# Project 1

# Diabetes 130-US hospitals for years 1999-2008 Data Set

Predicting hospital readmission for patients with diabetes using the dataset from UCI Machine Learning Repository

(<https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008>).

# Project 2

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | | Bike Sharing Dataset Data Set This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bike share system with the corresponding weather and seasonal information.  Interest : classic case for using regression |  | |

# Project 3

|  |  |
| --- | --- |
| |  | | --- | | Heart Disease Data Set | |

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to   
this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).   
  
The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.   
  
One file has been "processed", that one containing the Cleveland database. All four unprocessed files also exist in this directory.   
  
To see Test Costs (donated by Peter Turney), please see the folder "Costs"

**Data Source:**

<https://www.californiadgstats.ca.gov/downloads/#_csi_15_id>

**Independent variables:**

* <https://www.californiadgstats.ca.gov/downloads/#_csi_15_id>
* <https://www.californiadgstats.ca.gov/static/documents/interconnection_nem_pv/NEM_CurrentlyInterconnectedDataKey.b29667e204f3.xlsx>

Remarks : Might be a good way to learn clustering using K Means