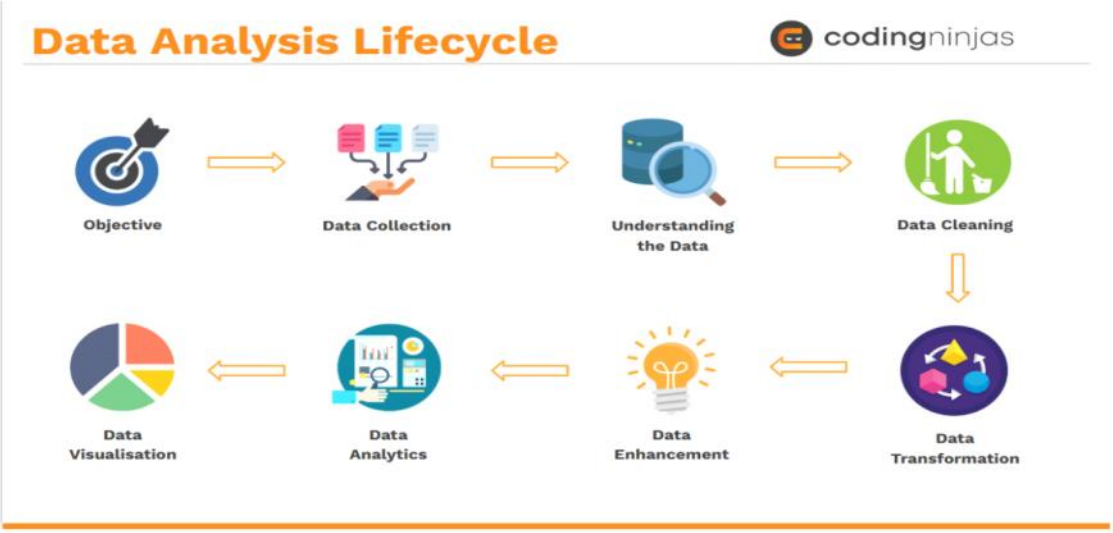
Case Study: US Healthcare



The US healthcare department has shared the hospital data for 100 hospitals which includes data like patient information, their admission information and the insurance information. The top executives from the department have a few questions on the data and would like you to analyse and present your findings to them. The analysis should derive meaning actionable insights and the presentation should be clear, complete and concise.

Problem Statement

Analyzing the Impact of Demographics and Admission Types on Healthcare Costs and Outcomes

Objective I

Identify the most prevalent medical conditions within different demographic groups

Understanding the data

- Creating a Data Dictionary
- Meaning of different attributes - *ChatGPT*
- Demographic attributes
- Distribution of data
- Importance of sample size
- Sample size analysis
- Best Practice - *Documentation of all steps and observations*

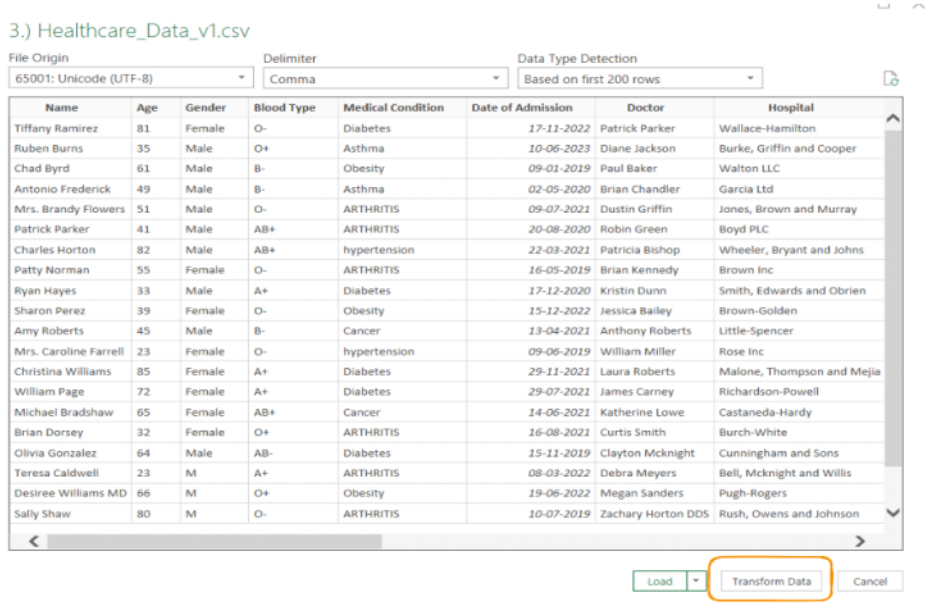
Summary

- Data Analysis Lifecycle
  - Objective
  - Data Collection
  - Understanding the Data
  - Data Cleaning
  - Data Transformation
  - Data Enhancement
  - Data Analytics
  - Data Visualisation

What's Next?

codingninjas

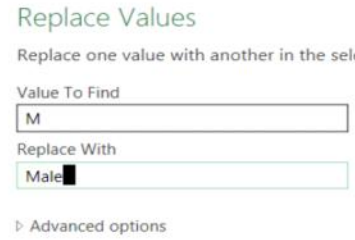
- Data Cleaning
  - Identify inconsistencies in the data
  - Fix inconsistent data
  - NULL value removal
  - Data Formatting
- Pre processing
  - Create additional columns required for analysis
- Excel Functions: *DATE, LEFT, RIGHT, PROPER, UPPER, LOWER, ISBLANK, SPLIT_TEXT*



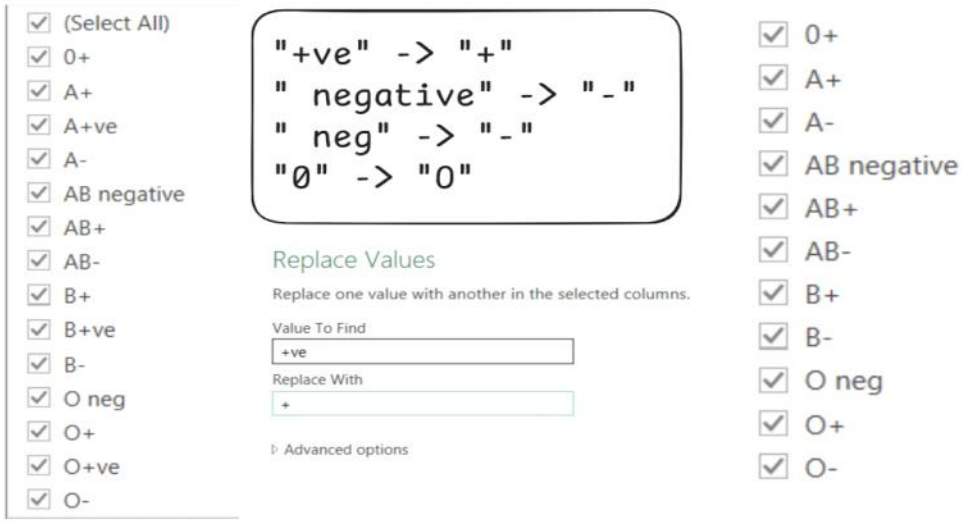
Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital
Tiffany Ramirez	81	Female	O-	Diabetes	17-11-2022	Patrick Parker	Wallace-Hamilton
Ruben Burns	35	Male	O+	Asthma	10-06-2023	Diane Jackson	Burke, Griffin and Cooper
Chad Byrd	61	Male	B-	Obesity	09-01-2019	Paul Baker	Walton LLC
Antonio Frederick	49	Male	B-	Asthma	02-05-2020	Brian Chandler	Garcia Ltd
Mrs. Brandy Flowers	51	Male	O-	ARTHRITIS	09-07-2021	Dustin Griffin	Jones, Brown and Murray
Patrick Parker	41	Male	AB+	ARTHRITIS	20-08-2020	Robin Green	Boyd PLC
Charles Horton	82	Male	AB+	hypertension	22-03-2021	Patricia Bishop	Wheeler, Bryant and Johns
Patty Norman	55	Female	O-	ARTHRITIS	16-05-2019	Brian Kennedy	Brown Inc
Ryan Hayes	33	Male	A+	Diabetes	17-12-2020	Kristin Dunn	Smith, Edwards and Obrien
Sharon Perez	39	Female	O-	Obesity	15-12-2022	Jessica Bailey	Brown-Golden
Amy Roberts	45	Male	B-	Cancer	13-04-2021	Anthony Roberts	Little-Spencer
Mrs. Caroline Farrell	23	Female	O-	hypertension	09-06-2019	William Miller	Rose Inc
Christina Williams	85	Female	A+	Diabetes	29-11-2021	Laura Roberts	Malone, Thompson and Mejia
William Page	72	Female	A+	Diabetes	29-07-2021	James Carney	Richardson-Powell
Michael Bradshaw	65	Female	AB+	Cancer	14-06-2021	Katherine Lowe	Castaneda-Hardy
Brian Dorsey	32	Female	O+	ARTHRITIS	16-08-2021	Curtis Smith	Burch-White
Olivia Gonzalez	64	Male	AB-	Diabetes	15-11-2019	Clayton Mcknight	Cunningham and Sons
Teresa Caldwell	23	M	A+	ARTHRITIS	08-03-2022	Debra Meyers	Bell, Mcknight and Willis
Desiree Williams MD	66	M	O+	Obesity	19-06-2022	Megan Sanders	Pugh-Rogers
Sally Shaw	80	M	O-	ARTHRITIS	10-07-2019	Zachary Horton DDS	Rush, Owens and Johnson

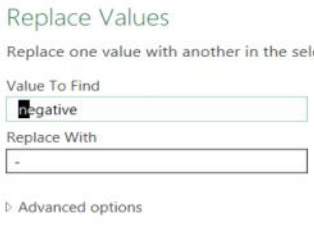
Health Insurance Portability and Accountability Act (HIPAA)


(Select All)
Aetna
Blue Cross
Cigna
Medicare
UnitedHealthcare


Replace Values
Replace one value with another in the selected columns.
Value To Find
M
Replace With
Male
Advanced options

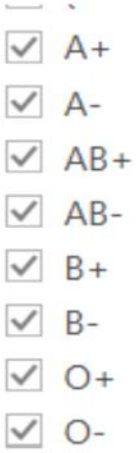

(Select All)
Female
M
Male
(Select All)
Female
Male


(Select All)
0+
A+
A+ve
A-
AB negative
AB+
AB-
B+
B+ve
B-
O neg
O+
O+ve
O-
"+ve" -> "+"
" negative" -> "-"
" neg" -> "-"
"0" -> "O"
Replace Values
Replace one value with another in the selected columns.
Value To Find
+ve
Replace With
+
Advanced options
0+
A+
A-
AB negative
AB+
AB-
B+
B-
O neg
O+
O-


Replace Values
Replace one value with another in the selected columns.
Value To Find
negative
Replace With
-
Advanced options


Replace Values
Replace one value with another in the selected columns.
Value To Find
neg
Replace With
-


Replace Values
Replace one value with another in the selected columns.
Value To Find
0
Replace With
O


A+
A-
AB+
AB-
B+
B-
O+
O-

DataFormat.Error: We couldn't convert to Number.
Details:
8I

Age column having some character
1 -> I , error arise while changing
the data type to whole number

Replace Values

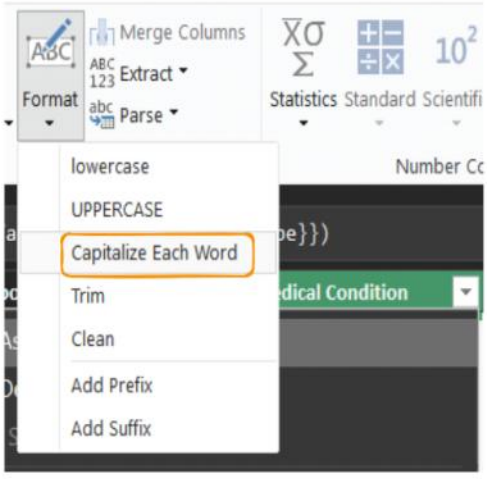
Replace one value with another in the selected columns.

Value To Find
I

Replace With
1

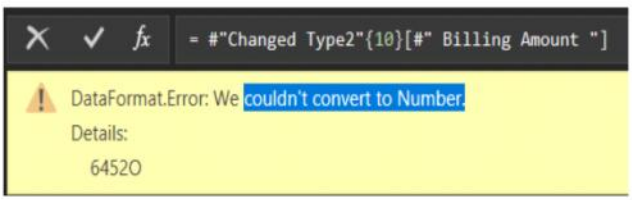
- [x] (Select All)
- [x] ARTHRITIS
- [x] Asthma
- [x] Cancer
- [x] Diabetes
- [x] hypertension
- [x] Obesity

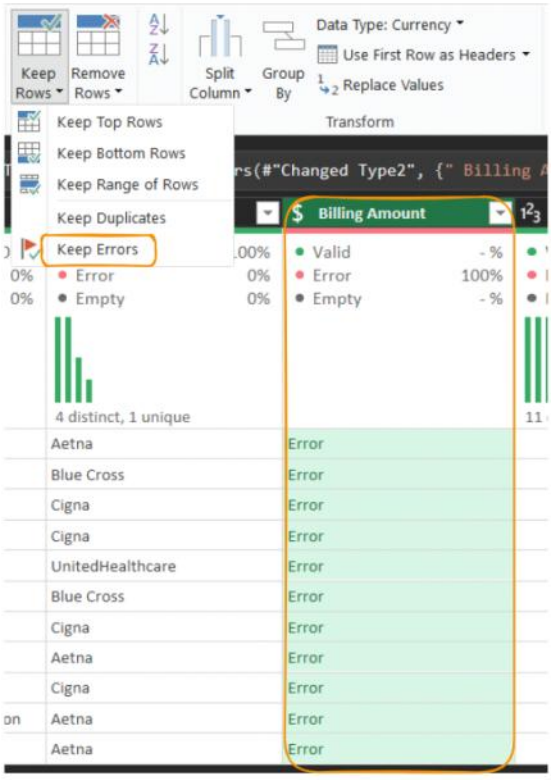
→ Format it into Capitalize each word.

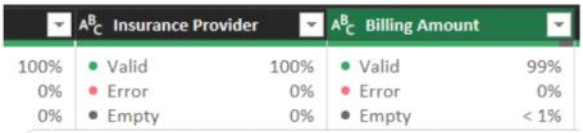


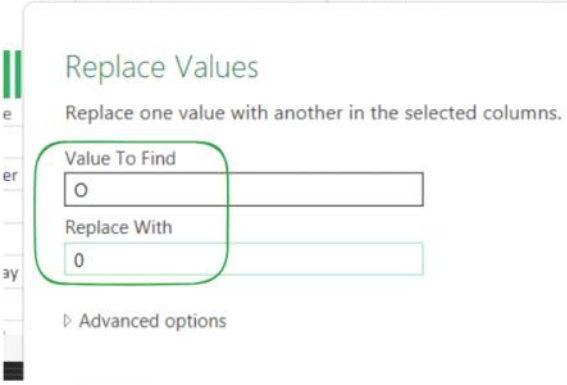
→

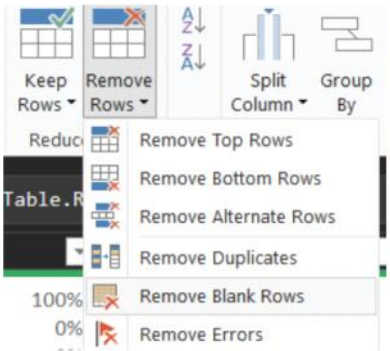
- [x] Arthritis
- [x] Asthma
- [x] Cancer
- [x] Diabetes
- [x] Hypertension
- [x] Obesity



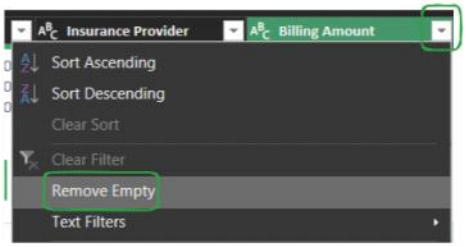


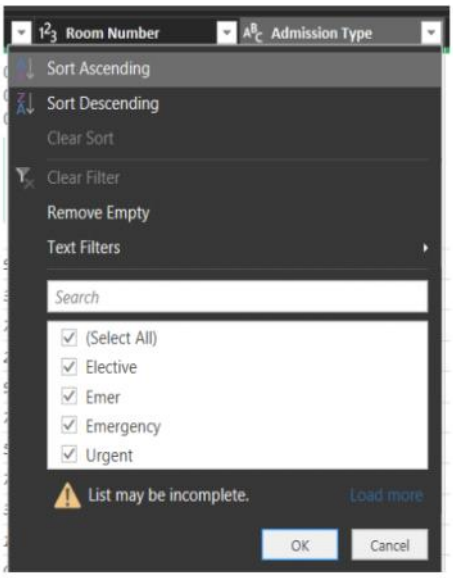






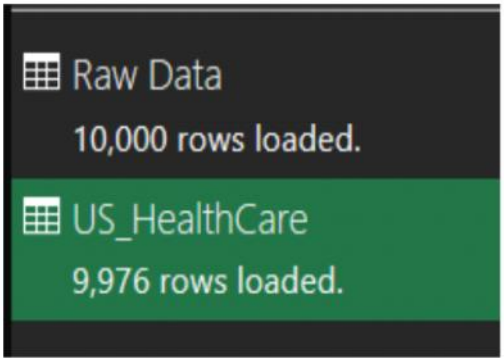
24 records on Billing Amount is having Missing Values,
Which are 0.24% of total record[10K].
On mutual discussion, we thought to remove it
because of impact anlaysis.




Room Number
Admission Type
Sort Ascending
Sort Descending
Clear Sort
Clear Filter
Remove Empty
Text Filters
Search
(Select All)
Elective
Emer
Emergency
Urgent
List may be incomplete.
Load more
OK
Cancel


Replace Values
Replace one value with another in the selected columns.
Value To Find
Emer
Replace With
Emergency
Advanced options


Search
(Select All)
Elective
Emergency
Urgent


Raw Data
10,000 rows loaded.
US_HealthCare
9,976 rows loaded.