Healthcare System

Exploratory-Data-Analysis

Submitted By:
Sachin Rajbhar & Ravindra Kanojiya

Submitted To:
Ms. Bushra Jamal Ma’am

Contents

1. Introduction
2. Library Files
3. Process
4. Findings
5. References

Introduction

In the context of healthcare, EDA provides a valuable lens to examine factors such as patient demographics, disease prevalence, treatment outcomes, healthcare utilization, and the effectiveness of interventions. By examining these factors, EDA enables healthcare professionals to identify risk factors, develop preventive strategies, optimize treatment protocols, and allocate resources efficiently.

Library Files

1. Pandas: Used for data manipulation & Analysis
2. Numpy: For Scientific computation
3. Datetime: A datetime module
4. Matplotlib: Create data Visualisation
5. Seaborn: For Visual themes (Based on matplotlib)

Process

1. Reading the dataset ‘Data.csv’ (using pandas)

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hipertension	Diabetes	Alcoholism	Handcap
0	2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z	62	JARDIM DA PENHA	0	1	0	0	0
1	5.589978e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z	56	JARDIM DA PENHA	0	0	0	0	0
2	4.262962e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z	62	MATA DA PRAIA	0	0	0	0	0
3	8.679512e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z	8	PONTAL DE CAMBURI	0	0	0	0	0
4	8.841186e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z	56	JARDIM DA PENHA	0	1	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
110522	2.572134e+12	5651768	F	2016-05-[illegible]	2016-06-[illegible]	56	MARIA ORTIZ	0	0	0	0	0

(DATASET LINK ***https://github.com/pik1989/blob/main/Data.csv)

2. Perform specific functions like info() ,shape, describe, columns, value_counts()

```
In [26]: new_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110527 entries, 0 to 110526
Data columns (total 13 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   Gender          110527 non-null  object
 1   ScheduledDay    110527 non-null  datetime64[ns]
 2   AppointmentDay  110527 non-null  datetime64[ns]
 3   Age             110527 non-null  int64
 4   Scholarship     110527 non-null  int64
 5   Hypertension    110527 non-null  int64
 6   Diabetes        110527 non-null  int64
 7   Alcoholism      110527 non-null  int64
 8   Handicap        110527 non-null  int64
 9   SMSReceived     110527 non-null  int64
 10  NoShow          110527 non-null  object
 11  sch_weekday     110527 non-null  int64
 12  app_weekday     110527 non-null  int64
dtypes: datetime64[ns](2), int64(9), object(2)
memory usage: 11.0+ MB
```

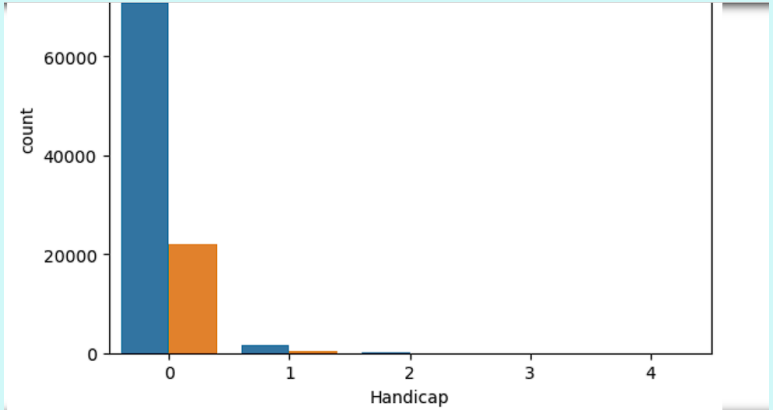
3. Data Cleaning

As we don't have any null records, there's no data cleaning required

4. Create a copy of base data for manipulation & processing

```
In [25]: new_data = base_data.copy()
```

5. Data Exploration



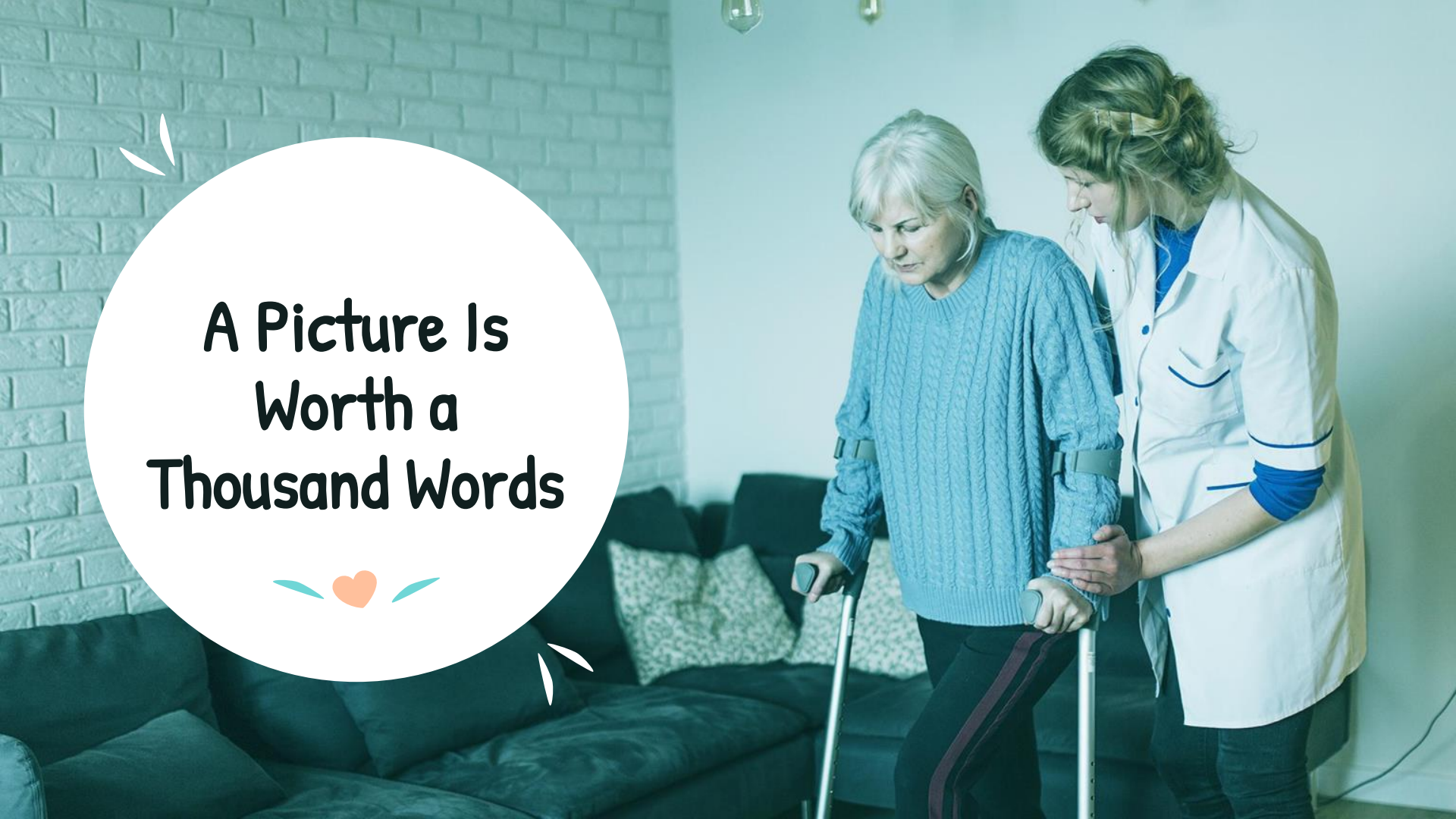
……Contd in jupyter file

Findings

1. Female patients have taken more appointments then male patients
2. Ratio of NoShow and Show is almost equal for age group except Age 0 and Age 1 with 80% show rate for each age group
3. Each Neighbourhood have almost 80% show rate
4. there is no appointments on sunday and on saturday appointments are very less in comparision to other week days

References

- Gosheets
- Wikipedia.org
- Google.com


A Picture Is
Worth a
Thousand Words

Do you have any questions?
sr4205@dseu.ac.in

Thanks