ECE5984 SP22 - Prof. Jones – HW 1

Due Tuesday, Feb 8, 2022 – 11:59 PM via Canvas

In this assignment you will perform some exploratory data analysis on a data file related to COVID-19 hospitalizations. You will use both Excel and Tableau to explore and understand the data, and to derive and show some data-driven insights. Here are the steps that you should follow.

1. Visit the following page at healthdata.gov: https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/uqq2-txqb . This data set is a current (updated weekly) record of hospitalizations and hospital impacts attributed to COVID-19. Download the tab-separated value (tsv) data file. To do this, choose the "Export" option and select the "TSV for Excel" file. The file will download; it's about 160 Mbytes.
2. Read the data file into Excel and save as an Excel workbook. To do this, open Excel, and browse and open the downloaded .tsv file. It will prompt you to choose how to import the file; it's delimited using tab characters. Once you open the file, save it as an Excel workbook (.xlsx extension).
3. Sometimes (as in this file), impossibly large values are used to indicate missing values (where no real data is available). For analysis, we would rather have missing values in the data. So, replace any impossibly large (positive or negative) values with empty cells. Make note in your submission of which values you replace, and how many replacements are made.
4. Calculate and display the (univariate only!) descriptive statistics for all of the columns in the data set. Do this in Excel, on a separate sheet of the workbook; you should calculate the following statistics (here is a sample):

Statistics	hospital_pk	collection_week	state	ccn
Mean	267614.4304	44309.64634	#DIV/0!	267571
Min	10001	44043	0	10001
Max	677297	44575	0	677297
Range	667296	532	0	667296
Median	260009	44309	#NUM!	260009
Mode	10108	44190	#N/A	50515
Variance	24511724326	24105.81784	#DIV/0!	2E+10
Std Deviation	156562.2059	155.2604838	#DIV/0!	156577
Quartile 1	140172	44176	#NUM!	140167
Quartile 2	260009	44309	#NUM!	260009
Quartile 3	390265	44442	#NUM!	390265

   See the hints below for a couple of tips. Copy the first six and the last six columns of your statistics and paste into your submission.
5. Save your Excel spreadsheet, and open Tableau. Use your (modified) Excel spreadsheet as the data source (the sheet with all of the data on it, not the summary statistics).
6. In Tableau, you are to answer the following questions; for each, paste the appropriate chart or table into your submission. For each chart or table, be sure to include any color legends or other labels. Where questions are posed below, include a chart or graph to support your answer.
   a. There are several hospital subtypes in this data. Prepare a table showing the average percentage used of inpatient hospital beds for each, and the average percent overall.
   b. Display a histogram of "Previous Day Admission Adult Covid Confirmed 7 Day Sum"; use bin size of 5.
   c. Display a line graph of "Icu Patients Confirmed Influenza 7 Day Avg" by week, for the period from April 2020 through December 2021.

d. Identify the five hospitals with the highest “Total Pediatric Patients Hospitalized Confirmed Covid 7 Day Avg”.
e. Identify the five states with the highest “Total Pediatric Patients Hospitalized Confirmed Covid 7 Day Avg”.
f. Create a calculated field called “InpatientBedsPctUsed” as the ratio of “Inpatient Beds Used 7 Day Avg” to “Inpatient Beds 7 Day Avg”. Display a map of this ratio by state for the 50 US states and Puerto Rico.
g. Which hospital has the most inpatient beds?
h. Which state has the highest number of influenza hospitalizations for the past seven days?

For your submission, paste all of the required information into a single Word (or pdf) file. Submit your file, along with your Tableau workbook, using Canvas. Do NOT submit your massive Excel workbook.

HINTS:

- In Excel, to calculate the mean of column A1 on the sheet called “data”, where the dataset has 400000 rows, use the following formula:
  `=MEAN(data!A$2:A$400000)`
- For this size of a dataset, it doesn’t matter much whether you calculate variance for a sample or the population – but use VAR.S() to be consistent with mine.
- In Tableau, you can generate maps by dragging longitude and latitude to the proper axis (column and row).
- Don’t hesitate to use “Show Me” in Tableau to see which graph formats are available.
- Sometimes, to get the format you need, a Measure needs to be converted to a Dimension, and vice versa.