Iowa's Coronavirus Relief Expenditures

Final Project Report

Zach McDaniel, Brody Adreon, Clint Krapfl, Yanni Mihalopoulos

Contents:

Executive Summary

 Using data provided by the state of Iowa that reports pandemic recovery state expenditures, this project identifies which state departments require Coronavirus relief funds in order to continue to operate effectively. The logistic regression model we created, using rattle, aims to predict the appropriate amount of funds allocated to each government entity.

Overview

The year 2020 has brought many challenges to our everyday lives, the most prominent of these would be the Coronavirus pandemic. Federal and State governments have been troubled with their ability to properly allocate funds where they are needed. This project will focus mainly on Iowa state legislature, and its fiscal plan on funding certain industries. Expenditures on pandemic relief in the United States has reached over $2.6T (DataLabs) and $1.25B in Iowa (Richardson). This gives better insight on how the federal government is dividing up money to state governments, and furthermore, how each state government chooses where aid relief funding is placed. Figure 1 below shows current active cases as reported by Johns Hopkins University & Medicine as of (12/2/2020). The size of each circle indicates a larger portion of population who currently has COVID-19. From this heat map we can say that countries filled with large circles will require more funding and spending to treat and help those who are sick with the coronavirus. This heat map gives data scientists who work in the government, a better representation of where funding must be given to have the most impact.

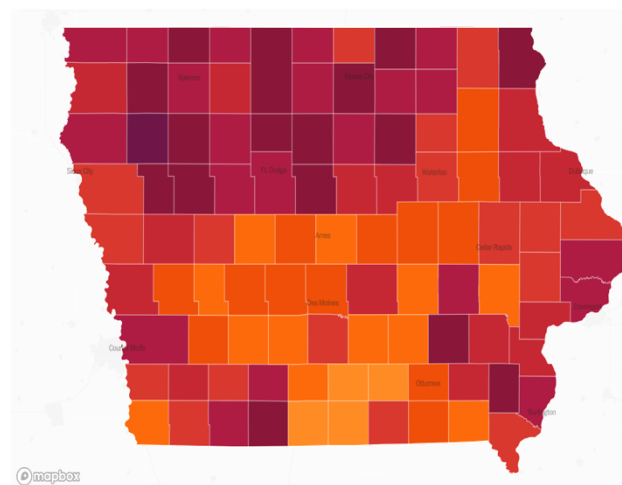*Figure 2. Iowa Covid Heat Map*   *Figure 1. Global Covid Heat Map*

Iowa government has many options as to where it decides pandemic relief will be situated. One thing that the governor looked to accomplish was compensating those out of a job due to the virus, with the federal

CARES act. Iowa government spent nearly $1B on providing financial relief for those out of a job (Clayworth). This is one of many examples of how local government can allocate money to help constituents during the pandemic. Figure 2 above shows Iowa coronavirus cases by county. Local government can take this info to better understand where to spend more money. This money will most likely be helping those who are unemployed, hospitals, and elderly homes in need of funds due to COVID related illness.

Business Goal

Iowa state government officials are interested in determining the appropriate amount and to which departments they should allocate their Coronavirus reliefs funds to stimulate productivity in a deprived global economy.

Potential clients are the Iowa state departments that are making the fund allocation decisions such as the Iowa Department of Agriculture & Land Stewardship, the Department of Human Services, the Workforce Development, and many more. This project will allow the state departments to operate more perceptively and equip these state departments with the information and insight they need to budget for these important financing contributions.

Data Mining Goal

The goal of this project is providing visual and descriptive information on how the state of Iowa allocates funds to certain departments across the state in the time of a pandemic. You will be able to see which departments use the most amount of money as well as the least. This will allow the reader to understand which departments required most of the given resources as well as creating a predictive model in case of future needs. Readers will understand which department(s) are the most critical in terms of pandemic relief.

Dataset Overview

The data we used was collected from iowadata.gov and pertains to all state government COVID-19 expenditures and contains granular data on each record including department, agency, unit, fund and more. Our dataset, "expense_data.csv", contains 4,982 instances, including 7 categorical variables, 2 numeric features, and 2 time/date features. The target variable is "amount" which is the dollar amount of each request. The highest amount is 294,000,000 and the lowest amount is -23,766,180.30.

Data Preparation

To prepare the data before analysis, we loaded it into RStudio and removed several features that did not add to our understanding of the data. The features removed were Record ID, State Department Number, Unit, Letter of Credit, Fund Code, and Object Class. After removing features, we coerced all features to their proper data types. Finally, we filtered the data to only contain cases from March through November,

since March was when cases started appearing in the United States. After the data was cleaned, it was exported as .csv file to be loaded into rattle.

Exploratory Analysis

For our exploratory analysis, we decided to investigate the number of requests per federal department and the spending per department. Our data was loaded into RStudio where we created a function to search the data and summary tables to understand it better. The function, called timeline, allowed us to search between a start data and end date to produce a summary table showing spending by federal department in the timeline. We also created a summary table that showed the frequency of requests by departments to understand who requests funds most frequently? Finally, scatter plots showing spending by department over time were created in r.

The diagram below represents the total amount of requests made by department, represented by a pie chart. The percent represents the total percent of fund requests made by each department. We chose to present this data to show the number of requests and where the requests are coming from. This gives data scientists the ability to interpret which departments have a higher volume of needs.
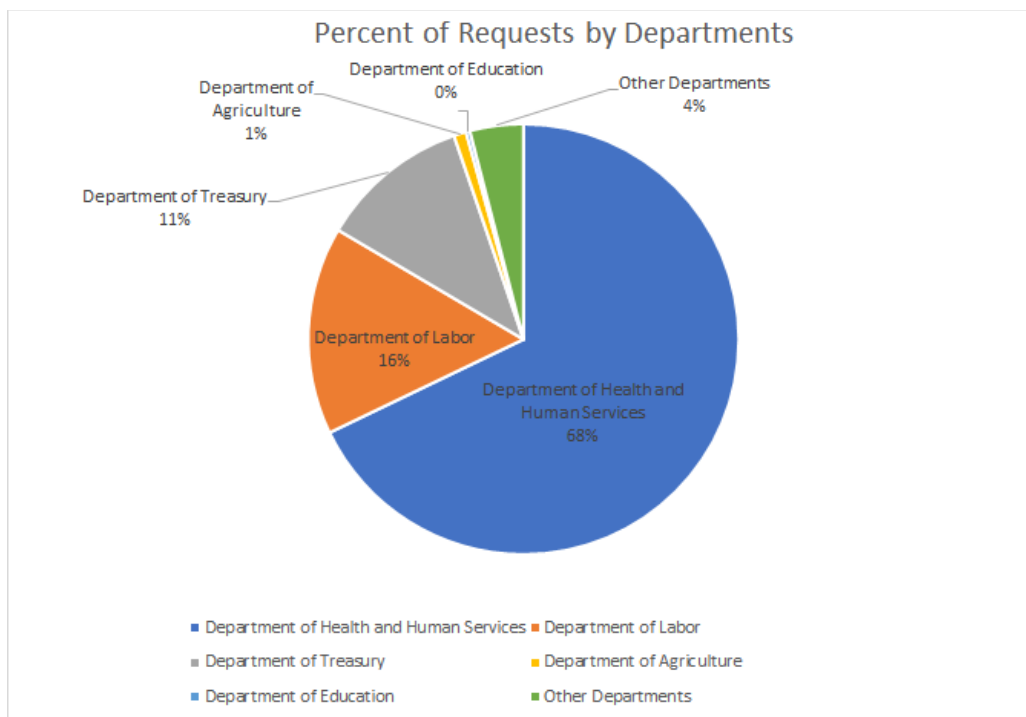

Other Departments includes the departments of transportation, election assistance commission, justice, humanities, museum and library, and homeland security whose expenses were not significant on their own relative to the other departments.

The figure below shows spending in Billions by the Federal Department in Iowa. We can see that the significant portion of spending falls within 3 major departments: Labor, Treasury, and Health and Human Services. The pie chart above showed the requests for funding, which leads us to the conclusion that the volume of requests is not correlated to the amount of funding. This can lead to conclusions that the Labor and Treasury Departments ask for money less frequently but receive much large lump-sums of money compared to the other departments.
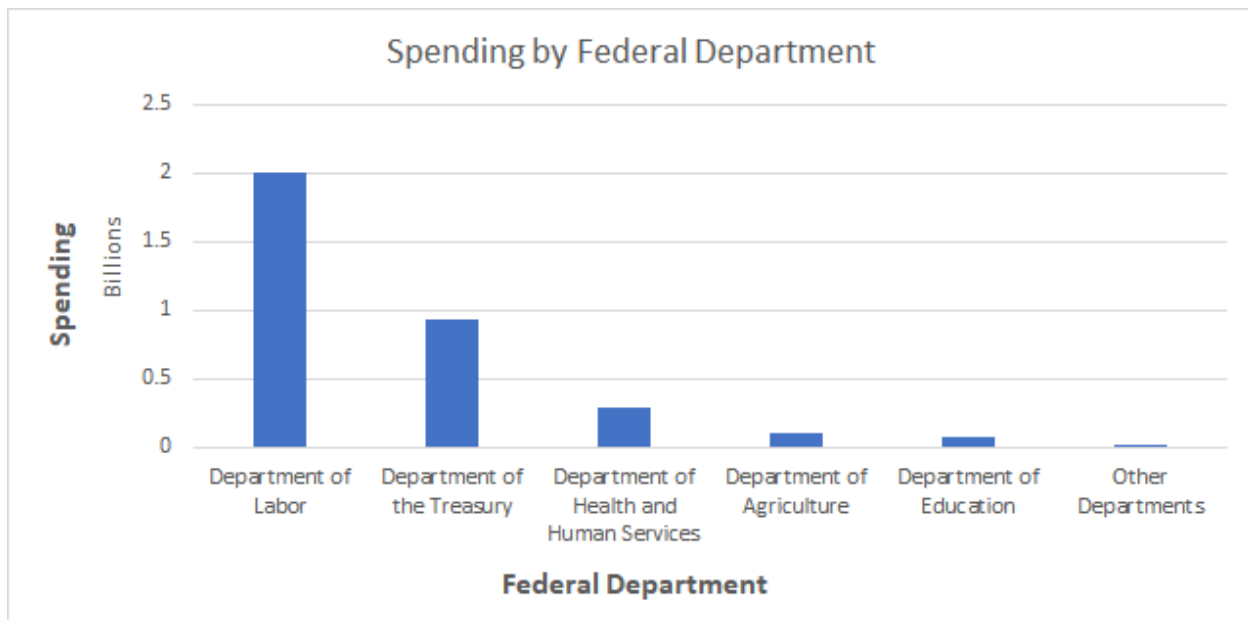
In the scatter plot below, we can better see how expenses were distributed from March through the end of November. Most lower cost expenses during the beginning of COVID-19 were from the Department of Health and Human Services, which entailed CDC, Medicare & Medicaid, as well as Administration for Children and Families. Around May, the Department of Labor and Department of Treasury come in when the government started distributing stimulus checks and supporting businesses from foreclosures. This graph only contains cases under one million dollars because it gave us a better look of the trends by the major departments.
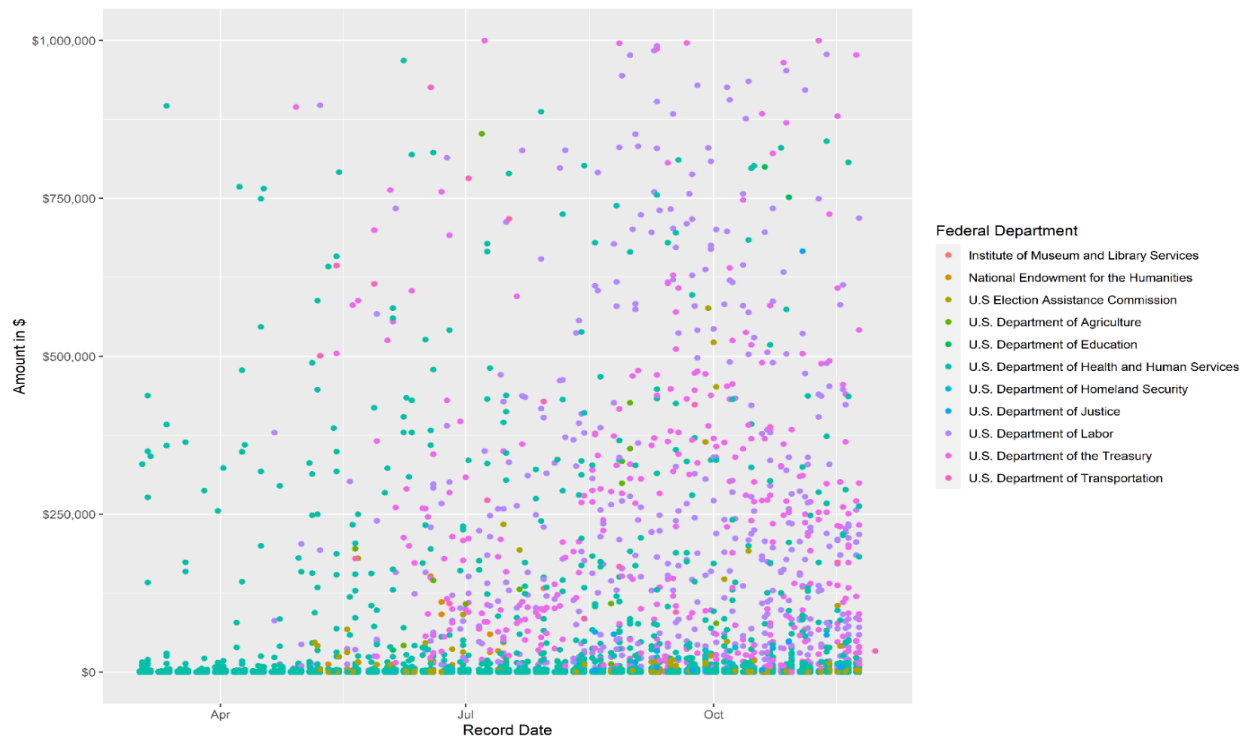
Key Findings:

During our exploratory analysis, we discovered several things about how spending was distributed among the departments, as well as how spending changed and increased over time. The three main departments that were significant were the departments of Health and Human services, Labor, and Treasury. Specifically, we found that the Department of Health and Human Services makes up almost 70% of total requests yet only 8.4% of total spending went to them. Incidentally, we discovered that the Department of Labor makes up 16% of the requests but received 59% of the spending with the Department of Treasury also receiving 27% of spending. Around the month of May, when people realized that COVID-19 was not going away, the government was required to increase spending for labor and treasury as workers needed stimulus to pay bills and businesses needed coverage to avoid foreclosure and evictions.

Model

When building our model, the intention was to create a predictor of amount by using department and agency info as the variables.

Linear Regression

The table below shows the linear regression output for predicting amount using the Federal Department and Fund as the input variables. To build this model, we identified department and fund as key variables resulting in the greatest account for variation in our results, which led us to ignoring all other features and recoding these two features as indicator variables.

This regression output re-affirms our discoveries in the explanatory analysis, showing the departments of Labor, Treasury, Agriculture, and Education as having high coefficients because those departments showed higher average amount requests than the other departments. After, trying varying combination of inputs for the regression, we found that department and fund resulted in the largest impact on predicting amount with the leas sized errors.

Example Linear Regression Output:

```
Coefficients: (6 not defined because of singularities)
                                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                    93061    2914243   0.032   0.9745
TIN_Federal.Department_National.Endowment.for.the.Humanities  510646    3242410   0.157   0.8749
TIN_Federal.Department_U.S.Election.Assistance.Commission     467749    3378458   0.138   0.8899
TIN_Federal.Department_U.S..Department.of.Agriculture        2693660    3149392   0.855   0.3924
TIN_Federal.Department_U.S..Department.of.Education          6912689    3322749   2.080   0.0376 *
TIN_Federal.Department_U.S..Department.of.Health.and.Human.Services  572214  3018365  0.190  0.8497
TIN_Federal.Department_U.S..Department.of.Homeland.Security   -87788    3125178  -0.028   0.9776
TIN_Federal.Department_U.S..Department.of.Justice             -17846    3258223  -0.005   0.9956
TIN_Federal.Department_U.S..Department.of.Labor                -2760    2969758  -0.001   0.9993
TIN_Federal.Department_U.S..Department.of.the.Treasury       629771    2977834   0.211   0.8325
TIN_Federal.Department_U.S..Department.of.Transportation     499124    3192394   0.156   0.8758
TIN_Fund_Biofuels.Relief.Fund                                985677    2149672   0.459   0.6466
TIN_Fund_Byrne.JAG                                               NA         NA      NA       NA
TIN_Fund_CARES.Act.Funding...DHS                            1310391    1217556   1.076   0.2819
TIN_Fund_Court.Tech...Modernization                         -719487    1640398  -0.439   0.6610
TIN_Fund_Csbg.Comm.Action.Agencies                          -547938    1313571  -0.417   0.6766
TIN_Fund_Disaster.Unemployment                              3966650     624395   6.353 2.39e-10 ***
TIN_Fund_Emergency.Mgt.Performance.Grt                           NA         NA      NA       NA
TIN_Fund_Eviction.and.Foreclosure                           -496026     809114  -0.613   0.5399
TIN_Fund_Food.Assistance..Temp.Emer                        -1837081    3055272  -0.601   0.5477
TIN_Fund_General.Fund                                       -560762     777779  -0.721   0.4710
TIN_Fund_HAVA.CARES                                         -508484    1818571  -0.280   0.7798
TIN_Fund_Hawk.I.Trust.Fund                                  -644739     927558  -0.695   0.4870
TIN_Fund_Idph.Gifts...Grants.Fund                           -594855     867069  -0.686   0.4927
TIN_Fund_Innovation...Commercialization.Development.Fund    1277169    5084597   0.251   0.8017
TIN_Fund_Iowa.Beginning.Farmer.Debt.Relief.Fund             334294    1886661   0.177   0.8594
TIN_Fund_Iowa.Farms..Schools.and.Communities.Broadband.Grant.Fund  -249792  2977834  -0.084  0.9332
TIN_Fund_Iowa.Hospital.COVID.19.Relief                    24277169    5084597   4.775 1.88e-06 ***
TIN_Fund_Iowa.Livestock.Producer.s.Relief.Fund             -405054    2003612  -0.202   0.8398
TIN_Fund_Iowa.Refugee.Service.Center                       -664536    5108440  -0.130   0.8965
TIN_Fund_IWD.Major.Federal.Programs                         -73518     813566  -0.090   0.9280
TIN_Fund_Library.Services...Technology                          NA         NA      NA       NA
TIN_Fund_Local.Government.Relief.Payments                  6531403    2149672   3.038   0.0024 **
TIN_Fund_Low.Income.Energy.Assistance                      -492577    2063368  -0.239   0.8113
TIN_Fund_Mine.And.Health.Safety                                 NA         NA      NA       NA
TIN_Fund_Movie.Theater.Relief.Fund                        3917169    5084597   0.770   0.4411
TIN_Fund_Non.Profit.Relief.Fund                            290591    2003612   0.145   0.8847
TIN_Fund_Office.of.Chief.Information.Officer               -205379    1709540  -0.120   0.9044
TIN_Fund_Public.Transit.Assistance                             NA         NA      NA       NA
TIN_Fund_Skilled.Workforce.Last.Dollar.Scholarship.Program.Fund  273040  2338883   0.117  0.9071
```

Methods & Evaluations

The following table documents the different methods that we used to predict amount, as well as the parameters for each model. For evaluating the model, we decided to use MAE, MSE, and R-Squared because we need to predict numerical outputs, and this allowed us to see the errors for each model. When looking at the evaluations, it is important to understand that the data set contains cases with dollar amounts ranging from negative numbers all the way up to $300,000,000, which resulted in us finding large values in our evaluation. We could not consider the high amount cases as outliers and they were legitimate cases and help us understand some of the biggest COVID-19 expenditures in 2020.

Linear Regression Models Tested: seed = 50

Linear Regression Inputs	MAE | MSE	R-Squared
<u>Department & Fund</u>	<u>MAE: 658654.7 | MSE: 8.95e+12</u>	<u>.585</u>
Department, agency, fund, & unit	MAE: 745278.117 | MSE: 8.99e+12	.58
All Features (- Unit Name)	MAE: 756592.4 | MSE: 9.03e+12	.58

We first started out by including all features of the dataset, besides Unit Name, which was not included due to it many levels compared to the rest of the features. In the process of finding our ideal model, we tested out several regressions using different amounts of input variables to determine which should be used in the model. In the end, we found that Federal Department and Fund were the most important indicator variables when predicting the dollar amount of an expense and resulted in smaller overall residuals. Although we settled on linear regression, several other methods were tested to select our model. The following table contains the evaluations of other methods we tried.

Other Methods Tested: seed = 50

Artificial Neural Network (ANN)	MAE | MSE
Nodes = 12	MAE: 757805 | MSE: 9.03e+12
Decision Tree	**MAE | MSE**
C = .0031	MAE: 712115.5 | MSE: 8.897e+12
Random Forest	**MAE | MSE**
Trees: 18, Variables: 13	MAE: 680283 | MSE: 8.93e+12

All models were built using optimal parameters selected during the modeling process, which resulted in the least sized residuals for each model.

We found that the random forest model to be similarly effective to our linear regression model, yet the regression still narrowly beat out the random forest with lesser residuals. In the end, our selected model

using department & fund as inputs had a mean absolute error of $658,654 and accounted for almost 59% of the variation in amount.

Some potential issues with the evaluation metrics are the general large amounts and several significantly large amount expenses. The nature of expenses for the government results in many similar groupings of expenses depending on current events, which overall makes the prediction of these numbers have a wider error range.

Summary

We discovered the key factors of Iowa government COVID-19 spending and how they changed over time. Specifically, we learned that the department of labor had the largest spending with 58% followed by Treasury with 27% and Health and Human Services with 8%. On the other hand, Health and Human Services was 68% of requests, with Labor and Treasury at 16% and 11%. Through our regression, we found the key funds that required the largest spending were hospital spending, unemployment, workforce, and relief funds. We also discovered a trend over time as Health and Human Services requests dominated the beginning of COVID-19 and around May the department of labor and treasury began spending large amounts as stimulus and eviction and foreclosure prevention plans were placed. Our analysis provides deeper understanding into the needs of the state of Iowa and the community's need trended over time. This insight will be very valuable moving forward as the government continues to plan its financing to the best of their ability.

Next Steps:

Use of model

This model can be used in a predictive analysis using projected federal department requests to calculate a large-scale estimation of total spending during a given time. This will allow the government to prepare budgets for future pandemic related crises.

Improvements to model

Data Quality: We used an 8-month time of data. As time progresses more long-term trends could develop.

References:

Clayworth, J. (2020, October 28). Iowa says alleged misallocation of $21 million in federal pandemic relief funds was permitted judgment call. Retrieved December 14, 2020, from https://www.desmoinesregister.com/story/news/2020/10/27/iowa-denies-misspent-21-million-coronavirus-relief-funds/3750251001/

COVID-19 Map. (n.d.). Retrieved December 07, 2020, from https://coronavirus.jhu.edu/map.html

Richardson, I. (2020, September 07). Iowa has received billions in federal COVID-19 funds: Here's where it's going, and who's crying foul. Retrieved December 07, 2020, from https://www.desmoinesregister.com/story/news/politics/2020/09/07/iowa-has-received-billions-covid-19-funding-heres-where-its-going/3402108001/

Robertson, I. (n.d.). Data Lab – The Federal Response to COVID-19 – U.S. Treasury. Retrieved December 07, 2020, from https://datalab.usaspending.gov/federal-covid-funding/