Field Name	Data Type	Changes (if any)	Description (if Needed
Encounter_ID	Int64 (Number-whole)	N/a	The ID number associated with the patient's visit
Patient_nbr	Int64 (Number-whole)	N/a	Patients ID number
Race	String	Assigned each distinct value with a numerical value ? = 0 **African American** = 10 **Asian** = 2 **Caucasian** = 3 **Hispanic** = 4 **Other** = 5	*Categorical data* Ordered based on alphabetical order
Gender	String	Assigned: Unknown/invalid = 0 Female = 1 Male = 2	*Categorical data*
Age	String	Assigned each group as a numerical Value: **0-10** = 1 **10-20** = 2 **20-30** = 3 **30-40** = 4 **40-50** = 5 **50-60** = 6 **60-70** = 7 **70-80** = 8 **80-90** = 9 **90-100** = 10	*Numerical data*
Weight	N/a	Removed due to a lot of data being unknown	N/a
Admission_type_id	Int64 (Number-whole)	N/a	*Categorical data* Refer to dataset IDs_mapping
discharge_disposition_id	Int64 (Number-whole)	N/a	*Categorical data* Refer to dataset IDs_mapping

admission_source_id	Int64 (Number-whole)	N/a	*Categorical data* Refer to dataset IDs_mapping
time_in_hospital	Int64 (Number-whole)	N/a	*Numerical data* Time spent in a hospital in Hours Max: 14 Min: 1
payer_code	N/a	Removed due to lack of prevalence to dataset	*Categorical data*
medical_specialty	String	Turned specialties into numerical Values based on alphabetical order: ? = 0 AllergyandImmunology = 1 Anesthesiology = 2 Anesthesiology-Pediatric = 3 Cardiology = 4 Cardiology-Pediatric = 5 DCPTEAM = 6 Dentistry = 7 Dermatology = 8 Emergency/Trauma = 9 Endocrinology/Trauma = 10 Endocrinology-Metabolism = 11 Family/GeneralPractice = 12 Gastroenterology= 13 Gynecology = 14 Hematology = 15 Hematology/Oncology = 16 Hospitalist = 17 InfectiousDiseases = 18 InternalMedicine = 19 Nephrology = 20 Neurology = 21 Neurophysiology =22 Obsterics&Gynecology-GynecologicOnco = 23 Obstetrics = 24 ObstetricsandGynecology = 25 Oncology = 26 Ophthalmology = 27 Orthopedics = 28 Orthopedics-Reconstructive = 29 Osteopath = 30 Otolaryngology = 31 OutreachServices = 32 Pathology = 33 Pediatrics = 34	*Categorical data* Different departments/specialties that help to differentiate which specialty is having the most readmissions

		Pediatrics-AllergyandImmunology = 35 Pediatrics-CriticalCare = 36 Pediatrics-EmergencyMedicine =37 Pediatrics-Endocrinology = 38 Pediatrics-Hematology-Oncology = 39 Pediatrics-InfectiousDiseases = 40 Pediatrics-Neurology = 41 Pediatrics-Pulmonology = 42 Perintatology = 43 PhysicalMedicineandRehabilitation = 44 PhysicianNotFound = 45 Podiatry =46 Proctology =47 Psychiatry =48 Psychiatry-Addictive = 49 Psychiatry-Child/Adolescent = 50 Psychology = 51 Pulmonology = 52 Radiologist = 53 Radiology = 54 Resident = 55 Rheumatology = 56 Speech = 57 SportsMedicine = 58 Surgeon = 59 Surgery Cardiovascular = 60 Surgery-Cardiovascular = 61 Surgery-Colon&Rectal = 62 Surgery-General = 63 Surgery-Maxillofacial = 64 Surgery-Neruo = 65 Surgery-Pediatric =66 Surgery-Plastic = 67 Surgery-PlasticwithinHeadandNeck = 68 Surgery-Thoracic = 69 Surgery-Vascular = 70 SurgerySpecialty = 71 Urology = 72	
num_lab_procedures	Int64 (Number-whole)	N/a	*Numerical data* Number of lab procedures done to patient: Blood test, Urine, other substance
num_procedures	Int64 (Number-whole)	N/a	*Numerical data* Number of Procedures done of a patient: can be surgery

num_medications	Int64 (Number-whole)	N/a	*Numerical data* Number of medication the patient is taking
number_outpatient	Int64 (Number-whole)	N/a	*Numerical data*
number_emergency	Int64 (Number-whole)	N/a	*Numerical data*
number_inpatient	Int64 (Number-whole)	N/a	*Numerical data*
diag_1	String (Number-decimal)	N/a	*Categorical data* LOINC code

diag_2	String (Number-decimal)	N/a	*Categorical data* LOINC code
diag_3	String	N/a	*Categorical data* LOINC code
number_diagnoses	Int64 (Number-whole)	n/a	*Numerical data* Number of diagnosis patient has
max_glu_serum	String	Changed values to a numerical format to allow for proper analysis **None** = 0 **Norm** = 140 **>200** = 200 **>300** = 300	*Categorical data* **Units: mg/dL** **Patients with higher than >200 and >300 both have diabetes**. **Normal reading would be average levels 140**
A1Cresult	String	Changed values to a numerical format to allow for proper analysis **None** = 0 **Norm** = 1 **>7** = 7 **>8** = 8	*Categorical data* **Units: mg/dL** **One of the most important category since all drugs affects this reading**

metformin	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
repaglinide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
nateglinide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
chlorpropamide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
glimepiride	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage

			Steady= medication stayed at same dose **Up** = Increase in dosage
acetohexamide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
glipizide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
glyburide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
tolbutamide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Steady** = 1	*Categorical data* Medication **No**= does not take medication **Steady**= medication stayed at same dose
pioglitazone	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2	*Categorical data* Medication **No**= does not take medication **Down** = decrease in

		Up = 3	dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
rosiglitazone	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
acarbose	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
miglitol	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
troglitazone	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Steady** = 1	*Categorical data* Medication **No**= does not take medication **Steady**= medication stayed at same dose
tolazamide	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1	*Categorical data* Medication **No**= does not take medication

		Steady = 2 **Up** = 3	**Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage

examide	String	Removed column since there is only one distinct value which would not help predict an outcome within an analysis	N/a
citoglipton	String	Removed column since there is only one distinct value which would not help predict an outcome within an analysis	N/a
insulin	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
glyburide-metformin	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in dosage **Steady**= medication stayed at same dose **Up** = Increase in dosage
glipizide-metformin	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Steady** = 1	*Categorical data* Medication **No**= does not take medication **Steady**= medication stayed at same dose
glimepiride-pioglitazone	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Steady** = 1	*Categorical data* Medication **No**= does not take medication **Steady**= medication

			stayed at same dose
metformin-rosiglitazone	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Steady** = 1	*Categorical data* Medication **No**= does not take medication **Steady**= medication stayed at same dose
metformin-pioglitazone	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Steady** = 1	*Categorical data* Medication **No**= does not take medication **Steady**= medication stayed at same dose
change	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Ch** = 1	*Categorical data* Changes in medication?
diabetesMed	String	Changed values to a numerical format to allow for proper analysis: **No** = 0 **Yes** = 1	*Categorical data* Do they have diabetes medication?
readmitted	String	Changed values to a numerical format to allow for proper analysis: **NO** = 0 **>30** = 1 **<30** = 2	*Categorical data* How long until they readmitted? **Before, after 30 days, or not at all**