Federal Spending & The Dept. Of Energy

Nicholas Staich

Agenda

Introduction

Data Information

Visulaizations

Model Description

Reccomendations

Topic One: Introduction

Introduction

- I wanted to understand the Federal Spending data better
- Focused on the DOE
- Spend in the DOE is an investment into future energy technologies
- Looking at the proportion of Spend in DOE can indicate rate of progress towards a more advanced national and global energy profile.

Data Set

ORIGIN:

- Data Source: DataUSA
- File: Data USA Cart.csv
- 5 initial features
- Observations: 6557

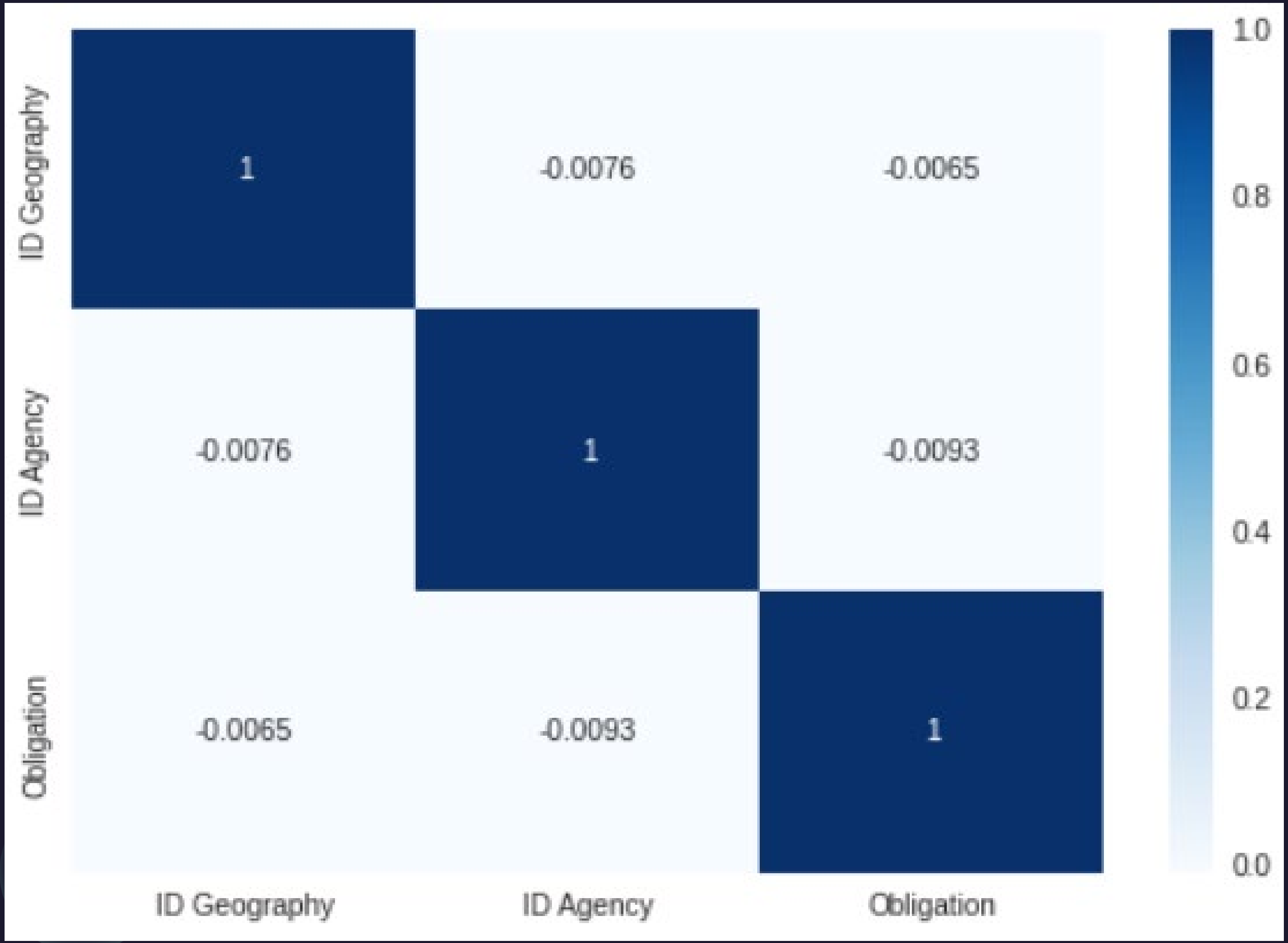
Initial Observations

- Total Gov't Spend (2017):
- $2,040,479,396,498.97
- US DOE Spend (2017)
- $27,513,206,449.87 (1.3%)
- Combined Military Spend (2017):
- $387,789,396,562 (19%)
- Defense Logistics Agency
- $27,253,153,800.05

FINAL DF.SHAPE

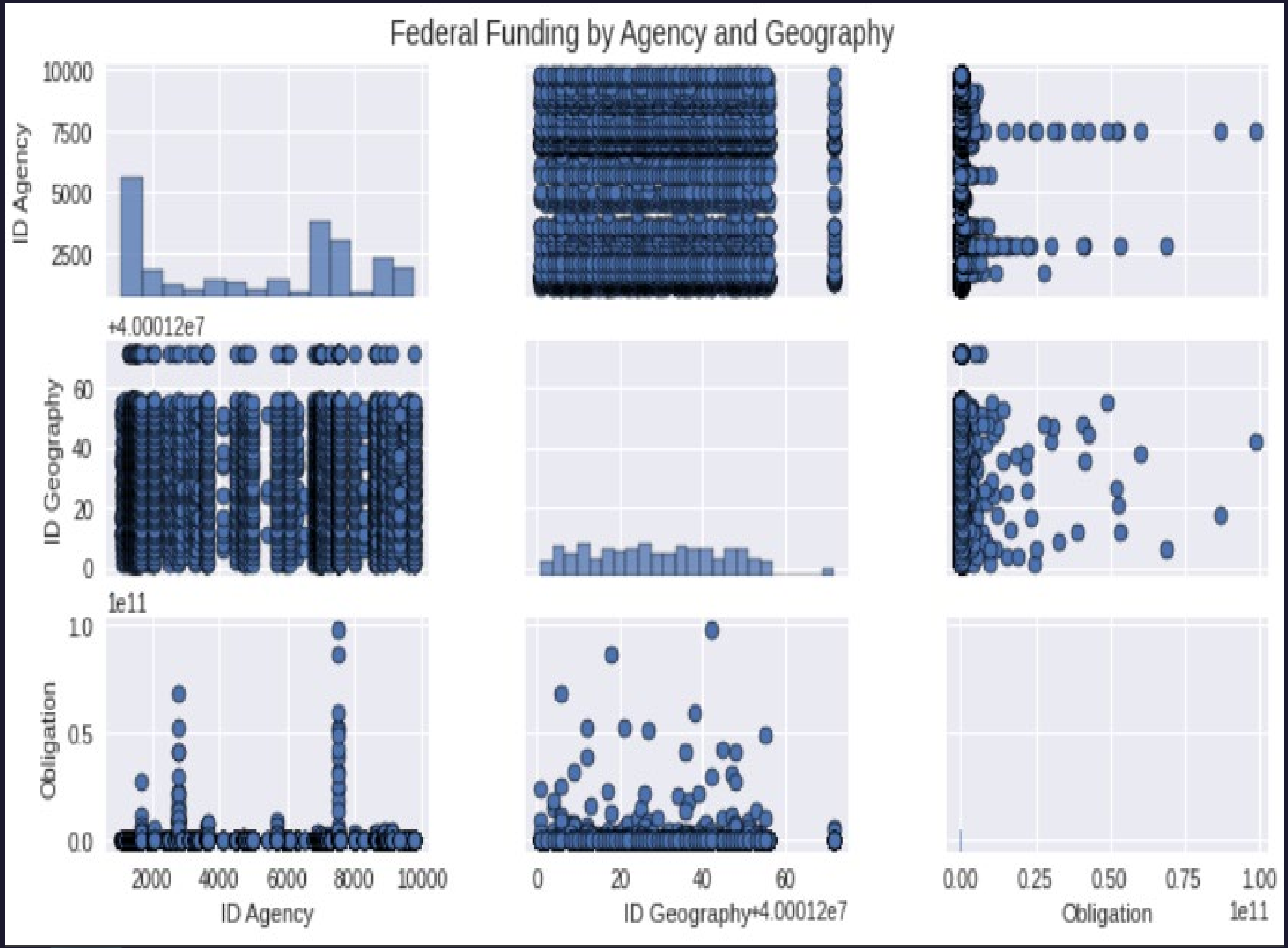
- (6557, 3) after EDA
- Main Feature Takeaways
  - 207 Agencies
  - 52 Geographic locations

Topic Three: Visualizations



Correlation Matrix

- Based on the Correlation Matrix, there is little correlation between the features.
- There still needs to be further analysis to see if there are any points worth investigation.



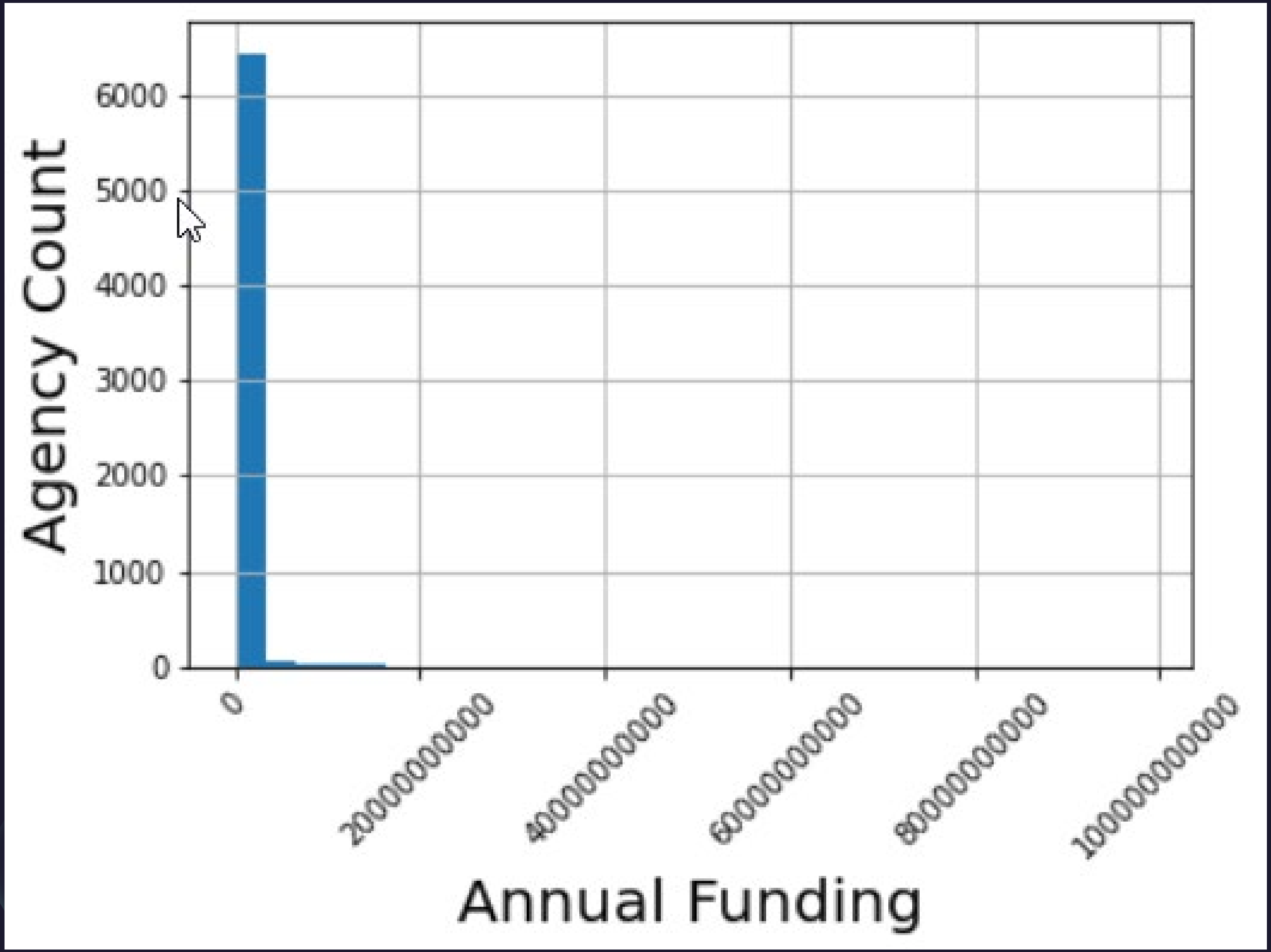
Interesting Distribution of Funds:

- Agencies around the 7500 level receive far more funding relative to other agencies
- Agencies around the 2500 level receive far more funding relative to other agencies

Topic Four: Model Description

```python

# create model architecture
input_dim = X_train_proc2.shape[1]

reg_model3 = Sequential()

reg_model3.add(Dense(251, input_dim=input_dim, activation='relu'))

#reg_model3.add(Dropout(.5))
kernel_regularizer=keras.regularizers.l2(0.001)
reg_model3.add(Dense(300, activation='sigmoid'))

early_stopping = EarlyStopping(patience = 5)
kernel_regularizer=keras.regularizers.l2(0.0001)
reg_model3.add(Dense(250, activation='sigmoid'))

reg_model3.add(Dropout(.5))
early_stopping = EarlyStopping(patience = 5)
kernel_regularizer=keras.regularizers.l2(0.0001)
reg_model3.add(Dense(200, activation='relu'))

reg_model3.add(Dropout(.5))
kernel_regularizer=keras.regularizers.l2(0.0001)
reg_model3.add(Dense(50, activation='relu'))

#reg_model3.add(Dropout(.5))
kernel_regularizer=keras.regularizers.l2(0.0001)
reg_model3.add(Dense(25, activation='relu'))

reg_model3.add(Dense(1, activation='linear'))

reg_model3.summary()
```

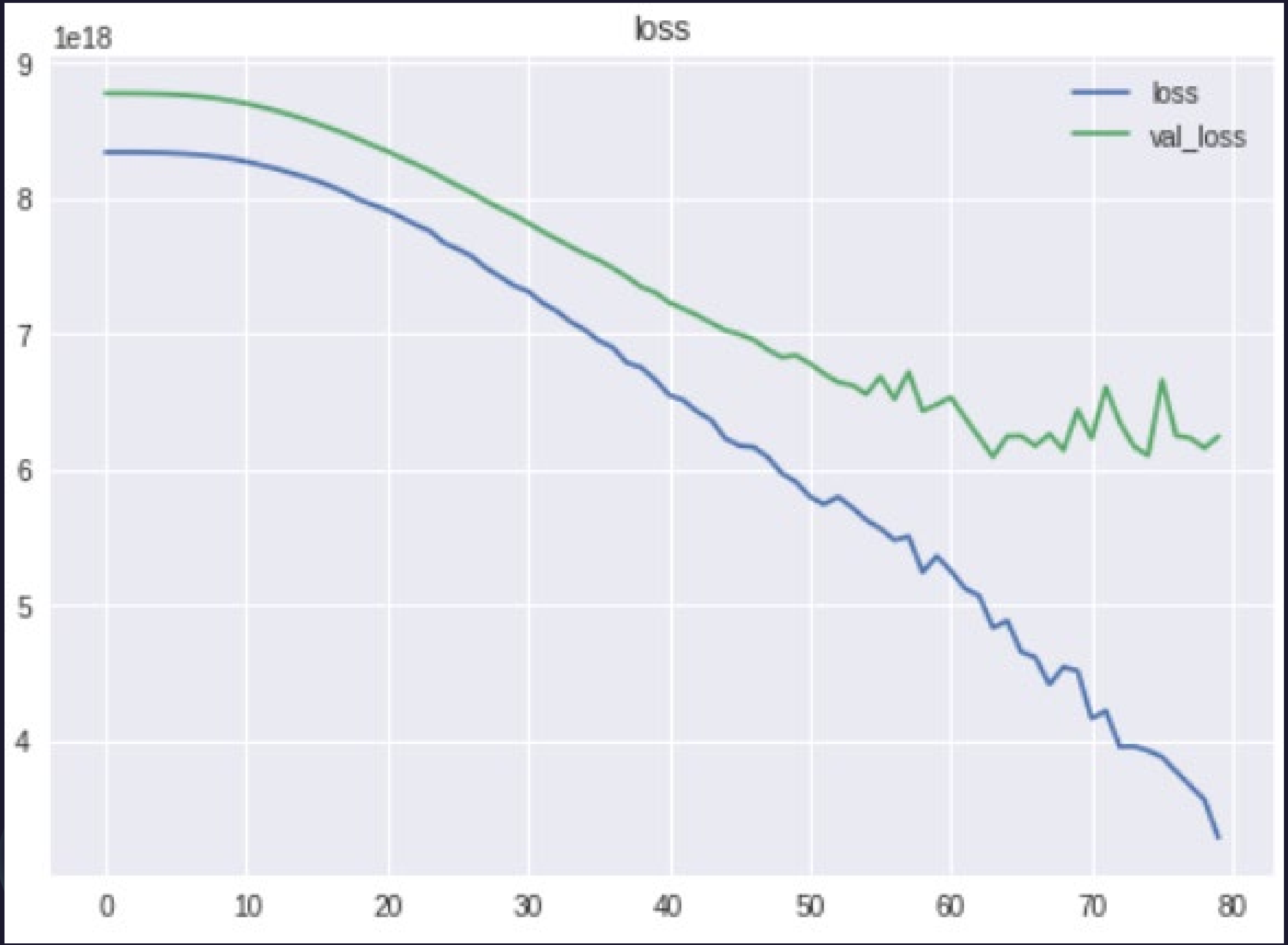
- Model Selected was Sequential()
- There is a relatively uniform gradient in the distribution of funds by geography, which one would expect.
- The histogram of obligations by Agency shows that most agencies receive very little funding
- 5 hidden layers
- Use of Sigmoid to offset overweight of positive deltas
- Used regularization to maintain smoothness
- Dropouts assisted with gradient



- Funding Breakdown
- This chart shows something amazing:
- The disparity between agencies and funding levels
- Almost 6000 Federal agency offices receive low to no annual funding.
- A small, non-visible minority took huge proportions of the available budget.

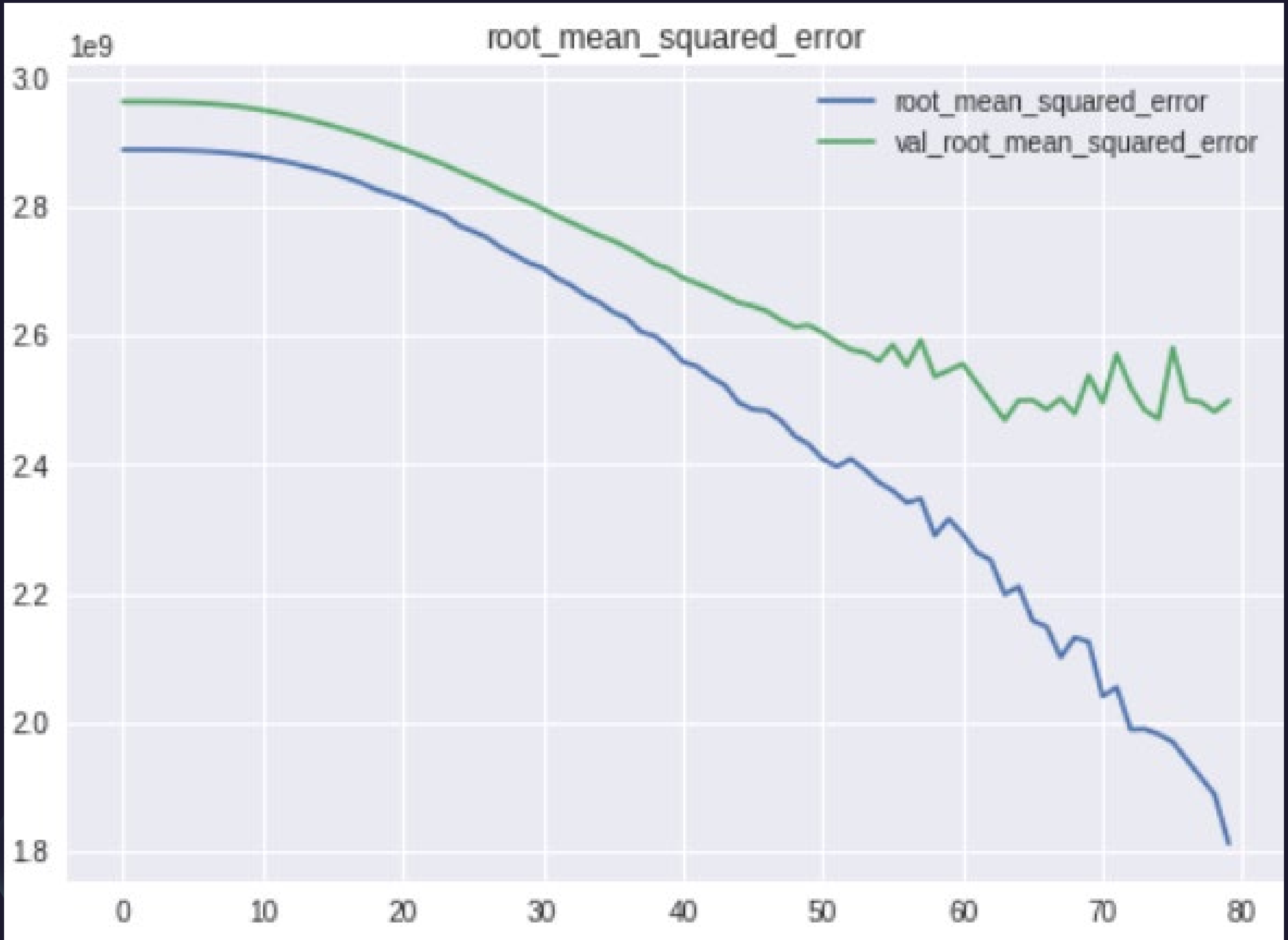
Agency	
Centers for Medicare and Medicaid Services	687665667131.05005
Social Security Administration	606352947124.80005
Department of the Navy	92445245668.16000
Under Secretary for Benefits/Veterans Benefits Administration	77393474082.00000
Department of the Army	62684553714.48000
Department of the Air Force	52762096380.28000
Under Secretary for Health/Veterans Health Administration	50483540685.00000
Department of Education	39539576671.30000
Federal Highway Administration	37183586470.43000
Department of Energy	27513206449.87000
Defense Logistics Agency	27253153800.05000
National Institutes of Health	26593656955.34000
Department of Veterans Affairs	24767286448.29000
Assistant Secretary for Public and Indian Housing	21669623303.82000
Federal Emergency Management Agency	13598632911.22000
Assistant Secretary for Community Planning and Development	12364569752.51000
Defense Health Agency (DHA)	12309897209.56000
National Aeronautics and Space Administration	11991007733.95000
Assistant Secretary for Housing--Federal Housing Commissioner	9471040758.59000
Railroad Retirement Board	9353925930.69000

- Funding Breakdown
- The department of Energy is already #10 on the list of highest funded government agencies
- Despite this, the agency only received $27B in Funding.
- Military funding was $5.5B in the same year
- We spend 10x more on military spending than we do on developing our energy infrastructure.



- Model # 3
- The Loss profile selected was ’mse’
- The losses remained smooth initially, but the added weights caused the models to diverge over time.

- Model # 3
- The MAE shows that both models were turning up together,
- Both sets executed a massive correction around epoch 60
- The training set continued to learn after the correction
- The test set did not.



- Model # 3
- The Loss profile selected was 'mse'
- The losses remained smooth initially, but the added weights caused the models to diverge over time.

Model Performance:

```
final RMSE: 2499062835.959598
final MAE: 282615772.00248593
final R2: 0.2809823085896217
```

The model is presently only achieving 28% accuracy.

Further tuning is likely to improve accuracy.

Additional Clustering methodologies should be considered.

This model should not be considered "Production Ready"

It can be used with a factoring value to "test run" various spending hypotheses.

Topic Five: Recommendations

Recommendations:

- 1. Continue to tune and develop model to allow for full determination of annual allocation by geography and agency.

- Once developed, model projections can be used to projected larger scale, higher cost programs with multi-year durations.

- Look for ways to increase overall funding.

Staich-FederalSpending.zip

Thank You

Presenter name

Email address

Website address