Trevor Kahl

Applied Data Science Portfolio

# Introduction

When starting the Applied Data Science program in the Fall of 2018 data was still pieces in an Excel sheet. Working while taking classes in the Applied Data Science has shown that data in my company can be used to great value. The opportunity to utilize the data at hand in spreadsheets and turn it into meaningful pieces of business value has come from joining this program. Based on the learning objectives outlined below it my goal to show with various projects from my time in the program that these objectives were met and exceeded.

# Learning Objectives

* Describe a broad overview of the **major practice areas** of data science.
* **Collect and organize** data.
* **Identify patterns** in data via visualization, statistical analysis, and data mining.
* **Develop alternative strategies** based on the data.
* **Demonstrate communication skills** regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization.
* Synthesize the **ethical dimensions** of data science practice (e.g., privacy).

# IST 652

## Initial Data Exploration – Collecting and Organizing the Data

Our team chose the Wine Reviews Dataset <https://www.kaggle.com/zynicide/wine-reviews>) as source data for our Scripting project. This is a typical data science data source but one that gives the ability to provide multiple angles of insight on. According to Lulie Halstead, CEO of Wine Intelligence, wine consumptions has increased from 9.3 times a month in October to 9.7 times a month by lockdown in March 2020. Yet…oversupply of grapes due to wine overproduction (<https://fox8.com/news/experts-say-wine-prices-could-drop-due-to-oversupply-of-grapes>) has led to lower wine prices. The combination has led to some affordable wine drinking opportunities as well as some interesting questions and decisions on wine.

The Kaggle wine review data set we chose includes 164,000 observations and 14 attributes. These include number, country, description (review) of the wine, designation, points, price, province, region1, region2, taster name, taster twitter handle, title, variety, and winery. The original source of this dataset is WineEnthusiast (<https://www.winemag.com/>). The data was scraped during the week of June 15th, 2017.

The data consists of 10 fields:

* *Numbers*: the location of the wine in the dataset
* *Points*: the number of points WineEnthusiast rated the wine on a scale of 1-100
* *Title*: the title of the wine review, which may also contain the vintage
* *Variety*: the type of grapes used to make the wine
* *Description*: a few sentences from a sommelier describing the wine
* *Country*: the country that the wine is from
* *Province*: the province or state that the wine is from
* *Region 1*: the wine growing area in a province or state
* *Region 2*: sometimes there are more specific regions specified within an area
* *Winery*: the winery that made the wine
* *Designation*: the vineyard within the winery where the grapes are from
* *Price*: the cost for a bottle of the wine
* *Taster Name*: name of the person who tasted and reviewed the wine
* *Taster Twitter Handle*: Twitter handle for the person above

### Preparation

The Kaggle wine review data set we chose includes 164,000 observations and 14 attributes. As part of our cleaning and preparation we reduced the data size to 25,241 observations and 11 attributes. This was due to the volume of data, the limited amount of time, and the limits (code and cost) of geocoding at scale.

We removed “taster\_name”, “taster\_twitter\_handle”, and “title”. Additionally, we removed some observations where data was significantly incomplete, replaced NaN in regions with None, and filled in some gaps in pricing with averages of wines of similar variety, price point, and region. We also removed wines priced over $300.

## What is the most popular wine on Twitter? – Developing an alternative method at analyzing the data

This sentiment analysis was done as an alternative method of analyzing the data opposed from the typical built models on the wine data. For the data completed for the sentiment analysis each category was taken directly off Twitter via API calls. Each tweet was found based on the hashtag of wine that analysis was to be done on. Each call was made to call and search for 2500 tweets. Once the tweets were called using the API they were stored in an empty dictionary. Once in the dictionary they were then parsed to remove special characters and other items.

Each tweet was then parsed for characters to then go over the sentiment of the tweet. In order to determine the sentiment of the tweet they had to determine a library to use for sentiment analysis. They used the Textblob library that was built on top of the NLTK package. Textblob uses movie review that were determined either positive or negative. Then based on this each tweet is passed through a polarity function. This then determines a polarity of the tweet between –1 and 1. Looking at this data through the lens of Twitter gave the group a better sense of the real time of the wines and also an alternative method of looking at our base data.

Then based on the polarity the tweet is determined either positive, negative, or neutral which means the tweet is at 0 for its polarity. Then given the polarity they are then able to determine the percentage of each positive, negative, and neutral tweets based on the number of tweets that are being analyzed in each case.

They looked at the varietals of the initial dataset that ranked high in average scores and searched for hashtags based on those varietals. They began with looking at #wine on Twitter in order to get a feel for how people feel about the drink with no varietal.

#Wine  
 Positive tweets percentage: 30.337078651685392 %

Negative tweets percentage: 2.247191011235955 %

Neutral tweets percentage: 67.41573033707866 % \

With this they could see that wine was overall seen as a neutral drink to Twitter. They then moved to looking at the varietals and began with #syrah. In the dataset they saw that the average rating of this wine was 93.33, which was one of the higher rated.

#Syrah  
 Positive tweets percentage: 30.337078651685392 %

Negative tweets percentage: 2.247191011235955 %

Neutral tweets percentage: 67.41573033707866 % \

This was similar to what they saw with wine which they found to interesting. Another interesting item to note was they saw that some of the negative tweets may not be considered negative by themselves, but the sentiment analysis saw them as such. They then looked at #malbec which had an average of 93.

#Malbec  
 Positive tweets percentage: 18.88888888888889 %

Negative tweets percentage: 1.1111111111111112 %

Neutral tweets percentage: 80.0 % \

#malbec seemed to show that less people to enjoy the wine according to the sentiment. It seemed that many users felt just alright given the choice to drink. They then choose to look at #pinotnoir which had a middle rating of 89.1. They chose to look this up given the relative popularity of the varietal in society.

#PinotNoir  
   
 Positive tweets percentage: 36.7816091954023 %

Negative tweets percentage:2.2988505747126435 %

Neutral tweets percentage: 60.91954022988506 % \

#pinotnoir showed the most positive reaction from Twitter and the least neutral reaction. Following this they then went to look into the least popular varietal of #lambrusco which had an average rating of 83.

#lAMBRUSCO  
 Positive tweets percentage: 28.571428571428573 %

Negative tweets percentage: 2.857142857142857 %

Neutral tweets percentage: 68.57142857142857 % \

This varietal was not a surprise as it followed its rating and did not impress the users of Twitter. They had planned to look further into the vineyards of that had high ratings but gathering tweets on these locations was not feasible via Twitter and they could not move forward with sentiment analysis of those items.

# IST 707

International organizations, such as the United Nation’s High Commissioner for Refugees (UNHCR), were created for the sole purpose of responding to this refugee crisis. Through private donations and UN member nation burden-sharing, UNHCR seeks to: ease pressure on host countries, enhance refugee self-reliance, expand access to third country solutions and support setting conditions conducive for a safe return to host countries. As funding and involvement from member states decreases, UNHCR must effectively prioritize flashpoint locations to maximize their response options amidst increasing uncertainty.

## **Analysis and Models**

**The data gathered for this project differed from the previously discussed as it was data that was not used for competitions purposes but governmental. This gave us as a group a more open-ended end goal and not those predetermined by others. Initially, data was pulled from the UNHCR’s website. The original data set was comprised of 15 tables (sheets) in Excel each with similar attributes. For the purposes of this project, only 5 tables were used: 1 and 2, which covered Destination country (where applications for asylum are submitted)) totals, and 3-5, which covered Origin country (where applicants come from) totals. All tables include a base of similar attributes with raw aggregate totals per year for application numbers from 2010-2014, with countries in rows. The tables also had an Aggregate Total and Scaled Total columns already included. The scaled attributes included applications per 1,000 inhabitants and applications per 1 USD/GDP per capita. Each total was then in another preset ordinal rank column for all countries in the table. The most cleaning that had to occur with the initial data was converting Rank attributes (labeled clearly in 6 different attributes) from numeric to categorial.**

|  |  |
| --- | --- |
| **Original Attributes (UNHCR)** | |
| **Attribute** | **Description** |
| **Country** | **List of country name (same in both Origin and Destination tables)** |
| **2010** | **Total aggregate count of applications either to or from the country (depending on table) in the year 2010** |
| **2011** | **Total aggregate count of applications either to or from the country (depending on table) in the year 2011** |
| **2012** | **Total aggregate count of applications either to or from the country (depending on table) in the year 2012** |
| **2013** | **Total aggregate count of applications either to or from the country (depending on table) in the year 2013** |
| **2014** | **Total aggregate count of applications either to or from the country (depending on table) in the year 2014** |
| **Total** | **Total numeric count of applications from 2010-2014** |
| **Annual change ’14-‘13** | **Percent change in applications from 2013 to 2014** |
| **Rank2014** | **Categorical ranking of high number of applications for just 2014** |
| **Rank2010-2014** | **Categorical ranking of high number of applications for 2010-2014** |
| **Per1000Inhabitants2014** | **Scaled number of applications per 1000 inhabitants in the country (AKA a control for population) for 2014** |
| **Per1000Inhabitants2010-2014** | **Scaled number of applications per 1000 inhabitants in the country (AKA a control for population) for 2010-2014** |
| **Rank1000\_2014** | **Categorical ranking of 1000 inhabitant scaled applications for just 2014** |
| **Rank1000\_10-14** | **Categorical ranking of 1000 inhabitant scaled applications for 2010-2014** |
| **Per1USDGSP2014** | **Scaled number of applications per 1 USD/GSP per capita in the country (AKA control for GDP) for 2014** |
| **Per1USDGSP2010-2014** | **Scaled number of applications per 1 USD/GSP per capita in the country (AKA control for GDP) for 2010-2014** |
| **RankGDP\_2014** | **Categorical ranking of applications per 1 USD/GDP per capita for just 2014** |
| **RankGDP\_2010-2014** | **Categorical ranking of applications per 1 USD/GDP per capita for 2010-2014** |

**In addition to relabeling the Rank attributes as categorical, the data had to be cleaned to remove the top and bottom chunk of rows for each table. The initial format of the sheets included supplementary information and instructions on how to analyze the results. While useful, it resulted in several NAs when read into R and were removed when imported as data frames for the purpose of this project.**

**After initial analysis, which is to be discussed below, it was determined that additional data would need to be pulled in order to better make sense of the differences in scaling for different countries. The addition of the data from the Freedom House aimed to provide greater insight onto why different countries would have a high number of applications. The next pull of data from Freedom House includes the following indices for each country:**

|  |  |
| --- | --- |
| **Appended Attributes (Freedom House)** | |
| **Attribute** | **Description** |
| **Region** | **Geographical region, as a factor (Europe or Asia)** |
| **Free** | **Indication of freedom status as Free, Not Free, or Partially Free (F, NF, PF)** |
| **PoliticalRights** | **Numeric rating from 1-7 (1 most free, 7 least free)** |
| **CivilLiberties** | **Numeric rating from 1-7 (1 most free, 7 least free)** |
| **Governance** | **Function of government, score from 1-12 (1 poor, 12 best)** |
| **Expression** | **Freedom of expression and belief rating 1-16 (1 poor, 16 best)** |
| **Electoral** | **Electoral process rating 1-12 (1 poor, 12 best)** |
| **PolPluralism** | **Political pluralism and participation rating 1-12 (1 poor, 12 best)** |
| **IndividualRights** | **Personal authority and individual rights rating 1-16 (1 poor, 16 best)** |
| **RuleofLaw** | **Rule of law rating 1-16 (1 poor, 16 best)** |
| **Organization** | **Association and organizational rights 1-12 (1 poor, 12 best)** |
| **PolRights** | **Score from 1-40 (total of Electoral, PolPluralism, and Government)** |
| **Liberties** | **Score from 1-60 (total of Expression, Organization, IndividualRights, and RuleofLaw)** |

Once the data from the UNHCR and Freedom House were located, they needed to be consolidated. In order to properly do this, the data was matched on a key. The key was country in this case and the data was merged in Excel rather than in R. The Freedom House Index values were appended to the original data of the UNHCR. Once the tables were properly merged in Excel it needed to be cleaned in order to analyze for future modeling.

The rows that were either totals or did not have a country related were removed from the tables. Then the data was given proper headers to then be able to easily fix column types. All data was read in as “character” and needed to be changed to either factor or numerical. Country, Region, Free, Political Rights, and Civil Liberties were converted to factors while all other variables were transformed to numerical.

Once all columns were associated to their proper type, another data frame was created to make cuts within the data either to be binned. Binning was done with three to four bins and was based on the range of the data in each of the data. These separate data frames allowed for flexibility for modeling that will be shown in the below analysis.

## **EDA – Identifying Patterns based on the Data**

**Initial evaluative analysis of the data revealed a lot of what would be expected. This would give us a good way to show patterns that can been in the data via visualization. Since the UNHCR data was pulled from a site conducting a report in 2014, part of their goal was to compare across years how the total number of applications was affected. The below first set of EDA includes histograms for 2010 and 2014 to compare the change in years. Each has an obvious right skew with the majority of applications residing in the lower range per country, however, as years go on the bin widths widen. The comparison of the bin sizes and max and mins of each histogram per year demonstrates a near tripling in the total number of applications from 2010 to 2014.**

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Initial analysis was also conducted on the aggregate totals and scaled total values. The below are depictions of the top 10 ranked countries for each category for origin countries, with the top two in royal blue:**

A picture containing drawing

Description automatically generated

Figure 1: Total Aggregate Applications

A picture containing drawing

Description automatically generated

Figure 2: Total per 100 Inhabitants

A screenshot of a cell phone

Description automatically generated

Figure 3: Total per1 USD/GDP

**From the creation of the above visualizations, it was clear the Germany and the U.S. were quite high-volume destination countries, but more peculiarly, Turkey was listed as a country with a high-volume of applications in both the aggregate and GDP scaled models.**

Once application totals were inspected, further analysis using the Freedom House indices was conducted to better understand the behavior of the country scores for both the origin and destination countries. Destination countries tended to have most labeled as F (red) and very few as PF (blue) or NF (green). The only NF was, as expected, Turkey. The below are visual representations of the total scores for PoliticalRights and CivilLiberties for destination countries.

A picture containing people

Description automatically generated

Figure 4: Destination Political Rights Scores

A screenshot of a cell phone

Description automatically generated

Figure 5: Destination Civil Liberties Scores

The same visualizations were created for the origin countries – however, the results are a bit more scattered across the board – there are two countries labeled “Free”, Ghana and India, that are useful to note in modeling.

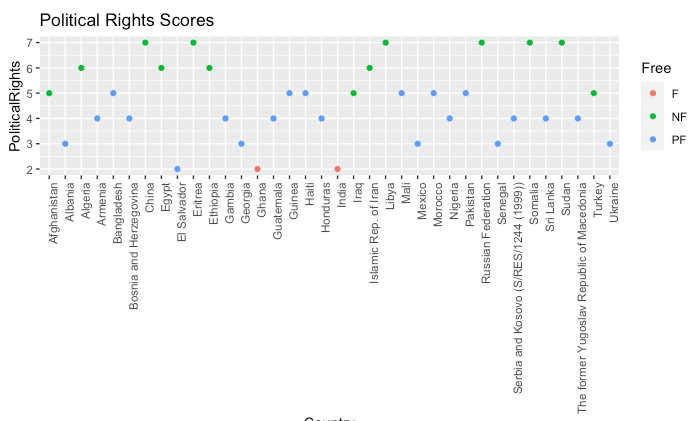
****

Figure 6: Origin Political Rights Scores

A picture containing clock

Description automatically generated

Figure 7: Origin Civil Liberties Scores

## Models - Developing multiple alternative methods at analyzing the data

Association Rule Mining

The purpose of Association Rule Mining (ARM) is to determine interesting and viable relations in variables to see if the data can be explained by certain variables. This was a chance for our team to find another way to analyze the data that was presented in an interesting way. In the case of this analysis, the team was interested in looking at where refugees were going to and originating from. With this the team was interested if the variables could define a trend that could be pointed towards with a strong and reliable correlation towards predicting trends.

The trends that come from the rules can be identified with reason that if something occurs then it can be explained by certain factors. The reliability of these rules can be deemed from the confidence, support, and lift. With these values the rules can be identified to be trusted for future analysis when collecting further data on more countries.

## Results

*ARM Assessment*

For the Association Rule Mining the team initially looked to run rules on the tables that involved the information of where the asylum seekers were sending their applications. The below rules will show that given then various other methods that were discussed that an alternative method was determined in order to analyze based on the data set.

lhs rhs support confidence lift

{2010=small, Rule of Law=mid-low} => {2014=small}    0.2 0.8 1.10

This rule shows that if the country had a small 2010 of applications submitted and low rule of law then the outcome would usually result in a small number of applications in 2014. This rule shows decent support and confidence. The lift is promising at 1.10 and shows that it can be generally dependent.

lhs rhs support confidence lift

{Free=F,Governance=mid-high}     => {2013=small}       0.3    0.8 1.07

The above rule shows that if the Freedom Index has indicated that the country is Free and their Governance is mid-high then they would have a small number of applications in 2013. The confidence is fairly high at 80 percent, but the lift again is showing that this is a rule that can be trusted.

lhs rhs support confidence lift

{2013=small-high,Liberties Score=F}   => {2012=small-high} 0.1    1    5

This rule shows that if the country had a small-high number of applications in 2013 and a Liberties score of Free then their 2012 was likely small-high. The small support and confidence of 1 is somewhat questionable in this case. The high lift at 5 too also can point that this rule is important, but should be taken with a grain of salt as there may have only be a handful of countries that follow this.

Next, the team wanted to see if when looking at where the asylum seekers were coming from, results would differ from the above rules that were generated from where the applications were being put in.

lhs rhs support confidence lift

{2013=highest,PolPluralism=mid} =>  {PolRights=mid-low}    0.36  0.8    1.9

This rule is showing that if the PolRights is mid-low then it is likely that the country had a highest 2013 and mid PolPluralism. This is interesting to show what can make up some of the Freedom Index values from both Freedom Index and UNHCR variables. The support at 36 percent and confidence of 80 percent are good. The lift of 1.9 is also a good indicator that this rule can be trusted.

lhs rhs support confidence lift

{2014=highest,Expression=low}    =>    {Free=NF}        0.36  0.86    2.2

This rule is showing that if 2014 had a highest number of seekers from the country and a low Expression then the country would be considered Free. The support and confidence show strong reason to believe in this rule. The lift also supports both the support and confidence and show that the rule can be valid.

lhs rhs support confidence lift

{Expression=low,Liberties Score=NF}    =>             {Governance=low}    0.36  0.86    1.9

This final rule is saying that if the Governance of the country is considered low then they have a strong chance of having a low Expression and a Not Free Liberties value. The support and confidence match the rule above and as previously discussed show good value to the rule. Then the lift is a bit lower than the previous but is still valid at 1.9 and shows strong promise to mark this rule as reliable.

From here the top rules with confidence, support, and then lift were run. With these rules the team took the top 5 from the where the seeker was submitting their applications and where they were from. With this, it gave a good indicator into the various effects on the data.

*Confidence Asylum Applications Submitted*

lhs rhs support confidence lift

{Governance=high,Rule of Law=mid-high} => {PolRights=high} 0.2 1 1.8

{Governance=high,Rule of Law=mid-high} => {PolPluralism=high} 0.2 1 1.4

{Governance=high,Rule of Law=mid-high} => {Political Rights=1} 0.2 1 1.4

{Governance=high,Rule of Law=mid-high} => {Liberties Score=F} 0.2 1 1.3

{Governance=high,Rule of Law=mid-high} => {Organization=high} 0.2 1 1.2

*Support Asylum Applications Submitted*

lhs rhs support confidence lift

{Organization=high,Liberties Score=F} => {Free=F} 0.78 1.00 1.1

{Free=F,Liberties Score=F} => {Organization=high} 0.78 1.00 1.2

{Free=F,Organization=high} => {Liberties Score=F} 0.78 0.91 1.2

{Political Rights=1,Organization=high} => {Free=F} 0.72 1.00 1.1

{Free=F,Political Rights=1} => {Organization=high} 0.72 1.00 1.2

*Lift Asylum Applications Submitted*

lhs rhs support confidence lift

{Expression=mid,Rule of Law=mid-low} => {Civil Liberties=2} 0.2 1 4.4 8

{2012=small,Expression=mid,Rule of Law=mid-low} => {Civil Liberties=2} 0.2 1 4.4 8

{2014=small,Expression=mid,Rule of Law=mid-low} => {Civil Liberties=2} 0.2 1 4.4 8

{2013=small,Expression=mid,Rule of Law=mid-low} => {Civil Liberties=2} 0.2 1 4.4 8

{2011=small,Expression=mid,Rule of Law=mid-low} => {Civil Liberties=2} 0.2 1 4.4

*Confidence Origin of Asylum Applications Submitted*

lhs rhs support confidence lift

{PolRights=mid-high,IndRights=mid-high} => {Civil Liberties=low} 0.21 1 2.5

{PolRights=mid-high,IndRights=mid-high} => {2014=highest} 0.21 1 1.0

{Organization=mid,IndRights=mid-high} => {Civil Liberties=low} 0.21 1 2.5

{Civil Liberties=low,IndRights=mid-high} => {2014=highest} 0.24 1 1.0

{2014=highest,IndRights=mid-high} => {Civil Liberties=low} 0.24 1 2.5

*Support Origin of Asylum Applications Submitted*

lhs rhs support confidence lift

{2013=highest,Liberties Score=NF} => {2014=highest} 0.64 1.00 1.0

{2014=highest,Liberties Score=NF} => {2013=highest} 0.64 1.00 1.1

{Civil Liberties=mid,Liberties Score=NF} => {2013=highest} 0.61 1.00 1.1

[{2013=highest,Civil Liberties=mid} => {Liberties Score=NF} 0.61 1.00 1.6

{2013=highest,Liberties Score=NF} => {Civil Liberties=mid} 0.61 0.95 1.6

*Support Origin of Asylum Applications Submitted*

lhs rhs support confidence lift

{PolPluralism=low,Rule of Law=low} => {PolRights=low} 0.27 1 3.3

{Free=NF,PolPluralism=low} => {PolRights=low} 0.33 1 3.3

{Civil Liberties=low,Expression=mid,Organization=mid} =>{Liberties Score=PF} 0.24 1 3.3

{Civil Liberties=low,Organization=mid,Rule of Law=mid-low =>{Liberties Score=PF} 0.24 1 3.3

{Free=NF,PolPluralism=low,Rule of Law=low} => {PolRights=low} 0.27 1 3.3

With a smaller data set these rules show some variation in that they look to be viable, but they can be deceiving with such good confidence numbers they were not overwhelmed by the outcomes. They also wanted to see the rules on the tables when the left-handed side was set to Free on the table of the Asylum applications submitted. The for the origin of the seekers they set the left-handed side to be Not Free. The rules are straightforward and common sense but still important to note when making further inspections into the countries.

*LHS Free Asylum Applications Submitted*

lhs rhs support confidence lift

{Free=F} => {Organization=high} 0.85 0.97 1.1

{Free=F} => {Liberties Score=F} 0.78 0.89 1.1

*LHS Not Free Origin of Asylum Applications Submitted*

lhs rhs support confidence lift

{Free=NF} => {Governance=low} 0.39 1 2.2

{Free=NF} => {Organization=low} 0.39 1 1.8

# IST 623 – Synthesizing Ethical Dimensions of Data

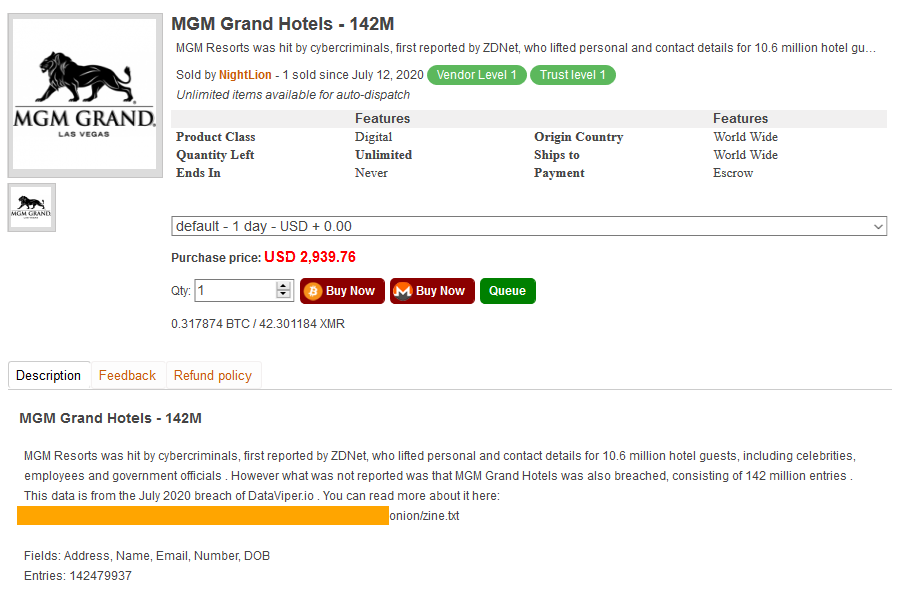
In July of 2019 a major data breach exposed MGM’s customer personal data after a hacker gained access to their cloud server. With the outlook of this project, it is important to note that all data from previous projects was publicly available. It comes in these lines where the source of data must be known. This specific project shows that data can be found on the internet and sourced through unethical ways.

Although MGM being hacked in 2019 the news broke in February of 2020 that over 10 million customers data was breached. The customer’s personal information was offered as a free download in hacking forums. Of the 10 million accounts that were breached high profile clients such as Justin Bieber and Jack Dorsey were among them. It was reported by MGM that users were informed of this breach of their information in 2019, but accounts from customers were contrary to this.

Graphical user interface, text, application

Description automatically generated

As MGM dealt with the initial breach word broke that even more security threats loomed. Another hack occurred from the security company DataViper. A Hacker claimed to have spent three months inside DataViper servers while exfiltrating databases. Around 8,200 databases were breached and the top 50 were put up for sale on the dark web. The owner of DataViper believed that the hacker was selling their own database’s and not real data. This led to 142,479,937 MGM hotel guests accounts breached and put up for a price just over $2,900. MGM spokesperson also pointed out that "the vast majority of data consisted of contact information like names, postal addresses, and email addresses”. There is a potential this could have been larger than MGM was willing to admit, but this was not made public aside from notifying the affected customers.



Following this incident MGM has shown little resolve and initiative to fix the breached customer data. This does however lead to potential to offer insight into how to combat these miscues in the future. Based on these incidents customers are more prone to phishing attacks and other attacks. It was the outcome of the group that two-factor authentication should be put into place and that proper communication from these companies should be put into place. Other basic methods to help mitigate the breach would be the below. Although these are basic steps to take, they are crucial in ensuring safe practices.

1. Change passwords
2. If you don’t already have it, set up credit monitoring
3. Practice good cybersecurity habits
4. Keep a record of your response
5. Stay alert

# Conclusion

In the 18 months of my education in the Applied Data Science program I was able to learn multiple programming languages after never having any hands-on time previously with them. As data is an ever-growing entity in the world today with smart connected devices created more data that what was previously known. With this program it has shown me that data can be broken down and analyzed in order to determine the best application of that specific set.

As this program has taught me data is more than a csv file or a json file. It must be obtained, scrubbed, explored, modeled and then finally interrupted. All these pieces then align to create a coherent picture to your stakeholder in order to then give a clear and concise message of how to effectively provide contextualized results.