Data to Decision

Course URL:
http://pinformatics.org/phpm631

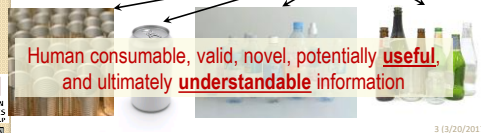
What is Data Science?

- Other words
  - Knowledge Discovery & Data mining (KDD)
  - Business Intelligence / Business Analytics
- Collecting and refining information from many sources
- Analyzing and presenting the information in useful ways
- So people can make better business decisions

Data Science
Knowledge Discovery & Data mining (KDD)

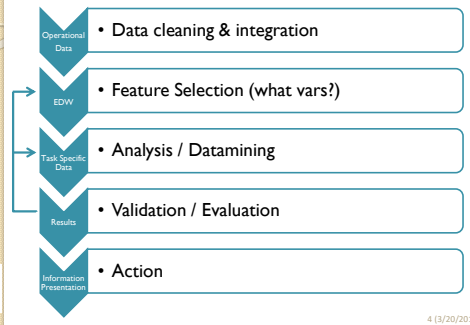
Big Data : operational data

KDD
Clean, Merge, Reprocess

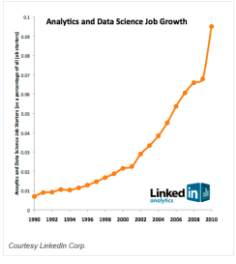


Human consumable, valid, novel, potentially **useful**, and ultimately **understandable** information

KDD Process



Job market of data scientists



Courtesy LinkedIn Corp.

- statisticians will be the next sexy job
  - Google Chief Economist Hal Varian
- shortage of 190,000 data scientists by the year 2019
  - McKinsey Global Institute

Take 10 minutes to
Download and read lab 6 & assignment 6

Lab 6: Tableau

- DUE in One Week (not tonight)
- Download the lab
- Follow the instructions and do them
- If you have questions, email for office hour

Assignment 6
Recommended Action Plan

- [Build a Team] Form a group of 3 to work together.
- [Problem Statement] Understand the question: Where would you recommend starting a new health services facility (e.g. hospital, physician office, FQHC, etc.)? What type of facility? Why?
- [Understand the Decision Maker] Assume you are writing the report to a board who will make the final determination about the new facility. You can make any assumptions about the makeup, characteristics, or nature of the board (e.g., they are a board for a hospital system with certain characteristics or they are a board for a local government unit with certain regional characteristics, etc.). When any assumption is made, you are required to state these assumptions in an appendix (which will not be part of the official report to the board) that will be submitted to me. So if there is any part of the assumptions that you need to convince the board of your proposal, remember to include these data/facts in the main report.
- [Feature selection] Figure out what data you need to answer the question
- [Data Analysis] Find the data you need, and analyze the data to answer the question based on evidence.
- [Develop a Story] Using the data you analyzed to arrive at your answer, develop a story in the form of a report. Use effective graphs and visualizations to convince the board to make the decision you are proposing. Make sure to review the requirements sections.

Assignment 6
Required Submissions I

- Write a report to answer the following question: Where would you recommend starting a new health services facility (e.g. hospital, physician office, FQHC, etc.)? What type of facility? Why?
- Progress Report : This is an informal memo to me. Submit a word document.
  - It needs to include information about
    - Team members
    - The tentative list of data (features) you plan to analyze. Include the source of the data. If you have a URL for the data, include it. If not, explain where you got the data, or plan to get the data.
    - [Optional] any plans on the data analysis you have

Assignment 6
Required Submissions II

- Final Report
  - The report should include an executive summary of no more than 350 words
  - The report must include at least 5 figures
  - The full report should be no more than 5 pages (not including the appendix)
  - [Appendix 1] Include an excel file, one sheet per figure that has the data that generated the figure
  - [Appendix 2] For each sheet (table in the excel file in appendix 1), write a SQL query that would generate the table, the data, needed for the report. You are not required to actually run the real SQL statements. But instead, write out the following for each SQL statement. You may do this either in word or excel.
    - Name the table (source of the data): then list the variables in the table (you do not need to list ALL the variables if the table is large. Just list the variables you need, plus a few more).
    - Include 1 row of data. The purpose is to demonstrate what kind of data each column holds.
    - Write the SQL query that would generate the data you needed for your report, the tables you included in the excel file (appendix 1), assuming you have the table in a database
    - List the name of the sheet in the excel file submitted in appendix 1, that holds the table you will generate with the SQL query
    - *NOTE: if you used a website to generate what you needed, it is very likely that your point and click resulted in a similar SQL query statement that was run behind the website.*
  - [Appendix 3] Remember to include the assumptions you made about the decision making board

SQL: Assignment 5

Course URL:
http://pinformatics.org/phpm631

Assignment 5

- Put your name IN the assignment file name
- Grading in assignment was lenient
  - Groups of 2 (not 3)
  - Will lose some points if you alter the base table
    - No ALTER or UPDATE
  - Will lost some points if not follow direction
    - Create & use VIEW (via programming)
    - Only query. No answer, when asked to give an answer
- Exam grading will be more strict

Q2: Create View

```
CREATE VIEW panel as
SELECT
      providers.fname AS dr_first,
      providers.lname AS dr_last,
      patients.fname,
      patients.lname,
      patientID
FROM providers, patients
WHERE
 providers.providerID=patients.primary_dr
ORDER BY
 providers.providerID;
```

13 (3/20/2017)

What's wrong ?

- Find two things that are wrong below

```
SELECT * FROM patients
ORDER BY lname;
WHERE race='H';
```

14 (3/20/2017)

ORDER matters
SQL – Structured Query Language

- Every statement yields a table of values as output
  - Sometimes there's only one row in the table!

Keyword	parameters
select	columns and/or expressions
from	Tables
where	conditions on the rows
group by	group rows together
order by	order the rows
;	

15 (3/20/2017)

Q1-Q3

- Most people got
- Many did not use views as instructed

16 (3/20/2017)

Q4.1: What's wrong

- How many visits did Lila Autry have?
- Who saw her on each visit?
- (*Hint: which columns, from which tables do you need to answer this question?)*

```
SELECT * FROM Visits WHERE patientID = "103";
```

17 (3/20/2017)

Q4.1 Good answer

- 1 visit
- "WILLIAM" "GAINES"

```
SELECT * FROM patients WHERE lname='AUTRY';

SELECT patientID FROM visits
WHERE patientID='103';

SELECT fname, lname FROM providers
WHERE providerID= '2' ;
```

18 (3/20/2017)

Q4.1 part 1: Better answer

- How many visits did Lila Autry have?
- 1 visit

```
SELECT patients.fname, patients.lname,
          patients.patientID,
COUNT(visits.visitID) AS visit
FROM patients, visits
WHERE patients.patientID = visits.patientID
      AND patients.fname = "LILA"
      AND patients.lname="AUTRY";
```

Q4.1 part 2: Better answer

- Who saw her on each visit?
- "WILLIAM" "GAINES"

```
SELECT visits.visitID,
       providers.fname, providers.lname
FROM patients, visits, providers
WHERE patients.patientID = visits.patientID
      AND visits.providerID = providers.providerID
      AND patients.fname = "LILA"
      AND patients.lname="AUTRY";
```

20 (3/20/2017)

Q4.1 part 2: Better answer: More interesting Example

- What happens if many visits?

```
SELECT visits.visitID,
       providers.fname, providers.lname
FROM patients, visits, providers
WHERE patients.patientID = visits.patientID
  AND visits.providerID = providers.providerID
  AND patients.lname= "JOHNSON";
```

21 (3/20/2017)

Q4.2 What is wrong?

- Who is her primary care doctor ?
- (the variable primary_dr holds this information)

```
SELECT primary_dr FROM patients
WHERE fname="LILA" AND lname="AUTRY";
```

22 (3/20/2017)

Q4.2 Good Answer

- Who is her primary care doctor ?
- (the variable primary_dr holds this information)
- DANA MAGOON

```
SELECT primary_dr FROM patients
WHERE fname="LILA" AND lname="AUTRY";

SELECT fname, lname FROM providers
WHERE providerID= '3' ;
```

23 (3/20/2017)

Q4.2 Better Answer

- Who is her primary care doctor ?
- (the variable primary_dr holds this information)
- DANA MAGOON

```
SELECT providers.fname, providers.lname
FROM patients, providers
WHERE patients.primary_dr= providers.providerID
      AND patients.fname = "LILA"
      AND patients.lname="AUTRY";
```

24 (3/20/2017)

Q4.3 Good Answer

- What were her diagnosis (all of them)? (*Hint: You will have to use the lookup table to find the meaning of the diagnosis code)*
- Don't forget to run the query to get dignosis

```
SELECT diag, diag_desc
FROM lu_diag
WHERE diag="6202";

SELECT diag, diag_desc
FROM lu_diag
WHERE diag="2859" ;
```

25 (3/20/2017)

Q4.3 Better Answer I

- What were her diagnosis (all of them)? (*Hint: You will have to use the lookup table to find the meaning of the diagnosis code)*

```
SELECT diag, diag_desc
FROM lu_diag
WHERE diag="6202" or diag="2859" ;
```

26 (3/20/2017)

Q4.3 Better Answer II

```
SELECT diag, diag_desc
FROM lu_diag
WHERE diag="6202" or diag="2859" ;

SELECT diag, diag_desc
FROM lu_diag
WHERE diag in ("6202", "2859") ;
```

27 (3/20/2017)

Q4.4 Answer I

- Were any medications prescribed?
- If so, how many and what?
- 3 prescriptions
- 861204-30,
- 562812-60, HISTAMINE H2-ANTAGONISTS
- 280808-12, OPIATE AGONISTS

```
SELECT medications.ahfs_cd,
      medications.ahfs_desc, medications.rx_quantity
FROM patients, medications
WHERE patients.patientID = medications.patientID
      AND patients.fname = "LILA"
      AND patients.lname = "AUTRY";
```

28 (3/20/2017)

Q4.4 Answer II

- Were any medications prescribed?
- If so, how many and what?
- 3 prescriptions
- 861204-30,
- 562812-60, HISTAMINE H2-ANTAGONISTS
- 280808-12, OPIATE AGONISTS

```
SELECT m.ahfs_cd, m.ahfs_desc, m.rx_quantity
FROM patients, medications m
WHERE patients.patientID = m.patientID
      AND patients.fname = "LILA"
      AND patients.lname = "AUTRY";
```

29 (3/20/2017)

Q4.4 Anything odd?

```
select patients.fname, patients.patientID,
      visits.diag1, visits.diag2, visits.diag3,
      lu_diag.diag_desc
from patients
inner join visits
on patients.patientID=visits.patientID
inner join lu_diag
on lu_diag.diag=diag1
      or lu_diag.diag=diag2
      or lu_diag.diag=diag3
where patients.fname='LILA';
```

30 (3/20/2017)

Why?

	fname	patientID	diag1	diag2	diag3	diag_desc
1	LILA	103	6202	2859		ANEMIA NOS
2	LILA	103	6202	2859		OVARIAN CYST NEC...

Q5: Debug?

- Write a query that displays patients who visited in the second quarter of 2012 together with the number of visits that quarter. Order the output by the number of visits (patients with the largest number of visits should be displayed first). Each patient name should appear once.
- *Hint: Join tables "patients" and "visits" and use Group By and Count. You'll also need to use a Where for the date range and Sort on (Order by) the Count.*

```
SELECT patientID, COUNT(visits.patientID)
FROM visits
GROUP BY patientID
ORDER BY COUNT(visits.patientID) DESC;
```


Q5: Debug?

- Write a query that displays patients who visited in the second quarter of 2012 together with the number of visits that quarter. Order the output by the number of visits (patients with the largest number of visits should be displayed first). Each patient name should appear once.
- How to add names?

```
SELECT patientID, COUNT(visits.patientID)
FROM visits
where dtl_qtr="2012QTR2"
GROUP BY patientID
ORDER BY COUNT(visits.patientID) DESC;
```


Q5: Answer

- Write a query that displays patients who visited in the second quarter of 2012 together with the number of visits that quarter. Order the output by the number of visits (patients with the largest number of visits should be displayed first). Each patient name should appear once.
- *Hint: Join tables "patients" and "visits" and use Group By and Count. You'll also need to use a Where for the date range and Sort on (Order by) the Count.*

```
SELECT patients.fname, patients.lname,
       visits.dtl_qtr, patients.patientID,
       COUNT(visits.visitID) AS visitcount
FROM patients, visits
WHERE patients.patientID=visits.patientID
      AND visits.dtl_qtr="2012QTR2"
GROUP BY visits.patientID
ORDER BY visitcount desc;
```

Q5: Answer II

```
select p.fname first_name, p.lname last_name,
      count(*) number_of_visits
from patients p
join visits v
on p.patientid=v.patientid
where v.dtl_qtr="2012QTR2"
group by p.patientid
order by count(*) desc;
```

Q5: Answer III

```
select p.fname first_name, p.lname last_name,
      count(*) number_of_visits
from patients p
join visits v
on p.patientid=v.patientid
where v.dtl_qtr="2012QTR2"
group by p.patientid
order by number_of_visits desc;
```

Q5. Debug?

```
SELECT visits.patientID, patients.fname,
       patients.lname, visits.dtl_qtr
       COUNT(visits.patientID) AS numberofvisits
FROM visits
INNER JOIN patients
ON visits.patientID=patients.patientID
WHERE dtl_qtr="2012QTR2"
GROUP BY patientID
ORDER BY numberofvisits DESC;
```

37 (3/20/2017)

Q5. Debug?

```
SELECT visits.patientID, patients.fname,
       patients.lname, visits.dtl_qtr,
       COUNT(visits.patientID) AS numberofvisits
FROM visits
INNER JOIN patients
ON visits.patientID=patients.patientID
WHERE dtl_qtr="2012QTR2"
GROUP BY patients.patientID
ORDER BY numberofvisits DESC;
```

38 (3/20/2017)

Q6: what's wrong

```
select patients.fname, patients.lname,
        billed, covered,
        sum(copay+pat_pd+insur_pd) as totalpaid,
        (billed - covered) as unpaid
from payments
join patients
on payments.patientID = patients.patientID
group by patients.patientID;
```

39 (3/20/2017)

Q6: what's wrong

```
select patients.fname, patients.lname,
        billed, covered,
        sum(copay+pat_pd+insur_pd) as totalpaid,
        (billed - covered) as unpaid
from payments
join patients
on payments.patientID = patients.patientID
group by patients.patientID;
```

40 (3/20/2017)

Q6. Debug

- Write a query that displays the **billed, covered,** total paid (=**copay+pat_pd+insur_pd**), and unpaid amount (=**covered**-total paid) for each patient. Save this as a permeant view so that you can use this again in the next task.
- *Hint: You'll need to calculate new columns for the total paid amount and the unpaid amount. You'll also have to Group By and Sum.*

```
CREATE VIEW PatientBalance AS
SUM(copay) + SUM(pat_pd) + SUM(insur_pd) AS
      TotalPaid
SUM(covered) - TotalPaid AS TotalUnpaid
FROM payments
GROUP BY patientID;
```

41 (3/20/2017)

Q6. Debug

- Write a query that displays the **billed, covered,** total paid (=**copay+pat_pd+insur_pd**), and unpaid amount (=**covered**-total paid) for each patient. Save this as a permeant view so that you can use this again in the next task.
- *Hint: You'll need to calculate new columns for the total paid amount and the unpaid amount. You'll also have to Group By and Sum.*

```
CREATE VIEW PatientBalance AS
SELECT
SUM(copay) + SUM(pat_pd) + SUM(insur_pd) AS
      TotalPaid ,
SUM(covered) - TotalPaid AS TotalUnpaid
FROM payments
GROUP BY patientID;
```

42 (3/20/2017)

Q6. Answer

- Write a query that displays the **billed, covered,** total paid (=**copay+pat_pd+insur_pd**), and unpaid amount (=**covered**-total paid) for each patient. Save this as a permeant view so that you can use this again in the next task.
- *Hint: You'll need to calculate new columns for the total paid amount and the unpaid amount. You'll also have to Group By and Sum.*

```
create view newdata as
select patientID, sum(billed) as billed,
      sum(covered) as covered,
      sum(copay+pat_pd+insur_pd) as paid,
      sum(covered)-sum(copay+pat_pd+insur_pd ) as unpaid
from payments
group by patientID;
```


Q6. Answer

- Write a query that displays the **billed, covered,** total paid (=**copay+pat_pd+insur_pd**), and unpaid amount (=**covered**-total paid) for each patient. Save this as a permeant view so that you can use this again in the next task.
- *Hint: You'll need to calculate new columns for the total paid amount and the unpaid amount. You'll also have to Group By and Sum.*

```
CREATE VIEW Money AS
SELECT patientID, SUM(billed), SUM(covered),
  SUM(copay+pat_pd+insur_pd) AS [Total Paid],
  SUM(covered-copay+pat_pd+insur_pd) AS [Unpaid Amount]
FROM payments
GROUP BY patientID;
```


Q6. What's wrong?

```
ALTER table payments ADD total_paid;
ALTER table payments ADD unpaid_amount;
UPDATE payments
      SET total_paid = copay + pat_pd + insur_pd;
UPDATE payments
      SET unpaid_amount = covered - total paid;
CREATE view billing AS
SELECT patientID, billed, covered,
      total_paid, unpaid_amount
FROM payments
GROUP BY payments.patientID;
```


Not changing base table

- ALTER
- UPDATE
- Normally do not have permission
- Query has new columns

Q7. Answer

- Using the view from Q6, write a query that displays which patients have not paid their billed amount yet, and what is the outstanding bill amount? Order the output by outstanding bill amount with large amount on top.
- *Hint: Join tables "patients" with the view and sort. You'll also need to use a Where to find the patients with outstanding bills. Try using greater than 1 cent.*

```
select patients.lname, patients.fname,
       patients.patientID, newdata.billed,
       newdata.covered, newdata.paid, newdata.unpaid
from patients, newdata
where newdata.unpaid>0.01
       and patients.patientID=newdata.patientID
order by unpaid desc;
```


Q7. Anything wrong?

```
SELECT billing.patientID, patients.fname,
patients.lname, billing.unpaid_amount
FROM patients
INNER JOIN billing
on billing.patientID=patients.patientID
WHERE unpaid_amount > 0.01
GROUP BY billing.patientID
ORDER BY unpaid_amount DESC;
```

Q7. Anything wrong?

- Why group by? Already one row per patientID

```
SELECT billing.patientID, patients.fname,
patients.lname, billing.unpaid
FROM patients
INNER JOIN billing
on billing.patientID=patients.patientID
WHERE unpaid > 0.01
GROUP BY billing.patientID
ORDER BY unpaid DESC;
```


Q7. What about <>0 ?

- <> 0 vs >0.1

```
SELECT billing.patientID, patients.fname,
patients.lname, billing.unpaid_amount
FROM patients
INNER JOIN billing
on billing.patientID=patients.patientID
WHERE unpaid_amount <> 0
GROUP BY billing.patientID
ORDER BY unpaid_amount DESC;
```


What's wrong?

```
CREATE VIEW [Hispanic Patients] AS
SELECT DISTINCT lname AS [Last Name],
      fname AS [First Name],
      race AS [Race]
FROM patients
WHERE race='H'
GROUP BY lname;
```


What's wrong?

```
CREATE VIEW [Hispanic Patients] AS
SELECT DISTINCT lname AS [Last Name],
      fname AS [First Name],
      race AS [Race]
FROM patients
WHERE race='H'
GROUP BY lname;
```


Table Operations

- Aggregate columns: col1 op col2 AS col3

col1	col2
a	d
b	e
c	f

→

col1	col2	col3
a	d	a+d
b	e	b+e
c	f	c+f

- Aggregate rows: Group BY

A
B
C

→ D

Where D=function(A,B,C)
Examples of function are
Sum(A,B,C) Avg(A,B,C) Max(A,B,C) Min(A,B,C) Count(A,B,C)

Conditional Operators

Operator	Description
=	Equal
<>	Not equal. Note: In some versions of SQL this operator may be written as !=
>	Greater than
<	Less than
>=	Greater than or equal
<=	Less than or equal
BETWEEN	Between an inclusive range
LIKE	Search for a pattern (approximately same)
IN	To specify multiple possible values for a column

Advanced conditionals (Optional)
Like & Wildcard

Wildcard	Description
%	A substitute for zero or more characters
_	A substitute for a single character
[charlist]	Sets and ranges of characters to match
[^charlist] or [!charlist]	Matches only a character NOT specified within the brackets

```
SELECT fname, lname
FROM patients
WHERE lname LIKE 's%' ;
```

Break at 7:15
Midterm 7:30-8:30