Airbnb New User Booking

Exploratory Business Analysis

by Valentin Plekhanov

John Bryce Data Analyst Program (February 2020)

Introduction.

Airbnb, Inc. is an online marketplace for arranging or offering lodging, primarily homestays, or tourism experiences. The company does not own any of the real estate listings, nor does it host events; it acts as a broker, receiving commissions from each booking. The company is based in San Francisco, California, United States. New users on Airbnb can book a place to stay in 34,000+ cities across 190+ countries.

By analyzing new user behavior and their environment, Airbnb can share more personalized content with its community, invest in advertising more efficiently, reduce average time to first order, and better meet demand.

Data Structure.

This data was officially introduced by Airbnb as part of a new hiring process for the analysis and machine learning department.
In this exploratory, given a list of real users along with their demographics, web session records, and some summary statistics. All the users in this dataset are from the USA.
The original dataset was obtained in CSV format, 3 mains files.
File 1 contains data about 213451new users such as:

- Age, gender information.
- Site activity.
- Activity dates.
- Booking destination.

***There are 12 possible outcomes of the destination country: 'US': USA, 'FR': France, 'CA': Canada, 'GB': Great British, 'ES': Spain, 'IT': Italy, 'PT': Portugal, 'NL': Nederland's, 'DE': Germany, 'AU': Australia, 'NDF' (no destination found), and 'other'. Please note that 'NDF' is different from 'other' because 'other' means there was a booking but is to a country not included in the list, while 'NDF' means there wasn't a booking.

File 2 contains data about summary statistics of users' age group, gender, country of destination. Total 10567737 entries.

File 3 data from web sessions log for users.

Analysis description.

Data was analyzed and visualized in Python, by using Jupyter Notebook. The syntax is annexed to this report.
The study is divided into 5 stages:

1. Analysis by personal details
   1.1 Age.
   1.2 Gender.
   1.3 Age and gender combination.
2. Analysis by booking environment.
   2.1 Paid marketing.
   2.2 Apps and Devices.
3. Site Analysis.
   3.1 Language.
   3.2 Sessions.
4. Trends Analysis.
5. Business Insights.

Research goals.

Main objective of the study: **New user analysis for strategic business decisions.**

Research.

First, was be found that the percent of users that never booking is **58.35%** from total new users. To understand what makes new users sign up for the site, but not to place an order, we will check the quality of new users in terms of **booking rate**.

1. Analysis by personal details.

1.1 Age

According to Airbnb practices, only an 18-year-old person can open an account.
In contrast, a company does not require age pre-registration and it is possible that new users 15 to 18-year old's can also place an order on their own. Therefore, for research, all customers who did not specify their age or age below 15 are defined as N/A (No age setting).
In addition, for the reliability of the study, users who have not defined their age in the range of 15-85 years will be associated with a group with no age setting(N/A).
As a result of company policy towards defining the age of a new customer, the percent new users without age is **41.5%.**
76.84% of total users without age no makes any order.

For the purpose of the study, we divided all users with age setting into 6 groups:

15-24
25-34
35-44
45-54
55-64
65+

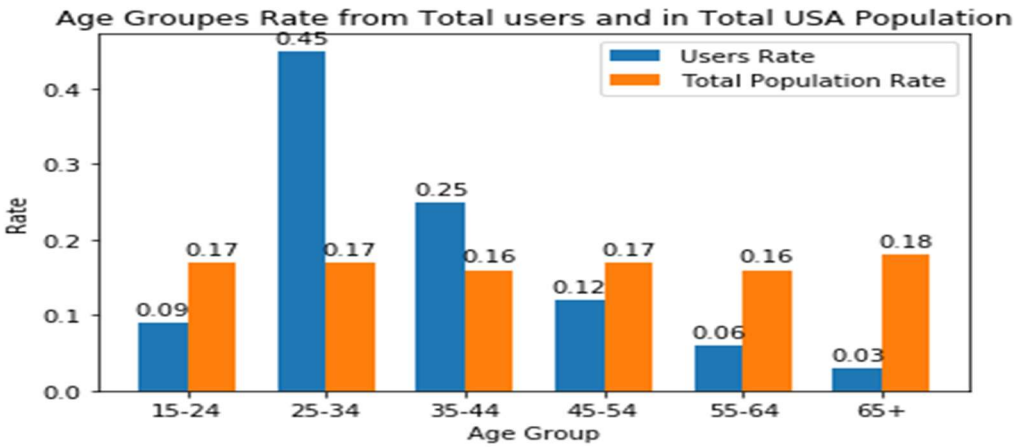


Figure 1.1 Participation rate of each age group of total users and total population in USA

For testing level of representation, the percentage of representation of each age group from all users is calculated. *
To compare the extent to which each age group is presented, a comparison was made with the representation of the same age group from a total USA population. **

* "Total users": Defined only as users who indicate their age at registration.

** "Total USA population": A total population over the age of 15.

Good representation among 25-34 age group users can be seen in 45% of new users, with the general population being only 17%. Compared to users in the 55-64 age group with 6% representation and over 65 who represent only 3%. Even among young people aged 15-24, the participation rate is only 6%, compared to 17% of the population.

Next, behavior should be checked for users who have already arrived on the site and placed an order. We will check order level and preferred destination.

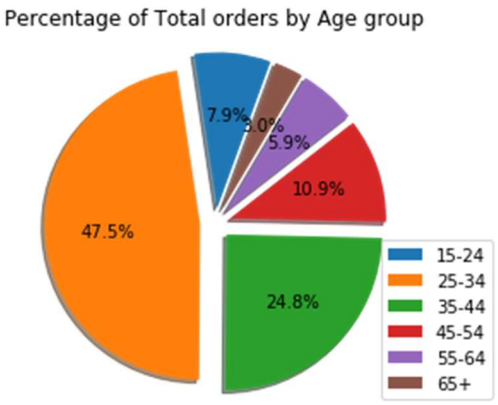


Figure 1.2 Distribution of all orders per age group.

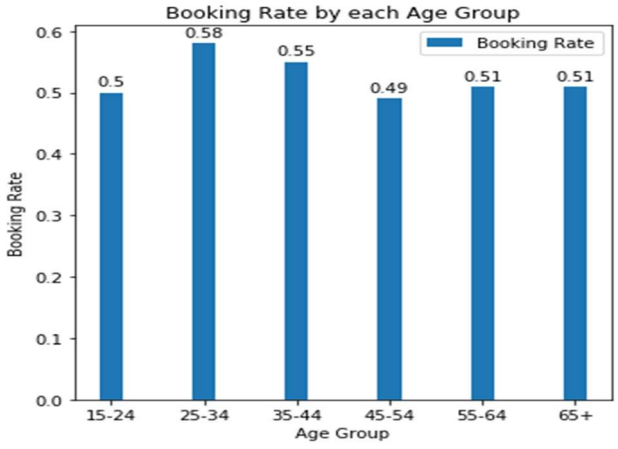


Figure 1.3 Booking Rate by each Age group.

.

age_group	country	rate
15-24	US	0.65
25-34	US	0.73
35-44	US	0.71
45-54	US	0.69
55-64	US	0.68
65+	US	0.64

age_group	country	rate
15-24	other	0.11
25-34	other	0.11
35-44	other	0.11
45-54	other	0.10
55-64	other	0.11
65+	other	0.12

age_group	country	rate
15-24	FR	0.07
25-34	FR	0.05
35-44	FR	0.06
45-54	FR	0.07
55-64	FR	0.06
65+	FR	0.08

Table 1.1 Top 3 countries destination by each group with booking rate.

Figure 1.3 shows that the order rates in each age group are between 50% and 58%. In addition, Table 1.1 shows that there are no significant differences in preferences between groups in order targets. Most new users choose as their first country destination in the USA.

Thus, when each age group has relatively similar results among the users who have already arrived, registered and placed their first order, it can be concluded that results in Figure 1.1 are as a result of low exposure to an ineffective advertising for the site or lack of dedicated advertising for age groups over 55 and young people up to age 24.

1.2 Gender

For the purpose of researching new users by gender, we will only extract users who specify their gender.
Percentage of users without sex definition is **44.96%**.
69.58% of total users without gender no makes any order.

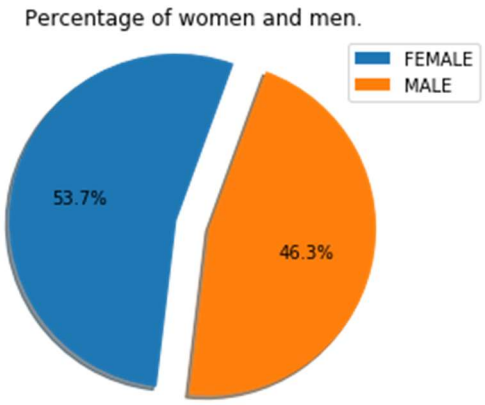


Figure 1.4 Gender distribution.

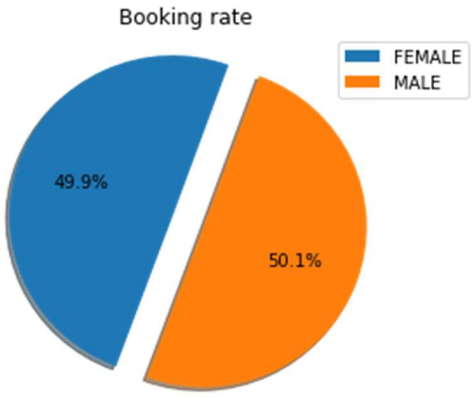


Figure 1.5 Booking rate by gender.

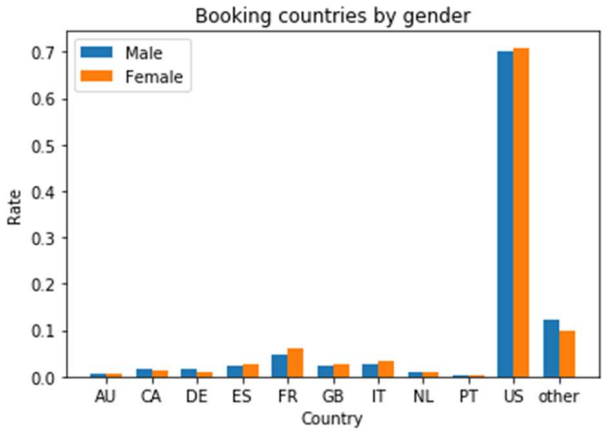


Figure 1.6 Preferred destination by gender.

As in the previous section, representation, order level and order preferences were also examined, by gender distribution.

Although women's representation is slightly higher than men, no significant differences were found to draw clear conclusions about the gender difference effect.

1.3 Age and gender preferences combination.

To attract more high-quality new users, deeper penetration is needed in analyzing new users based on age and gender integration. Dedicated advertising to each age group and to a specific gender in each group can produce more effective results and not only attract more new users, but high-quality new users who are more likely to get their first order.

First, will filter for new users with age and gender settings only.

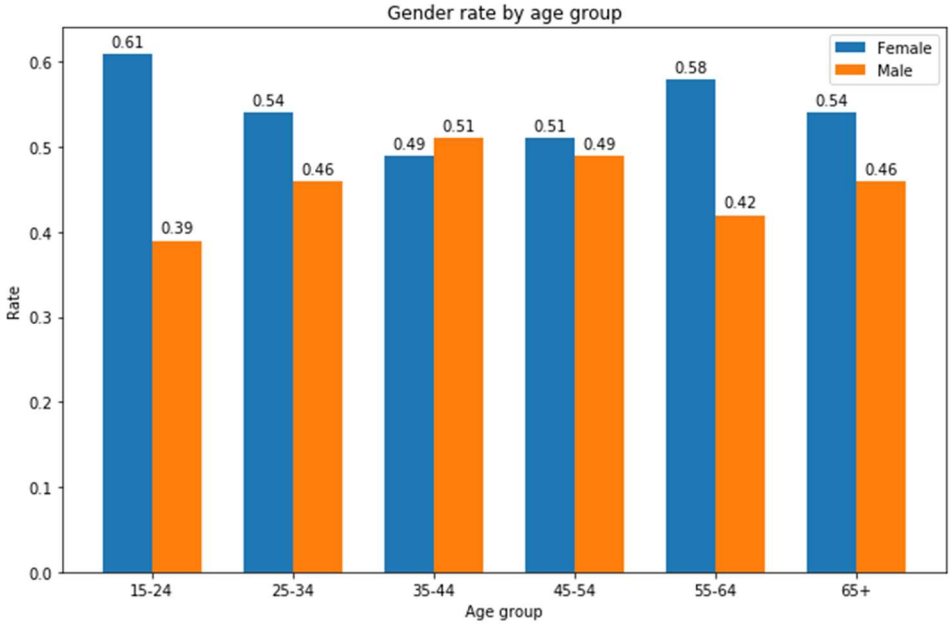


Figure 1.7 shows a particularly noticeable difference in the ratio of men to women among persons aged 15-24 (22%) and 55-64 (16%). In other age groups the ratio is relatively equal.

Figure 1.7 Division between women and men in each age group.

Next, we will review new user quality by order level in each age group and gender.

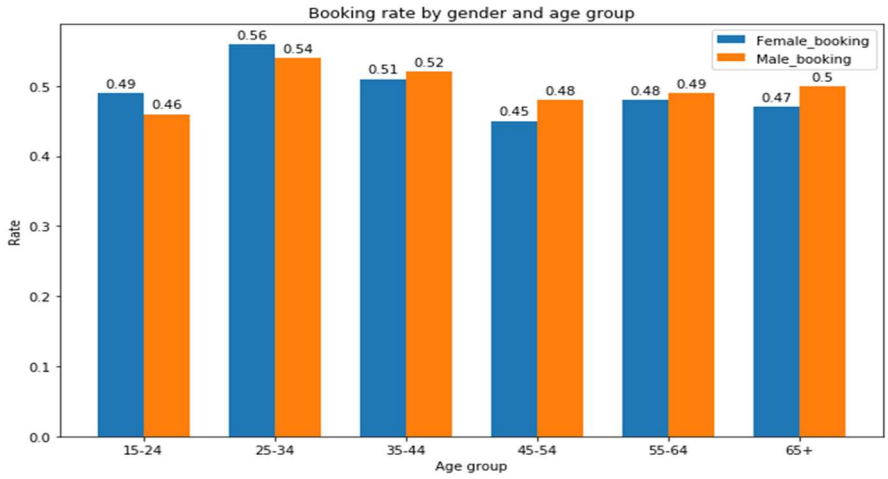


Figure 1.8 shows no significant differences in the level of invitation between women and men in each age group.

Figure 1.8 First booking rate women and men in each age group.

Research results by age and gender indicate that new users' quality in each age group is relatively the same. At the same time, the representation of general women among new users in certain age groups as seen in Figure 1.7 reaches differences up to 22%.

The conclusion is similar to the conclusion in section 1.1. Low efficiency in dedicated advertising leads to low attract of high-quality new users.

Section 5 Business Insights will provide recommendations on the results of analysis by personal details.

2. Analysis by booking environment.

A user environment analysis is necessary to understand how users come to the site, how they discover a site, and which marketing channels are more useful for attracting new users.

2.1 Paid marketing.

To analyze the effectiveness of the marketing channel, we will look at how many new users come from each channel, as opposed to how many users who come from the same channel eventually place an order.

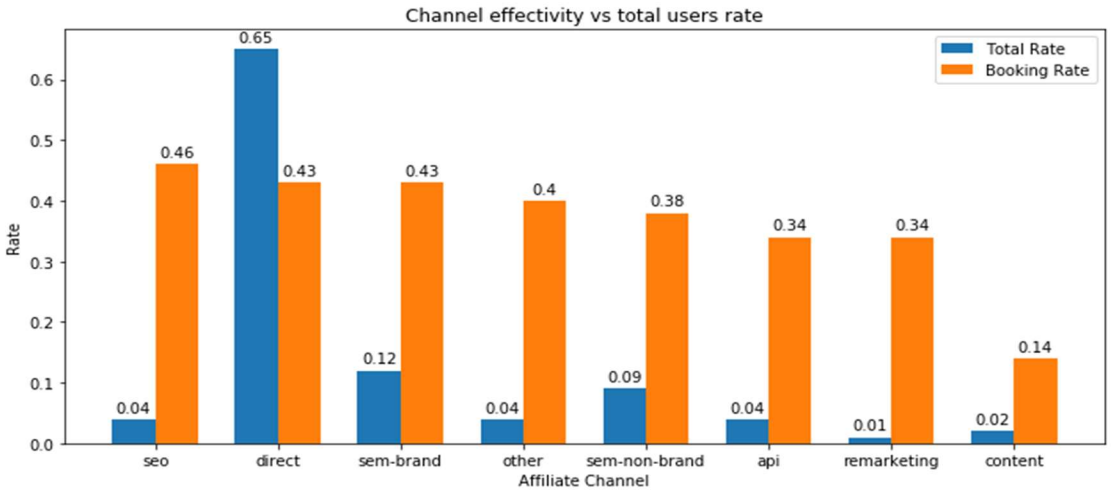


Figure 2.1 Marketing channel efficiency.

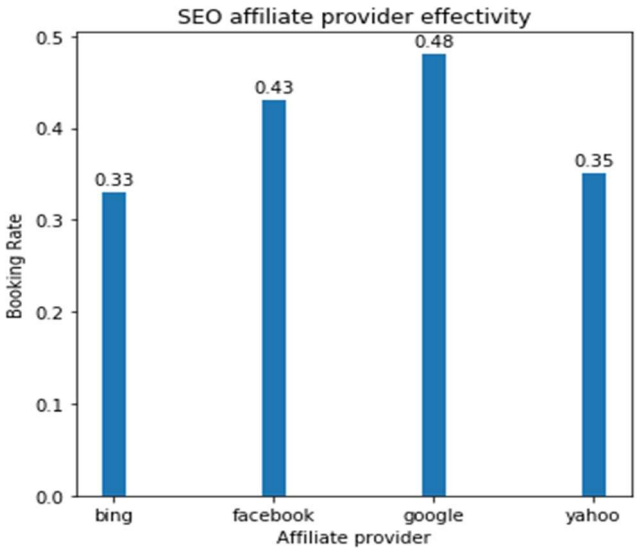
As we can see direct channel is the most popular.
Direct traffic can include visits that result from typing the URL directly into a browser, as the simple definition suggests. It's a channel with no marketing costs.
With 65% of the total traffic, users make reservations in 43% of cases.
However, we can see the high potential of the other two channels: SEO and SEM
So, the SEO channel through which only 4% of new users come has 46% of the effectiveness of the first booking. The SEM channel also effective with up to 43% of first booking.

2.1.1 SEO and SEM drilldown

How to define when effective and when not? In the **short term**, less than 50% of users who came through a paid marketing channel to place their first order may not seem effective, but in the **long-term** a new user who came through a paid channel and placed a reservation will become a regular user and the benefit will be greater.

To get a clearer picture of the affiliate channel, we will explore existing information about affiliate providers both SEO and SEM channels.

* Note: In order to neutralize accidental providers, let's filter out that we will only review providers that have expressed at least 5% of new users.



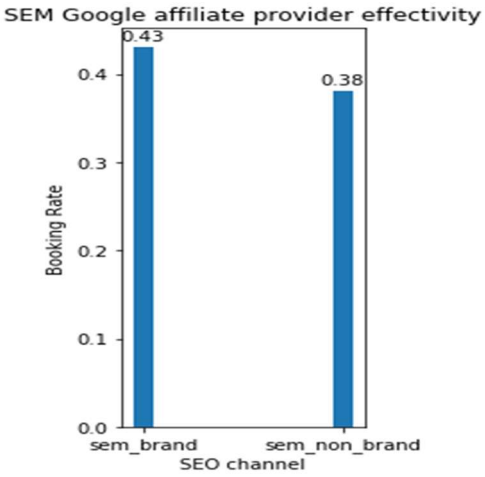


Figure 2.2 Drilldown SEO channel.
Figure 2.3 Drilldown SEM channel (SEM-brand and SEM-non-brand).

We can see on the SEO channel that Google (48%) and Facebook (43%) are the most effective providers over others.
For SEM, there is only Google provider that is effective. In this case, SEM-brand is a more efficient approach to recruiting new users who make a first booking.
In addition, given that SEO and SEM channels are paid in comparison to direct channels, a high order level with low current levels of participation makes these channels potentially effective in the long run to attract quality new users.

Section 5 Business Insights will provide recommendations on the results of analysis by paid marketing.

2.2 New users Apps and Devices.

When we look at how a new user enters the site for the first time, one can understand in which environment to invest in advertising or which website version is more effective.

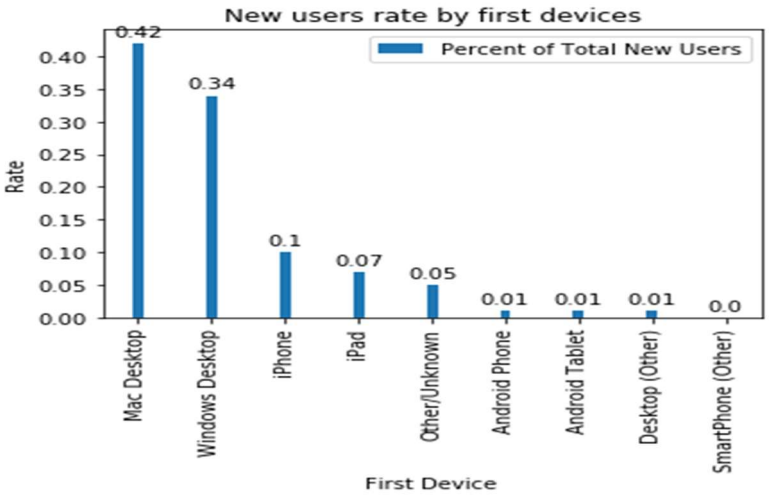
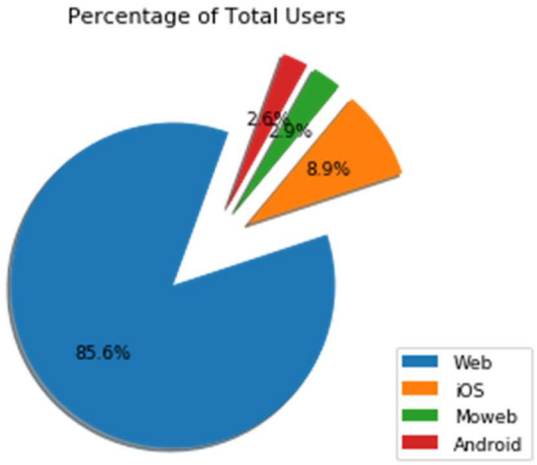


Figure 2.4 Distribution between first devices by number new users.



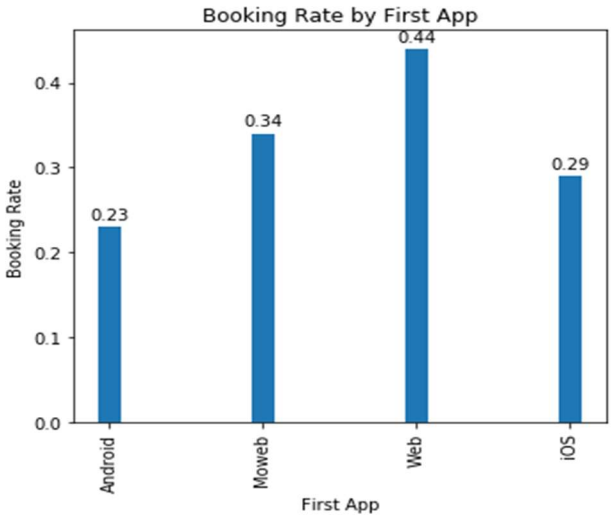


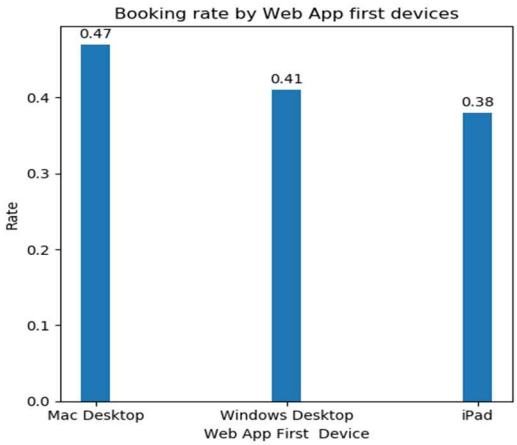
Figure 2.5 Distribution between first Apps.

Figure 2.6 Effectivity by first App.

Most (85%) new users log in via web site version and around 15 % via mobile site (App or Mobile) version.
If you look at application performance, you can see a standard website with 44% efficiency and a mobile application counter with an efficiency of about 30%.
It can be concluded that the mobile application is less known and does not attract new users due to its inconvenience or lack of refinement.

2.2.1 Apps drilldown.

Now that we see which way users come and which devices they prefer, we can test how effective they are to us in terms of the percentage of users who end up booking
In order to neutralize accidental devices, we will only review devices that have expressed at least 5% of users.



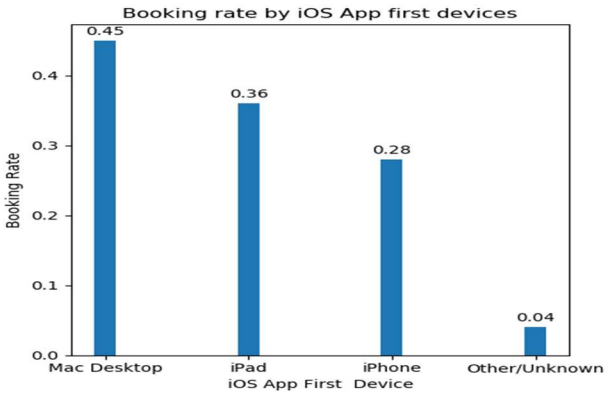
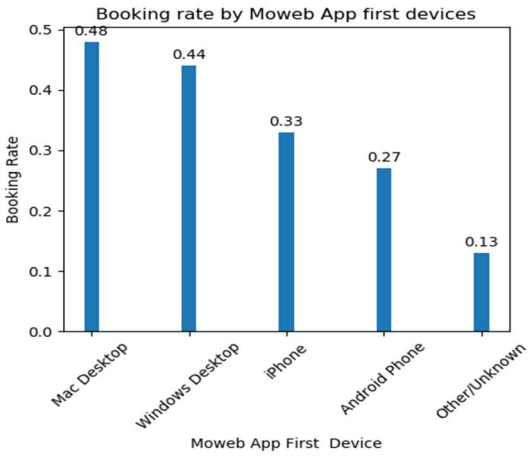


Figure 2.7 Efficiency rate distribution between Web App first devices.
Figure 2.8 Efficiency rate distribution between iOS App first devices.



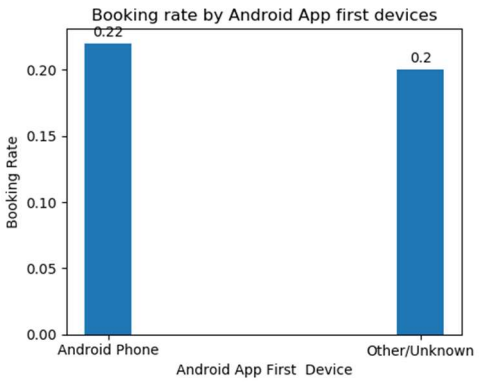


Figure 2.9 Efficiency rate distribution between Moweb App first devices.
Figure 2.10 Efficiency rate distribution between Android App first devices.

The low effectivity of mobile devices (iPhone, Android Phone) compared to home computers or notebooks (Mac, Windows) can be explained by the fact that users who are more inclined to make a booking prefer to do this through a computer, because web apps is more convenient.
We can also notice that even when entering from mobile devices, new users prefer to make their first booking via the web version of the site.
It may also mean that users make the first reservation more planned, and not spontaneously.

Section 5 Business Insights will provide recommendations on the results of analysis by Apps and Devices.

3. Site Analysis

The place where the user places the order is the site. A convenient site makes the user spend more time on the site and more likely to place an order!

3.1 Language

Airbnb has announced it is doubling the number of languages the platform supports—from 31 to 62. With 62 languages now supported, Airbnb says it will make the platform even more accessible to more than 4 billion native speakers. And not only will support for the new languages make it easier for people to book rooms on Airbnb, but the company points out that it will also enable economic opportunities in new regions for hosts who can now list properties in their native language.

The percent of users using English site version is **97%**
The percent of users using Other (non-English) site version is **3%**
* Reminder: All the users in this dataset are from the USA.
According CIA (Central Intelligence Agency) English only 78.2%, Spanish 13.4%, Chinese 1.1%, other 7.3% (2017 est.)
https://www.cia.gov/library/publications/the-world-factbook/fields/402.html#US

It is clearly seen how not used the potential users from the United States for which English is not native. Them according to statistics in more 20%, but only 3% use a non-English version.
We can also notice that among them the percentage of the first reservation is lower than among users of the English version.
This is probably due to the low popularity and inconvenience of non-English versions of the site.

Figure 3.1 Booking rate by site language version.

3.2 Sessions

The session dataset contains a sample of new of users. Using session analysis, we can understand how long new users spend time on the site, depending on their next steps. Next steps as a booking or not, and if so, to which country.
For convenience, we will use a scatter plot:

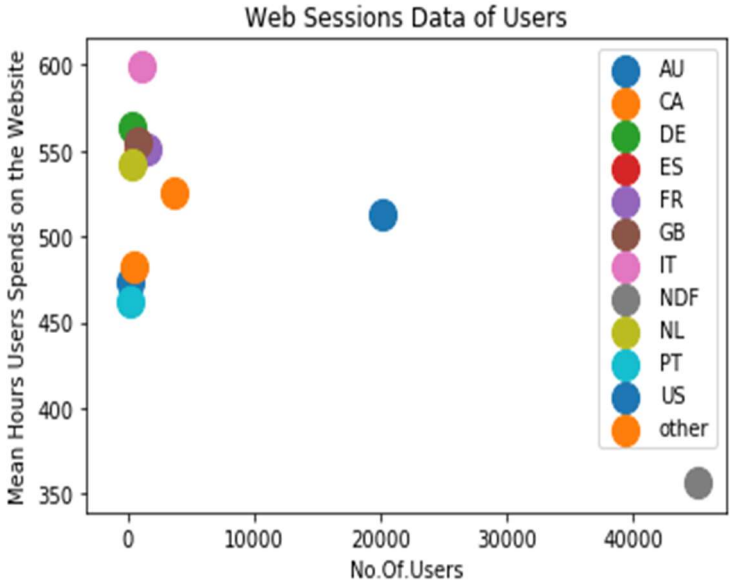


Figure 3.2 Web Sessions Data of Users.

We can see that those who did not make a booking at all spend much less time on the site than those who made the booking.

And if we look even deeper, we can see that the percentage of bookings on the site is reduced depending on the time that the user spends on the site.

Figure 3.3 clearly shows that with the increase in hours spent by a new user on the site, the likelihood that he will place an order increase.

Figure 3.2.1 The relationship between the likelihood of booking and time spent on the site.

We can say that new user who in a relatively short time do not find the necessary information leave the site without placing an order.
Thus, as already mentioned, it is important to quickly interest a new user on the site in order to increase the time spent on the site and, as a result, increase the likelihood of his first booking.

Section 5 Business Insights will provide recommendations on the results of Site Analysis.

4. Trends Analysis.

In this section, the behavior of new users will be analyzed by time periods. After analysis, we can understand trends, identify exceptions and improve resource allocation.

4.1 Years.

In order to analyze the trend by year, the same time intervals were taken 01/01/2010-30/06/2014.
To understand the quality of new users in terms of booking rate, two trends can be compared. Figure 4.1 shows that in 2010 more than half of new users place an order. Can be seen that before 2013, the number of new users increased dramatically, but the order level began to decline.
From this it can be concluded that in 2010-2013 many fewer quality users joined.
Probably the reason for this trend is that the site's visibility was good compared to the poor quality of the site.
As of 2013, can see that the number of new users remains the same as the level of orders has continued to decline.
Also be seen the relationship between the time from the first visit to the first booking. If in 2010 a new user placed the first order on average 27 days after the first visit with a high order level, then by 2013 this figure will increase to 47 days. Some improvement was observed in 2013.
The clear relationship between the increase in the number of days from the user's first impression to the first order and the order level is clear.
Newer users find it difficult to find relevant offers or information on the site.

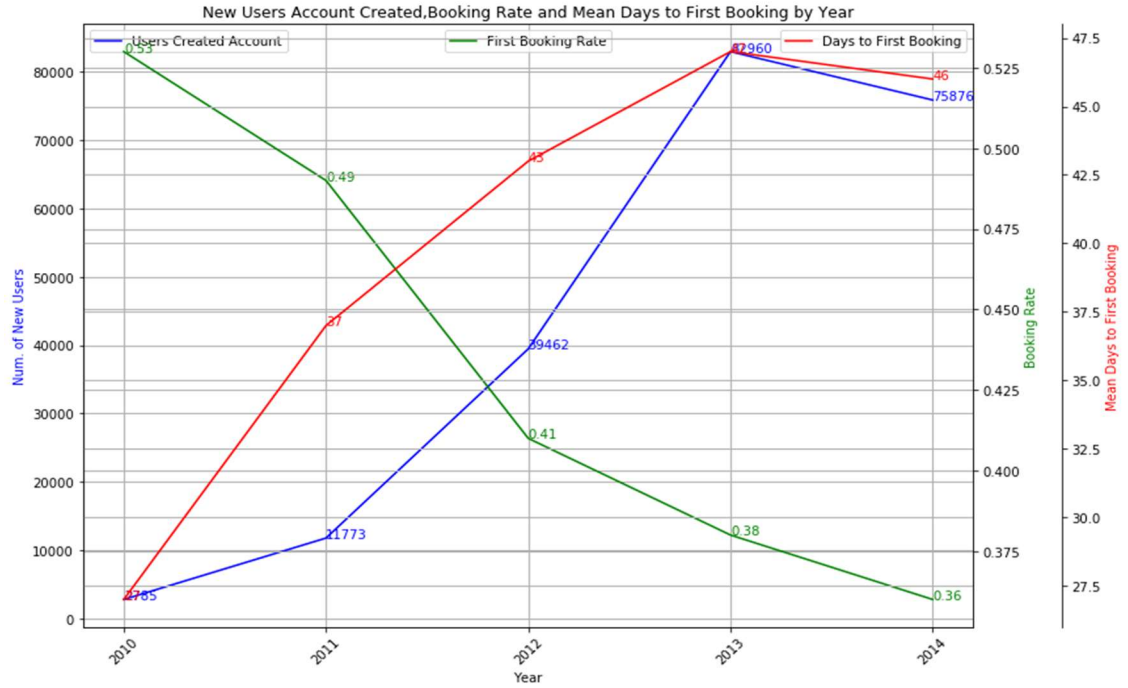


Figure 4.1 Trend New Users Account Created, Booking Rate and Mean Days to First Booking by Year. (01/01/2010-30/06/2014)

To check why there was a decrease in the quality of new users with an increase in their number, consider figures 4.2 and 4.3. You can immediately see a lot of new members joining in who do not indicate their age. However, without worrying about the quantity, their quality has fallen sharply since 2011, and despite the general tendency to improve since 2013, there is no improvement in this group. Only 19% booking rate by the end of 2014.

Of the groups that indicate their age data, groups 25-34 and 35-44 differ, that is, middle-aged people. With a general increase in quantity, their quality is relatively stable. Also, the youngest group between 15-24 lost about 25% in quality by 2012 with slight improvement thereafter.

The inclusion of new users from the older groups 55-64 and 65+ has hardly increased and maintains a low trend.

It can be noted that quantitative and qualitative indicators for age groups coincide with the conclusions of the 1st section (analysis based on personal data).

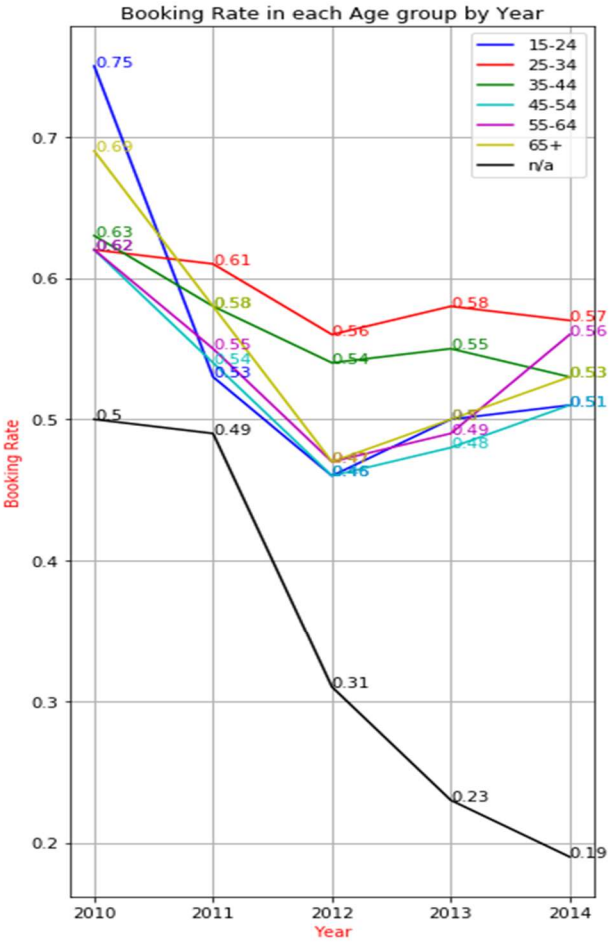


Figure 4.2. Booking rate by years drilldown.

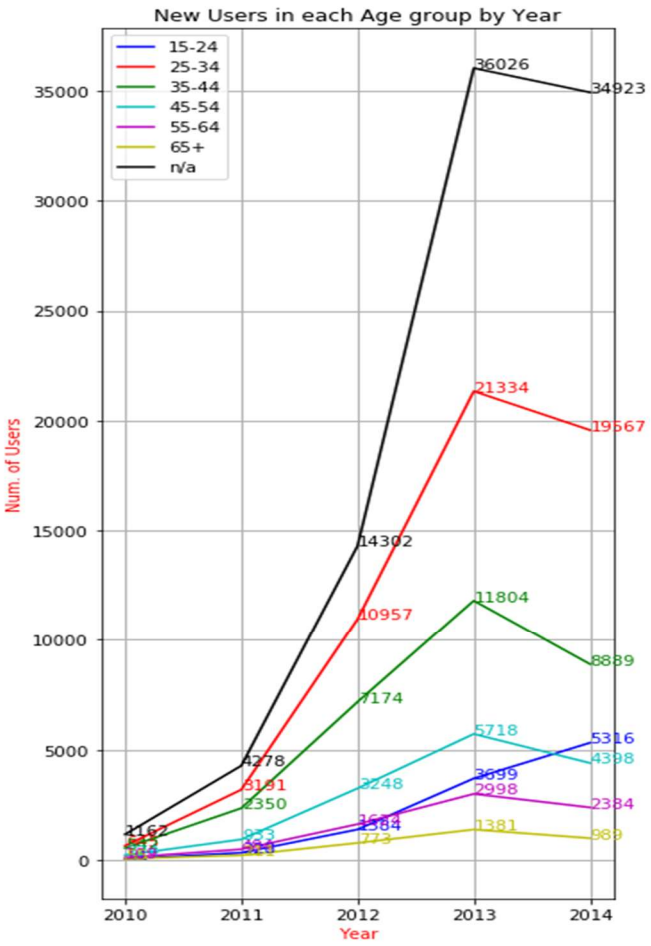


Figure 4.3 QTY New users by years drilldown

4.2 Months

Analyzing trends by month, it is difficult to draw conclusions about the quality of new users, but you can understand the behavior of users by month and season.

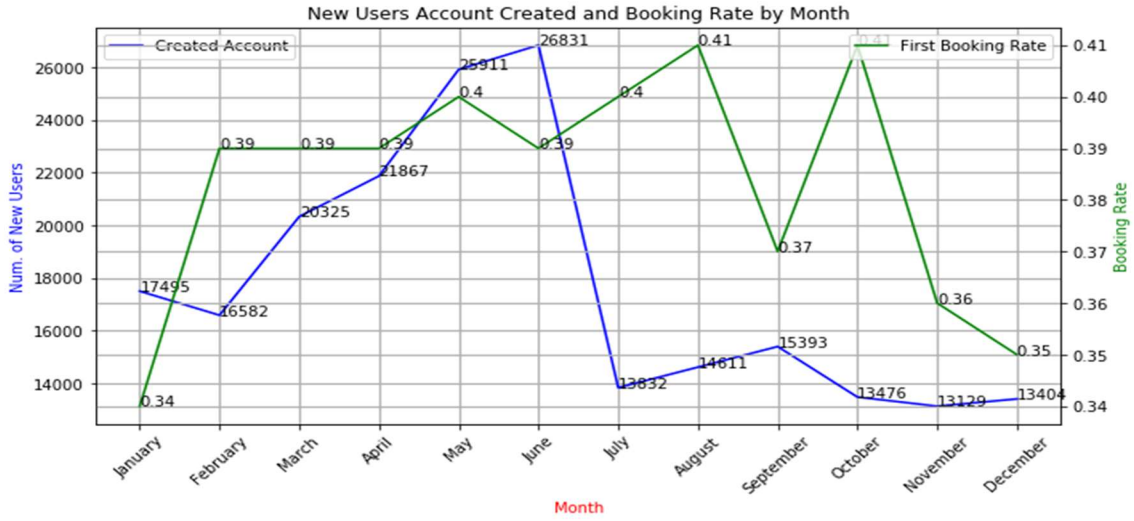


Figure 4.4 Trend new accounts created and booking rate by months

Starting in February (Figure 4.4) the new account creation trend is on the rise by the end of June with a sharp decline starting in June and reaching a low in early winter.

In contrast, the booking rate is on a sharp rise in June and at a relatively high level until October. As of November, the order rate remained virtually unchanged.
These trends show us that most new users place their first booking at the beginning of the summer, which means they are most likely to be booked for vacation rather than business.
In addition, it can be concluded that most bookings are made close to the date of departure and not several months ago.

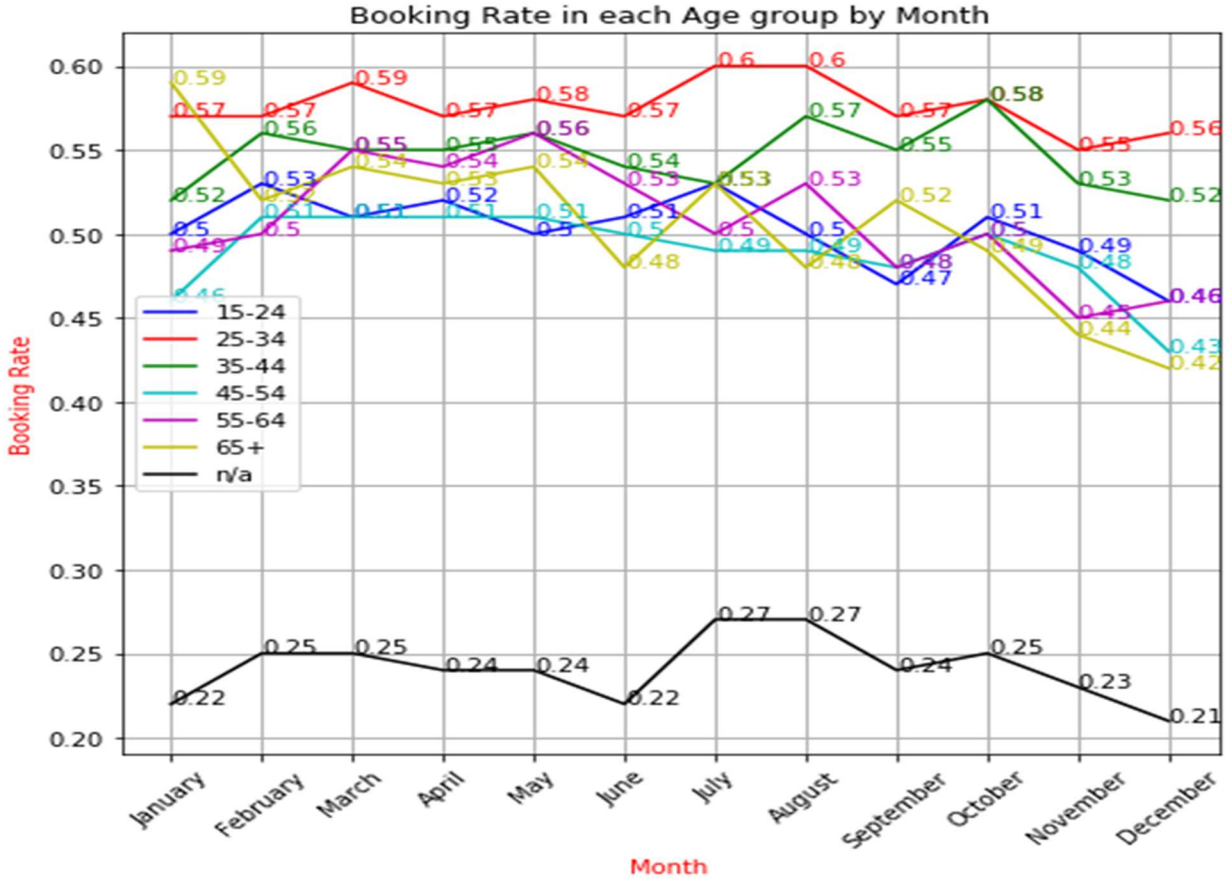


Figure 4.4 Booking rate by months and age group.

Figure 4.4 shows that there are no significant differences in the behavior of different age groups. The reservation level in the summer is slightly higher than in other months. In each group, similar trends persist throughout the year. Thus, it can be assumed that the level of booking in the summer months also increases due to a decrease in the total number of open accounts.

4.3 Weekday.

Can be seen notice trends in the behavior of new users also by day of the week. The trend chart shown in Figure 4.5 shows that most new accounts open on business days, maximum on Tuesday and at the weekend.
At the same time, orders are higher on Friday, that is, on the last business day.
It can be concluded that new users prefer to open accounts and possibly browse the site for a week even during business hours and make their first reservation at home.

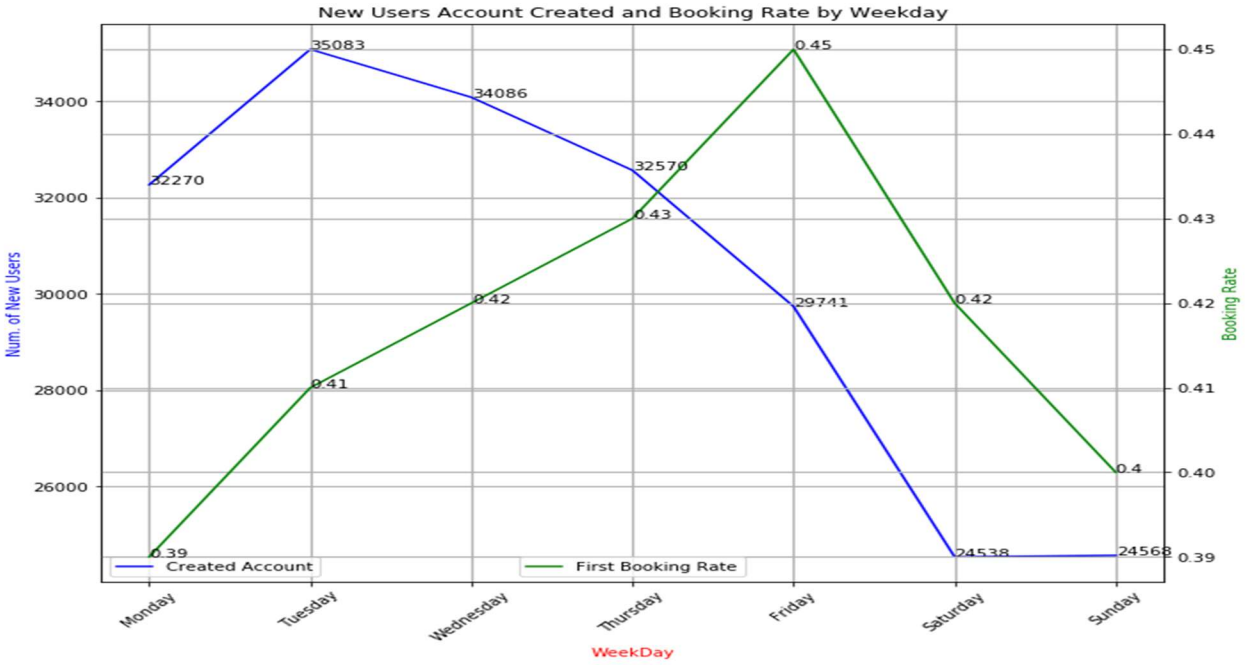


Figure 4.5 Trend new accounts created and booking rate by weekday.

5.Business Insights.

To attract more higher-quality new users and efficient use of advertising resources based on analysis, it is recommended:

1. Add More services to attract more people, the bigger portion of the users don't do any reservations at all after signing up.
2. Make mandatory the filling in of personal data on age and gender at registration.
3. Add more advertising to attract younger under 24 and older people 55+.
4. Advertise more online, as most subscriptions come directly, and not through any marketing agency. Use more high-performance channels SEO and SEM-brand.
5. Prioritize the mobile application in general and for iOS products in the above privacy.
6. Make the website more attractive and easier to use to increase the number of bookings by the time spent on the site.
7. Advertise to attract more non-English-speaking users improve the non-English version of the site.
8. Examine the behavior of users booking to France and Germany because they spend a lot of time on the site compared to other users.
9. Change the use of the policies that were used in late 2013 and early 2014, as they show that the growth in the number of created accounts is not growing.
10. Create offers for quick reservations to reduce the number of days from first visit to first reservation.
11. Give discounts / offers in the second half of the year to increase booking rate.
12. Increasing attractiveness for business orders. Most users belong to the age group 25-34 and have more reservations since the beginning of summer, this may indicate that the service is more often used for tourism purposes.