Casino Royale

Hayes Vavpetic
Computer Science
Data Mining 4502
hava4541@colorado.edu

Cameron Mattson
Computer Science
Data Mining 5502
cama1171@colorado.edu

Thomas Neal
Data Science and Math
Data Mining 4502
thne6271@colorado.edu

Michael Grodecki
Computer Science
Data Mining 4502
migr1903@colorado.edu

ABSTRACT

To find interesting patterns in Colorado casino tax data we implemented a variety of classification models, and analyzed multiple trends. Specifically, temporal data, spatial data, device data and device attribute data were segregated for this purpose.

One prediction model and two classification models were generated. The Neural networks classified the season and the casino location using aggregate slot machine attributes, while a linear regression predicted the AGP of aggregate slot data. Other interesting patterns were visualized from this data by generating various pie charts as well line graphs.

1 INTRODUCTION

Since casino gambling became legal in Colorado in 1990 it has become a significant economic pillar of the state. These casinos are located in three areas: Black Hawk, Central City, and Cripple Creek. Between them they generate over 800 million dollars in revenue annually. However, despite this, very few people know the nuances of how these casino generate there money. Analyzing how these casinos generate profit and how it has changed over time is of interest to several parties. For instance, casino owners can determine how to maximize casino profit by analyzing revenue trends. Since these casinos generate so much money in a small area, their revenue is a significant part of the local economy, so local residents have the incentive to care about the revenue that these casinos are generating. Additionally, casinos are a tightly regulated industry and thus analysis of their revenue can affect policy makers' decisions and the laws that govern them. We plan on analyzing available casino data and hope to generate conclusions that are relative to all parties mentioned.

2 RELATED WORK

In analyzing online casino gambling, one study proposed a method for clustering gamblers according to the ways in which they gambled. Another study conducted the same type of partition clustering method we are proposing, except this study compared k-means clustering with k-medoid clustering when evaluating a different data set [5]. The idea of analyzing our clusters using a confusion matrix has been implemented in a study attempting to cluster multiple documents [2]. However, we did not implement an additional algorithm in understanding how well our clusters performed as in this study. We also did not use a confusion matrix, since our data consisted of more than one class label. Our data was also labeled, so we determined that a supervised learning model would be a better choice for this project.

Other studies evaluated the performance of slot machines [1], the relationships between devices [4], the relationship between table games and casino revenue [4], and the demand for casino gaming [3]. The demand for casino gaming can also be determined using AGP (i.e. Adjusted Gross Proceeds, a casino's gross receipts, less winnings paid to wagerers), as a surrogate for revenue and performance. These studies were very similar to the types of analysis we performed using attributes such as AGP and device revenue.

3 METHODOLOGY

3.1 Initially Proposed Work / Main Tasks

Our goal with this project is to use Colorado state casino tax data to gain an understanding of current gambling trends in the state in order to provide meaningful conclusions to parties of interest. To do this we will start by preparing the data, which includes importing and cleaning. Then we will explore and create some general visualisations of the data to find meaningful trends. We will be using several analysis tools such as clustering, regression, and supervised learning to draw conclusions from the datasets.

3.2 About the Data

Since casino gambling is tightly regulated, the state of Colorado has kept in-depth monthly records on how much money these casinos are making and the devices that generate this profit. The Colorado department of revenue has made these records public from the years 2011 until 2020. This is the data we used for our analysis. The data consists of 9 excel files, one for each year. Each individual sheet contains a column that represents a specific location: Black Hawk, Central City, Cripple Creek, a fourth column for state totals and attributes that represent: device name, coins in, AGP (Adjusted Gross Proceeds), average daily AGP, hold percent, and the rest of the columns represent the value of these attributes per month.

To understand this data it is important to understand what these attributes represent. For example, the word 'device' refers to a machine or table game owned by the casino that generates revenue. This includes the various coin slots such as penny slots, dollar slots, high value slots, etc. This also include table games like blackjack and poker. For every casino the amount of each device in these categories is summed for each month. Then the attribute "coins in" or "<table game> drop" both refer to the total amount of money put into these devices each month. Similarly the attribute "AGP" refers to how much the casino made in profit off those devices in that month and "average daily AGP" is the average amount of money a group of devices made in the casino on any given day of that month. Lastly the attribute "hold %" refers to the percentage of the money that gamblers put in that the casino keeps as profit. The figure below is an example of the data attributes for further clarification.

	A	B	C	D
1	**2018-2019**	**Tax Year Basis**		
2			JULY	AUGUST
117	**Cripple Creek**	1¢ Slots	1,944	1,929
118	**Cripple Creek**	Coins In	86,538,069.41	86,158,169.76
119	**Cripple Creek**	AGP	7,049,998.43	7,094,108.06
120	**Cripple Creek**	Avg Daily AGP	116.99	118.63
121	**Cripple Creek**	Hold %	8.14%	8.23%
122				
123	**Cripple Creek**	5¢ Slots	129	129
124	**Cripple Creek**	Coins In	3,647,202.95	3,615,903.00
125	**Cripple Creek**	AGP	254,051.18	207,153.80
126	**Cripple Creek**	Avg Daily AGP	63.53	51.80
127	**Cripple Creek**	Hold %	6.96%	5.72%
128				

Figure 1: Example of a section of one of the excel sheets showing the attributes for the Cripple Creek's 1 and 5 cent slots during July and August for the 2018-2019 year

3.3 Tools

We used the following python libraries to help analyze the data throughout this project.

3.3.1 Numpy. Numpy is one of the libraries that we used to help clean and analyze the data that is going to being used throughout this project. Familiarity with Numpy is very common since we have used it in this class before and Numpy brings in a variety of different mathematical functions that give summaries about the data. Numpy provides functions that can be used on entire arrays to find means, averages, highs, lows, etc... for entire data sets that could be thousands of indexes long.

3.3.2 Pandas. Pandas is another library that we have used in this class before and one that has been very useful throughout this project. Pandas provides data frame objects that allow for fast and efficient data manipulation, functions for reading and writing data easily to and from CSV or excel files, which is what our data is presented as. Pandas allows for data to easily be inserted, deleted, sliced, indexed or reshaped through the use of the functions and data frames.

3.3.3 Matplotlib. Matplotlib is the final library that has been used so far in this class that we used throughout this project. Matplotlib brings functions in that make displaying data significantly easier while also allowing there to be many different ways that the data can be displayed.

3.3.4 Scipy. Scipy is a library that is all about data manipulation and analysis. Scipy has a subset of libraries that covers almost anything that could be used while working with data. One of the specific libraries that we imported are scipy.stats, a function that helps with creating regression models, linear models and time series analysis.

Keras and Tensorflow. Keras and Tensorflow were used to build, train, and validate our model. This allowed us to retrieve the categorical accuracy and categorical cross-entropy metrics for each epoch in both the training and validation sets.

3.3.5 Scikit-Learn. Scikit-Learn is a library that provides the functionality to cluster our data using a k-means clustering approach. Similarly we used Scikit-Learn to implement the k-means approach depending on the distribution of the data. Keras was also implemented to organize the training class data labels using a one-hot encoding so the labels could be represented as binary arrays.

3.4 Changing the Data

3.4.1 COVID Data. One of the biggest changes that had to be made to the data was cutting out all data points from March 2020 onward. When COVID swept through America, all casinos were completely shut down for March and April. After that only certain games at certain casinos were opened back up in order to comply with the new COVID guidelines. When graphs and tables were being made, the data from COVID disrupted the trends and patterns that we were trying to show, so as a result it was decided that the best move going forward was going to be to only analyze the data up until March 2020.

3.4.2 Location Data. In each of the dataframes, every attribute was evaluated for each of the three different casinos in Colorado as well as on an entire state level. In order to be able to analyze each of the three different casinos better we sliced the dataframes so that they would only reflect the data for each location. This allowed us to have a better look at how each casino performed specifically throughout the years and also allowed us to make good comparisons regarding AGP and coins in per casino compared to the entire state.

3.4.3 Seasonal Data. When looking for different trends and patterns, one trend that potentially showed promise was analyzing the casinos per each season. The dataframes were sliced up into four different seasons, summer, fall, winter and spring, that all contained three different months of the year and then they were analyzed.

4 EVALUATION

4.1 Data Visualization and Trend Hunting

4.1.1 AGP.

With our data clean and organized we began visualizing different attributes over time in figure 2 to see if any interesting trends appeared. The first trend looked at the annual total for Adjusted Gross Profit (AGP) over time: In figure 2, we see a consistent increasing trend in profit up until the start of COVID-19 where we then see a rapid decline in profit. This is expected because the casinos were forced to shut down due to COVID-19 restrictions and lock-downs. Another interesting trend appears when we analyze AGP by devices as in the following figure 3.

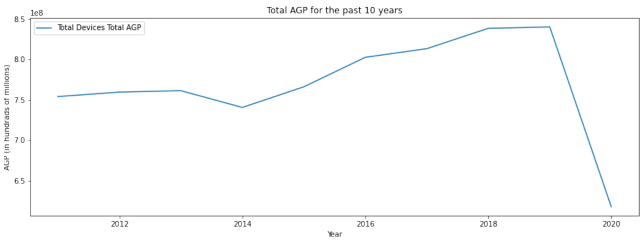

Figure 2: Profit produced state wide for the past decade

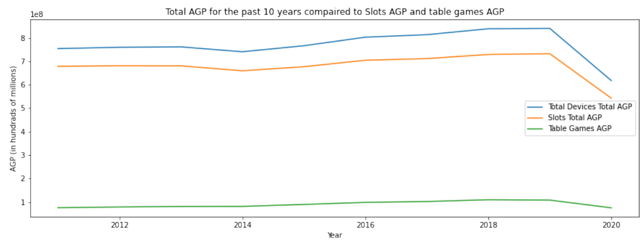

Figure 3: State wide profit totals for the different types of devices

In figure 3 we see that slot machines make the majority of the profit for the casinos as well as table games, whereas black jack or roulette, for example, only account for a small fraction of the statewide total profit. This was surprising to us because we expected most of the profit to come from big spenders at VIP table games. However, it seems to be the slot machines that make the majority of the money. There are more interesting patterns to be mined from this data. To analyze one of these patterns we decided to compare AGP by slot machine type to see if there was one slot type in particular that had a different relationship compared to other slot machines. These relationships are shown in figure 4. For

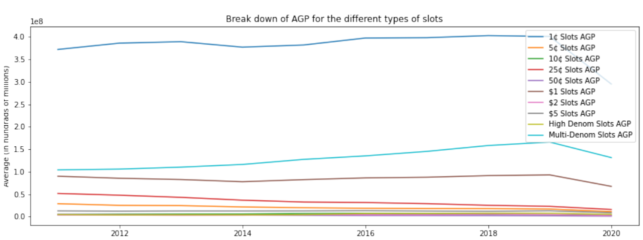

Figure 4: State wide profit totals for the different types of slot machines

highly profitable slots, penny slots or 1¢ slots make considerably more profit than any of the other slots. Comparing the penny slots to the other slot types, in any given year, the penny slots make almost 250 million more dollars annually than the other slots. We also see that most of the other categories for slots are falling out of favor while the multi-denomination slots are gaining in popularity based on the AGP. This could be due to the casinos implementing more electronic slots that accept a wider range of denominations as viable bets. Now, by similarly analyzing the different types of table games by AGP we also see the following trends fin figure 5.

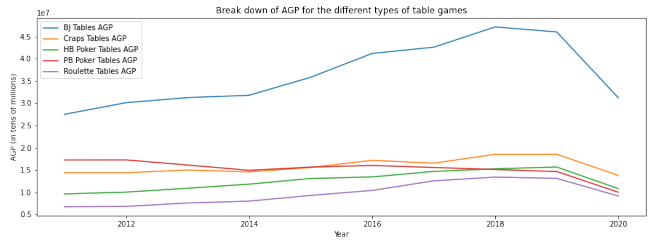

Figure 5: State wide profit totals for the different types of table games

In Figure 5, when looking at AGP for different slot types we measure in hundreds of millions. However, to properly see the data in Figure 6 we need to look at the scale in tens of millions, so immediately we know that these table games play less of a role in the overall profit when compared to the slots. To summarize, we have found that statewide casino profits were increasing up until the crash in 2020. We also learned those profits are mostly comprised of profits generated by slot machines and that table games make up a very small minority of the overall profits. Similarly, we know that of the slot machines, the 1¢ slots make almost double of any other slot machine type. To show these trends in a different way rather than considering the changes over time we found the average of these attributes across the ten years and created pie charts to show the proportion of AGP that each device type contributes to the whole. The first pie chart (figure 6) is a simple breakdown of AGP split between the slot machines and the table games: Here we see

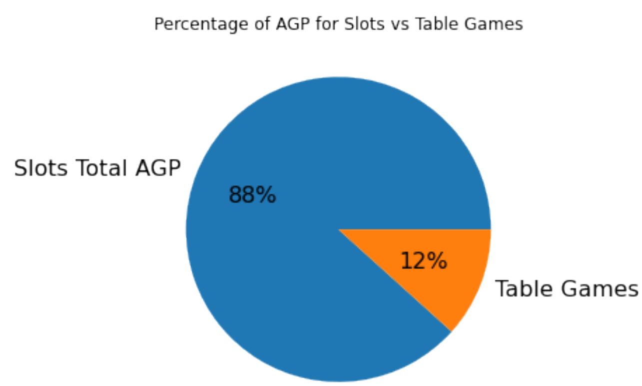

Figure 6: Ten year averaged AGP split by proportion

that the slots make up 88% of the total AGP on average and table games only account for 12%. To dig a bit deeper we also created a pie chart showing the contribution of each slot to the overall total AGP alongside the contribution by table games (figure 7). It should be noted to make the chart more readable $2 and $5 slots are included in the high denomination category.

Looking at figure 7 we can see that the 1¢ slots make about 50% of the overall profit for casinos. This is exactly what we expected from the previous trends we found.

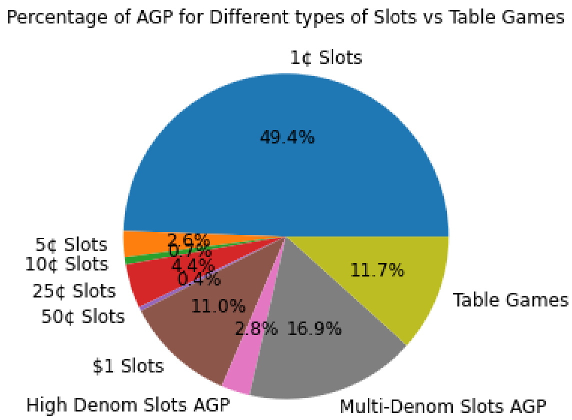

Figure 7: Comparing AGP of different slot devices and all table games

4.1.2 Hold Percent.
In order to gain a better understating of the trends we observed, we decided to look at the hold percent over time. The hold percent is calculated by $\frac{AGP}{\text{Money in}} * 100$. In other words it is the percentage of money that the device keeps. In figure 8 we decided to compare the hold percents of table games and slots, respectively.

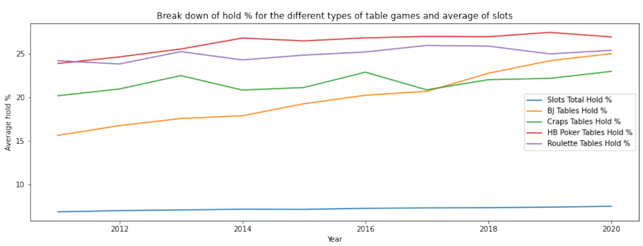

Figure 8: State wide hold percent for different table games and average of slots

Interestingly, what we observed is that the slots have a seemingly significant lower hold percent than that of any table game. This is unexpected because slots constitute the majority of the profit, this leads us to conclude that the amount of bets being placed on slots is significantly higher than the amount of bets placed on table games. We can also draw some basic conclusions, like if you are going to bet on table games, Craps is one of the best options because historically it has the lowest hold percent which means that it has the highest chance that the player will win. The following figure 9 shows the hold percent based on the different types of slot machines:

Here we see that the 1¢ slots have a higher hold percent, around 10% when compared to the other slots that are around 5-7%. This makes sense, because the casinos know that penny slots make a large percentage of their overall profit, so they would want to maximize their profit by maximizing the hold percent. This made us curious regarding the legalisation around slot machine win percentages, so we decided to do some investigating and according to the Colorado Department of Revenue, "By law, slot machines

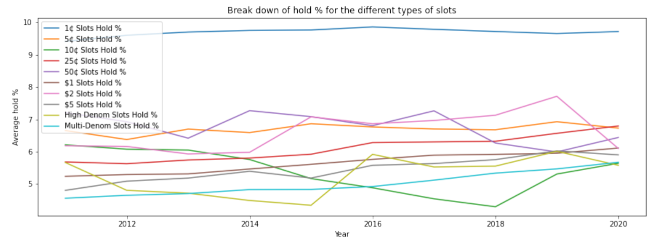

Figure 9: State wide hold percent for different slot machine types

must pay out between 80 percent and 100 percent, over the life of the machine. Most slot machines pay out around 90 percent, with higher denominations paying out higher than lower denominations. A quarter slot machine generally pays out more than a nickel slot, a dollar slot more than a quarter slot, etc. The percentage of payout on a slot machine is determined by a computer chip within the machine itself". This was the expected outcome, that the penny slots do payout less, however, by law they still must payout around 90%. Thus we conclude that if you want to play slots, the penny machines are the worst option for casino customers but the best option for casinos to make money.

4.1.3 Monthly data.
Previously we were looking at the annual totals for AGP, however, there is potentially a trend that can be revealed by looking at the monthly AGP such as in the following figure 10. Here we see some-

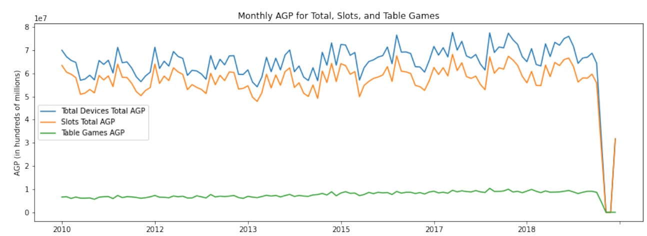

Figure 10: State wide monthly AGP

thing interesting about profit after the casinos reopened in 2020. They reopened with only the slot machines available and thus they accounted for all the the following profit. It makes sense that the casinos would prioritize reopening the slots, since historically they make a majority of of the casino's money. Outside of this we can see the emergence of another trend. There appear to be seasonal highs and lows and this is a trend that we decided to explore further.

4.1.4 Seasonal Trends. To analyze the seasonal trends we decided to find the average AGP for each month over the past ten years. We then considered the 'seasons' to be the following: spring(green)= March, April, May, summer(red)= June, July, August, fall(tan)=September, October, November, winter(blue)= December, January, February. We used this definition to create a color coded bar chart representing the average AGP for each month.

Here we see potential for a bi-seasonal trend with summer being the highest-earning season and winter being the lowest. However the clearer picture that is being painted is that there are certain

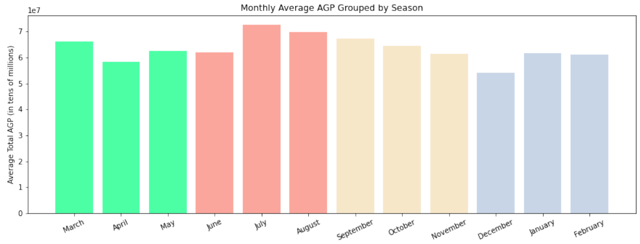

Figure 11: Ten year AGP averaged per month and color coded by season

months that consistently perform well or poorly. We can see on average July is the highest earning month across the state and December is the lowest. These highs and lows could be for any number of outside factors a cause could be that the summer months like July bring more people into the mountain areas around the casinos for reasons like hiking, rock climbing or visiting Rocky Mountain National Park located nearby. While in the winter most tourists come for the skiing and the major ski mountains are further away from the casinos. Additionally in December people typically have larger bills around the Holiday season and thus can't afford to lose money at the casinos. However, this is speculation and would require additional data and further analysis to prove.

4.2 Cluster Analysis

We clustered a subset of the data to determine if the computed clusters would represent the slot machines using AGP and Coins In as attributes. We determined that clusters could not represent the slot machines when we clustered by AGP and Coins In, since the data for different slot machines overlaps as shown by comparing Figure 4 and Figure 5 below. However, we may have been able to separate the data by applying a kernel function to the data, or increasing the dimensionality of the points, which would further separate the points.

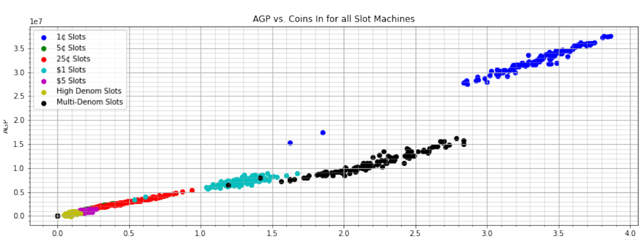

Figure 12: Slot machine data as AGP and Coins In

4.3 Regression Analysis

We wanted to implement a linear regression to show a trend line through the data and potentially make predictions using our model. We decided to create a linear model where coins in, the amount of money that goes into a slot machine, is the independent variable and the AGP is the dependent variable. From these variables we get a very linear plot and thus the linear regression is a great fit. However, if we take some time to think about this model we realizes that this isn't really an interesting or useful model. Basically all we conclude is that revenue is a good predictor of profits witch makes instinctive sense. Especially when these devices have consistent hold percents, so for differing amounts of money going in the expected profit will always be relative and thus this model is redundant. However, we are okay with this fact as our goal wasn't just to try and find a trend or make a prediction, but also to learn how to implement and interpret a linear regression. We think we were successful in this latter goal and are confident we could implement more interesting linear regressions in the future.

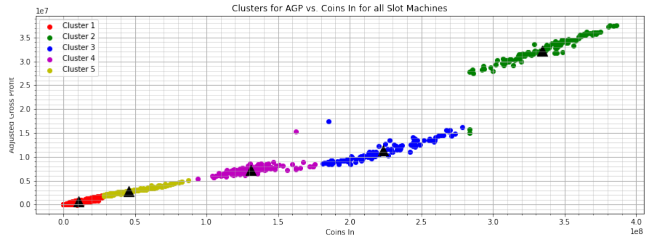

Figure 13: Clusters of slot machines using AGP and Coins In

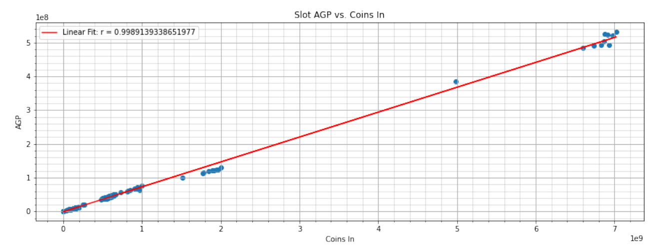

Figure 14: Linear regression of slot profits vs slot revenue

4.4 Supervised Learning

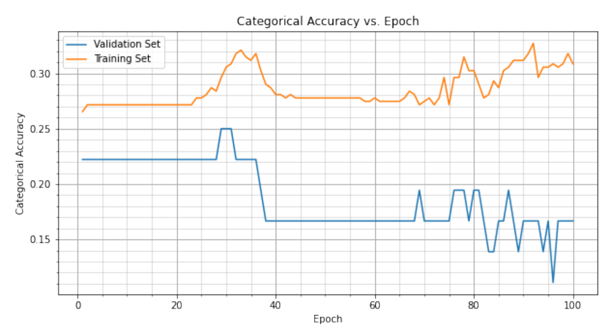

Figure 15: Categorical Accuracy vs. Epoch of seasonal neural network classifier

The plot of the seasonal neural network's categorical accuracy above seems to indicate that the model's categorical accuracy is not improving. Rather, the categorical accuracy of the validation set tends to decline with more epochs, while the categorical accuracy of the training set increases slightly with more epochs. However, both sets categorical accuracy values fluctuate around some baselines rather than monotonically increasing or decreasing.

In the plot above, the validation set's categorical cross-entropy

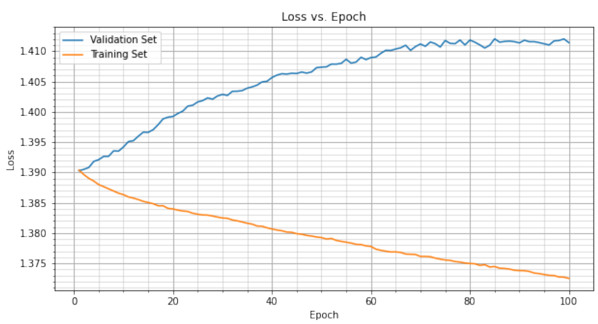

Figure 16: Categorical Cross-Entropy Loss vs. Epoch of seasonal neural network classifier

loss of the seasonal neural network tends to increase, and then plateaus around 75 epochs. Whereas the training set's categorical cross-entropy tends to decrease with epoch.

The graph above shows the categorical accuracy of the spatial neu-

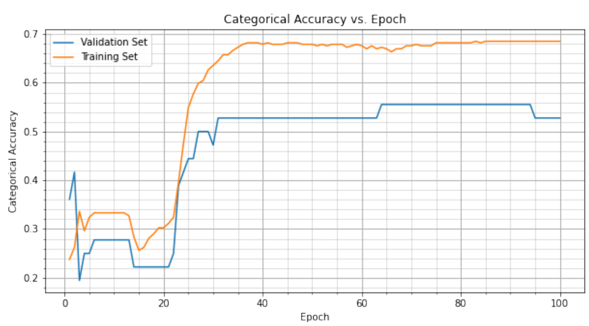

Figure 17: Categorical Accuracy vs. Epoch of spatial neural network classifier

ral network classifier with each epoch. In the plot, the categorical accuracy of both the validation set and the training set increase rapidly at around epoch 25. For the remaining epochs, the categorical accuracy remain relatively constant for both sets.

The graph above shows the categorical cross-entropy loss of the spatial neural network classifier with each epoch. The loss of both sets decrease with each epoch.

Comparing Figure 17 and Figure 18, we notice that in Figure 18 the model is learning after each epoch, since the categorical cross-entropy losses of both sets are decreasing. However, in Figure 18, the categorical accuracy of both sets seems to be staying constant after a few epochs. We suspected that the model may be memorizing from the dataset rather than finding useful patterns that can be applied to future representative datasets. However, we now believe that this pattern is the result of how the categorical accuracy

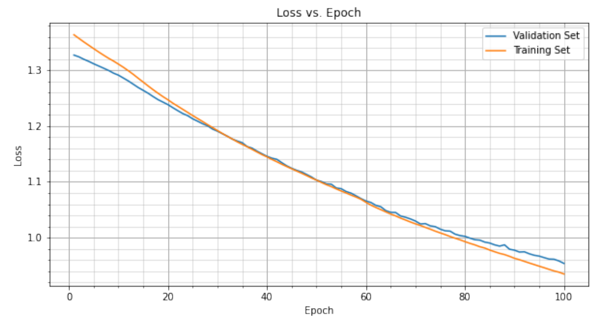

Figure 18: Categorical Cross-Entropy Loss vs. Epoch of spatial neural network classifier

function is defined relative to the categorical cross-entropy loss function.

5 DISCUSSION

5.1 Milestone Review

Throughout the course of the project we set several different milestones for ourselves in order to keep the project on track to be completed on time.

5.1.1 Data Importing, Separating and Cleaning. The first significant milestone we set was in mid October and it was to have all the data imported, separated into different dataframes and cleaned. By this milestone, we downloaded all of our data, off of the website for the Colorado Department of Revenue, in the form of excel sheets and imported them into the project and into dataframes using the pandas library. After all the data had been imported we cleaned the data by filtering out NULL and 0 values and by slicing out the parts that we did not deem usable.

5.1.2 Initial Data Exploration. The second milestone that we set was towards the end of October and it was to have explored the data and look for interesting trends, patterns or attributes that we wanted to have analyzed throughout all of the data. When the milestone approached we had decided to mostly focus on AGP (Adjusted Gross Proceeds) and coins in as the main attributes. It was at this milestone too that we made the decision to cut out the data from March 2020 forward because of COVID.

5.1.3 Data Clustering. The third milestone we set was at the beginning of November about one week after the second milestone. By this time we wanted to have the data clustered. This was completed and figure 13 and figure 14 show the clusters that were created.

5.1.4 All Major Analysis. The fourth and final milestone that we set was right at the beginning of December and it was to have performed all major analysis for the data. By this point we had completed all of the analysis that we wanted to do including cluster analysis, location-based analysis, seasonal analysis and regression analysis. From this point forward the only additional analysis of the data was through a neural network, a supervised learning model.

5.2 Choices

In any determining there are always choices that we make through out the process that can have major impact on the outcome. For the sake of transparency we want to walk through some of the decisions we made, why we made them, and the impact they could have had on the project. An example of this is how we decided to cut out data after march 2020 to mitigate Covid-19 restrictions from skewing our results. When reflecting on this paper we realized that a lot of our analysis revolves around AGP at the state wide level. However, we also concluded that this makes sense for two reasons. First off since we were trying to draw real conclusions that would be meaningful for people involved in the casino industry focusing on showing trends regarding profits is ideal and this is exactly what AGP shows us. Additionally we learned that all of these casinos are very close to each other, almost walking distance, so looking at each location individually gives redundant result so we focused on the data at a state wide level. Another important choice was the type of supervised learning method we chose. We chose to use a neural network because one member on our team had some prior experience with it while another was really interested in learning about it. It may not have been the best choice for the problem we were investigating, however it was the best choice regarding what we wanted to learn through completing this project. There are countless other choice that we made along the way, but these are just some examples to show our though t process behind each decision.

5.3 Lessons Learned

This project has been a big learning experience for everyone on the team. We all came in with different levels of coding and data science experience but all came out of it a little bit better in both categories. One of the biggest lessons we learned is that it is never worth it to save time at the start by not properly cleaning the data. It will always take more time later in the project because you will start to run into errors caused by problems in the data set. So, always make sure the data is clean and tidy before attempting analysis. Another important lesson we learned is that is is important to consider the background information and outside factors that may effect your data. The perfect example of this is how we anticipated using the different locations a major area to explore. However, in reality these casinos are very close to each other so looking at the different locations is arbitrary to drawing the big picture conclusion we were aiming for.

5.4 Future Work

Given more time or extra resources there is plenty of information that can still be mined from this data. On interesting analysis that could be performed from here would be to revisit the idea of seasonal trend, but instead look at annul trends. From the previous bar chat (Figure 11) We saw that there appears to be a trend across the different months. We could graph each month's AGP across the 10 years in a scatter plot. From here we could see if a linear regression model fits the data and if it suggest and annual cycle of profit flow. Hopefully this would result in a more interesting model than the previous linear regression we already implemented. Other supervised learning methods could have been applied to more attributes in the data sets to find a model that captures the data better. Similarly, if we used relevance analysis, we could have determined which attributes we wanted to focus on in each analysis method we performed. We also could have explored other variations of our neural network model, as well as performed hyper-parameter optimization to optimize our model.

CONCLUSION

We looked at the past 10 years of tax review provided by the Colorado Department of Review with the goal of finding interesting trends that could lead to useful analysis and conclusions for several different groups. We identified four major groups we hoped to draw relevant conclusions for: casino owners, locals, policy makers, and visitors. In order to get to the point where we could make conclusions we first needed to import, clean, and tidy all the data. From there we began visualizing our data to find any trends that stood out to us. Lastly, we implemented several different analysis techniques such as regression, clustering, and supervised learning in the form of a neural network. While not all of these analysis were useful or relevant to answering questions regarding our target audience, they all helped compete our other goal of increasing our understanding of data mining though there implementation. There are several conclusions that we were able to draw from our findings. For casino owners, we would recommend that they continue to focus on penny slots as they have the highest payout of any slot machine and make up about 50% of the profits every year. Additionally we would recommend they invest more into multi-denomination slots. We saw from figure 4 that these multi-denomination slots have been increasing in popularity and profit making more than any other slot type. As for our findings for locals, this is the area where we draw the least concrete conclusions. Our one recommendation would be to understand the annual flow of popularity across the different months because we suspect this must correlate with the amount of tourist visiting the area. More visitors means more money moving around the local economy and understanding this cycle can help small business owners capitalize on the crowds. As for the policy makers, the major findings that we found policy makers must already be aware of because we found laws that directly address our conclusions. For example the laws around hold percent regarding slot machines. Our one recommendation would be to update these polices to account for the multi-denomination slots that are increasing in popularity. Lastly, the group that most interested us the visitors to the casinos. Our recommendations to some one looking for their best option when gambling at the casino is to go with unorthodox slot machine types. By this we mean the 10 or 50 cent machines. Historically, these machines have the lowest hold percent and thus pay out more often than other slot machines. We would recommend against playing penny slots, even though it feels like you are wasting less money each pull it's actually works because they have a higher hold percent. If slot machines don't interest you and you want to play a table game we would recommend Craps because historically it has the most consistently low hold percent amongst the table games. Despite these recommendations, after all the casino profit analysis we have done we know that these casinos make hundreds of millions of dollars a year, and the only way to ensure you win is not to play.

REFERENCES

[1] Anthony F Lucas, William T Dunn, Wesley S Roehl, and Guy M Wolcott. 2004. Evaluating slot machine performance: A performance-potential model. *International Journal of Hospitality Management* 23, 2 (2004), 103–121.

[2] AK Santra and C Josephine Christy. 2012. Genetic algorithm and confusion matrix for document clustering. *International Journal of Computer Science Issues (IJCSI)* 9, 1 (2012), 322.

[3] Richard Thalheimer and Mukhtar M Ali. 2003. The demand for casino gaming. *Applied Economics* 35, 8 (2003), 907–918.

[4] Richard Thalheimer and Mukhtar M Ali. 2008. Table games, slot machines and casino revenue. *Applied Economics* 40, 18 (2008), 2395–2404.

[5] Mahendra Tiwari and Randhir Singh. 2012. Comparative investigation of k-means and k-medoid algorithm on iris data. *International Journal of Engineering Research and Development* 4, 8 (2012), 69–72.

A APPENDIX

A.1 Honor Code Pledge

We hereby pledge that we did not abstain from the honor code.

A.2 Team Member Contribution

A.2.1 Cameron Mattson. In this project, I cleaned and organized all of the data. This included concatenating all of the excel spreadsheets and creating the corresponding dataframes. I also created the regression, created the clusters, and created the neural network models. For each of these models I created plots. I explained the plots as well as the models. I contributed to the evaluation section, the tools section, the discussion section, the related work section, the abstract section, and the introduction section to name a few. I contributed about 40-50 % to this project.

A.2.2 Thomas Neal. I started this project by suggesting we analyze casino data and finding the data set that we ultimately choose to use. In this report I am responsible for writing the following sections: 1 (Introduction), 3.1 (initially proposed work), 3.2 (about the data), 4.1 (data visualisation and trend hunting), 4.3 (regression analysis), 5.2 (choices), 5.3 (lessons learned), 5.4 (future work), and conclusion. I also created every graph within these sections (with the exception of the regression, that was made by Cameron). Overall, I would say I did 45-50% of the code between helping to clean, tidy, summarise and visualize the data. I was also a major editor on the final report and all prior reports. Lastly, I helped to create the presentations by building a majority of slides and structuring so it flowed well.

A.2.3 Hayes Vavpetic. In this report I am responsible for writing sections: 3.3 (Tools) and 3.4 (Changing the data). I helped create all the presentations as well as write the reports. I also assisted in cleaning the data and some of the code.

A.2.4 Michael Grodecki. In this project I am responsible for generating and editing plots, including the AGP per slot per casino graphs and analysis. I helped with proofreading / editing all of the presentations and reports as well as contributing to the final analysis.

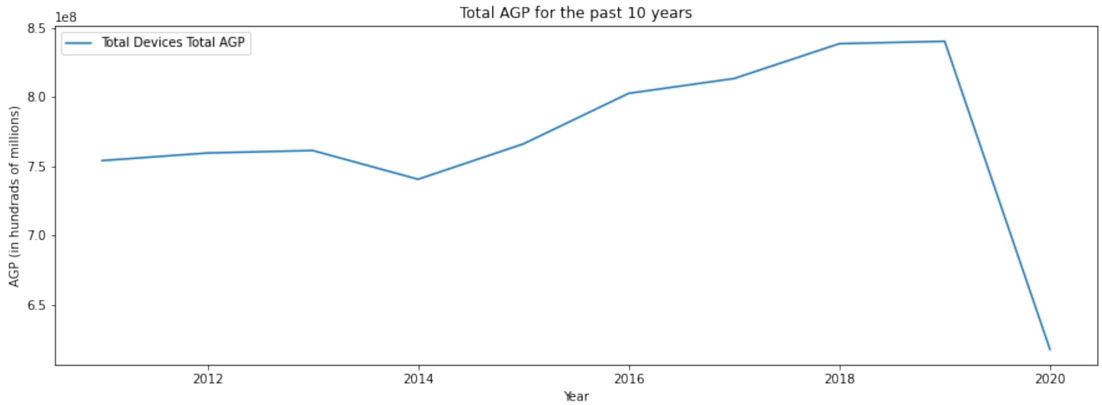

Figure A2: Profit produced state wide for the past decade

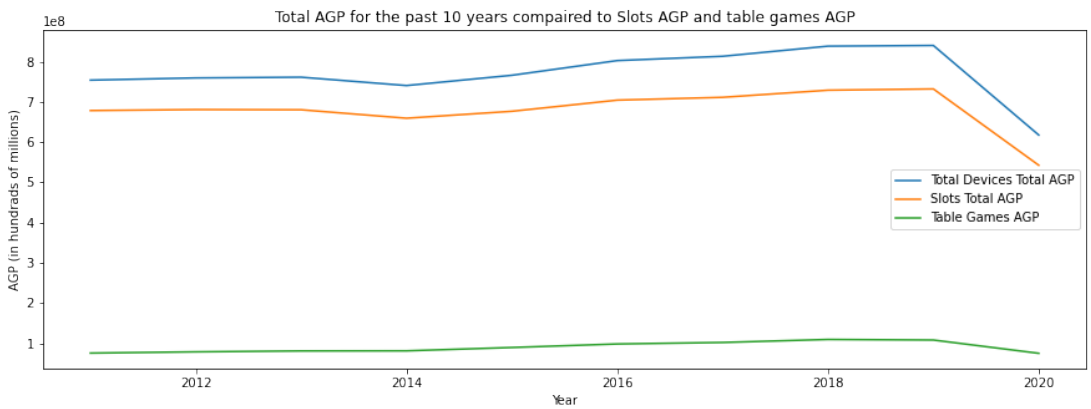

Figure A3: State wide profit totals for the different types of devices

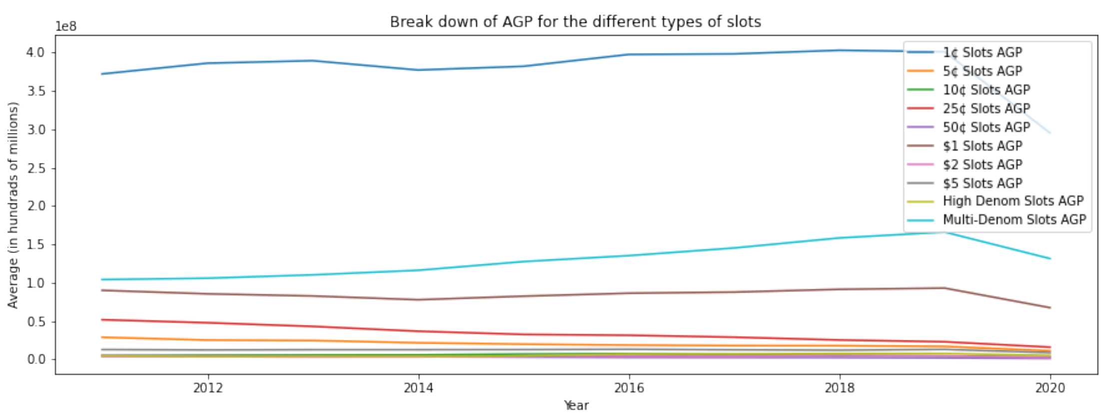

Figure A4: State wide profit totals for the different types of slot machines

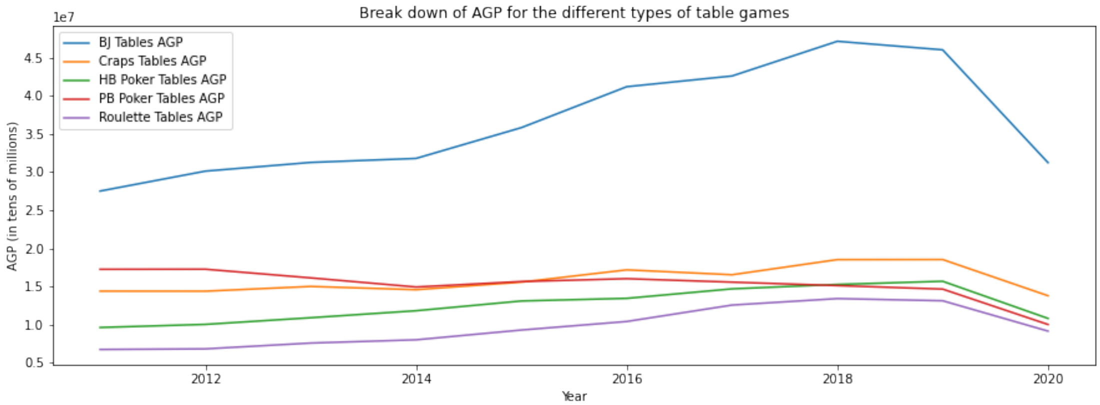

Figure A5: State wide profit totals for the different types of table games

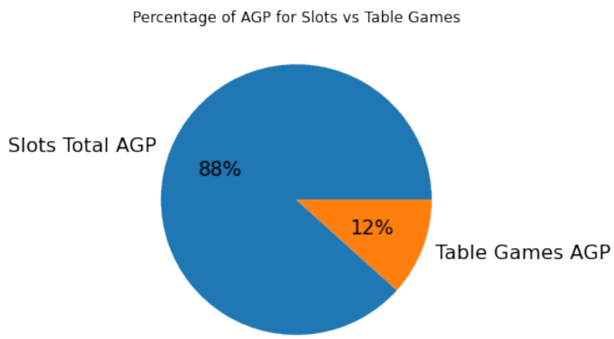

Figure A6: Ten year averaged AGP split by proportion

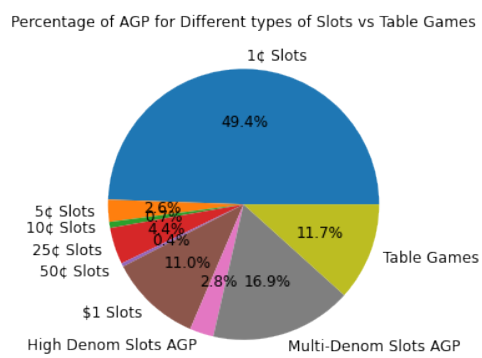

Figure A7: Ten year averaged AGP split by proportion

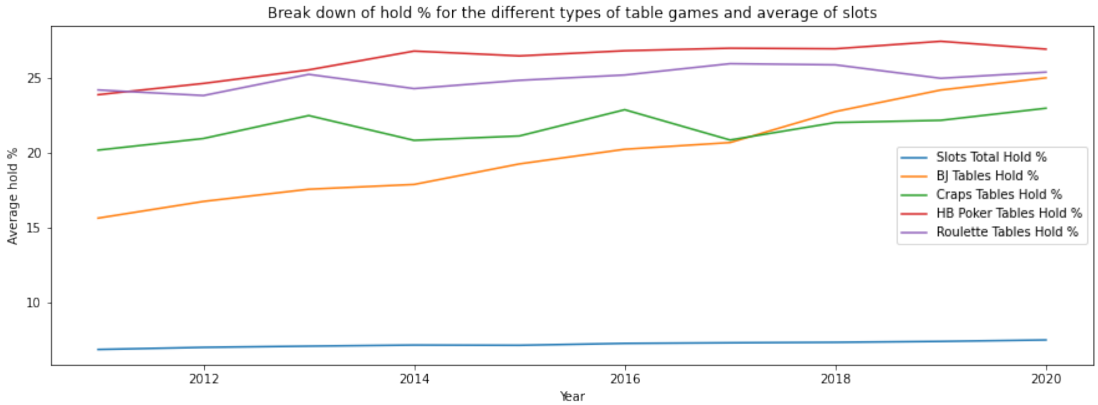

Figure A8: State wide hold percent for different table games and average of slots

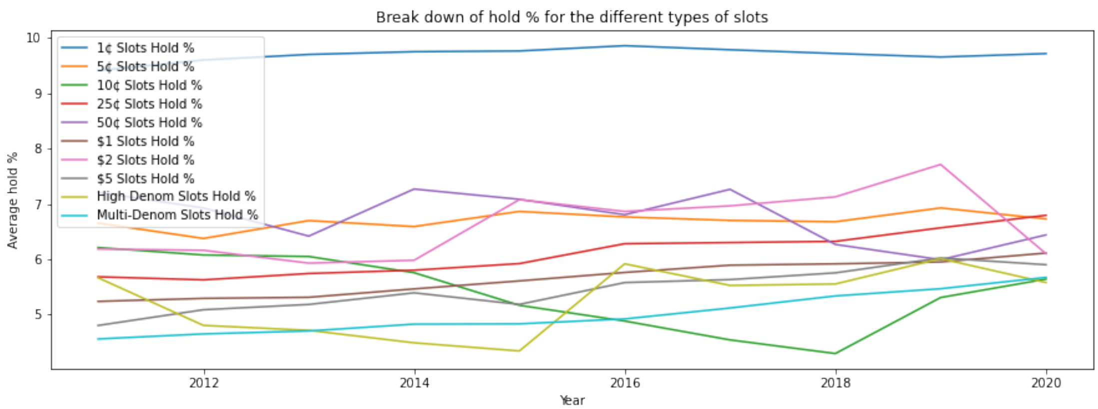

Figure A9: State wide hold percent for different slot machine types

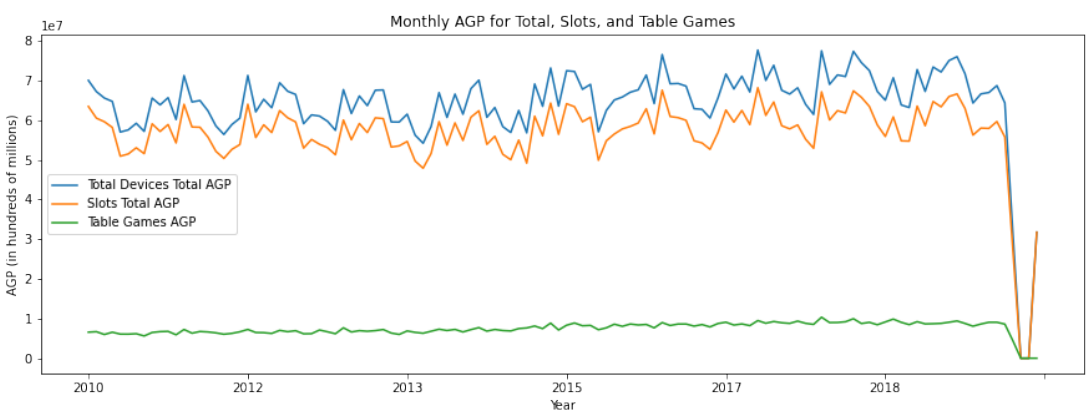

Figure A10: State wide monthly AGP

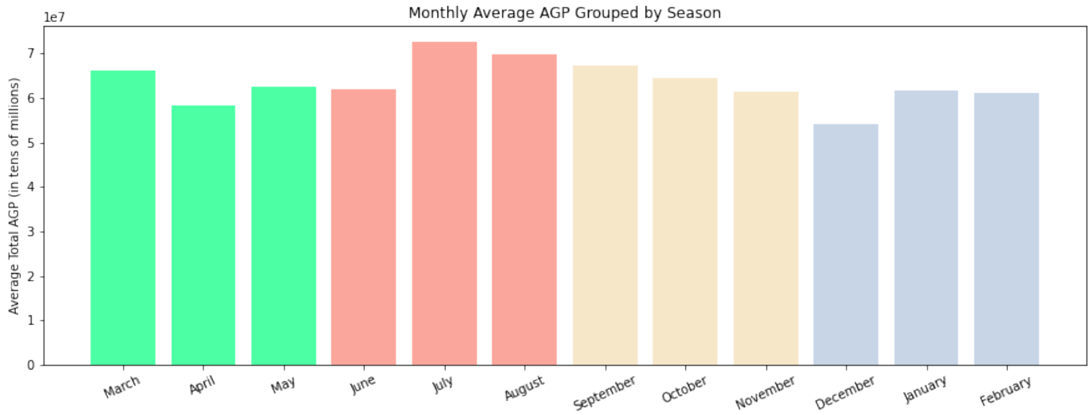

Figure A11: Ten year AGP averaged per month and color coded by season

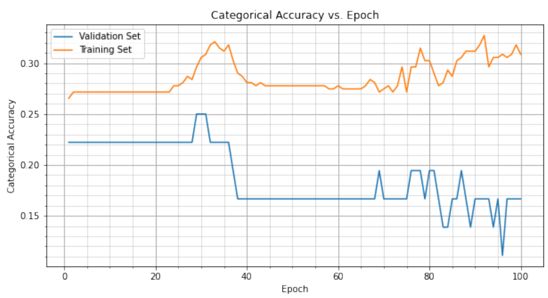

Figure A15: Categorical Accuracy vs. Epoch of seasonal neural network classifier

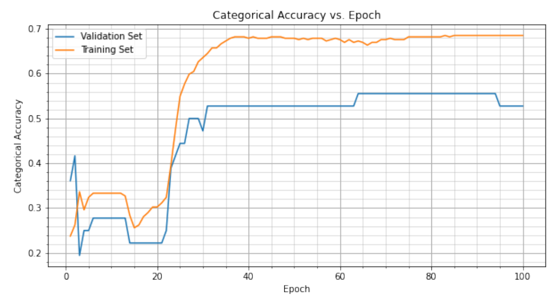

Figure A17: Categorical Accuracy vs. Epoch of spatial neural network classifier

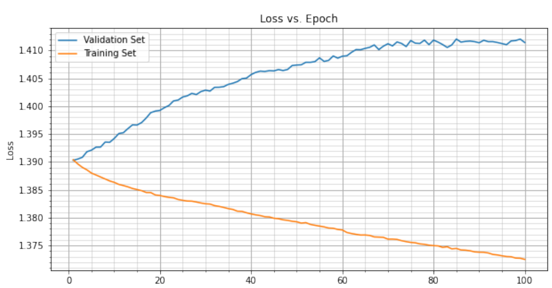

Figure A16: Categorical Cross-Entropy Loss vs. Epoch of seasonal neural network classifier

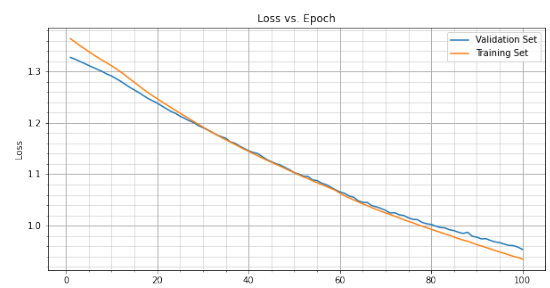

Figure A18: Categorical Cross-Entropy Loss vs. Epoch of spatial neural network classifier