**Assessment 1 – report**

Table of variables comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Variable Name | Description | Pandas data type | Mismatch |
| 1 | encounter\_id | Unique identifier of an encounter | int64 | OK – leave as is  Will not be used in analysis |
| 2 | patient\_nbr | Unique identifier of a patient | int64 | OK – leave as is  Will not be used in analysis |
| 3 | race | Race of the patient | Object type (string) | OK, but replace ‘?’ with ‘Other’ to align with unknown values  Treat as nominal |
| 4 | gender | Gender | Object type (string) | OK, but drop single Unknown variable.  Treat as nominal |
| 5 | age | Age quantile | Object type (string) | There are no missing values but consider converting to an interval type,  Treat as ordinal. |
| 6 | weight | Weight in pounds | Object type (string) | 96% if values are missing. Drop this variable |
| 7 | admission\_type\_id | Identifier corresponding to 9 distinct admission types (see IDs\_mapping in Appendix 1) | int64 | This is fine. It is a mapping variable. |
| 8 | discharge\_disposition\_id | Identifier corresponding to 29 distinct values (see IDs\_mapping in Appendix 1) | int64 | This is fine. It is a mapping variable. |
| 9 | admission\_source\_id | Identifier corresponding to 26 distinct values (see IDs\_mapping in Appendix 1) | int64 | This is fine. It is a mapping variable. |
| 10 | Length\_of\_stay | Number of days between admission and discharge | int64 | Looks good. No missing values. |
| 11 | payer\_code | Unique identifier assigned to each insurance company | Object type (string) | Convert ‘?’ to ‘Unknown’ |
| 12 | medical\_specialty | Indicates a specialty of the admitting physician, for example, cardiology, surgeon, etc. | Object type (string) | Convert ‘?’ to ‘Unknown’ |
| 13 | num\_lab\_procedures | Number of lab tests performed during the encounter | int64 | No missing values |
| 14 | num\_procedures | Number of procedures (other than lab tests) performed during the encounter | int64 | No missing values |
| 15 | num\_medications | Number of distinct generic medication names administered during the encounter | int64 | No missing values |
| 16 | number\_outpatient | Number of outpatient visits of the patient in the year preceding the encounter | Object type (string) | Drop the ‘?’ and convert the value to an integer. There are only 20 missing values. |
| 17 | number\_emergency | Number of emergency visits of the patient in the year preceding the encounter | Object type (string) | Drop the ‘?’ and convert the value to an integer. There are only 69 missing values. |
| 18 | number\_inpatient | Number of inpatient visits of the patient in the year preceding the encounter | Object type (string) | Drop the ‘?’ and convert the value to an integer. There are only 15 missing values. |
| 19 | diag\_1 | The primary diagnosis (coded as the first three digits of ICD9) | Object type (string) | These are nominal variables so can be treated as strings. |
| 20 | diag\_2 | Secondary diagnosis (coded as the first three digits of ICD9) | Object type (string) | These are nominal variables so can be treated as strings. |
| 21 | diag\_3 | Additional secondary diagnosis (coded as the first three digits of ICD9) | Object type (string) | These are nominal variables so can be treated as strings. |
| 22 | number\_diagnoses | Number of diagnoses entered into the system | int64 | No mismatch. Data are fine. |
| 23 | diabetes | Indicates if the patient’s primary diagnosis is diabetes or not.  Values include: “Yes” and “No” | Object (string) | No mismatch. Data are fine.  Treat as binary / boolean |
| 24 | max\_glu\_serum | Glucose serum test result.  Indicates the range of the result or if the test was not taken ( “none”) | Object type (string) | The data do not align with the description. The value represents range or ‘Norm’ suggesting results were in normal range.  Treat as ordinal. |
| 25 | A1Cresult | A1c test result.  Values include: ‘>8’ if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal” if the result was less than 7%, and “none” if not measured. | Object type (string) | Treat as ordinal. |
| 26 | metformin | These are 10 variables for diabetes medications. The values of these variables indicate whether the drug was prescribed or there was a change in the dosage.  Values include: “Up” if the dosage was increased during the encounter, “Down” if the dosage was decreased, “Steady” if the dosage did not change, and “No” if the drug was not prescribed. | Object type (string) | Treat as nominal. |
| 27 | repaglinide | Object type (string) | Treat as nominal. |
| 28 | nateglinide | Object type (string) | Treat as nominal. |
| 29 | chlorpropamide | Object type (string) | 1 missing variable – recode to No  Treat as nominal. |
| 30 | glimepiride | Object type (string) | Treat as nominal. |
| 31 | acetohexamide | Object type (string) | Drop the variable. All values are no. Variable will not be used |
| 32 | glipizide | Object type (string) | Treat as nominal. |
| 33 | glyburide | Object type (string) | Treat as nominal. |
| 34 | tolbutamide | Object type (string) | Drop the variable. All values are no. Variable will not be used |
| 35 | insulin | Object type (string) | Treat as nominal. |
| 36 | change | Change of medications.  Indicates if there was a change in diabetic medications (either dosage or generic name).  Values include: “Ch” and “No” | Object type (string) | Treat as binary / boolean |
| 37 | diabetesMed | Diabetes medications.  Indicates if there was any diabetic medication prescribed.  Values include: “Yes” and “No” | Object type (string) | Treat as binary / boolean |
| 38 | Readmitted | Days to inpatient readmission.  Values include: “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was  readmitted in more than 30 days, and “No” for no record of readmission. | Object type (string) | Treat as ordinal. Not readmitted, within 30 days, treat the >30 as greater than or equal. |
| 39 | single\_day\_admission | Indicates if this encounter is a single-day admission or not  Values include: “Yes” and “No” | Object type (string) | Treat as binary / boolean |