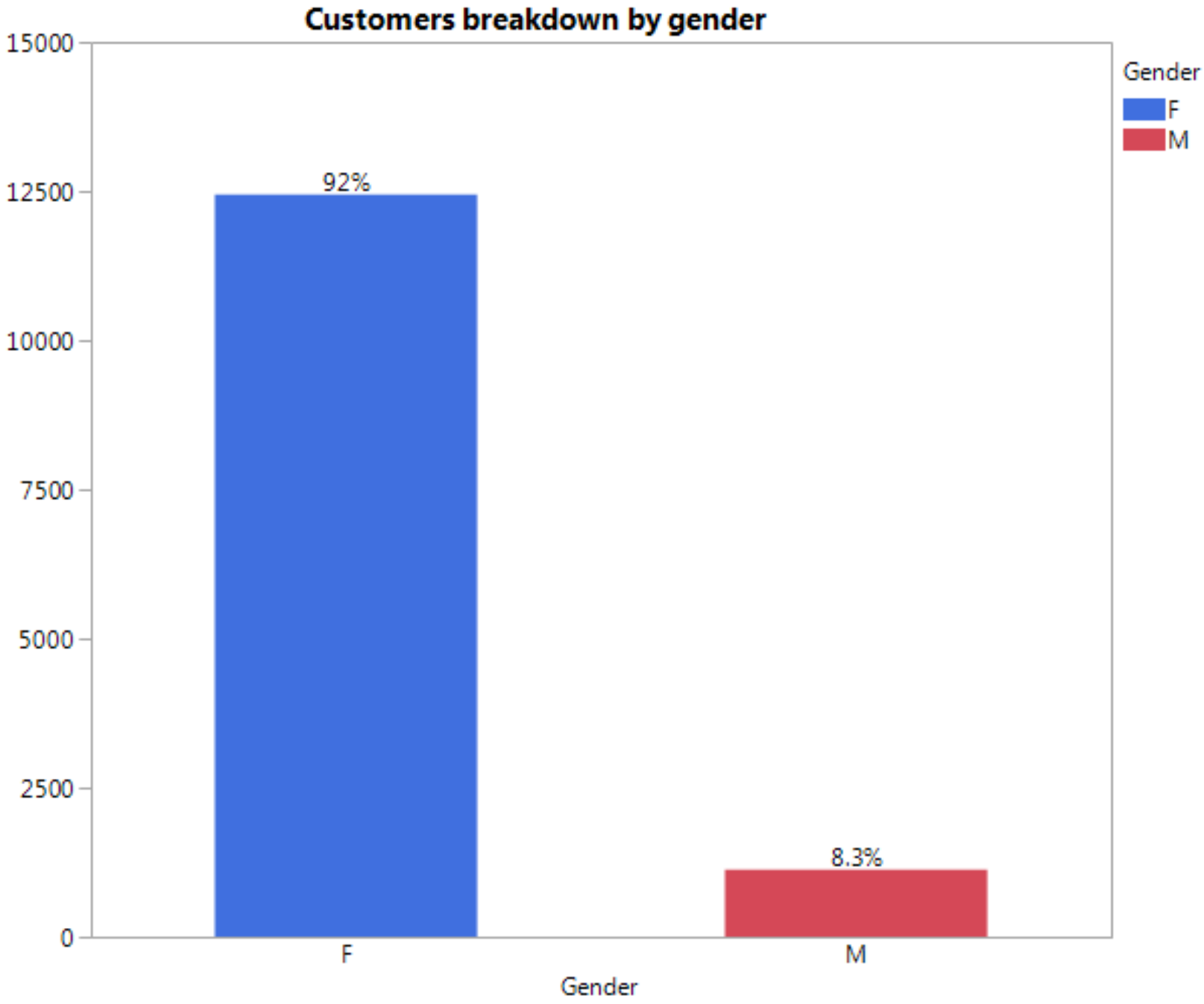
- Total profit of approx. -$98k

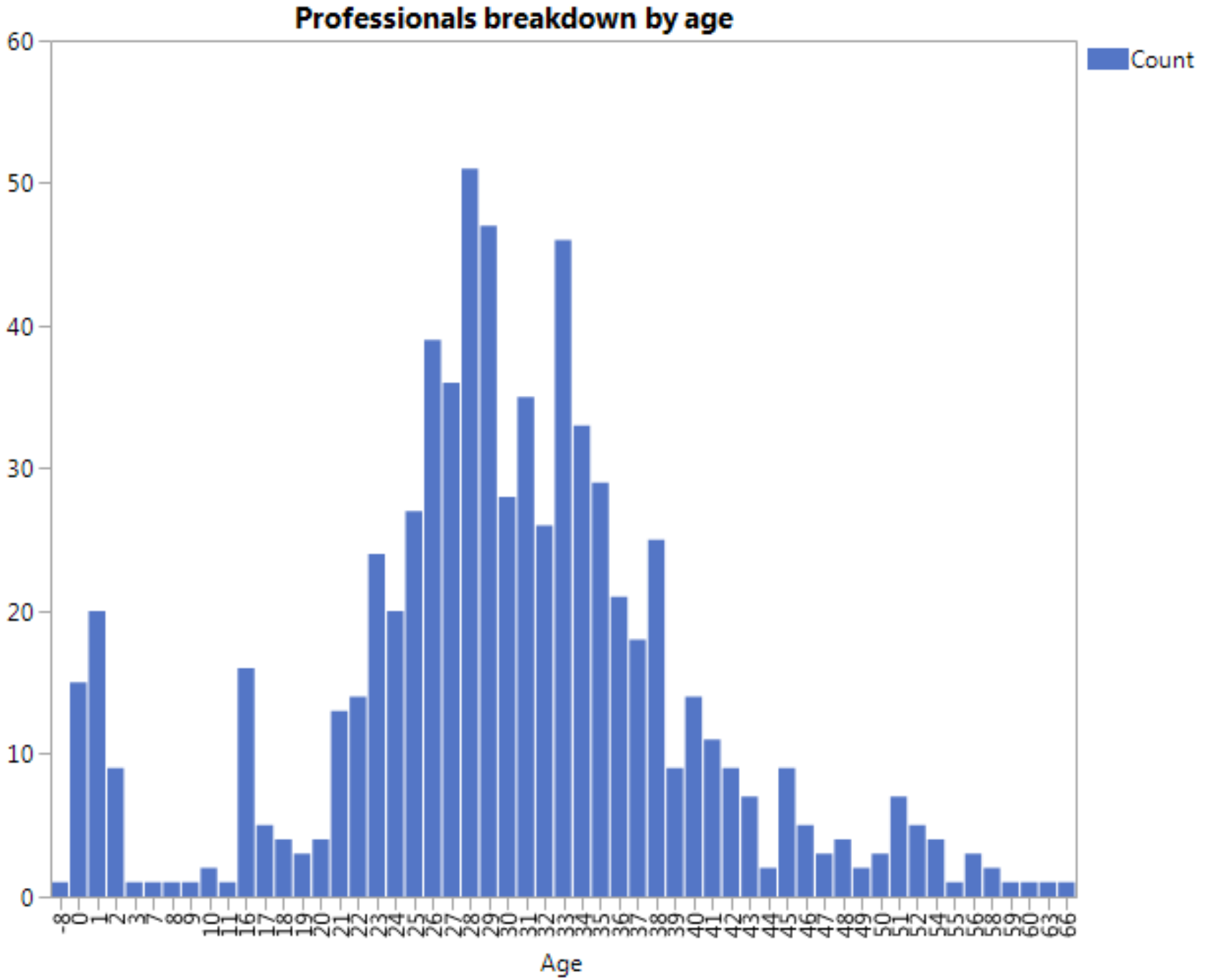
$$vanitee\ fee = final\ price\ - \frac{payout}{total\ fee} * final\ price - \frac{transaction\ fee}{total\ fee} * final\ price - \frac{cashback}{total\ fee} * final\ price$$

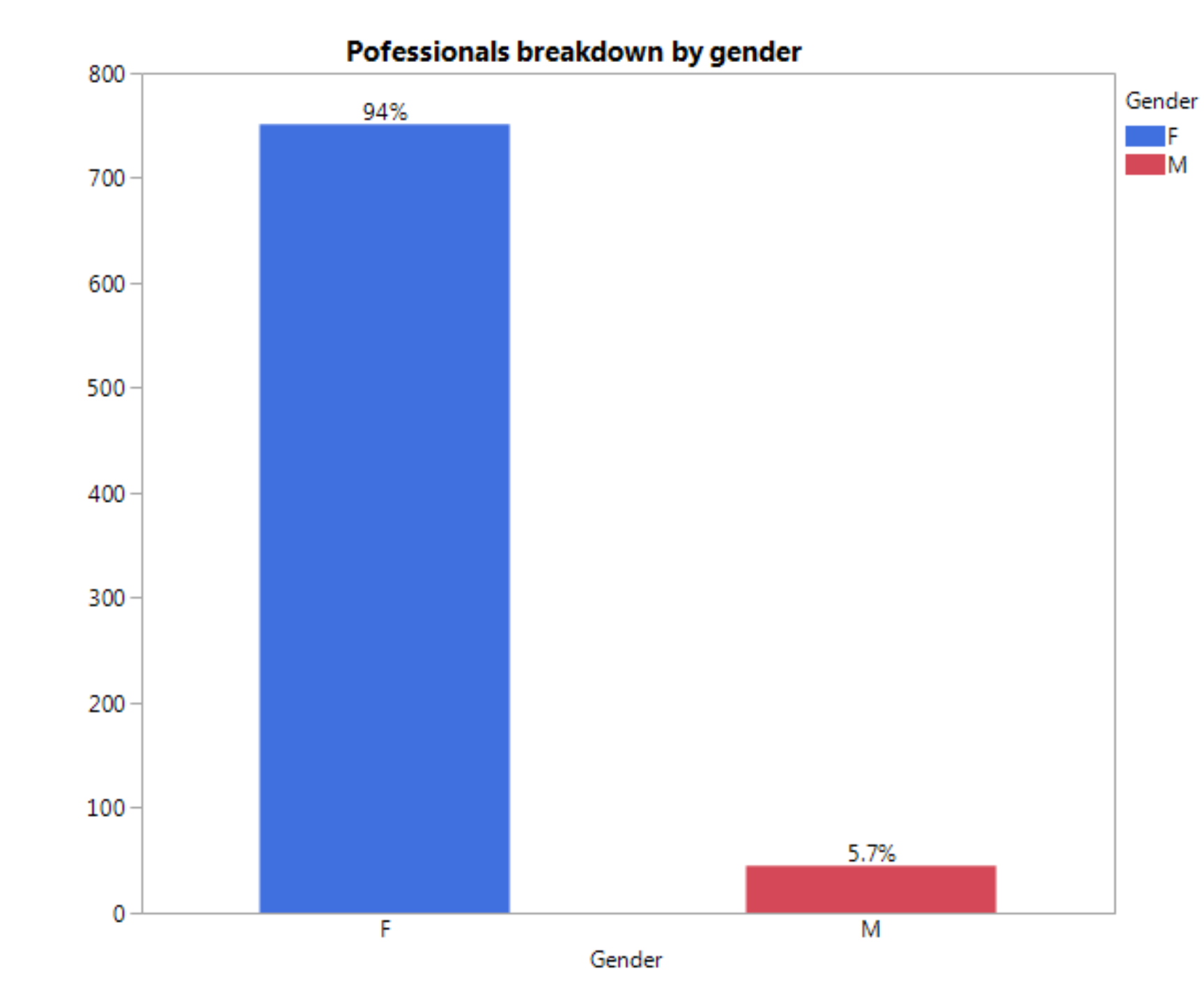
- Total profit of approx. -$85k

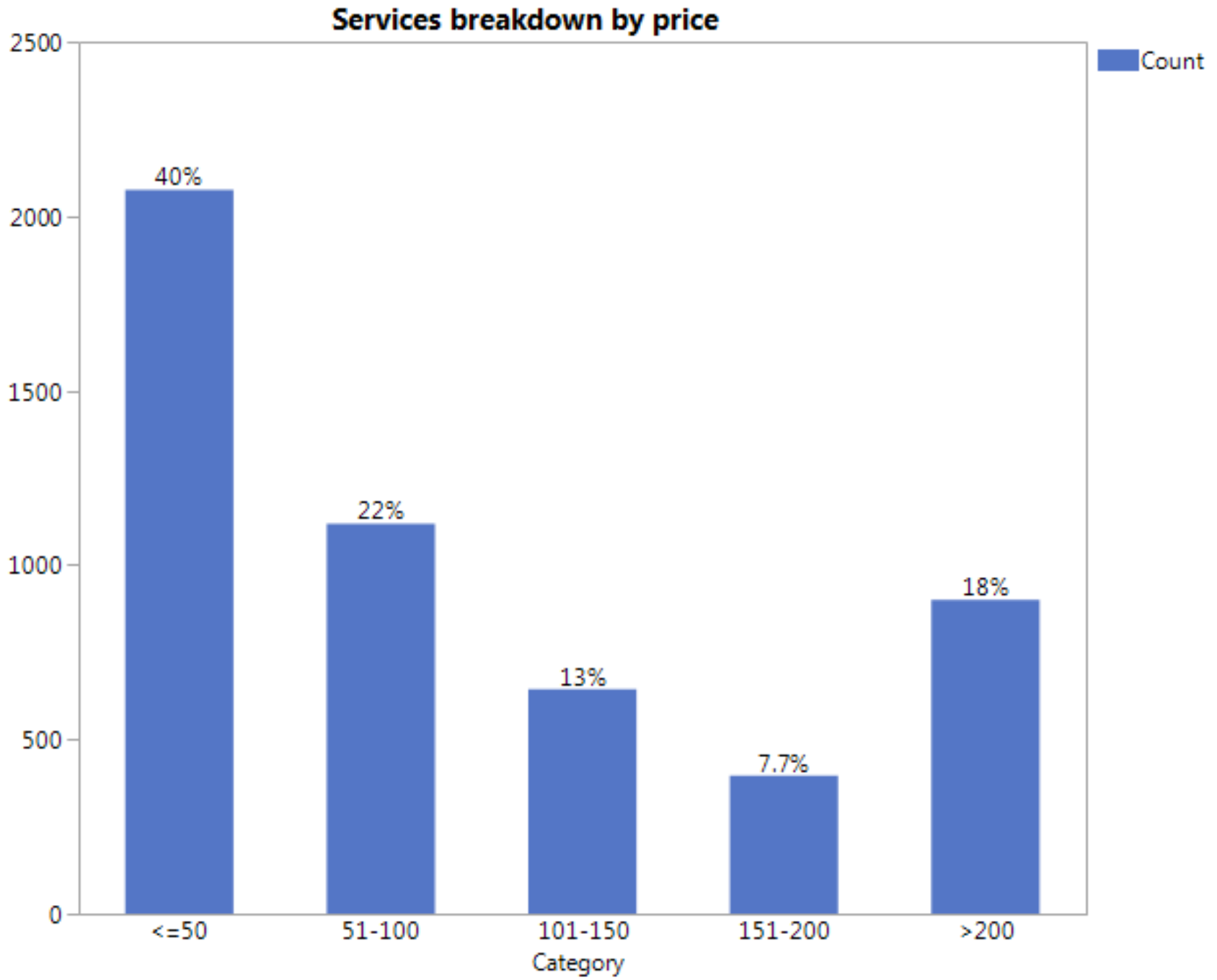

Users breakdown by type
Type
Non-professional
Professional
95%
4.9%
15000
10000
5000
0
Non-professional
Professional
Type

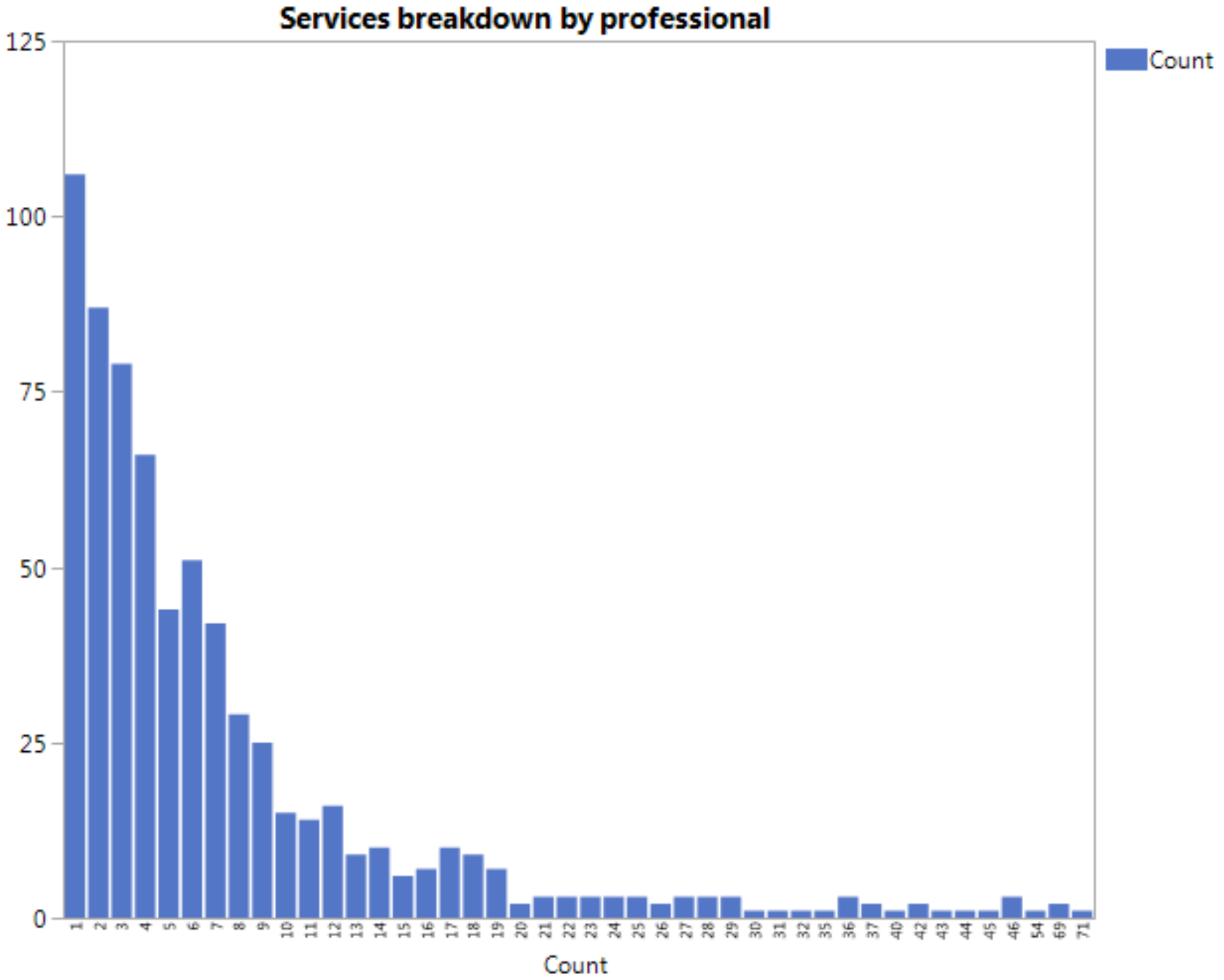

Customers breakdown by age
Count
800
700
600
500
400
300
200
100
0
-8
-6
-3
0
2
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
37
39
41
43
45
47
49
51
53
55
57
59
61
63
65
67
70
Age

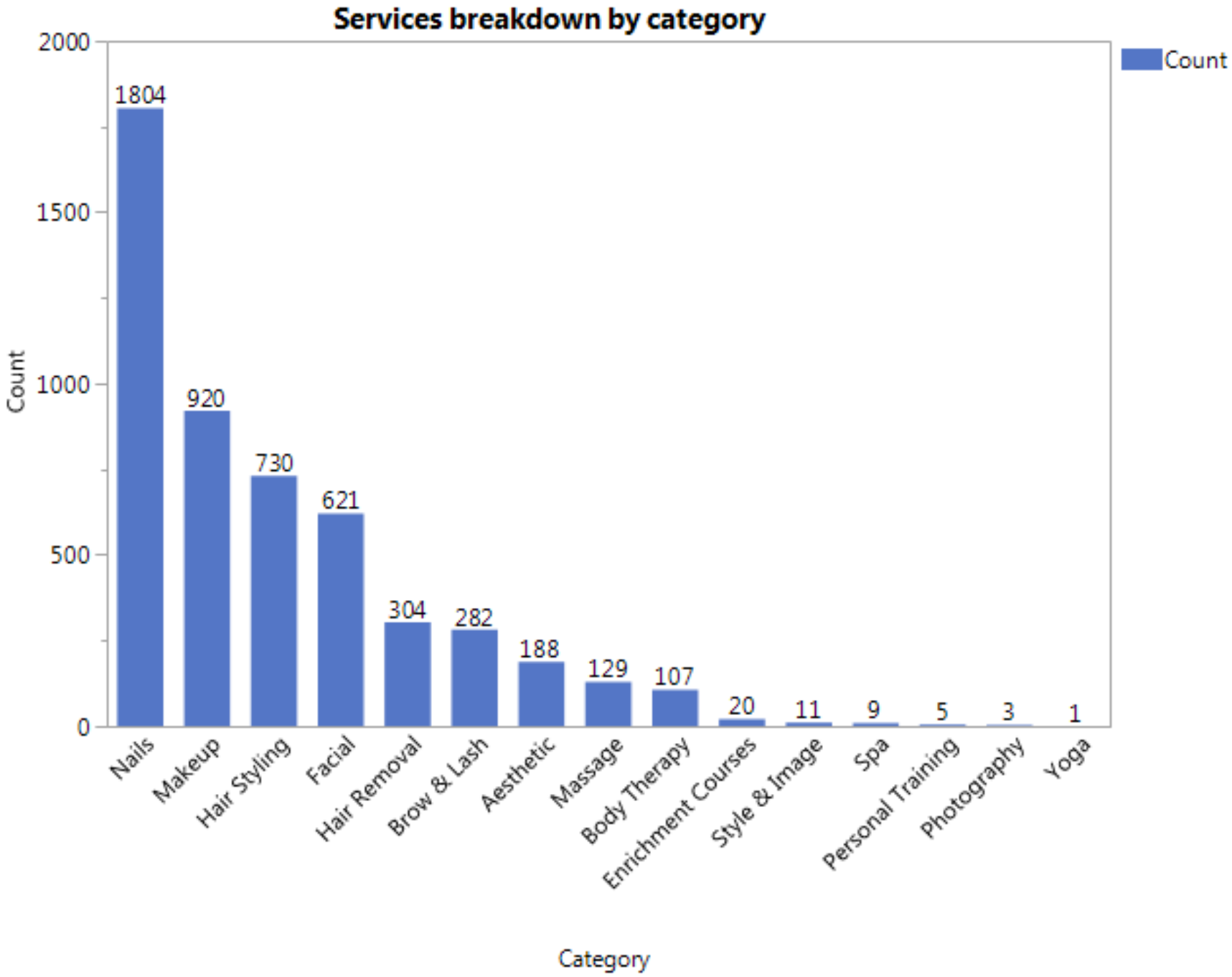

Customers breakdown by gender
15000
12500
10000
7500
5000
2500
0
92%
8.3%
F
M
Gender
Gender
F
M

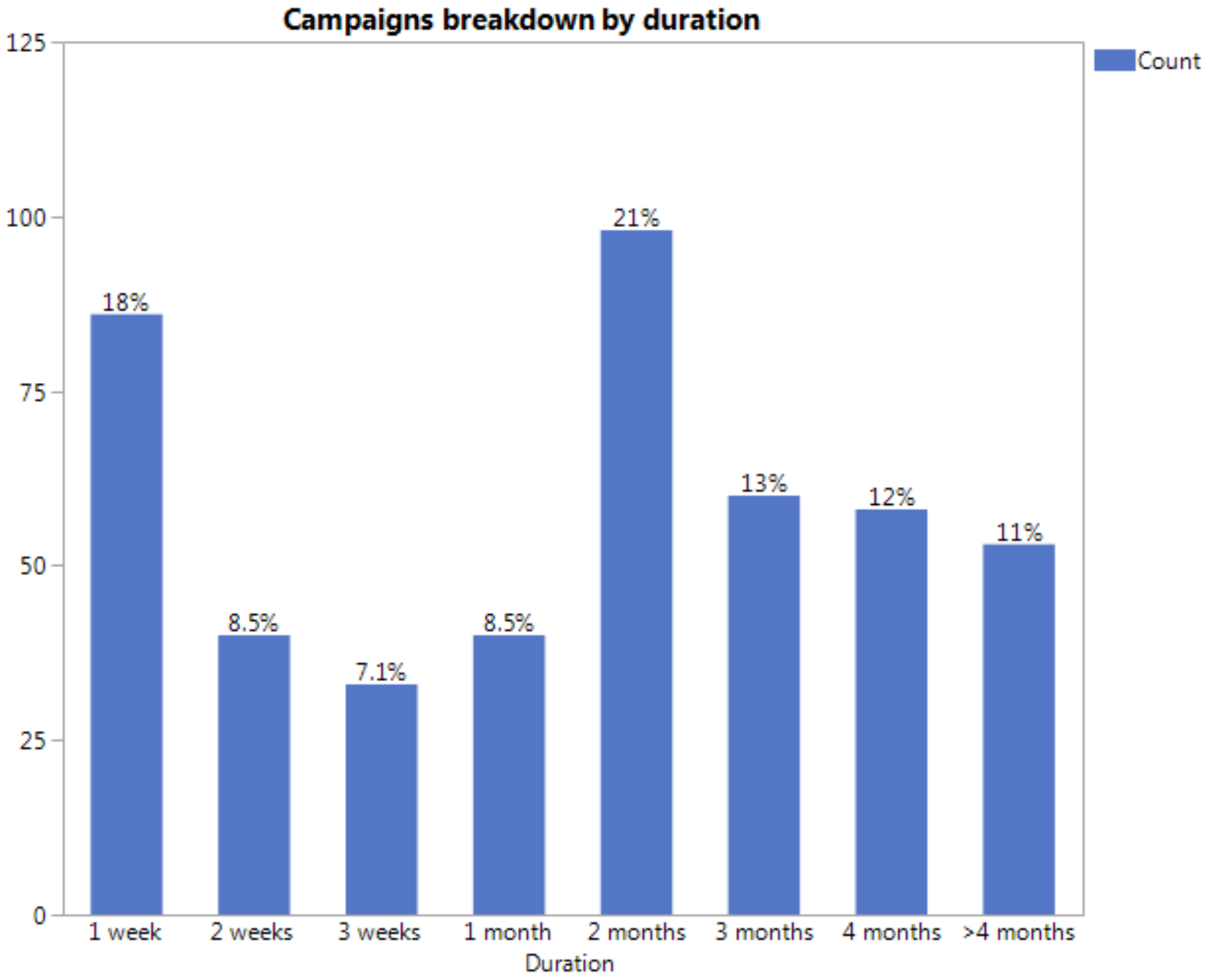

Professionals breakdown by age
Count
60
50
40
30
20
10
0
Age

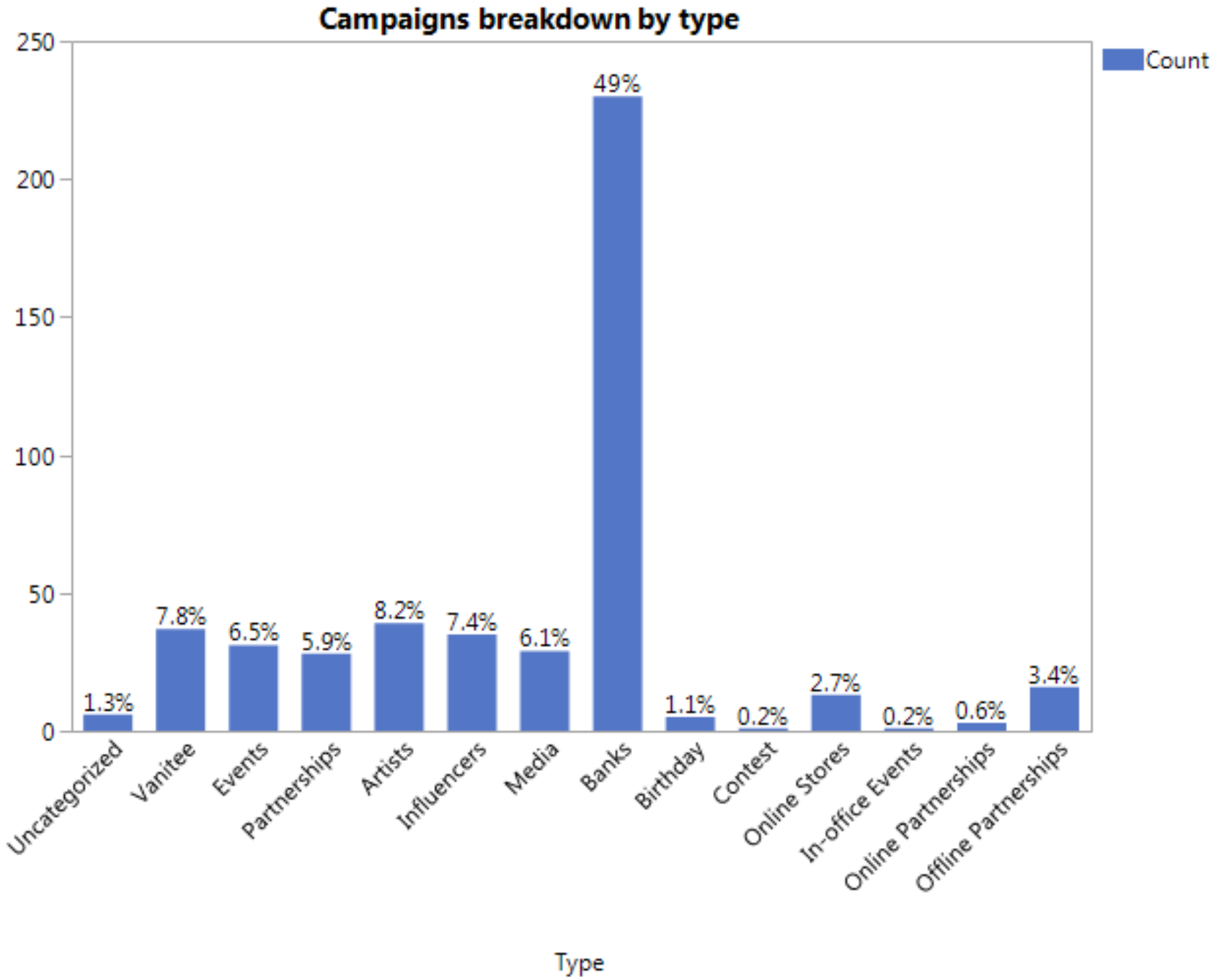

Pofessionals breakdown by gender
800
700
600
500
400
300
200
100
0
94%
5.7%
F
M
Gender
Gender
F
M
Gender

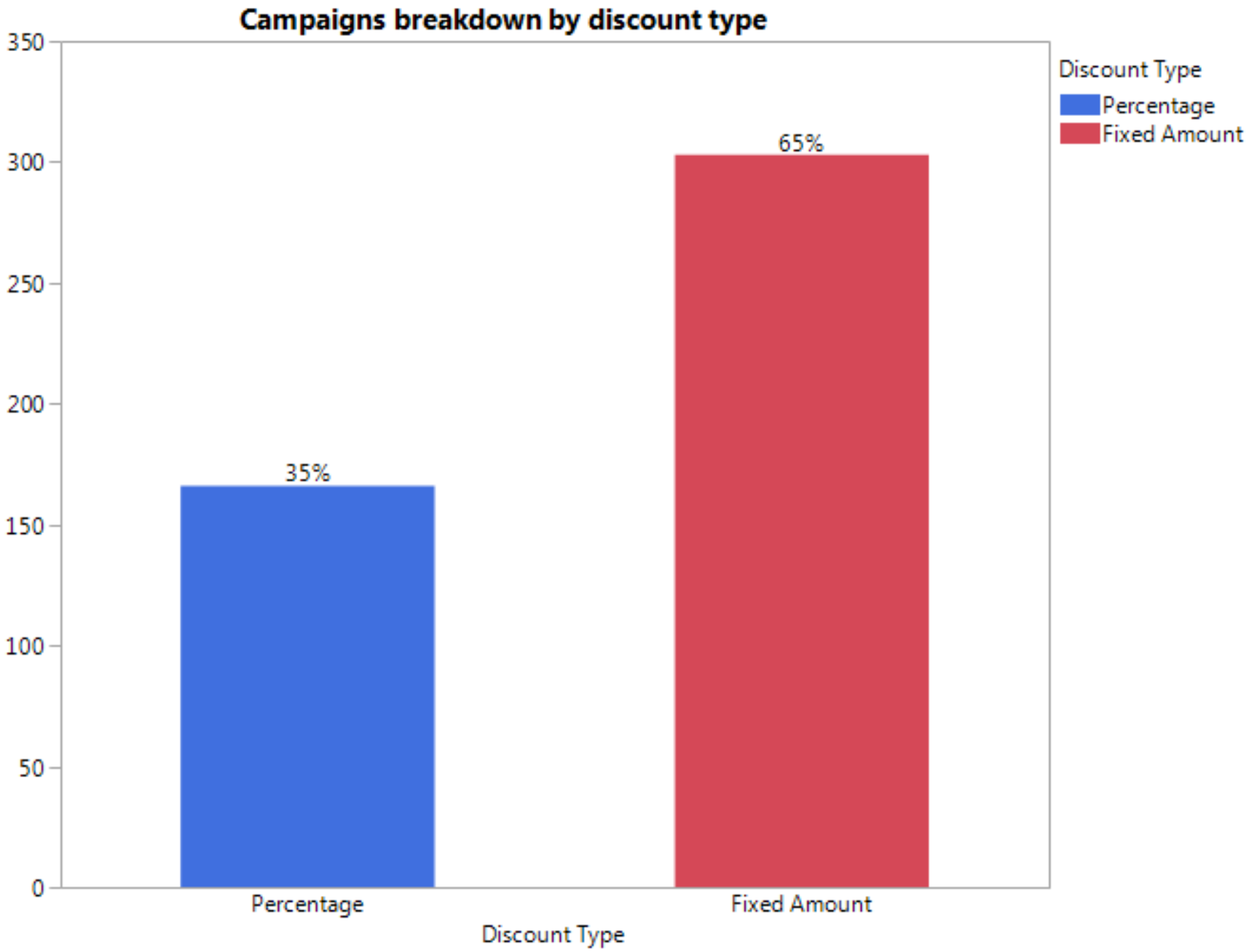

Services breakdown by price
Count
2500
2000
1500
1000
500
0
40%
22%
13%
7.7%
18%
<=50
51-100
101-150
151-200
>200
Category

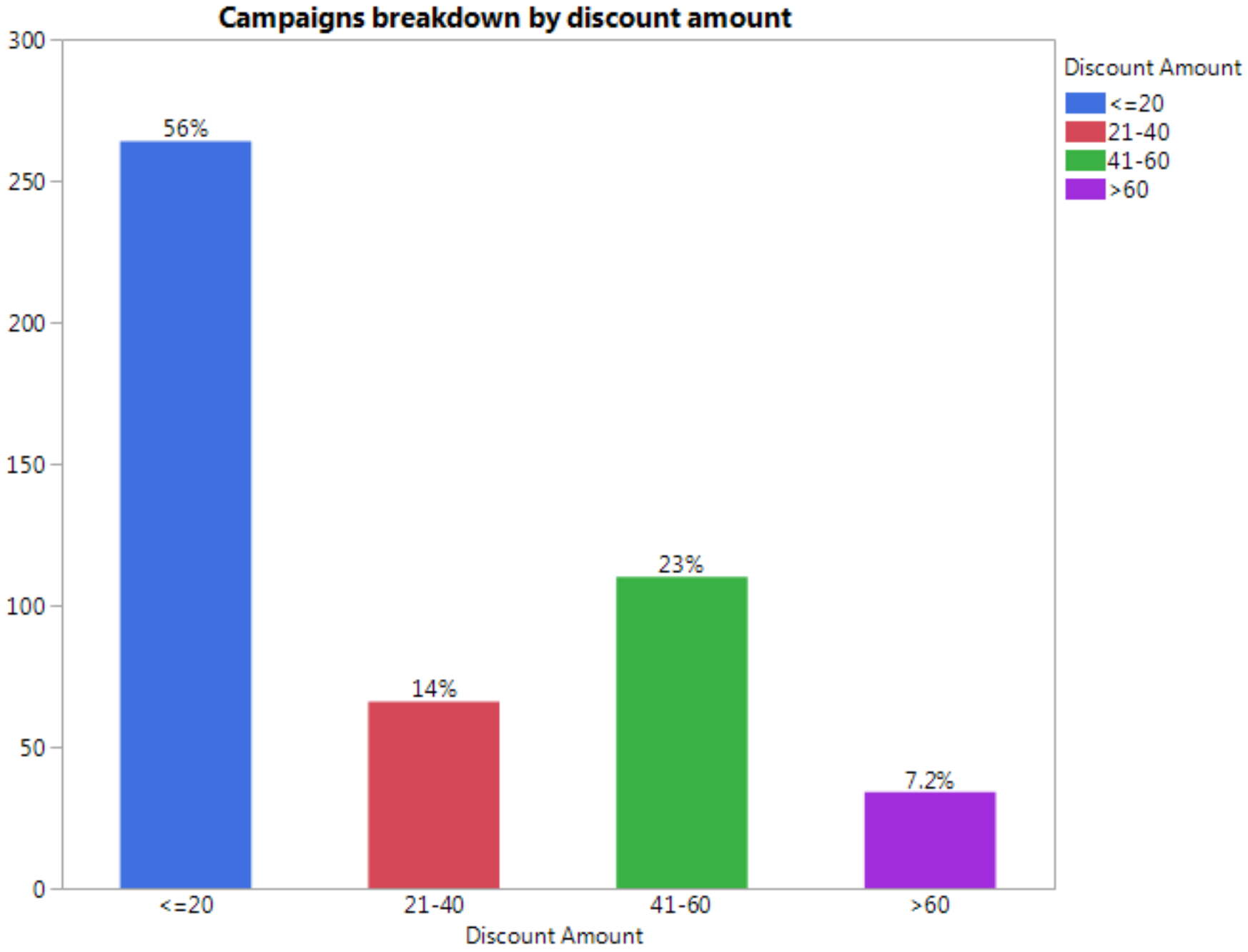

Services breakdown by professional
125
100
75
50
25
0
Count
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
35
36
37
40
42
43
44
45
46
54
69
71
Count

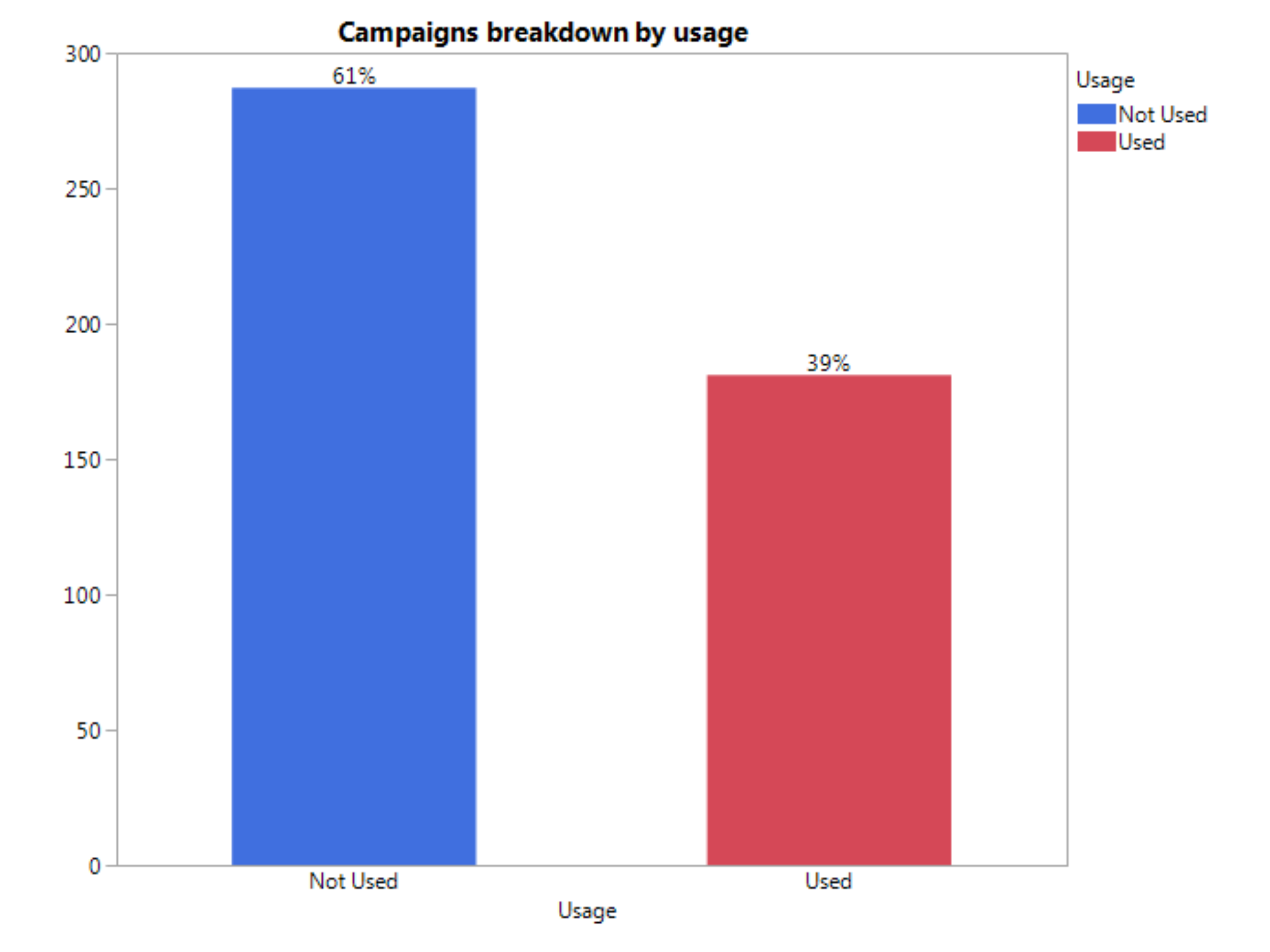

Services breakdown by category
Count
2000
1500
1000
500
0
1804
920
730
621
304
282
188
129
107
20
11
9
5
3
1
Nails
Makeup
Hair Styling
Facial
Hair Removal
Brow & Lash
Aesthetic
Massage
Body Therapy
Enrichment Courses
Style & Image
Spa
Personal Training
Photography
Yoga
Category

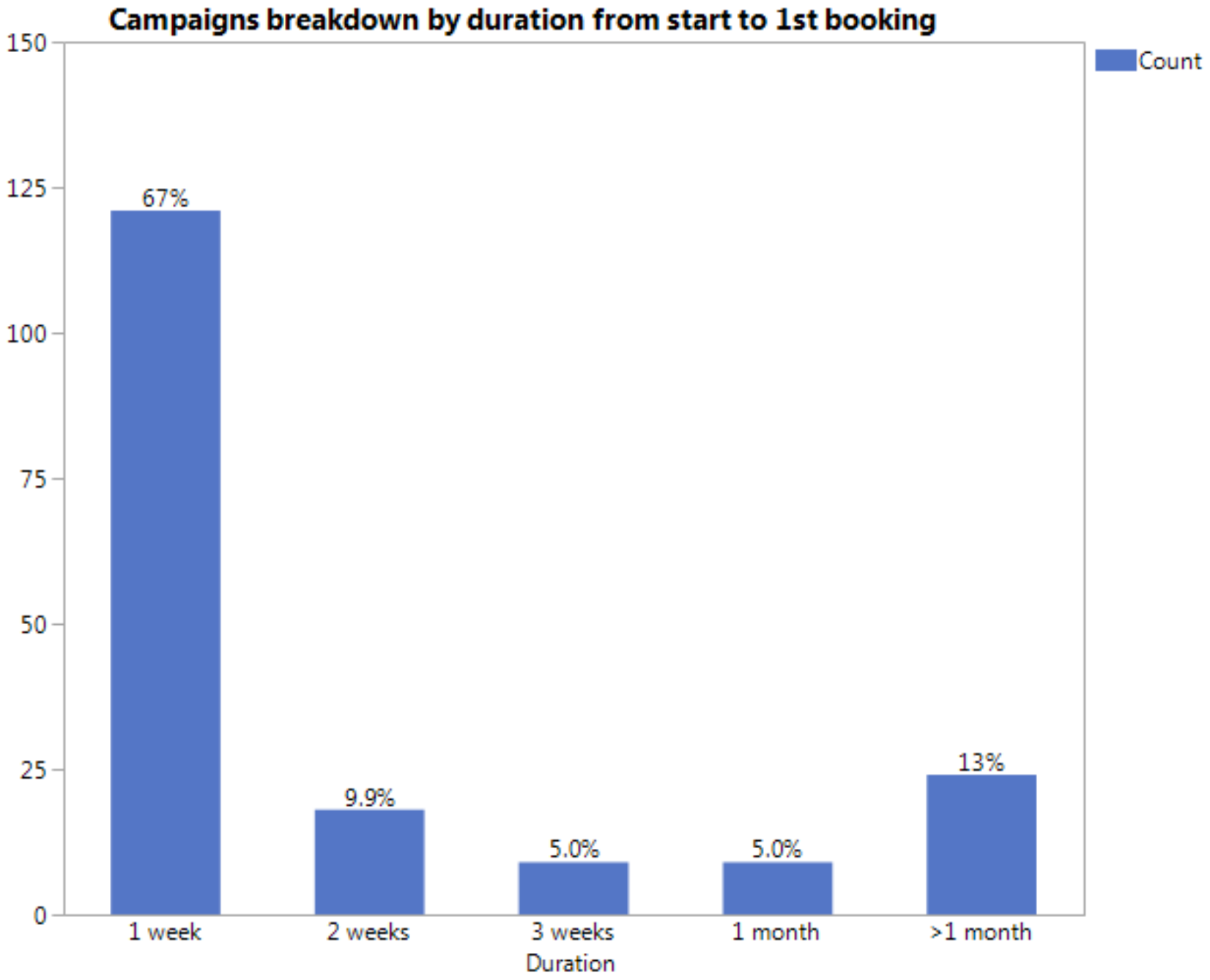

Campaigns breakdown by duration
125
100
75
50
25
0
Count
18%
8.5%
7.1%
8.5%
21%
13%
12%
11%
1 week
2 weeks
3 weeks
1 month
2 months
3 months
4 months
>4 months
Duration

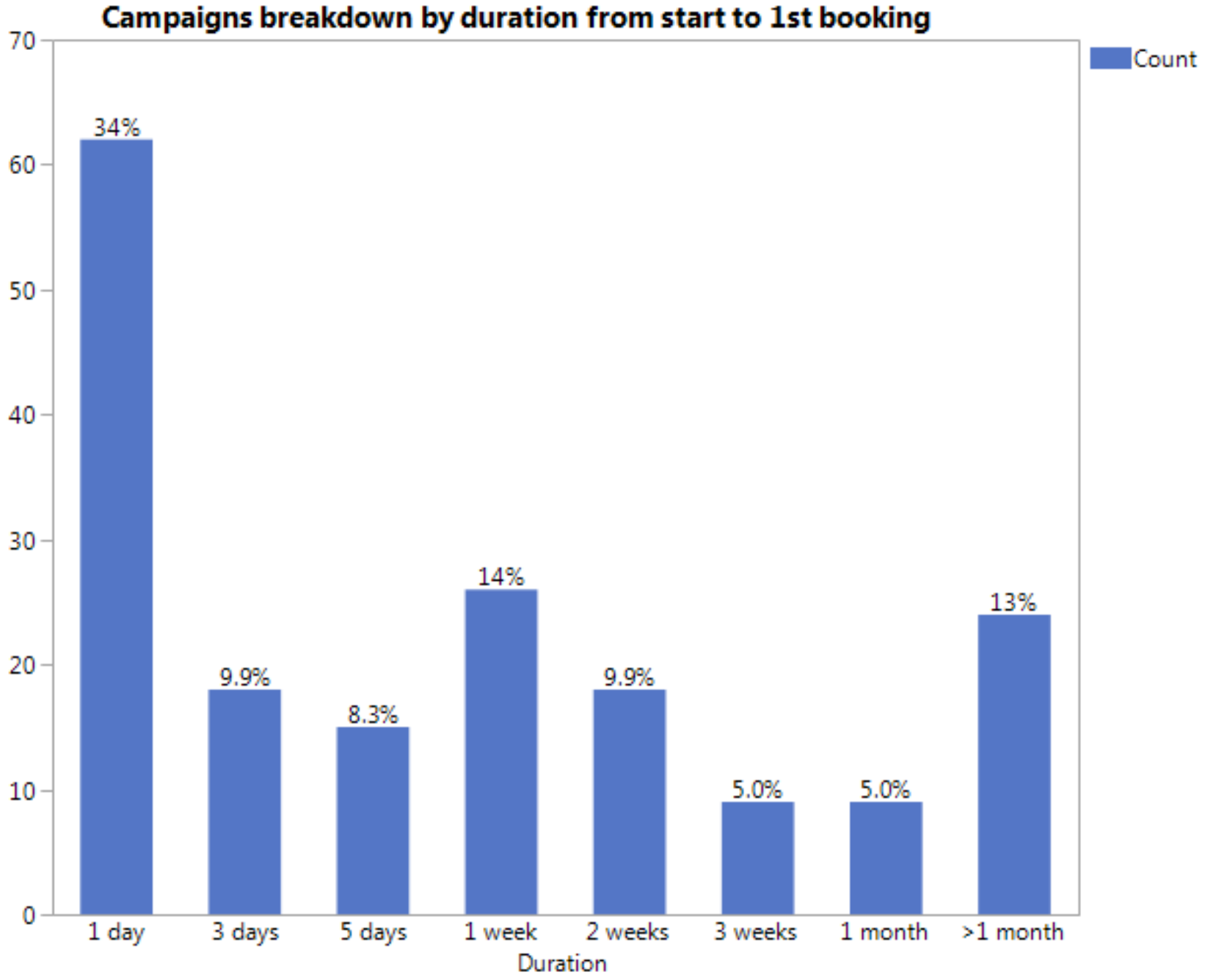

Campaigns breakdown by type
Count
250
200
150
100
50
0
49%
8.2%
7.8%
7.4%
6.5%
6.1%
5.9%
3.4%
2.7%
1.3%
1.1%
0.6%
0.2%
0.2%
Uncategorized
Vanitee
Events
Partnerships
Artists
Influencers
Media
Banks
Birthday
Contest
Online Stores
In-office Events
Online Partnerships
Offline Partnerships
Type


Campaigns breakdown by discount type
350
300
250
200
150
100
50
0
65%
35%
Percentage
Fixed Amount
Discount Type
Discount Type
Percentage
Fixed Amount


Campaigns breakdown by discount amount
300
250
200
150
100
50
0
56%
14%
23%
7.2%
<=20
21-40
41-60
>60
Discount Amount
Discount Amount
<=20
21-40
41-60
>60


Campaigns breakdown by usage
300
250
200
150
100
50
0
61%
39%
Not Used
Used
Usage
Usage
Not Used
Used


Campaigns breakdown by duration from start to 1st booking
Count
150
125
100
75
50
25
0
67%
9.9%
5.0%
5.0%
13%
1 week
2 weeks
3 weeks
1 month
>1 month
Duration


Campaigns breakdown by duration from start to 1st booking
Count
70
60
50
40
30
20
10
0
34%
9.9%
8.3%
14%
9.9%
5.0%
5.0%
13%
1 day
3 days
5 days
1 week
2 weeks
3 weeks
1 month
>1 month
Duration

Methodologies

- Cluster Analysis
  - Identify any clusters based on bookings patterns
  - Attempt to segment customers if there are any distinct clusters

- Survival Analysis
  - Predict when a booking will most likely occur during a campaign (fixed period)
  - Whether the type of campaign affects this occurrence

Scope of Work

Completed ✓	Removed ✗	In-Progress and New Tasks +
1. Data cleaning on all 7 tables by replacing missing values and filtering out duplicates 2. Identifying the preferred type of booking 3. The following completed tasks are based on bookings made online 4. Identifying the status at which most bookings fall under 5. Identifying the number of bookings made typically 6. Identifying the number of bookings made within the range of data used for each year, month, a combination of both year and month, and lastly week 7. Identifying the amount of bookings that were made recently 8. Identifying the bookings within a certain price range 9. Identifying the duration a booking is usually made from the time of account sign up to the first booking.	1. Conducting EDA on activeness of beauty professional	1. Analyse the breakdown of bookings based on the day it is checked out 2. Analyse the breakdown of individual services and the prices 3. Analyse the campaigns that have the most redemptions to follow up on finding the ideal duration and discount a campaign code should have 4. Analysis on the attractiveness of a professional 5. Transforming and normalising the data to prepare it for further analysis 6. Perform cluster analysis for customer segmentation.

Scope of Work

Completed ✓	Removed ✕	In-Progress and New Tasks +
9. Identifying the number of service(s) bookings typically have 10. Identifying the service that most bookings are made for 11. Identifying the amount of booking that utilise a campaign code and credits separately 12. Identifying the number of customers for each age and gender 13. Identifying the number of professionals for each age and gender 14. Identifying the number of services under different price ranges 15. Identifying the number of services a professional typically offers 16. Identifying the number of services provided in different categories 17. Identifying number of campaigns for various range of duration, type, discount type, discount amount, usage and duration from start to the first booking		7. Conduct survival analysis to find out the value of a customer and the effectiveness of a campaign 8. Written report, final presentation, poster, insights and recommendations to Vanitee.

Work Plan

	Tasks	Week -1	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16
Requirement Gathering	Confirm Project Sponsor																		
	Gather Requirements																		
Data Gathering	Gather Data from Sponsor																		
Project Proposal	Refining Project Scope																		
	Proposal Preparation																		
	Wiki Page Preparation																		
Research	Literature Research																		
Data Exploratory	Data Preparation																		
	Data Cleaning																		
	Data Exploration																		
Proposal Deadline - 15 January 2017 - Week 3																			
Data Exploratory	Generate Findings																		
Interim	Sponsor Relvew and Preparation																		
	Interim Report Preparation																		
	Interim Presentation Preparation																		
	Interim Wiki Update																		
Interim Practicum Presentation - 20 to 24 Feb 2017 - Week 8																			
Model Building	Data Normalisation & Transformation																		
	Cluster Analysis																		
	Survival Analysis																		
Insights & Recommendation	Create Visualisation from Analysis Results																		
	Generating Insights																		
	Formulate Recommendations																		
Final	Sponsor Review																		
	Abstract & Full Paper Preparation																		
	Final Wiki Update																		
	Final Presentation Preparation																		
	Final Paper Preparation & Submission																		
Final Practicum Presentation & Submission - 2 April & 16 April 2017 - Week 14 and 16																			
Conference	Undergraduate Conference on Data Analytics																		

Legend	
	Planned
	Actual
	Milestone

Work Plan

Tasks		Start Week	End Week	Planned Man Hours	Actual Man Hours	Team Members Involved	Status
Requirement Gathering	Confirm Project Sponsor	-1	0	4	4	All	Completed
	Gather Requirements	-1	0	2	2	Andrew	Completed
Data Gathering	Gather Data from Sponsor	-1	0	2	2	Andrew	Completed
Project Proposal	Refining Project Scope	0	2	3	5	All	Completed
	Proposal Preparation	0	2	20	20	All	Completed
	Wiki Page Preparation	0	2	10	10	Sarah	Completed
Research	Literature Research	1	3	5	8	All	Completed
Data Exploratory	Data Preparation	1	4	15	15	All	Completed
	Data Cleaning	1	5	15	20	All	Completed
	Data Exploration	2	5	25	35	All	Completed
Proposal Deadline - 15 January 2017 - Week 3							
Data Exploratory	Generate Findings	3	8	20	30	All	Completed
Interim	Sponsor Review	4	8	5	6	All	Completed
	Interim Report Preparation	4	8	30	25	All	Completed
	Interim Presentation Preparation	5	8	20	20	Andrew	Completed
	Interim Wiki Update	6	8	20	15	Sarah	Completed
Interim Practicum Presentation - 20 to 24 Feb 2017 - Week 8							
Model Building	Data Normalisation & Transformation	8	-	15	-	All	Not Started
	Cluster Analysis	8	-	20	-	Sarah	Not Started
	Survival Analysis	9	-	20	-	Andrew	Not Started
Insights & Recommendation	Create Visualisation from Analysis Results	10	-	30	-	All	Not Started
	Generating Insights	11	-	30	-	All	Not Started
	Formulate Recommendations	12	-	25	-	All	Not Started
Final	Sponsor Review	12	-	8		All	Not Started
	Abstract & Full Paper Preparation	12	-	45	-	All	Not Started
	Final Wiki Update	13	-	10	-	Sarah	Not Started
	Final Presentation and Poster Preparation	14	-	20	-	All	Not Started
	Final Paper Preparation & Submission	14	-	30	-	All	Not Started
Final Practicum Presentation & Submission - 2 April & 16 April 2017 - Week 14 and 16							
Conference	Undergraduate Conference on Data Analytics	16	-	8	-	All	Not Started

Thank you