

Data Quality Report

Profile Report

Column	Data Type	Missing Values	Unique Values	Mean	Median	Std Dev
PatientID	object	0	200	nan	nan	nan
Age	int64	0	68	53.15	53.5	22.18713240193753
Gender	object	0	6	nan	nan	nan
AdmissionDate	object	0	190	nan	nan	nan
DischargeDate	object	0	4	nan	nan	nan
Diagnosis	float64	20	180	102.06957123811553	103.24515052821329	17.980830125721706
LabResult1	float64	0	191	5.791406770537578	5.563801292632213	1.7722137082903038
LabResult2	object	21	3	nan	nan	nan
Medication	int64	0	200	100.5	100.5	57.879184513951124
Readmitted	object	0	2	nan	nan	nan

Validation Rules

Total Anomalies Detected (Both Methods)

Count: 0

Anomalies Detected Only by Isolation Forest

Count: 10

PatientID	Age	Gender	AdmissionDate	DischargeDate	Diagnosis	LabResult1	LabResult2	Medication	Readmitted	isolation_forest_anomaly	kmeans_cluster	kmeans_anomaly	combined_anomaly
6	41	Female	2023-01-06 00:00:00	2023-01-07 00:00:00	Asthma	nan	4.68368609536731	Med C	No	-1	0	1	0
34	24	Male	2023-02-03 00:00:00	2023-02-04 00:00:00	Healthy	129.41955207945455	11.534855101209908	Med B	No	-1	1	0	0
41	26	Unknown	2023-02-10 00:00:00	2023-02-11 00:00:00	Diabetes	nan	4.066875487785097	Med C	No	-1	0	1	0
54	23							nan	Yes	-1	1	0	0

		Female	2023-02-23 00:00:00	2023-02-24 00:00:00	COVID-19	86.31 8076 1508 5752	11.53 4855 1012 0990 8						
67	79	Male	2023-03-08 00:00:00	2023-03-09 00:00:00	Healthy	nan	8.415 8172 7438 4763	Med A	No	-1	0	1	0
164	19	Male	2023-06-13 00:00:00	2023-06-14 00:00:00	COVID-19	nan	3.436 5507 9194 5335 5	Med A	No	-1	0	1	0
183	33	Male	2023-07-02 00:00:00	2023-07-03 00:00:00	Hypertension	107.1 9371 3866 4963	11.53 4855 1012 0990 8	Med A	No	-1	1	0	0
184	76	Female	2023-07-03 00:00:00	2023-07-04 00:00:00	Asthma	123.7 4546 1909 3480 2	11.53 4855 1012 0990 8	Med B	No	-1	1	0	0
194	88	Female	2023-07-13 00:00:00	2023-07-14 00:00:00	Diabetes	nan	5.164 6011 2103 9952	Med A	No	-1	0	1	0
199	18	Female	2023-07-18 00:00:00	2023-07-19 00:00:00	Diabetes	61.40 6031 3792 3484	7.128 2281 2906 2666	nan	No	-1	0	1	0

Anomalies Detected Only by K-Means

Count: 0

Imputation Suggestions

- LabResult1: Use Mean
- Medication: Use Mode

Date Validation

PatientID: All values are valid dates

Age: All values are valid dates

Gender: Invalid date values detected: Unknown datetime string format, unable to parse: Unknown, at position 0

AdmissionDate: All values are valid dates

DischargeDate: All values are valid dates

Diagnosis: Invalid date values detected: Unknown datetime string format, unable to parse: COVID-19, at position 0

LabResult1: All values are valid dates

LabResult2: All values are valid dates

Medication: Invalid date values detected: Unknown datetime string format, unable to parse: MedC, at position 0

Readmitted: Invalid date values detected: Unknown datetime string format, unable to parse: No, at position 0

Date Sequencing Results

Columns PatientID and Age: Forward Sequences = 145, Backward Sequences = 55

Columns PatientID and AdmissionDate: Forward Sequences = 0, Backward Sequences = 200

Columns PatientID and DischargeDate: Forward Sequences = 0, Backward Sequences = 200

Columns PatientID and LabResult1: Forward Sequences = 87, Backward Sequences = 92

Columns PatientID and LabResult2: Forward Sequences = 196, Backward Sequences = 4

Columns Age and AdmissionDate: Forward Sequences = 0, Backward Sequences = 200

Columns Age and DischargeDate: Forward Sequences = 0, Backward Sequences = 200

Columns Age and LabResult1: Forward Sequences = 8, Backward Sequences = 172

Columns Age and LabResult2: Forward Sequences = 200, Backward Sequences = 0

Columns AdmissionDate and DischargeDate: Forward Sequences = 10, Backward Sequences = 190

Columns AdmissionDate and LabResult1: Forward Sequences = 180, Backward Sequences = 0

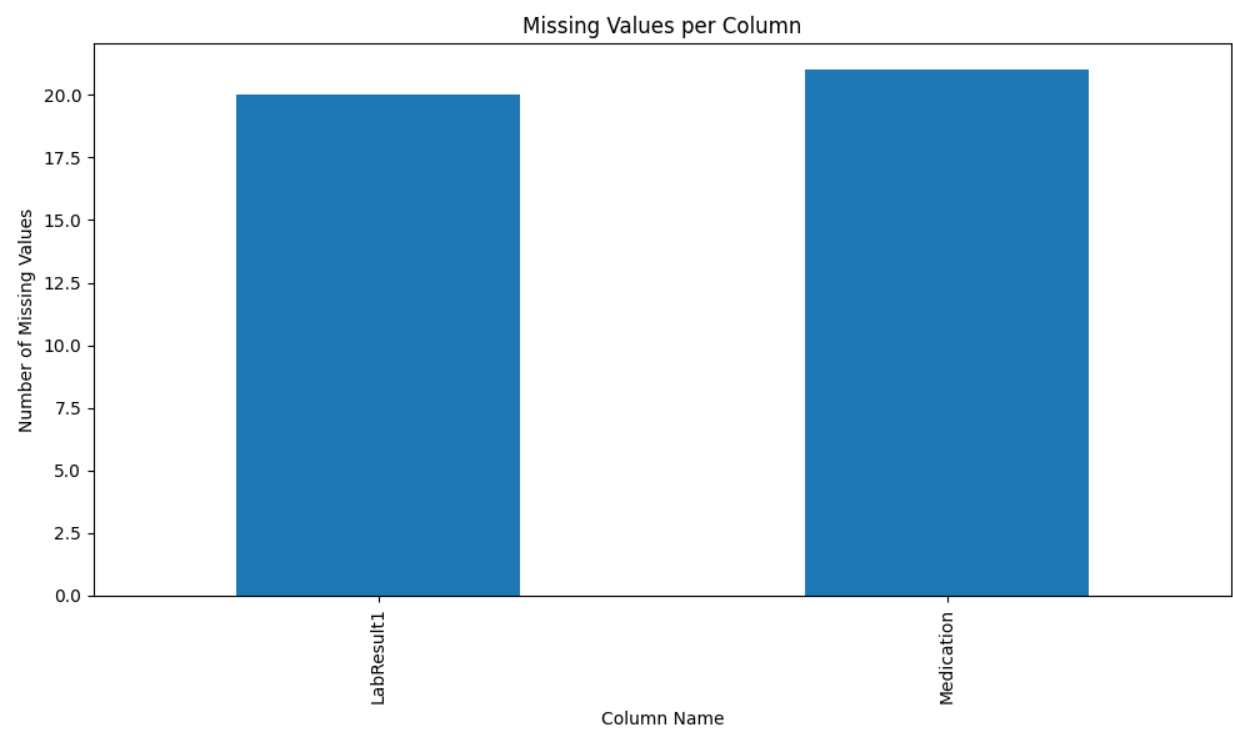
Columns AdmissionDate and LabResult2: Forward Sequences = 200, Backward Sequences = 0

Columns DischargeDate and LabResult1: Forward Sequences = 180, Backward Sequences = 0

Columns DischargeDate and LabResult2: Forward Sequences = 200, Backward Sequences = 0

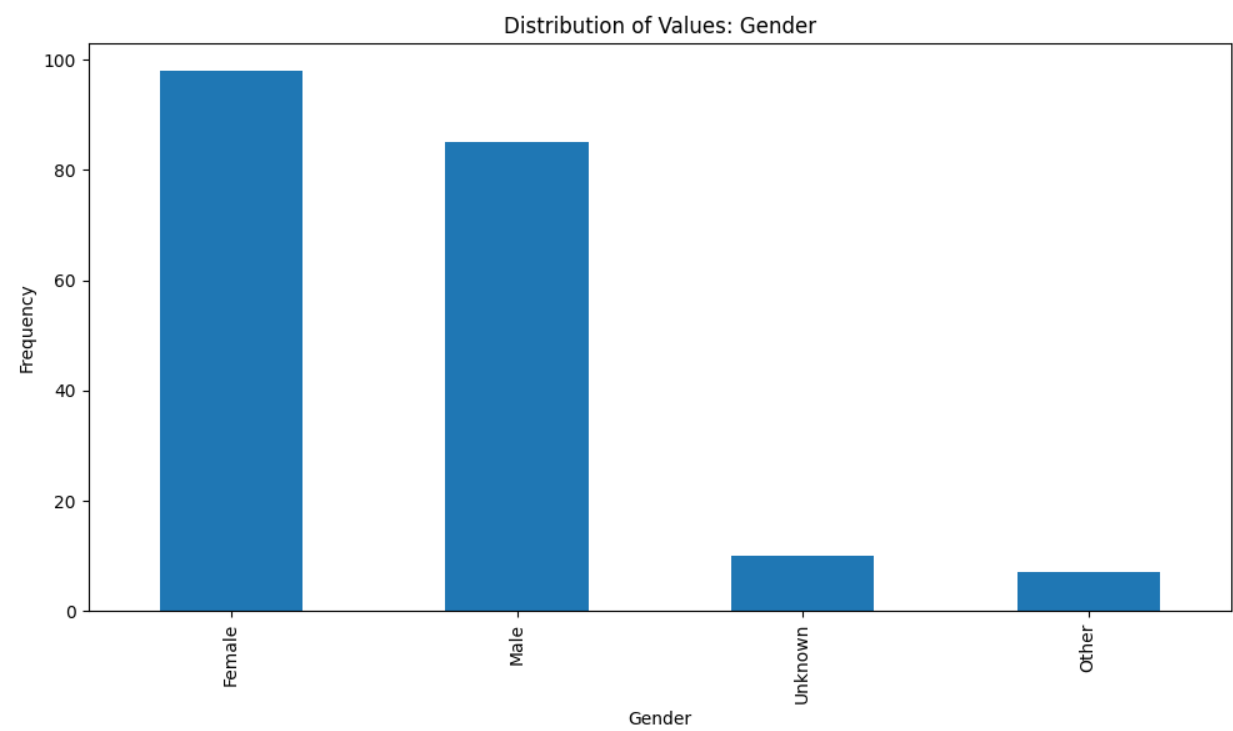
Columns LabResult1 and LabResult2: Forward Sequences = 180, Backward Sequences = 0

Missing Values

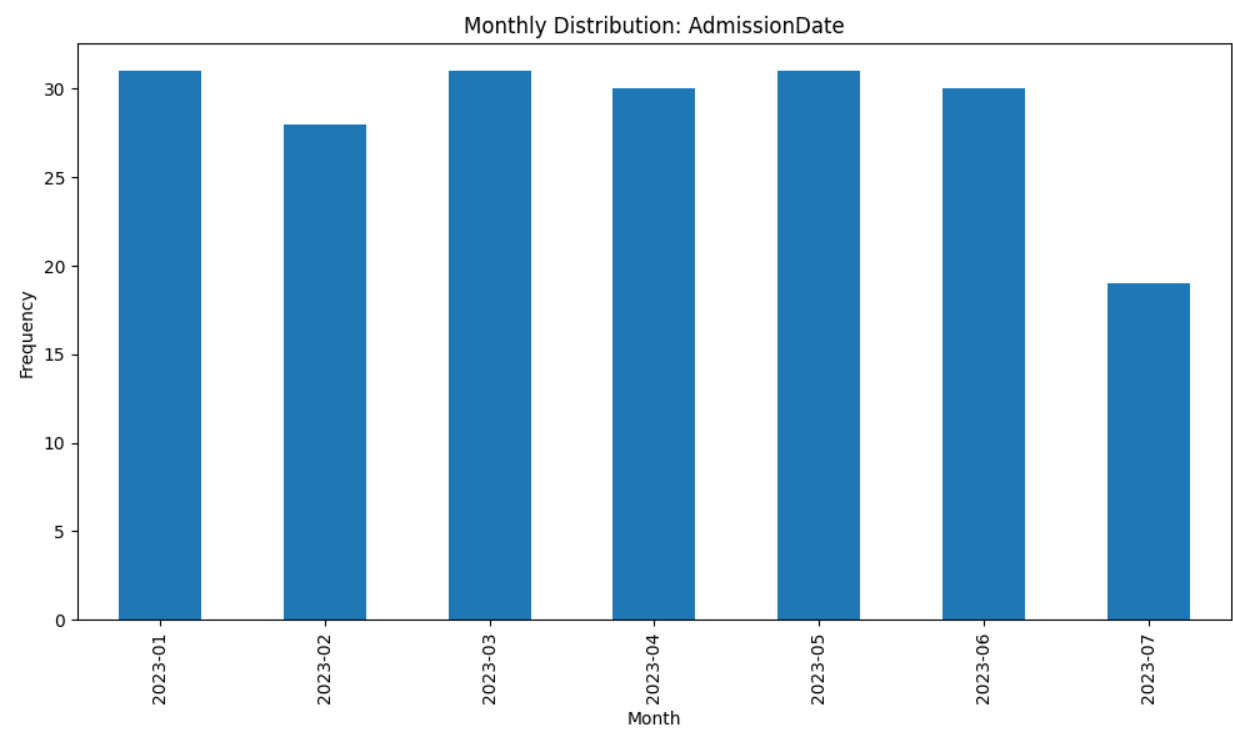


Value Distributions

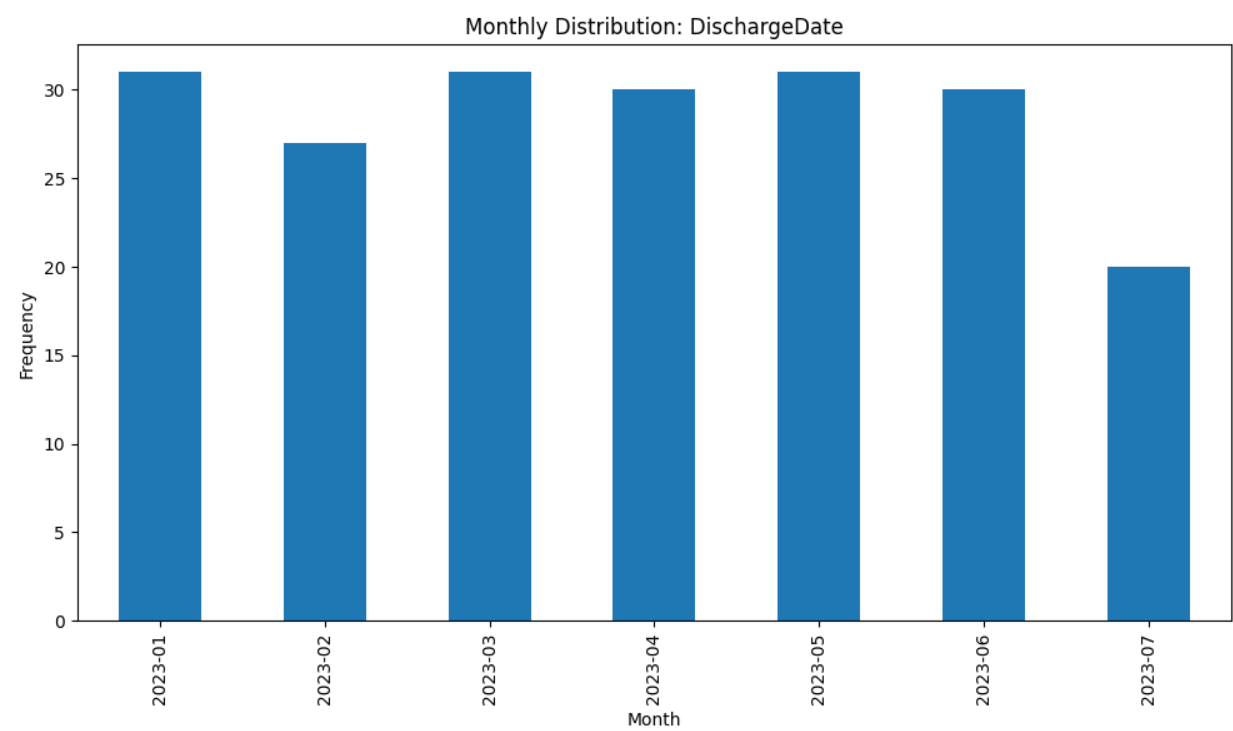
Gender



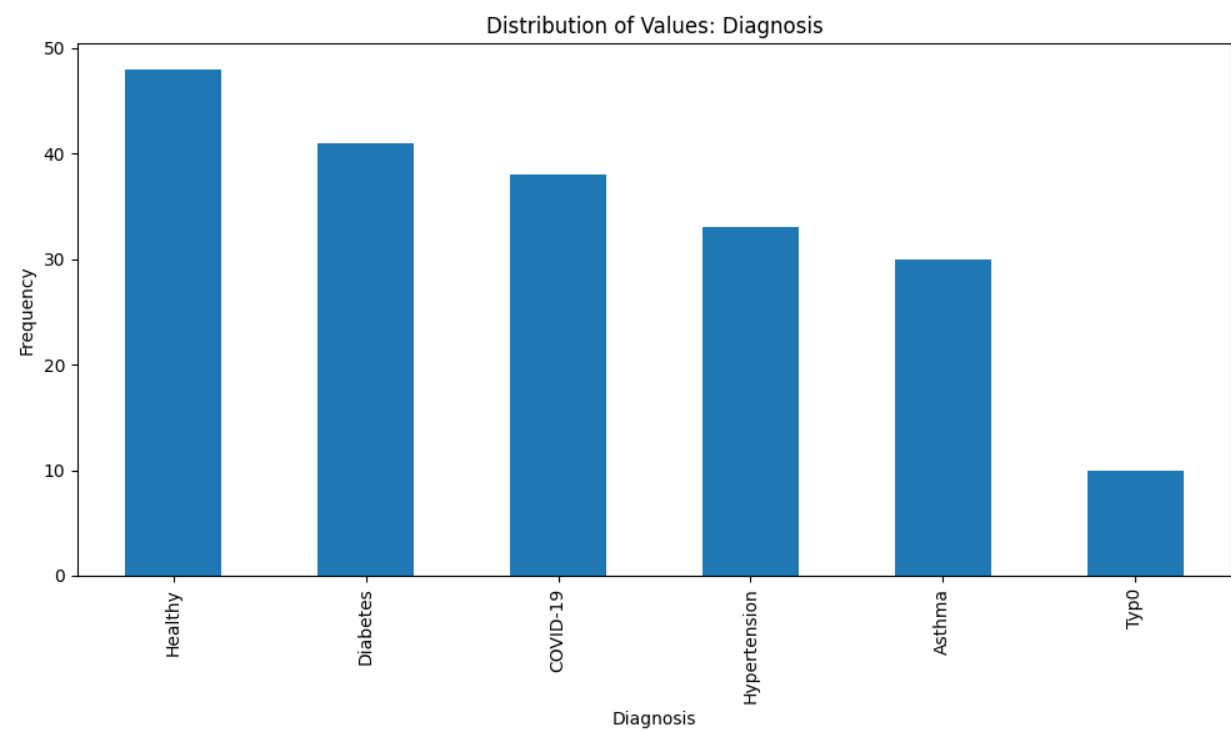
AdmissionDate



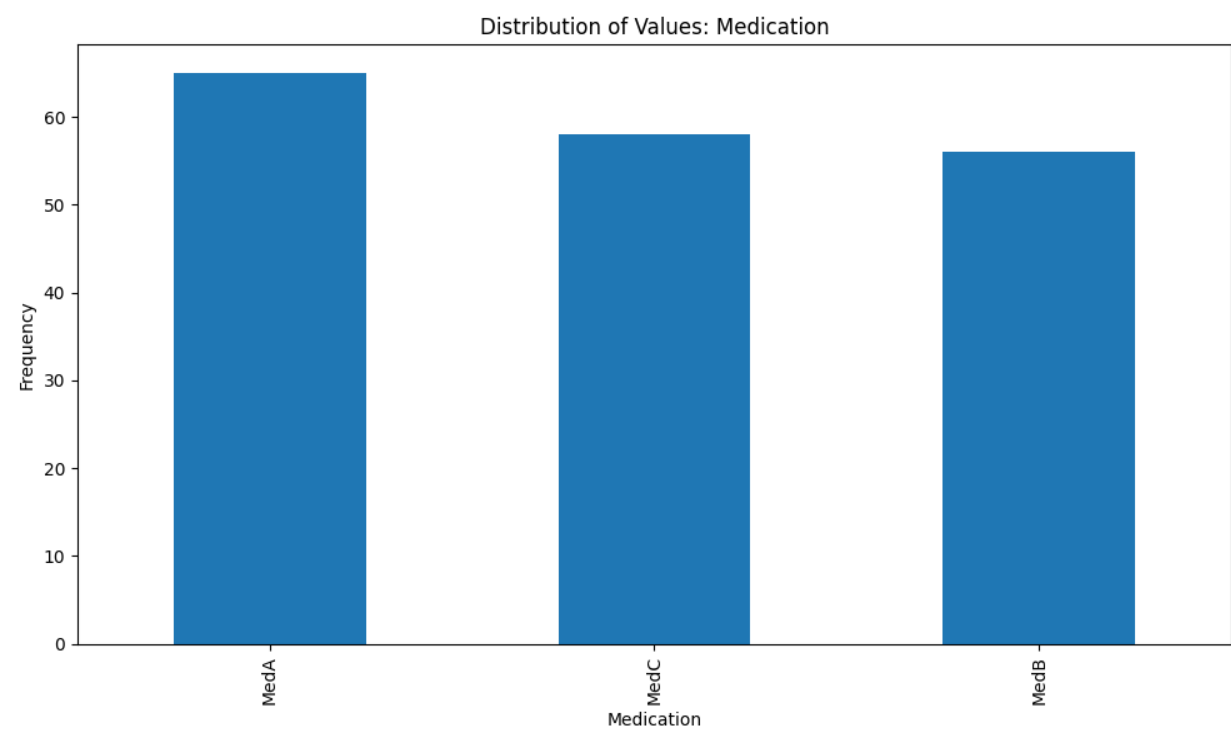
DischargeDate



Diagnosis



Medication



Readmitted

