```
In [2]:
         # Import our dependencies
         from sklearn.model_selection import train_test_split
         from sklearn.preprocessing import StandardScaler,OneHotEncoder, MinMaxScaler
         import pandas as pd
         import tensorflow as tf
         import numpy as np
         # Import our input dataset
         df = pd.read_csv('./pitcher_salaries_cleaned.csv')
         df.head()
Out[2]:
```

•	Year	Full Name	Age	Salary	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Faced by Pitcher	Games Finished	Weight	ı
0	1990	AbbottJim	23	185000	4.51	246	106	105	16	10	14	635	925	0	200	
1	1990	AbbottPaul	23	100000	5.97	37	23	25	0	0	5	104	162	0	185	
2	1990	AldredScott	22	100000	3.77	13	6	7	0	1	2	43	63	0	195	
3	1990	AndersonAllan	26	300000	4.53	214	95	82	20	7	18	566	797	0	178	
4	1990	AppierKevin	23	100000	2.76	179	57	127	13	12	8	557	784	1	180	
4																<b>&gt;</b>

### **Create Salary Brackets**

```
In [3]:
         # look at distribution of salaries (suppressing scientific notation)
         df['Salary'].describe().apply(lambda x: format(x, 'f'))
```

```
Out[3]: count
                      4937.000000
         mean
                   3011304.443387
                   4265619.190449
        std
        min
                    100000.000000
         25%
                    327000.000000
         50%
                    980000.000000
         75%
                   4000000.000000
                  33000000.000000
         max
         Name: Salary, dtype: object
```

```
In [5]:
         # create salary brackets and labels
         bins = [0, 499999, 4999999, 9999999, 34999999]
         labels = ['low', 'mid', 'high', 'top']
```

```
In [6]:
         # apply salary brackets
         df['Salary Bin'] = pd.cut(df['Salary'], bins=bins, labels=labels)
```

Out[6]: **Batters** Earned Strike Home Outs Faced Games Year Full Name Age Salary ERA Hits Wins Losses Pitched **Finished** Outs Runs Runs by Pitcher 0 1990 AbbottJim 23 185000 4.51 106 105 16 10 635 925 0 246 14 AbbottPaul 25 5 104 **1** 1990 23 100000 5.97 37 23 0 0 162 0 **2** 1990 AldredScott 100000 3.77 13 6 7 0 2 43 63 0 7 1990 AndersonAllan 26 300000 4.53 214 95 82 20 18 566 797 0 4 1990 23 100000 2.76 179 57 127 13 12 8 557 784 1 AppierKevin

	Year	Full Name	Age	Salary	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Faced by Pitcher	Games Finished
•••														
4932	2016	WorleyVance	29	2600000	3.53	84	34	56	11	2	2	260	365	13
4933	2016	WrightMike	26	510500	5.79	81	48	50	12	3	4	224	328	5
4934	2016	WrightSteven	32	514500	3.33	138	58	127	12	13	6	470	656	0
4935	2016	YoungChris	37	4250000	6.19	104	61	94	28	3	9	266	406	7
4936	2016	ZimmermannJordan	30	18000000	4.87	118	57	66	14	9	7	316	450	1
4937 rows × 20 columns														
4														•

Rattors

## **Encode Salary Bins column**

```
In [7]: # encode object features
    from sklearn.preprocessing import LabelEncoder
    le = LabelEncoder()
    encoded_df = df.copy()
    df['Salary Bin'] = le.fit_transform(df['Salary Bin'])

    df.head()
```

Out[7]:		Year	Full Name	Age	Salary	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Batters Faced by Pitcher	Games Finished	Weight I	1
	0	1990	AbbottJim	23	185000	4.51	246	106	105	16	10	14	635	925	0	200	
	1	1990	AbbottPaul	23	100000	5.97	37	23	25	0	0	5	104	162	0	185	
	2	1990	AldredScott	22	100000	3.77	13	6	7	0	1	2	43	63	0	195	
	3	1990	AndersonAllan	26	300000	4.53	214	95	82	20	7	18	566	797	0	178	
	4	1990	AppierKevin	23	100000	2.76	179	57	127	13	12	8	557	784	1	180	

```
In [11]: # create new df using only top features
df= df.filter(['ERA','Batters Faced by Pitcher','Outs Pitched','Games Finished','Strike Outs','Salary Bin
df.head()
```

Out[11]:		ERA	<b>Batters Faced by Pitcher</b>	Outs Pitched	<b>Games Finished</b>	Strike Outs	Salary Bin
	0	4.51	925	635	0	105	1
	1	5.97	162	104	0	25	1
	2	3.77	63	43	0	7	1
	3	4.53	797	566	0	82	1
	4	2.76	784	557	1	127	1

# Split Features/Target & Training/Testing Sets

• y variable: Our target variable, Salary Bin

• X variable: Our features

```
In [12]:
# Split our preprocessed data into our features and target arrays
y = df["Salary Bin"].values
X = df.drop(["Salary Bin"],1).values

# Split the preprocessed data into a training and testing dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)
```

C:\Users\alyss\anaconda3\envs\mlenv\lib\site-packages\ipykernel\_launcher.py:3: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only This is separate from the ipykernel package so we can avoid doing imports until

# Build and Instantiate StandardScaler object, then standardize numerical features

```
In [13]: # Create a StandardScaler instance
scaler = StandardScaler()

# Fit the StandardScaler
X_scaler = scaler.fit(X_train)

# Scale the data
X_train_scaled = X_scaler.transform(X_train)
X_test_scaled = X_scaler.transform(X_test)
```

### **Build Neural Net Framework**

HL1:

- 50 neurons
- activation fxn: relu

HL2:

- 40 neurons
- activation fxn: relu

HL3:

- 30 neurons
- activation fxn: relu

output layer:

- 4 neurons
  - same as number of salary bins, suggested from (https://machinelearningmastery.com/loss-and-loss-functionsfor-training-deep-learning-neural-networks/)
- activation fxn: softmax
  - suggested for multiclass classification problems per (https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/)

```
# Define the model - deep neural net
number_input_features = len(X_train[0])
hidden_nodes_layer1 = 50
hidden_nodes_layer2 = 40
hidden_nodes_layer3 = 4
```

```
nn = tf.keras.models.Sequential()
# First hidden Layer
nn.add(
    tf.keras.layers.Dense(units=hidden_nodes_layer1, input_dim=number_input_features, activation="relu")
)
# Second hidden Layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer2, activation="relu"))
# Third hidden Layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer3, activation="relu"))
# Output Layer
nn.add(tf.keras.layers.Dense(units=1, activation="softmax"))
# Check the structure of the model
nn.summary()
```

Model: "sequential"

Lavon (type)	Output Chang	Param #
Layer (type)	Output Shape	#
dense (Dense)	(None, 50)	300
dense_1 (Dense)	(None, 40)	2040
dense_2 (Dense)	(None, 4)	164
dense_3 (Dense)	(None, 1)	5
Total params: 2,509		

Total params: 2,509 Trainable params: 2,509 Non-trainable params: 0

### Compile the Model

```
# Compile the model
nn.compile(loss="CategoricalCrossentropy", optimizer="adam", metrics=["accuracy"])
```

#### Train the model

```
In [16]:
     # Train the model
     fit model = nn.fit(X train,y train,epochs=200)
     Epoch 1/200
     Epoch 2/200
     Epoch 3/200
     116/116 [============] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
     Fnoch 4/200
     116/116 [============ - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
     Epoch 5/200
     Epoch 6/200
     116/116 [============] - Os 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
     Epoch 7/200
     Epoch 8/200
     116/116 [============] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
     Epoch 9/200
     116/116 [============] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
```

```
Epoch 10/200
Epoch 11/200
Epoch 12/200
116/116 [============ ] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 13/200
Epoch 14/200
Epoch 15/200
Epoch 16/200
116/116 [============= ] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 17/200
Epoch 18/200
Epoch 19/200
Epoch 20/200
Epoch 21/200
116/116 [============ - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 22/200
Epoch 23/200
Epoch 24/200
Epoch 25/200
Epoch 26/200
116/116 [=============] - 0s 618us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 27/200
Epoch 28/200
116/116 [============== ] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 29/200
Epoch 30/200
Epoch 31/200
Epoch 32/200
116/116 [============= ] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 33/200
116/116 [============ ] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 34/200
Epoch 35/200
Epoch 36/200
Epoch 37/200
Epoch 38/200
Epoch 39/200
116/116 [============== ] - 0s 817us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 40/200
Epoch 41/200
Epoch 42/200
Epoch 43/200
Epoch 44/200
116/116 [=================== ] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 45/200
Epoch 46/200
Epoch 47/200
```

```
Epoch 48/200
Epoch 49/200
Epoch 50/200
116/116 [============= - 0s 600us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 51/200
Epoch 52/200
Epoch 53/200
Epoch 54/200
Epoch 55/200
Epoch 56/200
Epoch 57/200
Epoch 58/200
Epoch 59/200
116/116 [============ - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 60/200
Epoch 61/200
116/116 [============ ] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 62/200
Epoch 63/200
Epoch 64/200
Epoch 65/200
Fnoch 66/200
116/116 [============== ] - 0s 600us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 67/200
Epoch 68/200
Epoch 69/200
Epoch 70/200
Epoch 71/200
116/116 [============ ] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 72/200
Epoch 73/200
Epoch 74/200
Epoch 75/200
116/116 [=============] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 76/200
Epoch 77/200
116/116 [============== ] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 78/200
Epoch 79/200
Epoch 80/200
Epoch 81/200
Epoch 82/200
116/116 [=================== ] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 83/200
Epoch 84/200
Epoch 85/200
```

```
Epoch 86/200
Epoch 87/200
Epoch 88/200
116/116 [============ ] - 0s 644us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 89/200
Epoch 90/200
Epoch 91/200
Epoch 92/200
116/116 [============= ] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 93/200
Epoch 94/200
Epoch 95/200
116/116 [============ - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 96/200
Epoch 97/200
116/116 [============ ] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 98/200
Epoch 99/200
116/116 [============ ] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 100/200
Epoch 101/200
Epoch 102/200
116/116 [=============] - 0s 696us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 103/200
Epoch 104/200
116/116 [============== ] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 105/200
Epoch 106/200
Epoch 107/200
Epoch 108/200
116/116 [============= ] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 109/200
116/116 [============ - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 110/200
Epoch 111/200
116/116 [============== ] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 112/200
Epoch 113/200
Epoch 114/200
Epoch 115/200
Epoch 116/200
Epoch 117/200
Epoch 118/200
Epoch 119/200
Epoch 120/200
116/116 [=================== ] - 0s 600us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 121/200
Epoch 122/200
Epoch 123/200
```

```
Epoch 124/200
Epoch 125/200
Epoch 126/200
Epoch 127/200
Epoch 128/200
Epoch 129/200
Epoch 130/200
Epoch 131/200
Epoch 132/200
Epoch 133/200
Epoch 134/200
Epoch 135/200
116/116 [============= - 0s 661us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 136/200
Epoch 137/200
116/116 [============ ] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 138/200
Epoch 139/200
Epoch 140/200
Epoch 141/200
Epoch 142/200
Epoch 143/200
Epoch 144/200
Epoch 145/200
Epoch 146/200
116/116 [============== ] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 147/200
116/116 [============= - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 148/200
Epoch 149/200
116/116 [============== ] - 0s 609us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 150/200
Epoch 151/200
Epoch 152/200
Epoch 153/200
116/116 [============== ] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 154/200
Epoch 155/200
Epoch 156/200
Epoch 157/200
Epoch 158/200
116/116 [=================== ] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 159/200
Epoch 160/200
Epoch 161/200
```

```
Epoch 162/200
Epoch 163/200
Epoch 164/200
116/116 [============ ] - 0s 644us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 165/200
Epoch 166/200
Epoch 167/200
Epoch 168/200
Epoch 169/200
Epoch 170/200
Epoch 171/200
116/116 [============= - 0s 800us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 172/200
Epoch 173/200
Epoch 174/200
Epoch 175/200
116/116 [============= - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 176/200
Epoch 177/200
Epoch 178/200
Epoch 179/200
Epoch 180/200
116/116 [============== ] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 181/200
Epoch 182/200
Epoch 183/200
Epoch 184/200
116/116 [============= ] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 185/200
116/116 [============= - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 186/200
Epoch 187/200
116/116 [============== ] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 188/200
Epoch 189/200
Epoch 190/200
Epoch 191/200
116/116 [============== ] - 0s 645us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 192/200
Epoch 193/200
Epoch 194/200
116/116 [============= - 0s 670us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 195/200
Epoch 196/200
116/116 [=================== ] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 197/200
Epoch 198/200
Epoch 199/200
```

```
In [17]:
           # Evaluate the model using the test data
           model_loss, model_accuracy = nn.evaluate(X_test_scaled,y_test,verbose=2)
           print(f"Loss: {model_loss*100:.2f}%, Accuracy: {model_accuracy*100:.2f}%")
          39/39 - 0s - loss: 0.0000e+00 - accuracy: 0.3887 - 121ms/epoch - 3ms/step
          Loss: 0.00%, Accuracy: 38.87%
In [18]:
           # Create a DataFrame containing training history
           history_df = pd.DataFrame(fit_model.history, index=range(1,len(fit_model.history["loss"])+1))
           # Plot the loss
           history_df.plot(y="loss")
Out[18]: <AxesSubplot:>
                                                           loss
           0.04
           0.02
           0.00
          -0.02
          -0.04
                 Ó
                      25
                            50
                                  75
                                       100
                                            125
                                                  150
                                                       175
                                                             200
In [19]:
           # Plot the accuracy
           history_df.plot(y="accuracy")
Out[19]: <AxesSubplot:>
                                                        accuracy
          0.395
          0.390
          0.385
          0.380
          0.375
          0.370
          0.365
          0.360
                      25
                                      100
                                            125
                                                             200
 In [ ]:
```

Epoch 200/200