FRAUD DETECTION FROM MEDICARE CLAIMS

Presented By Anvesh

Healthcare Fraud: An Unseen Epidemic Impacting Medicare's Effectiveness

- **Problem Statement**
  - Healthcare fraud is a persistent issue in the US, with certain providers exploiting Medicare for personal gain. This problem limits Medicare's capacity to serve the healthcare needs of elderly and other qualifying individuals effectively.
  - Despite efforts by the Centers for Medicare and Medicaid Services (CMS) to minimize fraudulent activities, identifying patterns of fraudulent claims remains a challenge.
- **Objective**
  - Our goal is to examine patterns of fraudulent claims activity in the CMS Medicare dataset, using the list of fraudulent providers from LEIE.
  - We aim to identify specific features distinguishing fraudulent physicians from non-fraudulent ones and develop a classifier model for fraud detection.
- **Significance**
  - Addressing healthcare fraud is vital for the equitable distribution of Medicare resources, ensuring maximum reach and effectiveness of the program.
  - By leveraging data analysis and machine learning, we can proactively identify potential fraud cases, thus preserving resources for those truly in need.

What can be Medicare Frauds?

Medicare fraud and abuse can occur everywhere, increasing everyone's taxes and health care expenditures. Many instances include:

- A healthcare professional charges Medicare for goods or services you never received, such as billing you for a visit or a back brace you never received.
- A provider who bills Medicare twice for a good or service you only received once.
- A person who uses your Medicare card or number to submit false claims on your behalf.
- A business that proposes a Medicare drug plan to you that Medicare hasn't approved.

CMS has released datasets, including the Medicare Provider Utilization and Payment Data: Physician and Other Supplier.

The Office of the Inspector General provides a dataset of List of Excluded Individuals and Entities (LEIE), signifying fraudulent providers.

https://www.medicare.gov/basics/reporting-medicare-fraud-and-abuse

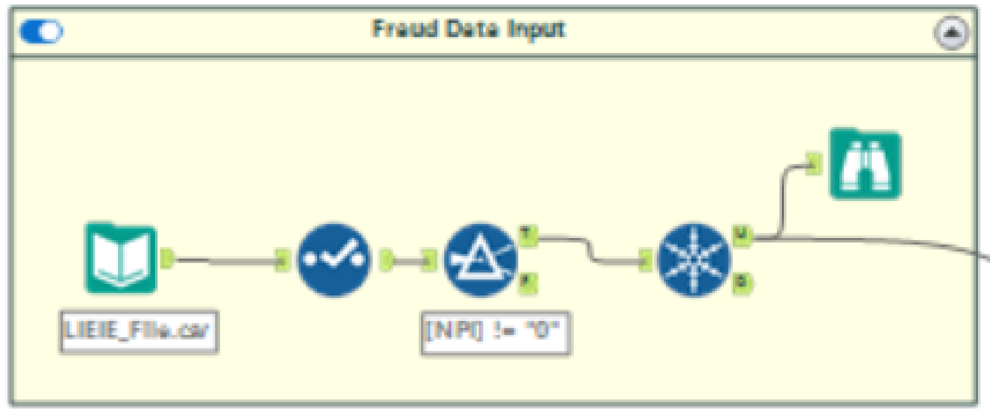
Joining the LEIE and Claims Data | Alteryx Workflow

The LEIE (List of Excluded Individuals/Entities) file contains comprehensive information about providers who have committed fraud.

Each provider is uniquely identified by their National Provider Identifier (NPI).

In this analysis, I am exclusively selecting the NPI column for further scrutiny using the Select tool in Alteryx.

Additionally, I'm employing the Filter tool to eliminate records with an NPI Code of zero, as all medicare claims data have a unique, non-zero NPI Code to identify provider.



In this phase, I am merging the Fraudulent Providers Data with the Claims Data using the "NPI" Column as the key for joining.

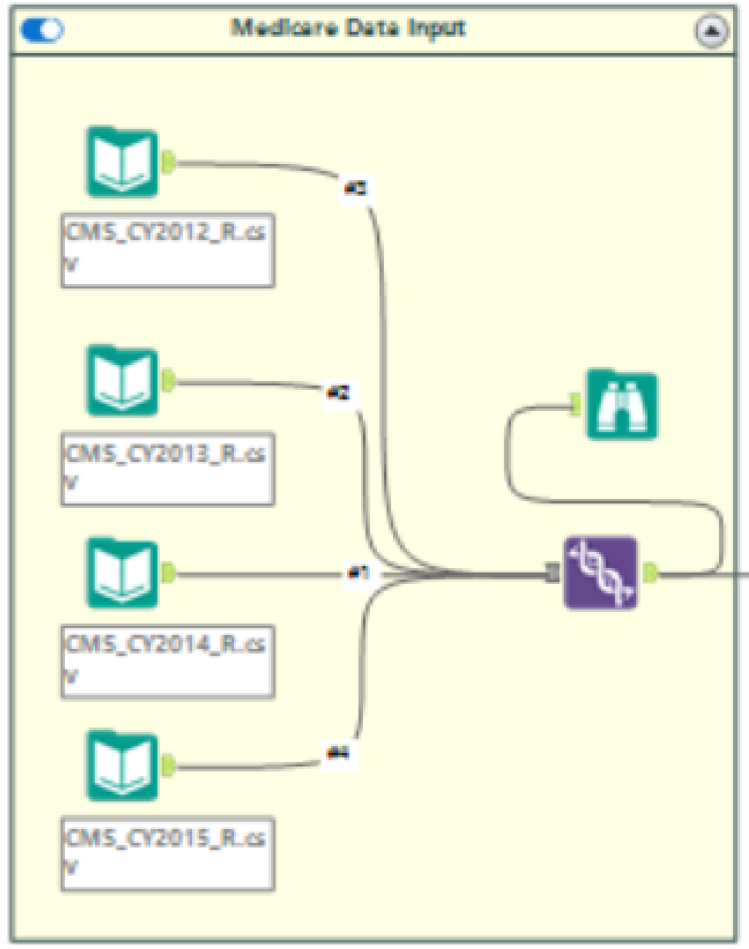
The 'J' Anchor of the Join tool in Alteryx outputs records that successfully matched from both the 'L' (Left) and 'R' (Right) inputs. These represent the fraudulent claims.

The 'R' Anchor outputs records from the 'R' input that didn't find a match in the 'L' input, signifying non-fraudulent claims.

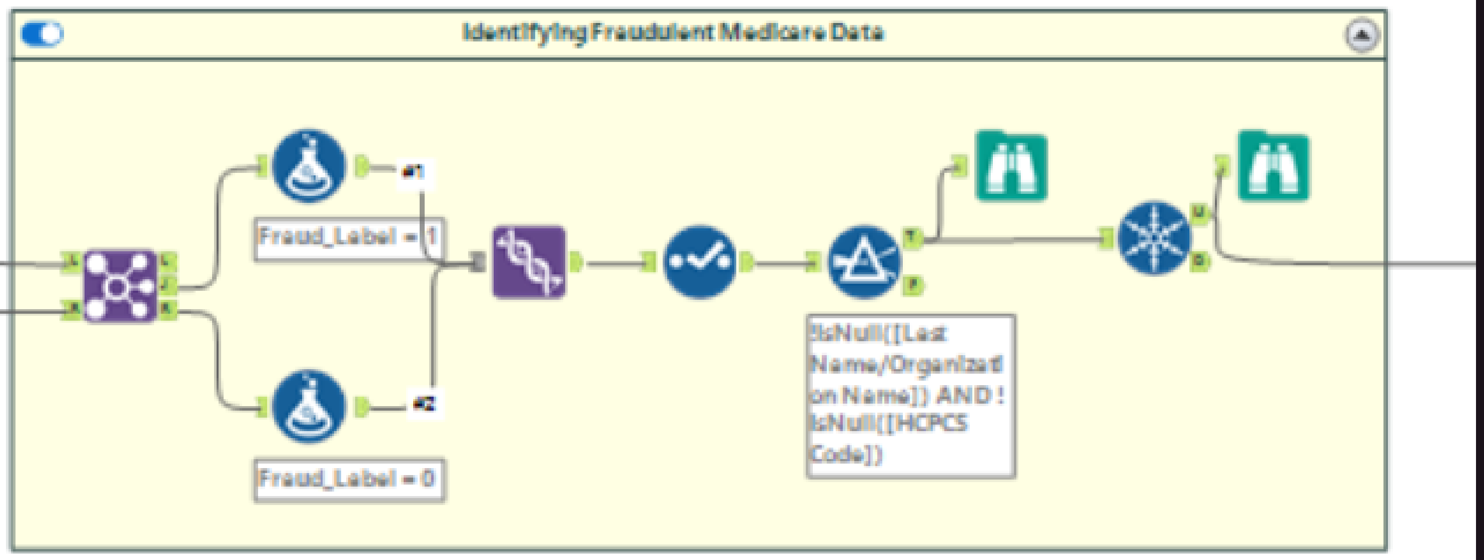
I'm employing the Formula tool to create a new column named 'Fraud_Label'. Data stemming from the 'J' Anchor is assigned a label of 1 (indicating fraud), while data from the 'R' Anchor is assigned a label of 0 (indicating non-fraud).

Next, I use the Union tool to combine both sets of labeled data. This is followed by removing any duplicate rows that may have been created due to the join operation.

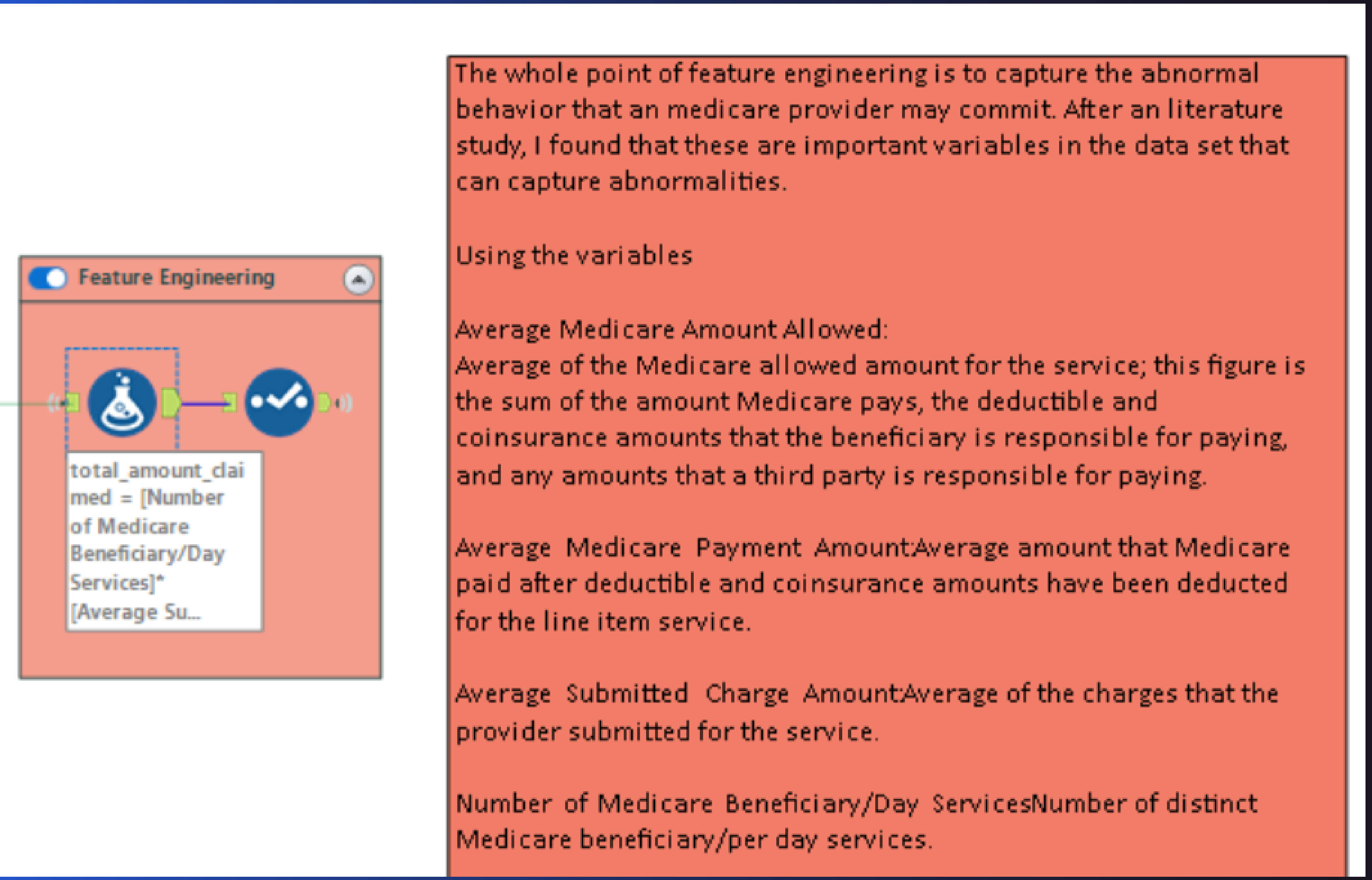
In my dataset, multiple claims are associated with the same provider. However, for my analysis, I only require a single row per provider. To achieve this, I'm selecting unique rows based on the combination of 'NPI' and 'HCPCS' codes.



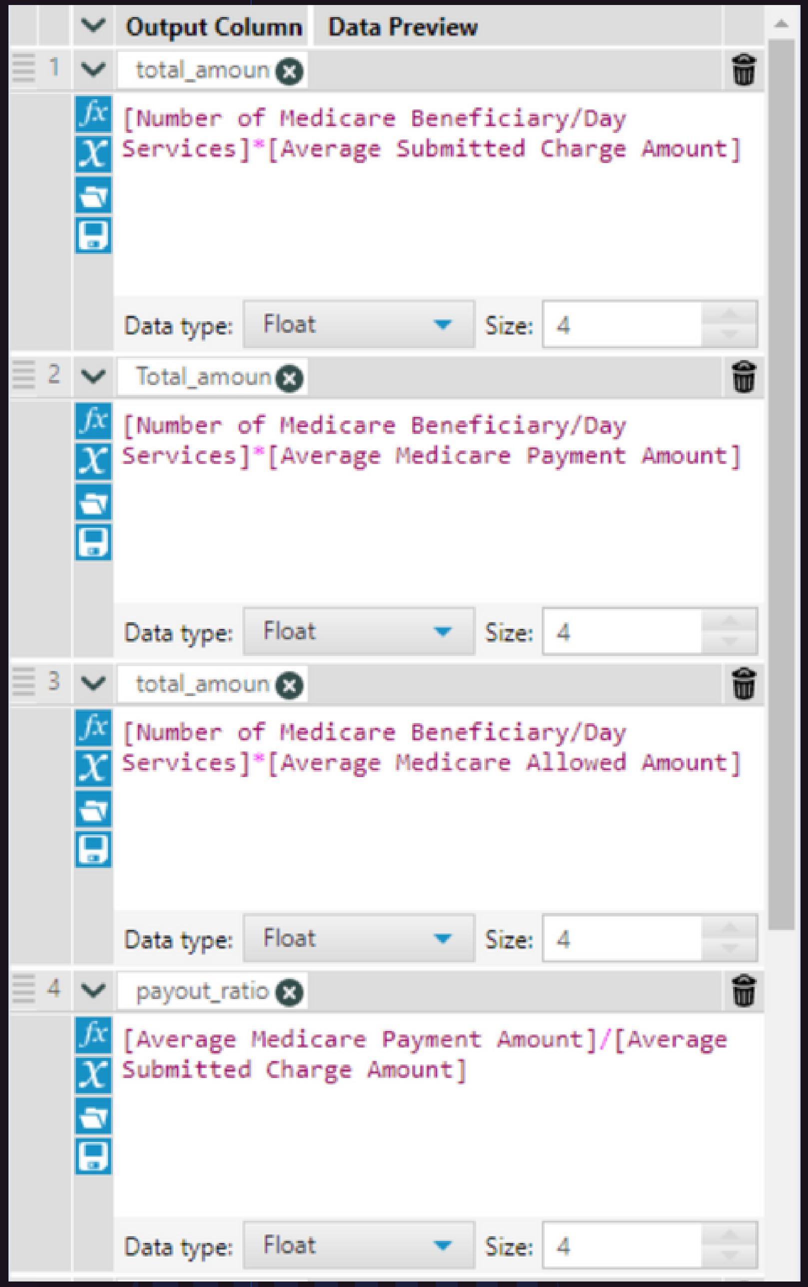
In this step, I'm utilizing the Union Tool in Alteryx to vertically stack the Medicare data files spanning the years 2012 through 2015. This process consolidates these annual datasets into a single comprehensive file for further analysis

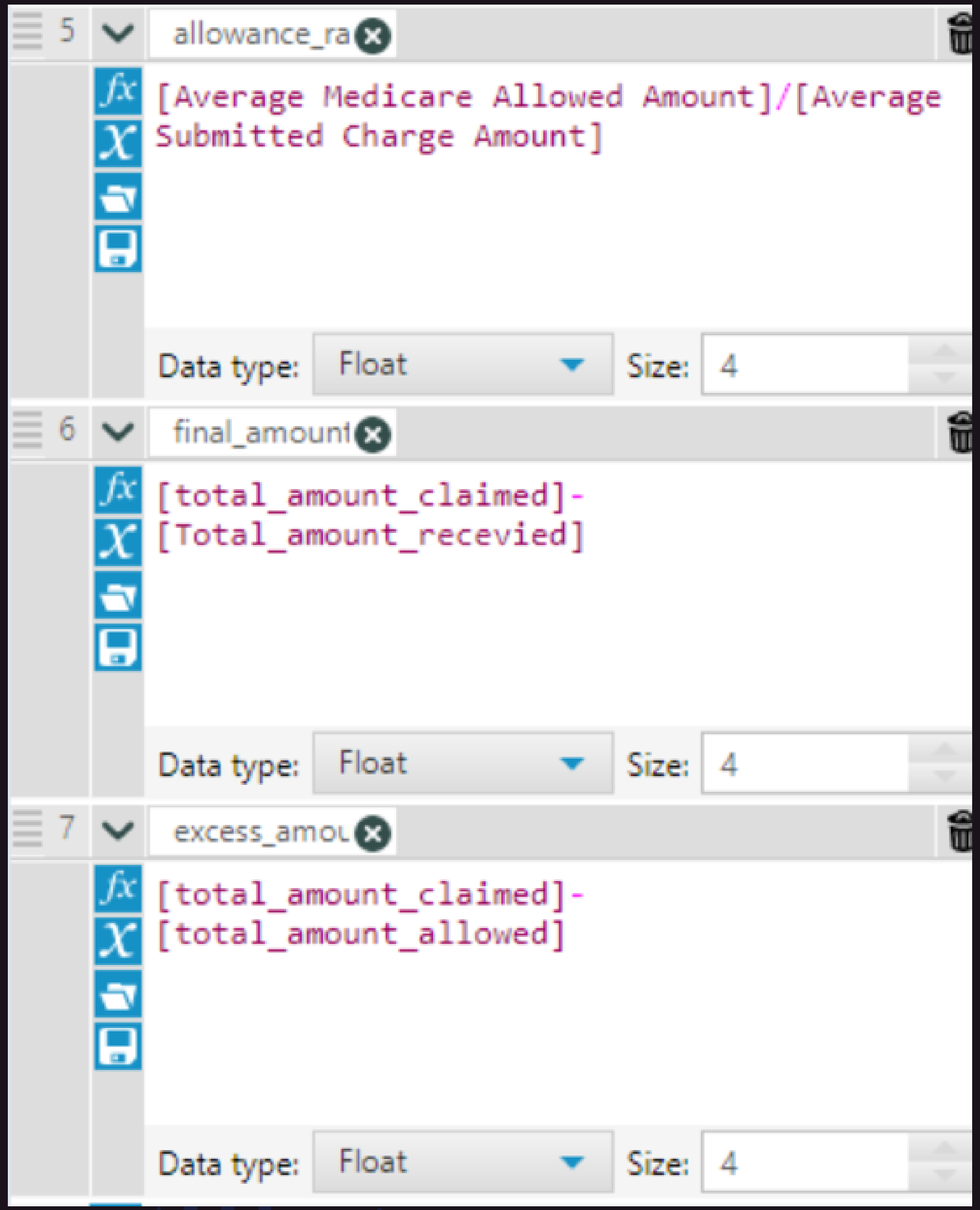


Feature Engineering | Alteryx Workflow



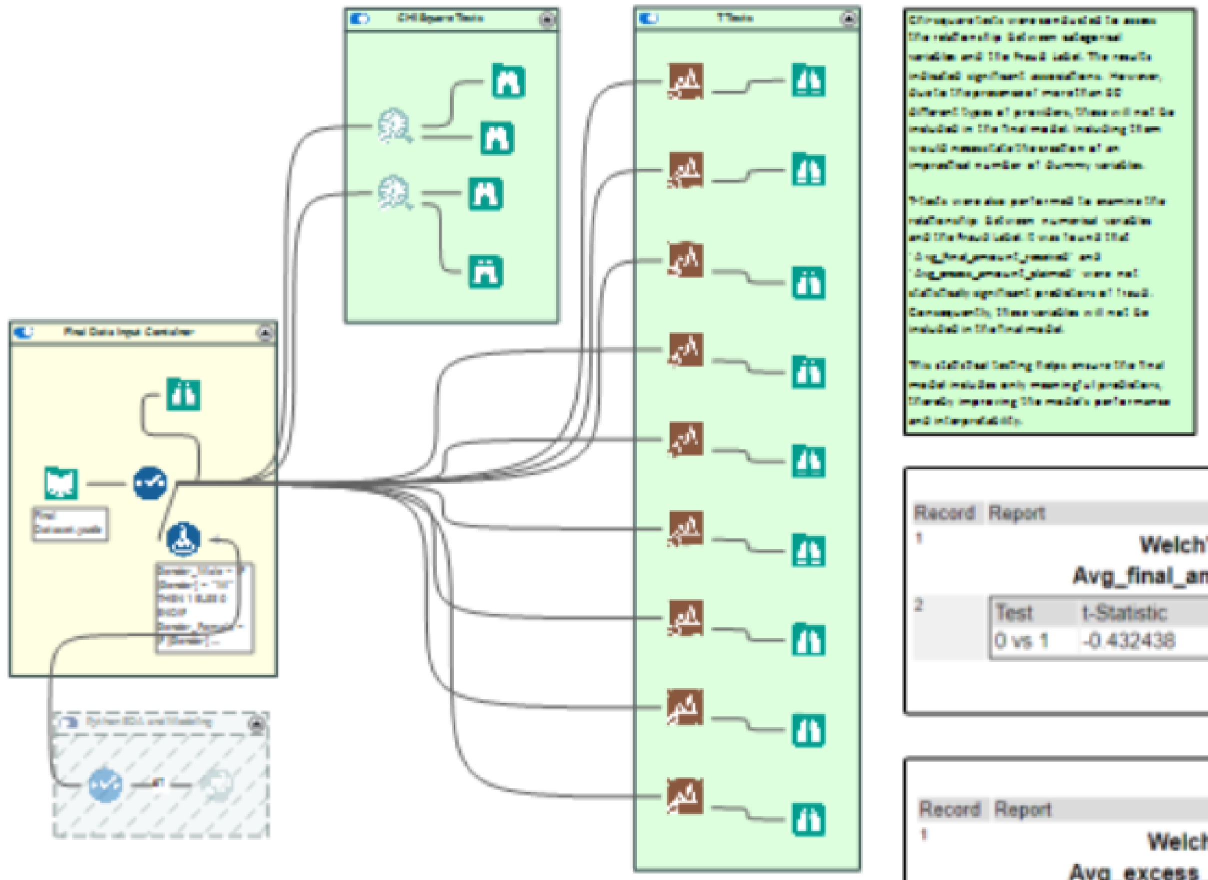
- total_amount_claimed
- Total_amount_recevied
- total_amount_alloweD
- payout_ratio
- allowance_ratio
- final_amount_recevied
- excess_amount_claimed





The whole point of feature engineering is to capture the abnormal behavior that an medicare provider may commit. After an literature study, I found that these are important variables in the data set that can capture abnormalities.

EDA and Model Building | Alteryx Workflow



Record | Report

1 **Welch's Two Sample t-test(s) of Avg_final_amount_recevied by Fraud_Label**

2

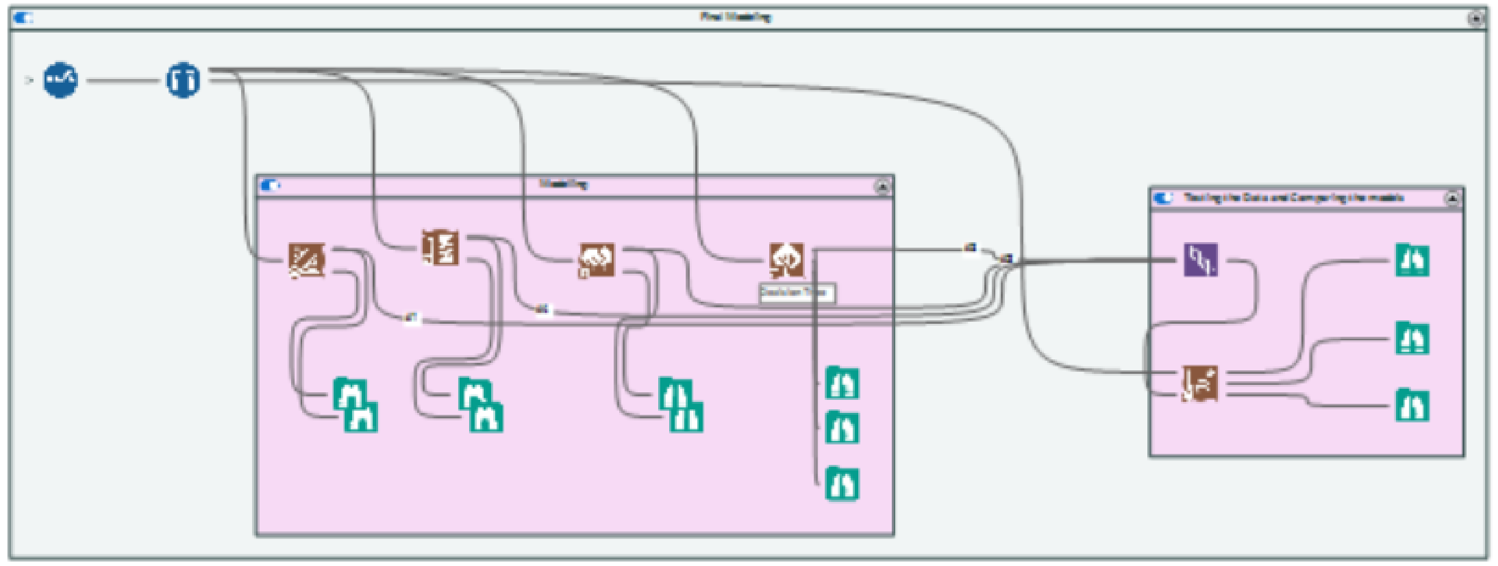
Test	t-Statistic	Degrees of Freedom	p-Value
0 vs 1	-0.432438	434.34	0.66564

Record | Report

1 **Welch's Two Sample t-test(s) of Avg_excess_amount_claimed by Fraud_Label**

2

Test	t-Statistic	Degrees of Freedom	p-Value
0 vs 1	0.5137	453.75	0.60771



Model Comparison Report

Fit and error measures

Model	Accuracy	F1	AUC	Accuracy_0	Accuracy_1
Support_Vector_Machine	0.7800	0.2326	0.6500	0.9573	0.1515
Random_Forest	0.7400	0.3390	0.7077	0.8632	0.3030
Decision Tree	0.7500	0.3697	0.6568	0.8675	0.3333
Test_Model	0.7633	0.3604	0.7164	0.8932	0.3030

Model: model names in the current comparison.

Accuracy: overall accuracy, number of correct predictions of all classes divided by total sample number.

Accuracy_[class name]: accuracy of Class [class name] is defined as the number of cases that are **correctly** predicted to be Class [class name] divided by the total number of cases that actually belong to Class [class name], this measure is also known as *recall*.

AUC: area under the ROC curve, only available for two-class classification.

F1: F1 score, 2 * precision * recall / (precision + recall). The *precision* measure is the percentage of actual members of a class that were predicted to be in that class divided by the total number of cases predicted to be in that class. In situations where there are three or more classes, average precision and average recall values across classes are used to calculate the F1 score.

- Python and Tableau for EDA
- T Tests and CHI Square using Alteryx
- Random Forest, Boosted, Decision Trees and SVM using Alteryx

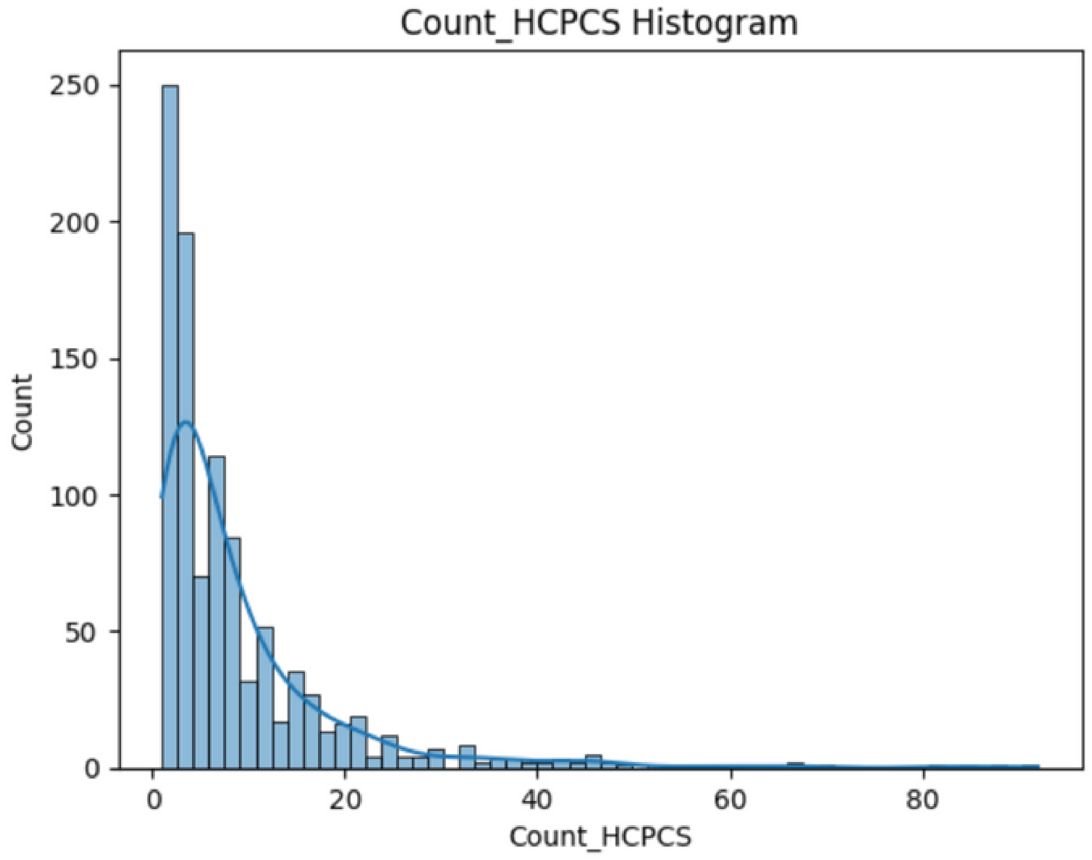
Data Set QA

- Most of the numerical values were highly skewed and were not normal

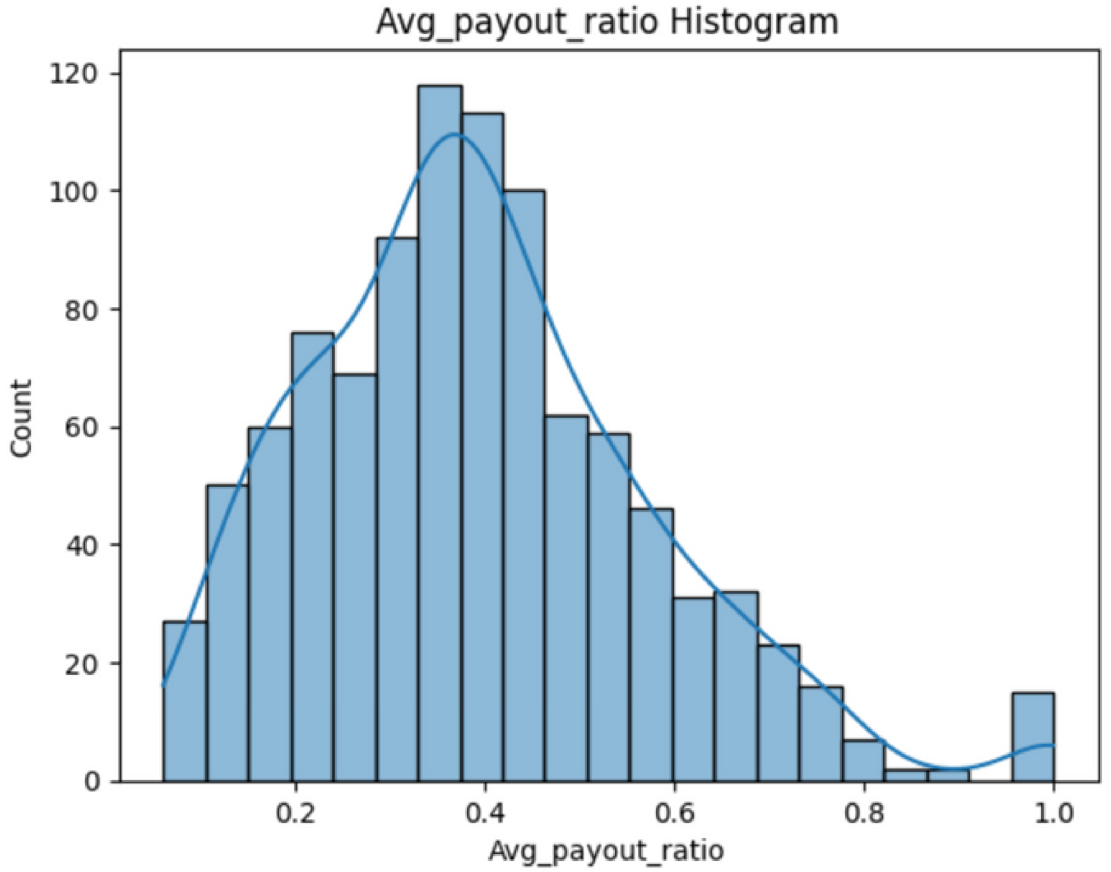
	Count_HCPCS	Avg_payout_ratio	Avg_Allowance_ratio	Avg_Final_Amount_recevied	Avg_Number of Medicare Beneficiaries	Avg_Number of Medicare Beneficiary/Day Services	Sum_total_amount_claimed	Sum_Total_amount_paid	Sum_Total_Amount_allowed
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	1.000000e+03	1.000000e+03
mean	8.776000	0.393823	0.514332	16444.665179	59.825377	143.956857	2.150381e+05	6.726330e+04	8.817577e+04
std	10.708244	0.180880	0.225564	36618.285232	59.420566	284.757244	6.161804e+05	1.407561e+05	1.802760e+05
min	1.000000	0.061101	0.083662	0.000000	11.000000	11.000000	2.210000e+02	3.900000e+01	3.900000e+01
25%	2.750000	0.264422	0.345560	3169.522588	25.663043	41.306818	2.119350e+04	7.706595e+03	1.082769e+04
50%	5.000000	0.377186	0.493178	7222.044224	43.784091	76.196429	7.897572e+04	2.510460e+04	3.450152e+04
75%	11.000000	0.498634	0.649388	16122.998575	73.471344	145.186688	2.019165e+05	7.118210e+04	9.410048e+04
max	92.000000	1.000000	1.000000	500559.594184	755.857143	3721.250000	1.308796e+07	2.155988e+06	2.702010e+06

Data Set QA

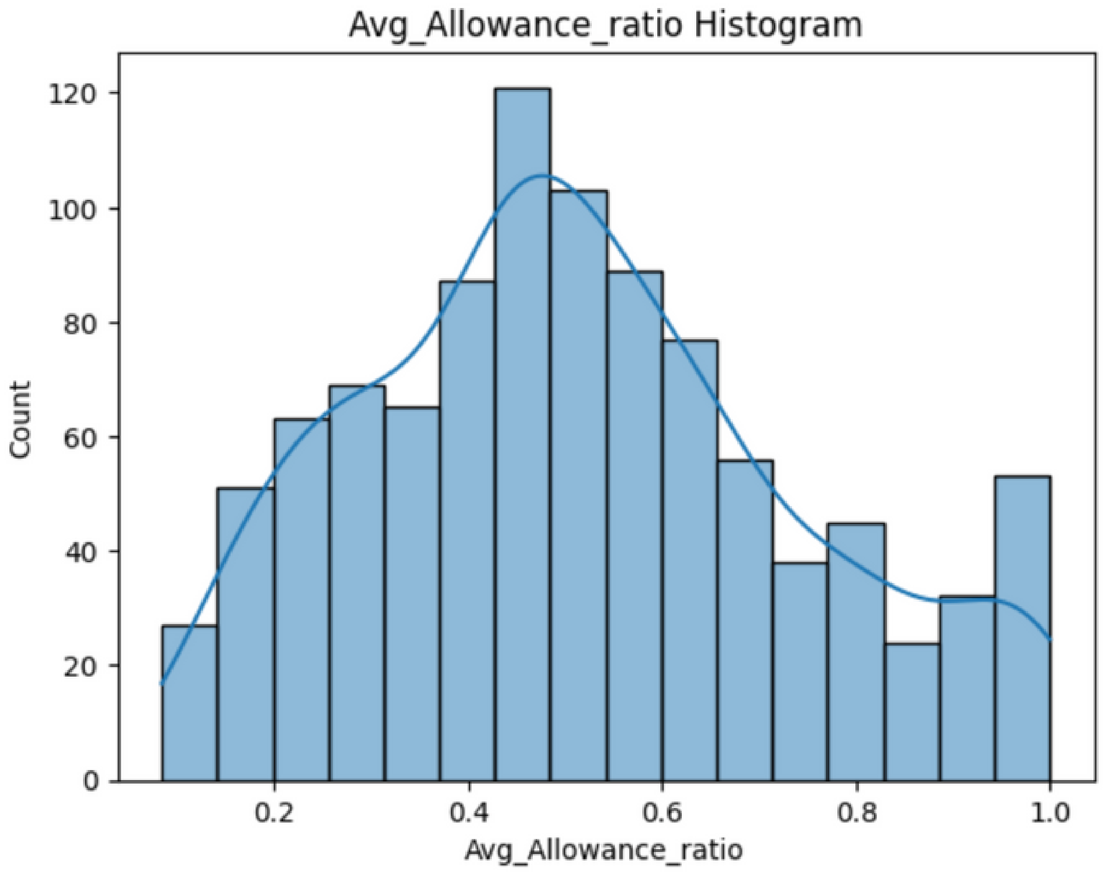
```
Column: Count_HCPCS
Skewness: 3.332112624712006
The distribution is highly right-skewed
Kurtosis: 15.540224902494042
The distribution is highly leptokurtic (peaked)
Shapiro-Wilk test p-value: 1.0183235940248446e-40
The data is not normally distributed
```



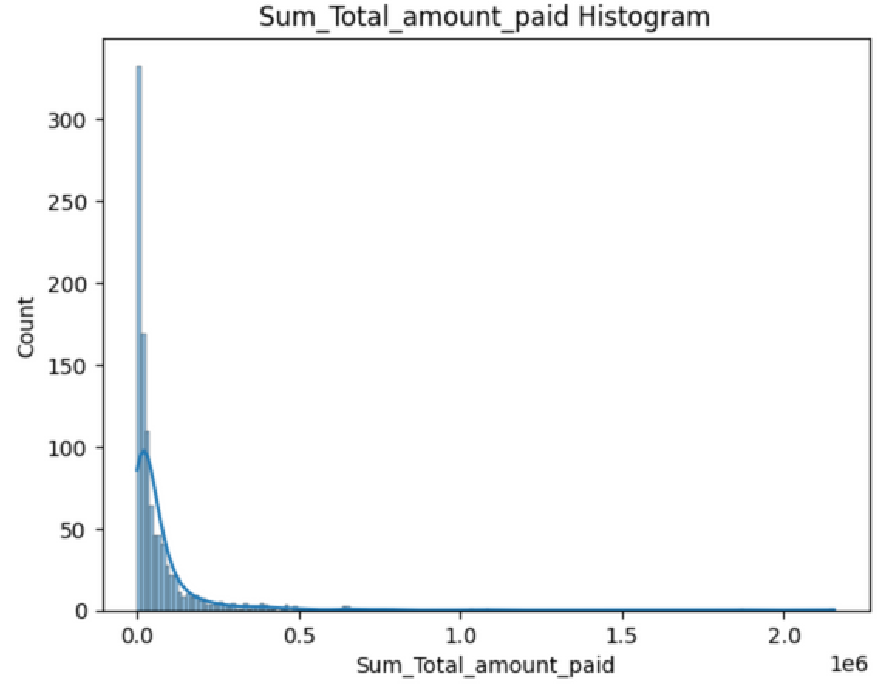

```
Column: Avg_payout_ratio
Skewness: 0.7159031058355508
The distribution is moderately right-skewed
Kurtosis: 0.6932629501340393
The distribution is approximately mesokurtic (normal)
Shapiro-Wilk test p-value: 1.585599768074309e-14
The data is not normally distributed
```

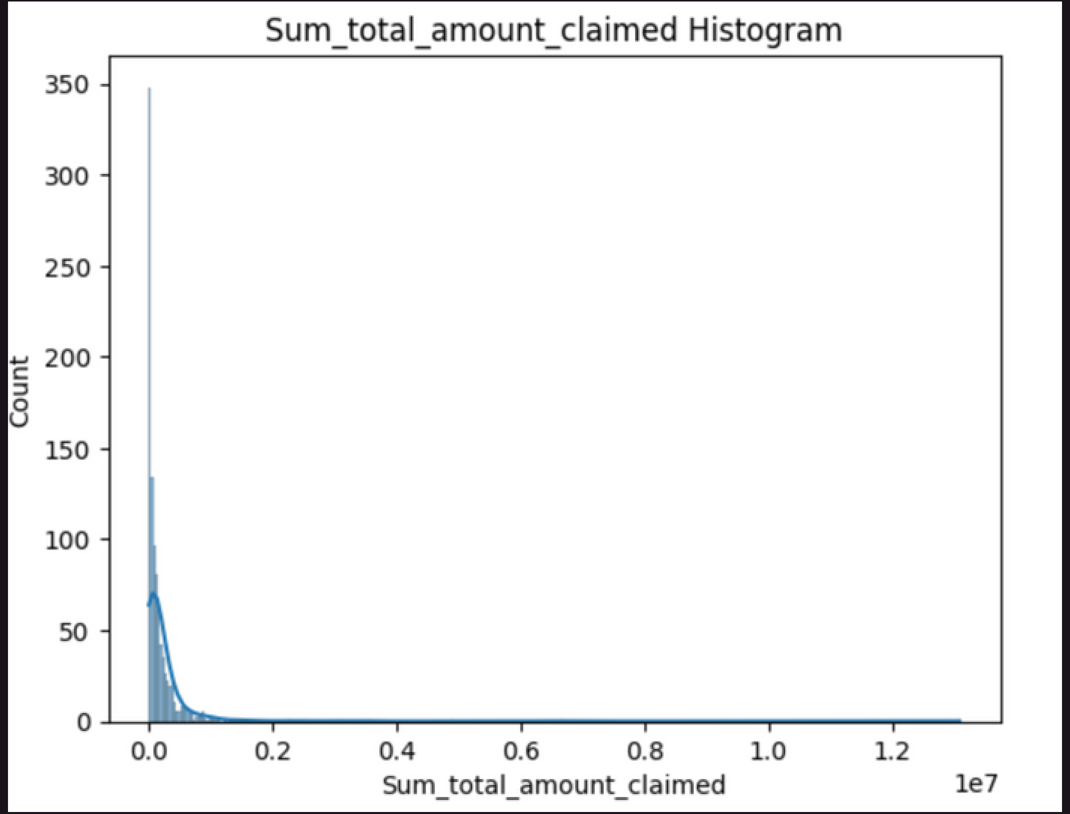



```
Column: Avg_Allowance_ratio
Skewness: 0.333417664603386
The distribution is approximately symmetric
Kurtosis: -0.5494814861282542
The distribution is approximately mesokurtic (normal)
Shapiro-Wilk test p-value: 5.008675765805215e-12
The data is not normally distributed
```

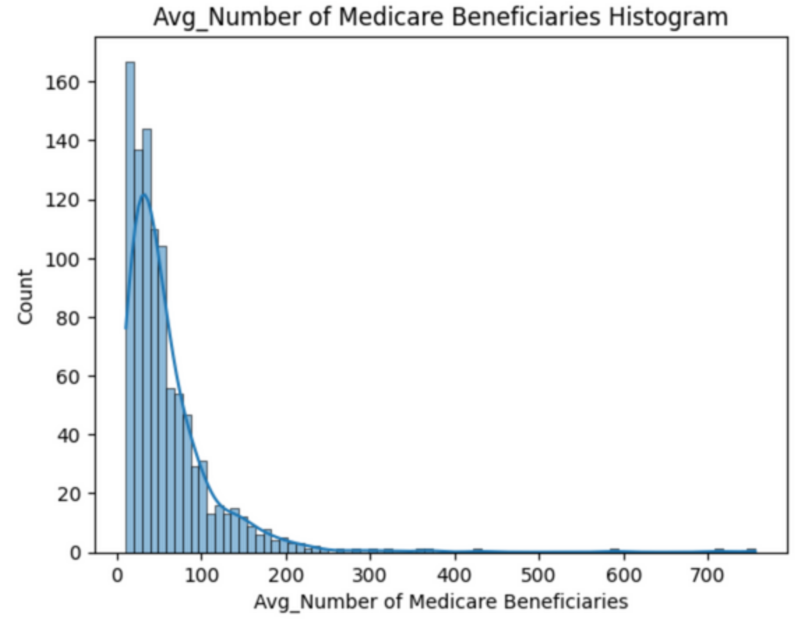


Column: Sum_Total_amount_paid
Skewness: 7.343742662836117
The distribution is highly right-skewed
Kurtosis: 82.58143900771745
The distribution is highly leptokurtic (peaked)
Shapiro-Wilk test p-value: 0.0
The data is not normally distributed

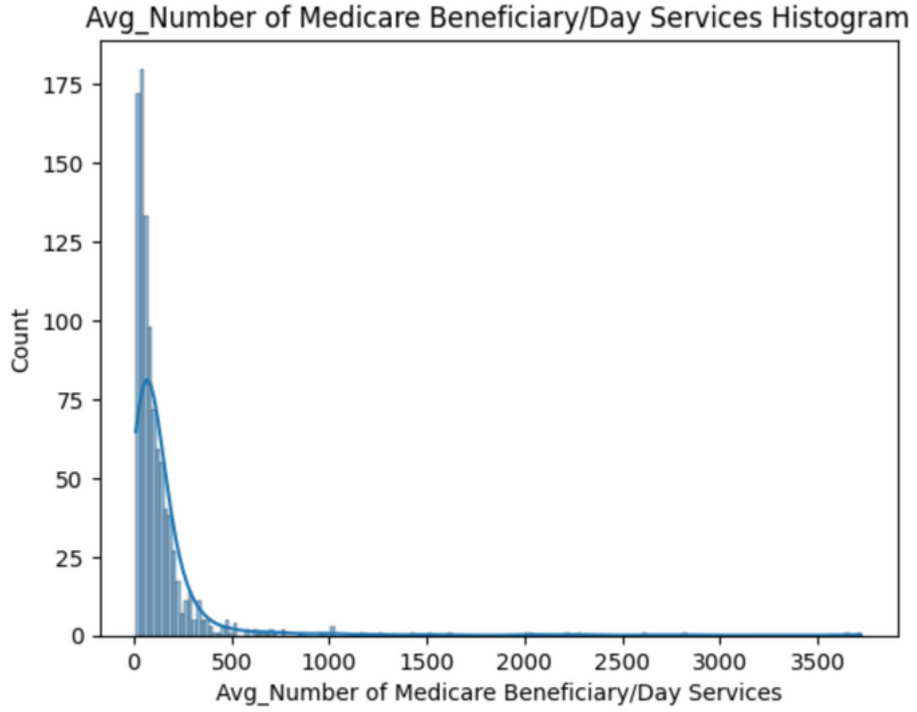




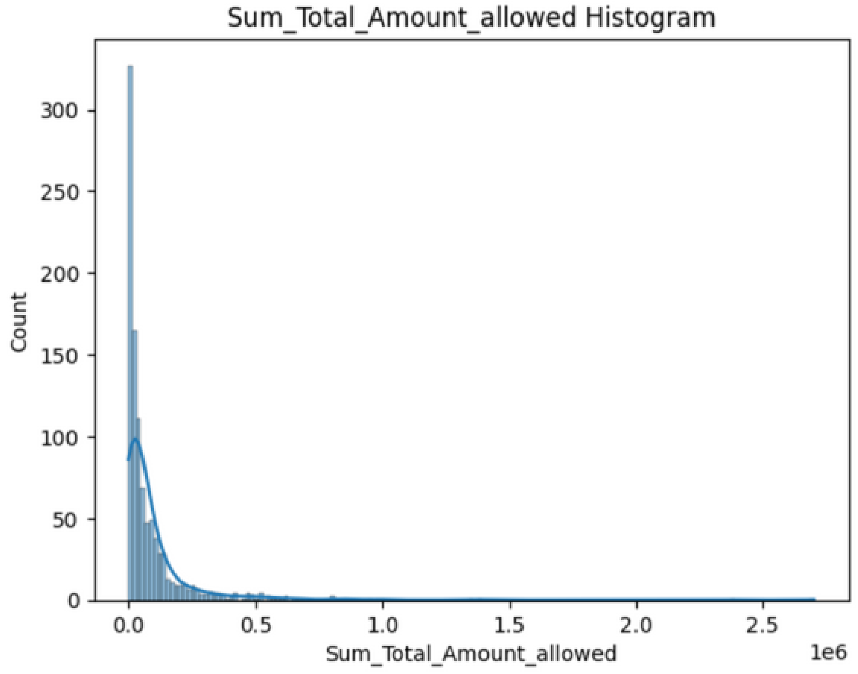
Column: Avg_Number of Medicare Beneficiaries
Skewness: 4.903196700861927
The distribution is highly right-skewed
Kurtosis: 42.00124286515312
The distribution is highly leptokurtic (peaked)
Shapiro-Wilk test p-value: 1.1169750059133117e-41
The data is not normally distributed



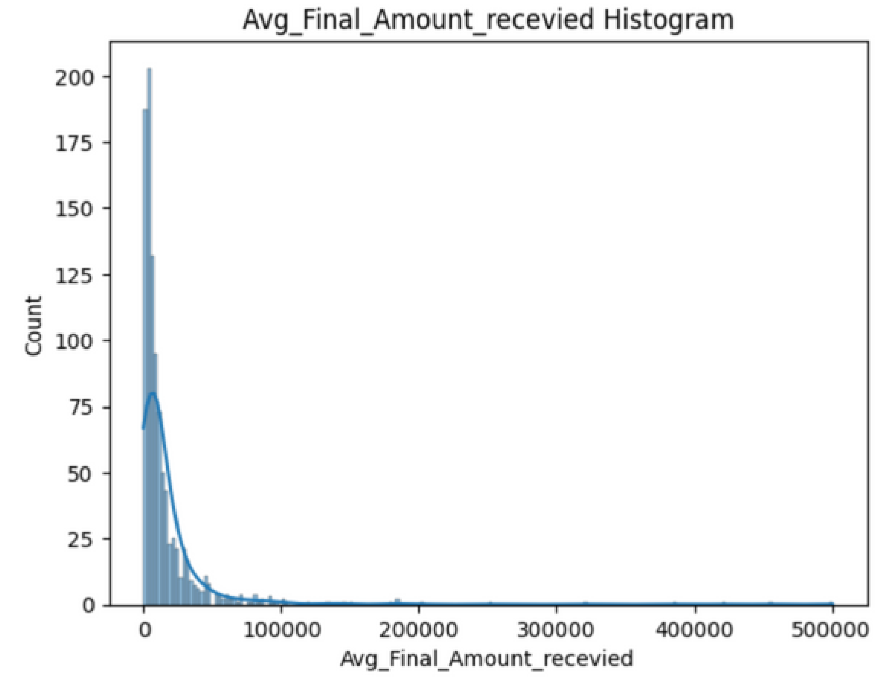
Skewness: 7.476916097970972
The distribution is highly right-skewed
Kurtosis: 71.11895719911658
The distribution is highly leptokurtic (peaked)
Shapiro-Wilk test p-value: 0.0
The data is not normally distributed



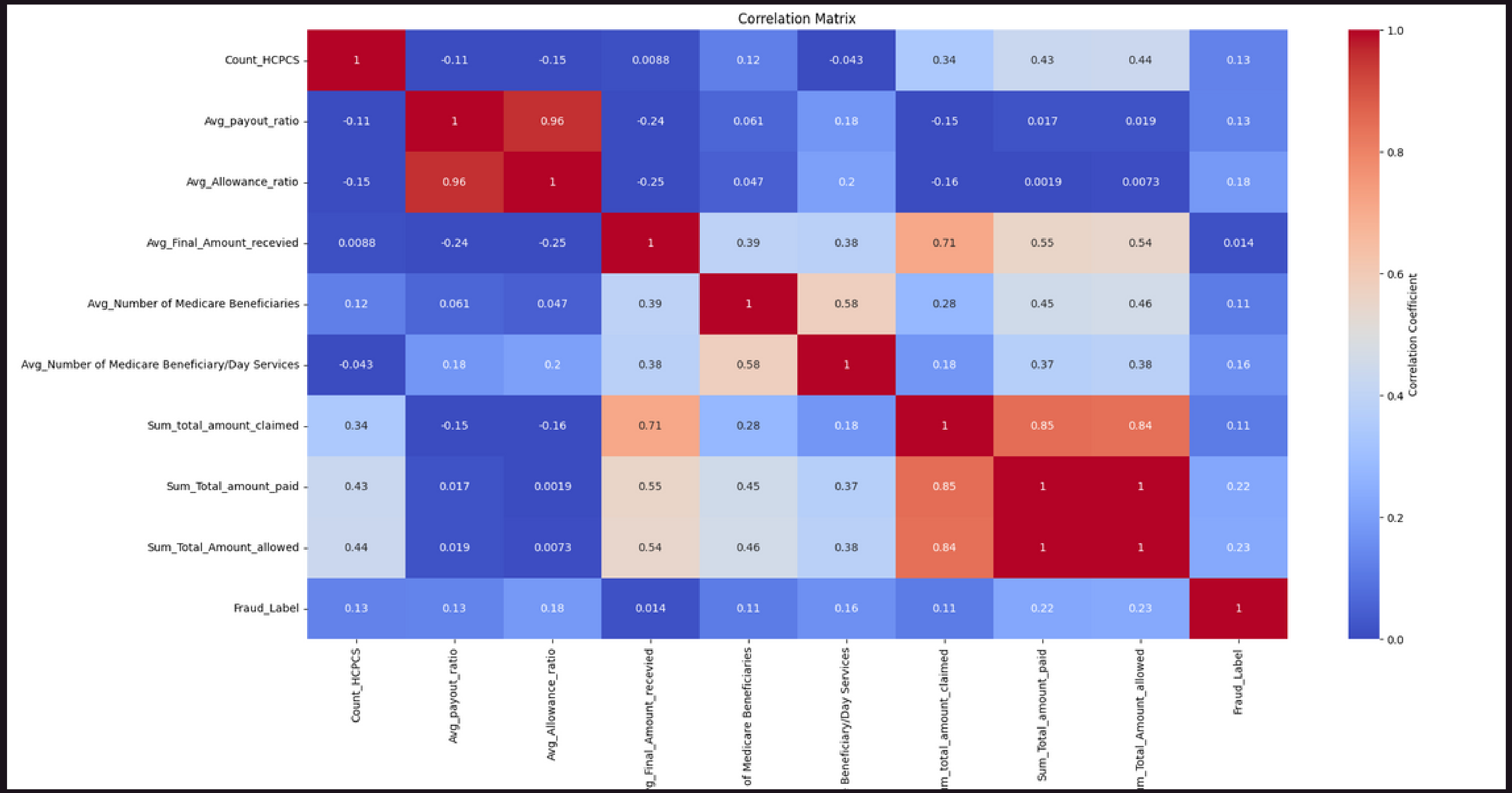
Column: Sum_Total_Amount_allowed
Skewness: 7.095681845501249
The distribution is highly right-skewed
Kurtosis: 77.37986496439925
The distribution is highly leptokurtic (peaked)
Shapiro-Wilk test p-value: 0.0
The data is not normally distributed



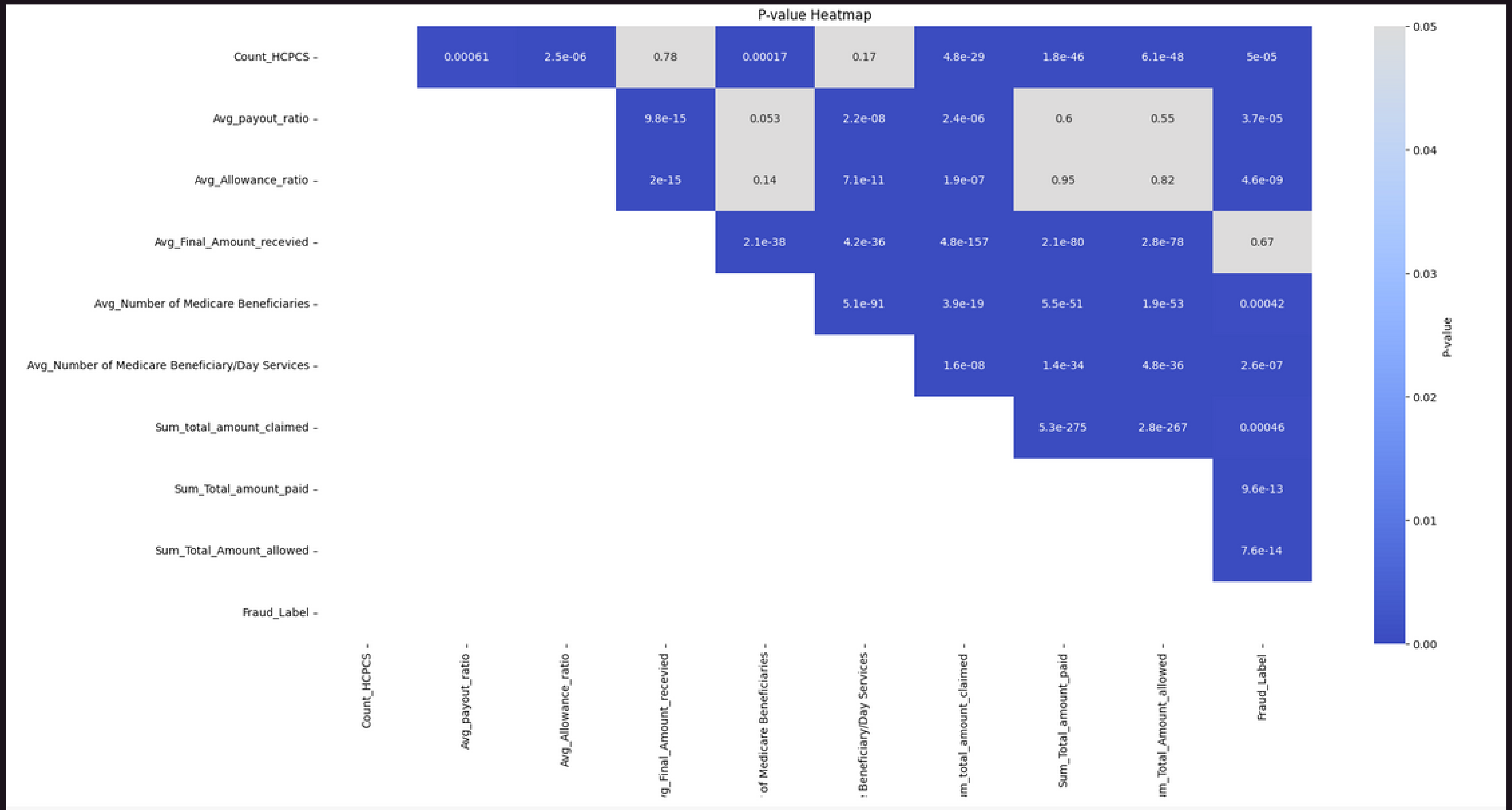
Column: Avg_Final_Amount_recevied
Skewness: 8.024640336217038
The distribution is highly right-skewed
Kurtosis: 82.93402402651303
The distribution is highly leptokurtic (peaked)
Shapiro-Wilk test p-value: 0.0
The data is not normally distributed



Data Set QA | Bi Variate Analysis



- Sum_Total_amount_claimed, Sum_Total_amount_paid, and Sum_Total_amount_allowed were highly correlated.
- Avg_Payout_Ratio and Avg_Allowance_Ratio were highly correlated.



Data Set QA | Bi Variate Analysis

Record	Report
1	**Welch's Two Sample t-test(s) of Avg_final_amount_recevied by Fraud_Label**
2	

Test	t-Statistic	Degrees of Freedom	p-Value
0 vs 1	-0.432438	434.34	0.66564

Record	Report
1	**Welch's Two Sample t-test(s) of Avg_excess_amount_claimed by Fraud_Label**
2	

Test	t-Statistic	Degrees of Freedom	p-Value
0 vs 1	0.5137	453.75	0.60771

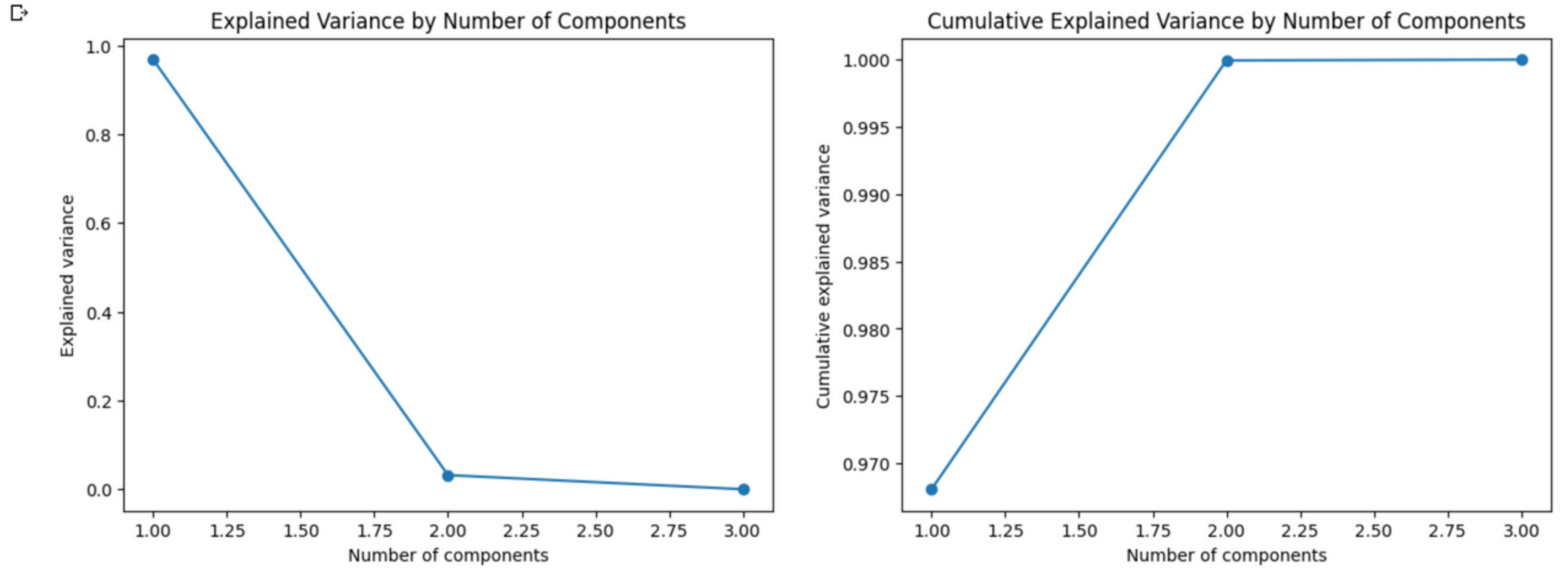
T-tests were also performed to examine the relationship between numerical variables and the Fraud Label. It was found that 'Avg_Final_amount_received' and 'Avg_excess_amount_claimed' were not statistically significant predictors of fraud. Consequently, these variables will not be included in the final model.

Data Set QA | Bi Variate Analysis | PCA

1 to 8 of 8 entries Filter

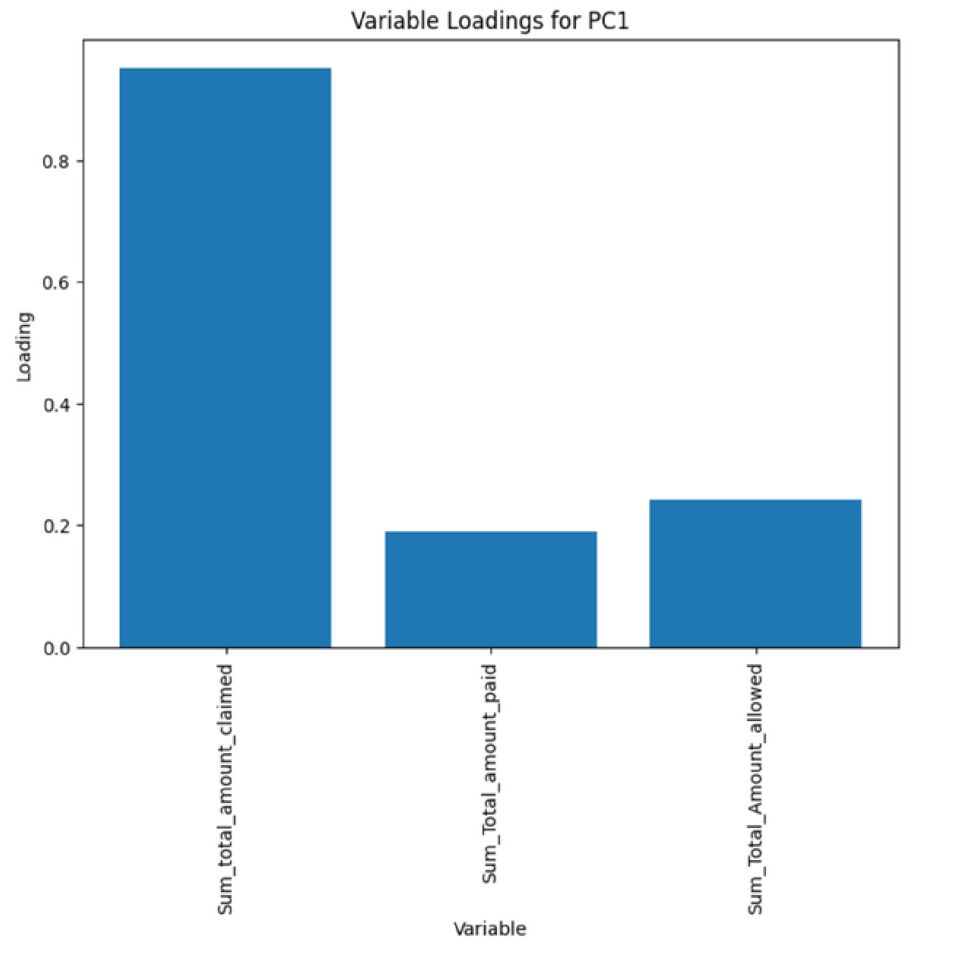
index	Sum_total_amount_claimed	Sum_Total_amount_paid	Sum_Total_Amount_allowed
min	221.0	39.0	39.0
count	1000.0	1000.0	1000.0
25%	21193.5	7706.595214843755	10827.6897583008
50%	78975.7247619629	25104.600818634048	34501.518310546904
mean	215038.1481562014	67263.29968508164	88175.76575987088
75%	201916.51086425775	71182.099029541	94100.47508287427
std	616180.4488146633	140756.09576939547	180275.97418805782
max	13087962.8164063	2155987.5447998	2702009.57373047

Show 25 per page
Like what you see? Visit the data table notebook to learn more about interactive tables.



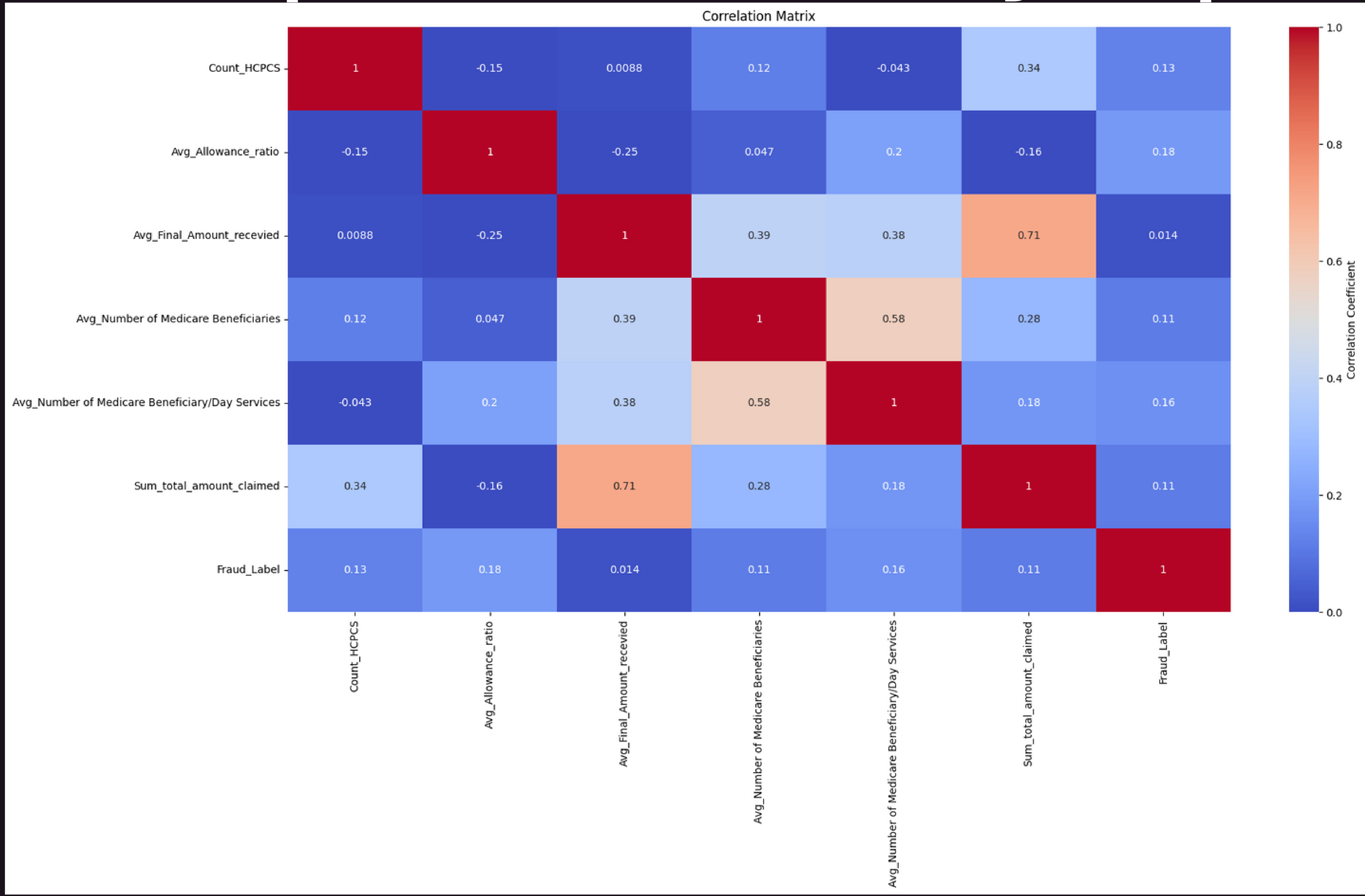

```
[40] # Select the number of components that explain at least 80-90% of the variance
cumulative_explained_variance = np.cumsum(pca.explained_variance_ratio_)
n_components = np.argmax(cumulative_explained_variance >= 0.8) + 1
print(f'Number of components to explain 80% of variance: {n_components}')

Number of components to explain 80% of variance: 1
```



SUM_TOTAL_AMOUNT CLAIMED and avg_allowance_ratio was used for further modelling.

Data Set QA | Bi Variate Analysis | PCA



Based on the Final Correlation matrix, the final features were selected:

Gender, Sum_total_amount_claimed, avg_number_of_medical services/day services, Count_HCPS

Feature Selection for Final Model

Claims Amount

- Avg_Allowance_Ratio
- Sum_Total_amount_claimed

Provider Info

- Gender
- Avg_Number of Medicare Beneficiary/Day Services
- Count_HCPCS

Considered Gender after performing the CHI Square test.

I should have grouped providers by type of service. Didn't do it.

Model Comparison Report

Fit and error measures

Model	Accuracy	F1	AUC	Accuracy_0	Accuracy_1
Support_Vector_Machine	0.7800	0.2326	0.6500	0.9573	0.1515
Random_Forest	0.7400	0.3390	0.7077	0.8632	0.3030
Decision Tree	0.7500	0.3697	0.6568	0.8675	0.3333
Test_Model	0.7633	0.3604	0.7164	0.8932	0.3030

Model: model names in the current comparison.

Accuracy: overall accuracy, number of correct predictions of all classes divided by total sample number.

Accuracy_[class name]: accuracy of Class [class name] is defined as the number of cases that are **correctly** predicted to be Class [class name] divided by the total number of cases that actually belong to Class [class name], this measure is also known as *recall*.

AUC: area under the ROC curve, only available for two-class classification.

F1: F1 score, 2 * precision * recall / (precision + recall). The *precision* measure is the percentage of actual members of a class that were predicted to be in that class divided by the total number of cases predicted to be in that class. In situations where there are three or more classes, average precision and average recall values across classes are used to calculate the F1 score.

Gradient Boosting is the test_model above.

Deals with skewed and outliers effectively. Reason for High AUC and Accuracy