

DATA MINING

MSBO Certification course

Janell Craig, Senior Business Analyst - Oakland Schools
Grace Damerow, Interim Director of Applications - Oakland Schools



Join at
slido.com
#3737 820

🔒 Passcode:
zwlv9k

slido



Describe your day in one or two words.

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

Data Mining

- Session Agenda
 - Define Data Mining
 - Sources of data
 - Data Mining tools/techniques
 - XML editors
 - Working with Access
 - Working with Excel
- Questions

slido




How do you define data mining?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.




Data Mining - Not Just Report Writing

- Data Mining is a process or set of processes which examine large amounts of data to discover valuable information
 - Focuses on finding relevant information in data sets which can then be used in analytics and predictive modeling
 - Data Mining processes involve using software, algorithms and/or statistical methods to:
 - Find patterns in data
 - Help answer questions and explain changes
 - Predict future trends
- 




Data Mining in Education

- Educational Data Mining extracts useful knowledge which could potentially affect teaching and learning processes
 - Districts collect, maintain and store large amounts of student data, which can inform decisions made to increase teaching and learning outcomes
 - Whether it be student achievement, effective district policy and procedures, or simply reviewing results of data governance policies, data mining in education can
 - Validate and evaluate systems and
 - Improve process quality
- 




Data Mining Techniques

- Classification
 - Classifies data items into categories (# of students w/A, B, C etc)
 - Clustering
 - Groups data into clusters that share common characteristics (grades on mid-term exams by group; Gender, Extra Curricular Activities)
 - Association Rule
 - Relationships between two or more data items (if, then); student engagement in classroom when compared to seating, course arrangement, content etc.
 - Anomaly detection/Outlier Analysis
 - Identifies rare or unexpected data items from expected results
 - Pattern Recognition
 - Looking for patterns in data
- 




Data Mining in Your Daily Life

- Turn to the table behind you - Meet someone new!
 - Introduce yourself and where you are from
 - To review the methods introduced, discuss the following
 - Identify two instances in professional or personal life you have experienced data mining techniques listed...
 - Classification (grouping)
 - Clustering (shared characteristics)
 - Association Rule (If, Then Analysis)
 - Anomaly/Outlier detection (breaks from expected results)
 - Pattern Recognition
- 



Primary Steps to Data Mining

1. Identify the Issue or Question
 2. Gather the Data
 3. Prepare the Data
 4. Mine the Data
 5. Interpret the Data
 6. Implement Change or Continue to Monitor
- 



Five Primary Steps to Data Mining

- Identify the Issue or Question
 - The why will always inform the what
 - What is “it” that you are looking for?
 - Understanding “what” determines expected results
 - Helps determine patterns, recognize outliers
- Examples
 - Can you name a recent data request from others or data you needed recently.
 - How did you Identify?




Five Primary Steps to Data Mining

- Gather the Data

- Understanding the issue determines the dataset(s) needed which will be used or compared against each other
- Know the data owners and where to get source data from
- Ensure datasets are equivalent (ie apples to apples)

- Examples

- Various examples of gathering data from sources will be discussed and detailed momentarily
- 



Five Primary Steps to Data Mining

- Prepare the Data
 - When gathering large amounts data, ensure data are checked for reasonableness, assessed for mistakes and standardized
 - If comparing two sets of data, ensure headings are created or renamed the same where appropriate
- Examples
 - Standardizing: Student demographic file source from MSDS, ethnicity is numeric (000000) where data from a another source might indicate “Female”, “F” or Database table code “4”.



Five Primary Steps to Data Mining

- Mine the Data
 - Determine the best method and tool to mine the data
- Examples
 - Various examples of mining data from sources will be discussed and detailed momentarily



Five Primary Steps to Data Mining

- Interpret the Data
 - Outcomes may be aggregated or detailed, interpreted, and/or presented to decision-makers that have largely been excluded from the data mining process to this point
- Examples
 - District administrators may compare virtual self paced learners in similar student classifications against in-person instruction?
 - Data Coordinators/State reporters may review data quality results from review to enhance data governance or data quality measures



Five Primary Steps to Data Mining

- Implement Change or Continue to Monitor
 - Ultimately, data mining processes will conclude change management or continuous monitoring based on the story the data tells.
- There is no right or wrong answer
 - Dig into the data
 - Use understanding of methods and tools to find patterns, outliers and differences

slido



Let's brainstorm DATA SOURCES you may use in your own data mining.

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.



Data Sources

- Gathering the Data - Data May Come from Any Source
 - Student Information System / Human Resource, Finance Systems
 - Michigan Data Hub
 - CEPI MSDS/REP Collection Submissions
 - MiSchool Data Portal
 - Others?
 - Learning Management, Assessment Systems



Data Sources - Student Information Systems

- Student Information Systems - Details not captured in any other place
 - Demographic Data
 - Attendance
 - Behavior and Discipline
 - Program Participation
 - Food Service
 - Special Education
 - Historical Data
 - Other?



Data Sources - Mi Data Hub

- MI Data Hub (<https://www.midatahub.org>)
 - Data in Hub typically consists of Ed-Fi Standardized data from SIS
 - Data are stored in what is called an Operational Data Store (ODS)
 - ODS's are school year based
 - Other Systems pull data from the district's data hub ODS
 - Some systems push data to the ODS (not many)
 - One example is State Assessment Data






Data Sources - CEPI

- CEPI Submission Downloads (<https://cepi.state.mi.us/msds/>)
 - Files submitted and edited during submission process
 - Contain data submitted for funding at a specific point in time
 - May differ from dataset pulled from a SIS
 - Data sets are not historical; submissions are purged on a rotating year basis



Data Sources - MiSchoolData

- MiSchoolData (<https://www.mischooldata.org/>)
 - Site maintained by CEPI
 - Public AND Private data
 - Primarily summarized, charted data
- 



Sources of Data

Connecting the dots.





Data Analysis Tools

slido



What does XML mean?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

XML Editors

- XML editors
 - Altova Spy (\$\$)
 - Notepad ++
 - There are other Free applications

XML Editors

(open sample file in Notepad++)

<https://bit.ly/XML1>

<https://bit.ly/XML2>




MS Access




Working with MS Access

MS Access

- Powerful database program
 - Can create tables queries, reports
 - Not included in all versions of MS Office
 - Very Steep learning curve vs Excel
- 



Difference Between Excel and Access

- Database Management Software that can store information
 - Relationships can be built on the data
 - Can be easy once learned to find and sort data being looked for
- 

MS Access

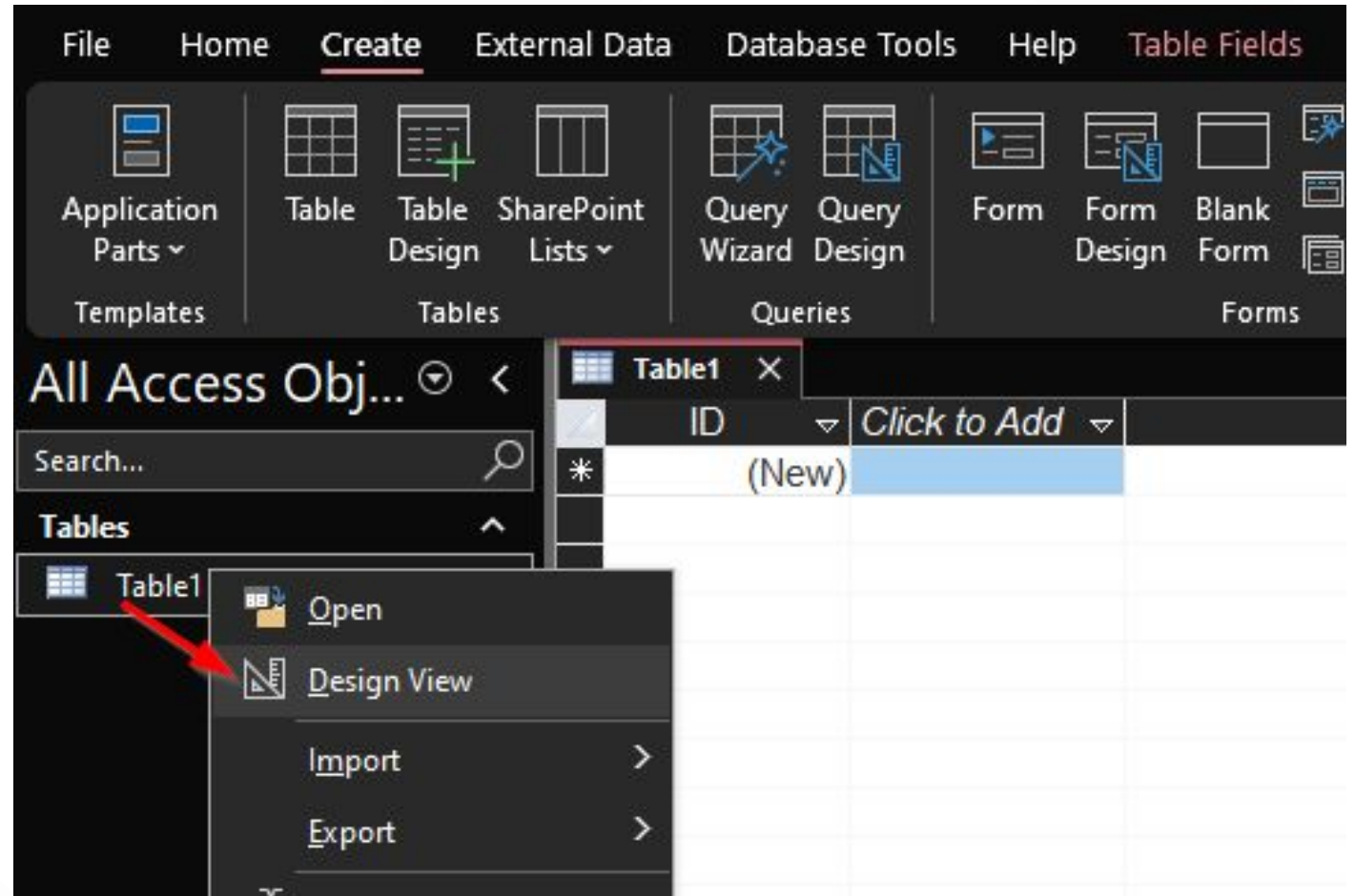
- Download Excel Sample Spreadsheet

<https://bit.ly/AccessDM>

- Open MS Access
- Create a blank database

Creating Tables

- Tables are used to store data, Queries are built from tables
- When creating a blank database, a simple table will be added
- Right-Click on Table1 and click Design view
- Save Table First



Creating Tables

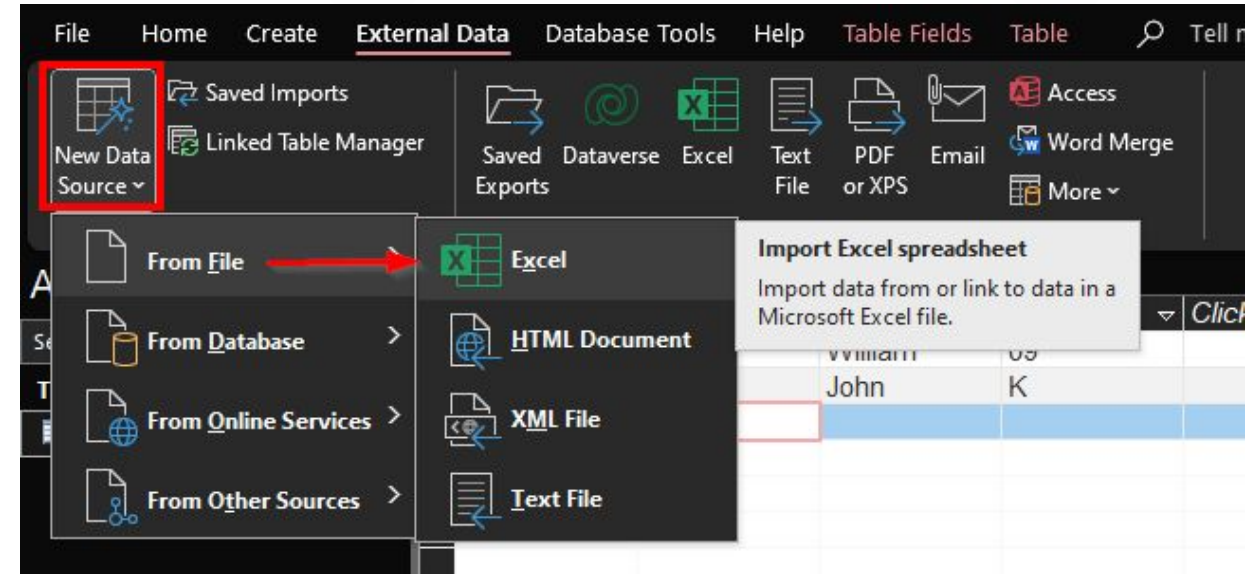
- Add Fields
- Set Data Type
- Use General Settings to format data entry
- Save Table
- Return to Datasheet View

Field Name	Data Type
ID	AutoNumber
Last Name	Short Text
First Name	Short Text
Grade	Short Text

General	
Field Size	2
Format	
Input Mask	
Caption	
Default Value	
Validation Rule	
Validation Text	
Required	Yes
Allow Zero Length	Yes


Importing Tables

- It is highly unlikely tables will be manually entered
- Access has a table import feature
- Click External Data, then New Data Source
- From File, then Excel






Importing Tables

- Import data for each workbook worksheet
 - First Row Contains headers
 - Use Default Field Options
 - Always choose primary key as a unique value (ID)
- 



Create Query

- Click Create, then Query Design
 - Add Tables you wish to pull data together from
 - Double Click the data fields to add to the query
 - Run and Save the Query
- 

Joins

- Joins allow you to determine which records to return
- Only those where the data match in one tables
- All data from both tables

Join Properties

Left Table Name: StudentList

Right Table Name: ProgramList

Left Column Name: suniq

Right Column Name: suniq

☒ 1: Only include rows where the joined fields from both tables are equal.

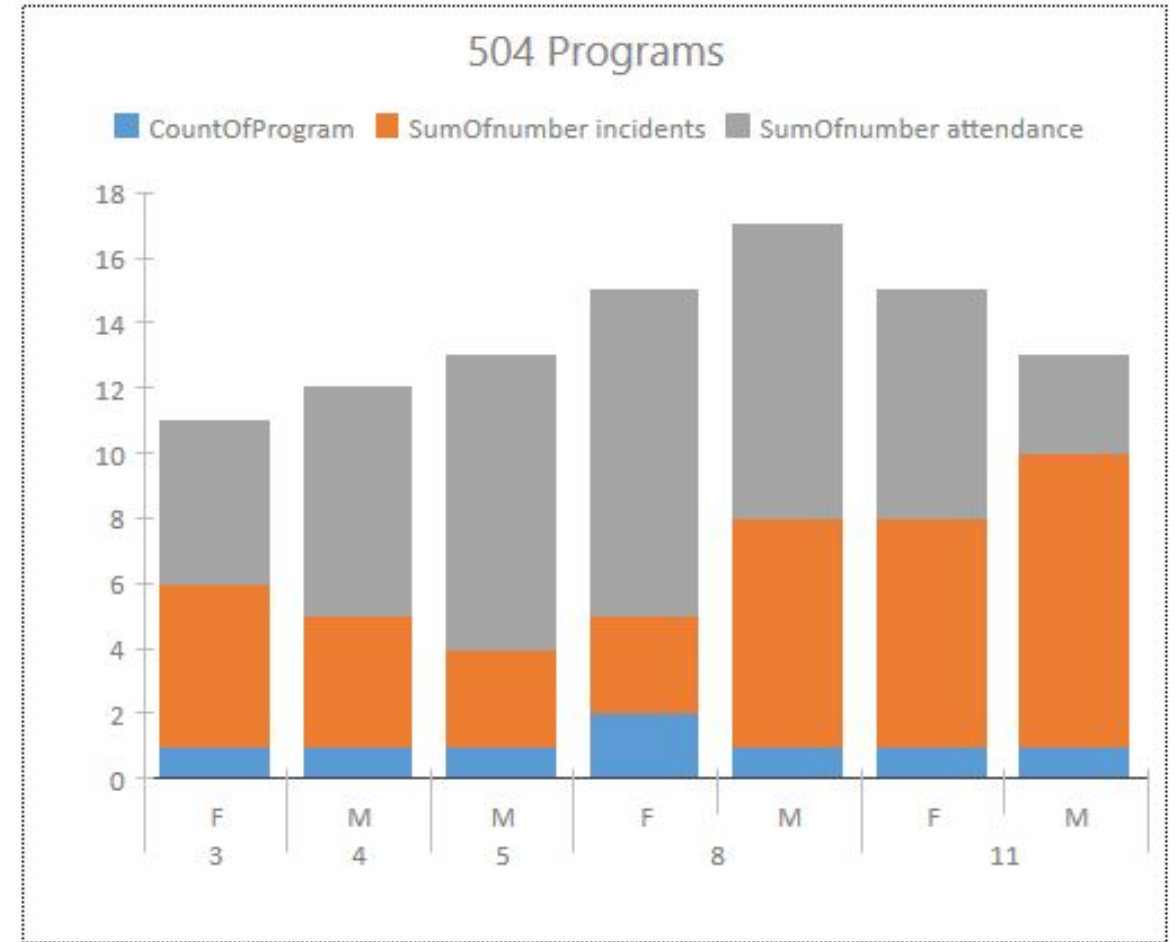
☐ 2: Include ALL records from 'StudentList' and only those records from 'ProgramList' where the joined fields are equal.

☐ 3: Include ALL records from 'ProgramList' and only those records from 'StudentList' where the joined fields are equal.

OK Cancel New

Create Charts

- Click Create Form Design
- Select Insert Chart; drag and draw chart on form
- From Chart Settings, Select the query or table to chart from
- Set Chart X and Y Axis values



MS Excel



<https://bit.ly/MSBO24-Excel>



Excel Resources

<https://bit.ly/MSBO24-Excel>

- [Auto Fill](#)
- [Flash Fill](#)
- [Comments](#)
- [Concatenate](#)
- [Conditional Formatting](#)
- [CountIf Formula](#)
- [Drop Down Lists](#)
- [Filters](#)
- [Advance Filter](#)
- [Find and Replace](#)
- [Pivot Table](#)
- [Protect Sheet](#)
- [Remove Duplicates](#)
- [Subtotal Function with Filters](#)
- [Subtotal Feature](#)
- [Split Text to Columns](#)
- [Split Screens](#)
- [Vertical Lookup](#)
- [XLookup](#)
- [If this, then this Formula And Formula & Nesting Formulas](#)
- [Switch](#)



Working with MS Excel

- Spreadsheet program
- Easy to use
- Pivot tables
- As powerful as Access, easier to use

<https://bit.ly/MSBO24-Excel>



Vlookup & Xlookup

Vlookup is a search for a value in the first column of a table range and return a corresponding value from another column in the same row. It stands for "Vertical Lookup" and is commonly used for data retrieval and analysis tasks.

- [Vertical Lookup](#)
- [XLookup](#)
- [MS XLookup info](#)

Concatenate

Combines text from different cells and free form text into one cell.

- [Concatenate](#)

Flash Fill

Flash Fill automatically fills your data when it senses a pattern. For example, you can use Flash Fill to separate first and last names from a single column, or combine first and last names from two different columns.

- [Flash Fill](#)

Subtotal Formula

Returns a subtotal in a list or database.

Function_num (includes hidden rows)	Function_num (ignores hidden rows)	Function
1	101	AVERAGE
2	102	COUNT
3	103	COUNTA
4	104	MAX
5	105	MIN
6	106	PRODUCT
7	107	STDEV
8	108	STDEVP
9	109	SUM
10	110	VAR
11	111	VARP

- VIDEO: [Subtotal Function with Filters](#)
- [MS Help](#)



Subtotal Feature

Sort

Save

Subtotal (Find under DATA, Outline)

- [Subtotal Feature](#)
- 



Pivot Table

You can use a PivotTable to summarize, analyze, explore, and present summary data. PivotCharts complement PivotTables by adding visualizations to the summary data in a PivotTable, and allow you to easily see comparisons, patterns, and trends.

- [Pivot Table](#)
- 



Final Thoughts

190 – Data Mining – Review the educational data tools of the trade. Learn about how and when to use Excel, Access, XML to submit, retrieve and use the educational data available from your Student Information System, MI School Data, other data warehouses and directly from CEPI.

Janell Craig, Senior Business Analyst - Oakland Schools

janell.craig@oakland.k12.mi.us

Grace Damerow, Interim Director of Applications - Oakland Schools

grace.rutkowski-damerow@oakland.k12.mi.us

