Analysis of Patient data

Overview of the Dataset:

PatientID - Unique identifier for a patient

AppointmentID - Generated ID for every appointment scheduled with the front office

Gender - Gender of the patient

ScheduledDay - Day the person calls front office for an appointment

AppointmentDay – Day of appointment

Age - Age of the patient

Neighborhood

Hypertension – binary for Hypertension

Diabetes- Binary for diabetes

Alcoholism- Binary for Alcoholism

Handicap- Binary (if the patient is handicapped)

SMS_received - binary for SMS sent for followup

No-show- Yes/ no if the patient has shown up for the appointment

Data Preprocessing:

Loaded the data to excel for preprocessing, analysis, and reporting

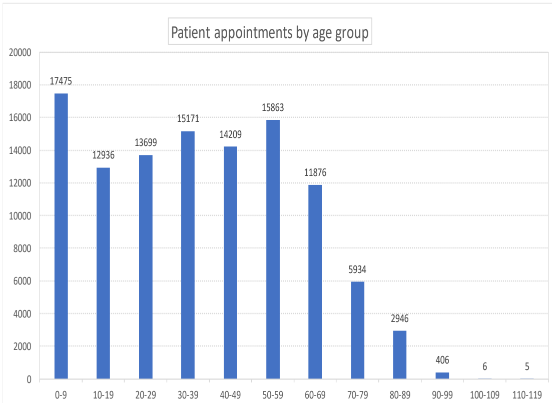
Steps used in preprocessing:

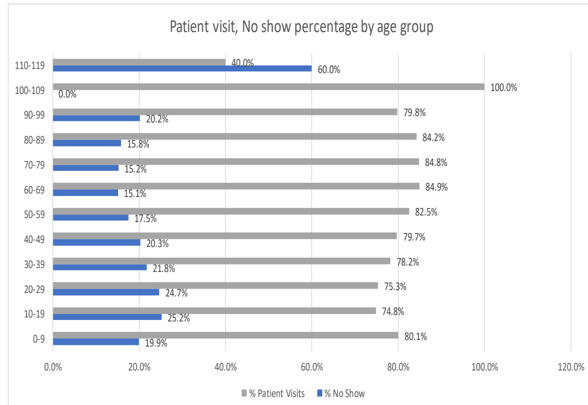
- Split Scheduled Day column using text to column functionality. Created scheduled day and scheduled time column for further analysis
- Split Appointment column using text to column functionality. Created appointment day and appointment time column for further analysis
- Used if statement to change no show column to 0 and 1
- Created a new column "Date difference" to calculate the difference between scheduled and the appointment day

Creating a Pivot table for analyses:

1) Appointment to no show ratio:
   Inference: Patients with most follow-ups tend to have few to no "No shows"
2) Age groups:
   Created age groups to analyze appointments booked by different age groups
   Inference: 0-9 ages book most of the appointments (Pediatrics) followed by 50-59 age groups
   This gives an interesting view on age groups. As children, one can assume that boys are being taken to the doctor by their moms or dads, and don't have much choice about showing up. In adulthood, however, there's plenty of patriarchal stigma about needing help or getting care, so they skip appointments. Then once people reach old age, it seems to even out again.

Age	Number of appointments	No Show	Patient Visits	% No Show	% Patient Visits
0-9	17475	3484	13991	19.9%	80.1%
10-19	12936	3257	9679	25.2%	74.8%
20-29	13699	3380	10319	24.7%	75.3%
30-39	15171	3300	11871	21.8%	78.2%
40-49	14209	2880	11329	20.3%	79.7%
50-59	15863	2776	13087	17.5%	82.5%
60-69	11876	1790	10086	15.1%	84.9%
70-79	5934	902	5032	15.2%	84.8%
80-89	2946	465	2481	15.8%	84.2%
90-99	406	82	324	20.2%	79.8%
100-109	6	0	6	0.0%	100.0%
110-119	5	3	2	60.0%	40.0%

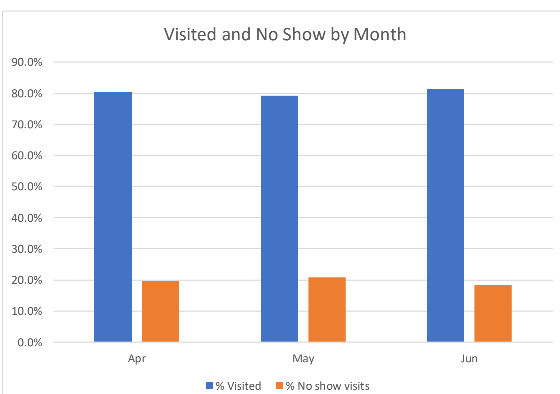


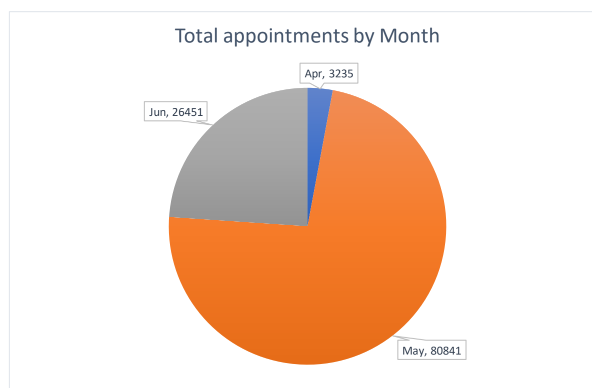


3) Monthly breakdown

In the given data most of the appointments were book in May. Data showed the least number of appointment booking in April, this may be due to imbalance in the dataset/ nature of the data shown in the dataset

Month	Patients visited	No Show Visits	Grand Total	% Visited	% No show visits
Apr	2602	633	3235	80.4%	19.6%
May	64037	16804	80841	79.2%	20.8%
Jun	21569	4882	26451	81.5%	18.5%





4) Patient population analysis

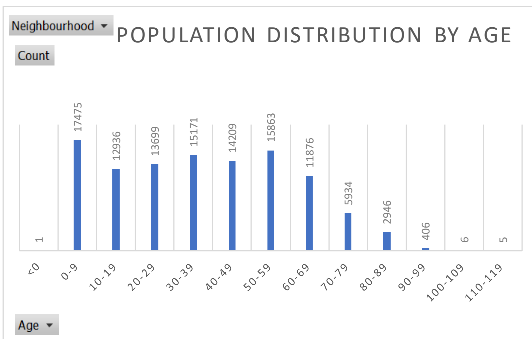
A deeper dive into the demographics of the patient

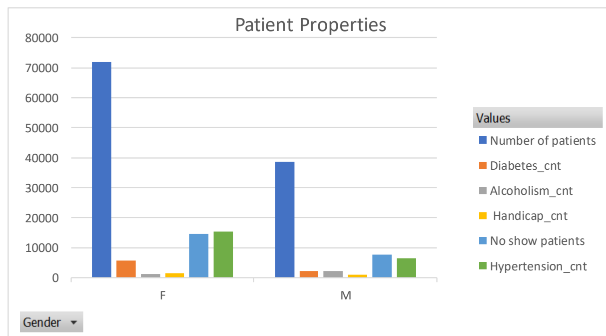
Certain health behaviors, such as alcohol use or smoking, but a lot of these attributes are just health conditions, like diabetes. There's a need to think more about what the expected impact of past behavior might be on no-shows, versus the impact of existing illness. Also, each illness existing is going to be different and has a different implication for appointment-keeping. (Perhaps visits to a doctor are more unpleasant or uncomfortable for certain illnesses?)

Neighbourhood (All)

Row Labels	Count
<0	1
0-9	17475
10-19	12936
20-29	13699
30-39	15171
40-49	14209
50-59	15863
60-69	11876
70-79	5934
80-89	2946
90-99	406
100-109	6
110-119	5
Grand Total	**110527**

Row Labels	Number of patients	Diabetes_cnt	Alcoholism_cnt	Handicap_cnt	No show patients	Hypertension_cnt
F	71840	5606	1223	1404	14594	15339
M	38687	2337	2137	1055	7725	6462
Grand Total	**110527**	**7943**	**3360**	**2459**	**22319**	**21801**





This is an exploratory analysis of the given data.

Finally,

Emphasizing on day related data might give more insights into no showcases and help the front office operate efficiently with more follow-ups, less no shows