ENHANCING AUDITOR DECISION-MAKING IN INDIAN FIRMS WITH A PREDICTION CLASSIFICATION MODEL

Team:

1. Srikanth Ranganathan (Student Number: 1009747794)
   Contact: srikanth.ranganathan@mail.utoronto.ca

2. Joshua Samadi (Student Number: 1005627403)
   Contact: josh.samadi@mail.utoronto.ca

1- Explain what problem you are going to solve using this dataset. Provide a brief overview of your problem statement.

We aim to address the challenge of identifying potential instances of fraudulent financial activities among firms audited by the Auditor Office in India during the years 2015 and 2016. Leveraging a dataset comprising 776 data points, we intend to construct a predictive classification model. This model will utilize 26 input variables, including Sector, Sector_Score, Location, money_value, district loss, and various Risk types, among others. By deploying this predictive model, auditors can make informed decisions during the auditing process, gaining valuable insights into whether a particular firm may exhibit signs of suspicious financial behavior. The classification model will serve as a practical tool for auditors, offering a reference point to assess and identify potentially fraudulent activities within audited firms based on the provided input variables.

2- Explain your dataset. Explore your dataset and provide at least 5 meaningful charts/graphs with an explanation.

3- Do data cleaning/pre-processing as required and explain what you have done for your dataset and why.

The audit_risk dataset has 776 rows and 27 columns. There are 26 feature columns and 1 target column (Risk)

```
[8]  1  df=pd.read_csv("audit_risk.csv")
     2  df.head()
```

✓ <1 sec

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOTAL	numbers	...	RiSk_E	History	Prob	Risk_F	Score	Inherent_Risk	CONTROL_RISK	Detection_Risk	Audit_Risk	Risk
0	3.89	23	4.18	0.6	2.508	2.50	0.2	0.500	6.68	5.0	...	0.4	0	0.2	0.0	2.4	8.574	0.4	0.5	1.7148	1
1	3.89	6	0.00	0.2	0.000	4.83	0.2	0.966	4.83	5.0	...	0.4	0	0.2	0.0	2.0	2.554	0.4	0.5	0.5108	0
2	3.89	6	0.51	0.2	0.102	0.23	0.2	0.046	0.74	5.0	...	0.4	0	0.2	0.0	2.0	1.548	0.4	0.5	0.3096	0
3	3.89	6	0.00	0.2	0.000	10.80	0.6	6.480	10.80	6.0	...	0.4	0	0.2	0.0	4.4	17.530	0.4	0.5	3.5060	1
4	3.89	6	0.00	0.2	0.000	0.08	0.2	0.016	0.08	5.0	...	0.4	0	0.2	0.0	2.0	1.416	0.4	0.5	0.2832	0

5 rows × 27 columns

```
[10]  1  df.shape
```

✓ <1 sec

```
(776, 27)
```

Below table provides statistical summary for all the features and target variables.

compute1assign5 · Kernel idle CPU 0% RAM 2%

```
1  df.describe().T
```

[7] ✓ <1 sec

	count	mean	std	min	25%	50%	75%	max
Sector_score	776.0	20.184536	24.319017	1.85	2.3700	3.8900	55.5700	59.8500
PARA_A	776.0	2.450194	5.678870	0.00	0.2100	0.8750	2.4800	85.0000
Score_A	776.0	0.351289	0.174055	0.20	0.2000	0.2000	0.6000	0.6000
Risk_A	776.0	1.351029	3.440447	0.00	0.0420	0.1750	1.4880	51.0000
PARA_B	776.0	10.799988	50.083624	0.00	0.0000	0.4050	4.1600	1264.6300
Score_B	776.0	0.313144	0.169804	0.20	0.2000	0.2000	0.4000	0.6000
Risk_B	776.0	6.334008	30.072845	0.00	0.0000	0.0810	1.8405	758.7780
TOTAL	776.0	13.218481	51.312829	0.00	0.5375	1.3700	7.7075	1268.9100
numbers	776.0	5.067655	0.264449	5.00	5.0000	5.0000	5.0000	9.0000
Score_B.1	776.0	0.223711	0.080352	0.20	0.2000	0.2000	0.2000	0.6000
Risk_C	776.0	1.152964	0.537417	1.00	1.0000	1.0000	1.0000	5.4000
Money_Value	775.0	14.137631	66.606519	0.00	0.0000	0.0900	5.5950	935.0300
Score_MV	776.0	0.290979	0.159745	0.20	0.2000	0.2000	0.4000	0.6000
Risk_D	776.0	8.265434	39.970849	0.00	0.0000	0.0180	2.2350	561.0180
District_Loss	776.0	2.505155	1.228678	2.00	2.0000	2.0000	2.0000	6.0000
PROB	776.0	0.206186	0.037508	0.20	0.2000	0.2000	0.2000	0.6000
RiSk_E	776.0	0.519072	0.290312	0.40	0.4000	0.4000	0.4000	2.4000
History	776.0	0.104381	0.531031	0.00	0.0000	0.0000	0.0000	9.0000
Prob	776.0	0.216753	0.067987	0.20	0.2000	0.2000	0.2000	0.6000
Risk_F	776.0	0.053608	0.305835	0.00	0.0000	0.0000	0.0000	5.4000
Score	776.0	2.702577	0.858923	2.00	2.0000	2.4000	3.2500	5.2000
Inherent_Risk	776.0	17.680612	54.740244	1.40	1.5835	2.2140	10.6635	801.2620
CONTROL_RISK	776.0	0.572680	0.444581	0.40	0.4000	0.4000	0.4000	5.8000
Detection_Risk	776.0	0.500000	0.000000	0.50	0.5000	0.5000	0.5000	0.5000
Audit_Risk	776.0	7.168158	38.667494	0.28	0.3167	0.5556	3.2499	961.5144

Below table provides features type for each column. The dataset has 23 column of double-precision Float64 real number, 3 columns of 64-bit integer data and 1 object column.

```
1  df.dtypes
```

[8] ✓ <1 sec

```
Sector_score      float64
LOCATION_ID        object
PARA_A            float64
Score_A           float64
Risk_A            float64
PARA_B            float64
Score_B           float64
Risk_B            float64
TOTAL             float64
numbers           float64
Score_B.1         float64
Risk_C            float64
Money_Value       float64
Score_MV          float64
Risk_D            float64
District_Loss       int64
PROB              float64
RiSk_E            float64
History             int64
Prob              float64
Risk_F            float64
Score             float64
Inherent_Risk     float64
CONTROL_RISK      float64
Detection_Risk    float64
Audit_Risk        float64
Risk                int64
dtype: object
```

As per following table there is only one null value in Money_Value column

```
1   df.isnull().sum()
```

[21] ✓ <1 sec

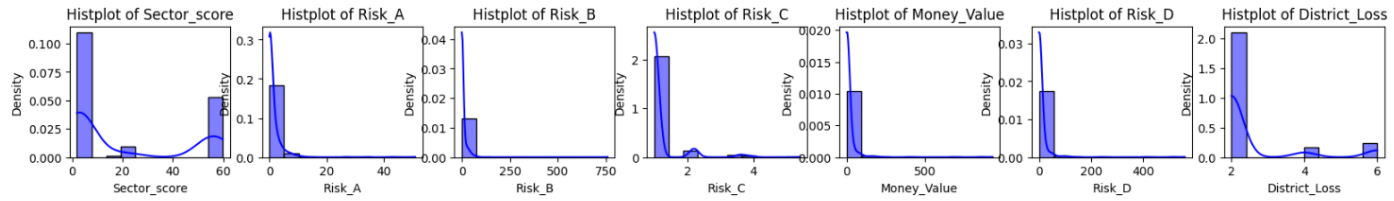
```
Sector_score        0
LOCATION_ID         0
PARA_A              0
Score_A             0
Risk_A              0
PARA_B              0
Score_B             0
Risk_B              0
TOTAL               0
numbers             0
Score_B.1           0
Risk_C              0
Money_Value         1
Score_MV            0
Risk_D              0
District_Loss       0
PROB                0
RiSk_E              0
History             0
Prob                0
Risk_F              0
Score               0
Inherent_Risk       0
CONTROL_RISK        0
Detection_Risk      0
Audit_Risk          0
Risk                0
dtype: int64
```

Below Histogram charts demonstrate distribution type across different features. it is observed that most of the features has non-normal distribution with some features distribution scattered across the range.

```
#below Histogram charts demonstrate distribution type across different features. it is observed that most of the features has non-normal distribution with some features distributes scattered across the range.
plt.figure(1,figsize=(20, 2))
n=0
for x in ['Sector_score', 'Risk_A', 'Risk_B', 'Risk_C', 'Money_Value', 'Risk_D', 'District_Loss']:
    n+=1
    plt.subplot(1,7,n)
    plt.subplots_adjust(hspace=0.2, wspace=0.2)
    sns.histplot(df[x], bins=10, kde=True, color='blue', stat='density')
    plt.title ('Histplot of {}'.format(x))
plt.show
```

✓ 1 sec

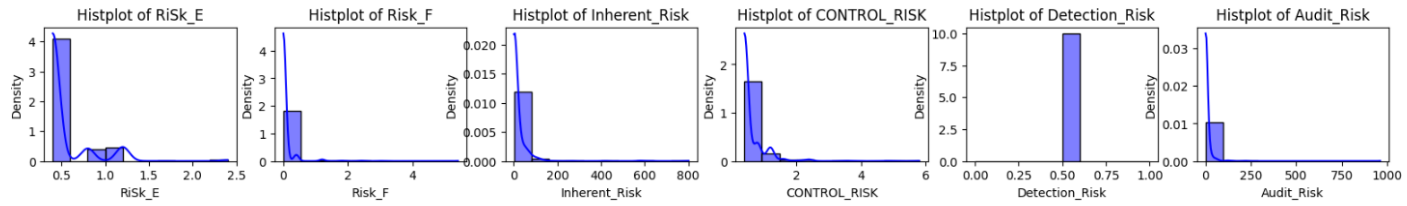
```
<function matplotlib.pyplot.show(close=None, block=None)>
```




```
plt.figure(1,figsize=(20, 2))
n=0
for x in ['RiSk_E', 'Risk_F', 'Inherent_Risk', 'CONTROL_RISK','Detection_Risk', 'Audit_Risk']:
    n+=1
    plt.subplot(1,6,n)
    plt.subplots_adjust(hspace=0.2, wspace=0.2)
    sns.histplot(df[x], bins=10, kde=True, color='blue', stat='density')
    plt.title ('Histplot of {}'.format(x))
plt.show
```

✓ 1 sec

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

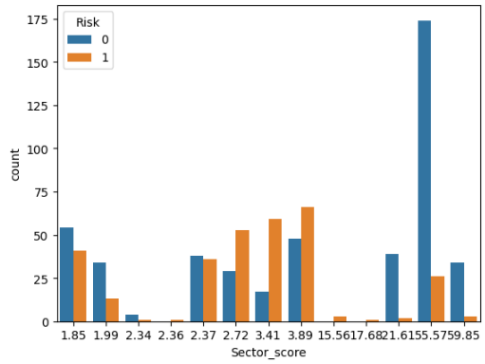


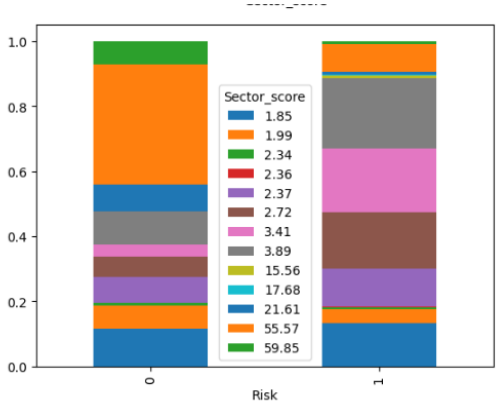
In order to get more insight from some of the features, we plotted cross table between some features and target column (“Risk”). The first feature is Sector_Score which sector score 2.37, 2.72, 3.41 and 3.89 sectors has demonstrated the highest amount of being fraudulent.

```
Code  Markdown

1  Risk_crosstab = pd.crosstab (df["Risk"], df["Sector_score"], margins=False)
2  sns.countplot(x="Sector_score", hue="Risk", data=df)
3  Risk_norm = Risk_crosstab.div(Risk_crosstab.sum(axis=1) ,axis=0)
4  Risk_norm.plot(kind = 'bar', stacked =True)
[25]
```

```
<Axes: xlabel='Risk'>
```

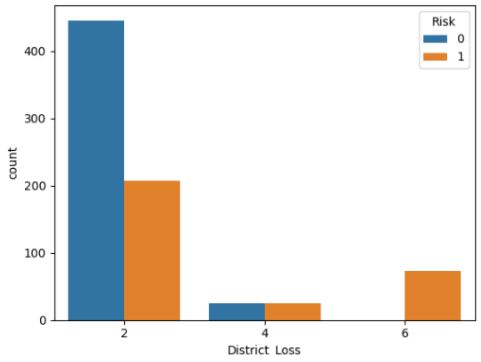


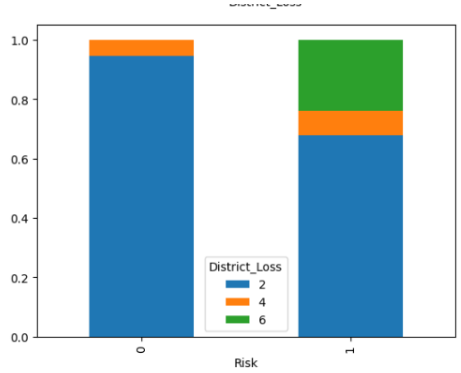


For District_Loss, District #2 has the highest amount of being fraudulent firm.

```
1  #For District_Loss, District #2 has the highest amount of being fraudulent firm.
2  Risk_crosstab2 = pd.crosstab(df["Risk"], df["District_Loss"], margins=False)
3  sns.countplot(x="District_Loss", hue="Risk", data=df)
4  Risk_norm2 = Risk_crosstab2.div(Risk_crosstab2.sum(axis=1) ,axis=0)
5  Risk_norm2.plot(kind = 'bar', stacked =True)
[9]  <1 sec
```

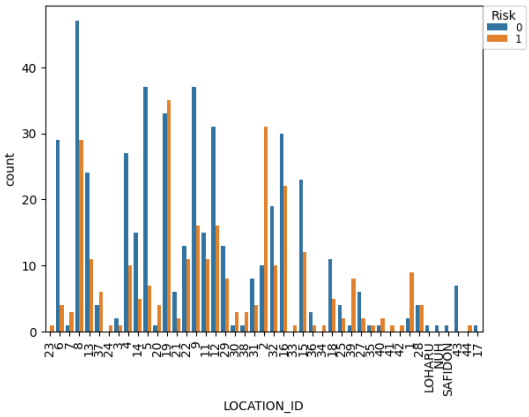
```
<Axes: xlabel='Risk'>
```





For Location_ID, Locations #8, #19 and #2 have this highest amount of fradulant firms.

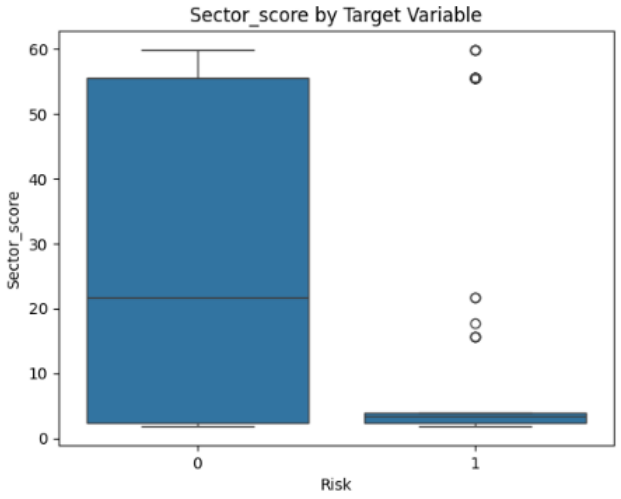
```
1  #For Location_ID, Locations #8, #19 and #2 have this highest amount of fradulant firms.
2  Risk_crosstab3 = pd.crosstab(df["Risk"], df["LOCATION_ID"], margins=False)
3  sns.countplot(x="LOCATION_ID", hue="Risk", data=df)
4  plt.legend(title='Risk', bbox_to_anchor=(1, 1), loc='upper left', ncol=1, fontsize='small', title_fontsize='medium', labelspacing=0.01, borderaxespad=0)
5  plt.xticks(rotation=90)
6  plt.show()
[34]  <1 sec
```

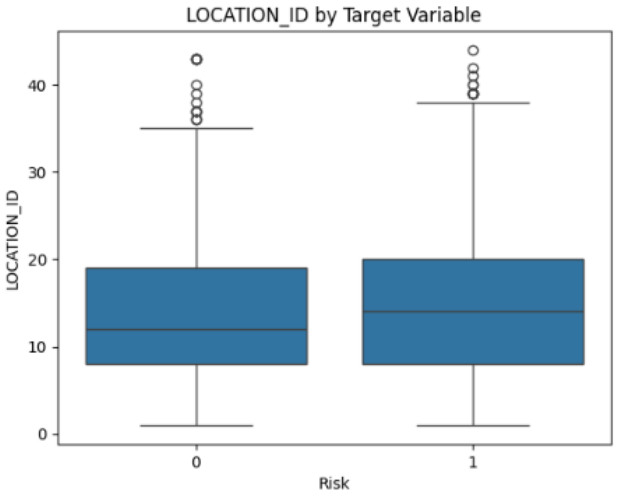


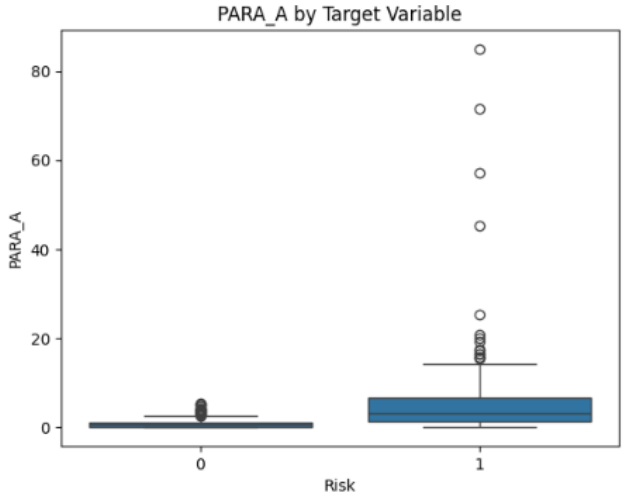
Below boxplot demonstrate the spread and skewness of different features for dataset

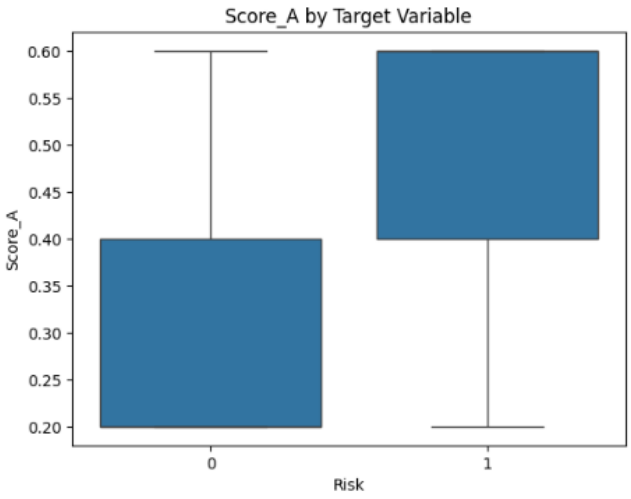
```
for feature in df.columns:
    if feature != 'Risk':
        sns.boxplot(x='Risk', y=feature, data=df)
        plt.title(f'{feature} by Target Variable')
        plt.show()
```

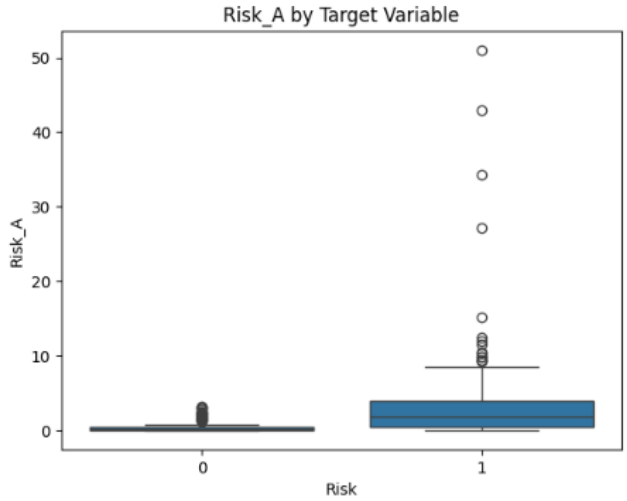
[57] ✓ 3 sec

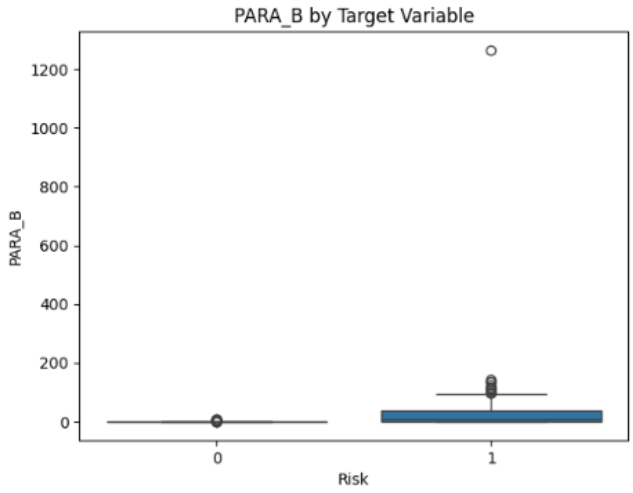


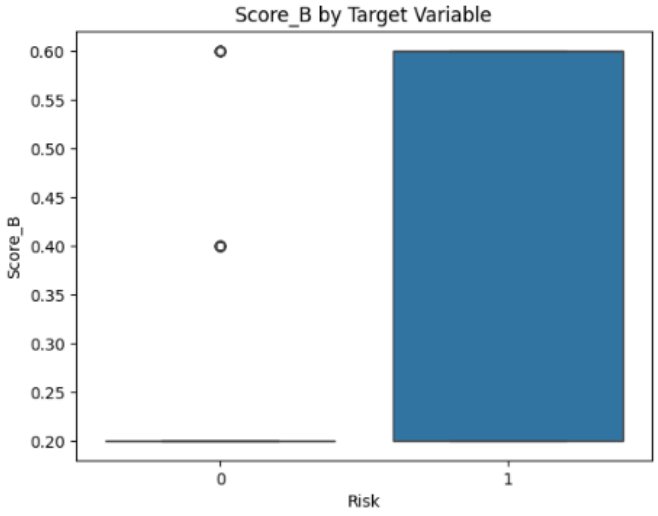


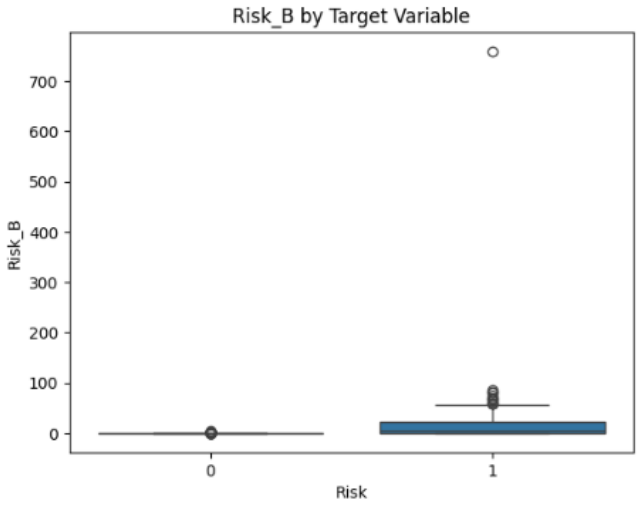


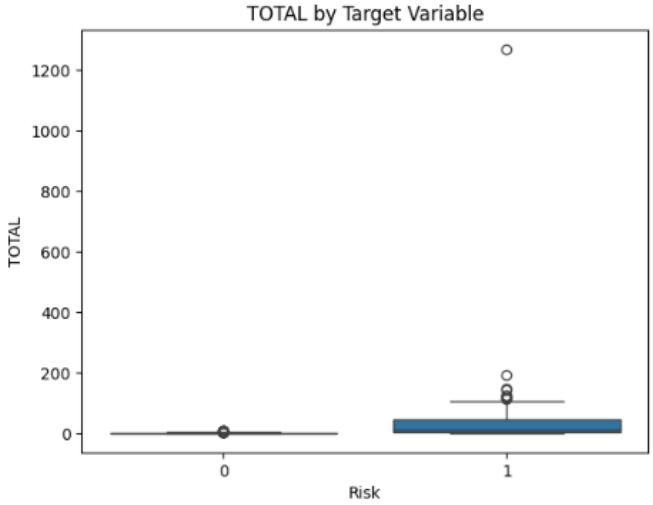


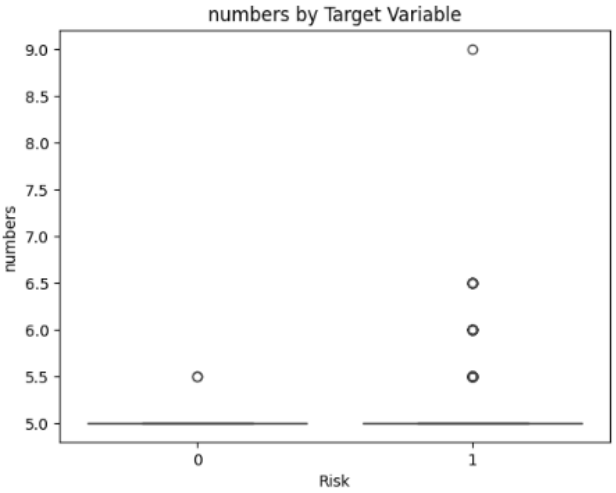


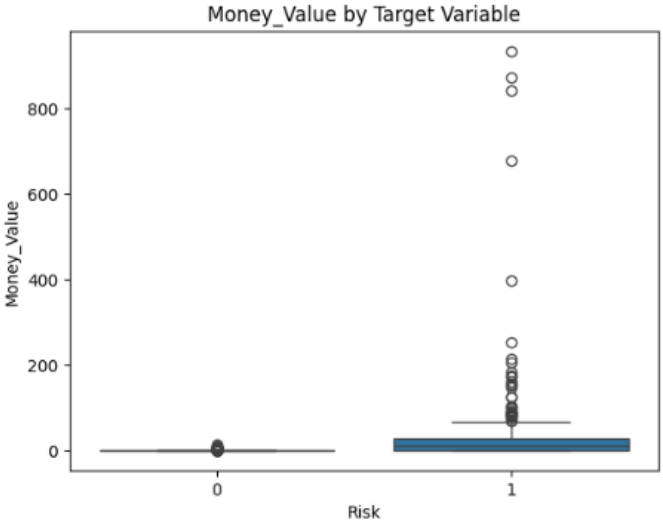


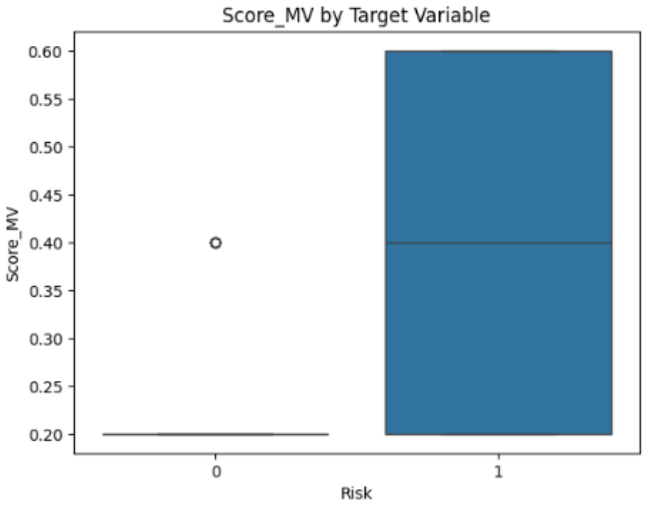

Score_B by Target Variable
Score_B
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0
1
Risk

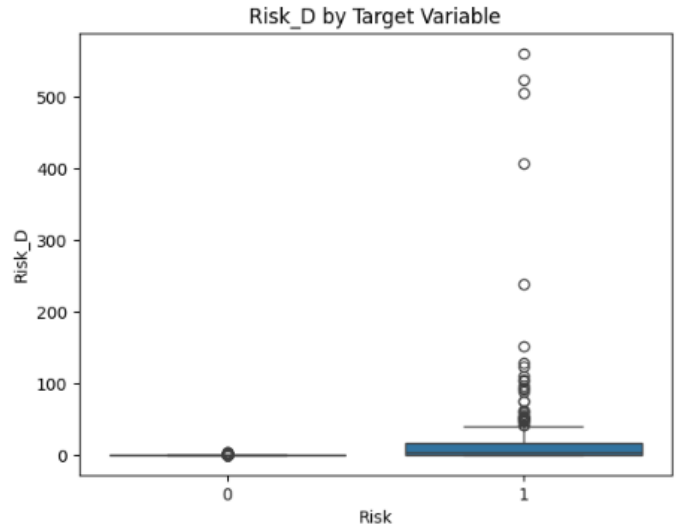

Risk_B by Target Variable
Risk_B
700
600
500
400
300
200
100
0
0
1
Risk

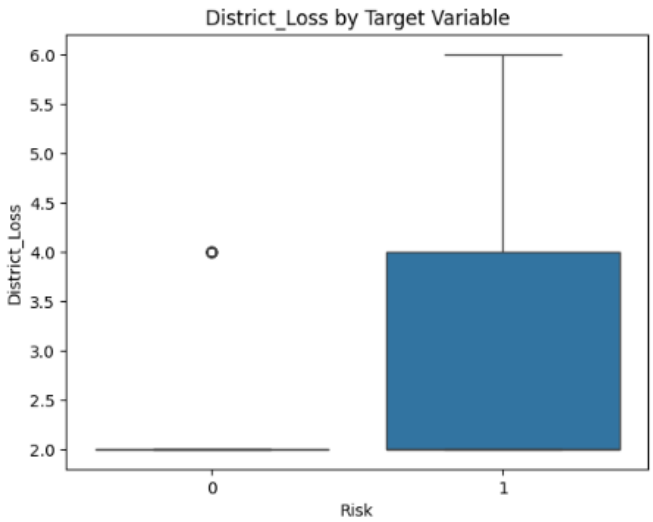

TOTAL by Target Variable
TOTAL
1200
1000
800
600
400
200
0
0
1
Risk

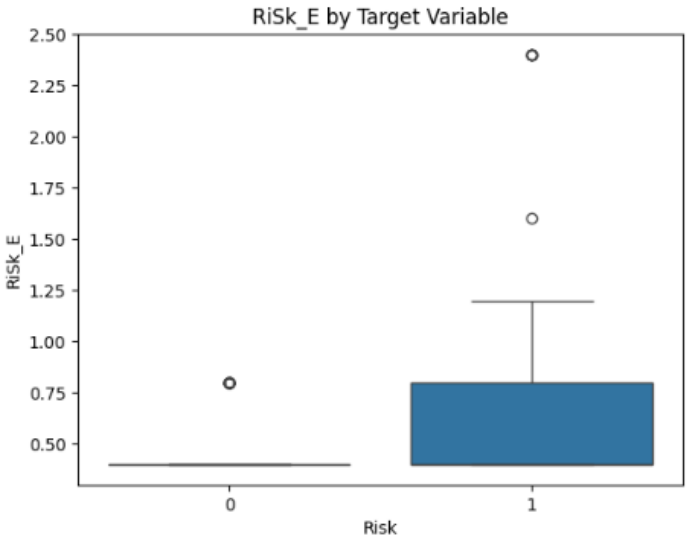

numbers by Target Variable
numbers
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
0
1
Risk

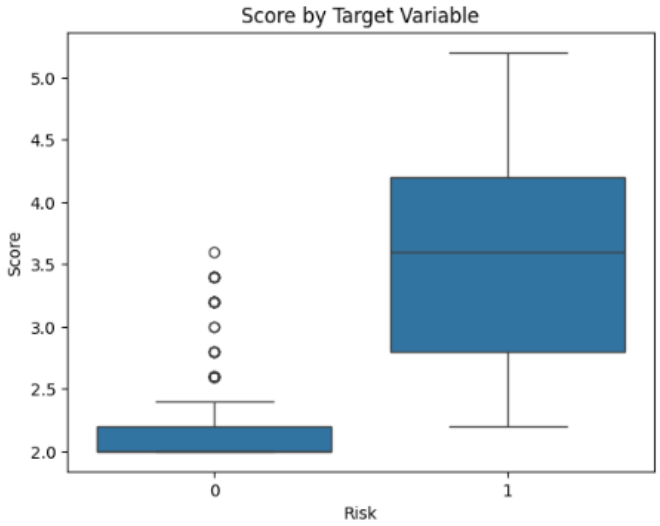

Money_Value by Target Variable
Money_Value
800
600
400
200
0
0
1
Risk

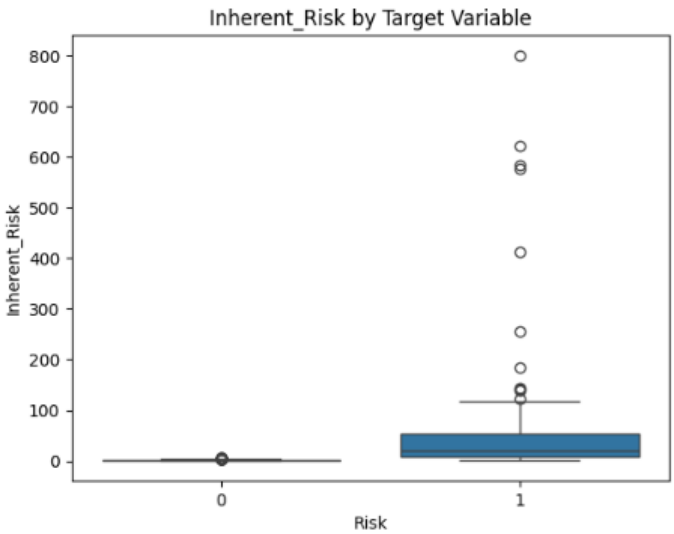

Score_MV by Target Variable
Score_MV
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0
1
Risk

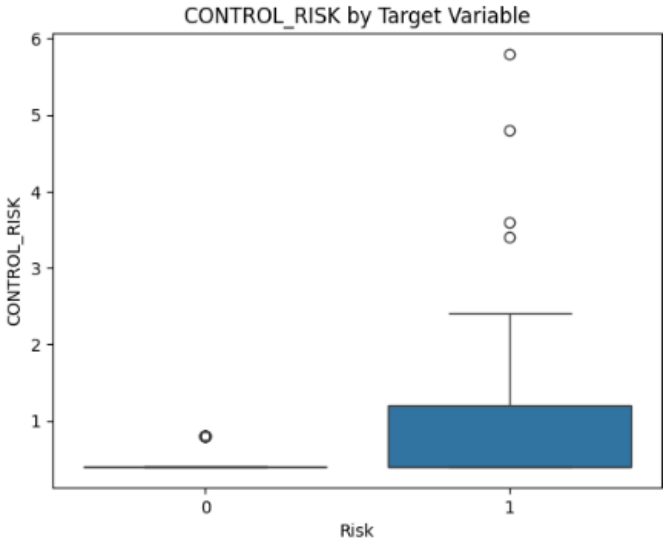










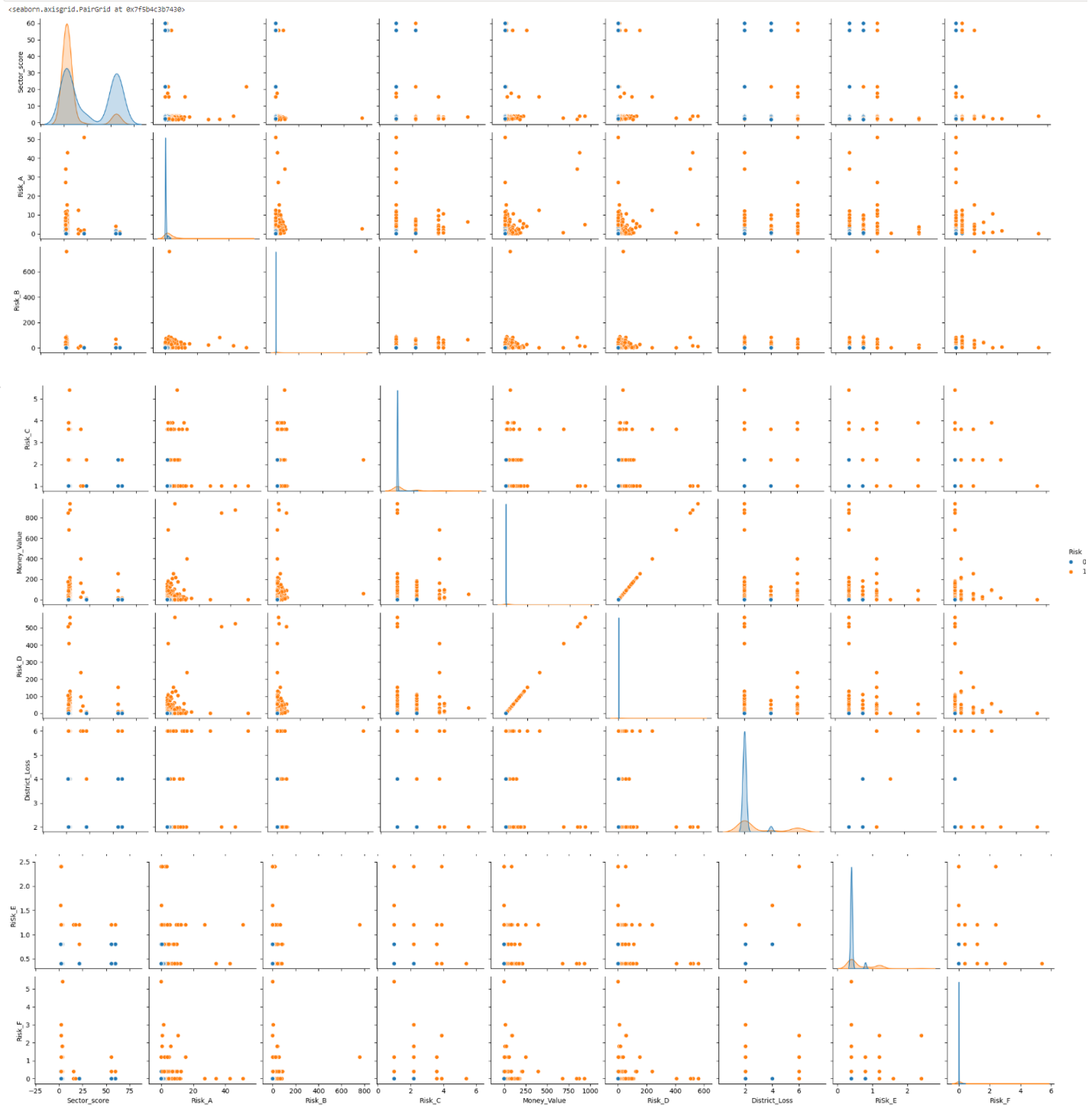


Pairplot and correlation matrix below demonstrated strong correlation between following features

- Risk_A & PARA_A
- Risk_B & TOTAL
- TOTAL & PARA_B

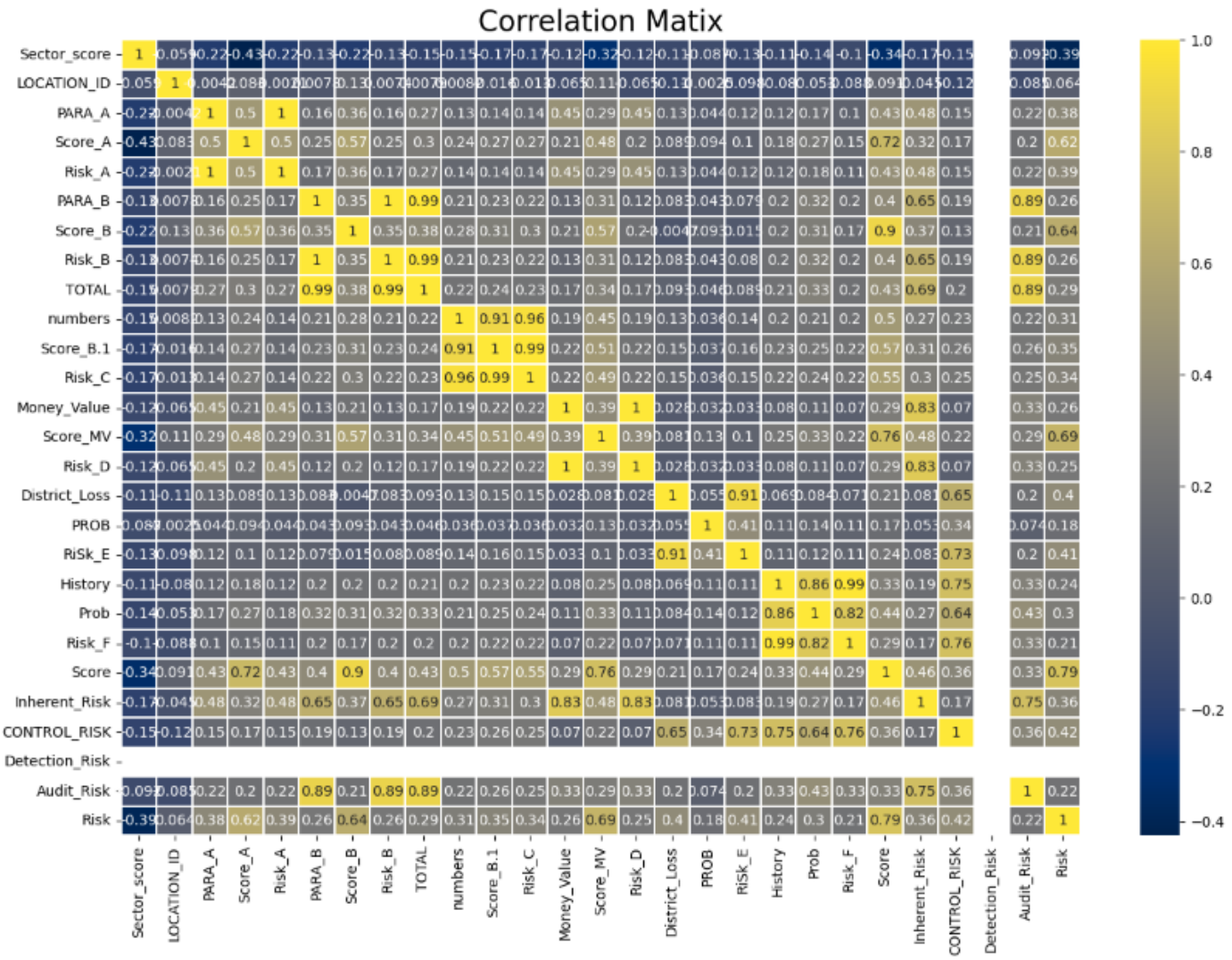
- Score_B.1 & Risk_C
- Risk_C & Number
- Risk_B.1 & Number
- Risk_D & Money Value
- Risk_E & District_Loss
- Prob & history
- History & Risk_F
- Risk_B and Score

```
df1=df.drop(columns=['PARA_A', 'LOCATION_ID', 'Score_A', 'PARA_B', 'Score_B', 'TOTAL', 'numbers', 'Score_B.1', 'Score_MV', 'PROB', 'History', 'Prob', 'Score', 'Inherent_Risk', 'CONTROL_RISK', 'Detection_Risk', 'Audit_Risk' ])
sns.pairplot(df1, hue="Risk")
```



```
df['LOCATION_ID'] = pd.to_numeric(df['LOCATION_ID'], errors='coerce')
df.fillna(df.mean(), inplace=True)
plt.figure(figsize=(15, 10))
sns.heatmap(df.corr(), cmap= 'cividis', annot = True, linewidths=0.01)
plt.title('Correlation Matix', fontsize = 20)
plt.show()
```

✓ 1 sec



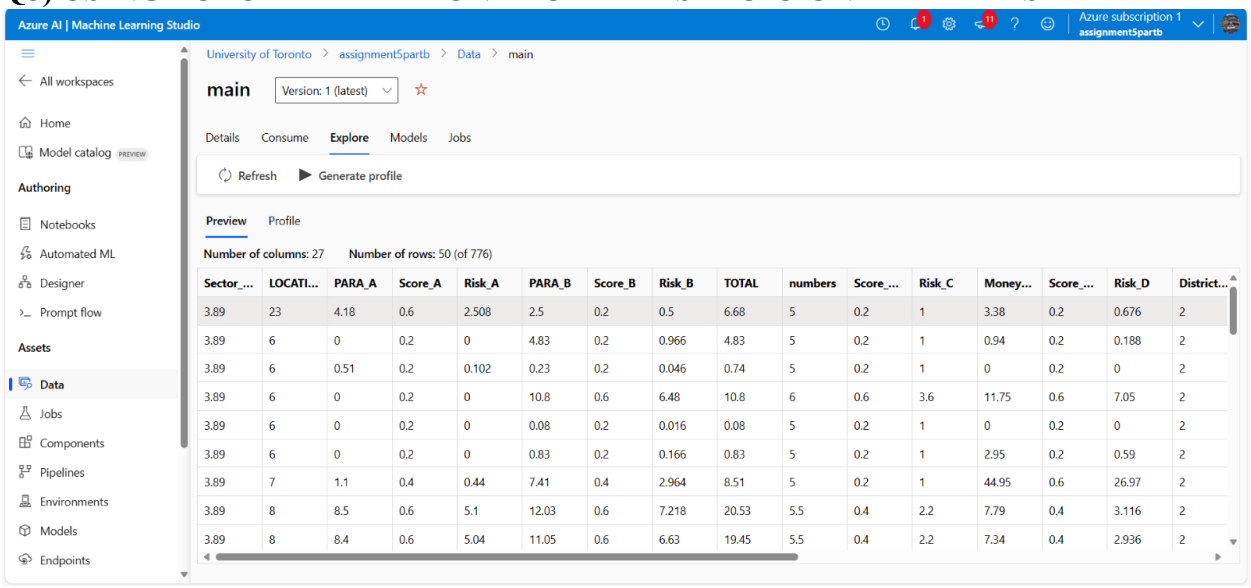
In order to clean the dataset, rows that contain objects are imputed with number.
correlated features are getting dropped from the dataset to avoid complexity of the algorithm and increasing the risk of errors. New dataset with 14 feature and 1 target columns will be used for machine learning purpose.

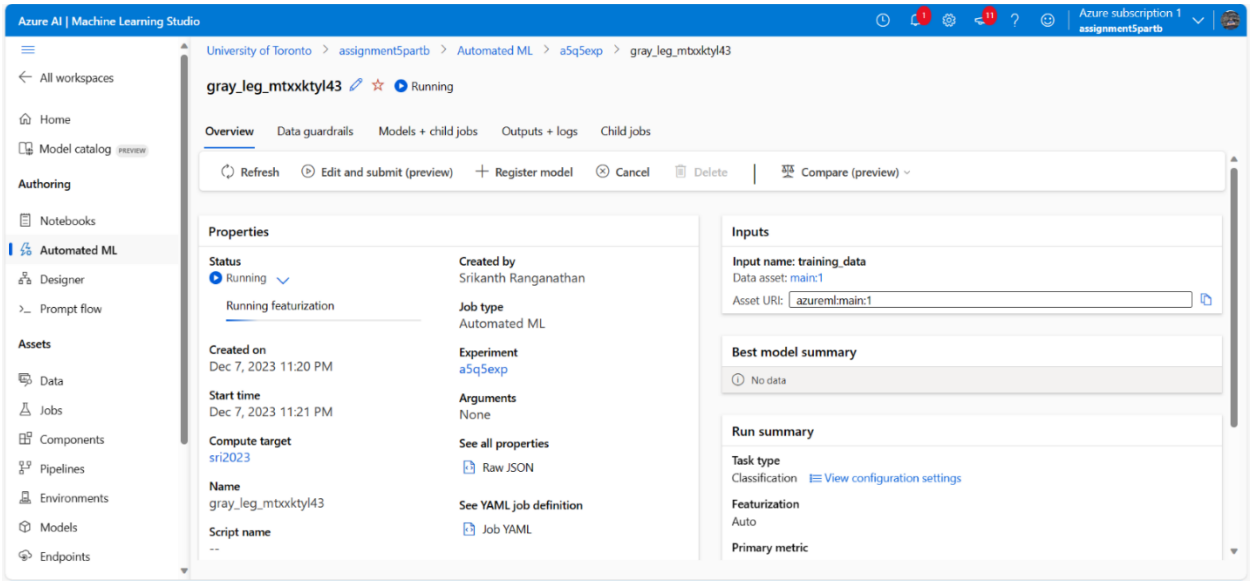
```
# In order to clean the dataset, non value rows are imputed with number
# correlated features are getting dropped from the dataset to avoid complexity of the algorithm and increasing the risk of errors.
df['LOCATION_ID'] = pd.to_numeric(df['LOCATION_ID'], errors='coerce')
df.fillna(df.mean(), inplace=True)
df2=df.drop(columns=['District_Loss', 'Money_Value', 'PARA_A', 'PARA_B', 'TOTAL', 'Score','Score_B.1', 'numbers' , 'PROB', 'Prob', 'History' ,'Audit_Risk'])
df2.describe()
```

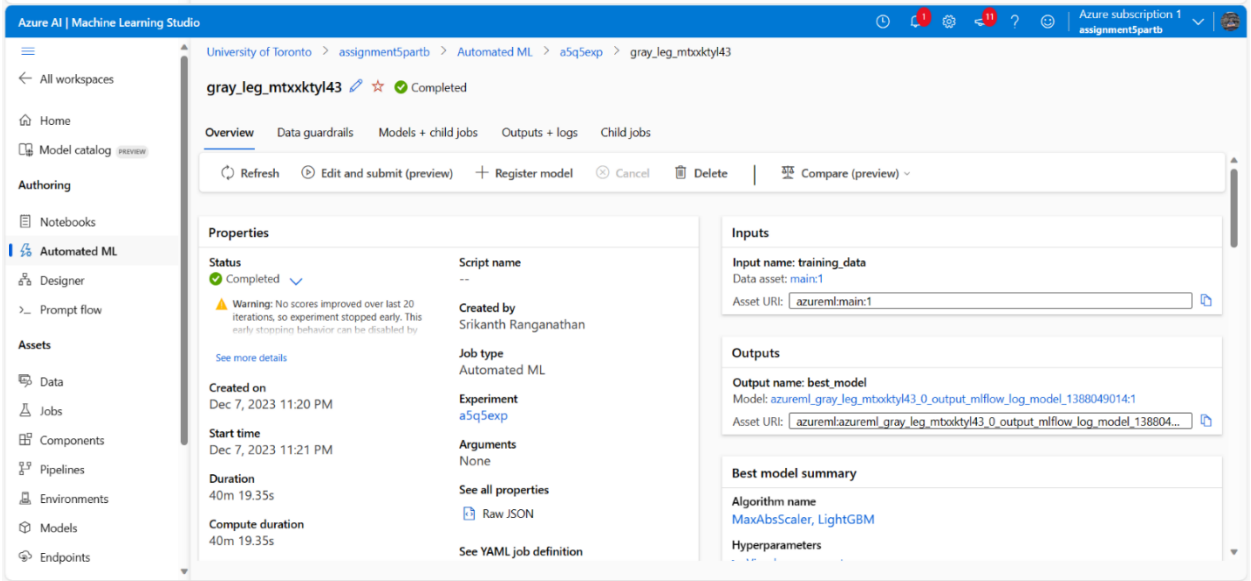
[69] ✓ <1 sec

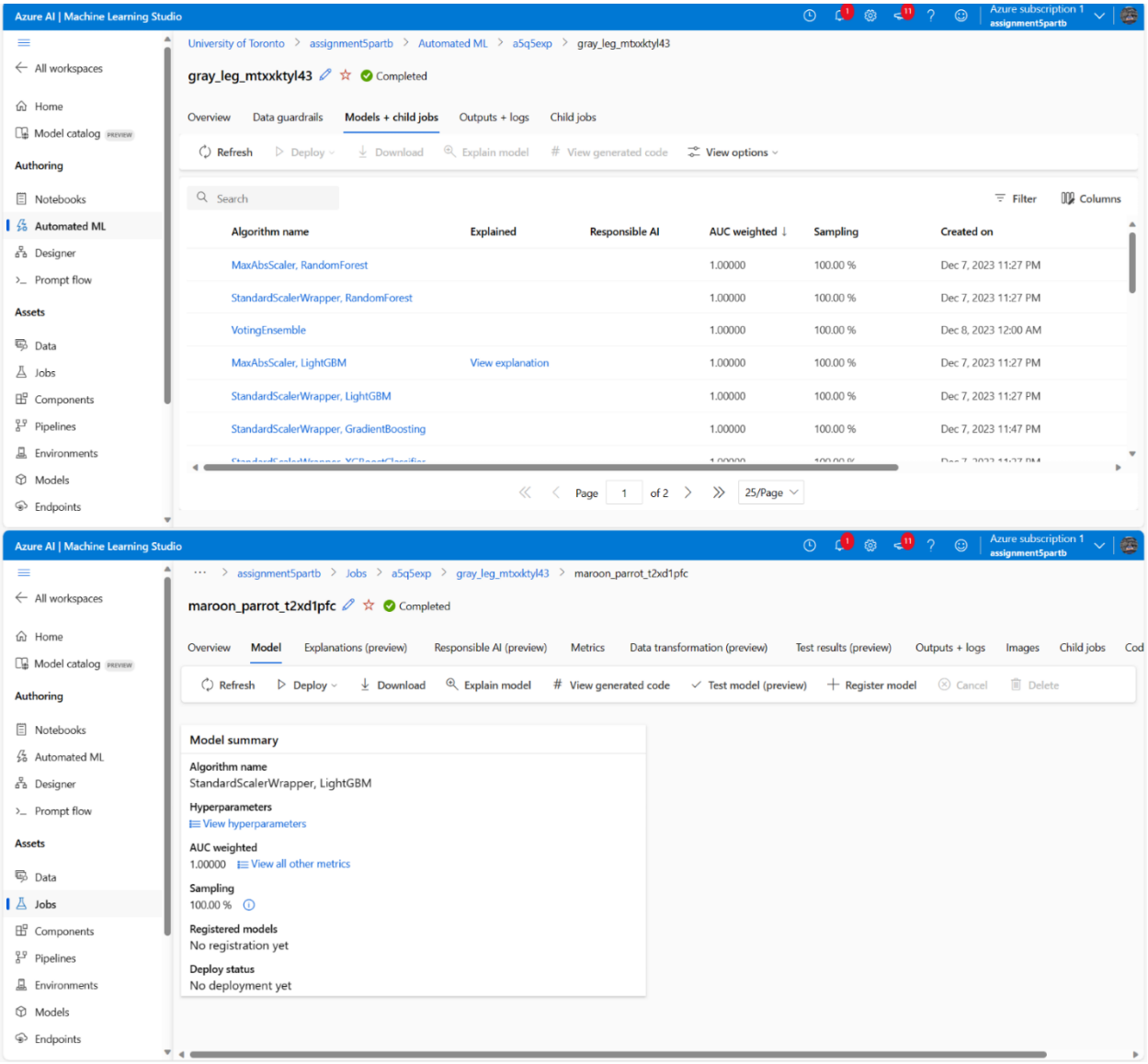
	Sector_score	LOCATION_ID	Score_A	Risk_A	Score_B	Risk_B	Risk_C	Score_MV	Risk_D	RiSk_E	Risk_F	Inherent_Risk	CONTROL_RISK	Detection_Risk	Risk
count	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.0	776.000000
mean	20.184536	14.856404	0.351289	1.351029	0.313144	6.334008	1.152964	0.290979	8.265434	0.519072	0.053608	17.680612	0.572680	0.5	0.393041
std	24.319017	9.872154	0.174055	3.440447	0.169804	30.072845	0.537417	0.159745	39.970849	0.290312	0.305835	54.740244	0.444581	0.0	0.488741
min	1.850000	1.000000	0.200000	0.000000	0.200000	0.000000	1.000000	0.200000	0.000000	0.400000	0.000000	1.400000	0.400000	0.5	0.000000
25%	2.370000	8.000000	0.200000	0.042000	0.200000	0.000000	1.000000	0.200000	0.000000	0.400000	0.000000	1.583500	0.400000	0.5	0.000000
50%	3.890000	13.000000	0.200000	0.175000	0.200000	0.081000	1.000000	0.200000	0.018000	0.400000	0.000000	2.214000	0.400000	0.5	0.000000
75%	55.570000	19.000000	0.600000	1.488000	0.400000	1.840500	1.000000	0.400000	2.235000	0.400000	0.000000	10.663500	0.400000	0.5	1.000000
max	59.850000	44.000000	0.600000	51.000000	0.600000	758.778000	5.400000	0.600000	561.018000	2.400000	5.400000	801.262000	5.800000	0.5	1.000000

Q5) USING AUTOMATED ML ON AZURE ML STDUIO ON THE DATASET

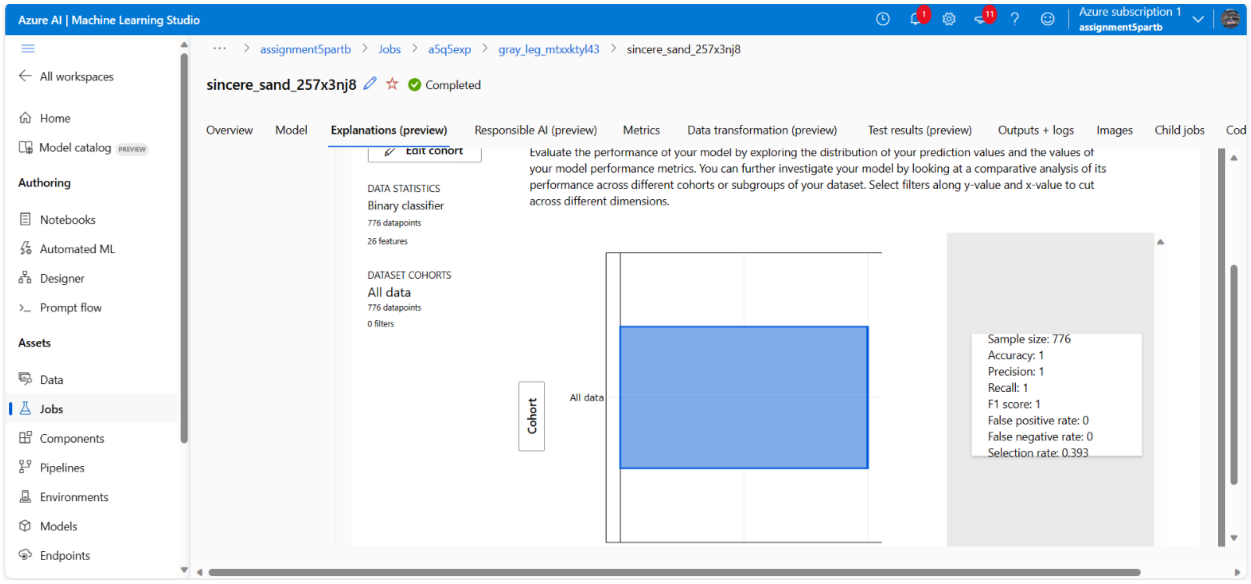








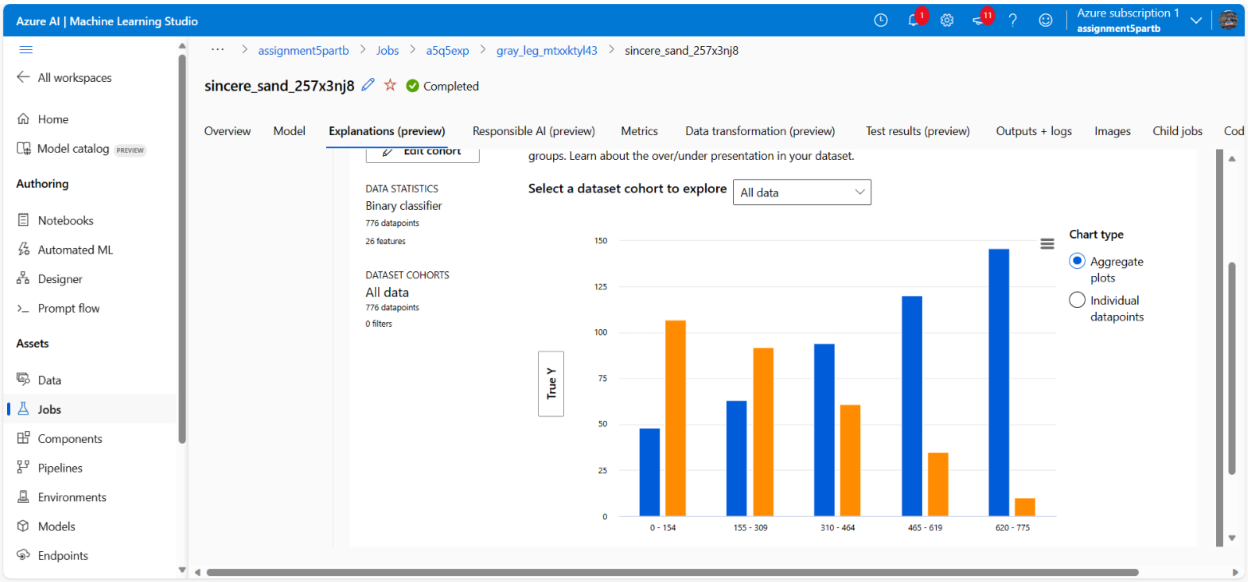
It can be noted that the LightGBM (Light Gradient Boosting Machine) has been concluded as the best model.

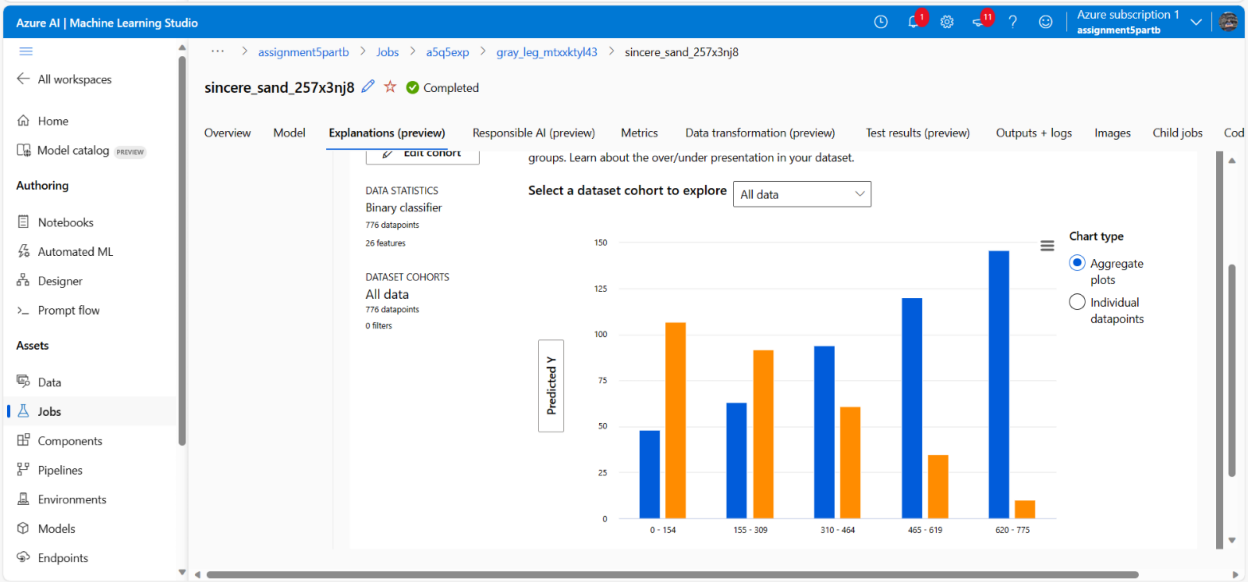


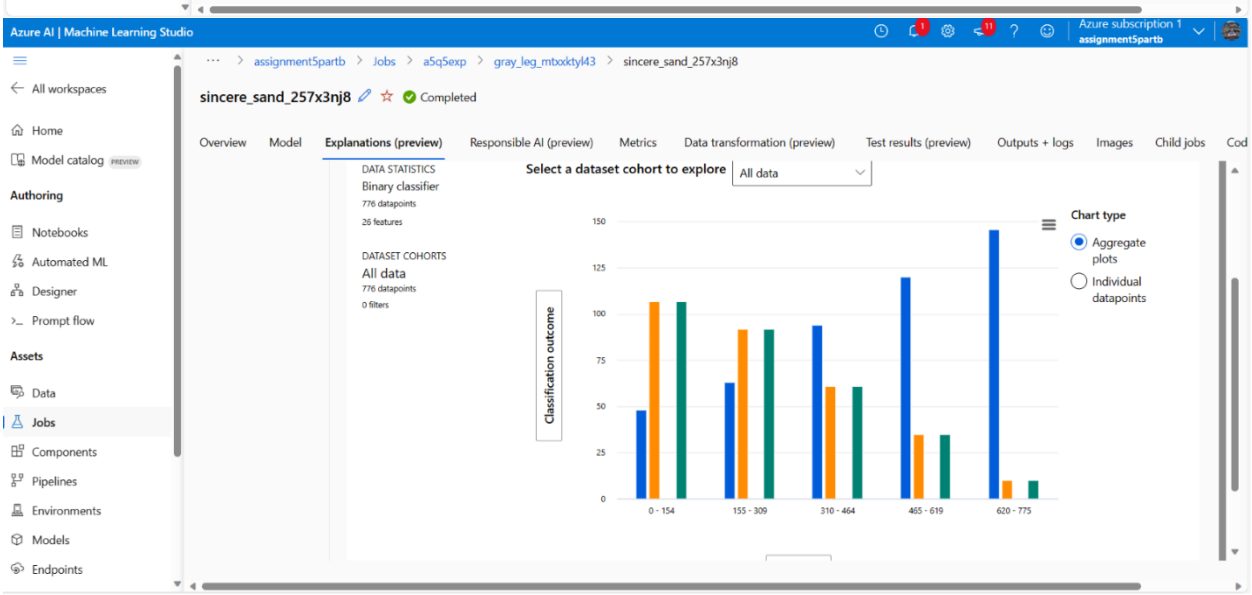
It is very clear as to why this model has been concluded as the best model, it is because of the resulting metric scores such as accuracy, precision, recall, and f1 score, all of them being 1.

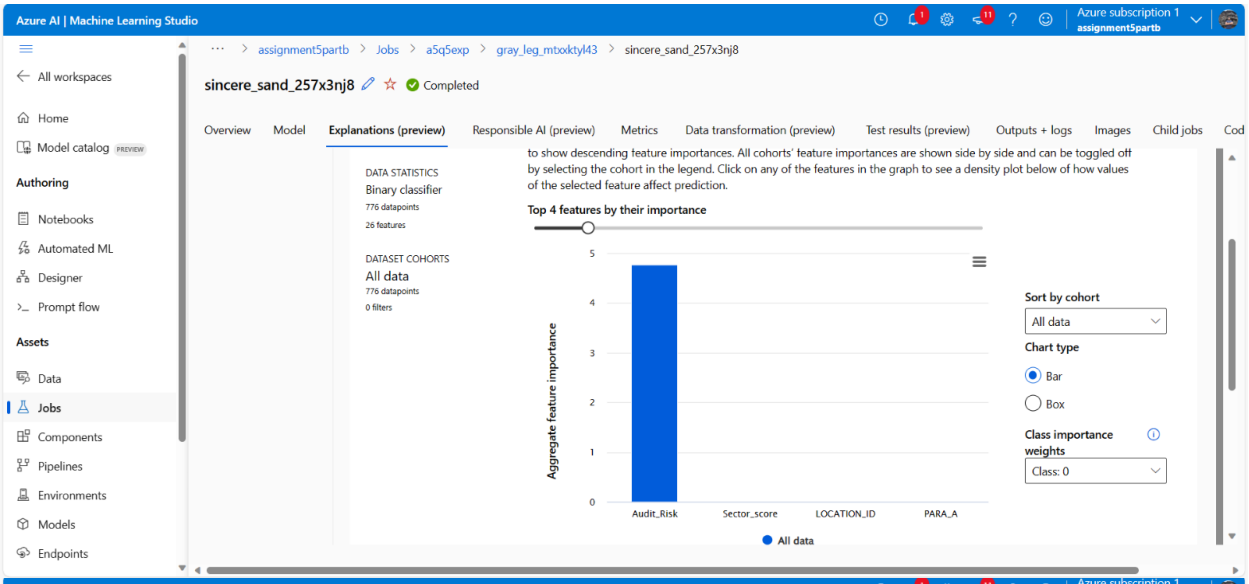
The FP and FN rates are also 0.

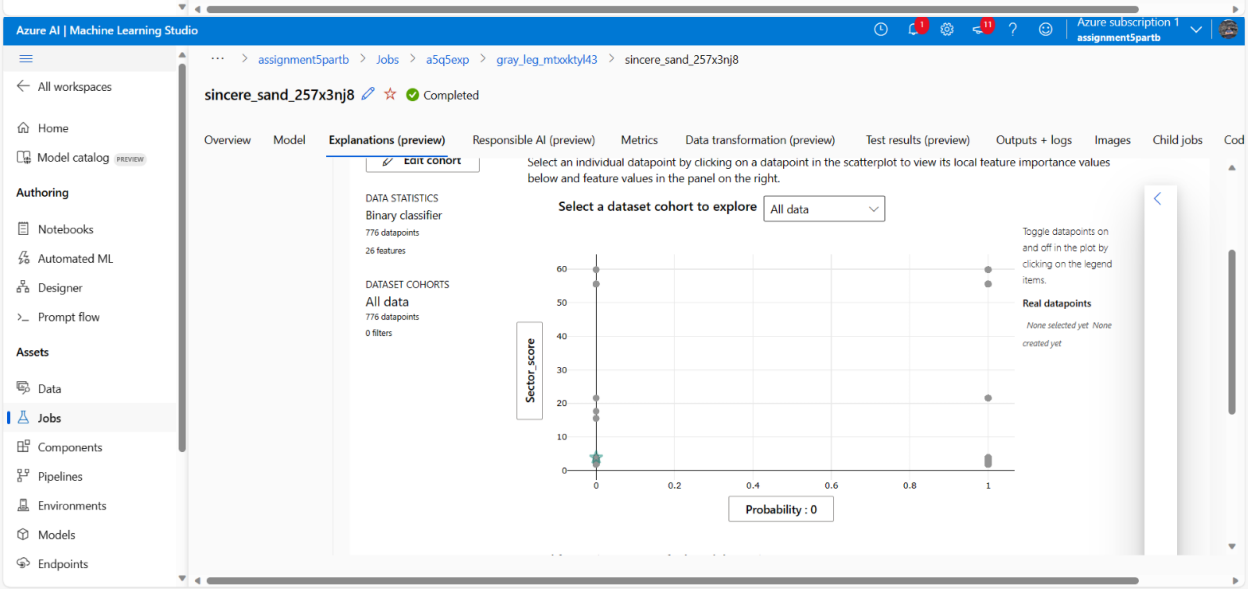
Here are some plots that further validate this:

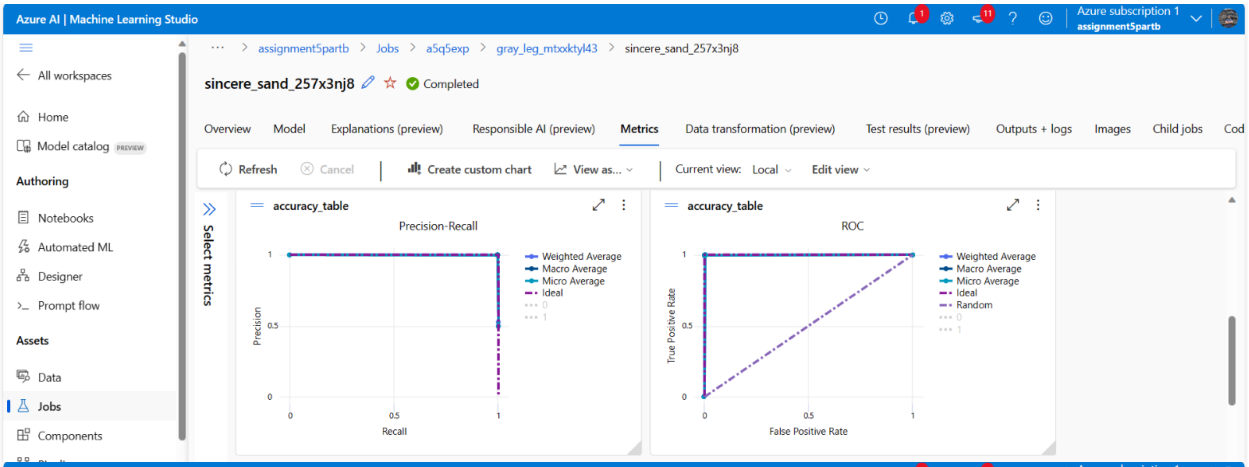

Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtoxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
Edit cohort
groups. Learn about the over/under presentation in your dataset.
DATA STATISTICS
Binary classifier
776 datapoints
26 features
DATASET COHORTS
All data
776 datapoints
0 filters
Select a dataset cohort to explore
All data
Chart type
Aggregate plots
Individual datapoints
True Y
150
125
100
75
50
25
0
0 - 154
155 - 309
310 - 464
465 - 619
620 - 775

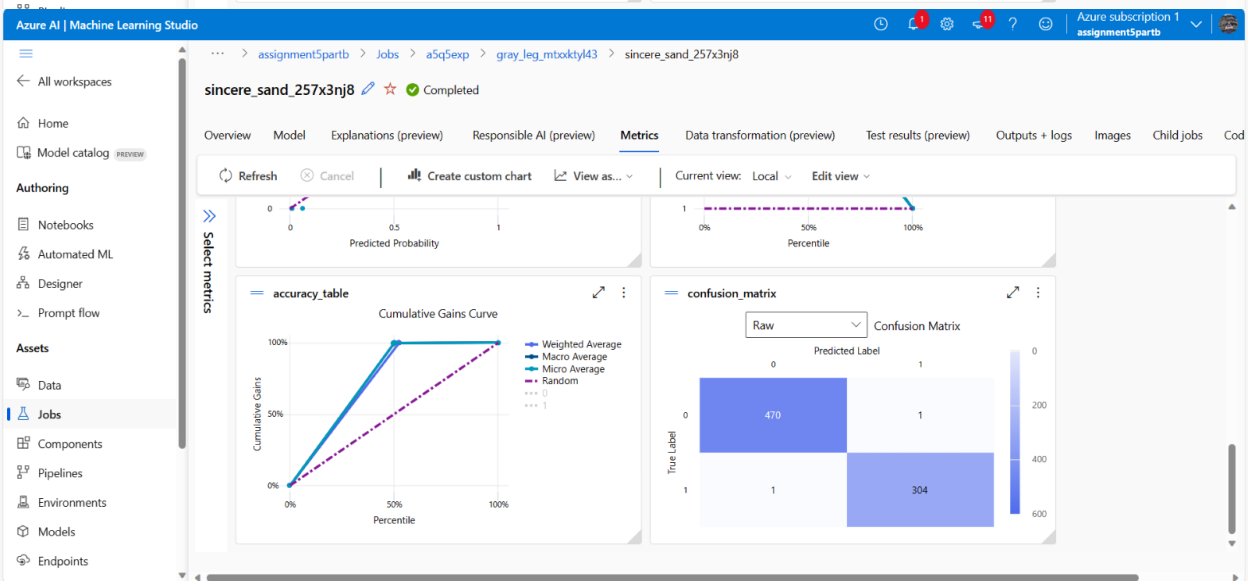

Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtoxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
groups. Learn about the over/under presentation in your dataset.
DATA STATISTICS
Binary classifier
776 datapoints
26 features
DATASET COHORTS
All data
776 datapoints
0 filters
Select a dataset cohort to explore
All data
Chart type
Aggregate plots
Individual datapoints
Predicted Y
150
125
100
75
50
25
0
0 - 154
155 - 309
310 - 464
465 - 619
620 - 775

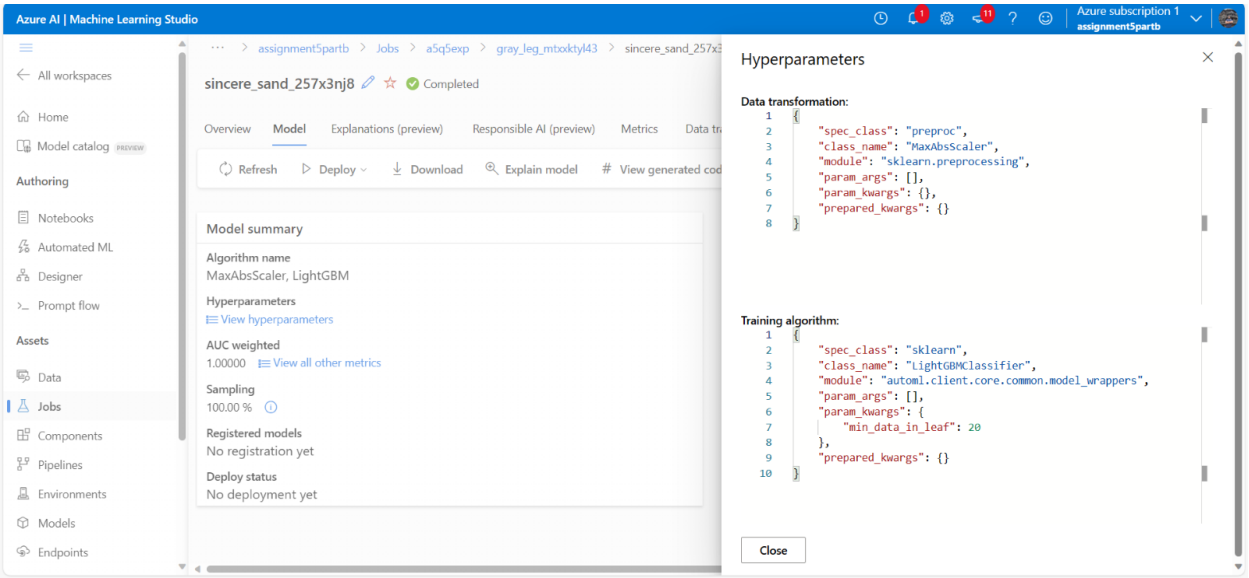

Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtoxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
DATA STATISTICS
Binary classifier
776 datapoints
26 features
DATASET COHORTS
All data
776 datapoints
0 filters
Select a dataset cohort to explore
All data
Chart type
Aggregate plots
Individual datapoints
Classification outcome
150
125
100
75
50
25
0
0 - 154
155 - 309
310 - 464
465 - 619
620 - 775

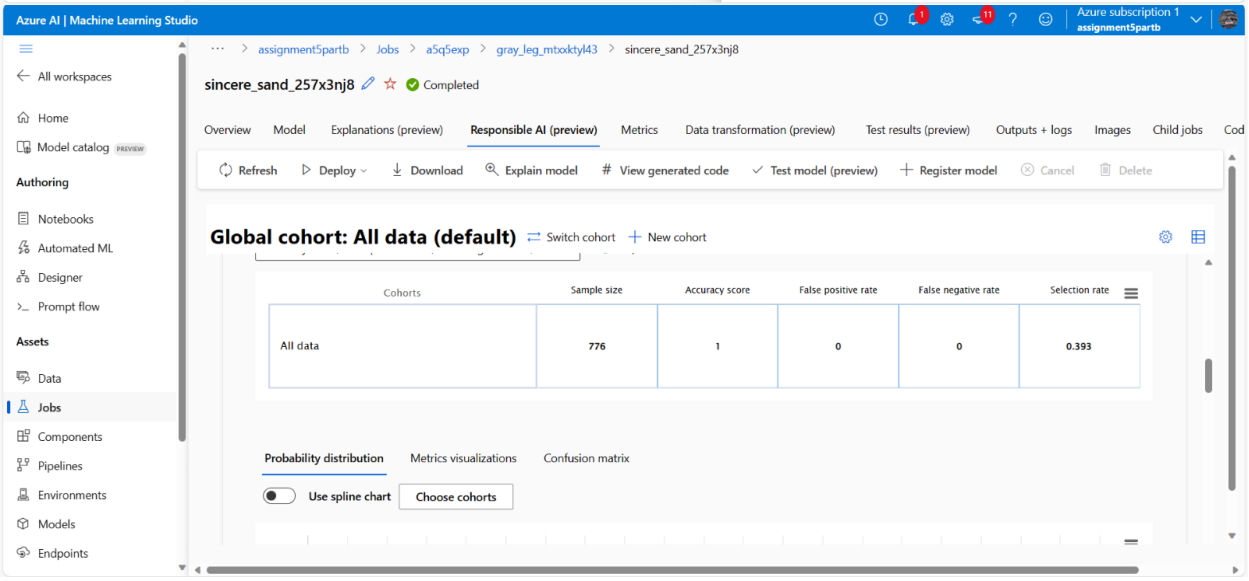

Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtoxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
DATA STATISTICS
Binary classifier
776 datapoints
26 features
DATASET COHORTS
All data
776 datapoints
0 filters
to show descending feature importances. All cohorts' feature importances are shown side by side and can be toggled off by selecting the cohort in the legend. Click on any of the features in the graph to see a density plot below of how values of the selected feature affect prediction.
Top 4 features by their importance
Aggregate feature importance
Audit_Risk
Sector_score
LOCATION_ID
PARA_A
All data
Sort by cohort
All data
Chart type
Bar
Box
Class importance weights
Class: 0

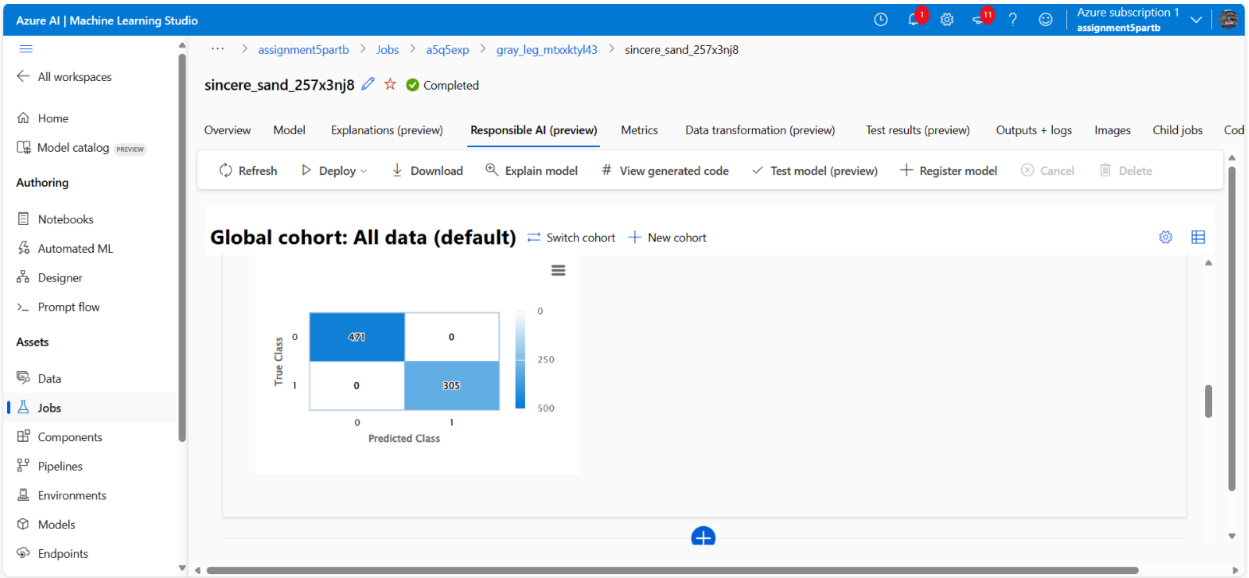

Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtoxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
Edit cohort
DATA STATISTICS
Binary classifier
776 datapoints
26 features
DATASET COHORTS
All data
776 datapoints
0 filters
Select an individual datapoint by clicking on a datapoint in the scatterplot to view its local feature importance values below and feature values in the panel on the right.
Select a dataset cohort to explore
All data
Sector_score
Probability : 0
Toggle datapoints on and off in the plot by clicking on the legend items.
Real datapoints
None selected yet None created yet

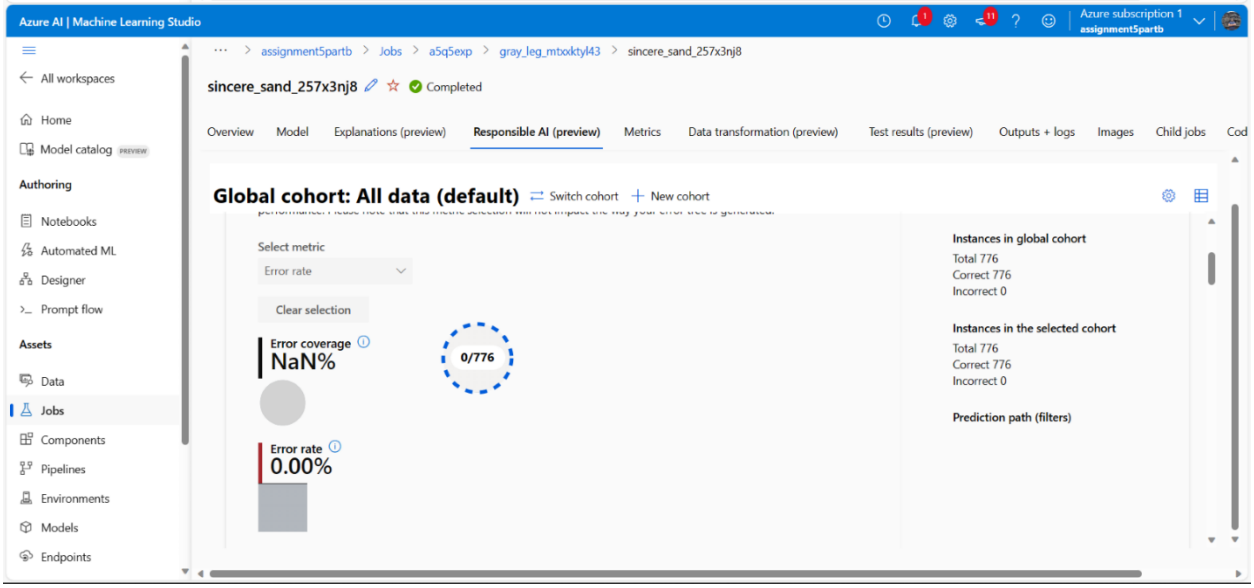

Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
assignment5partb > Jobs > a5q5exp > gray_leg_mtxxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
Refresh
Cancel
Create custom chart
View as...
Current view: Local
Edit view
Select metrics
accuracy_table
Precision-Recall
Precision
Recall
Weighted Average
Macro Average
Micro Average
Ideal
accuracy_table
ROC
True Positive Rate
False Positive Rate
Weighted Average
Macro Average
Micro Average
Ideal
Random


Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
assignment5partb > Jobs > a5q5exp > gray_leg_mtxxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
Refresh
Cancel
Create custom chart
View as...
Current view: Local
Edit view
Select metrics
Predicted Probability
Percentile
accuracy_table
Cumulative Gains Curve
Cumulative Gains
Percentile
Weighted Average
Macro Average
Micro Average
Random
confusion_matrix
Raw
Confusion Matrix
Predicted Label
True Label
470
1
1
304


Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
assignment5partb > Jobs > a5q5exp > gray_leg_mtxxktyl43 > sincere_sand_257x3
sincere_sand_257x3nj8
Completed
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Refresh
Deploy
Download
Explain model
Model summary
Algorithm name
MaxAbsScaler, LightGBM
Hyperparameters
View hyperparameters
AUC weighted
1.00000
View all other metrics
Sampling
100.00 %
Registered models
No registration yet
Deploy status
No deployment yet
Hyperparameters
Data transformation:
{
"spec_class": "preproc",
"class_name": "MaxAbsScaler",
"module": "sklearn.preprocessing",
"param_args": [],
"param_kwargs": {},
"prepared_kwargs": {}
}
Training algorithm:
{
"spec_class": "sklearn",
"class_name": "LightGBMClassifier",
"module": "automl.client.core.common.model_wrappers",
"param_args": [],
"param_kwargs": {
"min_data_in_leaf": 20
},
"prepared_kwargs": {}
}
Close


Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtxxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
Refresh
Deploy
Download
Explain model
View generated code
Test model (preview)
Register model
Cancel
Delete
Global cohort: All data (default)
Switch cohort
New cohort
Cohorts
Sample size
Accuracy score
False positive rate
False negative rate
Selection rate
All data
776
1
0
0
0.393
Probability distribution
Metrics visualizations
Confusion matrix
Use spline chart
Choose cohorts


Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtoxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
Refresh
Deploy
Download
Explain model
View generated code
Test model (preview)
Register model
Cancel
Delete
Global cohort: All data (default)
Switch cohort
New cohort
True Class
Predicted Class
471
0
0
305
0
250
500


Azure AI | Machine Learning Studio
Azure subscription 1
assignment5partb
assignment5partb > Jobs > a5q5exp > gray_leg_mtoxktyl43 > sincere_sand_257x3nj8
sincere_sand_257x3nj8
Completed
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Overview
Model
Explanations (preview)
Responsible AI (preview)
Metrics
Data transformation (preview)
Test results (preview)
Outputs + logs
Images
Child jobs
Global cohort: All data (default)
Switch cohort
New cohort
Select metric
Error rate
Clear selection
Error coverage
NaN%
0/776
Error rate
0.00%
Instances in global cohort
Total 776
Correct 776
Incorrect 0
Instances in the selected cohort
Total 776
Correct 776
Incorrect 0
Prediction path (filters)