diabetes[, c(6:15)]

Autogenerated data summary from dataMaid

2024-03-21 21:43:52.495032

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	101766
Number of variables	10

Checks performed

The following variable checks were performed, depending on the data type of each variable:

	character	factor	labelled	haven labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×	×		×
Identify prefixed and suffixed whitespace	×	×	×	×				
Identify levels with < 6 obs.	×	×	×	×				
Identify case issues	×	×	×	×				
Identify misclassified numeric or integer variables	×	×	×	×				
Identify outliers					×	×		×

Please note that all numerical values in the following have been rounded to 2 decimals.

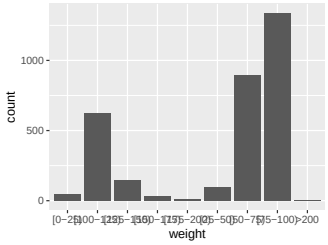
Summary table

	Variable class	# unique values	Missing observations	Any problems?
weight	character	10	96.86 %	×
admission_type_id	integer	8	0.00 %	
discharge_disposition_id	integer	26	0.00 %	
admission_source_id	integer	17	0.00 %	×
time_in_hospital	integer	14	0.00 %	
payer_code	character	18	39.56 %	×
medical_specialty	character	73	49.08 %	×
num_lab_procedures	integer	118	0.00 %	×
num_procedures	integer	7	0.00 %	
num_medications	integer	75	0.00 %	×

Variable list

weight

Feature	Result
Variable type	character
Number of missing obs.	98569 (96.86 %)
Number of unique values	9
Mode	“[75-100)”

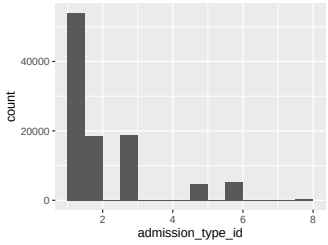


- Note that the following levels have at most five observations: ">200".

---

admission_type_id

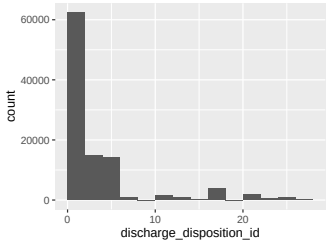
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	8
Median	1
1st and 3rd quartiles	1; 3
Min. and max.	1; 8



---

discharge_disposition_id

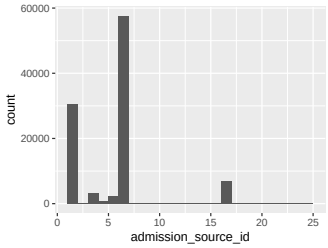
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	26
Median	1
1st and 3rd quartiles	1; 4
Min. and max.	1; 28



---

admission_source_id

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	17
Median	7
1st and 3rd quartiles	1; 7
Min. and max.	1; 25

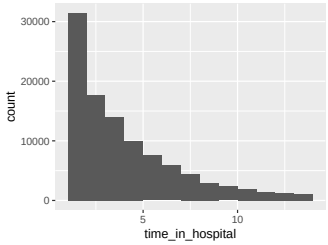


- Note that the following possible outlier values were detected: "8", "9", "10", "11", "13", "14", "17", "20", "22", "25".

---

time_in_hospital

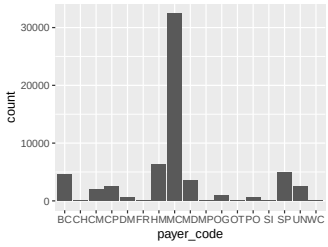
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	14
Median	4
1st and 3rd quartiles	2; 6
Min. and max.	1; 14



---

payer_code

Feature	Result
Variable type	character
Number of missing obs.	40256 (39.56 %)
Number of unique values	17
Mode	“MC”

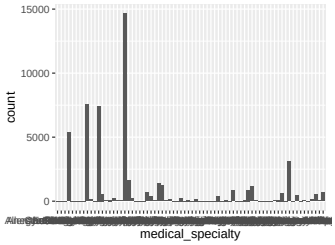


- Note that the following levels have at most five observations: "FR".

---

medical_specialty

Feature	Result
Variable type	character
Number of missing obs.	49949 (49.08 %)
Number of unique values	72
Mode	“InternalMedicine”

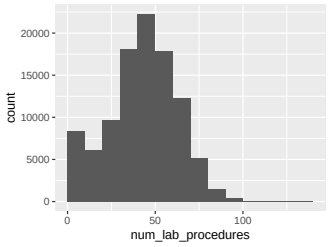


- Note that the following levels have at most five observations: "Dentistry", "Dermatology", "Neurophysiology", "Pediatrics-AllergyandImmunology", "Pediatrics-EmergencyMedicine", ..., "Psychiatry-Addictive", "Resident", "Speech", "SportsMedicine", "Surgery-PlasticwithinHeadandNeck" (4 values omitted).

---

num_lab_procedures

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	118
Median	44
1st and 3rd quartiles	31; 57
Min. and max.	1; 132

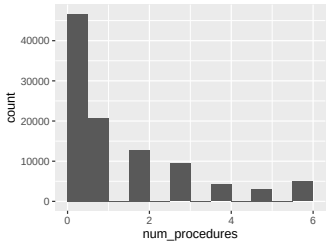


- Note that the following possible outlier values were detected: "91", "92", "93", "94", "95", ..., "120", "121", "126", "129", "132" (18 values omitted).

---

num_procedures

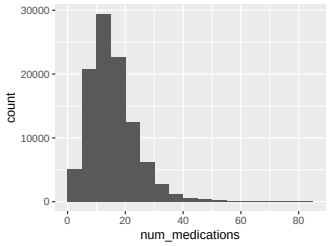
Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	7
Median	1
1st and 3rd quartiles	0; 2
Min. and max.	0; 6



---

num_medications

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	75
Median	15
1st and 3rd quartiles	10; 20
Min. and max.	1; 81



- Note that the following possible outlier values were detected: "41", "42", "43", "44", "45", . . . , "72", "74", "75", "79", "81" (25 values omitted).

Report generation information:

- Created by: Analytics Enlightened LLC (username: `redbe`).
- Report creation time: Thu Mar 21 2024 21:43:54
- Report was run from directory: `C:/Users/redbe/OneDrive/Documents/Data`
- dataMaid v1.4.1 [Pkg: 2021-10-08 from CRAN (R 4.3.1)]
- R version 4.3.1 (2023-06-16 ucrt).
- Platform: x86_64-w64-mingw32/x64 (64-bit)(America/New_York).
- Function call: `makeDataReport(data = diabetes[, c(6:15)])`