DATA DICTIONARY

1. id: Unique identifier for each record.
2. encounter_id: Unique identifier for each patient encounter.
3. patient_nbr: Unique identifier for each patient.
4. race: Patient's reported race.
5. gender: Patient's gender.
6. age: Patient's age group.
7. weight: Patient's weight (if available).
8. admission_type_id: Identifier for the type of admission (e.g., emergency, elective, etc.).
9. discharge_disposition_id: Identifier for the disposition of the patient upon discharge (e.g., home, hospice, etc.).
10. admission_source_id: Identifier for the source of admission (e.g., emergency room, referring physician, etc.).
11. time_in_hospital: Number of days the patient spent in the hospital.
12. payer_code: Code for the insurance provider.
13. medical_specialty: Specialty of the admitting physician.
14. num_lab_procedures: Number of laboratory procedures the patient underwent.
15. num_procedures: Number of non-lab procedures the patient underwent.
16. num_medications: Number of medications the patient was on.
17. number_outpatient: Number of outpatient visits the patient had.
18. number_emergency: Number of emergency visits the patient had.
19. number_inpatient: Number of inpatient visits the patient had.
20. diag_1: Primary diagnosis code.
21. diag_2: Secondary diagnosis code 1.
22. diag_3: Secondary diagnosis code 2.
23. number_diagnoses: Number of diagnoses associated with the patient encounter.
24. max_glu_serum: Maximum glucose serum test result.
25. A1Cresult: A1C test result.
26. metformin: Medication status for metformin.
27. repaglinide: Medication status for repaglinide.
28. nateglinide: Medication status for nateglinide.
29. chlorpropamide: Medication status for chlorpropamide.
30. glimepiride: Medication status for glimepiride.
31. acetohexamide: Medication status for acetohexamide.
32. glipizide: Medication status for glipizide.
33. glyburide: Medication status for glyburide.

34. tolbutamide: Medication status for tolbutamide.
35. pioglitazone: Medication status for pioglitazone.
36. rosiglitazone: Medication status for rosiglitazone.
37. acarbose: Medication status for acarbose.
38. miglitol: Medication status for miglitol.
39. troglitazone: Medication status for troglitazone.
40. tolazamide: Medication status for tolazamide.
41. examide: Medication status for examide.
42. citoglipton: Medication status for citoglipton.
43. insulin: Medication status for insulin.
44. glyburide.metformin: Medication status for combination of glyburide and metformin.
45. glipizide.metformin: Medication status for combination of glipizide and metformin.
46. glimepiride.pioglitazone: Medication status for combination of glimepiride and pioglitazone.
47. metformin.rosiglitazone: Medication status for combination of metformin and rosiglitazone.
48. metformin.pioglitazone: Medication status for combination of metformin and pioglitazone.
49. change: Whether there was a change in diabetes medications.
50. diabetesMed: Whether the patient was on diabetes medications.
51. readmitted: Whether the patient was readmitted to the hospital.

FEATURE DESCRIPTION

1. id: Unique identifier for each record. This column may not have a direct medical significance but can help keep track of individual data points.

2. encounter_id: Unique identifier for each patient encounter. This column helps identify and differentiate different visits or encounters for the same patient.

3. patient_nbr: Unique identifier for each patient. This column helps track and distinguish different patients within the dataset.

4. race: Patient's reported race. This could be used to analyze disparities in healthcare outcomes among different racial groups.

5. gender: Patient's gender. Gender can play a role in medical diagnoses, treatments, and outcomes analysis.

6. age: Patient's age group. This can help categorize patients into age ranges and analyze how medical conditions and treatments vary with age.

7. weight: Patient's weight (if available). Weight is crucial for medication dosing and assessing overall health.

8. admission_type_id: Identifier for the type of admission. This can be used to analyze the distribution of different admission types and their impact on outcomes.

9. discharge_disposition_id: Identifier for the disposition of the patient upon discharge. This could help study where patients are sent after their hospital stay and assess readmission rates based on these dispositions.

10. admission_source_id: Identifier for the source of admission. This helps analyze where patients are coming from when they enter the hospital.

11. time_in_hospital: Number of days the patient spent in the hospital. This could provide insights into the severity of the patient's condition and its impact on the length of stay.

12. payer_code: Code for the insurance provider. This could be used to analyze healthcare utilization and outcomes based on different insurance providers.

13. medical_specialty: Specialty of the admitting physician. This could help understand how different medical specialties are associated with specific patient populations or conditions.

14. num_lab_procedures: Number of laboratory procedures the patient underwent. This could indicate the extent of diagnostic testing done for a patient.

15. num_procedures: Number of non-lab procedures the patient underwent. This might include surgeries or other interventions.

16. num_medications: Number of medications the patient was on. This can provide insight into the complexity of the patient's medical regimen.

17. number_outpatient: Number of outpatient visits the patient had. This could indicate the frequency of post-discharge follow-up visits.

18. number_emergency: Number of emergency visits the patient had. This could reflect the patient's overall health status and utilization of emergency services.

19. number_inpatient: Number of inpatient visits the patient had. This could reflect the patient's overall hospitalization history.

20. diag_1, diag_2, diag_3: Primary and secondary diagnosis codes. These codes represent the medical conditions the patient was diagnosed with during the encounter.

21. number_diagnoses: Number of diagnoses associated with the patient encounter. This could give an indication of the complexity of the patient's medical condition.

22. max_glu_serum: Maximum glucose serum test result. This could indicate the patient's glucose control.

23. A1Cresult: A1C test result. This test measures a patient's average blood sugar levels over the past few months, giving insight into diabetes management.

24. metformin, repaglinide, etc.: Medication status columns. These columns indicate whether the patient was on specific medications, which can be used to study treatment adherence and effectiveness.

25. change: Whether there was a change in diabetes medications. This could indicate adjustments made to the patient's treatment regimen.

26. diabetesMed: Whether the patient was on diabetes medications. This gives an overall view of the patient's medication status.
27. readmitted: Whether the patient was readmitted to the hospital. This could be used for readmission prediction and analyzing factors associated with readmission.