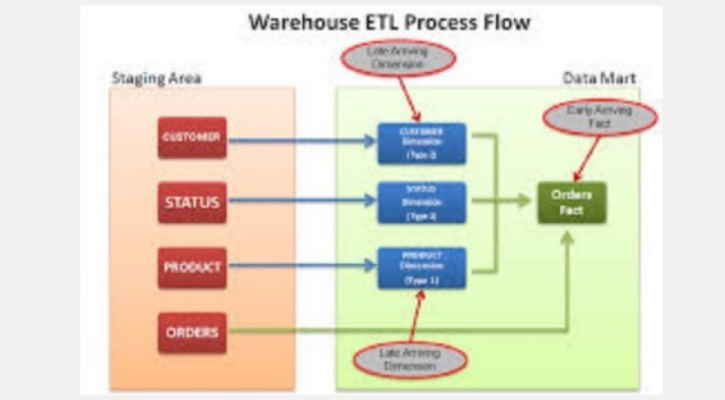
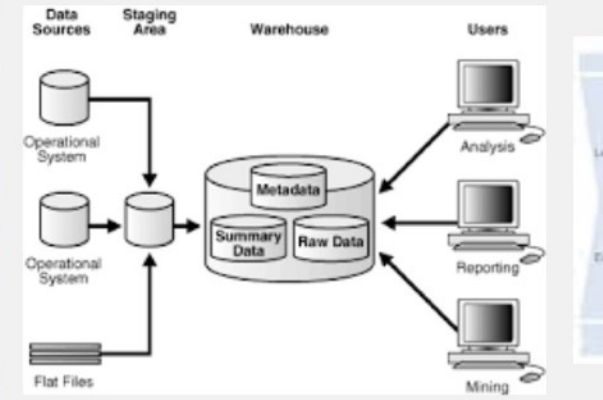
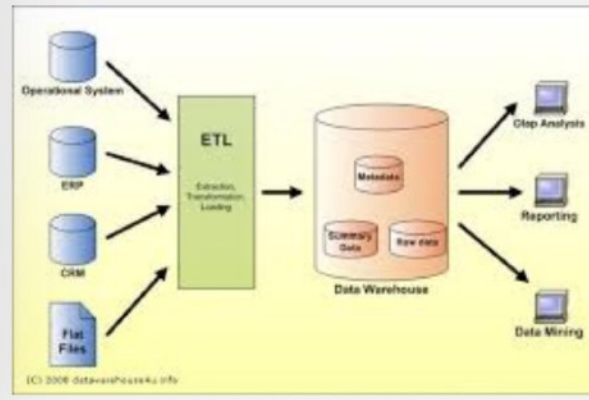
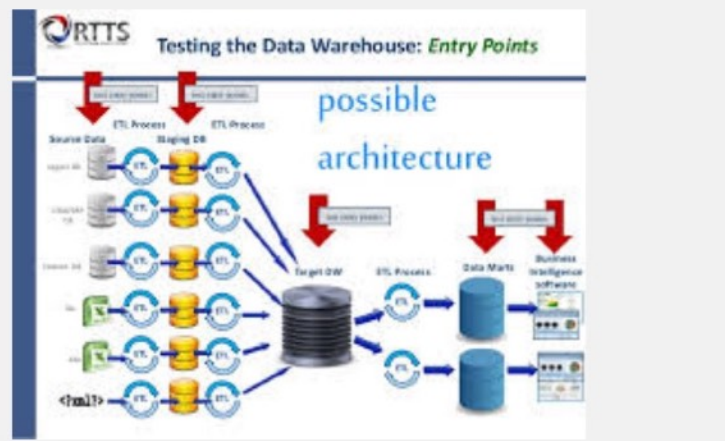
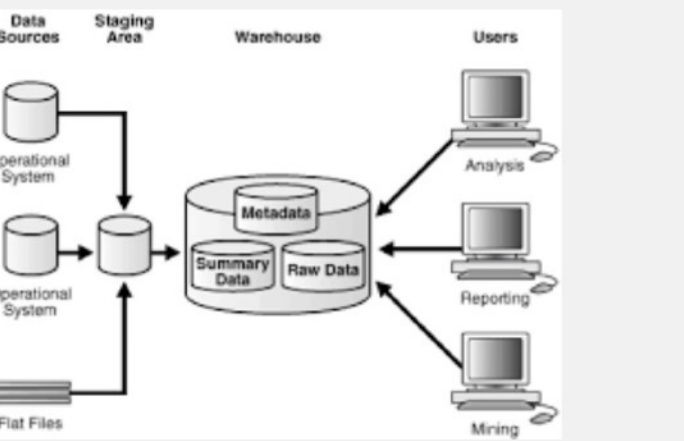
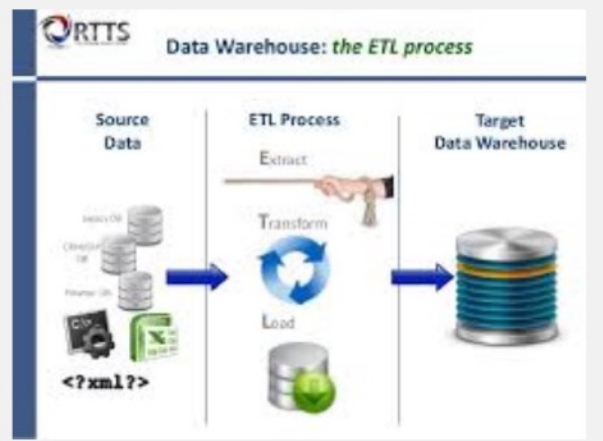
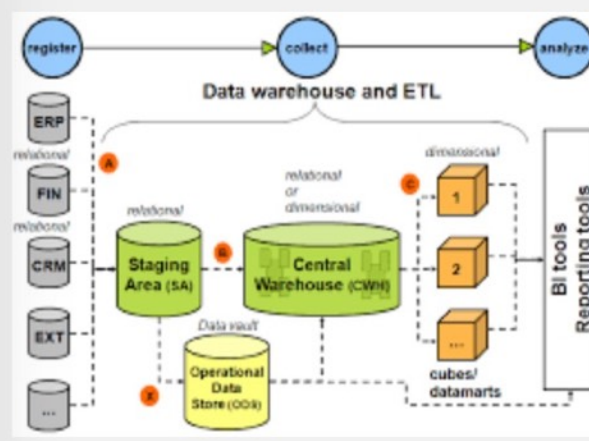
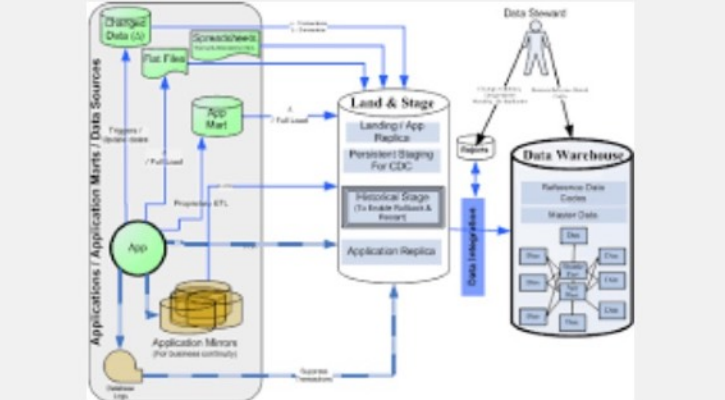
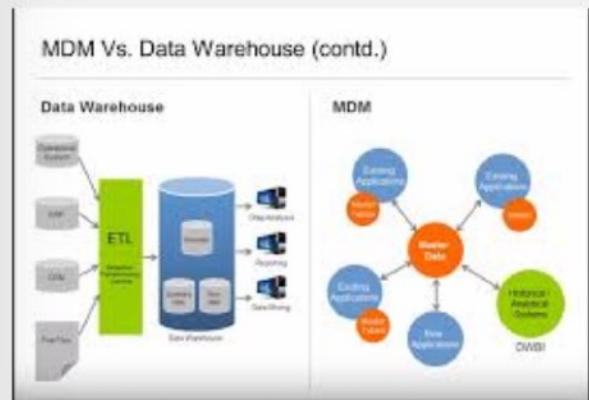


# Edmonton Data Management Meetup

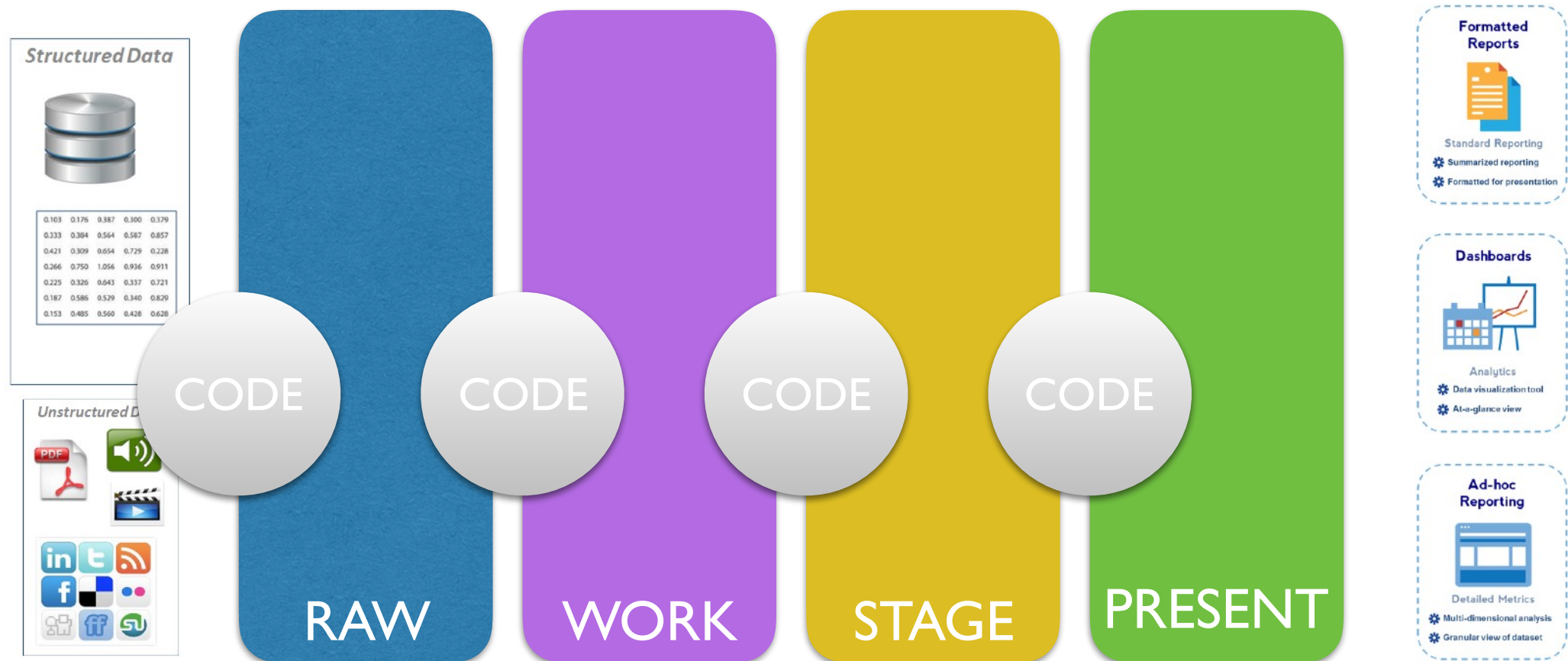
20170501

# Agenda

- Member presentations
  - Always open to short, casual talks
  - Something you're knowledgeable in, interested in, or a past/current work project that you'd appreciate feedback about
- The Edmonton Open Data Project
  - Our base / template project on Data Management
  - A self-directed case study - **we** are the client
  - Walk through a Data Management process



# Generic Terms

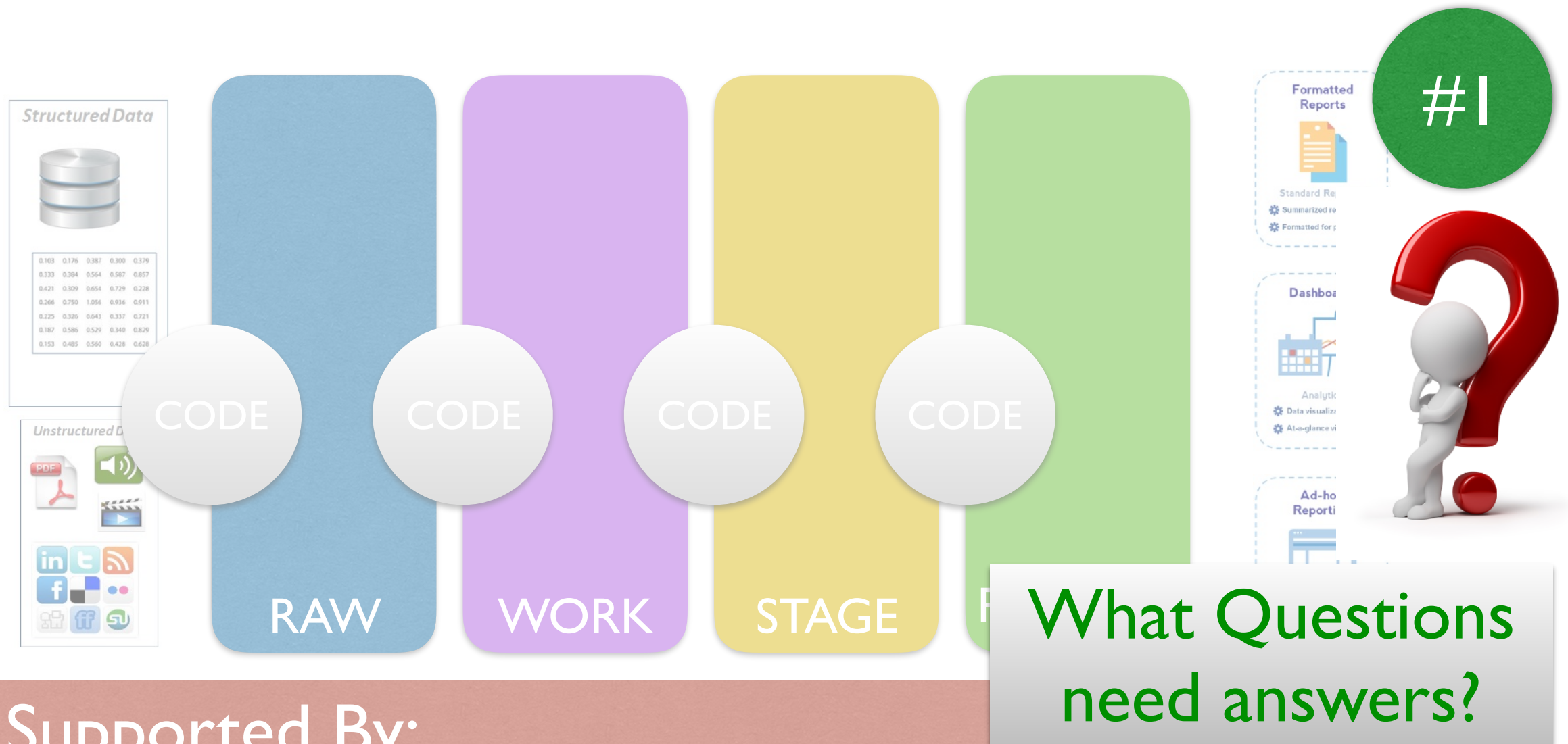


## Supported By:

- MetaData
- Security
- Logging
- Profiling
- Versioning
- Archiving
- Audibility
- Operations
- Testing
- User Group



# Generic Terms



## Supported By:

- MetaData
- Security
- Logging
- Profiling
- Versioning

- Archiving
- Audibility
- Operations
- Testing
- User Group

# Generic Terms

#2



CODE

CODE

CODE

CODE

W

WORK

STAGE

PRESENT

What data is required?  
Is it available?

By:

- MetaData
- Security
- Logging
- Profiling
- Versioning

- Archiving
- Audibility
- Operations
- Testing
- User Group

#1



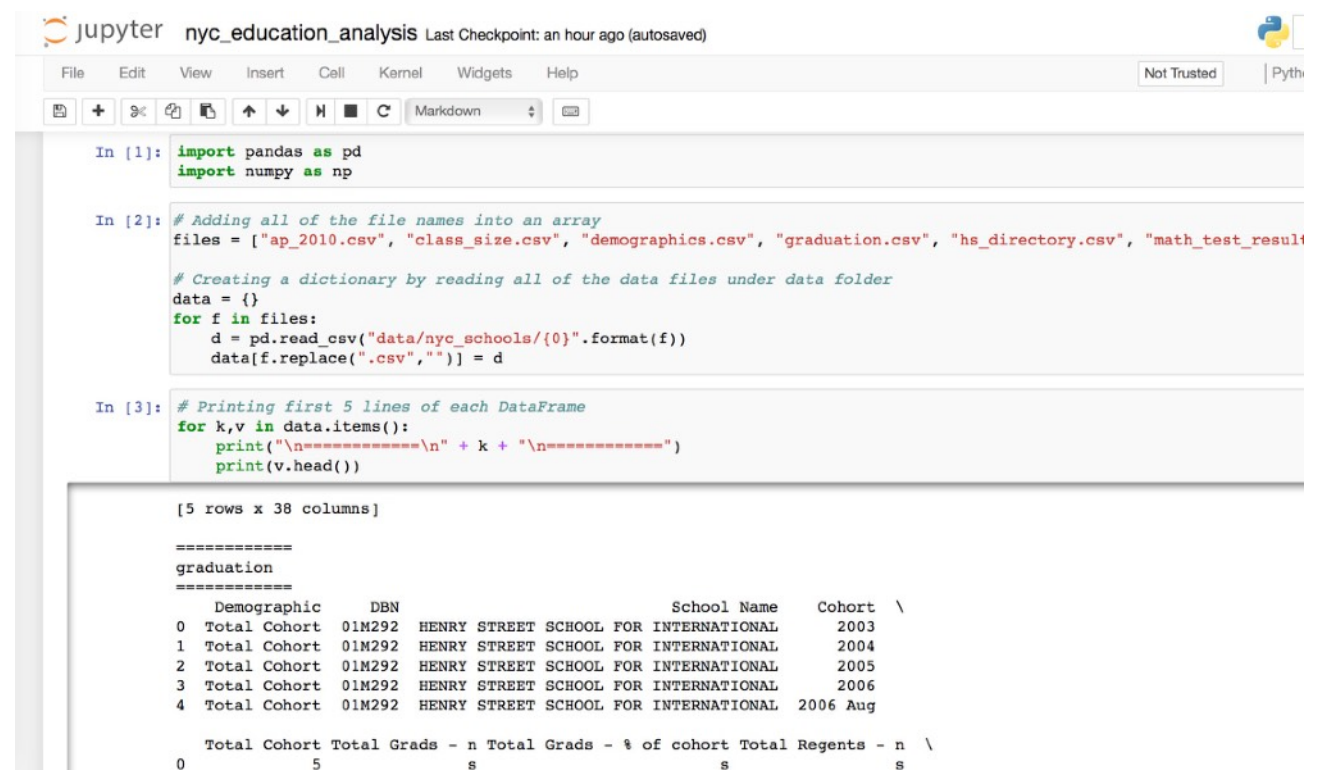
# Generic Terms





# Example

- NYC OpenData Example
- **QUESTION:**  
What is the fairness of the SAT testing in NYC schools?



```
jupyter nyc_education_analysis Last Checkpoint: an hour ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help
In [1]: import pandas as pd
import numpy as np

In [2]: # Adding all of the file names into an array
files = ["ap_2010.csv", "class_size.csv", "demographics.csv", "graduation.csv", "hs_directory.csv", "math_test_results.csv"]

# Creating a dictionary by reading all of the data files under data folder
data = {}
for f in files:
    d = pd.read_csv("data/nyc_schools/{0}".format(f))
    data[f.replace(".csv", "")] = d

In [3]: # Printing first 5 lines of each DataFrame
for k,v in data.items():
    print("\n===== \n" + k + " \n=====")
    print(v.head())

[5 rows x 38 columns]

=====
graduation
=====
   Demographic  DBN  School Name  Cohort \
0 Total Cohort  01M292  HENRY STREET SCHOOL FOR INTERNATIONAL  2003
1 Total Cohort  01M292  HENRY STREET SCHOOL FOR INTERNATIONAL  2004
2 Total Cohort  01M292  HENRY STREET SCHOOL FOR INTERNATIONAL  2005
3 Total Cohort  01M292  HENRY STREET SCHOOL FOR INTERNATIONAL  2006
4 Total Cohort  01M292  HENRY STREET SCHOOL FOR INTERNATIONAL  2006 Aug

Total Cohort Total Grads - n Total Grads - % of cohort Total Regents - n \
0 5 s s s s
```



# Plan

- Tonight

- Work on the first 2 steps
  - Identifying some interesting Questions
  - Begin associating data sources

- Following Meet-Ups

- Process data sources
  - cleaning & restructuring
  - developing new datasets that support questions
- Investigate Support requirements
  - share ideas and discuss methods and techniques
- There **may** be homework ;-)



Make sense?

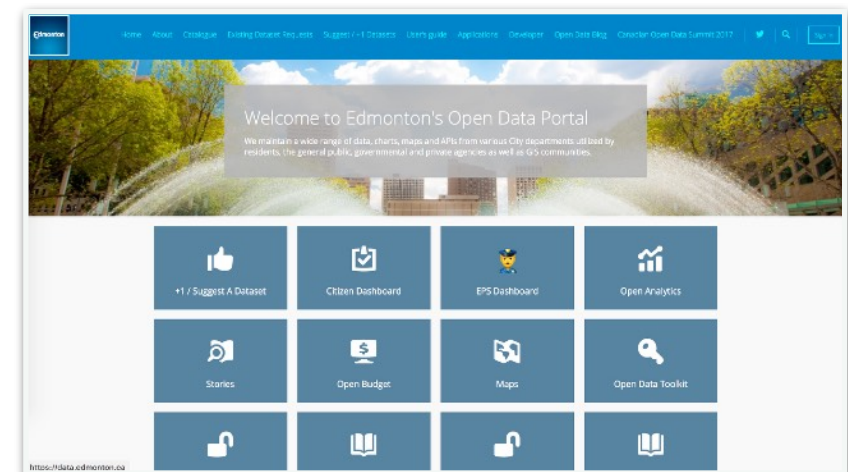
# Questions

- Arrange into groups
- Start brainstorming questions
  - Edmonton focussed
  - No bad questions
    - just may not find data to be answer them all

Edmonton Data Science Meetup 2017-05-01 Meeting		Group members: _____
Data Management Project Brainstorming		
Questions / Topics of Interest	Types of Data Needed	Does Data Exist?

# Data

- Start looking through YEG Open Data
  - Or other sources you know of
- What data would help answer the question
  - Search by category and tag
- <https://data.edmonton.ca>







# Discussion