***Featherman’s Analytics Adventure’s******Final project***

You have become empowered and enlightened with new data management technical skills, perspective, and much-needed experience. Your analytical skills have improved such that you can better interpret what a dataset reveals regarding business performance. You might even be developing intuition regarding business data analysis, how to answer the questions of business managers. You even have been cautioned about some rookie mistakes and passed that hurdle.

Give yourself a pat on the back, you have learned a lot about data management, data integration, data analysis and data presentation! Our semester-long adventure is coming to its end, you have come far grasshopper, but you have one last step before this segment of journey is complete. It is time for you to walk alone and demonstrate your learning.

Indeed, the next step down the yellow brick road path to prosperity is for you to practice your analytical skills using a scenario that interests you and demonstrate your level of learning and conscientiousness.

Now you are asked to choose your own research project, present your analysis. Because analysts often present their findings to audiences (a favorite interviewing technique) you need more experience and to demonstrate your competence and usefulness and a data analyst. You are advised to now start your portfolio of original projects. IF you publish links to your original projects (ie on your careerbuilder and linked in resume’s) then recruiters can verify that you indeed have the skills and experience, and used your time well in the MIS@WSU program.

You are now asked to perform research of your choice, business setting or otherwise. You need to find your own dataset and problem to analyze, this may be the hardest part of the project! You probably will not find the exact data you want, so find the data available and ask appropriate questions to better understand the data. In past semesters, student projects included analysis of economic data, building permits, Seattle immigration patterns, financial news, public safety, shark bites, rainfall, and sports metrics.

***Possible Sources of data:***  
1. [Seattle data](https://data.seattle.gov/)

2. [US Data](https://www.data.gov/)

3. [Other datasets](https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/#14785ac4b54d)

4. [Kaggle.com](https://www.kaggle.com/)

5. You can just google the dataset you are looking for such as [shark attack data](http://www.sharkattackfile.net/incidentlog.htm)

***Process***

* Consider a phenomenon that interests you.
* Find data on that phenomenon. Repeat as necessary until you find an in-depth dataset.
* Download data either into your SSMS database, or onto your PC or Cloud.
* Begin your analysis of the problem using any combination of Excel, SSMS, Report Builder, Tableau or PowerBI.
* Start at a macro general level to make sense of the data then examine the phenomenon over time if possible. This gives you a sense of the past, present, future of the phenomenon.
* Identify the dimensions in the dataset that you can use to slice up the data, and to made hierarchies for drilling down into the data. For example, such as geographic which could generate a region/state/county hierarchy.
* Make [groups](https://www.youtube.com/watch?v=yK0JBwN_9o0) in the data as possible. For example, group customers into target markets, group states into regions, and group teams into conferences. Looking at data at different levels of granularity (i.e., team - WSU vs. conference – PAC-12) will often reveal different pattern in the data. Changing the granularity helps the analyst better understand the scope/size/localization of the problem.
* Build a set of charts, tables and maps that explain the phenomenon. You do not need 30 charts, but if you are having fun go ahead. Focus on the big issues in the dataset. Learning to provide what the report consumer needs to know about the problem NOT everything you know about the problem is the goal.
* If you use maps of streets or districts, please add street-view pictures from maps.google.com or similar webservice.
* Find and integrate more data as possible.
* Use the charts to form your interpretation and understanding of the phenomenon. Add more data hierarchies or slicers to the charts to better explain the data and focus the reader’s attention. Find new perspective on your dataset such as zooming on one facet of the problem (such as one district, or year) to analyze your dataset, but refine, and present your findings and recommendations.
* Refine your set of visuals to increase fonts, add textbox callouts and arrows to bring attention to areas that are important. Please make all fonts large enough to be informative.
* Add title slide, executive summary, and recommendations/key take-aways as slides 1,2,3. Then add 10 slides of visuals. Make the last few slides, repeat of the recommendations/key take-aways, limitations of the research (limitations in your data). Add a slide on next steps/future research. Mention the data you need for further analysis of the phenomenon.
* Check the Turn-In comments below for specifics.

***Turn-in***

Create your set of charts and then make a video screencast where you click through the charts and verbally provide your analysis and interpretation. Upload to the provided dropbox;  
  
a) your .tdsx (Tableau), .pbix (powerBI), or .xlsx (Excel) file or published URL, and   
b) your video

**Academic requirements**  
You may work alone or in a group of 2.  
Higher grades given for in-depth work; lower grades given to over-simplistic analysis. Deductions applied for grammar/spelling and formatting errors.  
Shoot for 12-15 slides total and a 10-minute presentation.

**Title:**

Hello, My name is Dylan, and for my project I analyzed data on significant individual drug seizure reports from the United Nations Office on Drug and Crime. I first became interested in the world of drug trafficking, smuggling, and seizures from years of watching Drugs Inc on National Geographic, so I was glad to find a dataset like this one that provided so much information on a topic I as interested in. This dataset is comprised of over 600,000 different drug seizures reported from all around the world from 2011 to 2016.

**Executive Summary:**

This dataset shows the necessity to keep better records of drug seizures in order to successfully cut down on the quantity of drugs trafficked into different nations. This includes simple things like maintaining descriptive regions, substance origins, and smuggling methods. The case of Ephedrine shows how new drugs can spring up seemingly overnight and take over nations. The number of drug seizures grows by the year and had doubled by 2016 from the numbers in 2011, so it shows that now more than ever the growing importance to keep good records to stop the next drug epidemic from overtaking nations and their people. A standardized system should be put in place in order to ensure uniform record keeping.

**Figure 1:**

The first thing I did with this dataset to get a general feel for the data was look at the what the most common drugs being seized throughout the world were. Figure 1 shows the seizures of the top 10 drugs in the world by region, by number of kilograms seized. Various substances have different metrics for how the amount is recorded in this dataset which makes it difficult to compare quantities, so for this figure I went with kilograms since that was the most common measurement metric. The data shows that Cannabis Herb and Cannabis Resin are the most frequently seized substances in Africa and Asia, with the Americas following a similar trend with Cannabis Herb as the number one most seized, but Cocaine and Cocaine derivative and byproducts being the second, third, and fourth most seized. Further analysis of this data shows that this outlier of cocaine in the Americas is due almost entirely to Columbia and Peru which produce the overwhelming majority of the world’s cocaine supply.

The data shows that Oceania as a region is a major outlier as well which barely shows up as a blip on this chart. This led me to ask the question: what is going on Oceania that causes it to be such a major outlier on the global scale when it comes to the prevalence of the top 10 most seized substances in the world not having a foothold there.

**Figure 2:**

Figure 2 represents the total amount of different substances seized in Oceania in grams. What this data shows is that the top five most seized substances in Oceania from 2011 to 2016 were Ephedrine, Methamphetamine, Cannabis Herb (marijuana), cannabis, and Pseudoephedrine. Immediately the data showed that Oceania has a very different drug landscape when compared to the four other regions represented by the data. Cannabis herb and cannabis overlapped with the other regions, but Ephedrine, Methamphetamine, and Pseudoephedrine were major outliers. On their own Ephedrine is often used to prevent low blood pressure during anesthesia, and pseudoephedrine is used to relieve nasal congestion caused by colds and allergies. However in the production of methamphetamine, pseudoephedrine and ephedrine are the active ingredients.

**Figure 3:**

Figure 3 represents where the top 5 most seized drugs in Oceania came from. It was at this point in my analysis when I realized something interesting about the dataset and how it organizes countries within regions. Typically countries commonly included in the Oceania region include New Zealand, Australia, Fiji, Samoa, and a number of other island nations. However, this dataset only includes a sole country: New Zealand. Upon learning this the previous discrepancies in quantities made more sense when comparing one single country which represents an entire region in the dataset against other regions which contain dozens of large nations. However, despite finding this, I was now presented with a new phenomenon that was the amphetamine market in New Zealand. Figure 3 shows that Methamphetamine, Ephedrine, and Pseudoephedrine in New Zealand are all heavily importing from China, Hong Kong, and other Asian countries while cannabis almost entirely originates from within New Zealand itself.

**Figure 4:**

Figure 4 further takes a look at the evolution of New Zealand’s three main amphetamine substances. The data shows that from January 2011 through January 2016 that Pseudoephedrine was the primary seized substance. After this date though Ephedrine and Methamphetamine begin to skyrocket. The data suggests that there may have been some sort of shift in the drug market at this time in somewhere in Asia that either made Methamphetamine and Ephedrine more readily available, pseudoephedrine less available, or a combination of the two.

**Figure 5:**

This is further made visible in Figure 5. Figure 5 shows the seizures of methamphetamine, ephedrine, and pseudoephedrine in New Zealand that originated from China. The chart shows the dominance of Pseudoephedrine as the preferred substance to smuggle as early as 2012, with cases rising quickly all the way through 2013. However there is a cut off in 2014 and 2015 before Ephedrine and Methamphetamine quickly take over as the highest growing seized substances originating from China. In June 2016, only 6 months after the first recorded seizure, Ephedrine passes Pseudoephedrine for total grams seized.

**Figure 6:**

Figure 6 takes a closer look at the known worldwide sources of ephedrine from the first year that data is available from this dataset, 2011, and the final year that data is available, 2016. On the left side the data shows that Ephedrine was not nearly the phenomenon it came to be by 2016, with it only coming from two countries, Russia, and Pakistan. In 2011 Russia had produced .02% of the world supply, and Pakistan produced 3.08%. By 2016 neither Russia or Pakistan had any cases of Ephedrine seizures originating from their countries, but instead east asia had taken over as the global producer. In 2016 China produced 36.45% of the total supply of Ephedrine seized. The data shows how Ephedrine went from almost entirely unseized and trafficked in 2011, to becoming a growing market by 2016 in east asia, as well as parts of western Europe.

**Figure 7:**

Figure 7 is the final figure I had prepared, and represents the production sources of ephedrine, both known and unknown. The chart on the left includes unknown sources as well as null values. This means that when the seizure was reported, they either did not know where the Ephedrine had come from, or they simply did not put a value at all for that field. These unknown origins make up a staggering 83.1% of all Ephedrine seizures. When these values are excluded, the chart on the right shows that the top 3 producers are China with 34.6%, Hong Kong with 22.41%, and Malysia with 14.71%. Together these three countries make up 71.72% of the world known sources of Ephedrine.

**Conclusion:**

Based on this data, there are a few key takeaways. The first of which is that it would behoove government agencies to keep better records on drug seizures. Through exploring this data I found myself going down the rabbit hole of the case of Ephedrine, and I am left with several questions, the most glaring being where it comes from. If I was a government power I would want to know where substances are originating from so I could better prepare my drug enforcement agency to limit the trafficking of it into my country. The amount of unknown data and pieces of entries left blank leaves a lot of holes and vulnerabilities. By focusing on a few key metrics, this dataset could become a lot more descriptive. A standardized way of recording drug seizures would go a long way in making it easier for governments to crack down on drug trafficking as a whole.

The second key takeaway from this data is that substances can seemingly spring up and flood nations overnight as seen with Ephedrine. As governments start to crack down on certain substances, others pop up in their place. People are very creative with what they use to synthesize and create drugs, in this case prescription stimulants and cold medicine to make meth, and in other cases not explored here, using poppy seeds and poppy straw for opium. The list of controlled substances appears to grow by the day, and if nations want to be successful in stopping the importation and trafficking of these substances, better records need to be kept in order to cut things off at the source.

**Limitations:**

This dataset came with a few unique challenges that made it difficult to analyze, specifically when it came to comparing the units for the amount of each drug in each seizure. This dataset provides eleven different measures for different drugs including things such as kilograms, grams, plants, liters, and other various units of measurement. Beyond this, there is no consistency in how the drugs were reported. For example a substance may be reported as 14,000 grams by one agency, and 14 kilograms by another. The lack of consistent units meant that the quantities reported were more abstract and would benefit greatly from being standardized for comparison purposes.

Another limitation of this dataset is the lack of data and poor recording of certain metrics. It would have been really interesting to analyze metrics such as where substances were hidden, but this was often not recorded for many cases. A similar issue is present with the production country for many substances. As seen with the Ephedrine phenomenon, in over 80% of seizures the originating country is unknown. It is well recorded where the substances were being smuggled from, but it is unclear in many cases where the substance originated from.

Finally, the way that regions are set up leaves a lot to be desired from the dataset. As seen throughout this presentation, New Zealand is the only country in the entire region of Oceania in this dataset. North America and South America are also combined into a single region labeled Americas which makes it more difficult to analyze any differences between the two. The dataset does provide subregions which divides continents into several regions, but the same issue persists with New Zealand being on its own in terms of the subregion Oceania.

**Future Research:**

While this dataset provides a great amount of detail about individual seizures including when it happened, what country it happened in, where the substances originated from, where they were hidden, and various other metrics, there is more data to be integrated for further analysis of different phenomena. The United Nations Office on Drug and Crime also has data on drug retail and wholesale prices, as well as drug retail and wholesale purities. However, this data is not made publicly available in full, rather they present it with their own charts. It would be interesting to do further analysis using these metrics to look into things such as the effect of purity on wholesale value, and the different values of each substance that are seized every year to get more of an idea on the drug economy for different regions and countries. While the dataset I used for this project does a great job at making it easy to find how much of each substance is seized, its hard to get a grasp on the value of what is being seized. It would then be interesting to see the relationship between the purities of each substance in each region or country to see what effect if any it has on the smuggling habits and various metrics that the original dataset I used contains.

**Contact:**

If you have any questions about this analysis, data, or otherwise, please feel free to reach out to me at my wsu email address listing on screen. Thank you.

**Tips**:

1. Try to avoid rookie mistakes such as using counts rather than averages, reading into gaps in data, and saying more than ‘the data suggests that…”
2. Remember just because you have data, does not mean that you have found the truth. There are always hidden factors affecting your data (e.g., macro-economic forces) that you cannot see. So, use phrases like ‘the data suggests’ rather act as if your findings are factual.
3. When you present please make each chart or map full screen (hide toolbars, fields, development windows, etc.)
4. Do not use dark color text on a dark color background (might look nice on your screen but is illegible on the big screen).
5. Charts should have < 10 columns or slices of the pie.

***Rubric – grading criteria***  
**Format** - analysis is clear for the audience to comprehend. Dates are in order (Monday – Sunday, January to December, etc.) and analysis is professional looking (refined). Use slicers and hierarchies, and stacked column charts (where the legend gives more information) as much as possible to allow more in-depth analysis  
**Content** – the above-mentioned sections included  
**Depth** - Analysis that is limited in depth and quality will receive lower grades. Higher grades for more depth, such as bringing in a second dataset, and related news headlines. While depth is important, there is no need to tell your audience everything you know, rather tell them what they need to know, to inform and persuade.  
 **In-person presentation?** *Include these additional requirements* **Attire** – Please dress business casual, you will feel and present with more confidence and credibility.   
**Professionalism** – presentation is not read from note cards, presenter takes content seriously. Seek to not look like a student, rather transcend and present like a professional (this means practice your presentation)  
**Accuracy** – content analyzed correctly, expressions correct, no rookie mistakes.  
**Enthusiasm** – you can make the audience care or yawn, higher grades for somehow being compelling  
  
***From Featherman***

You are now independently walking the path of your Analytics Adventure. Your support systems are intact. Do not expect to start and finish this project in 48 hours, rather spread the analysis over days so you can reflect on what the dataset is communicating. Have some fun with this assignment. Post the URL of your published project so that recruiters and opportunity can find you.