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** = 1 **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** = 10 **10-20** = 20 **20-30** = 30 **30-40** = 40 **40-50** = 50 **50-60** = 60 **60-70** = 70 **70-80** = 80 **80-90** = 90 **90-100** = 100	*Categorical data* Not based on a continuous value but rather groupings
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	N/a Will perform dummy encoding on	*Categorical data* Different departments/specialties that help to differentiate which specialty is having the most readmissions
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 (Number-decimal)	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	Removed column since there is only one value that is different from the rest	*N/a*
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 **Up** = 3	*Categorical data* Medication **No**= does not take medication **Down** = decrease in 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 **Stead**y = 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 **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

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	*Categorical data*

		for proper analysis: **No** = 0 **Down** =1 **Steady** = 2 **Up** = 3	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	Removed column since there is only one value that is different from the rest	*N/a*
metformin-pioglitazone	String	Removed column since there is only one value that is different from the rest	N/a
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* Did they have diabetes medication:
readmitted	String	Changed values to a numerical format to allow for proper analysis: **NO** = 1 **>30** = 1 **<30** = 2	*Categorical data* How long until they readmitted? **Two datasets one with both no readmissions/ readmission and another with only readmissions** **Made the values for**

			both no and >30 as 1 since after 30 day readmissions don't mean much for analysis