Data analysis project

This project is comprised of four separate parts. The first three parts will analyze aggregated data obtained from <http://www.healthdata.gov> and <https://data.medicare.gov/>. This data has been downloaded and when necessary incorporated into basic databases amenable to Microsoft Access 2010. The last (fourth) part will focus on generating an (E-R diagram) for a model database to use in a specific healthcare setting.

For the first two parts, insert your answers within this document, then upload the entire document to the appropriate dropbox by the deadline. For the last two parts, turn in your hospital report as well as your E-R diagram into the same dropbox by the deadline. Please review the appropriate Lesson 4 Powerpoint presentation and the Practice Section below to ensure you have the necessary background to evaluate and utilize appropriate Database structure and Structured Query Language (SQL). Please note that when starting out many of you may rely upon the ‘Query Wizard’ available in Microsoft Access, which is fine if it works for you. However, you may want to inspect your code using ‘SQL View’ under ‘Query Design’ to see how your query is built. Using ‘SQL View’ and executing your own SQL code will prove easier for some of these exercises.

Part 1 (15 points)

Download HHA\_database.accb, open in Microsoft Access 2010. This file is a database derived from an Excel spreadsheet available at <http://www.healthdata.gov>. The database contains responses to the CMS survey of U.S. Home Health Agencies (both nationally and abroad) regarding their services. All U.S. HHAs are listed in the HHA\_Providers and HHA\_Services tables, while agencies represented in the HHA\_surveys table are the only ones to offer complete response to the following questions:

| **HHA\_Survey\_Questions** | |
| --- | --- |
| **ID** | **Field1** |
| 1 | How often the home health team began their patients care in a timely manner 93.7% |
| 2 | How often the home health team taught patients (or their family caregivers) about their drugs 97.90% |
| 3 | How often the home health team checked patients risk of falling 99.50% |
| 4 | How often the home health team checked patients for depression 97.90% |
| 5 | How often the home health team determined whether patients received a flu shot for the current flu season 77.60% |
| 6 | How often the home health team determined whether their patients received a pneumococcal vaccine (pneumonia shot) 80.60% |
| 7 | How often the home health team checked patients for pain |
| 8 | How often the home health team treated their patients pain 82.1% |
| 9 | How often the home health team took doctor-ordered action to prevent pressure sores (bed sores) 95.5% |
| 10 | How often the home health team included treatments to prevent pressure sores (bed sores) in the plan of care |
| 11 | How often the home health team checked patients for the risk of developing pressure sores (bed sores) |
| 12 | How often patients got better at walking or moving around 73.20% |
| 13 | How often patients got better at getting in and out of bed 71% |
| 14 | How often patients got better at bathing 75.90% |
| 15 | How often patients had less pain when moving around 76.40% |
| 16 | How often patients breathing improved 75.30% |
| 17 | How often patients got better at taking their drugs correctly by mouth 63.40% |
| 18 | How often patients receiving home health care needed urgent, unplanned care in the ER without being admitted 12.90% |
| 19 | How often home health patients had to be admitted to the hospital 15.90% |
|  |  |

1. **Which zip code(s) have more than 100 HHAs?**

SELECT ZIP, COUNT(\*)

FROM HHA\_Providers

GROUP BY ZIP

HAVING COUNT(\*) >= 100;



1. **Among the positive indicators, which question do HHAs reflect the best performance? the worst?**

**<SQL code>**

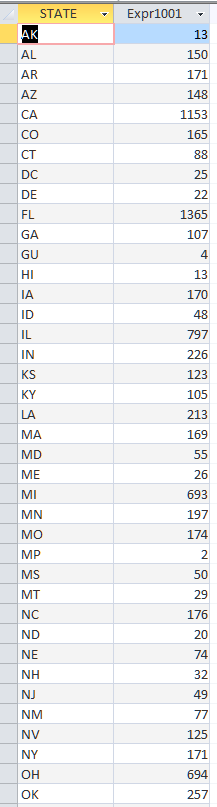
**<Results>**

1. **How many HHAs are there per state?**

SELECT STATE, COUNT(\*)

FROM HHA\_Providers

GROUP BY STATE;



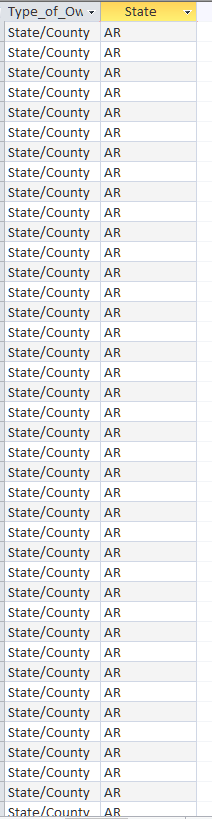
1. **What state has the highest number of ‘State/County’ HHAs?**

SELECT Type\_of\_Ownership, State

FROM HHA\_Providers

WHERE Type\_of\_Ownership = 'State/County'

ORDER BY State DESC;



\*\*I counted who had the most and it was AR.

1. **What is the average length of certification among HHAs in each state?(Hint: do not ignore # signs in DateDiff function and use 1/1/2013 as endpoint for certification)**

**<SQL code>**

**<Results>**

Part 2 (20 points)

Download and open Part\_B\_National\_Summary.accB in Microsoft Access 2010. This is a database generated from spreadsheets for 2009 and 2010 downloaded from <http://www.healthdata.gov>. The Medicare Part B national datasets are summarized by meaningful Healthcare Common Procedure Coding/Current Procedural Terminology (HCPC/CPT) code ranges. Each data set displays the allowed services, allowed charges, and payment amounts by HCPCS/CPT codes and prominent modifiers. In addition, there is a README file (PartBNationalSummaryReadmeFile2010.pdf) available in this unit that provides additional information about the datasets. Using SQL statements, answer the following questions:

1. Which Description has the most HCPCS codes?

<SQL code>

<Results>

1. What was the difference between Payment and Allowed Charges for HCPCS codes associated with ‘INTEGUMENTARY’ in 2009?

<SQL code>

<Results>

1. Name the HCPCS codes with the five (5) highest number of allowable services in 2009.

<SQL code>

<Results>

1. Determine and report total payments associated ‘cardiovascular’ procedures in 2010.

<SQL code>

<Results>

1. What are the top five (5) HCPCS as determined by Total Payment in 2010? Report the procedure corresponding to each HCPCS code to a level greater than available in the database. That is, look up the HCPCS codes to reveal precisely what they are coding for.

Taken from the CMS site:  "Each year, in the United States, health care insurers process over 5 billion claims for payment. For Medicare and other health insurance programs to ensure that these claims are processed in an orderly and consistent manner, standardized coding systems are essential. The HCPCS Level II Code Set is one of the standard code sets used for this purpose. The HCPCS is divided into two principal subsystems, referred to as level I and level II of the HCPCS. Level I of the HCPCS is comprised of CPT (Current Procedural Terminology), a numeric coding system maintained by the American Medical Association (AMA). The CPT is a uniform coding system consisting of descriptive terms and identifying codes that are used primarily to identify medical services and procedures furnished by physicians and other health care professionals."

So...the idea is for you to find out that CPT code 66984 is for cataract surgery....the HCPCS codes in this database are Level 1, CPT codes.

There are various resources available on the web, such as (but not limited to) http://www.cms.gov/apps/physician-fee-schedule/search/search-criteria.aspx OR **just Google CPT or HCPCS + your five digit code**and you will get the information.  DO NOT pay anything for this information.

<SQL code>

<Results>

Part 3 (45 points)

Parts 1 & 2 are intended to sharpen your SQL skills, while part 3 is intended for you to apply those skills by coming up with your own questions that help you address the issue of where and what type of hospital care might present opportunity (ies) for expansion in WI...also, what would be areas that are not good ideas for expansion.  Part 3 is intended to have students demonstrate application of SQL for searching a database; that said, you don't have to join 8 tables or anything, but you do need to provide evidence of your SQL queries used to retrieve data for your report.  It is reasonable to first query: how many hospitals are in WI?  What types of hospitals are in WI among those identified in the first query?  What about HCAHPS scores among these hospitals assessing patient satisfaction? etc. These are relatively straightforward queries that allow you to assess the healthcare market in WI. If it helps, consider this part of the project as sort of a business exercise.  That is, identifying gaps in current hospital care based upon the data in this database.  To identify gaps, one could ask questions about the types of services rendered...maybe there are not enough or too many childrens hospitals, maybe not enough or too many heart care facilities...etc.  It is about surveying the market in WI and proposing what type of hospital would be successful in the current WI market.  Looking at HCAHPS, you are identifying hospitals in areas that perhaps are not providing the type of customer service that are necessary to be competitive and successful...​

One can obtain the database under Lesson 4, Assignments, #5 where you will find Hospital.zip (containing a compressed version of Hospital.mdb...you need to uncompress it) and Hospital.pdf.  The Hospital.pdf file contains information for understanding the data available in Hospital.mdb.  Look at page 8 of 79 to see what each table in the database is named.  Subsequent documentation reveals the data contained in each table.  This data includes AHRQ survey results, which are a series of questions that hospitals report on regarding quality of care as well as HCAHPS survey results.  Then, you make your own SQL statements to generate recommendations regarding the WI hospital market **and write a 3-5 page report (Font: Times New Roman; 12 font; double-spaced; figures, tables, and SQL code do NOT count towards page requirement) of hospitals in Wisconsin based upon these analyses.  Include evidence of your SQL queries in your report.** Your grade for this exercise will be determined both by the nature of the questions you utilize to assess current measures of hospitals in Wisconsin, as well as the data you retrieve and summarize. Submit your report to the Data Analysis Project DropBox on the course web-site.

Part 4 (20 points)

Lastly, you have subsequently been hired as a consultant for a hospice start-up (Review Chapter 13 of the Peden textbook for hospice care data requirements). Construct an E-R Diagram for your database model company based upon your understanding of data requirements for hospice care. Assistance with developing an E-R diagram is available via a link on the D2L web-site (and [here](https://www.youtube.com/watch?v=-fQ-bRllhXc); <https://www.youtube.com/watch?v=-fQ-bRllhXc>). Also, Pratt and Last, Chapter 1 Figures 1-7, 1-20, and 1-27 offer E-R diagrams for databases used throughout their textbook and are reflective of the expectations for this part of the project. Please note that you do NOT need Microsoft Access or any other software to complete this exercise; rather, you simply need to submit a diagram. Your grade for this exercise will be determined by the recognition of the data requirements for patients and providers in hospice, and by the database structural model. Upon completion, submit your work to the Data Handling Dropbox on the course web-site