HEALTHCARE DATASET CHECKPOINT

INTRODUCTION

Data analysis is the process of inspecting, cleansing, transforming, and modelling data to uncover useful information, inform conclusions, and support decision-making. It encompasses multiple techniques and approaches and is applied across various fields such as business, science, healthcare, and social sciences.

In today's data-driven world, data analysis plays a key role in enabling evidence-based decisions and enhancing operational efficiency.

In the healthcare sector, data analytics is especially vital. It allows organizations to evaluate and develop practitioners, detect anomalies in scans, and predict disease outbreaks, as highlighted by Harvard Business School. Furthermore, data analytics can reduce costs, enhance business intelligence, and most importantly- support better care decisions for patients.

For instance, data analysis helps track patient health conditions, improve treatment plans for specific medical issues, and identify trends to anticipate future health events. An example is the digital storage of vaccination records, which helps governments identify and reach unvaccinated populations. In the past, such records were stored manually, making retrieval time-consuming and labour-intensive-especially in multi-specialty hospitals where different departments (e.g., Paramedics, Physicians, Nurses, Lab Technicians, and Specialists) operate under varied record-keeping systems.

Digital record-keeping simplifies data retrieval and enables faster access to patient history when needed. Tools such as Excel, SQL, Python assist in data cleansing, transformation, and validation, while visualization tools like Power BI, Tableau, and Pivot Charts in Excel make data more understandable for healthcare decision-makers.

SUMMARY ABOUT THE PROJECT

The healthcare dataset used in this study was sourced from Kaggle.com and contains various patient-related data including: Name, Age, Gender, Doctor, Hospital, Name, Admission & Discharge Dates, Medical Condition, Insurance Provider, Medication, Test Results, Room Number.

Using this dataset, I performed Descriptive Analytics in SQL. In healthcare, descriptive analytics uses historical patient data to derive insights about trends, benchmarks, and care outcomes.

Various steps involved in the process of analysing given dataset with SQL;

1) To verify any actions required for cleaning the data from the provided dataset, tables are examined and columns are updated.
2) To make it easier to read, columns specifically [Name] were changed from irregular cases to correct cases. The [Doctor] column is examined to see if the same case mistake as the [Name] field occurs.
3) To see the available and missing data clearly, null values are examined. Daniel Brooks, the patient, does not have information on his room number, medical condition, medication, discharge date, or test results.

4) Number of patients consulted doctor
    a. Based on gender:

Gender	COUNT_GEN
Male	3467
Female	3465

    b. Based on blood type:

Gender	Blood_Type
857	A-
886	A+
867	AB-
886	AB+
834	B-
885	B+
835	O-
882	O+

c. Each blood type based on Gender

Blood_Type	Gender	Gender_count
A-	Male	452
	Female	405
A+	Male	430
	Female	456
AB-	Male	436
	Female	431
AB+	Male	424
	Female	462
B-	Male	418
	Female	416
B+	Male	475
	Female	410
O-	Male	396
	Female	439
O+	Male	436
	Female	446

5) The majority of hospitalized patients come from a wide range of age groups.

a. The following lists the average age group of hospitalized patients by gender:

Age_AVG(in General)
51

Gender	Age_AVG
Male	51
Female	52

b. Standard Deviation of the patient age are calculated to check the variation in the age:

Age_stdev
19.7664800688837

Gender	Age_stdev
Male	19.9186736471249
Female	19.6099245501932

6) Patient suffering different conditions are listed below:

- Diabetes
- Cancer
- Arthritis
- Asthma
- Hypertension
- Obesity

a. Number of patients suffering from different diseases and disorders are given below:

Medical_Condition	NO_OF_PEOPLE
Diabetes	1156
Cancer	1159
Arthritis	1160
Asthma	1161
Hypertension	1154
Obesity	1142

b. Number of patients suffering from different diseases and disorders based on Gender are given below:

Medical_Condition	Gender	NO_OF_PEOPLE
Arthritis	Female	597
	Male	563
Asthma	Female	568
	Male	593
Cancer	Female	584
	Male	575
Diabetes	Female	564
	Male	592
Hypertension	Female	581
	Male	573
Obesity	Female	571
	Male	571

7) Medicines being administered to the patient for different diseases or disorders are listed below:

Medical condition: DIABETES		
No_of_Patients	Medication	Test_Results
67	Aspirin	Abnormal
63	Aspirin	Inconclusive
75	Aspirin	Normal
88	Ibuprofen	Abnormal
87	Ibuprofen	Inconclusive
70	Ibuprofen	Normal
86	Lipitor	Abnormal
77	Lipitor	Inconclusive
80	Lipitor	Normal
73	Paracetamol	Abnormal
62	Paracetamol	Inconclusive
71	Paracetamol	Normal
72	Penicillin	Abnormal
93	Penicillin	Inconclusive
92	Penicillin	Normal

Medical condition: CANCER		
No_of_Patients	Medication	Test_Results
1	NULL	NULL
62	Aspirin	Abnormal
76	Aspirin	Inconclusive
78	Aspirin	Normal
90	Ibuprofen	Abnormal
73	Ibuprofen	Inconclusive
61	Ibuprofen	Normal
99	Lipitor	Abnormal
76	Lipitor	Inconclusive
69	Lipitor	Normal
74	Paracetamol	Abnormal
79	Paracetamol	Inconclusive
79	Paracetamol	Normal
85	Penicillin	Abnormal
74	Penicillin	Inconclusive
83	Penicillin	Normal

Medical condition: ARTHRITIS		
No_of_Patients	Medication	Test_Results
84	Aspirin	Abnormal
81	Aspirin	Inconclusive
68	Aspirin	Normal
79	Ibuprofen	Abnormal
82	Ibuprofen	Inconclusive
72	Ibuprofen	Normal
75	Lipitor	Abnormal
72	Lipitor	Inconclusive
69	Lipitor	Normal
76	Paracetamol	Abnormal
81	Paracetamol	Inconclusive
79	Paracetamol	Normal
88	Penicillin	Abnormal
83	Penicillin	Inconclusive
70	Penicillin	Normal

Medical condition: Asthma		
No_of_Patients	Medication	Test_Results
85	Aspirin	Abnormal
78	Aspirin	Inconclusive
65	Aspirin	Normal
72	Ibuprofen	Abnormal
78	Ibuprofen	Inconclusive
94	Ibuprofen	Normal
96	Lipitor	Abnormal
58	Lipitor	Inconclusive
70	Lipitor	Normal
70	Paracetamol	Abnormal
78	Paracetamol	Inconclusive
68	Paracetamol	Normal
81	Penicillin	Abnormal
82	Penicillin	Inconclusive
84	Penicillin	Normal

Medical condition: Obesity		
No_of_Patients	Medication	Test_Results
88	Aspirin	Abnormal
88	Aspirin	Inconclusive
78	Aspirin	Normal
72	Ibuprofen	Abnormal
87	Ibuprofen	Inconclusive
79	Ibuprofen	Normal
63	Lipitor	Abnormal
69	Lipitor	Inconclusive
72	Lipitor	Normal
76	Paracetamol	Abnormal
74	Paracetamol	Inconclusive
79	Paracetamol	Normal
70	Penicillin	Abnormal
69	Penicillin	Inconclusive
78	Penicillin	Normal

Medical condition: Hypertension		
No_of_Patients	Medication	Test_Results
79	Aspirin	Abnormal
61	Aspirin	Inconclusive
80	Aspirin	Normal
85	Ibuprofen	Abnormal
76	Ibuprofen	Inconclusive
66	Ibuprofen	Normal
78	Lipitor	Abnormal
95	Lipitor	Inconclusive
87	Lipitor	Normal
80	Paracetamol	Abnormal
77	Paracetamol	Inconclusive
64	Paracetamol	Normal
71	Penicillin	Abnormal
77	Penicillin	Inconclusive
77	Penicillin	Normal

8) Number of Patients who gets benefits from different health insurance company

COUNT_name	Insurance_Provider
1408	Cigna
1380	Medicare
1400	Blue Cross
1371	UnitedHealthcare
1373	Aetna

COUNT_MC	Medical_Condition	Insurance_Provider
208	Obesity	Aetna
232	Hypertension	Aetna
228	Diabetes	Aetna
229	Cancer	Aetna
236	Asthma	Aetna
240	Arthritis	Aetna
239	Obesity	Blue Cross
237	Hypertension	Blue Cross
227	Diabetes	Blue Cross
242	Cancer	Blue Cross
226	Asthma	Blue Cross
229	Arthritis	Blue Cross
241	Obesity	Cigna
230	Hypertension	Cigna
230	Diabetes	Cigna
242	Cancer	Cigna
243	Asthma	Cigna
222	Arthritis	Cigna
234	Obesity	Medicare
229	Hypertension	Medicare
242	Diabetes	Medicare
215	Cancer	Medicare
236	Asthma	Medicare
224	Arthritis	Medicare
220	Obesity	UnitedHealthcare
226	Hypertension	UnitedHealthcare
229	Diabetes	UnitedHealthcare
231	Cancer	UnitedHealthcare
220	Asthma	UnitedHealthcare
245	Arthritis	UnitedHealthcare

9) Total billing amount from the given dataset is calculated as $ 177579772.15 and their distribution over different categories are given below.

a. Based on insurance provider

COUNT_MC	BILL_PER_IP	Insurance_Provider	AVG_AMT/PATIENT
1371	35381451.95	UnitedHealthcare	25533.8084303977
1380	35490162.5	Medicare	25717.509057971
1408	35951602.27	Cigna	25718.5319357143
1400	36005944.71	Blue Cross	25807.0400802334
1373	34750610.72	Aetna	25309.9859577567

b. Categorized based on Test result

Test result = Normal		
COUNT_MC	**BILL_PER_IP**	**Medication**
485	11945141.98	Penicillin
447	11025069.02	Lipitor
445	11921954.45	Aspirin
442	10848751.24	Ibuprofen
440	11285299.94	Paracetamol
Total = 2259	57026216.63	

Test result = Abnormal		
COUNT_MC	**BILL_PER_IP**	**Medication**
497	12737064.01	Lipitor
487	12323460.58	Ibuprofen
467	12197154.28	Penicillin
465	12101533.65	Aspirin
450	11672553.79	Paracetamol
Total = 2366	61031766.31	

Test result = Inconclusive		
COUNT_MC	BILL_PER_IP	Medication
483	12531186.15	Ibuprofen
478	12280137.28	Penicillin
451	12136611	Paracetamol
447	11455989.01	Lipitor
447	11088180.37	Aspirin
Total = 2306	59492103.81	

10) Duration (in Days) that patients spent in hospital for treatment.

Duration_Count	Duration (in Days)
1	2130
234	30
201	29
243	28
213	27
234	26
233	25
228	24
233	23
237	22
239	21
245	20
238	19
244	18
238	17
230	16
228	15
233	14
220	13
249	12
217	11
214	10
240	9
238	8
234	7
232	6
228	5
232	4
234	3
214	2
228	1

11) No. of people grouped based on Admission Type

No_of_People	Admission_Type
2274	Emergency
2301	Urgent
1	NULL
2356	Elective

12) No. of people grouped based on Admission Type AND Insurance Provides

No_of_People	Admission_Type	Insurance_Provider
464	Elective	Aetna
463	Emergency	Aetna
446	Urgent	Aetna
463	Elective	Blue Cross
480	Emergency	Blue Cross
457	Urgent	Blue Cross
489	Urgent	Cigna
456	Emergency	Cigna
463	Elective	Cigna
472	Elective	Medicare
440	Emergency	Medicare
468	Urgent	Medicare
435	Emergency	UnitedHealthcare
1	NULL	UnitedHealthcare
494	Elective	UnitedHealthcare
441	Urgent	UnitedHealthcare

13) Average age of people admitted as per Admission Type

No_of_People	Admission_Type	AVG_Age	STDEV_Age
2356	Elective	51	19.7626813340586
2301	Urgent	51	19.5803617026063
2274	Emergency	52	19.9621901737033
1	NULL	32	NULL

14) Blood type needed in Emergency cases

Blood_Type	count	Admission_Type
AB-	298	Emergency
O+	294	Emergency
B+	291	Emergency
O-	289	Emergency
A+	288	Emergency
B-	276	Emergency
AB+	272	Emergency
A-	266	Emergency

15) Emergency cases based on gender

Blood_Type	count	Gender
O+	158	Female
AB-	155	Female
A+	154	Female
O-	152	Female
B+	146	Male
B+	145	Female
AB-	143	Male
A-	143	Male
AB+	143	Female
B-	140	Male
O-	137	Male
B-	136	Female
O+	136	Male
A+	134	Male
AB+	129	Male
A-	123	Female

16) Admission per year

a. No. of. patient admitted per year

count	YEAR
935	2019
1429	2020
1332	2021
1420	2022
1319	2023
496	2024

b. Based on Admission type

count	Admission_Type	YEAR
307	Elective	2019
321	Urgent	2019
307	Emergency	2019
0	NULL	2019
465	Urgent	2020
479	Emergency	2020
485	Elective	2020
441	Elective	2021
448	Emergency	2021
443	Urgent	2021
426	Emergency	2022
483	Urgent	2022
511	Elective	2022
449	Elective	2023
447	Emergency	2023
423	Urgent	2023
163	Elective	2024
167	Emergency	2024
166	Urgent	2024

c. Discharged Patients per year

C_D_P	YEAR
935	2019
1429	2020
1332	2021
1420	2022
1319	2023
496	2024

d. Bill accounted per year

BILL	YEAR
24410527.29	2019
36199046.48	2020
34048231.31	2021
36678293.17	2022
33676660.88	2023
12567013.02	2024

e. Medical condition per year

M_C_Y	YEAR	Medical_condition
169	2019	Arthritis
166	2019	Asthma
160	2019	Cancer
156	2019	Obesity
145	2019	Diabetes
140	2019	Hypertension

M_C_Y	YEAR	Medical_condition
252	2020	Obesity
251	2020	Hypertension
243	2020	Asthma
233	2020	Cancer
230	2020	Diabetes
220	2020	Arthritis

M_C_Y	YEAR	Medical_condition
235	2021	Hypertension
227	2021	Diabetes
222	2021	Asthma
219	2021	Cancer
218	2021	Arthritis
211	2021	Obesity

M_C_Y	YEAR	Medical_condition
235	2021	Hypertension
227	2021	Diabetes
222	2021	Asthma
219	2021	Cancer
218	2021	Arthritis
211	2021	Obesity

M_C_Y	YEAR	Medical_condition
231	2023	Arthritis
222	2023	Hypertension
221	2023	Cancer
219	2023	Diabetes
213	2023	Asthma
213	2023	Obesity

M_C_Y	YEAR	Medical_condition
91	2024	Arthritis
89	2024	Asthma
84	2024	Diabetes
83	2024	Obesity
76	2024	Cancer
73	2024	Hypertension