Savvy Coders DAP Cohort Capstone Guidelines

Business Intelligence Capstone

A capstone might be called a culmination project, senior thesis, or a final exhibition. ... The capstone exercise is intended to apply all of the knowledge and skills you've gained over the course in one assignment. This capstone exercise will focus on the following:

- Analysis of raw data
- Visualization of the raw data
- Driving insights out of pre-existing dashboards

The capstone will not be a paper or thesis, but a live demonstration of the knowledge gained using such tools as python, Excel and Tableau. The student will be able to demonstrate their level of expertise in a manner that is befitting a budding Data Analyst/Scientist.

Analysis of Data

The student will pick and choose from many different publicly available datasets; collect and wrangle that data; and have the option to illustrate a particularly step in the data analytics process by generating raw data in a programmatic fashion.

The six steps of Data Wrangling entail: -

- Discovery
- Structuring
- Cleaning
- Enriching
- Validating
- Publishing

Programmatic Fashion

Using python, generate up to 100,000 rows of data at various dimensionality in order to satisfy several use cases (both positive and negative) for inclusion in the complete capstone exercise.

The student should be able to walk through the code, providing a straightforward explanation of the various use cases, and specifically point out initial inputs and expected results.

Be creative with this as your audience will not want to look at a record level, tabular file (hint – a bit of programmatic visualization never hurt anybody).

Internet Available

There are many websites that supply various datasets, that have been used by students and professionals in the field, to develop techniques for analysis of raw data.

The student will need to select a prepared dataset, give a short statement about the contents of the dataset and define several use cases (positive and negative) that will make use of the information to generate actionable insights.

Here are a few resources to find datasets:

https://community.tableau.com/s/question/0D54T00000G557ESAR/data-sets

https://www.kaggle.com/datasets

http://statweb.stanford.edu/~sabatti/data.html

https://opendatainception.io/

https://www.stats.govt.nz/large-datasets/csv-files-for-download/

Again, the student should be able to walk through the code, providing a straightforward explanation of the various use cases, and specifically point out initial inputs and expected results. Be creative with this as your audience will not want to look at a record level, tabular file (hint – a bit of programmatic visualization never hurt anybody).

Organization of Data

After the student has selected a dataset that is of interest to them, the next step in this exercise is to become truly familiar with the details that are used to describe the data in a more statistical manner.

Descriptive statistics can be useful for two purposes:

1. to provide basic information about variables in a dataset and
2. to highlight potential relationships between variables.

The three most common descriptive statistics can be displayed graphically or pictorially. It is recommended that the student use the more common descriptive statistics to describe or summarize the characteristics of a sample or data set.

Descriptive statistics consists of three basic categories of measures:

- measures of central tendency of the data set (mean, median, mode)
- measures of variability (or spread) describe the dispersion of the data set (variance, standard deviation)
- and frequency distribution describe the occurrence of data within the data set (count)

It is recommended as possible to include some other descriptive statistics to give a better understanding of the three basic stats.

Here is are hints as to how common descriptive statistics could be used on your sample dataset –

https://www.investopedia.com/terms/d/descriptive_statistics.asp

https://corporatefinanceinstitute.com/resources/knowledge/other/descriptive-statistics/.

Visualization of Data

Now that the student has a better understanding of the dataset based on descriptive statistics, it is time to generate a few charts, graphs and even reports to enlighten users on some of the insightful aspects of the data.

We will use Tableau for our data presentation -- to generate sheets or simple dashboards with three or four graphics to tell us your data story.   Remember, you are not looking for perfection in your graphics at this point in the capstone. The purpose of this exercise is to think fast and act quickly to generate a couple of interesting artifacts.

Driving Actionable Insights

This is the final step in the process of illustrating that the student has grasped the concepts associated with data analytics/business intelligence.

Tableau provides a stunning set of data visualization examples from around the world created with Tableau Public (the licensed software used for this class).

There is a "Viz of the Day" as well as many "Featured Vizs" that the student will be able to choose from for the exercise.

Discovering actionable insights from your data, involves a methodology.   Use the concepts of the Data Analytics methodology to peruse the Tableau Gallery and find one or two "Vizs" that allow you to demonstrate your prowess.   Remember to keep it simple and be sure it is engaging!

https://public.tableau.com/en-us/gallery/?tab=viz-of-the-day&type=viz-of-the-day

Presentation of the Data Story

The final step in the capstone exercise will be to walk the talk and talk the walk.

You worked very smart, and hard, on putting together a first-class story from scratch and now it is time to "broadcast it. Data and Data Analysis is only useful when insightful findings are published and made available for decision makers to use. Use your final project to tell everybody your Data Story!

This might seem a bit daunting, but I assure you that these are the typical types of activities that are performed on a daily basis in the Data Analytics/Business Intelligence field.