Well hello, as you already know my name is Paweł Grzeszczyk and I would like to present you my case study considering retention rate analysis. In the beginning I would like to mention that retention rate is defined as the percentage of users who visited the platform on a given day and returned within the next 3 days.

***NEXT***

Here you can see a brief agenda. Main purpose of this presentation is to

I have asked myself a few questions and tried to answer them. These questions are:

* How big is retention rate in a given timeframe?
* What is influencing retention rate? That is categories, sub-categories and channels

And I have also conducted a case study with a hypothesis: ios users have higher retention rate than mobile web users.

***NEXT***

At the beginning, I would like to briefly show you an executable summary with my main insights:

* Average total retention rate is on surprisingly high level (almost 70%),
* Retention rate is relatively stable during this period of time,

Confirmation for these two points can be seen on the graph to the right where bars are showing number of users who visited the platform on a given day and a number of users who returned within next 3 days. Scale for them is on the right side of the graph. A line is showing a retention rate and it’s scale is on the left side of the graph.

* Apartments is a category with the biggest number of visits and it came out to be the one with the highest retention rate (68,87%),
* Sub-categories with the highest average retention rate are Apartments for Rent (65,32%), Apartments for Sale (63,38%) and Houses for Sale (62,85%),
* Android is the channel from which the vast majority of users come and it also has the biggest retention rate (80.27%).

***NEXT***

Now let’s deep dive into the daily total retention rate. It has to be said that the data I had available concern period from 1st of June 2021 to 10th of June 2021, so because we’re interested in a 3 day retention rate I could calculate it up to 7th of June 2021.

To the right you can see a plot that you have already seen on the pervious slide and a table presenting daily retention rate values. Based on it a few conclusions can be made, that is:

* Both retention level and number of users who visited a platform on a given day are stable in this period of time,
* Average retention rate is equal to almost 70%, once more I think it’s worth mentioning how high it is. In my opinion a 3-day retention rate of 40% is a fairly good result, whereas 70% seems to be good for 5 or maybe even 7 day.
* Retention rate reached its peak on 04 of June 2021, reaching the value of 72,49% and it’s 2,5 percentage points above the average

***NEXT***

When it comes to analysing category we have to ask ourselves a question: Are users visiting only one category? Based on the pie plot we can tell that half of the users who visited a platform in a given period searched only for one category. As the number of categories visited by user increases, the percentage share decreases. In the end all 6 categories were visited by only 0.3% of users.

***NEXT***

Knowing that we can analyse retention rates in categories. There are 6 main categories and they are divider further for rent and sale. In the table to the right you can see main sub-categories that I’ve chosen among all of them based on that they were most popular in the dataset.

In order to perform analysis accurately it has to be split into rent and sale.

* In Rent:
  + Users are most likely to come back when they are looking for Apartments (65.32%),
  + Lowest retention rate is for Commercials - there is also a very few of them in the dataset.
* Sale:
  + Apartments and Houses have almost equal retention rate (~63%), what makes them categories with the highest score when it comes to sale,
  + Users are least likely to come back in Commercials and Plots.

Biggest retention rate in Apartments and Houses may be caused by many factors:

* Offer for these categories is known to be dynamically changing, especially nowadays, so people may feel need to come back more often and check whether something new appeared on the market.

***NEXT***

Next important factor influencing retention rate is channel form which user is coming from. There are 4 channels available in a dataset: android, ios, mobile\_web and desktop\_web.

We can already see that android is the most popular channel in the dataset (62% of users came from there), next is ios – almost 20% and desktop\_web and mobile\_web sum up to nearly 20%. Android and ios apps have way higher retention rates than web versions for mobile and desktop. Main reason for this may be that using them is just more convenient, you always have access to them in your pocket. What is more it may be easier to stay logged on the mobile app instead of having to log on the website over and over again, so it’s easier for us to track users behaviour.

We can also pose a hypothesis that the retention rate is so high because majority of users come from apps, where retention rate is expected to be higher than for web.

***NEXT***

Now I’ll walk you through a case study. My first step was to pose a null and an alternative hypothesis:

Null hypothesis: ios users have smaller or equal retention rate than mobile web users.

Alternate hypothesis: ios users have higher retention rate than mobile web users.

In the tables we can see the data for both channels mentioned in the hypothesis. We can already see that average retention rate for ios is almost two times bigger than retention rate for mobile web. Based on that sample we can already expect that we’ll be rejecting null hypothesis.

***NEXT***

I already knew that I want to perform a student’s t-test for two independent samples and in order to do that I had to assure that two conditions are met:

1. Both samples are normally distributed.

To check it I performed Shapiro test and for both samples I did not reject null hypothesis saying that they are normally distributed.

1. Variances of these samples are homogenous (that is, they are similar).

Here I performed Levene’s test where the result based on p-value showed that there is again not enough evidence to reject null hypothesis saying that variances are equal.

With both conditions met I could perform the Student’s t-test for two independent samples.

Because we’re interested in checking whether ios retention rate is not only not equal, but greater we have to check p-value divided by two.

In the end, p-value basically equal to 0 proved that at this significance level there is enough evidence to conclude that the average retention rate for ios users is higher than average retention rate for mobile web users.

What concludes the case study.

Thank you very much.

**(If data was not normally distributed I would use a Kruskal-Wallis-Test)**

**(If I had more samples I would have to use ANOVA)**

**What might have causes this 70%?**

Big share of mobiles may cause bigger retention rate.

**KPI – Key Performing Indicators.**

**KPI’s for Otodom.**

For otodom. I think it very much depends on perspective:

For the company, revenue – we want to focus on stuff that generates us money: that is posting ads and promoting ads in many ways, so we may want to show the sellers that it’s worth to promote ads by calculating metrics showing that it’s actually worth it to promote ads for example share of users who click on promoted and not promoted ads. Metrics: seller retention, ad posting frequency, ad quality and suggestions to improve, reach of ads, we may want to prove that we’re the leaders on the market so it’s worth for seller to pick us instead of our competition.

This point actually applies to the seekers as well, we want them to be aware that our Real Estate database is actually the largest, the cleanest (that is there are no fake/scam ads) and it’s just nice and easy to use. Here it’s important to stay in touch with our customers, be aware of their needs and adjust our site to them. When it comes to metrics there are multiple of them: 3, 5, 7 days retention rate, share of users who text to sellers, share of users who click “show the phone number” button, percentage of users who search a specific feature, Daily Active Users, Monthly Active Users, User Engagement: Time spent watching ads, number of ads clicked by users, based on user behaviour we can suggest optimal real estates for them, key demographics, devices used