

CODEBOOK

This dataset is a subset of the following dataset:

- Prabhavalkar, Neha. (2020). AV: Healthcare Analytics II [Data Set] (<https://www.kaggle.com/nehaprabhavalkar/av-healthcare-analytics-ii>)

The original dataset contains information about COVID-19 hospitalizations including several characteristic variables regarding each patient. Each patient case is identifiable by `case_id`, although it was excluded in the processed dataset because it was not used as a predictor.

The processed dataset **patients_data**, read in from `data/processed/patients_data.csv`, contains the subset of variables from the original dataset that was included in the model. This dataset contains 244393 rows and 6 columns. There is no missing data for any of the patient records.

Variable Names	Variable Types	Description of Variable
stay	factor	Length of stay for the patient as categorized into one of four categories (0-10, 11-20, 21-30, 31-40 days)
admission_deposit	double	Deposit at the admission time
age	factor	Age of the patient as categorized into one of nine categories (0-10, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100)
severity_of_illness	factor	Severity of the illness recorded at the time of admission as categorized into one of three categories (extreme, moderate, minor)
visitors_with_patient	double	Number of visitors with the patient
ward_facility_code	factor	Code for the ward type as categorized into one of six categories (A, B, C, D, E, F)