Towards optimal audit selection for sales tax compliance

Trevor Chan1,2, Cheng-En Tan1,2, Jae Choi3, and Ilias Tagkopoulos1,2\*

1Department of Computer Science, University of California, Davis 95616, USA

2Genome Center, University of California, Davis 95616, USA

3California Department of Tax and Fee Administration (CDTFA)

\*Corresponding author (itagkopoulos@ucdavis.edu)

Supplementary materials and results

Table of Contents

[1. HYPOTHESIS TESTING 3](#_Toc42844242)

[2. TAX PREDICTION PIPELINE 3](#_Toc42844243)

[2.1 Overview 3](#_Toc42844244)

[2.2 Dataset 3](#_Toc42844245)

[2.2.1 Filtration 4](#_Toc42844246)

[2.2.2 Visualization 4](#_Toc42844247)

[2.3 Classification Model 4](#_Toc42844248)

[2.3.1 Incorporating All Features 5](#_Toc42844249)

[2.3.2 Without Filtration 5](#_Toc42844250)

[2.3.3 Normalization Methods 5](#_Toc42844251)

[2.4 Regression Model 6](#_Toc42844252)

[2.4.1 Single Regressor Model 6](#_Toc42844253)

[2.4.2 Ensemble Regressor Model 6](#_Toc42844254)

# **1. HYPOTHESIS TESTING**

We examined the following five hypotheses created by the California Department of Tax and Fee Administration’s (CDTFA) auditors:

1. Businesses with gross sales lower than a threshold (calculated as the average in area based on NAICS code, zip ID, city ID) have higher audit yield than businesses with gross sales higher than said threshold
2. Individual taxpayers are more likely to have positive audits compared with corporations
3. Businesses with the taxable sales/gross sales ratio less than a threshold have higher audit yield than businesses with said ratio higher than said threshold
4. Variance of reported gross sales (in a specific period) lower than some threshold have higher audit yield than variance of reported gross sales higher than said threshold
5. Late return penalties that are higher than a threshold have higher audit yield than late return penalties lower than said threshold

As our dataset consists of multiple tax returns per business (every entry for a given audit period had the same audit yield), to prevent overrepresentation of businesses with more return information, data was summed and normalized by the amount of returns per audit period. **Tables S7, S8, 1** and **2** show statistical validation of the hypotheses and their inverses. We note that while we cannot draw conclusions about hypothesis 3, certain threshold values of the taxable sales, gross sales ratio can be statistically shown to exhibit higher audit yields (**Table S10**).

# **2. TAX PREDICTION PIPELINE**

## **2.1 Overview**

Provided by a cohesive audit dataset by the California Department of Tax and Fee Administration (CDTFA), we have created a pipeline to generate a comprehensive compendium and predict tax audit yield for a given business. This occurs in several steps (**Figure 1**). First the data is preprocessed via refinement criteria, normalization and outlier filtration. Features are also filtered based on correlation with other features. Second, the preprocessed dataset is fed as input to a classifier trained on data with audit information to identify which audit leads may or may not be positive. Lastly, audits classified as positive are fed as input to a regressor trained on data with positive audit yield, and audit yield predictions are generated.

## **2.2 Dataset**

The California Department of Tax and Fee Administration provided us with a dataset of 108,162 unique tax records and 42 features. A more concise description of these features can be found in **Table S1** and **Figure S1.** Of these 42 features, only 25 were used in the final dataset, as it was decided by the CDTFA that this study would only concern single-location businesses. Additionally, datetime data was not particularly beneficial as we did not use a time-series approach to analyze data (**Table S2)** We hope to extend our study to accommodate these features in the future.

### **2.2.1 Filtration**

Both feature and outlier filtration were employed. Regarding feature filtration, Pearson correlation was used to measure the level of correlation between features. Features that exhibited high correlation (Pearson correlation greater than 0.95) were removed, as they do not contribute more to the data than their correlated counterparts [1] (**Figure S2**). Outlier filtration was done by first visualizing the distribution of each numerical feature (**Figures S7**), and then removing data above the 95% confidence interval for each feature. The purpose of data distribution visualization was to remove extreme data points that could detriment analysis. For example, in Total Gross Sales values range from $0 to $20,000,000, with most of the data points falling into the range of $0 to $5,000,000. Incorporating points outside of this range gives an inaccurate interpretation of the data and can produce predictive errors in analysis. **Figure S8** shows the same distribution graphs after these data points have been removed.

### **2.2.2 Visualization**

Data was first hierarchically clustered and plotted using Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE). First all returns were used, however this led to poor visualization and clustering results (**Figure S13**). Merging returns by audit period, as well as outlier removal improved clustering and visualization results. (**Figure S15, Figure 2**).Based on the generated clusters, association rule mining via the Apriori algorithm [2] was employed to extract rules for each cluster. **Table S4** shows the number rules extracted per cluster, while **Table S5** lists the top ten rules per cluster. As both PCA and t-SNE plots had difficulties exhibiting the characteristics of tax records with positive audit yield, feature-wise quantile normalization was applied to the dataset, orderings of the features were evaluated. Some clusters based on these orders can be observed in the t-SNE plot (**Figure S11**). Each cluster represents a specific order of the 20 numeric features. In **Figure S12,** each cluster in the t-SNE plot can be interpreted as the order of 20 numerical features in returns. For example, the most majority patterns (the red cluster in **Figure** **S12**) contain 6,894 returns with the ordering of 20 numerical features are as follows:

This means that in these 6,894 returns, all sales are taxable. In addition, there are no deductions, interest, or late penalty in these returns.

## **2.3 Classification Model**

All classification models were evaluated using ten-fold cross-validation on a dataset split 90%-10% between training/testing and validation sets. The Adam optimizer was used, with binary cross-entropy as the loss function [**3**]. L2 regularization was used to reduce overfitting. For early stopping, a delta of 0.0001 and patience, the number of epochs with no improvement after which training is stopped, of 10 was used [**4**]. Due to high class-imbalance, a combination of over-sampling and under-sampling was performed on each fold [**5**]. Random search was employed to find optimal hyperparameters on a feed-forward artificial neural network with an input layer where the number of input nodes was equivalent to the number of input features, 2 hidden layers, and one output layer with one output node. The sigmoid function was used as an activation for the nodes in each layer.

### **2.3.1 Incorporating All Features**

We first attempted to use all data features, this included all 143 city and zip code features. Of the 6,301 tax records, 3,688 records above the 95% confidence interval were removed and 38 features were removed with 0 standard deviation, giving a dataset of 2,351 tax records for 510 unique businesses. Parameter tuning with random search yielded 53 and 14 hidden nodes per hidden layer, as well as a batch size of 32 and a dropout rate of 40% were used. Training this architecture with the proposed dataset led to overfitting and poor results on the validation set (**Figure S19**). Model training and testing accuracy over all 10 folds was 97% with a standard deviation of 4%, and 63% with a standard deviation of 12%, respectively. The training and testing F1 scores were 0.67 and 0.28, while validation accuracy was 49% with an F1 score of 0.20.

### **2.3.2 Without Filtration**

We also experimented with a non-filtered approach on the dataset. All 6,301 tax records were used, along with all 37 features for 510 unique businesses. Parameter tuning with random search yielded 23 and 11 hidden nodes per hidden layer, and a batch size of 64 with a dropout rate of 30%. A model trained on this dataset yielded training and testing accuracies of 55% with standard deviation of 5 % and 53% with a standard deviation of 10%, respectively. The training and testing F1 scores were 0.38 and 0.35, while validation accuracy was 54% with an F1 score of 0.21. **Figures S20** and **S21** shows a significant decrease in performance when compared with a model trained on filtered data (**Figure 2**).

### **2.3.3 Normalization Methods**

Besides minmax, quantile and z-score normalization were also tested on the data. PCA and t-SNE were both conducted on the normalized data to better visualize the effects of normalization (**Figure S9** and **Figure S10**). Normalization methods were evaluated based on their effects on classifier performance. Tuned hyperparameters for normalizations can be found in **Table S12**.

## **2.4 Regression Model**

### **2.4.1 Single Regressor Model**

Like the audit lead classifier, a simple feed-forward artificial neural network was initially constructed for audit yield regression. It was evaluated using ten-fold cross-validation on a dataset split 90%-10% between training/testing and validation sets. The Adam optimizer was used with mean squared error as the loss function [**2**]. L2-regularization was used to reduce overfitting. For early stopping, a delta of 0.0001 and patience, the number of epochs with no improvement after which training is stopped, of 10 was used [**3**]. Random search was employed to find optimal hyperparameters on a feed-forward artificial neural network with a 30-node input layer, 2 hidden layers, and one output layer with one output node. The Rectified Linear Unit (ReLU) function was used as an activation for the nodes in each layer, with exception of the linear function which was used as an activation for the output node. Training and testing R-squared values were 0.10 and 0.05, respectively, while validation R-squared was 0.02. Due to data limitation, this model performed poorly (**Figure S23**) and our approach needed to be adapted.

### **2.4.2 Ensemble Regressor Model**

Data was split based on validated hypothesis rules and categorical variables to create four distinct ensemble models. For validated hypotheses 1 and 5, the gross sales and late penalty hypotheses, data was split above and below the mean of gross sales and late penalty values. Splits for categorical variables were done on business type and NAICS codes, and not on categorical variables such as city ID or zip code as there were too many variables with too few tax records. For splits based on business type, data was split between individual, corporation, and the remaining business entities. More splits were not done due to lack of data – there were 675 corporation tax records, 437 individual tax records, and 382 other tax records. Finally, for splits based on NAICS codes data was split between limited service restaurants, mobile food services, and all other NAICS codes. More splits were not done due to the lack of data – there were 550 limited-service tax records, 241 mobile-service tax records, and 207 remaining NAICS code records. The ensemble model based on NAICS codes performed the best, with a validation R-squared values of 0.30, while the ensemble model generated based on business type performed equally as well, with a validation R-squared of 0.29. The ensemble models based on hypotheses 1 and 5 yielded validation R-squared values of 0.15 and 0.12, respectively. Values of tuned hyperparameters were done by random search (**Table S13**).

**REFERENCES**

1. Biesiada J., Duch W. (2007) Feature Selection for High-Dimensional Data — A Pearson

Redundancy Based Filter. In: Kurzynski M., Puchala E., Wozniak M., Zolnierek A.

(eds) Computer Recognition Systems 2. Advances in Soft Computing, vol 45. Springer, Berlin, Heidelberg

2. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules", Proc. of the VLDB

Conference

3. Buja, A., Stuetzle, W. and Shen, Y. Loss functions for binary class probability estimation and

classification: structure and applications. manuscript, available at

www-stat.wharton.upenn.edu/~buja/

4. Lutz Prechelt. Early stopping-but when? Neural Networks: Tricks of the trade, pp. 553–553,

1998.

5. Chawla, N. et al. “SMOTE: Synthetic Minority Over-sampling Technique.” J. Artif.

Intell. Res. 16 (2002): 321-357.

**SUPPLEMENTARY FIGURES**

|  |  |
| --- | --- |
| **NAICS Code** | **North American Industry Classification System (NAICS)** |
| 722513 | Limited-Service Restaurants |
| 722511 | Full-Service Restaurants |
| 722310 | Food Service Contractors |
| 722330 | Mobile Food Services |
| 722515 | Snack and Nonalcoholic Beverage Bars |
| 722400 | Drinking Places (Alcoholic Beverages) – Type 1 |
| 722410 | Drinking Places (Alcoholic Beverages) – Type 2 |
| 722320 | Caterers |
| 722300 | Special Food Services |
| 722514 | Cafeterias, Grill Buffets, and Buffets – Type 1 |
|  |  |
| **Anonymized Selling Location City Name ID** | |
| 25 unique values | |
|  |  |
| **Anonymized Selling Location Zip Code ID** | |
| 113 unique values | |
|  |  |

|  |  |
| --- | --- |
| **Type of Business** | |
| IND | Individual |
| CORP | Corporation |
| LLC | Limited Liability Corporation |
| PART | Partnership |
| LIMITD | Limited |
|  |  |
| **Filing Frequency** | |
| SUTFA | Annual Filer (State Fiscal Year) |
| SUTM | Monthly Filer |
| SUTPQ | Quarterly w/ Prepay Filer |
| SUTQ | Quarterly Filer |

Figure S1: Possible values for the categorical variables in Registration Data – the number of unique values is shown for City ID and Zip ID.

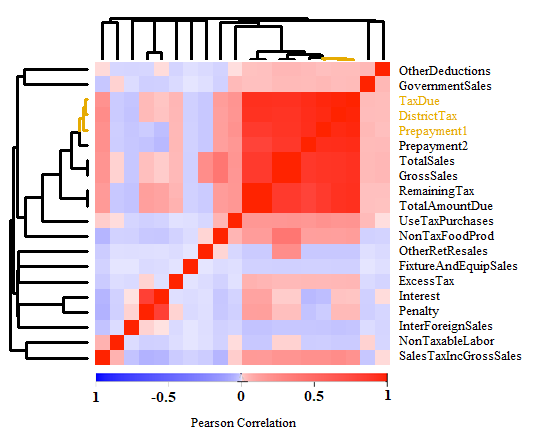
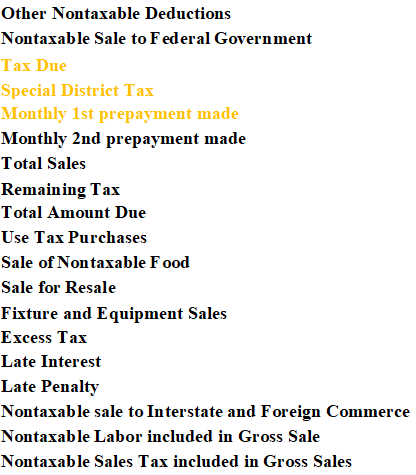


Figure S2: Heatmap showing clusters based on Pearson correlation between the various return data features. Clusters of size greater than one are colored both in label and dendrograms.

|  |  |  |
| --- | --- | --- |
| **Business Type** | **Occurrences** | **Percentage of Total** |
| CORP | 2941 | 46.68% |
| IND | 1965 | 31.19% |
| LIMITD | 24 | 0.38% |
| LLC | 1064 | 16.89% |
| PART | 307 | 4.87% |

Figure S3: Business Type frequency distribution

Figure S4: Filing Frequency frequency distribution

|  |  |  |
| --- | --- | --- |
| **Filing Frequency** | **Occurrences** | **Percentage of Total** |
| SUTFA | 12 | 0.19% |
| SUTM | 399 | 6.33% |
| SUTPQ | 5219 | 82.83% |
| SUTQ | 671 | 10.65% |

|  |  |  |  |
| --- | --- | --- | --- |
| **NAICS**  **Code** | **Definition** | **Occurrences** | **Percentage of Total** |
| 722514 | Cafeterias, Grill Buffets, and Buffets | 48 | 0.76% |
| 722320 | Caterers | 186 | 2.95% |
| 722400 | Drinking Places (Alcoholic Beverages) – Type 1 | 298 | 4.73% |
| 722410 | Drinking Places (Alcoholic Beverages) – Type 2 | 24 | 0.38% |
| 722310 | Food Service Contractors | 107 | 1.70% |
| 722000 | Full-Service Restaurants | 3616 | 57.39% |
| 722511 | Limited-Service Restaurants | 1556 | 24.69% |
| 722513 | Mobile Food Services | 126 | 2.00% |
| 722330 | Snack and Nonalcoholic Beverage Bars | 350 | 5.55% |
| 722515 | Special Food Services | 14 | 0.22% |

Figure S5: NAICS codes frequency distribution

|  |  |  |
| --- | --- | --- |
| **DasCityID** | **Occurrence** | **Percentage of Total** |
| C1 | 48 | 0.76% |
| C10 | 12 | 0.19% |
| C11 | 86 | 1.36% |
| C12 | 233 | 3.70% |
| C13 | 298 | 4.73% |
| C14 | 553 | 8.78% |
| C15 | 206 | 3.27% |
| C16 | 12 | 0.19% |
| C17 | 62 | 0.98% |
| C18 | 87 | 1.38% |
| C19 | 1753 | 27.82% |
| C2 | 270 | 4.29% |
| C20 | 1391 | 22.08% |
| C21 | 96 | 1.52% |
| C22 | 327 | 5.19% |
| C23 | 108 | 1.71% |
| C24 | 192 | 3.05% |
| C25 | 134 | 2.13% |
| C3 | 96 | 1.52% |
| C4 | 72 | 1.14% |
| C6 | 25 | 0.40% |
| C7 | 24 | 0.38% |
| C8 | 78 | 1.24% |
| C9 | 138 | 2.19% |

Figure S6: City ID frequency distribution

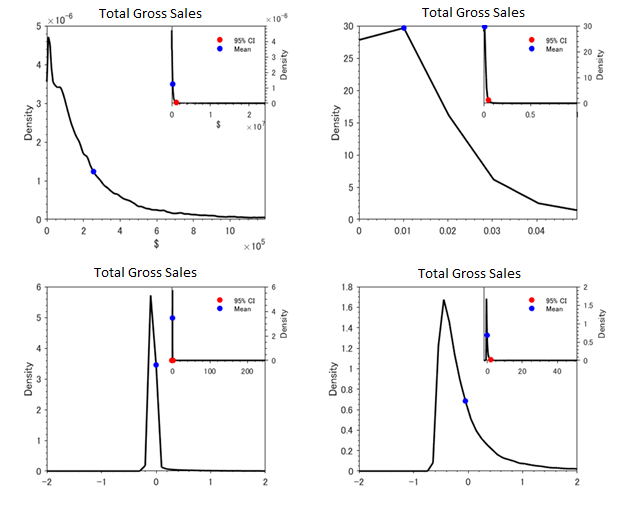
**

Figure S7: Data distribution of Total Gross Sales prior to filtration. Non-normalized (top left), minmax normalization (top right). quantile normalization (bottom left), and z-score normalization (bottom right) are shown above, with red and blue dots representing 95% confidence interval and mean respectively. Inset plots show the entire dataset while the main plots show the inset plots zoomed in to the 95% confidence interval range.

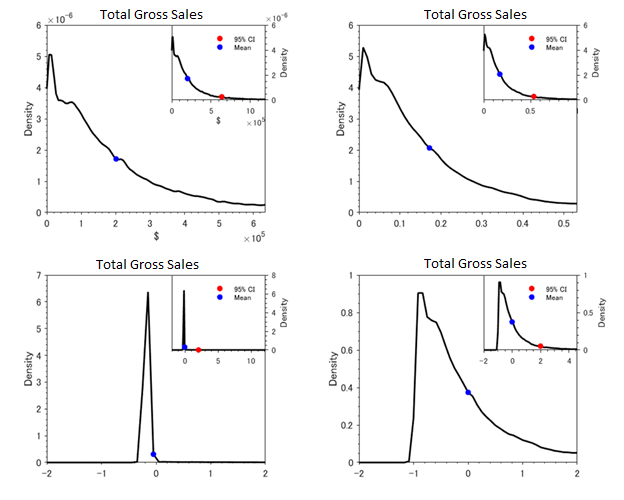


Figure S8: Data distribution of Total Gross Sales post-filtration. Non-normalized (top left), minmax normalization (top right). quantile normalization (bottom left), and z-score normalization (bottom right) are shown above, with red and blue dots representing 95% confidence interval and mean respectively. Inset plots show the entire dataset while the main plots show the inset plots zoomed in to the 95% confidence interval range.

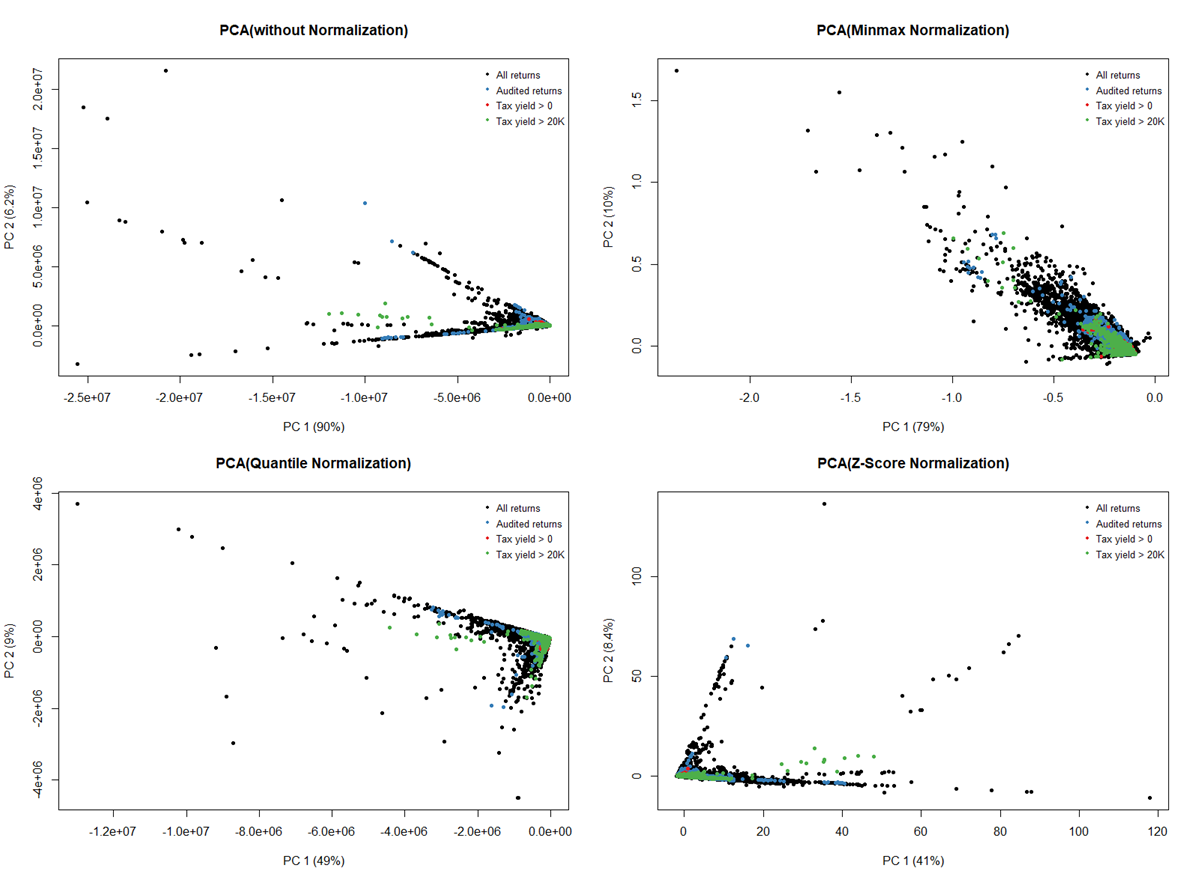


Figure S9: PCA visualization of all 93,413 returns without filtration. Non-normalized (top left), minmax normalization (top right). quantile normalization (bottom left), and z-score normalization (bottom right) are shown above. The returns with different audit outcome are in different color (black: not being audited; blue: being audited but get zero audit yield; red: being audited with audit yield greater than zero; green: being audited with audited yield greater than CDTFA’s threshold).

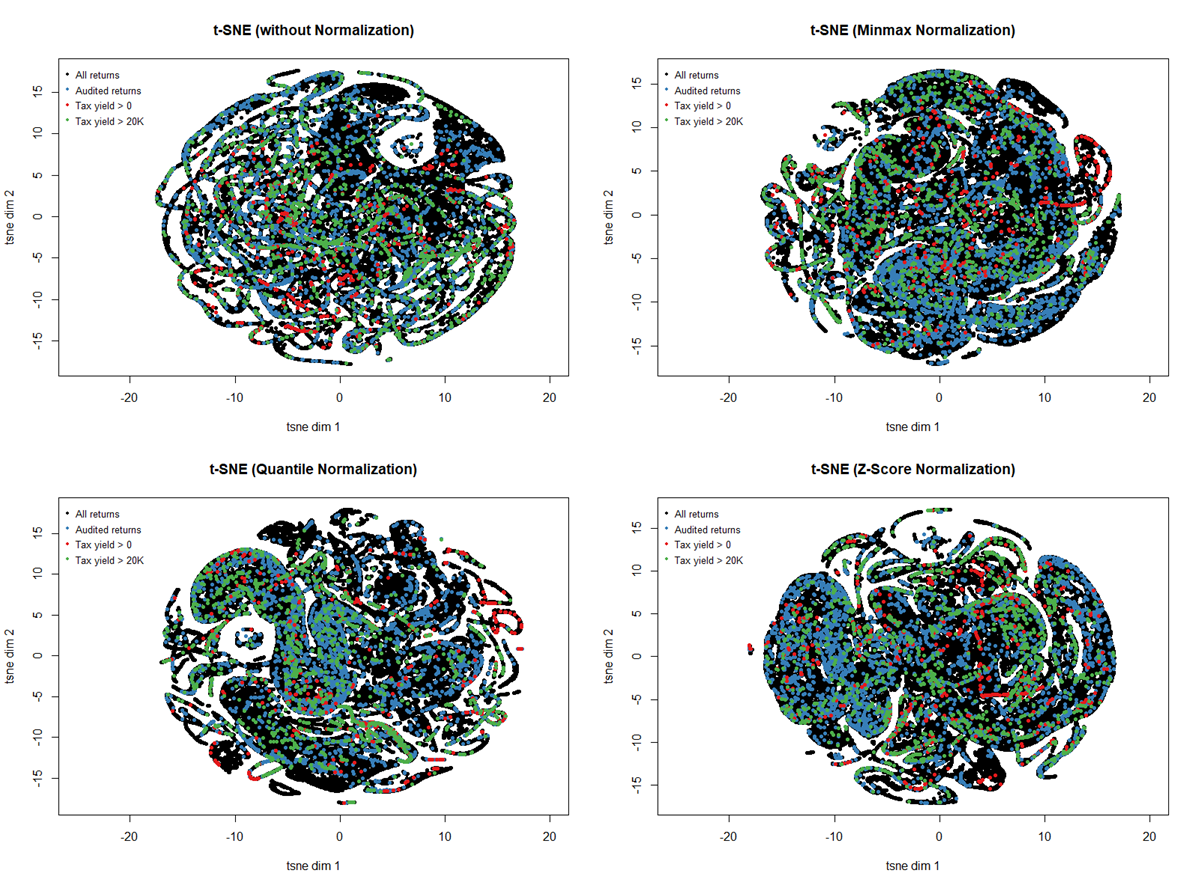


Figure S10: t-SNE visualization of all 93,413 returns without filtration. Non-normalized (top left), minmax normalization (top right). quantile normalization (bottom left), and z-score normalization (bottom right) are shown above. The returns with different audit outcome are in different color (black: not being audited; blue: being audited but get zero audit yield; red: being audited with audit yield greater than zero; green: being audited with audited yield greater than CDTFA’s threshold).

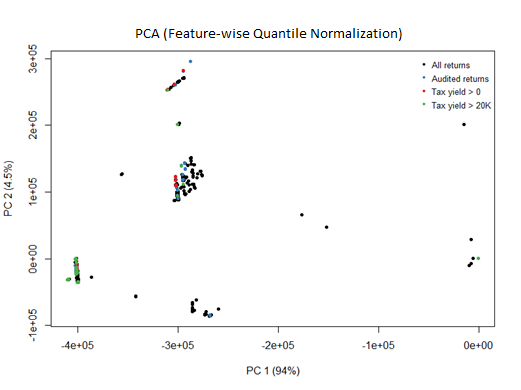
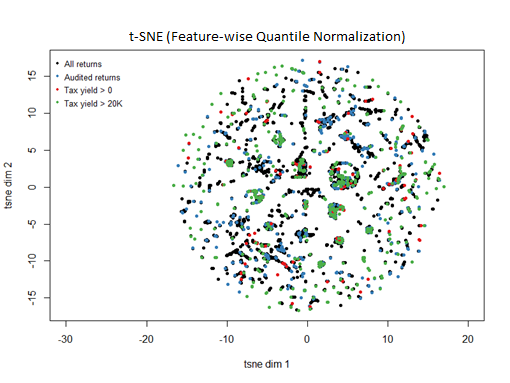


Figure S11: Visualization of all 93,413 returns without filtration (with feature-wise quantile normalization) by PCA (left) and t-SNE (right). The returns with different audit outcome are in different color (black: not being audited; blue: being audited but get zero audit yield; red: being audited with audit yield greater than zero; green: being audited with audited yield greater than CDTFA’s threshold).

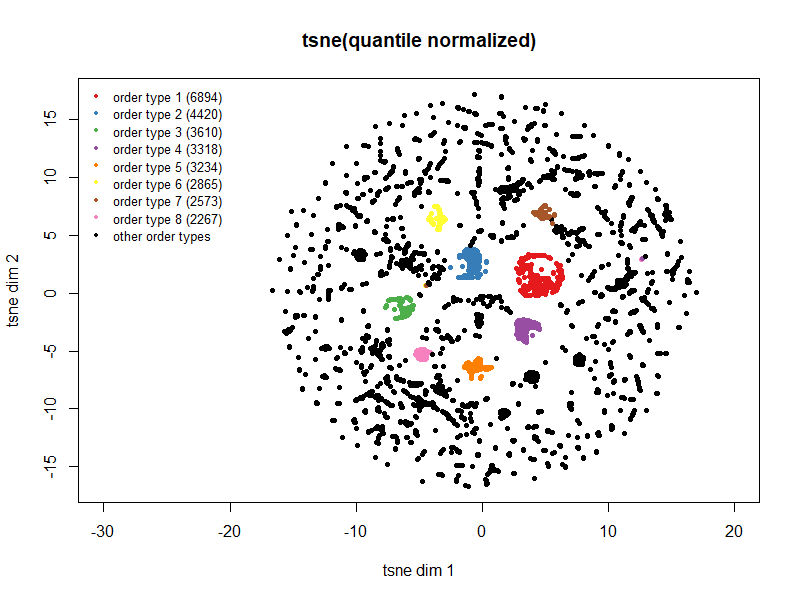


Figure S12: t-SNE plot of 93,413 feature-wise quantile normalized returns. The corresponding to the eight largest clusters of feature order are in represented by different colors.

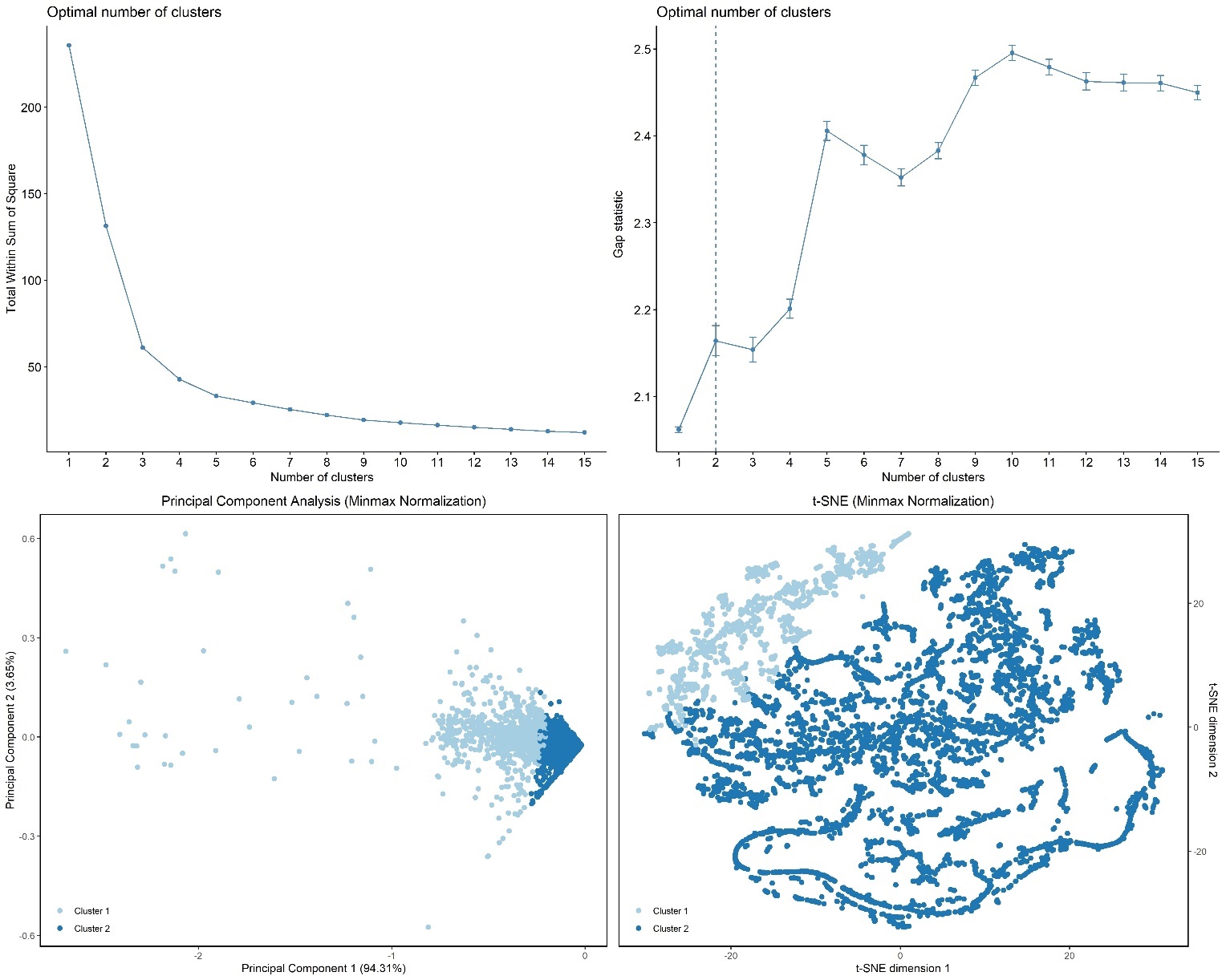


Figure S13: Elbow method (top left) and gap statistic method (top right) conducted on result of hierarchical clustering performed on minmax-normalized PCA data to find the number of clusters. Two clusters were selected as the optimal number of clusters. PCA (bottom left) and t-SNE (bottom right) visualization of the clustered data with minmax normalization post-filtration is shown.

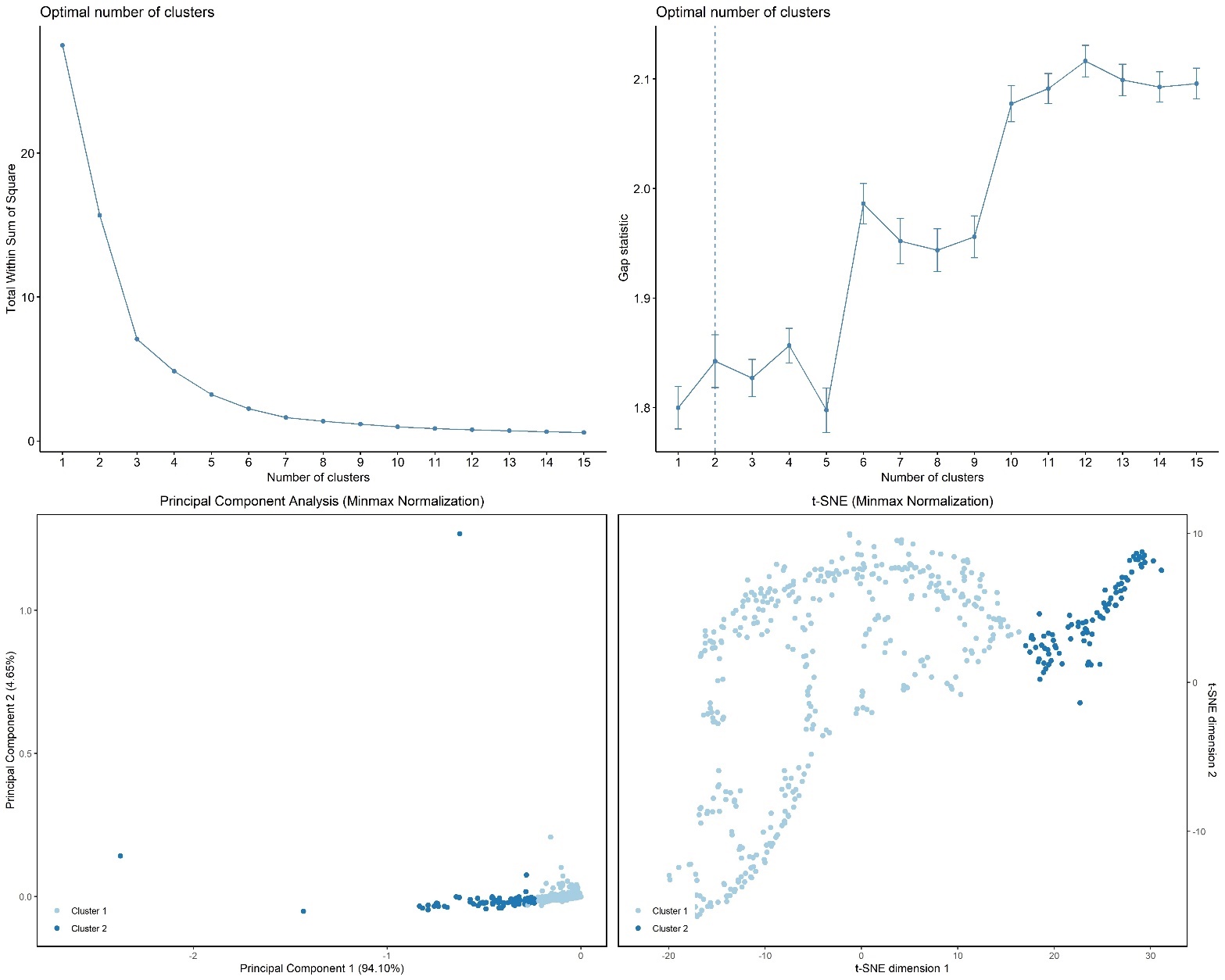


Figure S14: Elbow method (top left) and gap statistic method (top right) conducted on result of hierarchical clustering performed on audit period-merged minmax-normalized PCA data to find the number of clusters. Two clusters were selected as the optimal number of clusters. PCA (bottom left) and t-SNE (bottom right) visualization of the clustered data with minmax normalization post-filtration is shown.

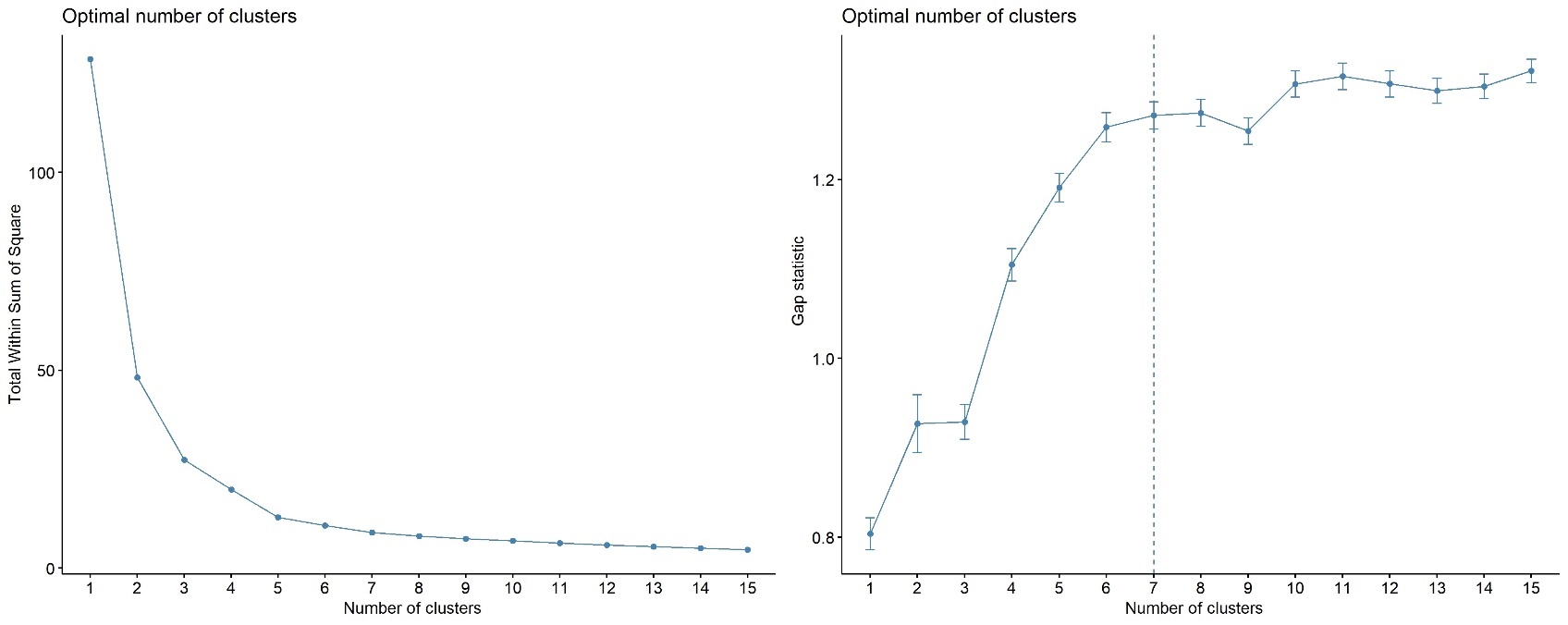


Figure S15: Elbow method (left) and gap statistic method (right) conducted on result of hierarchical clustering performed on audit period-merged PCA data to find the number of clusters. Seven clusters were selected as a conservative estimate for the optimal number of clusters.

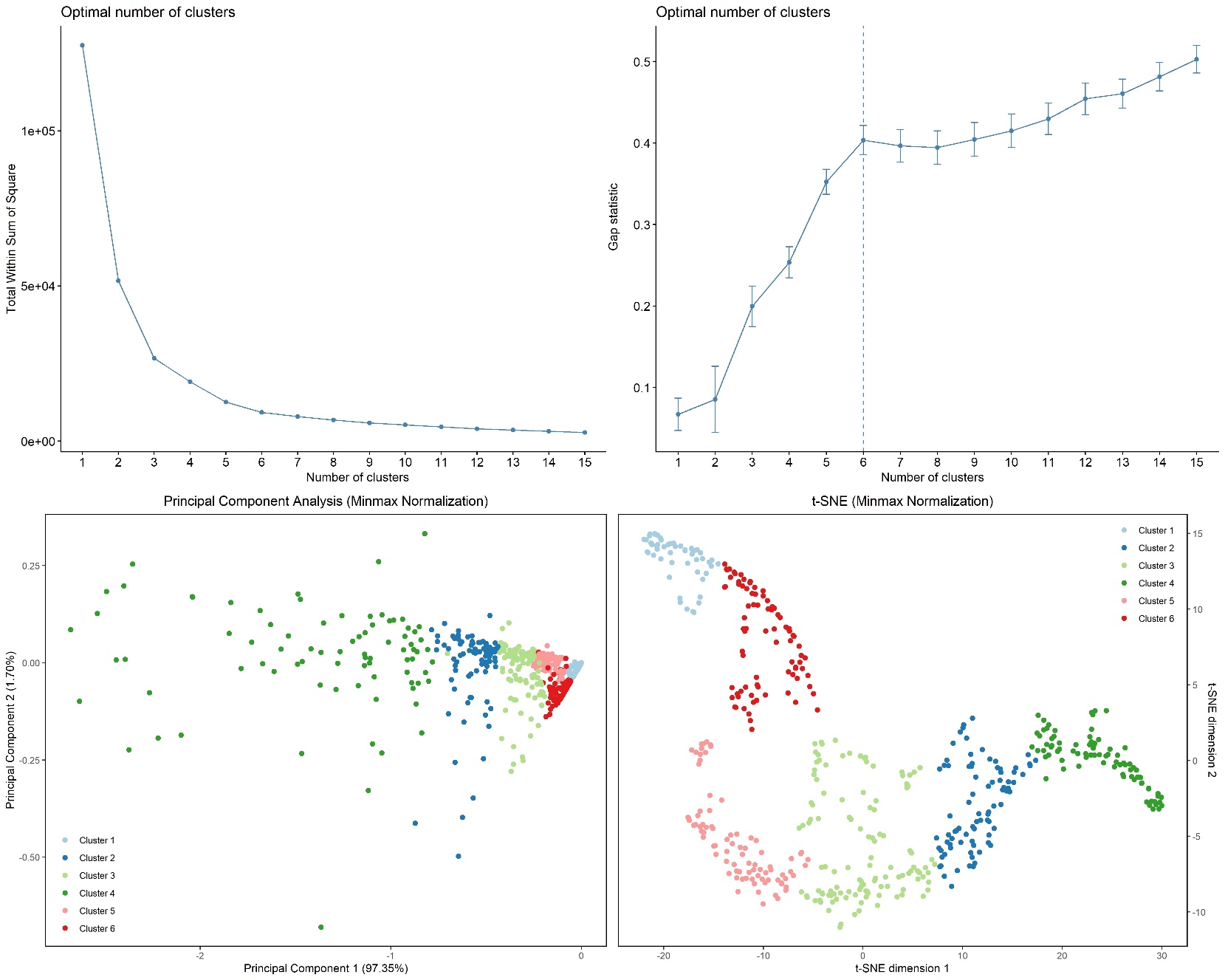


Figure S16: Elbow method (top left) and gap statistic method (top right) conducted on result of hierarchical clustering performed on audit period-merged minmax-normalized t-SNE data to find the number of clusters. Six clusters were selected as a conservative estimate for the optimal number of clusters. PCA (bottom left) and t-SNE (bottom right) visualization of the clustered data with minmax normalization post-filtration is shown.

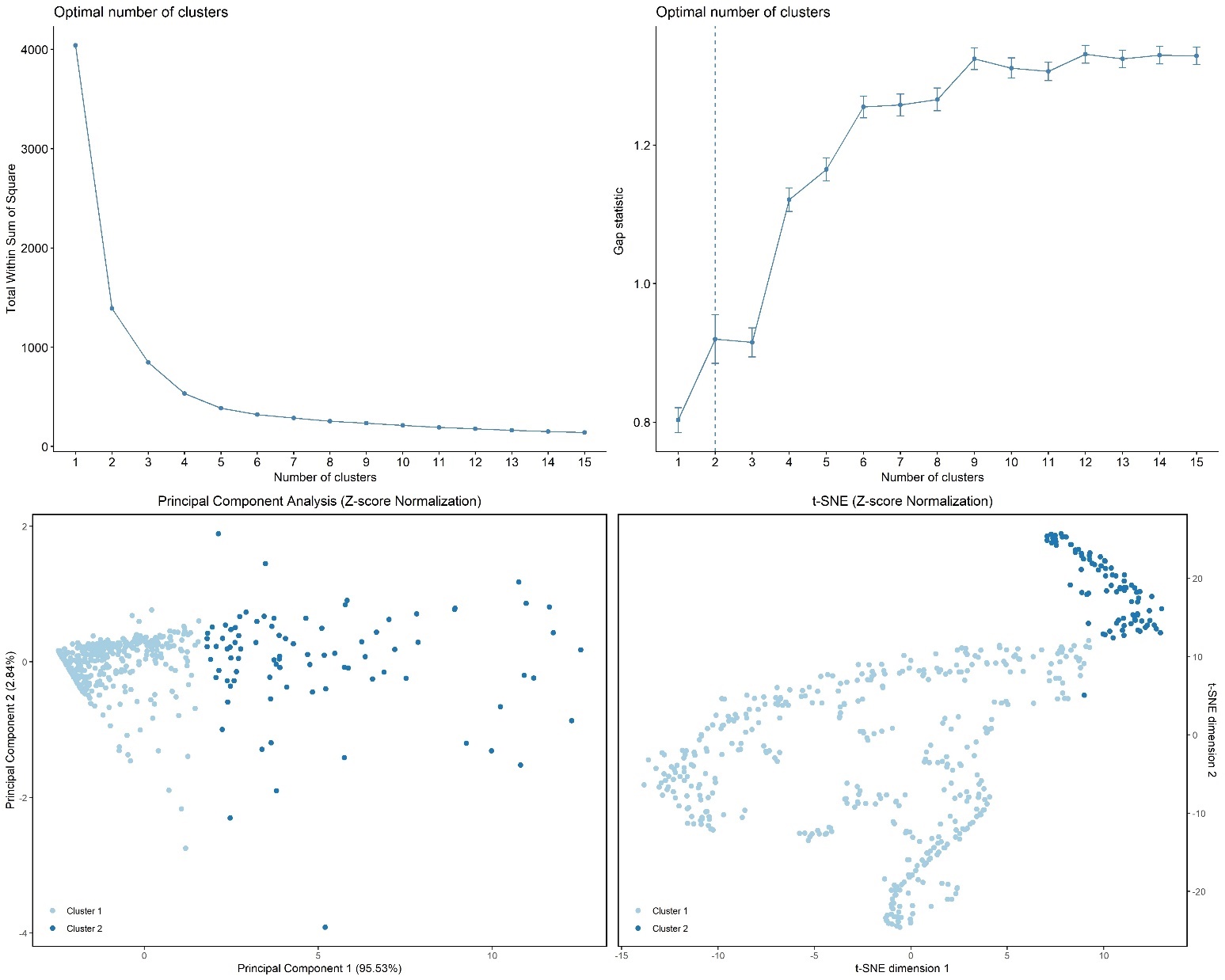


Figure S17: Elbow method (top left) and gap statistic method (top right) conducted on result of hierarchical clustering performed on audit period-merged z-score-normalized PCA data to find the number of clusters. Two clusters were selected as the optimal number of clusters. PCA (bottom left) and t-SNE (bottom right) visualization of the clustered data with z-score normalization post-filtration is shown.

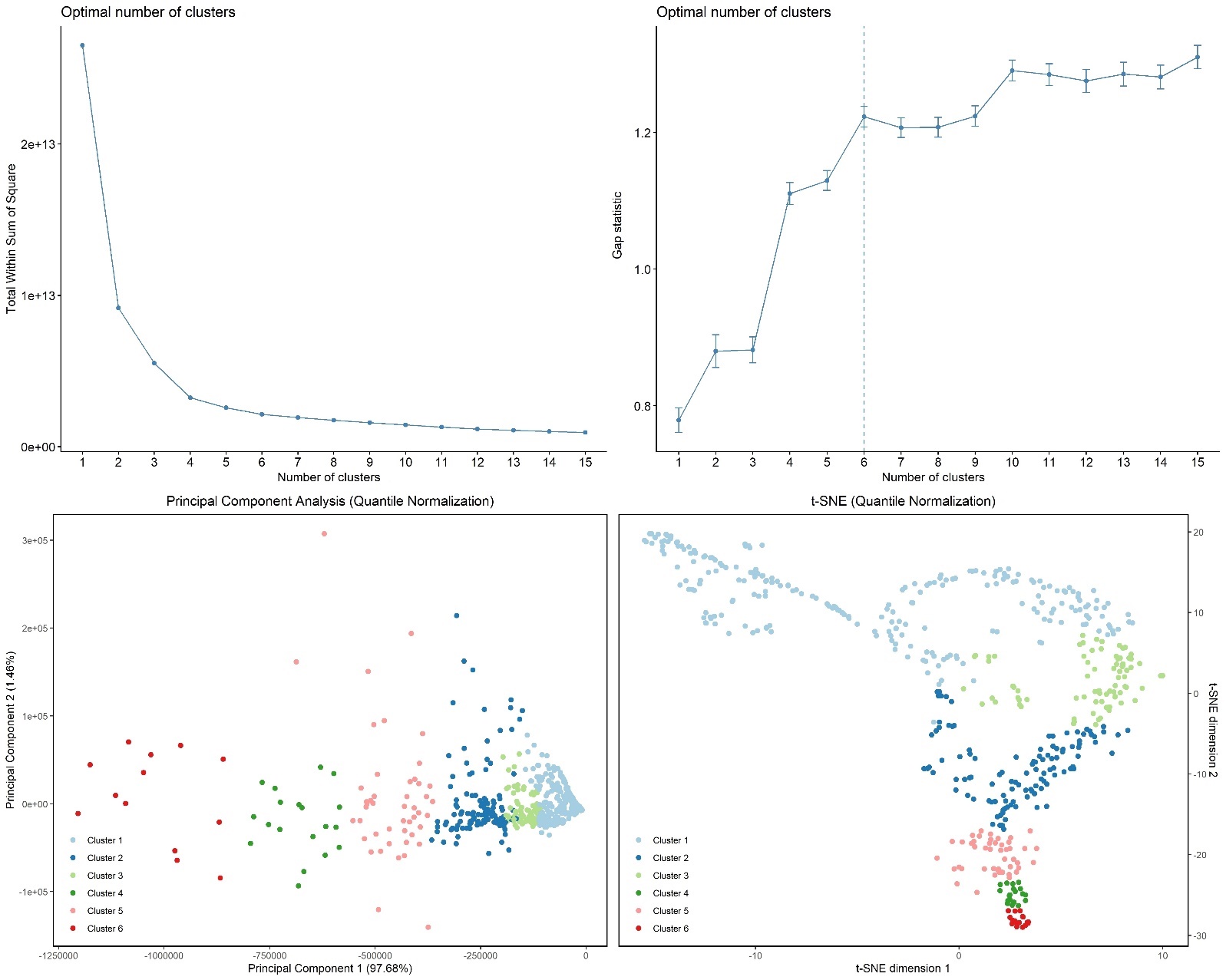


Figure S18: Elbow method (top left) and gap statistic method (top right) conducted on result of hierarchical clustering performed on audit period-merged quantile-normalized PCA data to find the number of clusters. Six clusters were selected as a conservative estimate for the optimal number of clusters. PCA (bottom left) and t-SNE (bottom right) visualization of the clustered data with quantile normalization post-filtration is shown.

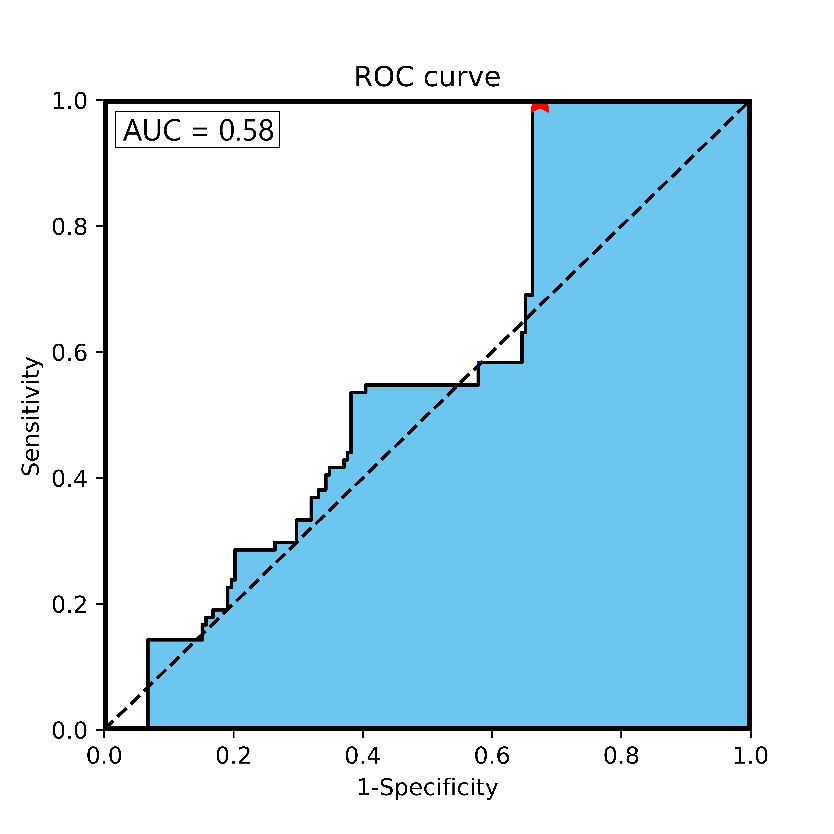
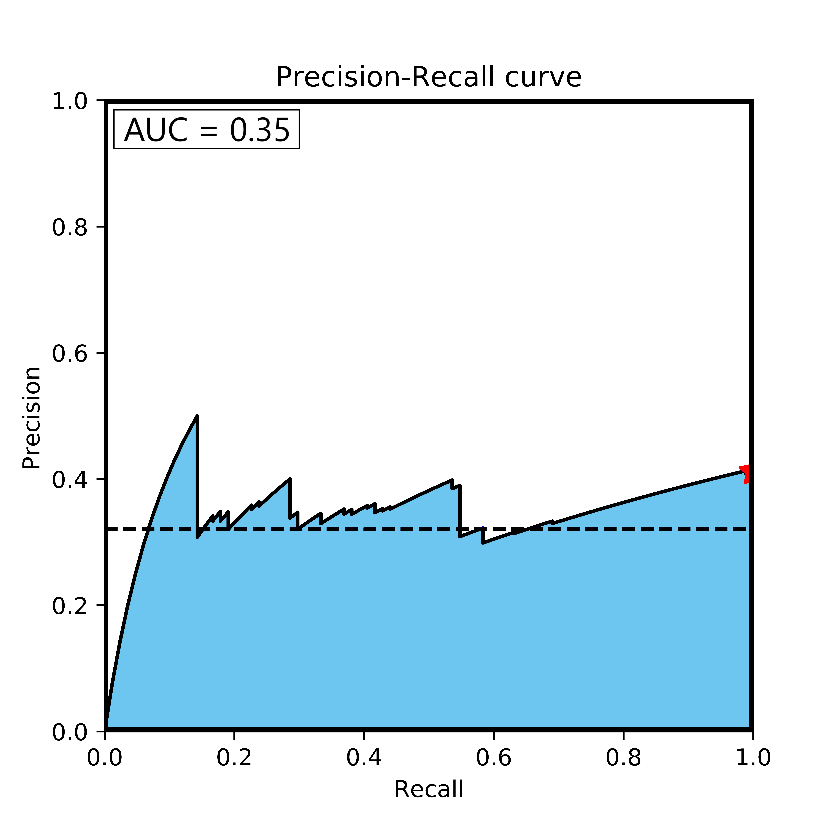
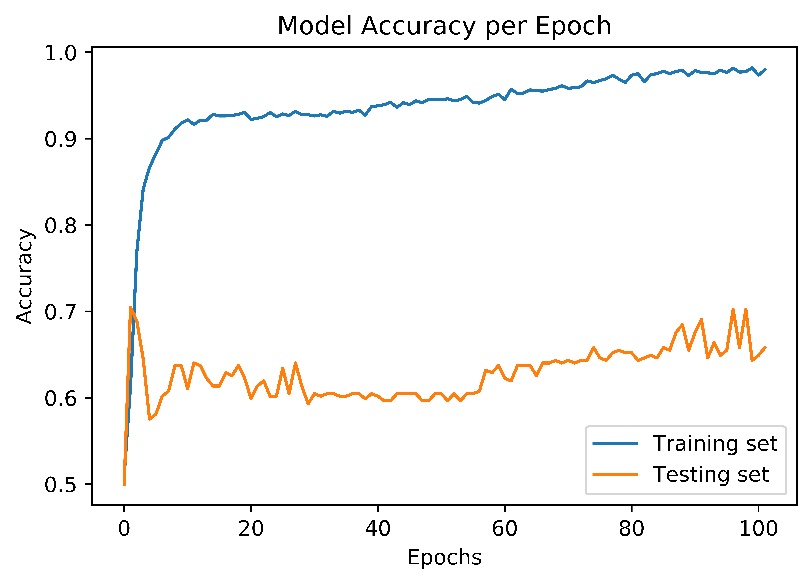


Figure S19: Accuracy (top left) and loss per epoch (top right) graphs for a single cross-validation fold using data with all 135 features exhibiting overfitting, as well as ROC (bottom left) and precision-recall (bottom right) for the validation set.

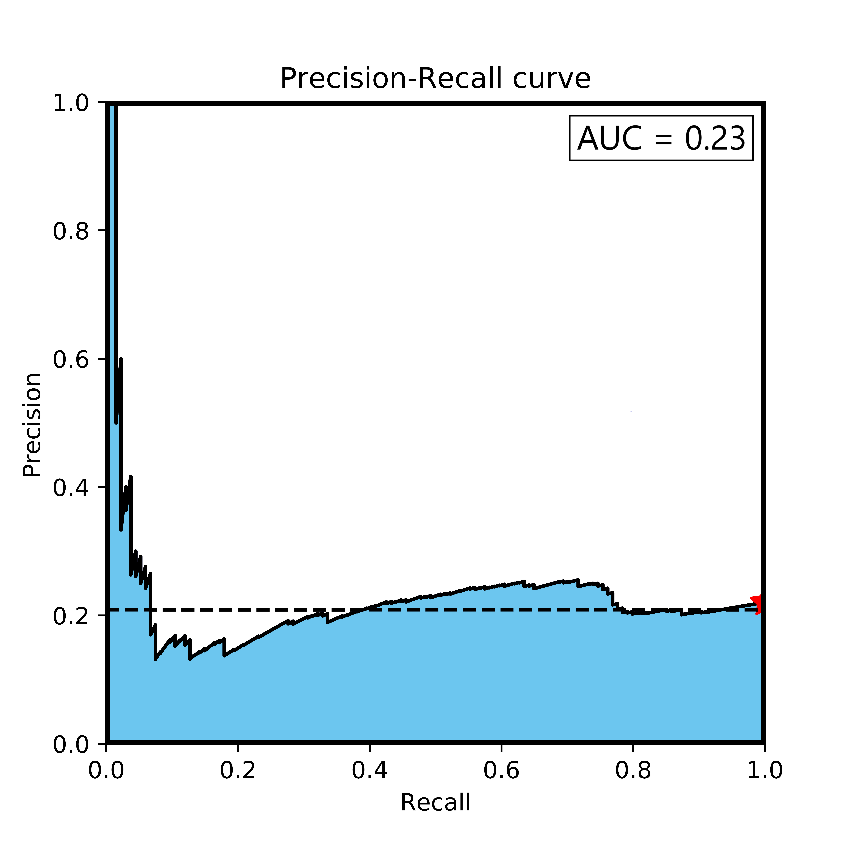
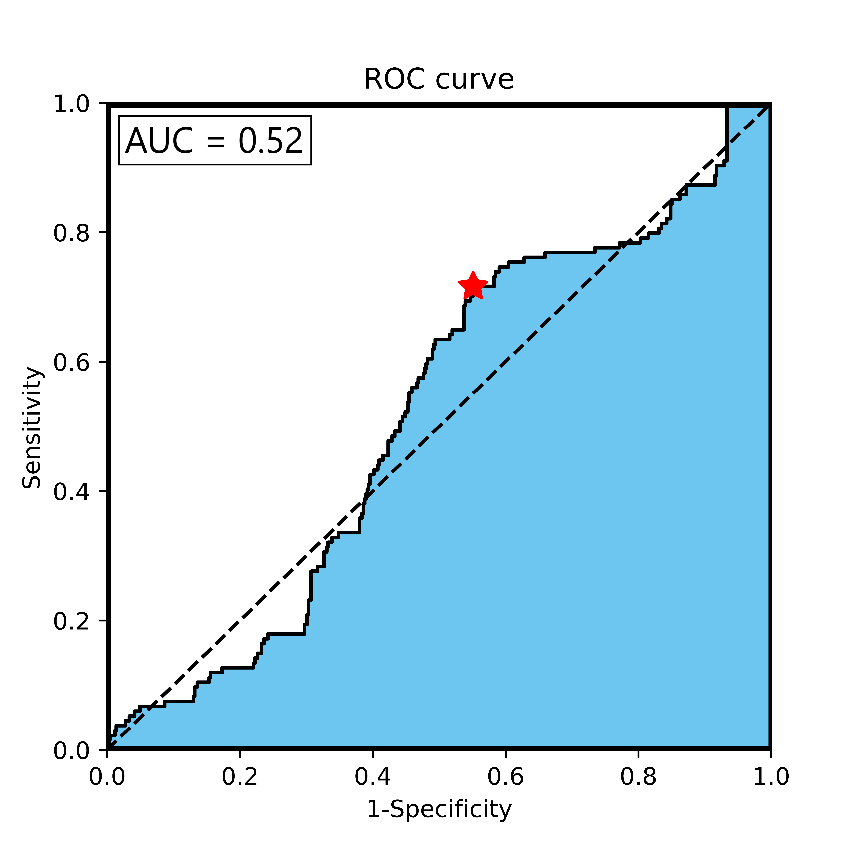


Figure S20: ROC (left) and precision-recall (right) curves for audit lead classifier without filtration of returns outside of the 95% confidence interval.

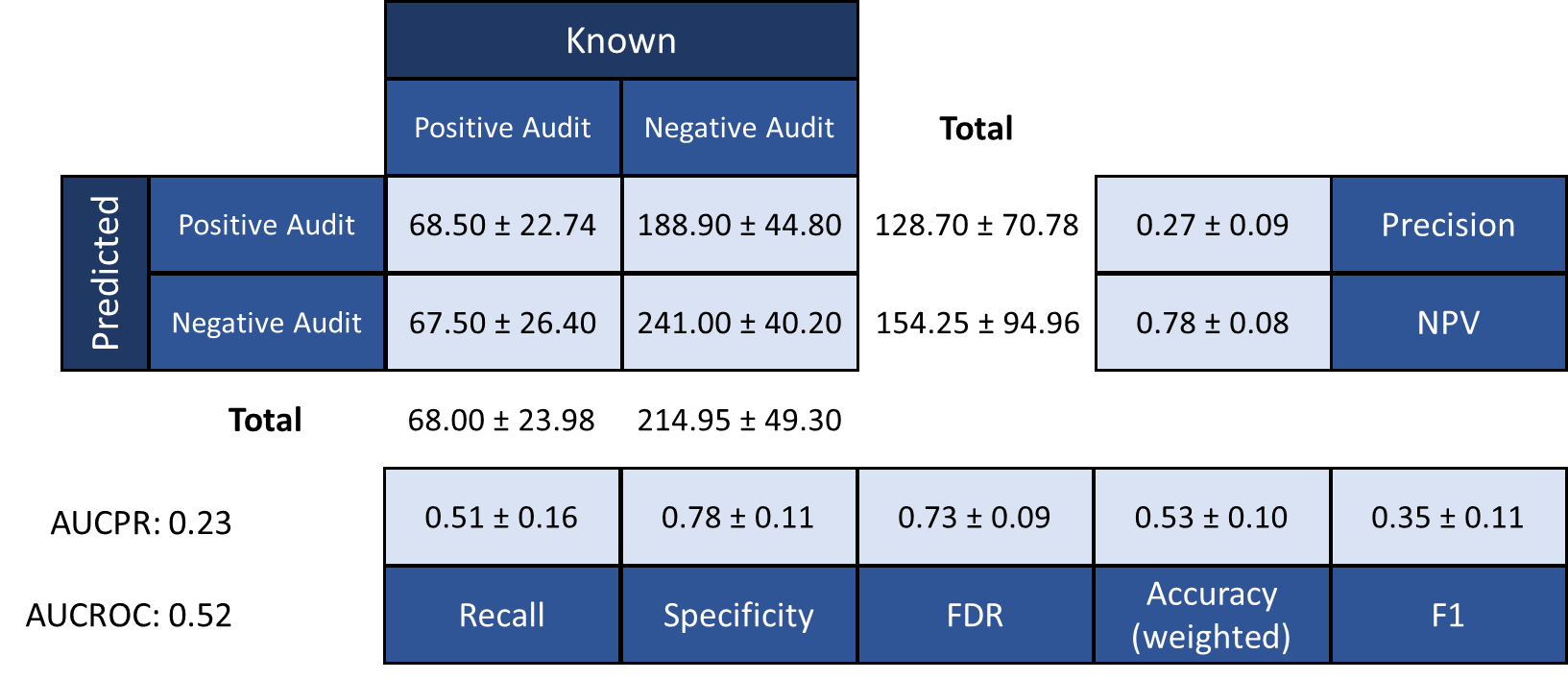
****

Figure S21: Confusion matrix for classifier trained without filtration of returns outside of the 95% confidence interval.

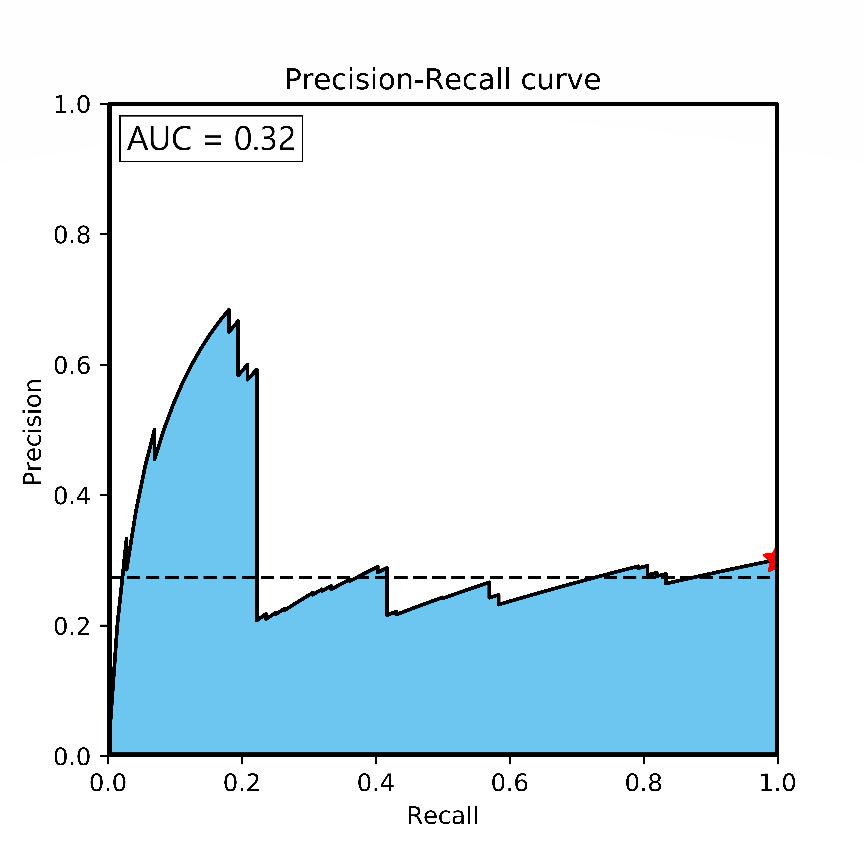
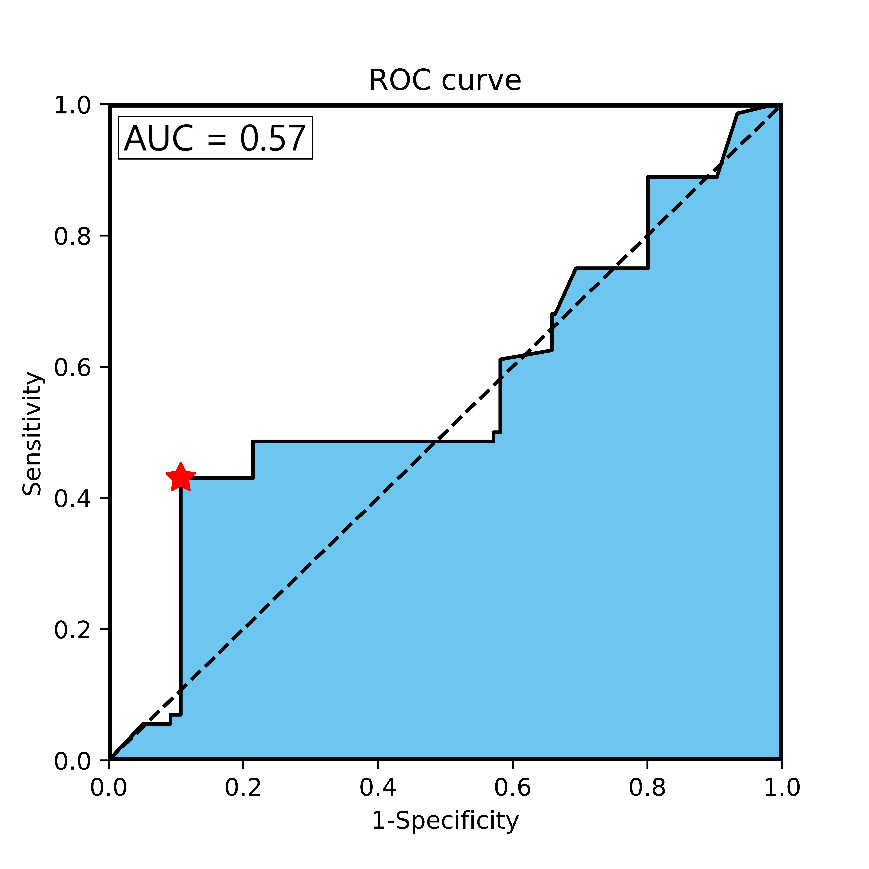
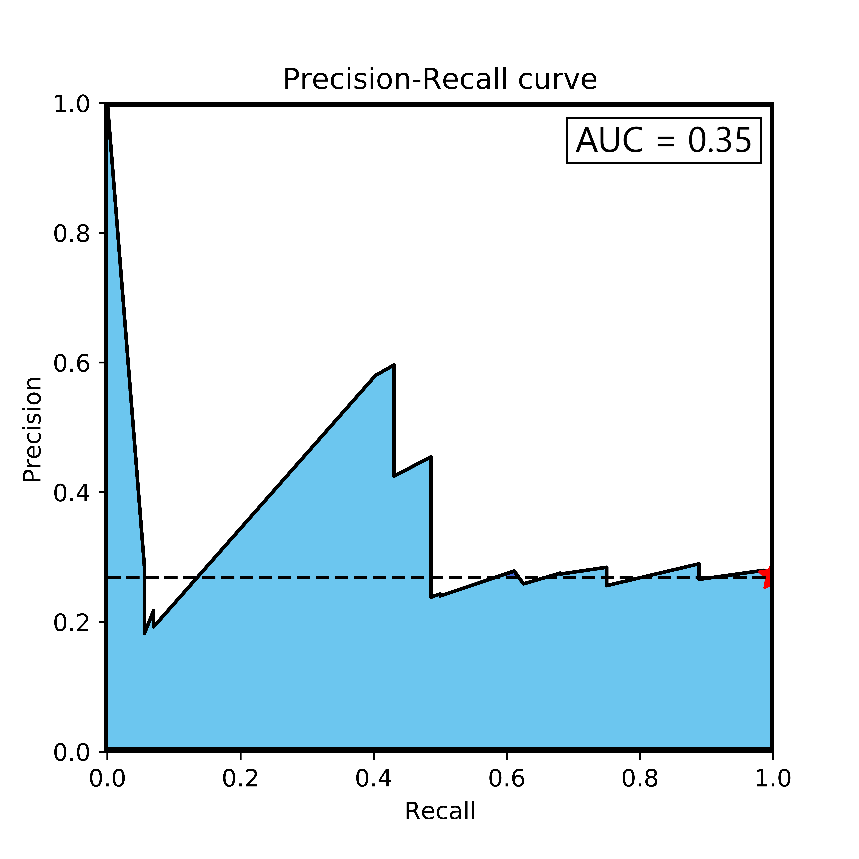
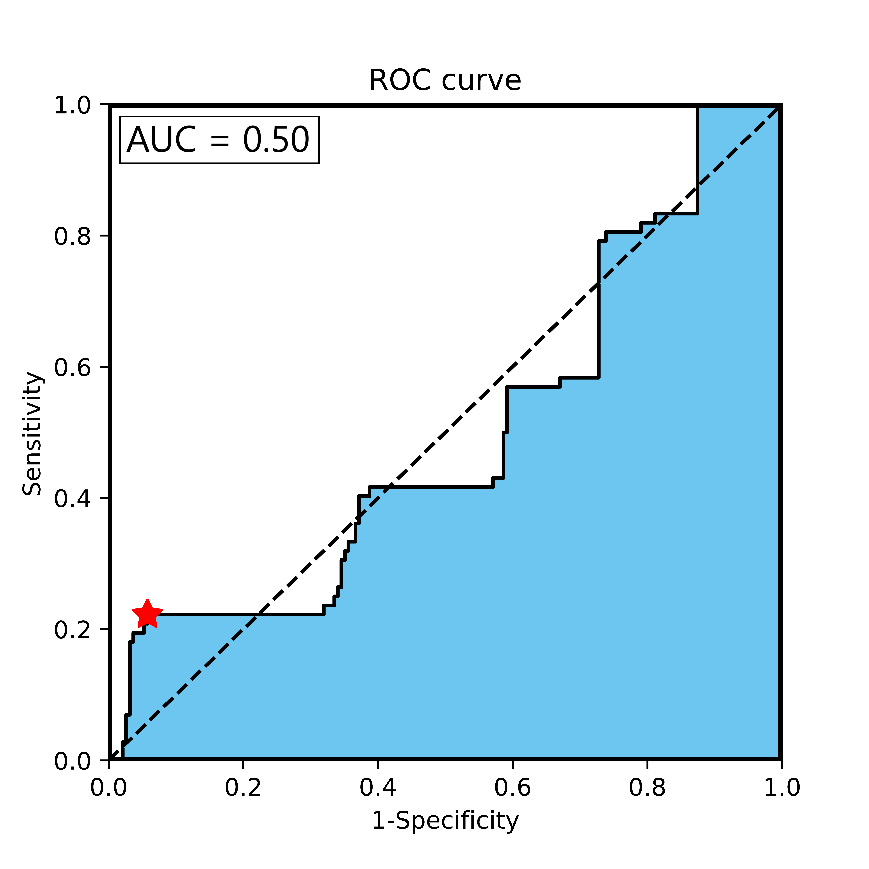
****

Figure S22: ROC and PR curves for classifier trained with z-score normalized data (top) and ROC and PR curves for classifier trained with quantile-normalized data (bottom).

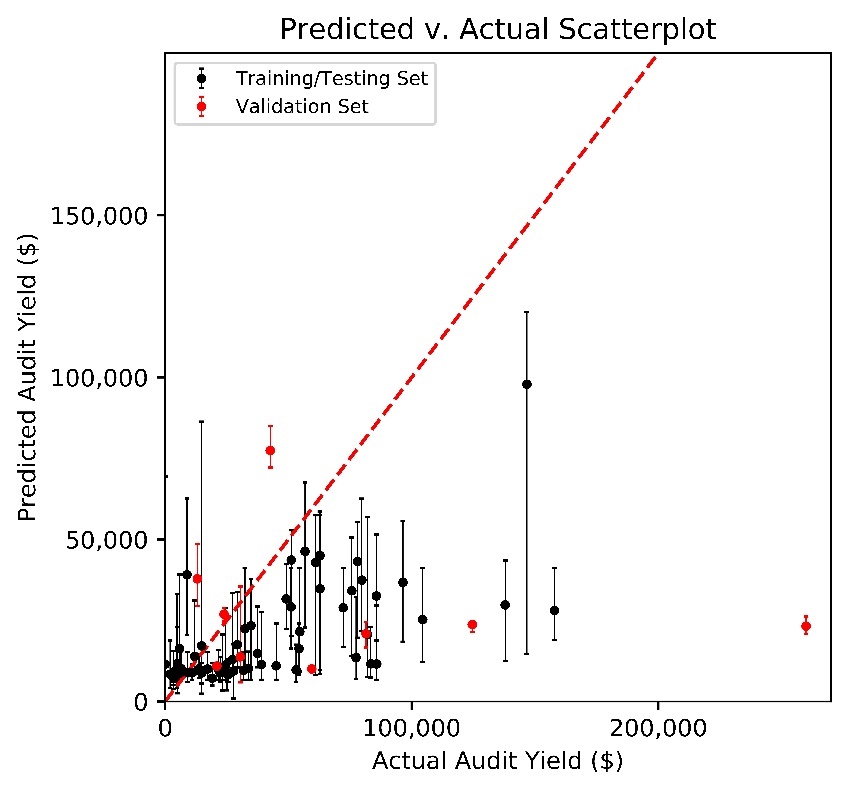
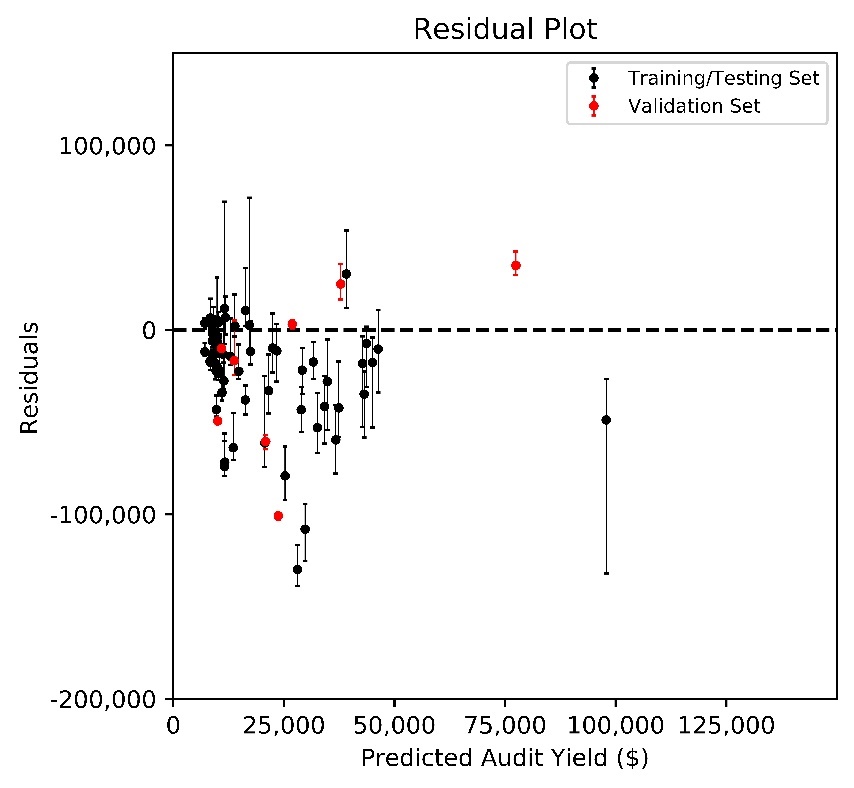
****

Figure S23: Predicted v. Actual (left) and Residual (right) plots for single regressor model. Error bars represent the prediction on multiple returns, where each point represents the mean of predictions for a single business.

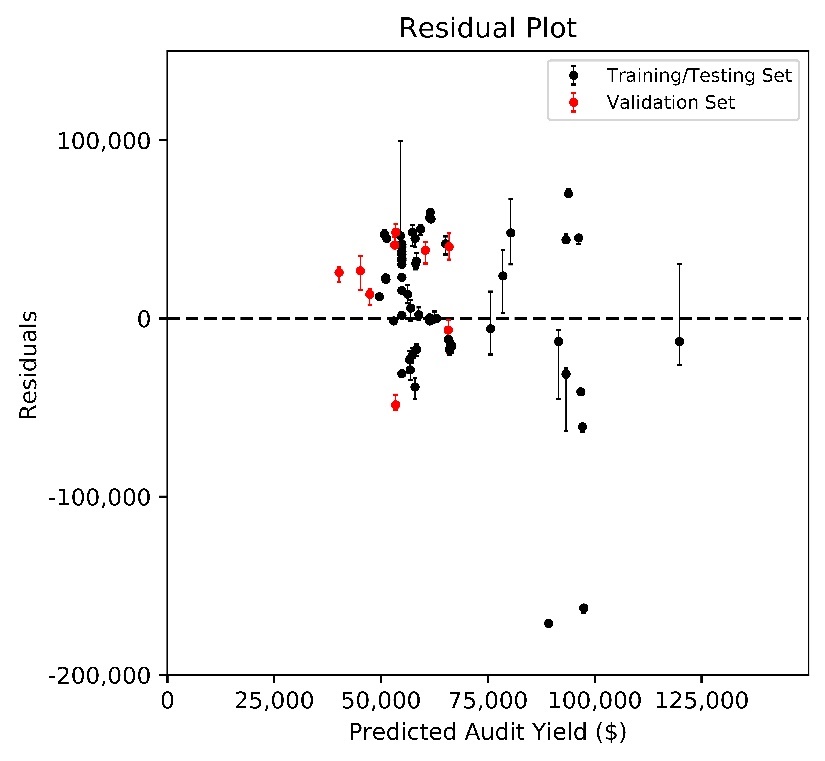
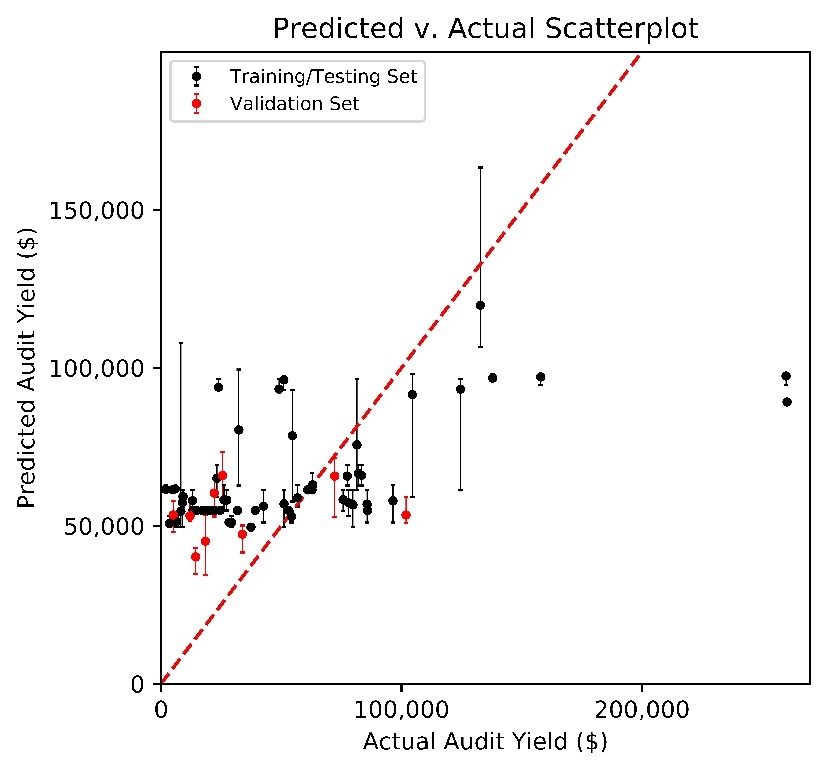
****

Figure S24: Predicted v. Actual (left) and Residual (right) plots ensemble model created with Hypothesis 1 splits. Error bars represent the prediction on multiple returns, where each point represents the mean of predictions for a single business.

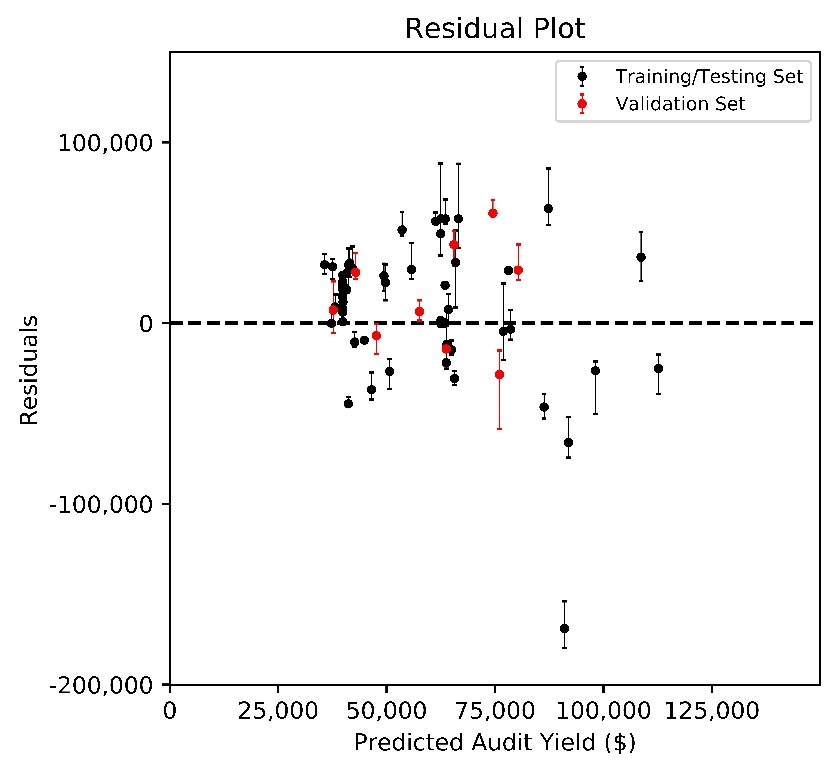
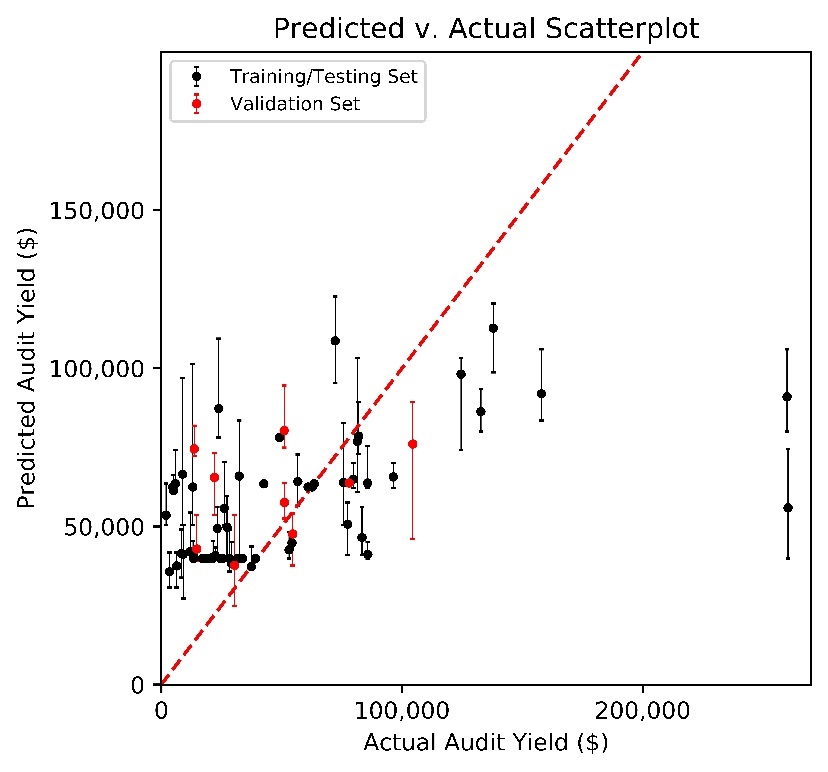
****

Figure S255: Predicted v. Actual (left) and Residual (right) plots ensemble model created with Hypothesis 5 splits. Error bars represent the prediction on multiple returns, where each point represents the mean of predictions for a single business.

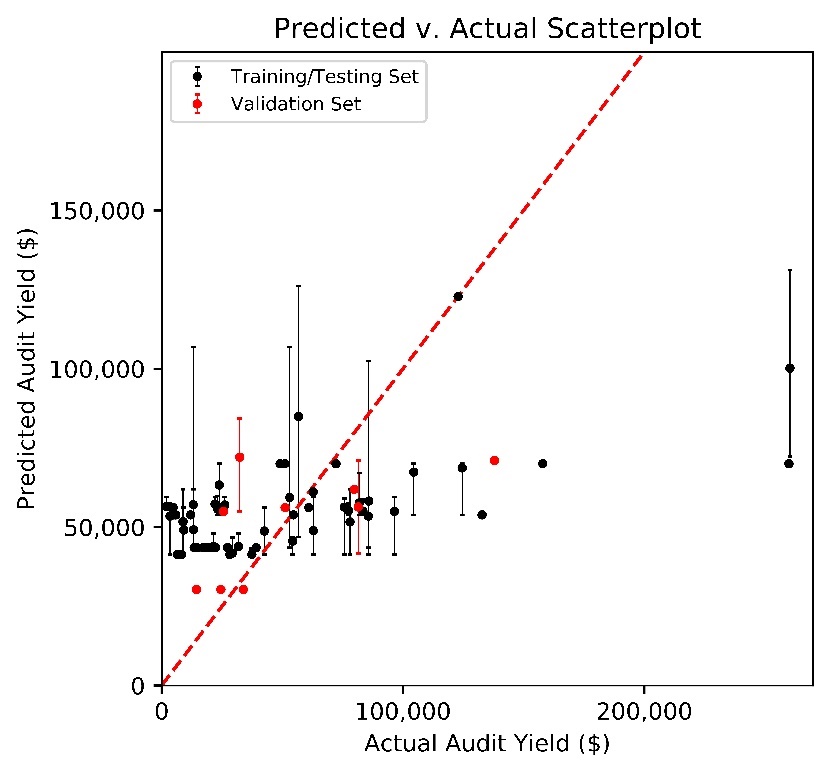
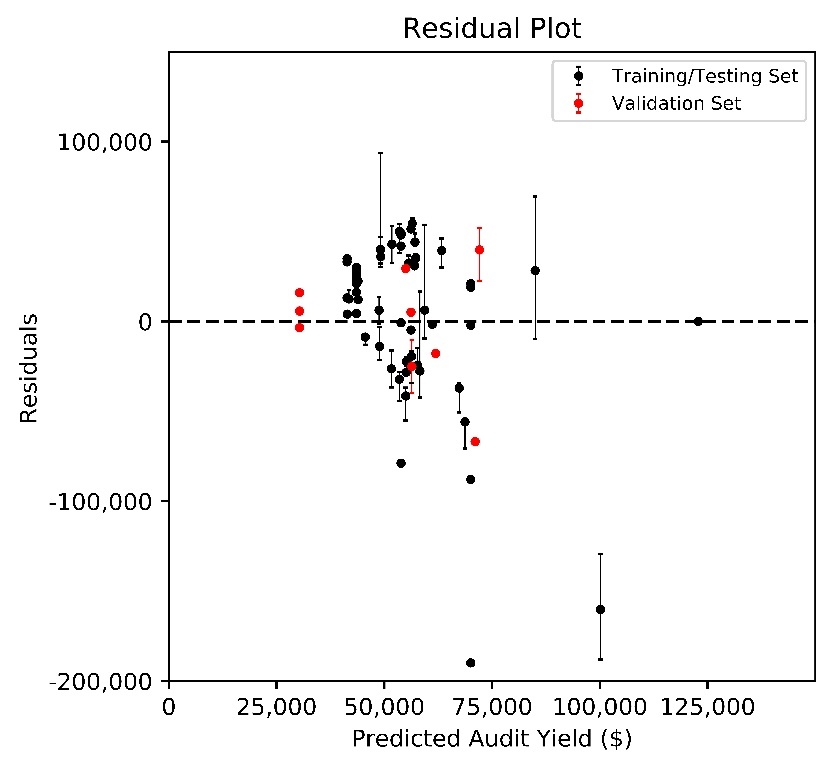
****

Figure S26: Predicted v. Actual (left) and Residual (right) plots ensemble model created with splits on business type. Error bars represent the prediction on multiple returns, where each point represents the mean of predictions for a single business.

**SUPPLEMENTARY TABLES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Index** | **Description** | **Data Type** | **Encoding** |
| 1 | Type of Business | Categorical | One-hot |
| 2 | Filing Frequency | Categorical | One-hot |
| 3 | North American Industry Classification System (NAICS) Code | Categorical | One-hot |
| 4 | Anonymized Primary Business Zip Code ID | Categorical | One-hot |
| 5 | Anonymized Primary City ID | Categorical | One-hot |
| 6 | Site ID | Numeric | None |
| 7 | Business Establishment Date | Date | None |
| 8 | Business Close Date | Date | None |
| 9 | Business Establish Start Date (location specific) | Date | None |
| 10 | Business Close Date (location specific) | Date | None |
| 11 | Anonymized Selling Location City Name ID | Categorical | One-hot |
| 12 | Anonymized Selling Location Zip Code ID | Categorical | One-hot |
| 13 | Return Filing Period | Date | None |
| 14 | Return Period Begin | Date | None |
| 15 | Return Period End | Date | None |
| 16 | Return Received Date | Date | None |
| 17 | Return Due Date | Date | None |
| 18 | Return Type | Categorical | None |
| 19 | Form Type | Categorical | None |
| 20 | Total Gross Sales (typically includes taxable and non-taxable) | Numeric | None |
| 21 | Use Tax Purchases | Numeric | None |
| 22 | Fixture and Equipment Sales | Numeric | None |
| 23 | Total Sales: Sum of Gross Sales + Use Tax Purchases + Fixture and Equipment Sales | Numeric | None |
| 24 | Sale for Resale | Numeric | None |
| 25 | Sale of Nontaxable Food | Numeric | None |
| 26 | Nontaxable Labor included in Gross Sales | Numeric | None |
| 27 | Nontaxable Sale to Federal Government | Numeric | None |
| 28 | Nontaxable sale to Interstate and Foreign Commerce | Numeric | None |
| 29 | Nontaxable Sales Tax included in Gross Sales | Numeric | None |
| 30 | Other Nontaxable Deductions | Numeric | None |
| 31 | Total Nontaxable Deductions | Numeric | None |
| 32 | Taxable Sales: Total Sale - Total Nontaxable Deductions | Numeric | None |
| 33 | Special District Tax | Numeric | None |
| 34 | Tax Due: Total State Tax + Special District Tax | Numeric | None |
| 35 | Excess Tax | Numeric | None |
| 36 | Monthly 1st prepayment made | Numeric | None |
| 37 | Monthly 2nd prepayment made | Numeric | None |
| 38 | Total Prepayments made for the quarter | Numeric | None |
| 39 | Remaining tax after prepayment | Numeric | None |
| 40 | Late Penalty | Numeric | None |
| 41 | Total Amount Due | Numeric | None |
| 42 | Late Interest | Numeric | None |

Table S1: List of data per data group, with a total of 42 features. Red text indicates the features that were removed and not used in analysis. Index 1-12 denote the Registration table data while Index 13-42 denote the Return table data

|  |  |  |
| --- | --- | --- |
| **Index** | **Data Field** | **Reason not Used** |
| 6 | Site ID | Only analyzing single location businesses |
| 7 | Business Establishment Date | Only active accounts in this dataset, not a useful piece of data |
| 8 | Business Close Date | Only active accounts in this dataset |
| 9 | Business Establish Start Date (location specific) | Only analyzing single location businesses |
| 10 | Business Close Date (location specific) | Only analyzing single location businesses |
| 11 | Anonymized Selling Location City Name ID | Only analyzing single location businesses |
| 12 | Anonymized Selling Location Zip Code ID | Only analyzing single location businesses |
| 13 | Return Filing Period | Could not find a way to use this data |
| 14 | Return Period Begin | Could not find a way to use this data |
| 15 | Return Period End | Could not find a way to use this data |
| 16 | Return Received Date | Could not find a way to use this data |
| 17 | Return Due Date | Could not find a way to use this data |
| 18 | Return Type | Only one possible entry, removed for redundancy |
| 19 | Form Type | Only analyzing single location businesses |
| 31 | Total nontaxable deductions | Linear combination of features 24-30 |
| 32 | Taxable Sales | Linear combination of features 23-30 |
| 38 | Total Prepayments | Linear combination of features 36-37 |

Table S2: List of 17 features not used, with a brief explanation as to why.

|  |  |  |
| --- | --- | --- |
| **Index** | **Feature** | **Percentage of 0’s** |
| 1 | Fixture and Equipment Sales | 99.97% |
| 2 | Nontaxable sale to Interstate and Foreign Commerce | 99.76% |
| 3 | Nontaxable Sale to Federal Government | 99.68% |
| 4 | Nontaxable Labor included in Gross Sales | 99.46% |
| 5 | Excess Tax | 98.51% |
| 6 | Other nontaxable deductions | 97.02% |
| 7 | Sale for Resale | 96.73% |
| 8 | Use Tax Purchases | 92.81% |
| 9 | Late Penalty | 90.89% |
| 10 | Late Interest | 90.83% |
| 11 | Sale of Nontaxable Food | 82.26% |
| 12 | Nontaxable Sales Tax included in Gross Sales | 62.34% |
| 13 | Monthly 2nd prepayment made | 32.69% |
| 14 | Monthly 1st prepayment made | 32.41% |
| 15 | Remaining Tax | 1.10% |
| 16 | Total Amount Due | 1.03% |
| 17 | Special District Tax | 0.98% |
| 18 | Tax Due | 0.78% |
| 19 | Total Gross Sales | 0.63% |
| 20 | Total Sales | 0.59% |

Table S3: Table showing numeric features ranked by sparsity – red denotes features that were removed (sparsity > 60%) before data visualization with PCA and t-SNE.

|  |  |
| --- | --- |
| **Cluster** | **Number of Rules found** |
| 1 | 160 |
| 2 | 19,266 |
| 3 | 8 |
| 4 | 4,693 |
| 5 | 5,187 |
| 6 | 12,597 |
| 7 | 3,458 |

Table S4: Number of rules found per cluster using the Apriori algorithm for association rule mining

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Index** | **Rules** | **Support** | **Confidence** | **Lift** | **Cluster** |
| 1 | Remaining Tax is between $1,892 and $1,976 => Total Amount Due is between $1,904 and $1,976 | 0.014 | 1 | 72.00 | 1 |
| 2 | Remaining Tax is between $2,855 and $2,883 => Total Amount Due is between $2,855 and $2,883 | 0.014 | 1 | 72.00 | 1 |
| 3 | Total Sales is between $140,100 and $151,400 => Total Gross Sales is between $140,100 and $151,400 | 0.014 | 1 | 72.00 | 1 |
| 4 | Total Gross Sales is between $79,370 and $81,820 => Total Sales is between $79,370 and $81,820 | 0.014 | 1 | 72.00 | 1 |
| 5 | Total Sales is between $106,900 and $109,700 => Total Gross Sales is between $106,900 and $109,700 | 0.014 | 1 | 72.00 | 1 |
| 6 | Total Sales is between $31,110 and $31,770 => Total Gross Sales is between $31,110 and $31,770 | 0.014 | 1 | 72.00 | 1 |
| 7 | Total Amount Due is between $6,323 and $7,418 => Remaining Tax is between $6,261 and $7,418 | 0.014 | 1 | 72.00 | 1 |
| 8 | Remaining Tax is between $4,099 and $4,122 => Total Amount Due is between $4,099 and $4,121 | 0.014 | 1 | 72.00 | 1 |
| 9 | Total Sales is between $38,040 and $38,530 => Total Gross Sales is between $38,040 and $38,530 | 0.014 | 1 | 72.00 | 1 |
| 10 | Remaining Tax is between $4,694 and $4,887 => Total Amount Due is between $4,694 and $4,887 | 0.014 | 1 | 72.00 | 1 |
| 11 | Total Gross Sales is between $296,960 and $297,490 => Total Sales is between $296,960 and $297,490 | 0.013 | 1 | 78.00 | 2 |
| 12 | Total Gross Sales is between $296,960 and $297,490 => District Tax is between $3,115 and $3,149 | 0.013 | 1 | 78.00 | 2 |
| 13 | Total Gross Sales is between $296,960 and $297,490 => Tax Due is between $21,859 and $22,191 | 0.013 | 1 | 78.00 | 2 |
| 14 | Total Gross Sales is between $296,960 and $297,490 => Monthly 1st prepayment made is between $6,058 and $6,110 | 0.013 | 1 | 78.00 | 2 |
| 15 | Total Gross Sales is between $296,960 and $297,490 => Monthly 2nd prepayment made is between $7,088 and $7,103 | 0.013 | 1 | 78.00 | 2 |
| 16 | Total Gross Sales is between $296,960 and $297,490 => Remaining Tax is between $7,767 and $7,779 | 0.013 | 1 | 78.00 | 2 |
| 17 | Total Gross Sales is between $296,960 and $297,490 => Total Amount Due is between $8,674 and $8,681 | 0.013 | 1 | 78.00 | 2 |
| 18 | District Tax is between $4,696 and $4,746 => Tax Due is between $25,919 and $26,005 | 0.013 | 1 | 78.00 | 2 |
| 19 | District Tax is between $4,696 and $4,746 => Monthly 1st prepayment made is between $8,712 and $8,774 | 0.013 | 1 | 78.00 | 2 |
| 20 | District Tax is between $4,696 and $4,746 => Monthly 2nd prepayment made is between $9,443 and $9,454 | 0.013 | 1 | 78.00 | 2 |
| 21 | Remaining Tax is between $4,426 and $4,441 => Total Amount Due is between $4,426 and $4,441 | 0.019 | 1 | 53.50 | 3 |
| 22 | Total Sales is between $217,440 and $218,570 => Total Gross Sales is between $217,400 and $218,600 | 0.019 | 1 | 53.50 | 3 |
| 23 | Total Gross Sales is between $492,500 and $638,400 => Total Sales is between $492,490 and $645,480 | 0.019 | 1 | 53.50 | 3 |
| 24 | Total Amount Due is between $3,559 and $3,648 => Remaining Tax is between $3,559 and $3,648 | 0.019 | 1 | 53.50 | 3 |
| 25 | District Tax is between $788 and $1,065 => Tax Due is between $5,509 and $8,440 | 0.019 | 1 | 53.50 | 3 |
| 26 | Total Gross Sales is between $102,500 and $103,400 => Total Sales is between $102,530 and $103,400 | 0.019 | 1 | 53.50 | 3 |
| 27 | Remaining Tax is between $4,905 and $4,913 => Total Amount Due is between $4,905 and $4,913 | 0.019 | 1 | 53.50 | 3 |
| 28 | Monthly 2nd prepayment made is between $0 and $30 => Monthly 1st prepayment made is between $0 and $30 | 0.028 | 1 | 35.67 | 3 |
| 29 | Total Gross Sales is between $960,000 and $961,800 => Total Sales is between $960,000 and $961,800 | 0.053 | 1 | 19.00 | 4 |
| 30 | Total Gross Sales is between $960,000 and $961,800 => District Tax is between $14,290 and $14,440 | 0.053 | 1 | 19.00 | 4 |
| 31 | Total Gross Sales is between $960,000 and $961,800 => Tax Due is between $86,348 and $86,847 | 0.053 | 1 | 19.00 | 4 |
| 32 | Total Gross Sales is between $960,000 and $961,800 => Monthly 1st prepayment made is between $28,937 and $29,039 | 0.053 | 1 | 19.00 | 4 |
| 33 | Total Gross Sales is between $960,000 and $961,800 => Monthly 2nd prepayment made is between $30,467 and $30,577 | 0.053 | 1 | 19.00 | 4 |
| 34 | Total Gross Sales is between $960,000 and $961,800 => Remaining Tax is between $26,827 and $26,847 | 0.053 | 1 | 19.00 | 4 |
| 35 | Total Gross Sales is between $960,000 and $961,800 => Total Amount Due is between $26,827 and $26,847 | 0.053 | 1 | 19.00 | 4 |
| 36 | Total Gross Sales is between $845,200 and $845,500 => Total Sales is between $845,200 and $845,500 | 0.053 | 1 | 19.00 | 4 |
| 37 | Total Gross Sales is between $845,200 and $845,500 => District Tax is between $9,668 and $9,701 | 0.053 | 1 | 19.00 | 4 |
| 38 | Total Gross Sales is between $845,200 and $845,500 => Tax Due is between $67,980 and $68,159 | 0.053 | 1 | 19.00 | 4 |
| 39 | Total Gross Sales is between $674,800 and $674,900 => Total Sales is between $674,800 and $674,960 | 0.048 | 1 | 21.00 | 5 |
| 40 | Total Gross Sales is between $674,800 and $674,900 => District Tax is between $8,435 and $8,440 | 0.048 | 1 | 21.00 | 5 |
| 41 | Total Gross Sales is between $674,800 and $674,900 => Tax Due is between $58,764 and $58,855 | 0.048 | 1 | 21.00 | 5 |
| 42 | Total Gross Sales is between $674,800 and $674,900 => Monthly 1st prepayment made is between $17,333 and $17,444 | 0.048 | 1 | 21.00 | 5 |
| 43 | Total Gross Sales is between $674,800 and $674,900 => Monthly 2nd prepayment made is between $18,417 and $18,471 | 0.048 | 1 | 21.00 | 5 |
| 44 | Total Gross Sales is between $674,800 and $674,900 => Remaining Tax is between $23,014 and $23,381 | 0.048 | 1 | 21.00 | 5 |
| 45 | Total Gross Sales is between $674,800 and $674,900 => Total Amount Due is between $23,014 and $23,381 | 0.048 | 1 | 21.00 | 5 |
| 46 | District Tax is between $8,185 and $8,190 => Tax Due is between $57,594 and $57,610 | 0.048 | 1 | 21.00 | 5 |
| 47 | District Tax is between $8,185 and $8,190 => Monthly 1st prepayment made is between $17,329 and $17,330 | 0.048 | 1 | 21.00 | 5 |
| 48 | District Tax is between $8,185 and $8,190 => Monthly 2nd prepayment made is between $20,194 and $20,222 | 0.048 | 1 | 21.00 | 5 |
| 49 | Total Gross Sales is between $538,790 and $539,640 => Total Amount Due is between $14,078 and $14,099 | 0.020 | 1 | 51.00 | 6 |
| 50 | Total Gross Sales is between $538,790 and $539,640 => Remaining Tax is between $14,078 and $14,099 | 0.020 | 1 | 51.00 | 6 |
| 51 | Total Gross Sales is between $538,790 and $539,640 => Monthly 2nd prepayment made is between $17,091 and $17,094 | 0.020 | 1 | 51.00 | 6 |
| 52 | Total Gross Sales is between $538,790 and $539,640 => Monthly 1st prepayment made is between $15,874 and $16,324 | 0.020 | 1 | 51.00 | 6 |
| 53 | Total Gross Sales is between $538,790 and $539,640 => Tax Due is between $47,040 and $47,170 | 0.020 | 1 | 51.00 | 6 |
| 54 | Total Gross Sales is between $538,790 and $539,640 => District Tax is between $6,735 and $6,740 | 0.020 | 1 | 51.00 | 6 |
| 55 | Total Gross Sales is between $538,790 and $539,640 => Total Sales is between $538,790 and $539,640 | 0.020 | 1 | 51.00 | 6 |
| 56 | Remaining Tax is between $16,361 and $16,688 => Monthly 2nd prepayment made is between $15,554 and $15,665 | 0.020 | 1 | 51.00 | 6 |
| 57 | Remaining Tax is between $16,361 and $16,688 => Monthly 1st prepayment made is between $13,921 and $13,925 | 0.020 | 1 | 51.00 | 6 |
| 58 | Remaining Tax is between $16,361 and $16,688 => Tax Due is between $45,840 and $46,110 | 0.020 | 1 | 51.00 | 6 |
| 59 | Total Gross Sales is between $1,278,800 and $1,281,000 => Total Amount Due is between $31,709 and $31,853 | 0.071 | 1 | 14.00 | 7 |
| 60 | Total Gross Sales is between $1,278,800 and $1,281,000 => Remaining Tax is between $31,709 and $31,853 | 0.071 | 1 | 14.00 | 7 |
| 61 | Total Gross Sales is between $1,278,800 and $1,281,000 => Monthly 2nd prepayment made is between $36,750 and $36,820 | 0.071 | 1 | 14.00 | 7 |
| 62 | Total Gross Sales is between $1,278,800 and $1,281,000 => Monthly 1st prepayment made is between $34,080 and $34,250 | 0.071 | 1 | 14.00 | 7 |
| 63 | Total Gross Sales is between $1,278,800 and $1,281,000 => Tax Due is between $102,670 and $103,320 | 0.071 | 1 | 14.00 | 7 |
| 64 | Total Gross Sales is between $1,278,800 and $1,281,000 => District Tax is between $14,679 and $14,708 | 0.071 | 1 | 14.00 | 7 |
| 65 | Total Gross Sales is between $1,278,800 and $1,281,000 => Total Sales is between $1,277,700 and $1,279,000 | 0.071 | 1 | 14.00 | 7 |
| 66 | District Tax is between $18,538 and $18,601 => Total Sales is between $1,491,700 and $1,497,600 | 0.071 | 1 | 14.00 | 7 |
| 67 | District Tax is between $18,538 and $18,601 => Total Gross Sales is between $1,485,200 and $1,490,800 | 0.071 | 1 | 14.00 | 7 |
| 68 | District Tax is between $18,538 and $18,601 => Tax Due is between $129,450 and $130,080 | 0.071 | 1 | 14.00 | 7 |

Table S5: Top ten associative mined rules per cluster with support, confidence, and lift measurements. Rules are separated by arrow delimiters between events A and B. Support denotes the probability of events A and B occurring together, while confidence denotes the conditional probability of B given A, and finally lift denotes the factor by which the cooccurrence of A and B exceed the probability of A and B occurring independently.

|  |  |
| --- | --- |
| **Index** | **Feature Ordering** |
| 1 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions > Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 2 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions = Remaining Tax = Total Amount Due > Tax Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 3 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due > Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 4 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions = Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 5 | Total Gross Sales = Total Sales > Sale of Nontaxable Food = Total Nontaxable Deductions > Taxable Sales > Remaining Tax = Total Amount Due > Tax Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 6 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due > Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 7 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Other Nontaxable Deductions > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 8 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 9 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Special District Tax > Late Penalty > Late Interest > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments |
| 10 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions = Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 11 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Late Penalty > Late Interest > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax |
| 12 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Total Prepayments > Monthly 1st prepayment made = Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 13 | Total Gross Sales = Total Sales = Taxable Sales > Remaining Tax = Total Amount Due > Tax Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 14 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Remaining Tax = Total Amount Due > Total Prepayments > Monthly 1st prepayment made = Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 15 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Total Prepayments > Monthly 1st prepayment made > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 16 | Total Gross Sales = Total Sales > Sale of Nontaxable Food = Total Nontaxable Deductions > Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 17 | Total Gross Sales = Total Sales = Taxable Sales > Remaining Tax = Total Amount Due > Tax Due > Special District Tax > Late Penalty > Late Interest > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments |
| 18 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Monthly 2nd prepayment made > Special District Tax > Late Penalty > Late Interest > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax |
| 19 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 20 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Total Prepayments > Monthly 1st prepayment made > Special District Tax > Monthly 2nd prepayment made > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 21 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Nontaxable Sales Tax included in Gross Sales = Total Nontaxable Deductions > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 22 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 23 | Total Gross Sales = Total Sales > Sale of Nontaxable Food > Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Late Penalty > Late Interest > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments |
| 24 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Late Penalty > Late Interest > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments |
| 25 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Monthly 2nd prepayment made > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 26 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 27 | Total Gross Sales = Total Sales > Taxable Sales > Sale of Nontaxable Food = Total Nontaxable Deductions > Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 28 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Total Prepayments > Monthly 2nd prepayment made > Special District Tax > Monthly 1st prepayment made > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 29 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Late Penalty > Late Interest > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments |
| 30 | Total Sales > Total Gross Sales > Taxable Sales > Sale of Nontaxable Food > Tax Due = Remaining Tax = Total Amount Due > Nontaxable Sales Tax included in Gross Sales > Special District Tax > Use Tax Purchases > Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 31 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 32 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made = Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 33 | Total Sales = Taxable Sales > Total Gross Sales > Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Use Tax Purchases > Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 34 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 35 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Monthly 1st prepayment made > Monthly 2nd prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 36 | Total Sales = Taxable Sales > Total Gross Sales > Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases > Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 37 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales > Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 38 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Remaining Tax = Total Amount Due > Monthly 1st prepayment made = Total Prepayments > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 2nd prepayment made = Late Penalty = Late Interest |
| 39 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Monthly 1st prepayment made = Monthly 2nd prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 40 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due > Nontaxable Sales Tax included in Gross Sales > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 41 | Total Sales = Taxable Sales > Total Gross Sales > Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases > Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 42 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 43 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made = Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 44 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Remaining Tax = Total Amount Due > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 45 | Total Gross Sales = Total Sales = Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 46 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Monthly 1st prepayment made = Monthly 2nd prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 47 | Total Gross Sales = Total Sales > Taxable Sales > Sale of Nontaxable Food > Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 48 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 49 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due > Nontaxable Sales Tax included in Gross Sales > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 50 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Monthly 2nd prepayment made > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 51 | Total Gross Sales = Total Sales > Sale of Nontaxable Food > Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 52 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 53 | Total Gross Sales = Total Sales > Sale of Nontaxable Food > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due = Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 54 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due > Nontaxable Sales Tax included in Gross Sales > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 55 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Monthly 1st prepayment made > Monthly 2nd prepayment made > Remaining Tax = Total Amount Due > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 56 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Monthly 2nd prepayment made > Monthly 1st prepayment made > Special District Tax > Remaining Tax = Total Amount Due > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 57 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Monthly 1st prepayment made > Remaining Tax = Total Amount Due > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 58 | Total Gross Sales = Total Sales > Taxable Sales > Nontaxable Sales Tax included in Gross Sales = Tax Due > Total Prepayments > Remaining Tax = Total Amount Due > Monthly 1st prepayment made > Monthly 2nd prepayment made > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Late Penalty = Late Interest |
| 59 | Total Gross Sales = Total Sales > Taxable Sales > Tax Due = Remaining Tax = Total Amount Due > Nontaxable Sales Tax included in Gross Sales > Special District Tax > Use Tax Purchases = Fixture and Equipment Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Other Nontaxable Deductions = Total Nontaxable Deductions = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Late Penalty = Late Interest |
| 60 | Total Gross Sales = Use Tax Purchases = Fixture and Equipment Sales = Total Sales = Sale for Resale = Sale of Nontaxable Food = Nontaxable Labor included in Gross Sales = Nontaxable Sale to Federal Government = Nontaxable sale to Interstate and Foreign Commerce = Nontaxable Sales Tax included in Gross Sales = Other Nontaxable Deductions = Total Nontaxable Deductions = Taxable Sales = Special District Tax = Tax Due = Excess Tax = Monthly 1st prepayment made = Monthly 2nd prepayment made = Total Prepayments = Remaining Tax = Late Penalty = Late Interest = Total Amount Due |

Table S6: List of several possible orderings of all numerical features

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Number of tax records** | **Number of audited records** | **Records with audit yield > $0** | **Records with audit yield > CDTFA threshold** | **Audit yield (> $0) rate** | **Audit yield (> CDTFA threshold) rate** | **Percentage of audited records** |
| 1 | 103 | 50 | 21 | 19 | 0.42 | 0.38 | 48.54% |
| 2 | 153 | 63 | 24 | 14 | 0.38 | 0.22 | 41.18% |
| 3 | 119 | 48 | 23 | 15 | 0.48 | 0.31 | 40.34% |
| 4 | 127 | 44 | 13 | 9 | 0.30 | 0.20 | 34.65% |
| 5 | 106 | 34 | 6 | 6 | 0.18 | 0.18 | 32.08% |
| 6 | 182 | 55 | 26 | 16 | 0.47 | 0.29 | 30.22% |
| 7 | 117 | 34 | 0 | 0 | 0.00 | 0.00 | 29.06% |
| 8 | 450 | 85 | 34 | 27 | 0.40 | 0.32 | 18.89% |
| 9 | 223 | 40 | 29 | 24 | 0.73 | 0.60 | 17.94% |
| 10 | 229 | 41 | 17 | 6 | 0.41 | 0.15 | 17.90% |
| 11 | 119 | 21 | 10 | 6 | 0.48 | 0.29 | 17.65% |
| 12 | 180 | 30 | 14 | 13 | 0.47 | 0.43 | 16.67% |
| 13 | 384 | 64 | 21 | 16 | 0.33 | 0.25 | 16.67% |
| 14 | 110 | 18 | 7 | 7 | 0.39 | 0.39 | 16.36% |
| 15 | 117 | 19 | 10 | 7 | 0.53 | 0.37 | 16.24% |
| 16 | 138 | 22 | 5 | 5 | 0.23 | 0.23 | 15.94% |
| 17 | 136 | 21 | 12 | 8 | 0.57 | 0.38 | 15.44% |
| 18 | 123 | 17 | 5 | 5 | 0.29 | 0.29 | 13.82% |
| 19 | 164 | 22 | 9 | 9 | 0.41 | 0.41 | 13.41% |
| 20 | 143 | 19 | 16 | 12 | 0.84 | 0.63 | 13.29% |
| 21 | 106 | 14 | 9 | 6 | 0.64 | 0.43 | 13.21% |
| 22 | 2,865 | 334 | 67 | 54 | 0.20 | 0.16 | 11.66% |
| 23 | 267 | 30 | 2 | 1 | 0.07 | 0.03 | 11.24% |
| 24 | 1,290 | 141 | 116 | 97 | 0.82 | 0.69 | 10.93% |
| 25 | 1,484 | 161 | 46 | 39 | 0.29 | 0.24 | 10.85% |
| 26 | 3,318 | 343 | 100 | 71 | 0.29 | 0.21 | 10.34% |
| 27 | 120 | 12 | 5 | 0 | 0.42 | 0.00 | 10.00% |
| 28 | 205 | 20 | 10 | 6 | 0.50 | 0.30 | 9.76% |
| 29 | 185 | 18 | 17 | 13 | 0.94 | 0.72 | 9.73% |
| 30 | 127 | 12 | 0 | 0 | 0.00 | 0.00 | 9.45% |
| 31 | 3,610 | 339 | 76 | 63 | 0.22 | 0.19 | 9.39% |
| 32 | 1,032 | 80 | 18 | 13 | 0.23 | 0.16 | 7.75% |
| 33 | 354 | 27 | 1 | 1 | 0.04 | 0.04 | 7.63% |
| 34 | 1,792 | 135 | 31 | 28 | 0.23 | 0.21 | 7.53% |
| 35 | 1,857 | 138 | 43 | 29 | 0.31 | 0.21 | 7.43% |
| 36 | 279 | 19 | 2 | 1 | 0.11 | 0.05 | 6.81% |
| 37 | 225 | 15 | 1 | 1 | 0.07 | 0.07 | 6.67% |
| 38 | 190 | 12 | 5 | 5 | 0.42 | 0.42 | 6.32% |
| 39 | 578 | 35 | 9 | 5 | 0.26 | 0.14 | 6.06% |
| 40 | 256 | 14 | 4 | 4 | 0.29 | 0.29 | 5.47% |
| 41 | 346 | 18 | 1 | 1 | 0.06 | 0.06 | 5.20% |
| 42 | 2,199 | 114 | 28 | 17 | 0.25 | 0.15 | 5.18% |
| 43 | 1,001 | 50 | 9 | 7 | 0.18 | 0.14 | 5.00% |
| 44 | 823 | 39 | 5 | 4 | 0.13 | 0.10 | 4.74% |
| 45 | 6,894 | 325 | 148 | 109 | 0.46 | 0.34 | 4.71% |
| 46 | 580 | 27 | 5 | 5 | 0.19 | 0.19 | 4.66% |
| 47 | 873 | 39 | 25 | 5 | 0.64 | 0.13 | 4.47% |
| 48 | 1,799 | 80 | 22 | 20 | 0.28 | 0.25 | 4.45% |
| 49 | 296 | 13 | 7 | 6 | 0.54 | 0.46 | 4.39% |
| 50 | 897 | 39 | 12 | 10 | 0.31 | 0.26 | 4.35% |
| 51 | 2,573 | 110 | 5 | 2 | 0.05 | 0.02 | 4.28% |
| 52 | 4,420 | 186 | 107 | 53 | 0.58 | 0.28 | 4.21% |
| 53 | 1,945 | 76 | 21 | 11 | 0.28 | 0.14 | 3.91% |
| 54 | 260 | 10 | 5 | 5 | 0.50 | 0.50 | 3.85% |
| 55 | 1,241 | 43 | 11 | 10 | 0.26 | 0.23 | 3.46% |
| 56 | 798 | 26 | 8 | 8 | 0.31 | 0.31 | 3.26% |
| 57 | 1,169 | 35 | 9 | 7 | 0.26 | 0.20 | 2.99% |
| 58 | 2,267 | 59 | 14 | 9 | 0.24 | 0.15 | 2.60% |
| 59 | 484 | 11 | 7 | 5 | 0.64 | 0.45 | 2.27% |
| 60 | 3,234 | 37 | 4 | 3 | 0.11 | 0.08 | 1.14% |

Table S7: Table showing tax record details regarding orderings listed in Table S6.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **p-value (Threshold = 0.05)** | **Selected Threshold for each bin (combination of CityID and NAICS)** | | | | | | |
| **ORIGINAL HYPOTHESES** | **0.25\*Mean** | **0.5\*Mean** | **0.75\*Mean** | **Mean** | **1.25\*Mean** | **1.5\*Mean** | **1.75\*Mean** |
| Reported gross sales lower than some threshold result in higher yield audits (vs. when reported gross sales is higher than the threshold) | < 10-2 | < 10-4 | < 10-4 | < 0.01 | 0.02 | 0.02 | 0.03 |
| Individual taxpayers are more likely to have positive audits compared to corporations | 0.17 | | | | | | |
| The ratio of is lower than some threshold result in higher yield audits | 0.96 | 0.94 | 0.98 | 0.58 | 0.07 | 0.05 | 0.10 |
| Variance of reported gross sales (in a specific period) lower than some threshold result in higher yield audits | 0.1 | 0.40 | 0.31 | 0.11 | 0.34 | 0.27 | 0.17 |
| Late return penalties higher than some threshold result in higher yield audits | < 10-7 | < 10-7 | < 10-6 | < 10-6 | < 10-7 | < 10-5 | < 10-4 |

Table S8: Wilcoxon-ranksum test applied to the hypotheses. Pink signifies a p-value > 0.05 (insignificant), while green signifies a statistically significant result (p-value < 0.05). Grayed-out cells mark non-significant results.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **p-value (Threshold = 0.05)** | **Selected Threshold for each bin (combination of CityID and NAICS)** | | | | | | |
| **INVERSE HYPOTHESES** | **0.25\*Mean** | **0.5\*Mean** | **0.75\*Mean** | **Mean** | **1.25\*Mean** | **1.5\*Mean** | **1.75\*Mean** |
| Reported gross sales higher than some threshold result in higher yield audits (vs. when reported gross sales is higher than the threshold) | 0.99 | ~ 1 | 0.99 | 0.98 | 0.98 | 0.98 | 0.97 |
| Individual taxpayers are less likely to have positive audits compared to corporations | 0.84 | | | | | | |
| The ratio of is higher than  some threshold results in higher yield audits | 0.04 | 0.06 | 0.02 | 0.42 | 0.93 | 0.95 | 0.91 |
| Variance of reported gross sales (in a specific period) higher than some threshold result in higher yield audits | 0.1 | 0.40 | 0.31 | 0.11 | 0.34 | 0.27 | 0.17 |
| Late return penalties lower than some threshold result in higher yield audits | ~ 1 | ~ 1 | ~ 1 | ~ 1 | ~ 1 | ~ 1 | ~ 1 |

Table S9: Wilcoxon-ranksum test applied to the inverse hypotheses. Pink signifies a p-value > 0.05 (insignificant). Grayed-out cells mark non-significant results. “~ 1” results denote probabilities arbitrarily close to 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | # Returns  (Audit Yield = CDTFA Threshold) | # Returns  (Audit Yield > CDTFA Threshold) | Ratio  (Audit Yield > CDTFA Threshold) | p-value  (HyperGeometric) | p-value  (Wilcoxon-ranksum) |
| 0.50 < 0.55 | 80 | 5 | 5.9% | > 1 | > 1-10-1 |
| 0.55 < 0.60 | 39 | 10 | 20.4% | > 1-10-1 | > 1-10-1 |
| 0.60 < 0.65 | 43 | 6 | 12.2% | > 1-10-1 | > 1-10-1 |
| 0.65 < 0.70 | 50 | 40 | 44.4% | < 10-6 | < 10-6 |
| 0.70 < 0.75 | 28 | 12 | 30.0% | > 1-10-1 | > 1-10-1 |
| 0.75 < 0.80 | 47 | 32 | 40.5% | < 10-4 | < 10-6 |
| 0.80 < 0.85 | 79 | 28 | 26.2% | > 1-10-1 | > 1-10-1 |
| 0.85 < 0.90 | 83 | 22 | 21.0% | > 1-10-1 | > 1-10-1 |
| 0.90 < 0.95 | 1,463 | 486 | 24.9% | > 1-10-2 | > 1-10-4 |
| 0.95 < 1.00 | 276 | 8 | 2.8% | > 1 | > 1 |
| 1 | 2,093 | 784 | 27.3% | < 10-10 | < 10-2 |

Table S10: Taxable transaction ratios and their corresponding audit yields. We observe that the ratio of taxable transaction to gross sales in three specific ranges produce higher audit yields. Even for the returns where all gross sales are taxable, about 27.3% of them produce audit yields > CDTFA Threshold (significantly higher than the average, 23.7%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Features** | **Used in Classification Model** | **Used in Regression Model** |
| 1 | Late Interest | Yes | Yes |
| 2 | Late Penalty | Yes | Yes |
| 3 | Total Gross Sales | Yes | Yes |
| 4 | Total Amount Due | Yes | Yes |
| 5 | Total Sales | Yes | Yes |
| 6 | Nontaxable Sales Tax included in Gross Sales | Yes | Yes |
| 7 | Tax Due | Yes | Yes |
| 8 | Excess Tax | Yes | No |
| 9 | Remaining Tax | Yes | Yes |
| 10 | Use Tax Purchases | Yes | Yes |
| 11 | Sale of Nontaxable Food | Yes | Yes |
| 12 | Monthly 2nd prepayment made | Yes | Yes |
| 13 | Filing Frequency - SUTFA | Yes | No |
| 14 | Filing Frequency - SUTQ | Yes | Yes |
| 15 | Filing Frequency - SUTPQ | Yes | Yes |
| 16 | Filing Frequency - SUTM | Yes | Yes |
| 17 | Business Type - LIMITD | Yes | No |
| 18 | Business Type - IND | Yes | Yes |
| 19 | Business Type - PART | Yes | Yes |
| 20 | Business Type - LLC | Yes | Yes |
| 21 | Business Type - CORP | Yes | Yes |
| 22 | NAICS Code - 722400 | Yes | Yes |
| 23 | NAICS Code - 722330 | Yes | Yes |
| 24 | NAICS Code - 722514 | Yes | No |
| 25 | NAICS Code - 722300 | Yes | No |
| 26 | NAICS Code - 722320 | Yes | Yes |
| 27 | NAICS Code - 722513 | Yes | Yes |
| 28 | NAICS Code - 722310 | Yes | Yes |
| 29 | NAICS Code - 722511 | Yes | Yes |
| 30 | NAICS Code - 722515 | Yes | Yes |
| 31 | NAICS Code - 722410 | No | Yes |

Table S11: Final dataset features used in training dataset for classification (left column) and regression (right column) models.

Table S12: Hyperparameters used for models for different normalization methods, along with evaluation results by which models were selected.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Tuned Hyperparameters** | | | | **Evaluation Metrics** | |
| **Normalization**  **Method** | **Batch**  **Size** | **Hidden Layer 1**  **(# nodes)** | **Hidden Layer 2**  **(# nodes)** | **Dropout** | **Accuracy** | **F1** |
| Z-score | 32 | 24 | 11 | 0.3 | 53% | 0.31 |
| Quantile | 64 | 26 | 7 | 0.2 | 55% | 0.34 |
| Minmax | 32 | 19 | 8 | 0.3 | 61% | 0.42 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Tuned Hyperparameters** | | | | | | **Evaluation Metrics** | | |
| **Ensemble Model** | **Features per Tree (%)** | **Learning**  **Rate** | **Samples per Tree (%)** | **Max depth** | **L1-regularization value** | **Number of Trees** | **Training R²** | **Testing R²** | **Validation R²** |
| Hypothesis 1 Model | 20 | 0.9 | 20 | 2 | 4 | 45 | 0.20 | 0.18 | 0.15 |
| Hypothesis 5 Model | 30 | 0.8 | 20 | 2 | 3 | 35 | 0.15 | 0.12 | 0.12 |
| NAICS code model | 30 | 1 | 20 | 2 | 4 | 50 | 0.31 | 0.28 | 0.30 |
| Business Type model | 30 | 1 | 40 | 3 | 4 | 45 | 0.29 | 0.28 | 0.28 |

Table S13: Hyperparameters used for different ensemble regressors, along with evaluation results by which models were selected.