

**PROJECT REPORT**

**ON**

**HOSPITAL READMISSION PREDICTION**

Done By:-

1. Meenavalli Sai Subhakar

2. Gurijala Sateesh

3. Ramini Manideep

INDEX :-

|  |  |  |
| --- | --- | --- |
| CHAPTER NO | TITLE | PAGENO |
| 1. | 1.1-Introduction  1.2-Objectives of Research  1.3-Problem Statement | 1-2 |
| 2. | Review of Literature | 3 |
| 3. | Data Collection | 4 |
| 4. | Methodology  1.Exploratory Data Analysis  2.Data Modelling | 5-8 |
| 5. | Findings and Suggestions | 8 |
| 6. | Conclusion | 9 |

**CHAPTER 1**

* 1. **INTRODUCTION**

In [computer science](https://en.wikipedia.org/wiki/Computer_science), **artificial intelligence** (**AI**), sometimes called **machine intelligence**, is [intelligence](https://en.wikipedia.org/wiki/Intelligence) demonstrated by [machines](https://en.wikipedia.org/wiki/Machine), in contrast to the **natural intelligence** displayed by humans. Colloquially, the term "artificial intelligence" is used to describe machines/computers that mimic "cognitive" functions that humans associate with other [human minds](https://en.wikipedia.org/wiki/Human_mind), such as "learning" and "problem solving".

As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the [AI effect](https://en.wikipedia.org/wiki/AI_effect). A quip in Tesler's Theorem says "AI is whatever hasn't been done yet." For instance, [optical character recognition](https://en.wikipedia.org/wiki/Optical_character_recognition) is frequently excluded from things considered to be AI, having become a routine technology. Modern machine capabilities generally classified as AI include successfully [understanding human speech](https://en.wikipedia.org/wiki/Natural_language_understanding), competing at the highest level in [strategic game](https://en.wikipedia.org/wiki/Strategic_game) systems (such as [chess](https://en.wikipedia.org/wiki/Chess) and [Go](https://en.wikipedia.org/wiki/Go_(game))), [autonomously operating cars](https://en.wikipedia.org/wiki/Autonomous_car), intelligent routing in [content delivery networks](https://en.wikipedia.org/wiki/Content_delivery_network), and [military simulations](https://en.wikipedia.org/wiki/Military_simulations).

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.

.

* 1. **OBJECTIVES OF RESEARCH**

The main aim of the project is to predict if the patient who is suffering with the diabetes will readmit into the hospital within the 30 days or not based on past information. If the patient readmitted into the hospital within the period of time then we can predict the rating of the hospital. Based on the Readmission rate, the care and quality of hospital will be checked. So our main aim is to predict the readmission rate of the hospital.

* 1. **Problem Statement**

In this project we will demonstrate how to build a model predicting readmission for patients with diabetes in Python using the following steps

* data exploration
* data preprocessing
* data training and testing
* model evaluation

**CHAPTER 2 : REVIEW OF LITERATURE**

Reduction of preventable hospital readmissions that result from chronic or acute conditions like stroke, heart failure, myocardial infarction and pneumonia remains a significant challenge for improving the outcomes and decreasing the cost of healthcare delivery in the United States. Patient readmission rates are relatively high for conditions like heart failure (HF) despite the implementation of high-quality healthcare delivery operation guidelines created by regulatory authorities. Multiple predictive models are currently available to evaluate potential 30-day readmission rates of patients. Most of these models are hypothesis driven and repetitively assess the predictive abilities of the same set of biomarkers as predictive features. In this manuscript, we discuss our attempt to develop a data-driven, electronic-medical record-wide (EMR-wide) feature selection approach and subsequent machine learning to predict readmission probabilities. We have assessed a large repertoire of variables from electronic medical records of heart failure patients in a single center. The cohort included 1,068 patients with 178 patients were readmitted within a 30-day interval (16.66% readmission rate). A total of 4,205 variables were extracted from EMR including diagnosis codes (n=1,763), medications (n=1,028), laboratory measurements (n=846), surgical procedures (n=564) and vital signs (n=4). We designed a multistep modelling strategy using the Naïve Bayes algorithm. In the first step, we created individual models to classify the cases (readmitted) and controls (non-readmitted). In the second step, features contributing to predictive risk from independent models were combined into a composite model using a correlation-based feature selection (CFS) method. All models were trained and tested using a 5-fold cross-validation method, with 70% of the cohort used for training and the remaining 30% for testing. Compared to existing predictive models for HF readmission rates (AUCs in the range of 0.6-0.7), results from our EMR-wide predictive model (AUC=0.78; Accuracy=83.19%) and phenome-wide feature selection strategies are encouraging and reveal the utility of such data driven machine learning. Fine tuning of the model, replication using multi-center cohorts and prospective clinical trial to evaluate the clinical utility would help the adoption of the model as a clinical decision system for evaluating readmission status.

**CHAPTER 3: DATA COLLECTION**

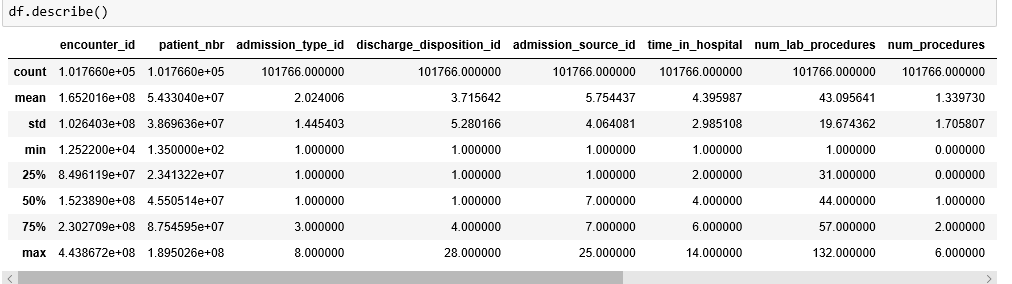
The data that is used in this project originally comes from the UCI machine learning repository (<https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008)>. The data consists of over 100000 hospital admissions from patients with diabetes from 130 US hospitals between 1999-2008.

**CHAPTER 4:** **METHODOLOGY**

**4.1 : IMPORTING THE LIBRARIES**

****

**4.2: DATA DESCRIPTION**

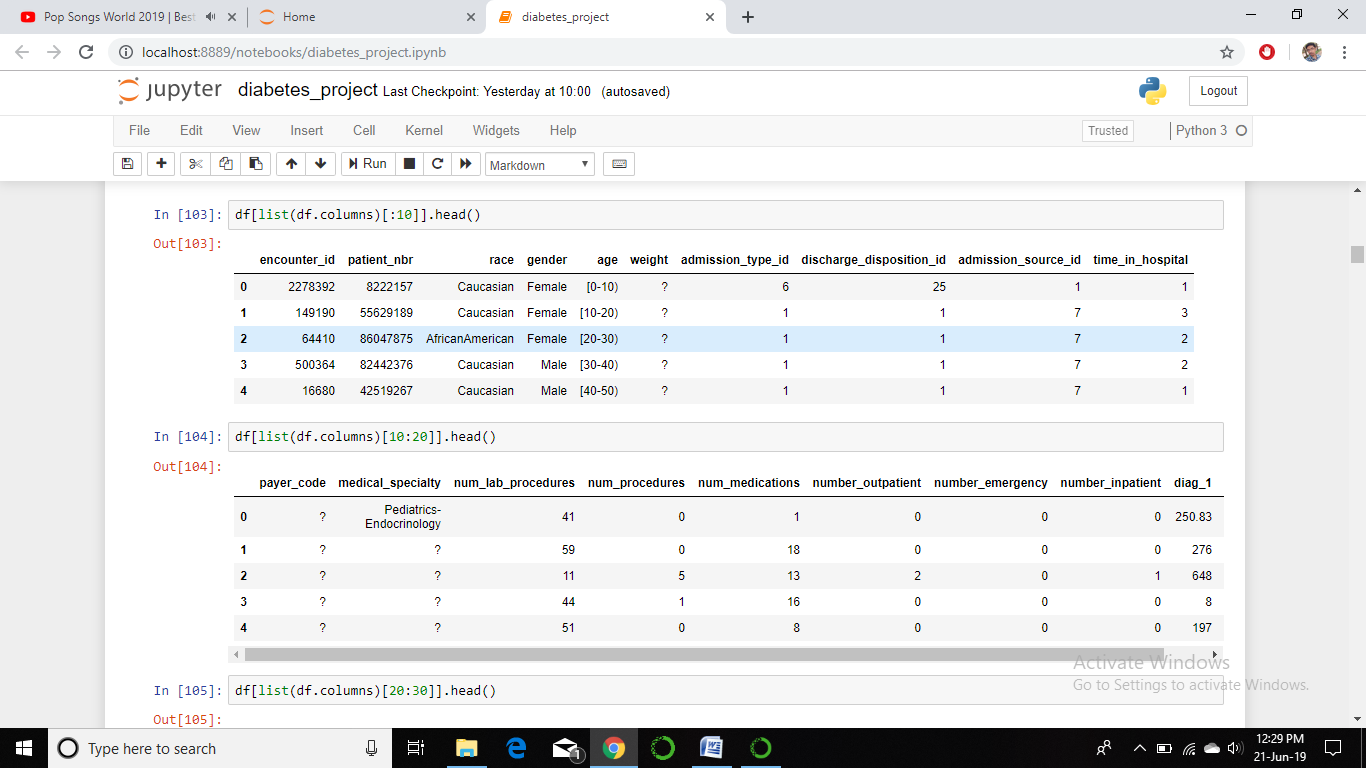
****

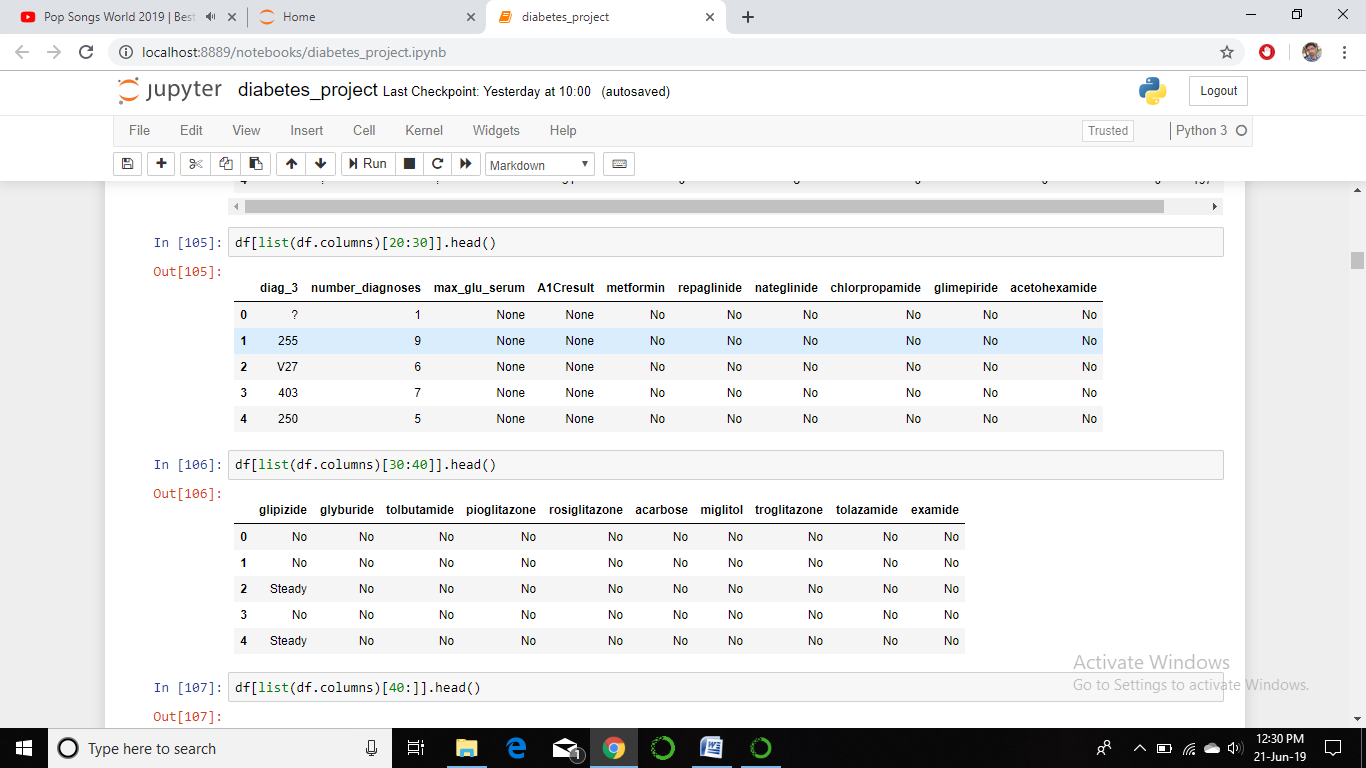
**4.3: TO DISPLAY THE TABLE**

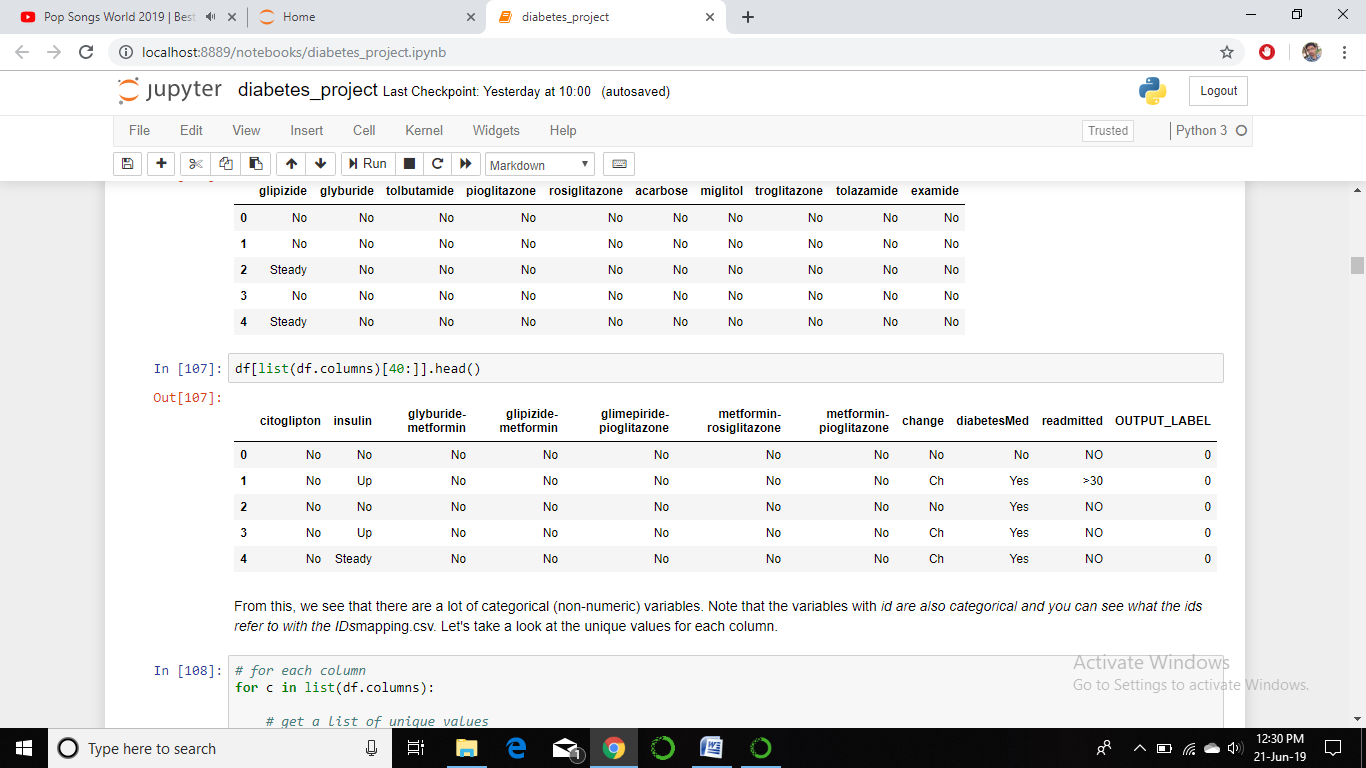
There is some missing data that are represented with ?. We will deal with this in the next section.

The most important column here is **readmitted**, which tells us if a patient was hospitalized within 30 days, greater than 30 days or not readmitted.

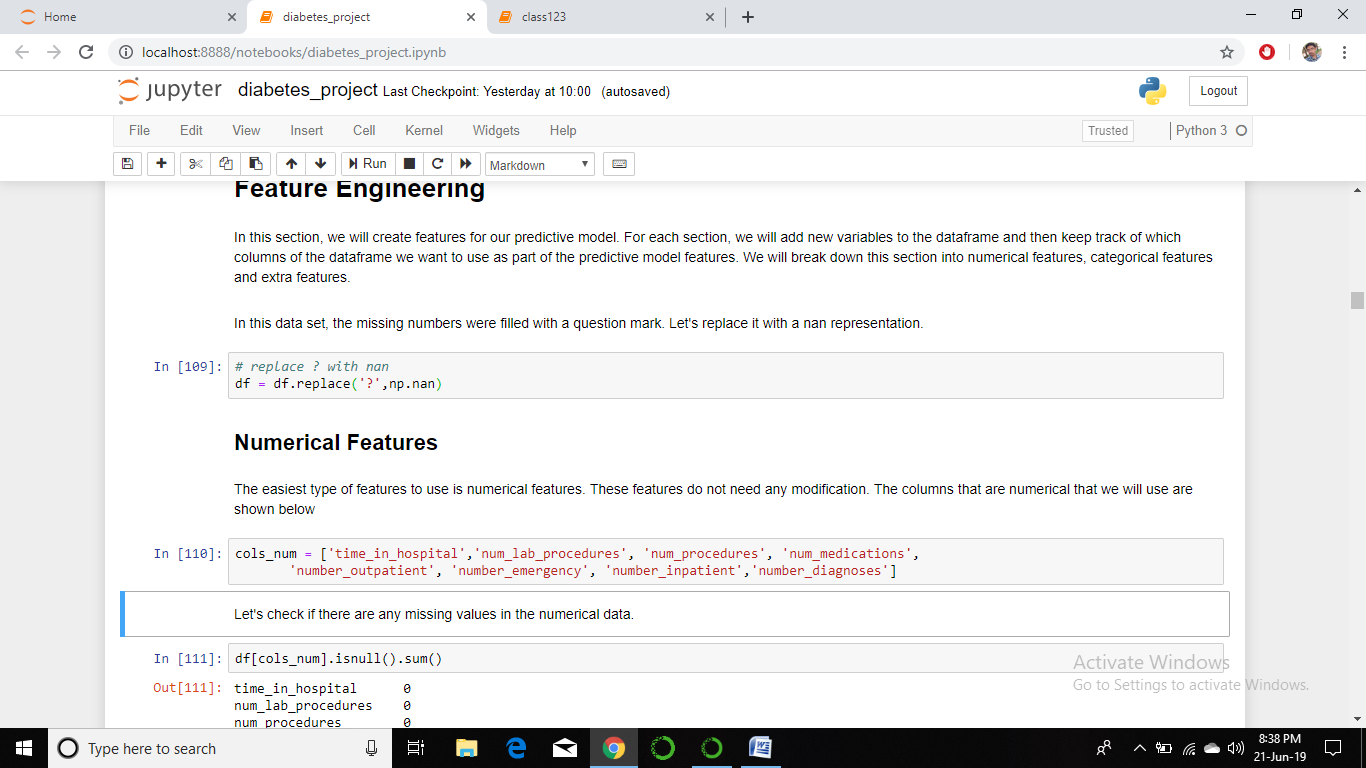
Pandas doesn't allow you to see all the columns at once, so let’s look at them in a group of 10.





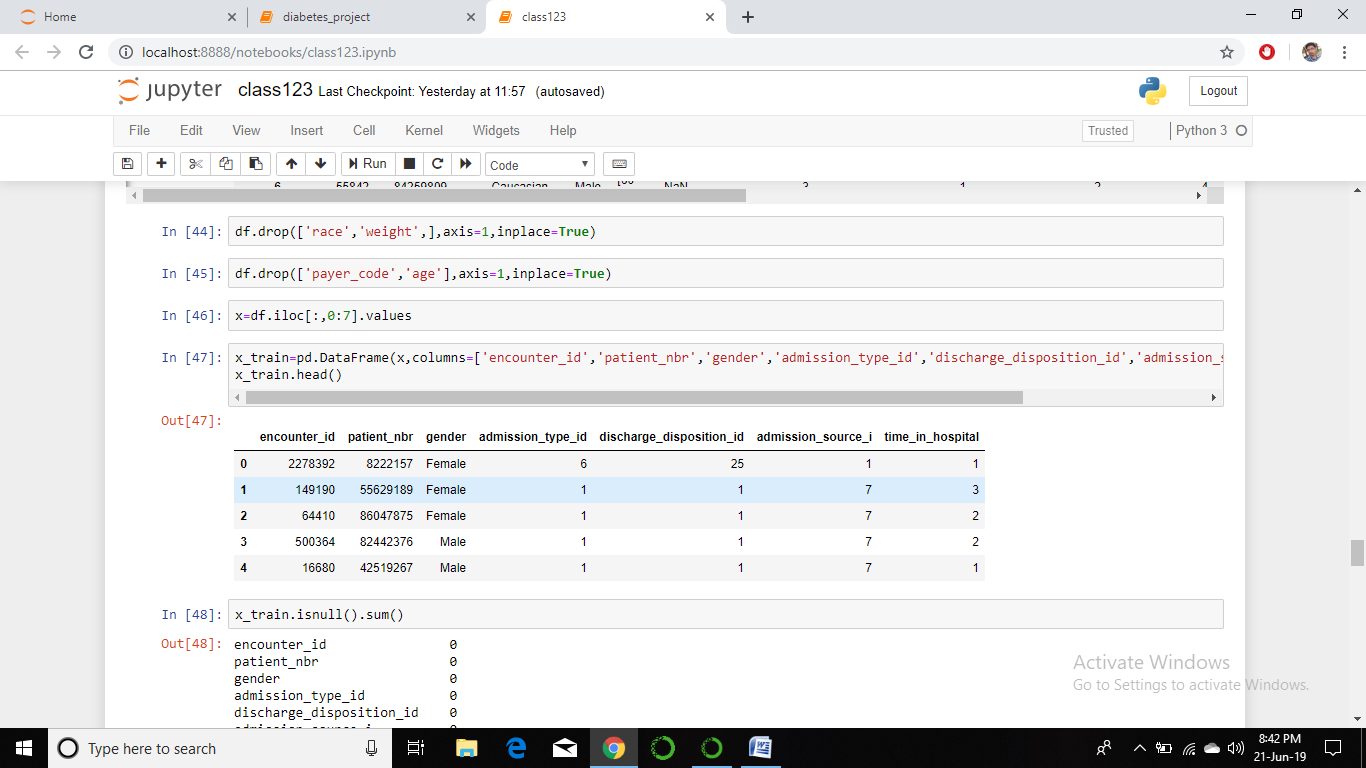


**4.4: REPLACING ‘?’ WITH nan**

****

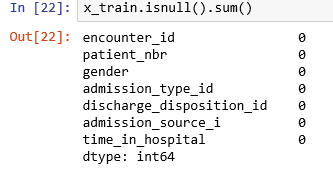
**4.5: DROPPING UNNECESSARY COLUMNS**

Dropping unnecessary columns and creating a training set with required data that are required for training the model so that we can predict the readmission rate.

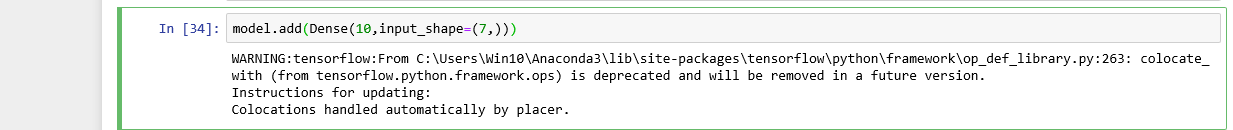
****

**4.6: TO CHECK THE DATA HAS NULL VALUES OR NOT**

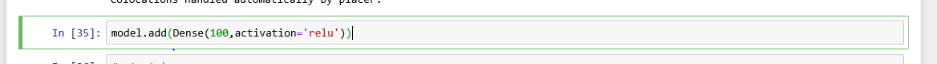
Checking the training set that we had created in above step contains any null values or not.

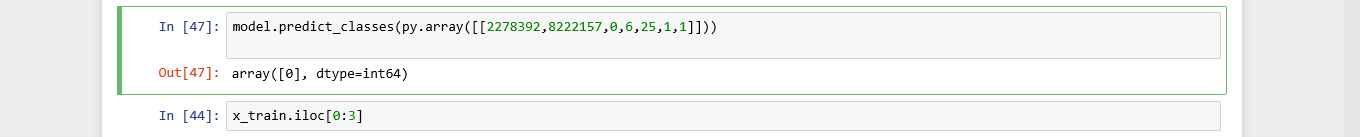
****

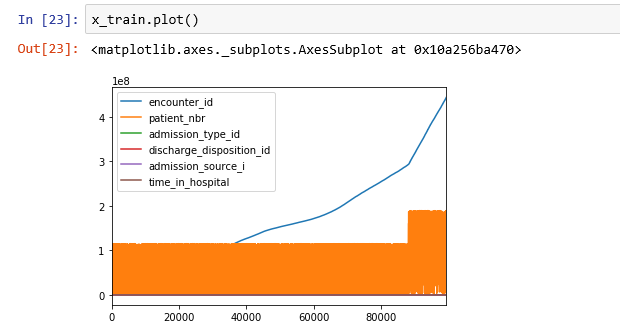
**4.7 : ADDING THE DENSE LAYERS**

****

**4.8: ACTIVATING FUNCTION**

****

**4.9: PREDICTING THE VALUES**

**4.10: PLOT **

**CHAPTER 5: FINDINGS AND SUGGESTIONS**

As we have done this project with a limited dataset, for a better result we can take a larger dataset to predict the hospital caring and also a better suggestion is that as we had stuck through a single model for prediction you can check all models available and can approach through best model that gives best accuracy by training the dataset.

The dataset in taken from the university of California(UoC) ,as diabetes is the common problem and the prediction will be better in that case you can have different dataset on different problems based on your interest.

**CHAPTER 6: CONCLUSION**

Through this project, we created a machine learning model that is able to predict the patients with diabetes with highest risk of being readmitted within 30 days. The best model was a gradient boosting classifier with optimized hyper parameters. The model was able to catch 58% of the readmissions and is about 1.5 times better than just randomly picking patients. Overall, I believe many healthcare data scientists are working on predictive models for hospital readmission.