Product ideas:

1. Perhaps redirect people with higher likelihood of not showing up to end of day (show them only certain slots before they even book appointment)
2. Have a waitlist system – ask people who are willing to waitlist for certain days at end of day (given that we can redirect people who are likely to not show up at end of day, or at the same block in the day)
3. Increase utilization!

Comments on plots (24th apr):

1. For number of patients per date, if the conclusion is that Sundays are outliers, should we also plot using the # of patients VS week of day? -> yes, but the time analysis is done later in the eda

2. “After 60 years old, people begin to die, and hence there are less patients. We see on the representation of the expected living time of people in Brazil.” -> brazil’s life expectancy is 73.8 on average. Don’t know if we want to say this? -> I am not saying that the life expectancy is 60 years old but that it’s comprised between 60 and 100, I will clarify

3. for plot “showing number of patients depending showing up depending on age”, is the y axis the ratio or number of patients? If it’s the latter, is it in the thousands? Also, I’d suggest that we do the “violin graphs” that Raphael talked about – **I think box plots give us the information but it takes time to interpret; the violin graphs will let us see intuitively where the weight of the points are ->** the name of the axis were reversed, I think that’s why you found it difficult to interpret the graph

4. The plot for ratio vs age is really good in terms of showing us the trend, but shall we cut off the x-upper limit at 90? Because beyond that it’s just erratic -> yes, but I feel that it’s better to show it and explain why it’s not significant and why we should interpret only the first part of the graph.

5. I think the “number of patients depending showing up depending on age and gender”, with 2 colors graph, looks nice but takes a little more time to interpret. Because the “ratio” is dependent on the overall height of the graph, and just looking at height of the red section only tells us absolute number of no-shows. **This graph might not tell us a lot about the ratio just by inspection**

If we want to keep the graph, maybe we can say that we see the ratio for men not showing up increases gradually when they hit 20+ years old till 60, then falls? But for women, the # of no-shows spike when they hit 20+ years old and then decreases sharply once they hit 60 (**working age probably?**) -> I don’t agree on this point

6. Need to update the second set of graphs for health conditions to have “count” instead of “ratio” as Y-axis. But on this health conditions analysis section, I think it’s worthy to know the show/no-show ratio of people WITH health conditions? I can’t interpret by looking at the graphs on first glance. **Because, even though there might be very small numbers of people who are sick, but I think if the ratio of no-shows are high, it can be a strong indicator.** -> Let’s discuss about that, I am not sure I get your point

7. Also shall we consider the possibility that people who have diabetes are likely to have hypertension too? I think there might be some correlation there. A joint probability thing might be another added layer of analysis. -> No, if we do that we’ll have to look for correlations for all variables, instead we could try to do some feature selection (forward feature selection, .. )