



Database Design

WV K-12 Education Problem

Topics

- Determine if Excel or Access are best suited for a dataset
- Design a relational database including tables and fields
- Identify appropriate table relationships

Background Information

This project includes information on West Virginia K-12 schools from 2012 to 2020.

Instructions

IMPORTANT: Complete the steps below in the order they are given. Completing the steps out of order may complicate the assignment or result in an incorrect result.

NOTE: This project will be done entirely in Microsoft Excel. This project involves designing a database, not creating one. Database creation is covered in a separate project.

1. Download and extract the provided Data Files ZIP file. It contains the following files for use in this assignment:



Database Design

WV K-12 Education Problem

- a. **education_ppdd_wvkep.xlsx** – Excel spreadsheet with information on West Virginia K-12 schools from 2012 to 2020 [1], [2].

Column Name	Type	Description
School Year	Text	School year for the data.
School ID	Text	State-assigned identifier for the school.
School Name	Text	Name of the school.
County	Text	County where school is located.
Managing District	Text	Name of district that managed the school for the school year.
Grades Served	Text	Grades served by the school for the school year.
Region	Text	Region where the district is located.
Population	Number	Population of the district.
Is Elementary School	Yes/No	If yes, the school is an elementary school for the school year.
Is Intermediate School	Yes/No	If yes, the school is an intermediate school for the school year.
Is Middle School	Yes/No	If yes, the school is a middle school for the school year.
Is High School	Yes/No	If yes, the school is a high school for the school year.
Is Preschool	Yes/No	If yes, the school is a preschool for the school year.
Is Alternative School	Yes/No	If yes, the school is an alternative school for the school year.
Dropout Rate	Percentage	Percentage of students who dropped out of school in the district for the school year.
Enrollment	Number	Number of students enrolled in the school for the school year.
Attendance Rate	Percentage	Average daily attendance rate for the school for the school year.
Graduation Rate	Percentage	Percentage of students who received a regular diploma in four years.
Students Tested	Percentage	Number of students who took state proficiency tests.
Math Proficiency	Percentage	Percentage of students who tested proficient in math.
Reading Proficiency	Percentage	Percentage of students who tested proficient in reading.
Closed	Yes/No	If yes, school is closed as of 2020-2021 school year.

2. Open the **education_ppdd_wvkep.xlsx** file to view and understand the data it contains.



Database Design

WV K-12 Education Problem

Determine if Excel or Access are best suited for a dataset

3. We need to determine if this dataset is better suited for a Microsoft Excel workbook or a Microsoft Access database. Access databases are more complex, but they can offer some advantages in organizing larger datasets.
 - a. Consider the following questions:
 - i. Are there multiple categories of data, and if so, is there a relationship between the different categories? If there are, the data can be split into multiple tables in an Access database and related together. If there are not, Excel might be a simpler way to organize the data.
 - ii. Are there large amounts of redundant data? If there are, Access may be better because its relational design can reduce redundancy and the chance of errors. If there is not, Excel may be better because it is simpler.
 - iii. Is there a need to be able to analyze only specific parts of the data? If there is, Access queries can be used to select a subset of the data. While filtering can be done in Excel, it is more cumbersome and relatively difficult to reuse search filters.
 - iv. Do charts need to be created to represent the data? If charts are needed, they can only be created in Excel. When Access is used, the data first must be exported to Excel before a chart can be made.
 - v. Will what-if analysis need to be performed? Excel includes features such as trendlines, Goal Seek, and scenarios to facilitate this process. Access is better suited to working on existing data and does not include built-in tools for what-if analysis.
 - b. Look at your answers from Step 3a. In many cases, there will be a clear pattern showing if Excel or Access should be used. If there is not, you may need to store your data in an Access database and export portions of it to Excel for detailed analysis.

For this particular dataset, Access is best suited to store the information. We have multiple categories of data, much of it redundant, we want to be able to analyze specific parts of it, and we don't need to create charts or perform what-if analysis.



Database Design

WV K-12 Education Problem

Identify field groupings

4. Review the types of information (fields/columns) that need to be stored in the database. Create groups based on related types of information.

In this dataset, the information broadly belongs to two categories: school information and district information.

Column Name	Field Grouping
School Year	Schools
School ID	Schools
School Name	Schools
County	Schools
Managing District	Schools
Grades Served	Schools
Region	Districts
Population	Districts
Is Elementary School	Schools
Is Intermediate School	Schools
Is Middle School	Schools
Is High School	Schools
Is Preschool	Schools
Is Alternative School	Schools
Dropout Rate	Districts
Enrollment	Schools
Attendance Rate	Schools
Graduation Rate	Schools
Students Tested	Schools
Math Proficiency	Schools
Reading Proficiency	Schools
Closed	Schools



Database Design

WV K-12 Education Problem

5. Look at the fields and groupings you have identified. See if they can be broken down further.

We can create subgroups for annual statistics for schools and another for annual statistics for districts.

Fields with changed groupings are highlighted in yellow below:

Column Name	Field Grouping
School Year	Schools – Annual Statistics
School ID	Schools
School Name	Schools
County	Schools
Managing District	Schools – Annual Statistics
Grades Served	Schools – Annual Statistics
Region	Districts
Population	Districts – Annual Statistics
Is Elementary School	Schools – Annual Statistics
Is Intermediate School	Schools – Annual Statistics
Is Middle School	Schools – Annual Statistics
Is High School	Schools – Annual Statistics
Is Preschool	Schools – Annual Statistics
Is Alternative School	Schools – Annual Statistics
Dropout Rate	Districts – Annual Statistics
Enrollment	Schools – Annual Statistics
Attendance Rate	Schools – Annual Statistics
Graduation Rate	Schools – Annual Statistics
Students Tested	Schools – Annual Statistics
Math Proficiency	Schools – Annual Statistics
Reading Proficiency	Schools – Annual Statistics
Closed	Schools



Database Design

WV K-12 Education Problem

Reorganize the fields

6. Review the field list. Are there fields that contain multiple distinct pieces of information that can be split into multiple fields?

There are no good candidates for being split into multiple fields in this dataset.

Divide fields into tables / Determine appropriate field types

7. Divide the field groupings into a series of tables. Identify appropriate data types and names for each field.

Table: <i>Districts</i>		
Field Name	Type	Description
Region	Short Text	Region where the school district is located.

Table: <i>DistrictStatistics</i>		
Field Name	Type	Description
Population	Number – Integer	Population of the district.
DropoutRate	Number – Percentage	Percentage of students who dropped out of school in the district for the school year.

Table: <i>Schools</i>		
Field Name	Type	Description
SchoolID	Short Text	State-assigned identifier for the school.
SchoolName	Short Text	Name of the school.
County	Short Text	County where the school is located.
Closed	Yes/No	If yes, school is closed as of 2020-2021 school year.



Database Design

WV K-12 Education Problem

Table: SchoolStatistics		
Field Name	Type	Description
SchoolYear	Short Text	School year for the data.
ManagingDistrict	Short Text	Name of district that managed the school for the school year.
GradesServed	Short Text	Grades served by the school for the school year.
IsElementarySchool	Yes/No	If yes, the school is an elementary school for the school year.
IsIntermediateSchool	Yes/No	If yes, the school is an intermediate school for the school year.
IsMiddleSchool	Yes/No	If yes, the school is a middle school for the school year.
IsHighSchool	Yes/No	If yes, the school is a high school for the school year.
IsPreschool	Yes/No	If yes, the school is a preschool for the school year.
IsAlternativeSchool	Yes/No	If yes, the school is an alternative school for the school year.
Enrollment	Number – Integer	Number of students enrolled in the school for the school year.
AttendanceRate	Number – Percentage	Average daily attendance rate for the school for the school year.
GraduationRate	Number – Percentage	Percentage of students who received a regular diploma in four years.
StudentsTested	Number – Percentage	Number of students who took state proficiency tests.
MathProficiency	Number – Percentage	Percentage of students who tested proficient in math.
ReadingProficiency	Number – Percentage	Percentage of students who tested proficient in reading.

Specify keys and junction tables

8. After splitting the fields into tables, we must add key fields and junction tables so that we can tie together information in different tables. Otherwise, for example, there would be no way for us to know which schools go with which enrollments.

For each table, we also must designate a primary key (single field) or composite key (multiple fields taken together) that can be used to uniquely identify each record. In some cases, it may be preferable to create a new AutoNumber-type field to serve as a primary key instead of using a composite key.

Newly added fields and tables are highlighted in yellow below.



Database Design

WV K-12 Education Problem

- a. In Excel, create 4 copies of the original *Education* sheet. Name each copied sheet after a table from below. Edit each sheet so it only contains the columns (fields) that are appropriate for that table.

Table: <i>Districts</i>		
Field Name	Type	Description
District	Short Text	Primary key. Name of the district.
Region	Short Text	Region where the district is located.

Table: <i>DistrictStatistics</i>		
Field Name	Type	Description
District	Short Text	Part of composite key. Name of the district.
SchoolYear	Short Text	Part of composite key. School year for the data.
Population	Number – Integer	Population of the district.
DropoutRate	Number – Percentage	Percentage of students who dropped out of school in the district for the school year.

Table: <i>Schools</i>		
Field Name	Type	Description
SchoolID	Short Text	Primary key. State-assigned identifier for the school.
SchoolName	Short Text	Name of the school.
County	Short Text	County where the school is located.
Closed	Yes/No	If yes, school is closed as of 2020-2021 school year.



Database Design

WV K-12 Education Problem

Table: SchoolStatistics		
Field Name	Type	Description
SchoolID	Short Text	Part of composite key. State-assigned identifier for the school.
SchoolYear	Short Text	Part of composite key. School year for the data.
ManagingDistrict	Short Text	Name of district that managed the school for the school year.
GradesServed	Short Text	Grades served by the school for the school year.
IsElementarySchool	Yes/No	If yes, the school is an elementary school for the school year.
IsIntermediateSchool	Yes/No	If yes, the school is an intermediate school for the school year.
IsMiddleSchool	Yes/No	If yes, the school is a middle school for the school year.
IsHighSchool	Yes/No	If yes, the school is a high school for the school year.
IsPreschool	Yes/No	If yes, the school is a preschool for the school year.
IsAlternativeSchool	Yes/No	If yes, the school is an alternative school for the school year.
Enrollment	Number – Integer	Number of students enrolled in the school for the school year.
AttendanceRate	Number – Percentage	Average daily attendance rate for the school for the school year.
GraduationRate	Number – Percentage	Percentage of students who received a regular diploma in four years.
StudentsTested	Number – Percentage	Number of students who took state proficiency tests.
MathProficiency	Number – Percentage	Percentage of students who tested proficient in math.
ReadingProficiency	Number – Percentage	Percentage of students who tested proficient in reading.



Database Design

WV K-12 Education Problem

Address redundant or inflexible fields

9. Look for cases where multiple fields store similar information. See if the tables can be redesigned to provide more flexibility.

Currently, the *SchoolStatistics* table has 4 fields containing different school types. The result is that these fields are redundant and confusing, and adding a new school type can cause inconvenience. By reorganizing the table, we can provide more flexibility. While, in some cases, the schools can work for more than elementary school, middle school, or high school, it is possible for them to be of multiple school types. For this situation, we must also include different combinations in school types.

Changes to the table are highlighted in yellow below.

- a. In Excel, edit the data on the *Schools* sheet to that it matches the format shown below.

Table: <i>SchoolStatistics</i>		
Field Name	Type	Description
SchoolID	Short Text	Part of composite key. School year for the data.
SchoolYear	Short Text	Part of composite key. Name of district that managed the school for the school year.
ManagingDistrict	Short Text	Name of district that managed the school for the school year.
GradesServed	Short Text	Grades served by the school for the school year.
SchoolType	Short Text	Type of school.
Enrollment	Number – Integer	Number of students enrolled in the school for the school year.
AttendanceRate	Number – Percentage	Average daily attendance rate for the school for the school year.
GraduationRate	Number – Percentage	Percentage of students who received a regular diploma in four years.
StudentsTested	Number – Percentage	Number of students who took state proficiency tests.
MathProficiency	Number – Percentage	Percentage of students who tested proficient in math.
ReadingProficiency	Number – Percentage	Percentage of students who tested proficient in reading.



Database Design

WV K-12 Education Problem

10. Now that we have consolidated our fields, we want to look for cases where we have complicated composite keys. In situations where you must use the combination of multiple fields to uniquely identify a record, it can sometimes be easier to have a single AutoNumber-type field serve as a primary key instead. For example, if we wanted to uniquely identify a person, would it be easier to use their Social Security number or the combination of their name, date and time of birth, location, parents, gender, weight, and height to be sure we are looking at the correct person?
- a. In the *SchoolStatistics* tables, there are two fields that collectively serve as a composite key to uniquely identify school statistics. It will be easier to instead have a single AutoNumber-type **ID** field serve as primary key.

In Excel, edit the data on the *SchoolStatistics* sheet by inserting a new **ID** column and sequentially numbering each record (e.g., 1, 2, 3).

Table: <i>SchoolStatistics</i>		
Field Name	Type	Description
ID	AutoNumber	Primary key. Unique identifier for the school statistics.
SchoolID	Short Text	School year for the data.
SchoolYear	Short Text	Name of district that managed the school for the school year.
ManagingDistrict	Short Text	Name of district that managed the school for the school year.
GradesServed	Short Text	Grades served by the school for the school year.
SchoolType	Short Text	Type of school.
Enrollment	Number – Integer	Number of students enrolled in the school for the school year.
AttendanceRate	Number – Percentage	Average daily attendance rate for the school for the school year.
GraduationRate	Number – Percentage	Percentage of students who received a regular diploma in four years.
StudentsTested	Number – Percentage	Number of students who took state proficiency tests.
MathProficiency	Number – Percentage	Percentage of students who tested proficient in math.
ReadingProficiency	Number – Percentage	Percentage of students who tested proficient in reading.

11. Identify cases where similar information is repeatedly entered. Fields where users repeatedly enter large amounts of text can be prone to errors such as misspellings.

In this example, the **SchoolType** field in the *SchoolStatistics* table is a likely source of errors since school types are spelled out. A better solution is to store the school type as a 3-letter abbreviation in the *Schools* table, and then create



Database Design

WV K-12 Education Problem

a *SchoolTypes* table to store school type abbreviations and their full names.

Changed fields are highlighted in yellow below. The *SchoolTypes* table is entirely new.

- a. In Excel, create a new sheet to represent the *SchoolTypes* table.

Table: <i>SchoolTypes</i>		
Field Name	Type	Description
SchoolTypeAbbrv	Short Text	Primary key. Abbreviation of the school type.
SchoolTypeName	Short Text	Name of the school type.

- b. Enter records for the school types below.

HINT: The *SchoolTypes* table will contain 11 records.

SchoolTypeAbbrv	SchoolTypeName
ELS	Elementary School
INS	Intermediate School
MDS	Middle School
HIS	High School
EMS	Elementary/Middle School
MHS	Middle/High School
EMH	Elementary/Middle/High School
OTH	Other
PRE	Preschool
ALS	Alternative School
TEC	Technical Center

- c. Update the *SchoolStatistics* table as shown below:

Table: <i>SchoolStatistics</i>		
Field Name	Type	Description
ID	AutoNumber	Primary key. Unique identifier for the school statistics.
SchoolID	Short Text	School year for the data.
SchoolYear	Short Text	Name of district that managed the school for the school year.
ManagingDistrict	Short Text	Name of district that managed the school for the school year.



Database Design

WV K-12 Education Problem

Table: <i>SchoolStatistics</i>		
Field Name	Type	Description
GradesServed	Short Text	Grades served by the school for the school year.
SchoolTypeAbbrv	Short Text	Abbreviation of the school type.
Enrollment	Number – Integer	Number of students enrolled in the school for the school year.
AttendanceRate	Number – Percentage	Average daily attendance rate for the school for the school year.
GraduationRate	Number – Percentage	Percentage of students who received a regular diploma in four years.
StudentsTested	Number – Percentage	Number of students who took state proficiency tests.
MathProficiency	Number – Percentage	Percentage of students who tested proficient in math.
ReadingProficiency	Number – Percentage	Percentage of students who tested proficient in reading.

Remove duplicate records

12. After your data has been split into multiple tables, look at your tables to see if there are any duplicate records such as multiple records in the *Districts* table for the same district. These should be eliminated.
 - a. In Excel, delete any duplicate records that are identical to another record.

Review final table structure and cleanup old data

13. Now that we have finished splitting the data into multiple tables and reorganizing the data to avoid duplication, take a moment to review the final table structure shown below.
 - a. In Excel, delete the original *Education* data sheet as it is no longer needed.

Table: <i>Districts</i>		
Field Name	Type	Description
District	Short Text	Primary key. Name of the district.
Region	Short Text	Region where the district is located.

Table: <i>DistrictStatistics</i>		
Field Name	Type	Description
District	Short Text	Part of composite key. Name of the district.
SchoolYear	Short Text	Part of composite key. School year for the data.
Population	Number	Population of the district.
DropoutRate	Percentage	Percentage of students who dropped out of school in the district.



Database Design

WV K-12 Education Problem

Table: <i>Schools</i>		
Field Name	Type	Description
SchoolID	Short Text	State-assigned identifier for the school.
SchoolName	Short Text	Name of the school.
County	Short Text	County where the school is located.
Closed	Yes/No	If yes, school is closed as of 2020-2021 school year.

Table: <i>SchoolStatistics</i>		
Field Name	Type	Description
ID	AutoNumber	Primary key. Unique identifier for the school statistics.
SchoolID	Short Text	State-assigned identifier for the school.
SchoolYear	Short Text	School year for the data.
ManagingDistrict	Short Text	Name of district that managed the school.
GradesServed	Short Text	Grades served by the school.
SchoolTypeAbbrv	Short Text	Abbreviation of the school type.
Enrollment	Number	Number of students enrolled in the school.
AttendanceRate	Percentage	Average daily attendance rate for the school.
GraduationRate	Percentage	For high schools, percentage of students who received a regular diploma in four years.
StudentsTested	Number	Number of students who took state proficiency tests.
MathProficiency	Percentage	Percentage of students who tested proficient in math.
ReadingProficiency	Percentage	Percentage of students who tested proficient in reading.

Table: <i>SchoolTypes</i>		
Field Name	Type	Description
SchoolTypeAbbrv	Short Text	Primary key. Abbreviation of the school type.
SchoolTypeName	Short Text	Name of the school type.



Database Design

WV K-12 Education Problem

Determine relationships

14. Now that the tables have been defined, we must determine the exact relationships that will tie the tables together. Relationships between two tables are based on (one or more) common fields that appear in both tables. All of the tables must be interconnected. By following one or more relationships, it should be possible to tie data from one table together with data in any other table.

Table 1	Table 2	Common Fields for Relationship
<i>Districts</i>	<i>DistrictStatistics</i>	District
<i>Schools</i>	<i>SchoolStatistics</i>	SchoolID
<i>Counties</i>	<i>Dropouts</i>	County
<i>DistrictStatistics</i>	<i>SchoolStatistics</i>	District and SchoolYear
<i>SchoolTypes</i>	<i>Schools</i>	SchoolTypeAbbrv

Grading Rubric

This assignment is worth 8 points. It will be graded by your instructor using this rubric:

Standard	Meets Requirements (8 points)	Does Not Meet Requirements (0 points)
Student made reasonable effort in correctly completing assignment.	Assignment is at least 70% complete and correct, or student contacted instructor for help on incorrect or incomplete items.	Assignment is less than 70% complete and correct, and student did not contact instructor for assistance on incorrect or incomplete items.

This rubric will be used for peer evaluation of this assignment:

Standard	Excellent	Satisfactory	Needs Improvement
Assignment is correct and complete.	Assignment is at least 90% complete and correct.	Assignment is 70%-89% complete and correct.	Assignment is less than 70% complete and correct.

References

- [1] "ZoomWV Data Dashboard," *West Virginia Department of Education*. Available: <https://zoomwv.k12.wv.us/>.
- [2] "County Population Totals: 2010-2019," U.S. Census Bureau, Washington, DC, Jun. 2020. Available: <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html>.