Project Analysis

I used the No-show Appointments document to analyze. I had three main questions to investigate:

1- Is there a correlation between the age of the patient and date of appointment?

   For this question, I added a column called “Date Difference” where it calculates the difference between the date of booking and the date of appointment. Then, I used a scatter plot to show the ages of the patients and the date difference. Based on this plot, there is a data quality issue where some patients are 0 years old. There is also a strong correlation between age and days of appointment, where older patients are more likely to book an appointment earlier in advance than younger patients.

2- Is there a correlation between the date of booking the appointment and patients arrival?

   For this question, I first calculated the percentage of the patient arrival for each date difference, to see how many patients who booked the appointment on the same day showed up to the appointment vs days later, I named it “Arrival Percentage”. Then, I used seaborn to make a line plot showing the desired data. From the plot, we can see that approximately 33% of the patients who arrived to their appointments have booked the appointment on the same day, which is the largest percentage of the patients in this data. This means that if a patient books their appointment in the same day they are more likely to show up.

3- Does receiving an SMS about the appointment factor into the patient showing up for the appointment?

   For this question, I used “groupby” to see how many patients received the SMS and how many did not. Then, I used it again to see how many patients who received the SMS showed up to the appointment and how many did not vs how many patients who did not receive the SMS showed up to the appointment and how many did not. From the data, we can see that the number of patients who did not receive an SMS is more than the number of patients who did. Furthermore, we can see that patients are more likely to show-up for the appointment if they received an SMS message.

Data Wrangling:

I turned the columns “ScheduledDay” and “AppointmentDay” to datetime.

Sources I used:

StackOverflow:

https://stackoverflow.com/questions/36762199/using-pandas-value-counts-and-matplotlib

https://stackoverflow.com/questions/332289/how-do-you-change-the-size-of-figures-drawn-with-matplotlib

https://stackoverflow.com/questions/33271098/python-get-a-frequency-count-based-on-two-columns-variables-in-pandas-datafra

https://stackoverflow.com/questions/55432724/how-can-i-calculate-number-of-days-between-two-dates

https://stackoverflow.com/questions/52385428/plot-point-markers-and-lines-in-different-hues-but-the-same-style-with-seaborn

https://stackoverflow.com/questions/32908315/could-not-interpret-input-error-with-seaborn-when-plotting-groupbys

https://stackoverflow.com/questions/19384532/get-statistics-for-each-group-such-as-count-mean-etc-using-pandas-groupby

Real Python:

https://realpython.com/visualizing-python-plt-scatter/

GeeksforGeeks:

https://www.geeksforgeeks.org/adding-new-column-to-existing-dataframe-in-pandas/

https://www.geeksforgeeks.org/how-to-combine-two-dataframe-in-python-pandas/