**Spring 2020 - ISM 6419: EXAM 2 Instructions**

* Download the 4 Datasets from CANVAS: DimCity.csv, DimState.csv, DimYear.csv, and FactCrime.csv
* Based on the Entity Relationship Diagram below, add all 4 of these datasets to PowerBI and establish their appropriate relationships.



* Create a “Page” in PowerBI and build a Visualization that tells a story from the data. Your “Page” should include at least 4 or more visualizations. Include at least 1 filter with one of your visualizations. Make sure you add appropriate descriptive text to describe your data visualization(s).
* Create a second “Page” in PowerBI. This time, create at least 2 Python scripts (similar to WEEK 4 Homework) using the Python Script editor that show visualizations of your choice using **matplotlib.pyplot**
* Save your PowerBI workbook.
* Complete all questions on the next page. Exchange your answers along with your PowerBI workbook with your group partner. Once you receive the answers and PowerBI workbook from your partner, use the Exam 2 Rubric sheet to evaluate your partner’s exam. Then, upload your PowerBI workbook, this document that includes your answers, and the Exam 2 Rubric document you filled out with your group partners evaluation to Exam 2 in CANVAS. In summary, you will upload 3 separate documents in Exam 2 within CANVAS.

**QUESTION 1:** What is the Story of your Visualization? Please be detailed for full marks.

My story tries to explain the variation in number of crimes and the population at a state level. It shows how different states with different populations have different numbers and types of crimes that are being committed. I have included a slicer for looking at the data on a city level. I have also provided a slicer for easy filtering of Chinese new years. The first visual is an ‘area’ graph which talks about the relationship between the total population and total crimes across the different years. I created a new column called *crimes* in the csv provided for finding the total number of crimes at each ‘city id+ year id’ level. The second graph is a ‘stacked column chart’ that tells the total crimes by state. The third visual is a ‘stacked bar chart’ that depicts the total population at a state level. This helps you to compare what the total population of a state is and the total number of crimes as per the population. The fourth visual is a ‘stacked bar chart’ and gives a break-down of the types of crime at a state level. It helps to identify the top crimes in a state. The user can easily relate these visuals since they are of the same type. The user can select any state in a visual and all other visuals are linked to filter on the specific state. Now you can study that state even closely. Once a user selects a specific city, the visuals change at city level. If the city with the same name exits in more than one state you can see all those states. Page 2 tries to compare the trend of the population vs crimes across all the states using a line to show crimes and bars to show population in the first visual. That helps in understanding the data at one glance. The second visual on page 2 talks about how the top 5 categories of crime have changed in the different Chinese years.

**QUESTION 2:** Who is the audience?

US citizens, Police, Other Law Enforcement departments, Government, Tourists and International Students.

**QUESTION 3:** What makes your visualization effective (ex. Numbers, Visuals)? Please be specific and detailed with this answer.

The first ‘area’ graph clearly explains how the total number of crimes are inversely proportional to the increasing years. That really captures the user’s attention to inspect for a similar trend at a state level. This is taken care of by the other visuals which explain the data at a state level. The ‘slicer’ even makes it easy for users to look at the different Chinese years separately. Also, the ‘slicer’ for specific cities can be used to understand data at a more granular level. The user can filter the data using these slicers effectively and all the filters stay visible all the time.

The visuals on page 2 are created using python. They take the story further and make it an immersive experience for the user to learn something more about the data by comparing different crimes across the different Chinese years.

**QUESTION 4:** How are you going to focus your audience's attention with your visualization? Be specific for full marks.

Page 1 talks about the number of crimes across years as compared to the total population at a state level. The user can filter at a city level using the 2 slicers provided on page 1 for each Chinese year. Page 2 then focuses of this comparison across states with a deeper look at the top crimes for different Chinese years. There is a clear connection between the two pages and similar graphs make it easier to understand data and focus audience’s attention.