



Exploring WEKA  
Database, Datawarehouse and dataset

# PL04

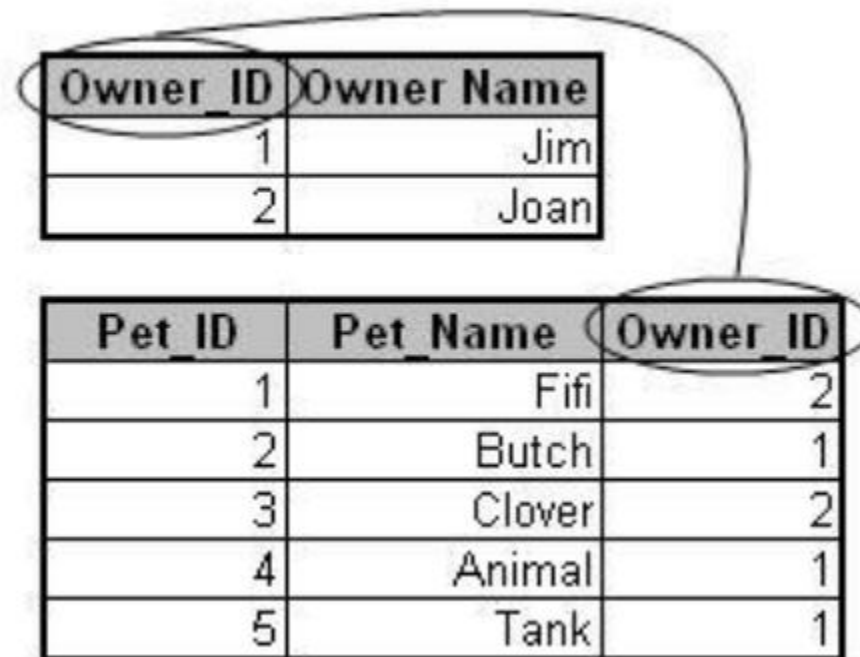


# Material

<http://hpeixoto.github.io/dc>

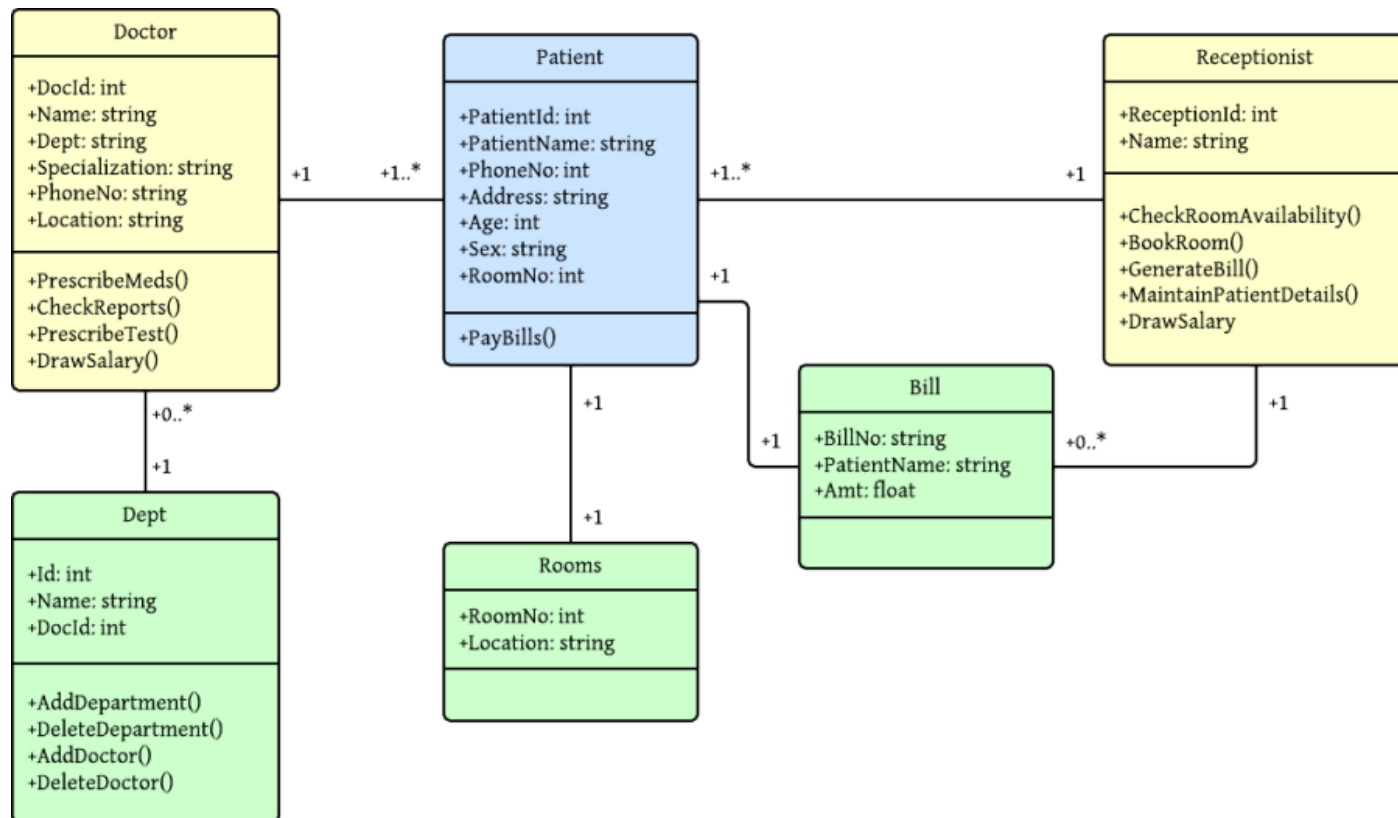


# DATA BASE





# DATA BASE





# DATA BASE

- Most relational databases which are designed to handle a high number of reads and writes (updates and retrievals of information) are referred to as **OLTP** (online transaction processing) systems.



# DATA BASE vs DATA WAREHOUSE

Owner_ID	Owner Name
1	Jim
2	Joan

Pet_ID	Pet_Name	Owner_ID
1	Fifi	2
2	Butch	1
3	Clover	2
4	Animal	1
5	Tank	1

Pet_ID	Pet_Name	Owner_Name
1	Fifi	Joan
2	Butch	Jim
3	Clover	Joan
4	Animal	Jim
5	Tank	Jim



# DATA WAREHOUSE

- Databases designed in this way are called OLAP systems
  - Online Analytical Processing
- Denormalization typically takes place at the time data is being copied out of transactional database.



# DATA WAREHOUSE

- Data mining is generally most effectively executed on data data sets, extracted from OLAP, rather than OLTP systems.





# DATA SET

- Subset of a database or a data warehouse.
- Creation of a data set includes:
  - Appending, combining, simplifying some data expressions
    - For exemple change data format



# TYPES OF DATA

- **Operational Data:**
  - the most elemental type of data, operational data, comes from transactional systems which record everyday activities;
- **Organizational Data:**
  - help to protect peoples' privacy, while still proving useful to data miners watching for trends in a given population.



# SECURITY & PRIVACY

**we should be aware that as we gather, organize and analyze data, there are real people behind the figures**

people have  
certain rights to privacy and protection against  
crimes such as identity theft



# BASE LINE EVALUATION

ZeroR

**Chooses the most common class and  
Classifies all instances with that value**



Exploring WEKA  
Database, Datawarehouse and dataset

# PL04