# Driverless AI Experiment: 7.Test

Generated by: user

Generated on: Wed Oct 27 09:15:52 2021

[Driverless AI Experiment: 7.Test 1](#_Toc55907577)

[Experiment Overview 1](#_Toc55907578)

[Data Overview 1](#_Toc55907579)

[Methodology 1](#_Toc55907580)

[Data Sampling 1](#_Toc55907581)

[Validation Strategy 2](#_Toc55907582)

[Model Tuning 2](#_Toc55907583)

[Feature Evolution 2](#_Toc55907584)

[Feature Transformations 2](#_Toc55907585)

[Final Model 2](#_Toc55907586)

[Alternative Models 2](#_Toc55907587)

[Deployment 2](#_Toc55907588)

[Partial Dependence Plots 2](#_Toc55907589)

[Shapley Values 2](#_Toc55907590)

[Appendix 3](#_Toc55907593)

## Experiment Overview

Driverless AI built 1 XGBoostGBMModel to predict *DEFAULT\_PAYMENT\_NEXT\_MONTH* given 24 original features from the input dataset *CreditCard\_Cat-train.csv*. This classification experiment completed in 1 minutes and 57 seconds (0:01:57), using 19 of the 24 original features, and 3 of the 3 engineered features.

### Performance

|  |  |
| --- | --- |
| **Dataset** | **AUC** |
| Internal Validation | 0.778 |
| Test Data | 0.796 |

### Driverless Settings

|  |  |  |  |
| --- | --- | --- | --- |
| **Dial Settings** | **Description** | **Setting Value** | **Range of Possible Values** |
| **Accuracy** | Controls accuracy needs of the model | 1 | 1-10 |
| **Time** | Controls duration of the experiment | 1 | 1-10 |
| **Interpretability** | Controls complexity of the model | 1 | 1-10 |

### System Specifications

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Address** | **System** | **System Memory** | **CPUs** | **GPUs** |
| http://127.0.0.1:12350 | Linux | 125 GB | 40 | 2 |

### Versions

|  |  |
| --- | --- |
| **Driverless AI version** | 1.10.1+local\_dev-placeholder-2240-g8bd2dc928a-dirty |
| **h2o4gpu version** | 0.4.2 |
| **h2o\_mli version** | 1.10.4 |
| **mojo2\_runtime version** | 2.7.3-master.362 |
| **procsy version** | 0.9.0 |
| **pydatatable version** | 1.1.0a2080 |
| **vis\_data\_server version** | 2.1.0-rc2 |

## Data Overview

This section provides information on the datasets used for the experiment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **data** | **file path** | **file size** | **number of rows** | **number of columns** |
| training | ./tmp/user/130321f4-35b5-11ec-b34c-ac1f6b6b49a6/CreditCard\_Cat-train.csv.1635181236.9401703.bin | 2.7 MiB | 23,999 | 25 |
| validation | Not provided | None | None | None |
| testing | ./tmp/user/0bbed88e-35b5-11ec-b34c-ac1f6b6b49a6/CreditCard\_Cat-test.csv.1635181224.7546077.bin | 691.3 KiB | 6,000 | 25 |

### Training Data

The training data consists of both numeric and categorical columns.

The summary of the columns is shown below:

#### Numeric Columns

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **name** | **logical\_type** | **storage\_type** | **min** | **mean** | **max** | **std** | **unique** | **freq of mode** |
| ID | id | int | 1.000 | 12,000.000 | 23,999.000 | 6,928.059 | 23,999 | 1 |
| LIMIT\_BAL | numeric, categorical, catlabel | int | 10,000.000 | 165,498.716 | 1,000,000.000 | 129,130.743 | 79 | 2,740 |
| AGE | numeric, categorical, catlabel | int | 21.000 | 35.381 | 79.000 | 9.271 | 55 | 1,284 |
| PAY\_0 | numeric, categorical, catlabel, ohe\_categorical | int | -2.000 | -0.002 | 8.000 | 1.127 | 11 | 11,732 |
| PAY\_2 | numeric, categorical, catlabel, ohe\_categorical | int | -2.000 | -0.123 | 8.000 | 1.201 | 11 | 12,543 |
| PAY\_3 | numeric, categorical, catlabel, ohe\_categorical | int | -2.000 | -0.155 | 8.000 | 1.204 | 11 | 12,576 |
| PAY\_4 | numeric, categorical, catlabel, ohe\_categorical | int | -2.000 | -0.212 | 8.000 | 1.167 | 11 | 13,250 |
| PAY\_5 | numeric, categorical, catlabel, ohe\_categorical | int | -2.000 | -0.253 | 8.000 | 1.137 | 10 | 13,520 |
| PAY\_6 | numeric, categorical, catlabel, ohe\_categorical | int | -2.000 | -0.278 | 8.000 | 1.158 | 10 | 12,876 |
| BILL\_AMT1 | numeric | int | -165,580.000 | 50,598.929 | 964,511.000 | 72,650.198 | 18,717 | 1,607 |
| BILL\_AMT2 | numeric | int | -69,777.000 | 48,648.047 | 983,931.000 | 70,365.396 | 18,367 | 2,049 |
| BILL\_AMT3 | numeric | int | -157,264.000 | 46,368.904 | 1,664,089.000 | 68,194.720 | 18,131 | 2,325 |
| BILL\_AMT4 | numeric | int | -170,000.000 | 42,369.873 | 891,586.000 | 63,071.455 | 17,719 | 2,547 |
| BILL\_AMT5 | numeric | int | -81,334.000 | 40,002.333 | 927,171.000 | 60,345.728 | 17,284 | 2,840 |
| BILL\_AMT6 | numeric | int | -339,603.000 | 38,565.267 | 961,664.000 | 59,156.501 | 16,906 | 3,258 |
| PAY\_AMT1 | numeric | int | 0.000 | 5,543.098 | 505,000.000 | 15,068.863 | 6,918 | 4,270 |
| PAY\_AMT2 | numeric | int | 0.000 | 5,815.529 | 1,684,259.000 | 20,797.444 | 6,839 | 4,362 |
| PAY\_AMT3 | numeric | int | 0.000 | 4,969.431 | 896,040.000 | 16,095.929 | 6,424 | 4,853 |
| PAY\_AMT4 | numeric | int | 0.000 | 4,743.657 | 497,000.000 | 14,883.555 | 6,028 | 5,200 |
| PAY\_AMT5 | numeric | int | 0.000 | 4,783.644 | 417,990.000 | 15,270.704 | 5,984 | 5,407 |
| PAY\_AMT6 | numeric | int | 0.000 | 5,189.574 | 528,666.000 | 17,630.719 | 5,988 | 5,846 |

#### Boolean Columns

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **name** | **logical\_type** | **storage\_type** | **min** | **mean** | **max** | **std** | **freq of max value** |
| DEFAULT\_PAYMENT\_NEXT\_MONTH | N/A | bool | False | 0.2237 | True | 0.4167 | 5,369 |

#### Categorical Columns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **name** | **logical\_type** | **storage\_type** | **unique** | **top** | **freq of top value** |
| SEX | categorical, catlabel, ohe\_categorical | str | 2 | female | 8,921 |
| EDUCATION | categorical, catlabel, ohe\_categorical | str | 4 | university | 11,360 |
| MARRIAGE | categorical, catlabel, ohe\_categorical | str | 4 | single | 12,876 |

### Shifts Detected

Driverless AI can perform shift detection between the training, validation, and testing datasets. It does this by training a binomial model to predict which dataset a record belongs to. For example, it may find that it is able to separate the training and testing data with an AUC of 0.8 using only the column: C1 as the predictor. This indicates that there is some sort of drift in the distribution of C1 between the training and testing data.

For this experiment, Driverless AI checked the train and test data for any shift in distributions but found none. This indicates that all the predictors/columns in the train and test data are from the same distribution.

## Methodology

This section describes the experiment methodology.

### Assumptions and Limitations

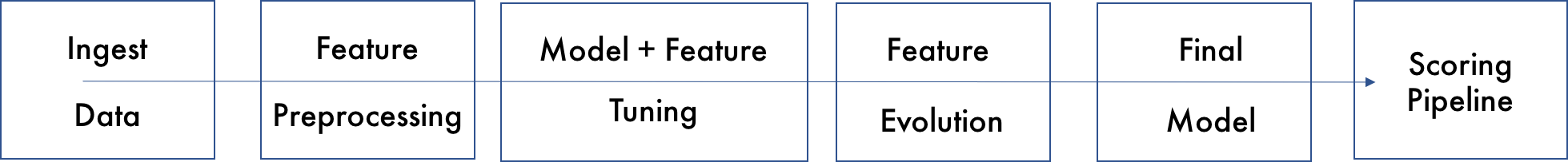
Driverless AI trains all models based on the training data provided (in this case: *CreditCard\_Cat-train.csv*). It is the assumption of Driverless AI that this dataset is representative of the data that will be seen when scoring.

Driverless AI may perform shift detection between the train and test data. If a shift in distribution is detected, this may indicate that the data that will be used for scoring may have distributions not represented in the training data.

For this experiment, Driverless AI performed shift detection but found no significant changes in the distribution of the train and test data.

### Experiment Pipeline

For this experiment, Driverless AI performed the following steps to find the optimal final model:



The steps in this pipeline are described in more detail below:

* **Ingest Data**
* detected column types
* **Feature Preprocessing**
* turned raw features into numeric
* **Model and Feature Tuning**

This stage combines random hyperparameter tuning with feature selection and generation. Features in each iteration are updated using variable importance from the previous iteration as a probabilistic prior to decide what new features to create. The best performing model and features are then passed to the feature evolution stage.

* found the optimal parameters for lightgbm, xgboost and constant models by training models with different parameters
* the best parameters are those that generate the largest **AUC** on the internal validation data
* 5 models trained and scored to evaluate features and model parameters
* **Feature Evolution**

This stage uses a genetic algorithm to find the best set of model parameters and feature transformations to be used in the final model. This experiment did not perform the Feature Evolution stage due to the experiment's configurations.

* **Final Model**
* created the best model from the feature engineering iterations
* no stacked ensemble is done due to accuracy or ensemble level settings (consider increasing accuracy or the ensemble\_level)
* **Create Scoring Pipeline**
* created and exported the MOJO and Python scoring pipeline
* MOJO Scoring Pipeline: user/h2oai\_experiment\_dd901dc8-3740-11ec-9ec3-ac1f6b6b49a6/mojo\_pipeline/mojo.zip
* Python Scoring Pipeline: user/h2oai\_experiment\_dd901dc8-3740-11ec-9ec3-ac1f6b6b49a6/scoring\_pipeline/scorer.zip

**Models for Optimization**

Driverless AI trained models throughout the experiment to determine the best parameters, model dataset, and optimal final model. The stages are described below:

|  |  |  |
| --- | --- | --- |
| **Driverless AI Stage** | **Timing (seconds)** | **Number of Models** |
| **Data Preparation** | 12.80 | 0 |
| **Model and Feature Tuning** | 9.88 | 5 |
| **Feature Evolution** | 1.25 | 0 |
| **Final Pipeline Training** | 31.01 | 7 |

### Experiment Settings

Below are the settings selected for the experiment by user. The Defined Parameters represent the high-level parameters.

**Defined Parameters**

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| is\_classification | True |
| enable\_gpus | True |
| seed | False |
| accuracy | 1 |
| time | 1 |
| interpretability | 1 |
| num\_prediction\_periods | None |
| num\_gap\_periods | None |
| is\_timeseries | False |
| is\_image | False |

These Accuracy, Time, and Interpretability settings map to the following internal configuration of the Driverless AI experiment:

|  |  |
| --- | --- |
| **Internal Parameter** | **Value** |
| data filtered | False |
| number of feature engineering iterations | 1 |
| number of models trained per iteration | 2 |
| early stopping rounds | 0 |
| monotonicity constraint | False |
| number of model tuning model combinations | 4 |
| number of base learners in ensemble | 0 |
| time column | [OFF] |

#### Details

* **data filtered:** Driverless AI may filter the training data depending on the number of rows and the Accuracy setting.
* for this experiment, the training data was not filtered.
* **number of feature engineering iterations**: the number of iterations performed of feature engineering.
* **number of models evaluated per iteration:** for each feature engineering iteration, Driverless AI trains multiple models. Each model is trained with a different set of predictors or features. The goal of this step is to determine which types of features lead to the largest AUC.
* **early stopping rounds:** if Driverless AI does not see any improvement after 0 iterations of feature engineering, the feature engineering step is automatically stopped.
* **monotonicity constraint:** if enabled, the models will only have monotone relationships between the predictors and target variable.
* **number of model tuning combinations:** the number of model tuning combinations evaluated to determine the optimal model settings for the lightgbm, xgboost and constant models.
* **number of base learners in ensemble:** the number of base models used to create the final ensemble.
* **time column:** the column that provides the time column. If a time column is provided, feature engineering and model validation will respect the causality of time. If the time column is turned off, no time order is used for modeling and data may be shuffled randomly (any potential temporal causality will be ignored).

## Data Sampling

In Driverless AI, data sampling is a pre-processing step that is done before model training begins; it is not related to sampling done during model training. Driverless AI does not perform data sampling unless the dataset is big or highly imbalanced. Whether a dataset is considered big depends on the experiment's accuracy setting and the *statistical\_threshold\_data\_size\_large* config.toml parameter.

Driverless AI did not perform any down sampling of the data.

## Validation Strategy

Driverless AI automatically split the training data to determine the performance of the model parameter tuning and feature engineering stages. For the experiment, Driverless AI split the data into 2/3 training and 1/3 validation using stratified sampling.

## Model Tuning

The table below shows the score and training time of the lightgbm, xgboost and constant models evaluated by Driverless AI. The table shows the parameter tuning models evaluated, ordered based on a combination of largest score and lowest training time.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **job order** | **booster** | **nfeatures** | **scores** | **training times** |
| 1 | gbtree | 22 | 0.7784 | 1.3246 |
| 2 | lightgbm | 23 | 0.7733 | 1.0477 |
| 0 | lightgbm | 22 | 0.7683 | 1.0696 |
| 3 | constant | 1 | 0.5 | 0.2331 |

More detailed information on the parameters evaluated for each algorithm is shown below.

### lightgbm tuning

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **tree method** | **grow policy** | **max depth** | **max leaves** | **colsample bytree** | **subsample** | **nfeatures** | **scores** | **training times** |
|  | depthwise | 6.0 | 64.0 | 0.8 | 0.7 | 23 | 0.7733 | 1.0477 |
|  | depthwise | 6.0 | 64.0 | 0.8 | 0.7 | 22 | 0.7683 | 1.0696 |

### gbtree tuning

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **tree method** | **grow policy** | **max depth** | **max leaves** | **colsample bytree** | **subsample** | **nfeatures** | **scores** | **training times** |
| gpu\_hist | depthwise | 6.0 | 64.0 | 0.8 | 0.7 | 22 | 0.7784 | 1.3246 |

### constant tuning

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **job order** | **booster** | **nfeatures** | **scores** | **training times** |
| 3 | constant | 1 | 0.5 | 0.2331 |

## Feature Evolution

The goal of the Feature Evolution stage is to determine the best features to use for the final model. This experiment did not perform the Feature Evolution stage due to the experiment's configurations.

**Dropped Features**

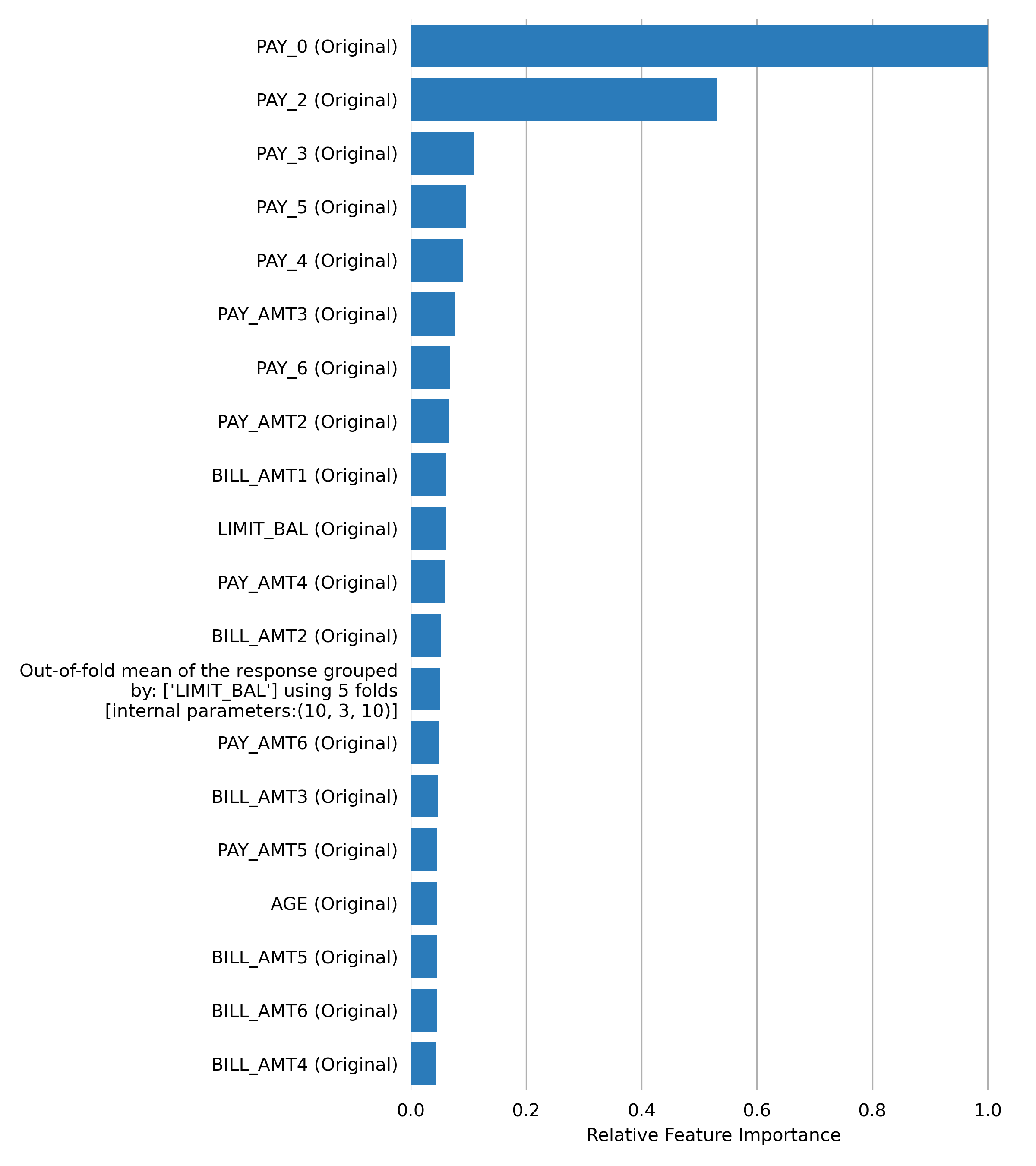
Below is the complete list of dropped features due to numerous reasons.

|  |  |
| --- | --- |
| **Name** | **Reason** |
| ID | DAI ID Columns |
| MARRIAGE | DAI Model Dropped |
| PAY\_AMT1 | DAI Model Dropped |
| SEX | DAI Model Dropped |

## Feature Transformations

The result of the Feature Evolution Stage is a set of features to use for the final model. Some of these features were automatically created by Driverless AI. The top features used in the final model are shown below, ordered by importance. The features in the table are limited to the top 50, restricted to those with relative importance greater than or equal to 0.003. If no transformer was applied, the feature is an original column.

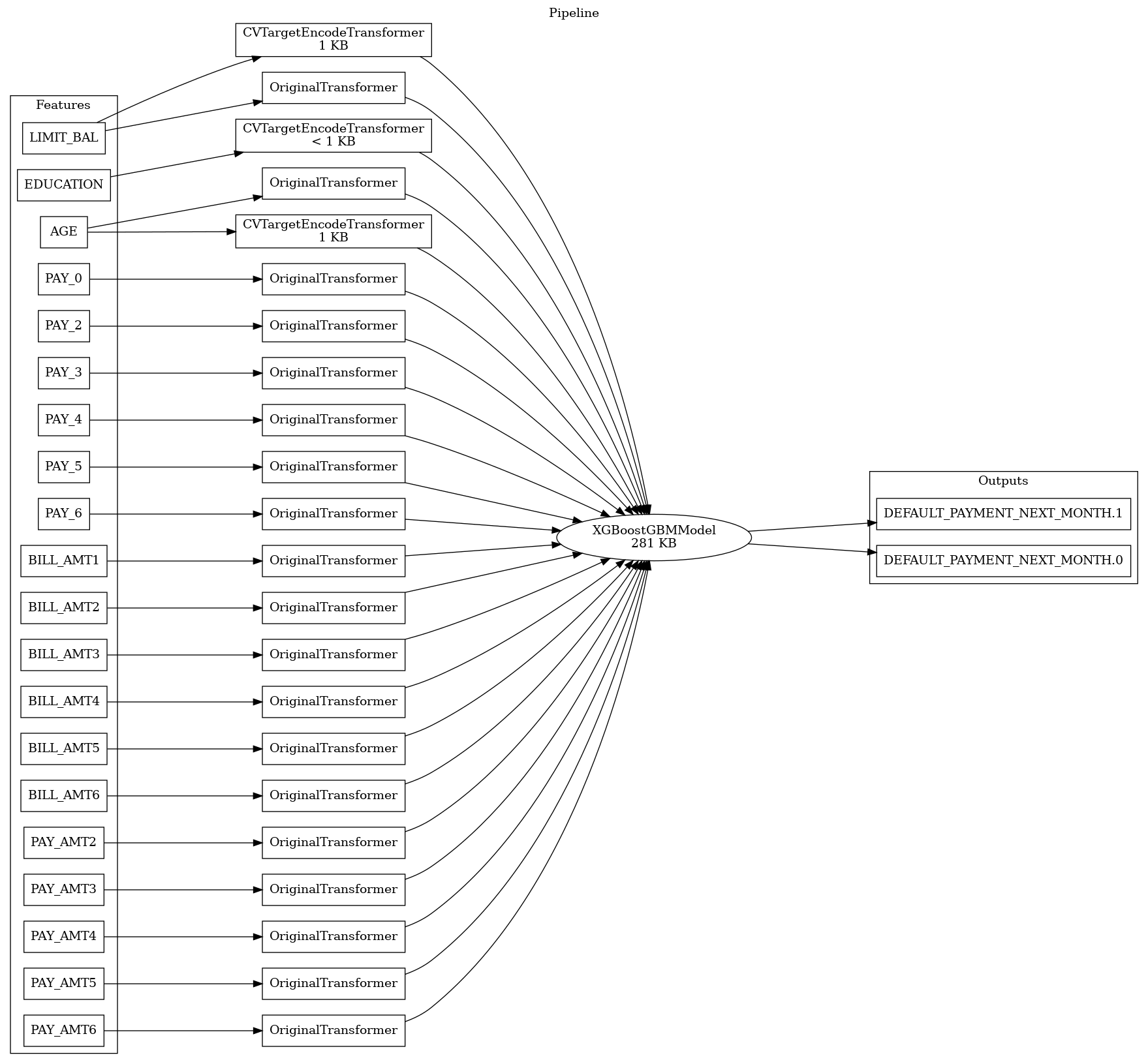
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Feature** | **Description** | **Transformer** | **Relative Importance** |
| 1 | 8\_PAY\_0 | PAY\_0 (Original) | None | 1.0 |
| 2 | 9\_PAY\_2 | PAY\_2 (Original) | None | 0.531 |
| 3 | 10\_PAY\_3 | PAY\_3 (Original) | None | 0.1108 |
| 4 | 12\_PAY\_5 | PAY\_5 (Original) | None | 0.0959 |
| 5 | 11\_PAY\_4 | PAY\_4 (Original) | None | 0.0908 |
| 6 | 16\_PAY\_AMT3 | PAY\_AMT3 (Original) | None | 0.0777 |
| 7 | 13\_PAY\_6 | PAY\_6 (Original) | None | 0.0678 |
| 8 | 15\_PAY\_AMT2 | PAY\_AMT2 (Original) | None | 0.0667 |
| 9 | 1\_BILL\_AMT1 | BILL\_AMT1 (Original) | None | 0.0611 |
| 10 | 7\_LIMIT\_BAL | LIMIT\_BAL (Original) | None | 0.0608 |
| 11 | 17\_PAY\_AMT4 | PAY\_AMT4 (Original) | None | 0.0589 |
| 12 | 2\_BILL\_AMT2 | BILL\_AMT2 (Original) | None | 0.0525 |
| 13 | 22\_CVTE: LIMIT\_BAL.0 | Out-of-fold mean of the response grouped by: ['LIMIT\_BAL'] using 5 folds [internal parameters:(10, 3, 10)] | Cross Validation Target Encoding | 0.0512 |
| 14 | 19\_PAY\_AMT6 | PAY\_AMT6 (Original) | None | 0.0483 |
| 15 | 3\_BILL\_AMT3 | BILL\_AMT3 (Original) | None | 0.0476 |
| 16 | 18\_PAY\_AMT5 | PAY\_AMT5 (Original) | None | 0.0457 |
| 17 | 0\_AGE | AGE (Original) | None | 0.0456 |
| 18 | 5\_BILL\_AMT5 | BILL\_AMT5 (Original) | None | 0.0454 |
| 19 | 6\_BILL\_AMT6 | BILL\_AMT6 (Original) | None | 0.0451 |
| 20 | 4\_BILL\_AMT4 | BILL\_AMT4 (Original) | None | 0.0446 |
| 21 | 20\_CVTE: AGE.0 | Out-of-fold mean of the response grouped by: ['AGE'] using 5 folds [internal parameters:(10, 3, 10)] | Cross Validation Target Encoding | 0.0441 |
| 22 | 21\_CVTE: EDUCATION.0 | Out-of-fold mean of the response grouped by: ['EDUCATION'] using 5 folds [internal parameters:(10, 3, 10)] | Cross Validation Target Encoding | 0.0432 |



## Final Model

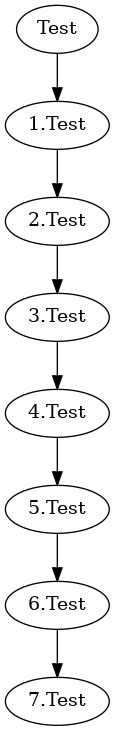
**Pipeline**

Final XGBoostGBMModel pipeline with ensemble\_level=0 transforming 20 original features -> 22 features in each of 1 models each of 3 fold hyperparameters averaged and re-fit as single model.:



**Model Lineage**

The following plot shows the experiment lineage for the current experiment 7.Test.



**Details**

* The fitted features of the final model are the best features found during the feature engineering iterations.
* The target transformer indicates the type of transformation applied to the target column.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Index** | **Type** | **Model Weight** | **Num Folds** | **Fitted features** | **Target Transformer** |
| 0 | XGBoostGBMModel | 1 | 3 | 22 | LabelEncoder |

* Model Index: 0 has a weight of 1 in the final ensemble

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Type** | **learning rate** | **subsample** | **tree method** | **max leaves** | **index** | **max depth** | **model class name** | **grow policy** | **colsample bytree** |
| XGBoostGBMModel | 0.05 | 0.7 | gpu\_hist | 64 | 0 | 6 | XGBoostGBMModel | depthwise | 0.8 |

For a complete list of the parameters of the final model, see the Appendix.

**Performance of Final Model**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Scorer** | **Optimized** | **Better score is** | **Final ensemble scores on validation (internal or external holdout(s)) data** | **Final ensemble standard deviation on validation (internal or external holdout(s)) data** | **Final test scores** | **Final test standard deviation** |
| AUC | \* | higher | 0.7775288 | 0.00328363 | 0.7964425 | 0.0071086 |
| ACCURACY |  | higher | 0.8179507 | 0.002053431 | 0.8335 | 0.004605334 |
| AUCPR |  | higher | 0.5450957 | 0.00503948 | 0.5769592 | 0.01520024 |
| F05 |  | higher | 0.5731741 | 0.005014345 | 0.5940356 | 0.01169257 |
| F1 |  | higher | 0.5406526 | 0.004655751 | 0.5546157 | 0.01105089 |
| F2 |  | higher | 0.6394723 | 0.00388114 | 0.6369657 | 0.009642211 |
| FDR |  | lower | 0.4761572 | 0.02418475 | 0.4810736 | 0.02418475 |
| FNR |  | lower | 0.441423 | 0.03047949 | 0.4044234 | 0.03047949 |
| FOR |  | lower | 0.1296925 | 0.004978913 | 0.1126017 | 0.005056502 |
| FPR |  | lower | 0.1463231 | 0.02305061 | 0.1476553 | 0.02305061 |
| GINI |  | higher | 0.5550575 | 0.00656726 | 0.5928849 | 0.01561697 |
| LOGLOSS |  | lower | 0.4344879 | 0.002722395 | 0.4072474 | 0.006150203 |
| MACROAUC |  | higher | 0.7775288 | 0.00328363 | 0.7964425 | 0.006431311 |
| MACROF1 |  | higher | 0.5406526 | 0.004655751 | 0.5546157 | 0.01166793 |
| MACROMCC |  | higher | 0.4164366 | 0.005580545 | 0.4374818 | 0.01080803 |
| MCC |  | higher | 0.4164366 | 0.005580545 | 0.4374818 | 0.01395377 |
| NPV |  | higher | 0.8703075 | 0.004978913 | 0.8873983 | 0.005208308 |
| PRECISION |  | higher | 0.5238428 | 0.02418475 | 0.5189264 | 0.03265311 |
| RECALL |  | higher | 0.558577 | 0.03047949 | 0.5955766 | 0.03047949 |
| TNR |  | higher | 0.8536769 | 0.02305061 | 0.8523447 | 0.02305061 |

**Validation Confusion Matrix**

*Threshold 0.2779437006*

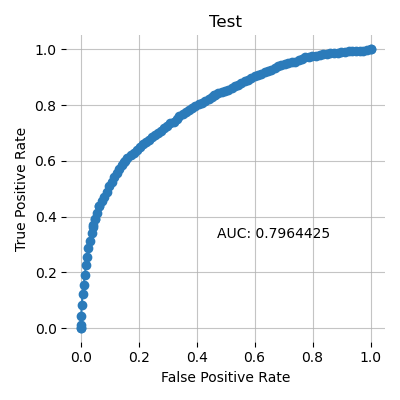
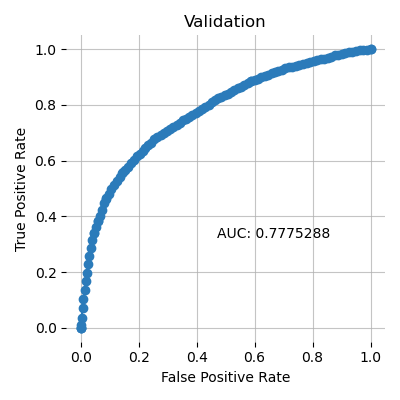
|  |  |  |  |
| --- | --- | --- | --- |
|  | **Predicted: 0** | **Predicted: 1** | **error** |
| Actual: 0 | 15,904 | 2,726 | 15% |
| Actual: 1 | 2,370 | 2,999 | 44% |

**Test Confusion Matrix**

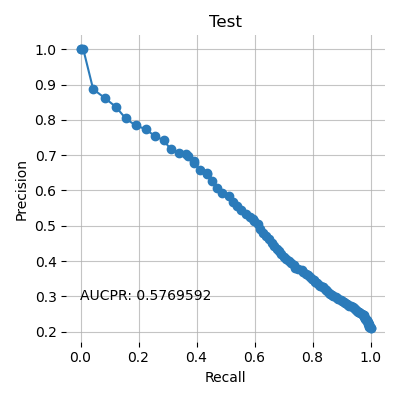
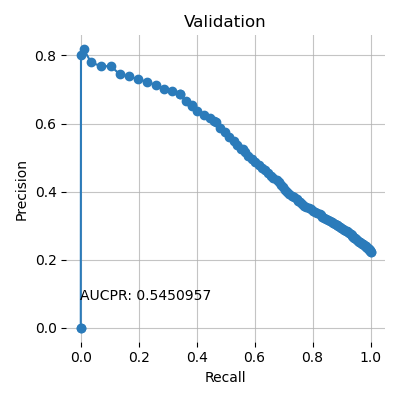
*Threshold 0.2779437006*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Predicted: 0** | **Predicted: 1** | **error** |
| Actual: 0 | 4,158 | 576 | 12% |
| Actual: 1 | 570 | 696 | 45% |

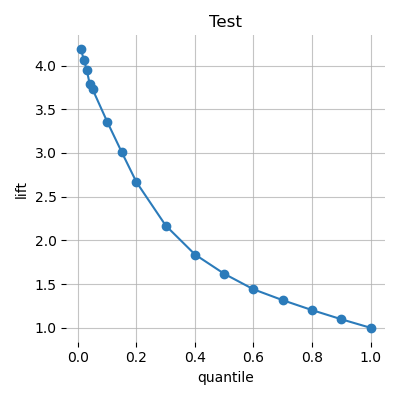
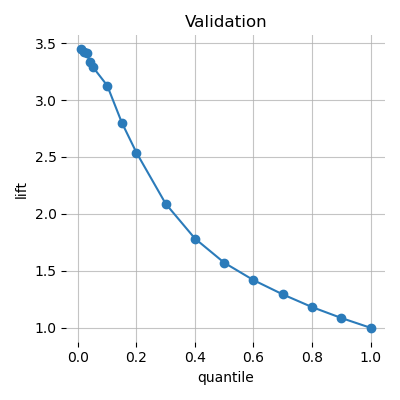
*Receiver Operating Characteristic Curve*



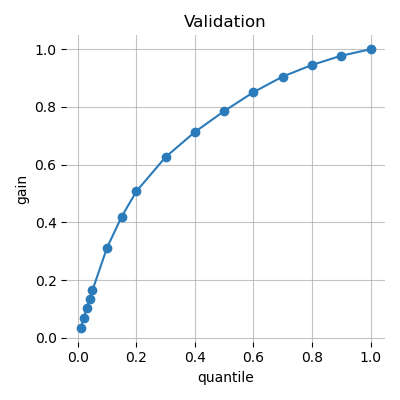
*Precision Recall Curve*



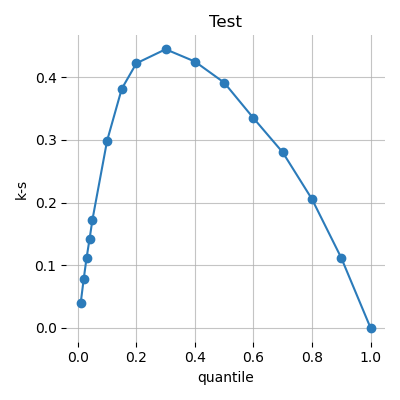
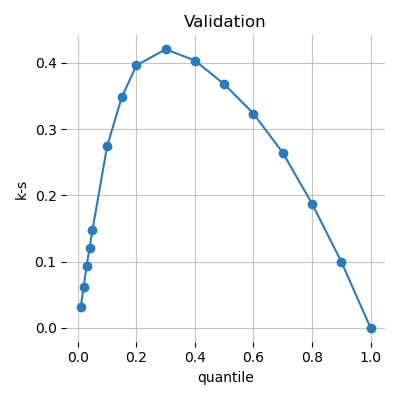
*Cumulative Lift*



*Cumulative Gains*



*Kolmogorov–Smirnov*



## Alternative Models

During the experiment, Driverless AI trained 5 alternative models. The following algorithms were evaluated during the Driverless AI experiment:

|  |  |  |  |
| --- | --- | --- | --- |
| **algorithm** | **package** | **version** | **documentation** |
| lightgbm | lightgbm | 3.2.1.99 | LightGBM, Light Gradient Boosting Machine. Contributors: https://github.com/microsoft/LightGBM/graphs/contributors. |
| gbtree | xgboost | 1.5.0-dev | XGBoost: eXtreme Gradient Boosting library. Contributors: https://github.com/dmlc/xgboost/blob/master/CONTRIBUTORS.md |
| constant | custom package | 1.10.1+local\_dev-placeholder-2240-g8bd2dc928a-dirty | reference model that predicts a constant aimed at minimizing the given scorer |

Driverless AI can evaluate an array of algorithms, including but not limited to XGBoost GBM, XGBoost Dart, XGBoost GLM, LightGBM, RuleFit, Tensorflow, and FTRL models. The table below explains why certain algorithms were not selected for the final model, if any.

|  |  |
| --- | --- |
| **algorithm** | **selection** |
| gblinear | algorithm not evaluated due to experiment configuration |
| decision tree | algorithm not evaluated due to experiment configuration |
| rulefit | algorithm not evaluated due to experiment configuration |
| tensorflow | algorithm not evaluated due to experiment configuration |
| ftrl | algorithm not evaluated due to experiment configuration |
| dart | algorithm not evaluated due to experiment configuration |
| lightgbm | not selected due to low performance during feature evolution stage |
| gbtree | selected for final model |

## Deployment

For this experiment, both Python and MOJO Scoring Pipelines are available for productionizing the final model pipeline for a given row of data or table of data.

### Python Scoring Pipeline

This package contains an exported model and Python 3.6 source code examples for productionizing models built using H2O Driverless AI. The Python Scoring Pipeline is located here:

* **user/h2oai\_experiment\_dd901dc8-3740-11ec-9ec3-ac1f6b6b49a6/scoring\_pipeline/scorer.zip**

The files in this package allow you to transform and score on new data in a couple of different ways:

* From Python 3.6, you can import a scoring module, then use the module to transform and score on new data.
* From other languages and platforms, you can use the TCP/HTTP scoring service bundled with this package to call into the scoring pipeline module through remote procedure calls (RPC).

### MOJO Scoring Pipeline

Note: The MOJO Scoring Pipeline is currently in a beta state. Updates and improvements will continue to be made in subsequent Driverless AI releases. The MOJO Scoring Pipeline is located here:

* **user/h2oai\_experiment\_dd901dc8-3740-11ec-9ec3-ac1f6b6b49a6/mojo\_pipeline/mojo.zip**

For completed experiments, Driverless AI converts models to MOJOs (Model Objects, Optimized). A MOJO is a scoring engine that can be deployed in any Java environment for scoring in real time.

## Partial Dependence Plots

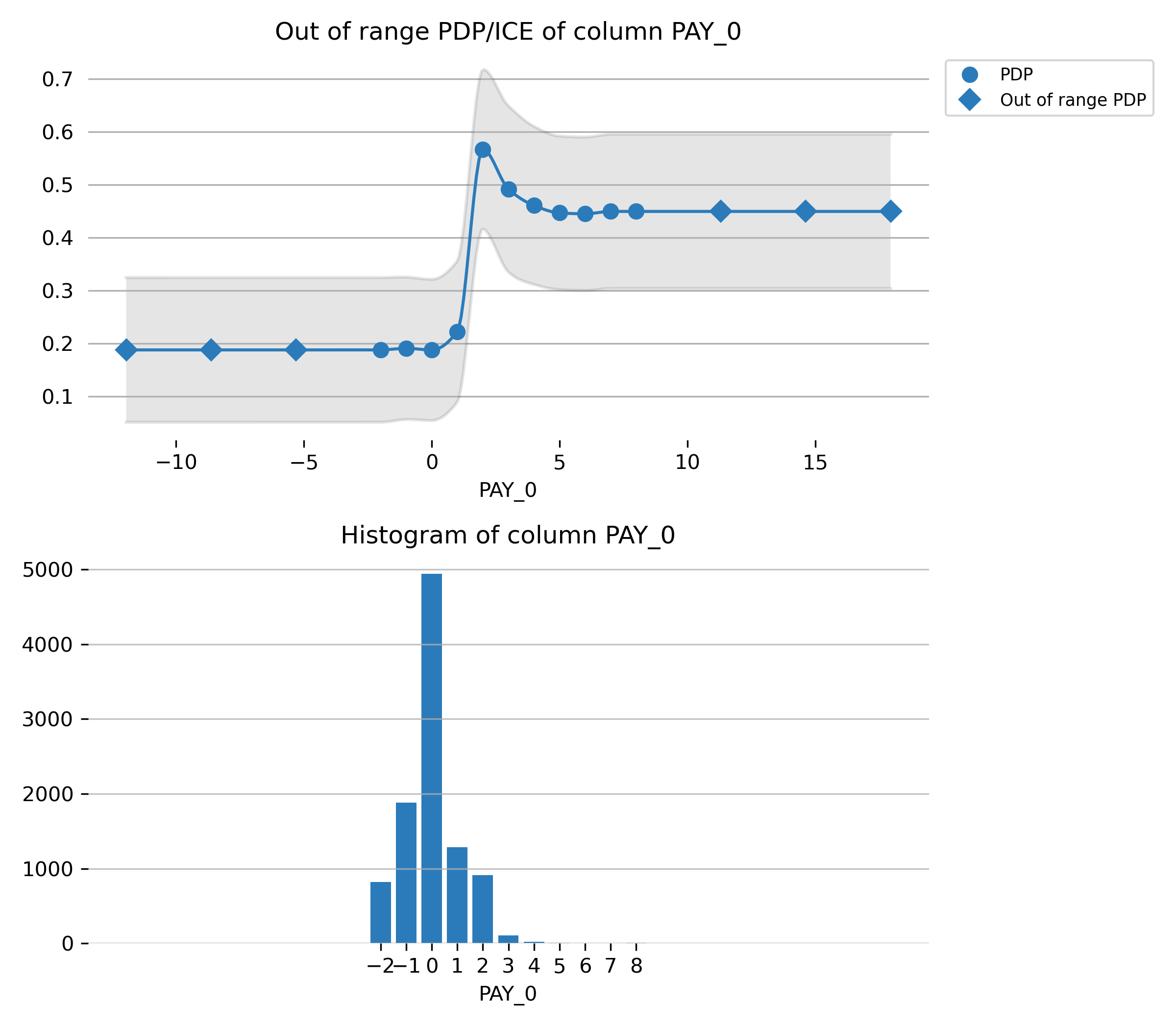
Partial dependence plots show the partial dependence as a function of specific values for a feature subset. The plots show how machine-learned response functions change based on the values of an input feature of interest, while taking nonlinearity into consideration and averaging out the effects of all other input features. Partial dependence plots enable increased transparency in a model and enable the ability to validate and debug a model by comparing a feature's average predictions across its domain to known standards and reasonable expectations.

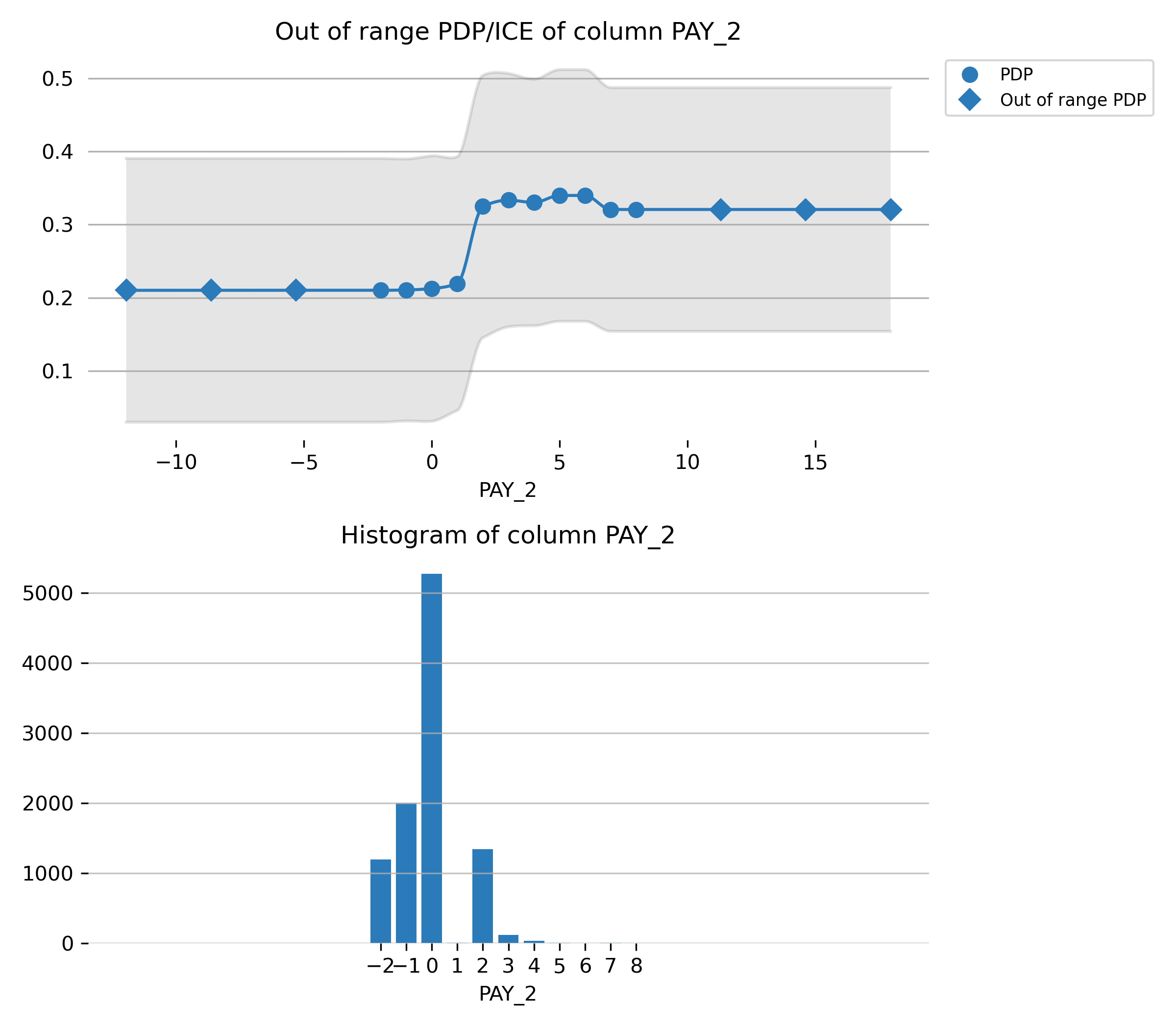
The partial dependence plots are shown for the top 20 original variables. The top 20 original variables are chosen based on their Component Based Variable Importance.

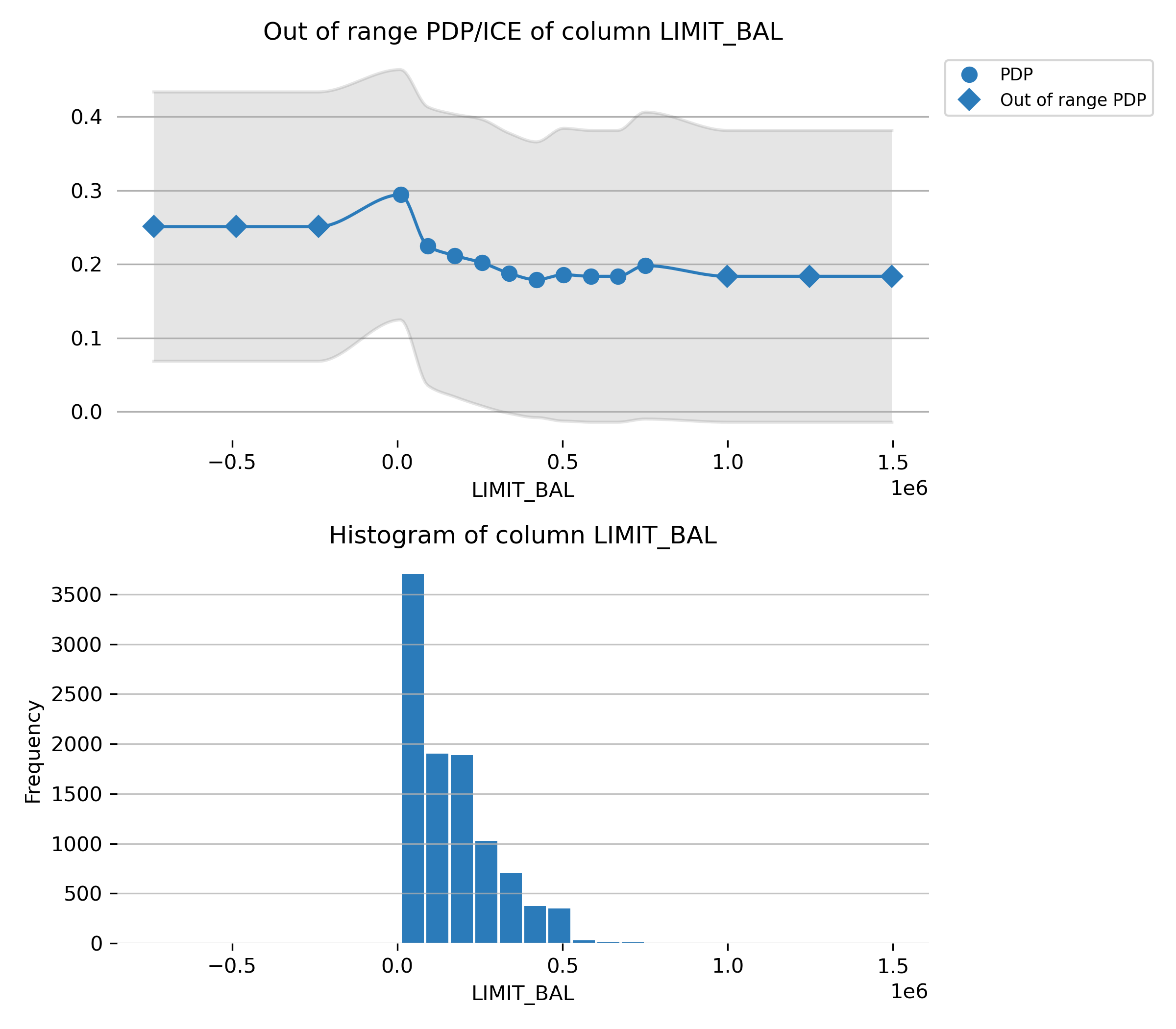
**Plot Details**

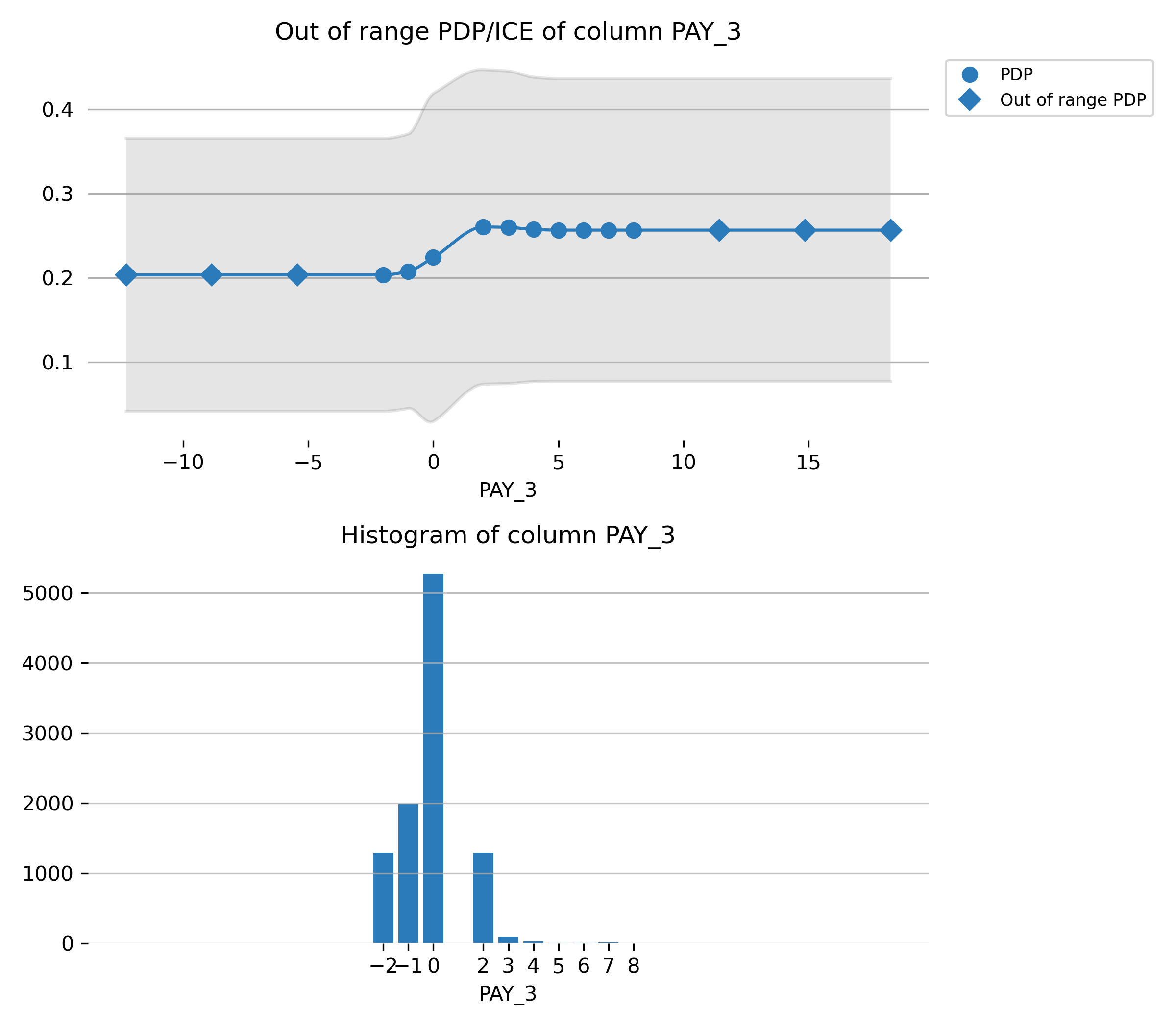
In the Driverless AI PDP, the y-axis represents the mean response, and a shaded region (for numeric features) or shaded bar (for categorical features) represents 1 standard deviation. Out-of-range PDP (diamond markers) represent values outside feature intervals seen in the data, unseen categorical values, or missing values.

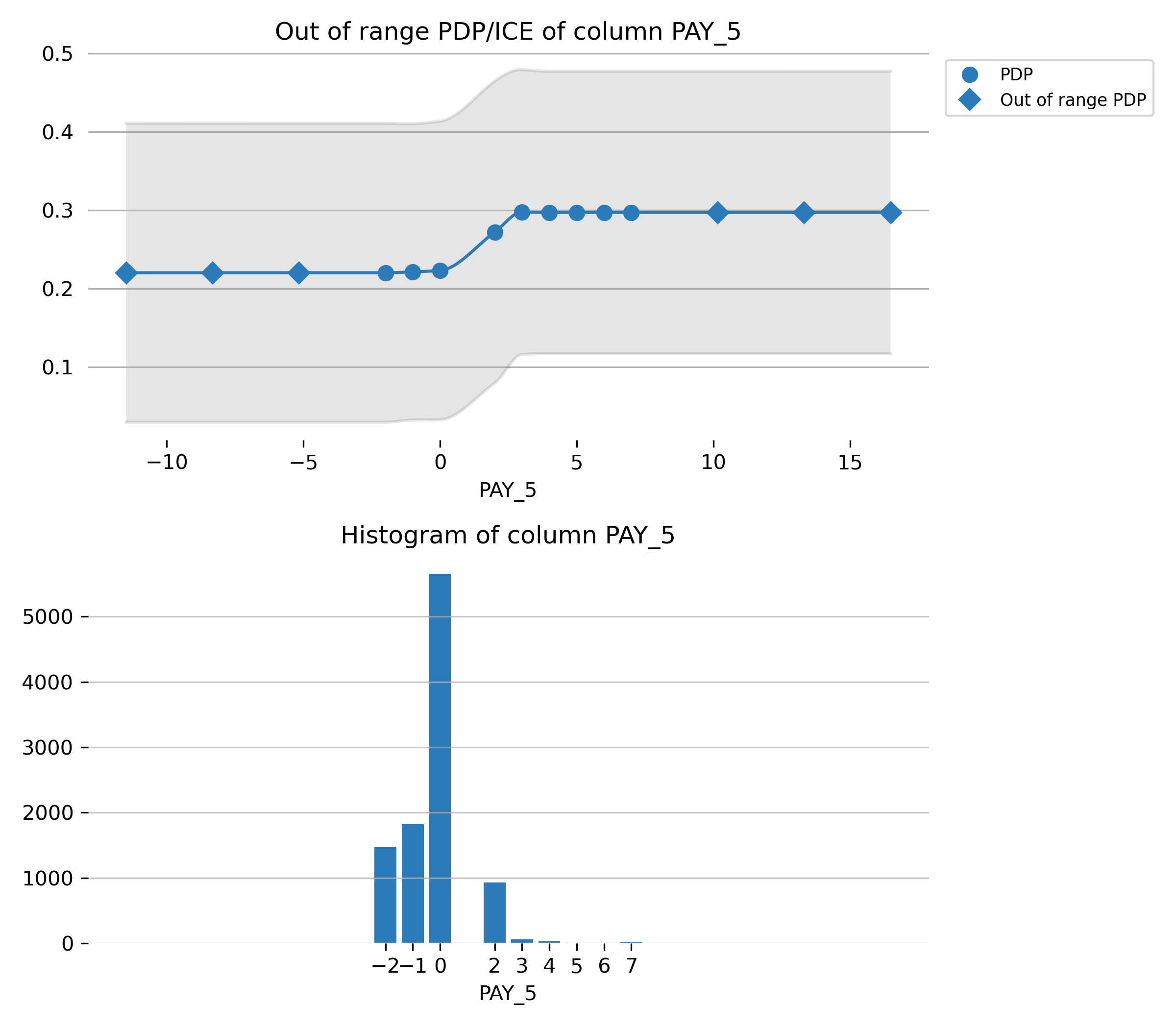
For continuous features, numeric values up to 3 standard deviations lower than the minimum training value and higher than the maximum training value are feed into the model. For categorical features, an unseen categorical value is feed into the model denoted by UNSEEN (if the categorical value "UNSEEN" already exists in the training data, the out-of-range is done on a value called "UNSEEN\_[x]," where x is some integer).

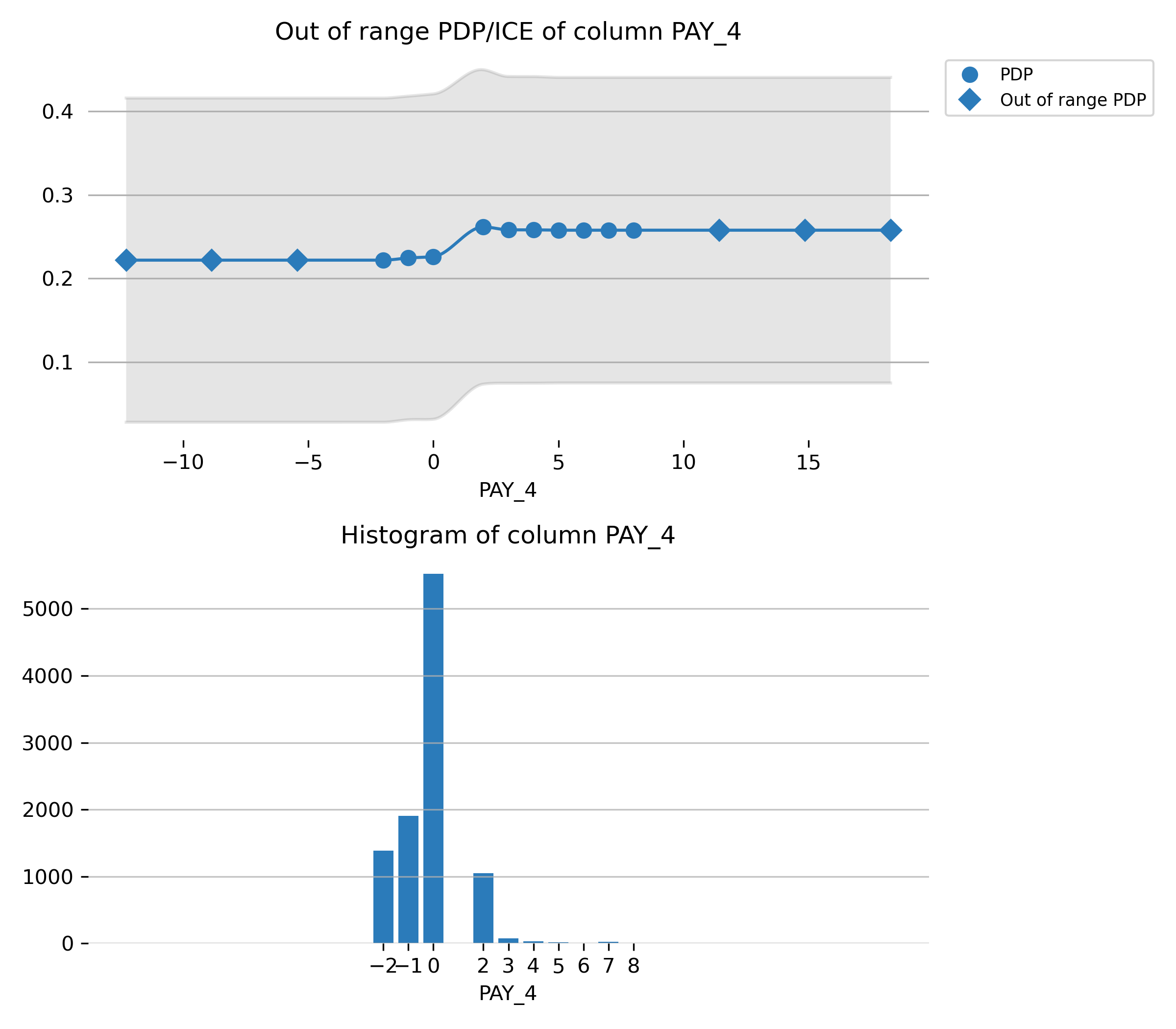
Feature **PAY\_0** 

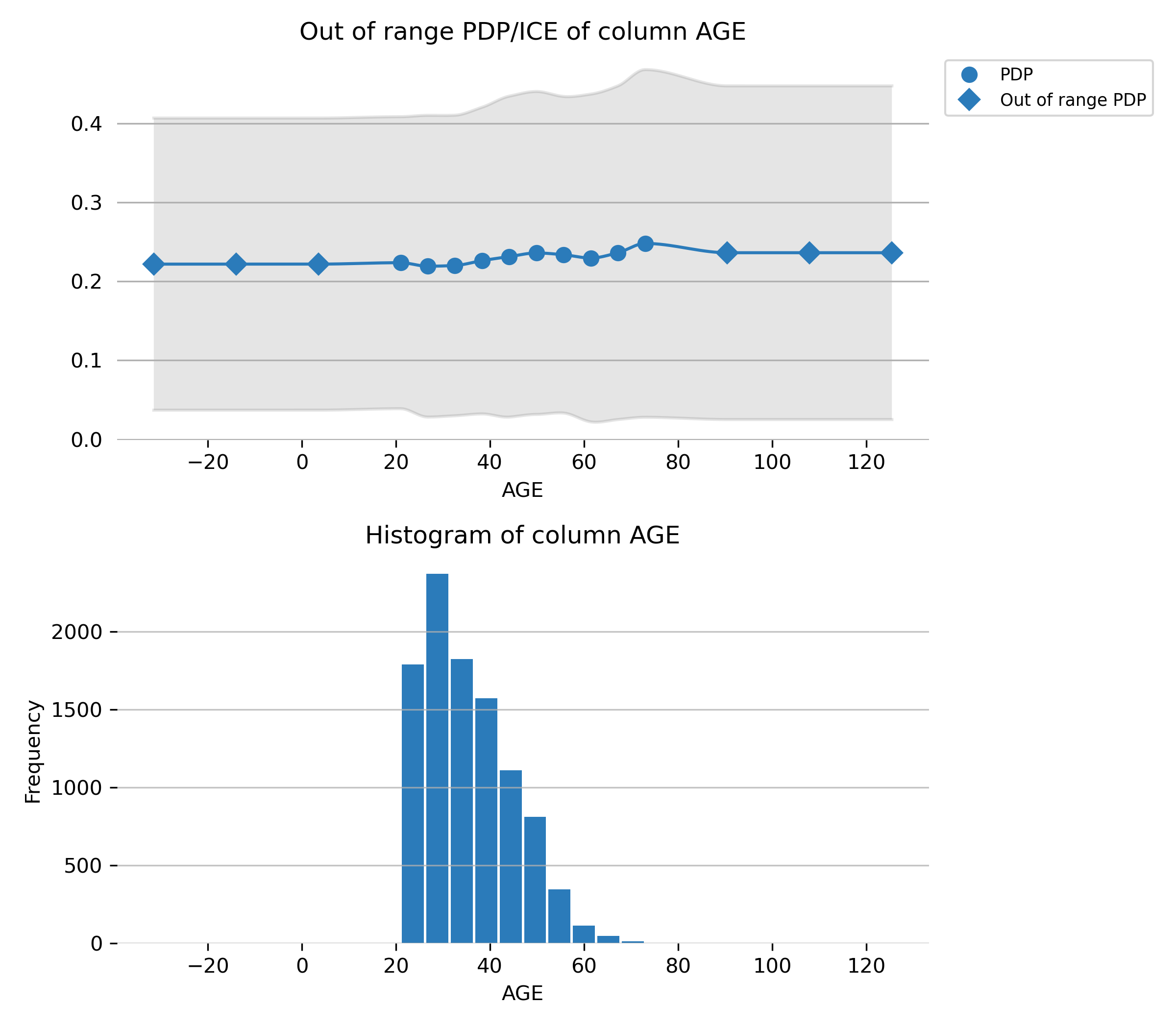
Feature **PAY\_2** 

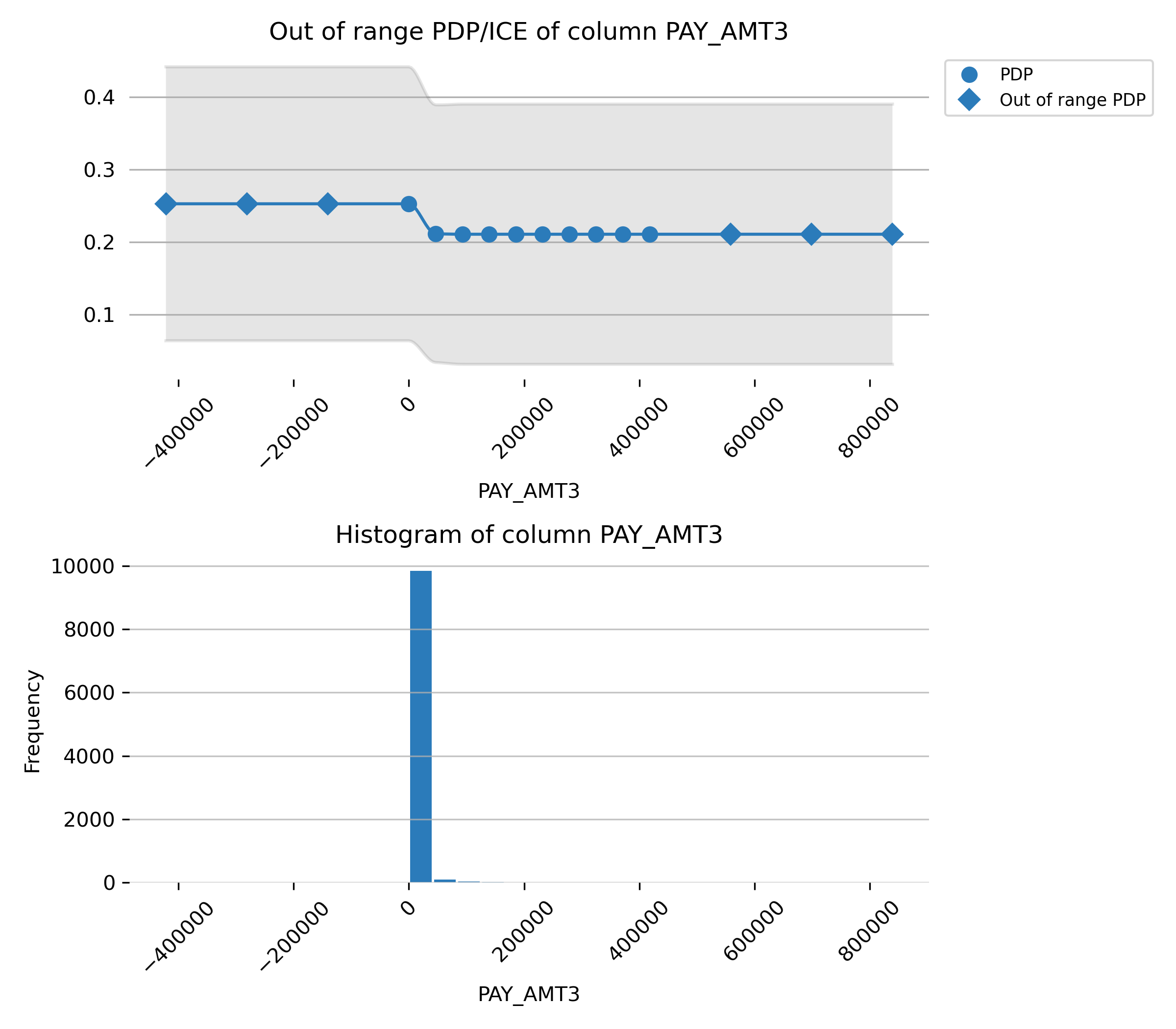
Feature **LIMIT\_BAL** 

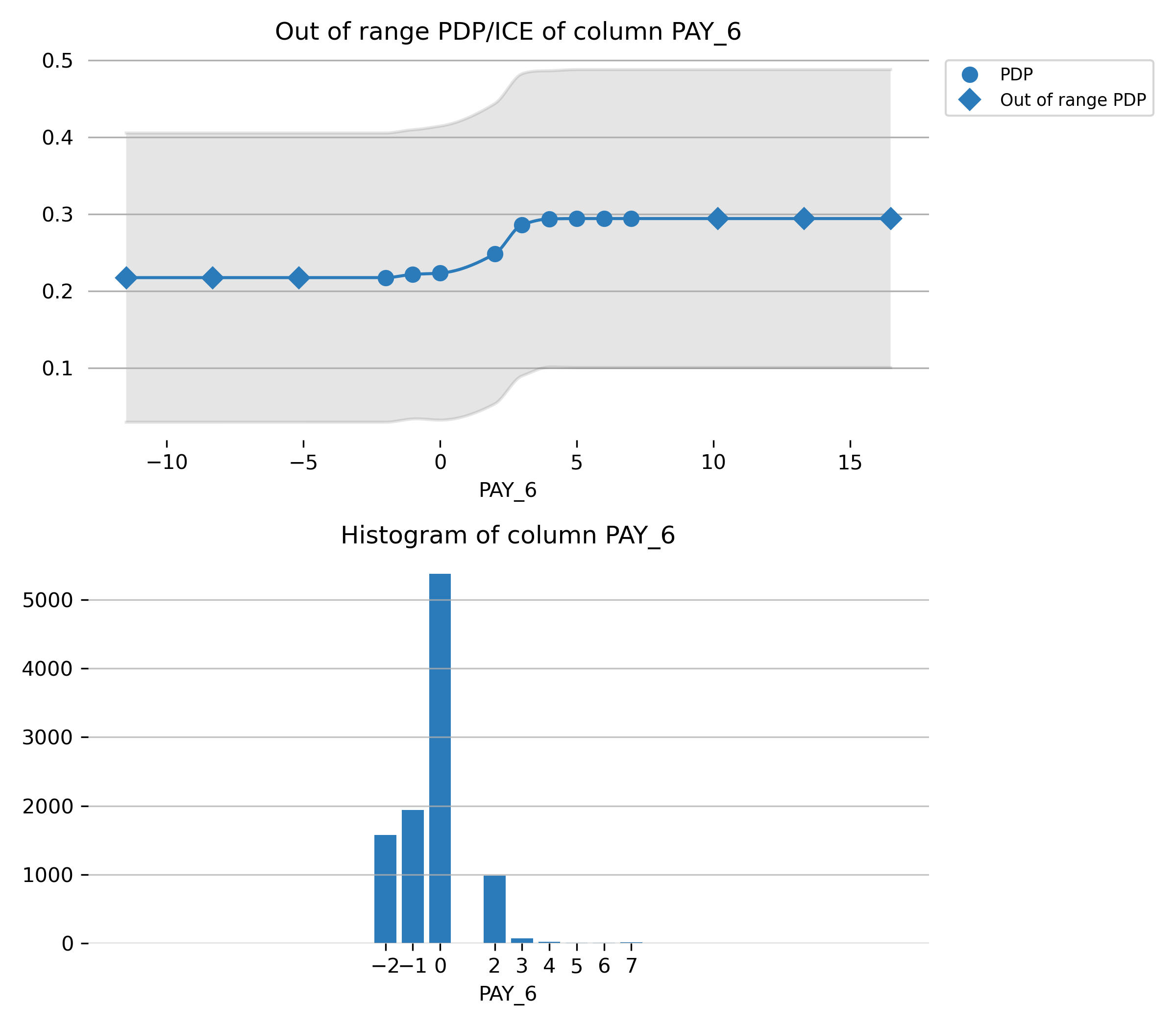
Feature **PAY\_3** 

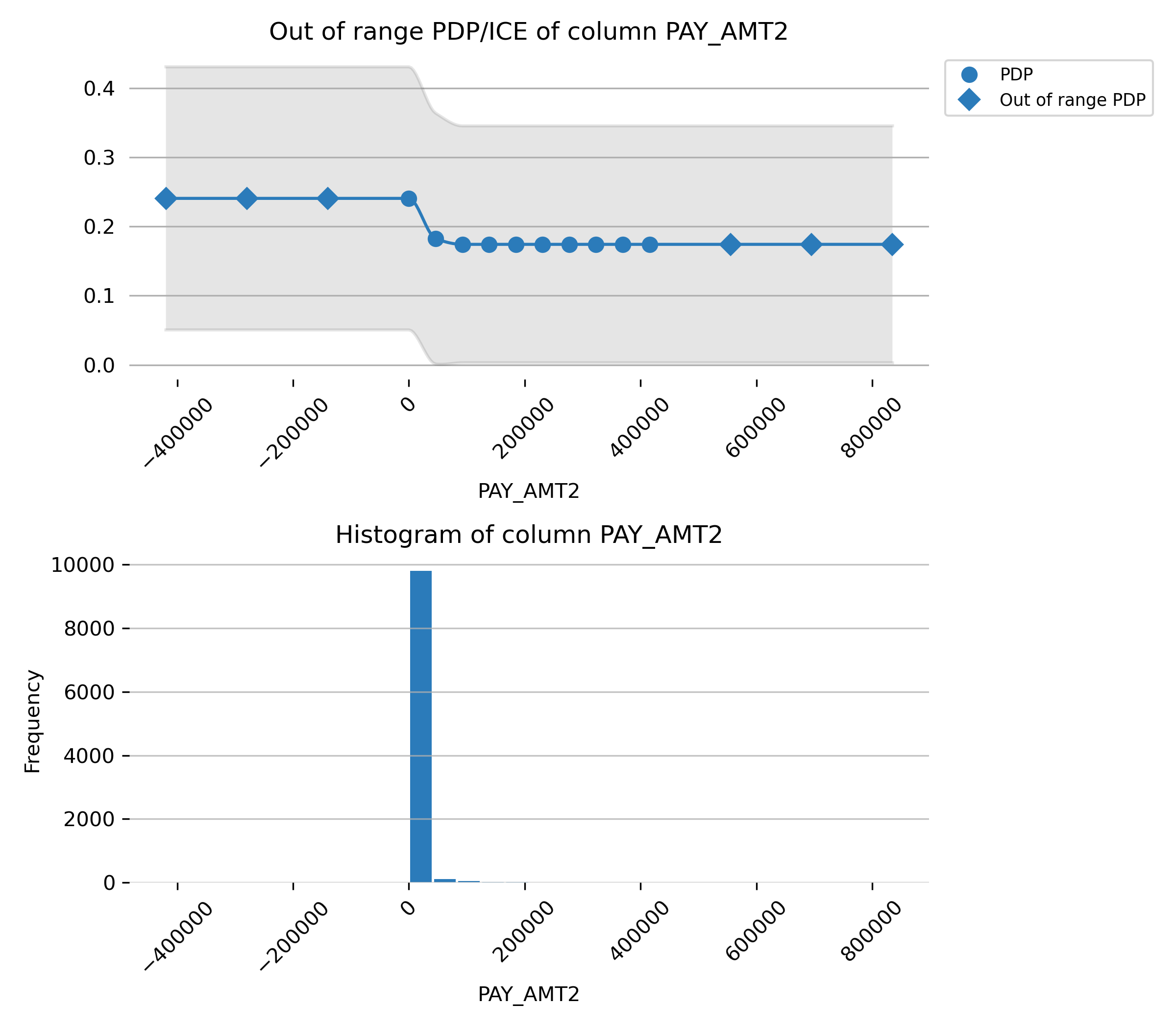
Feature **PAY\_5** 

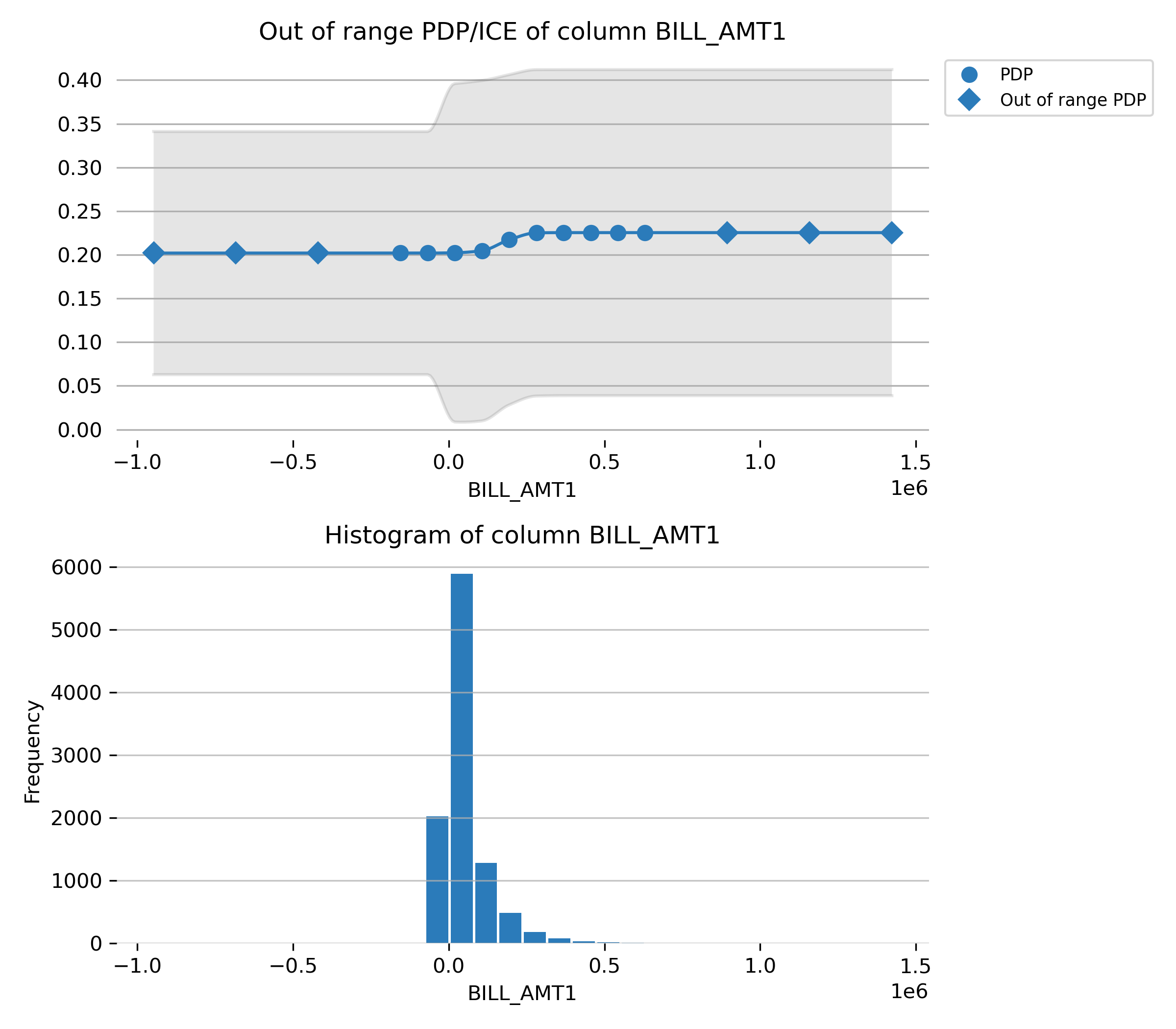
Feature **PAY\_4** 

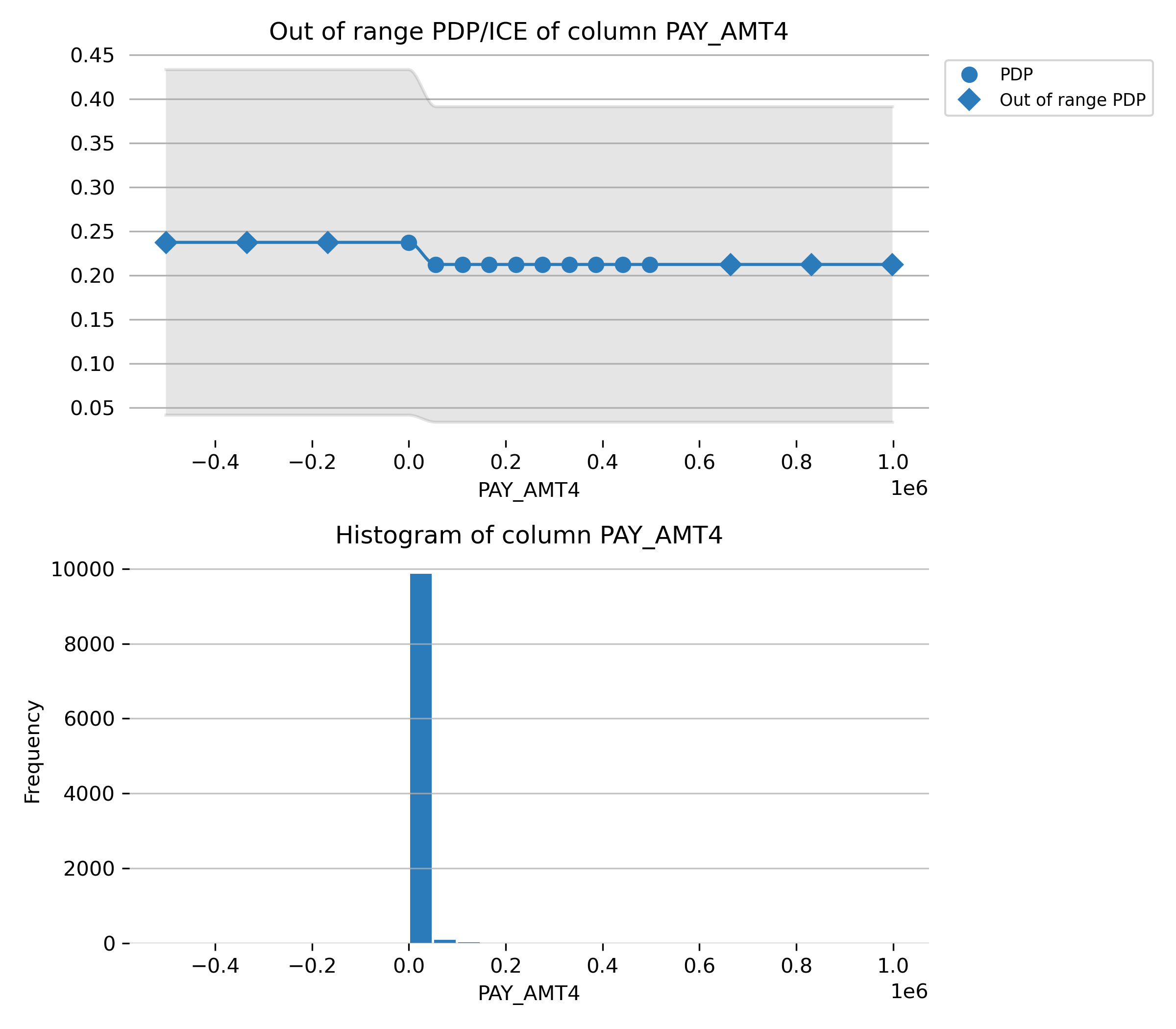
Feature **AGE** 

Feature **PAY\_AMT3** 

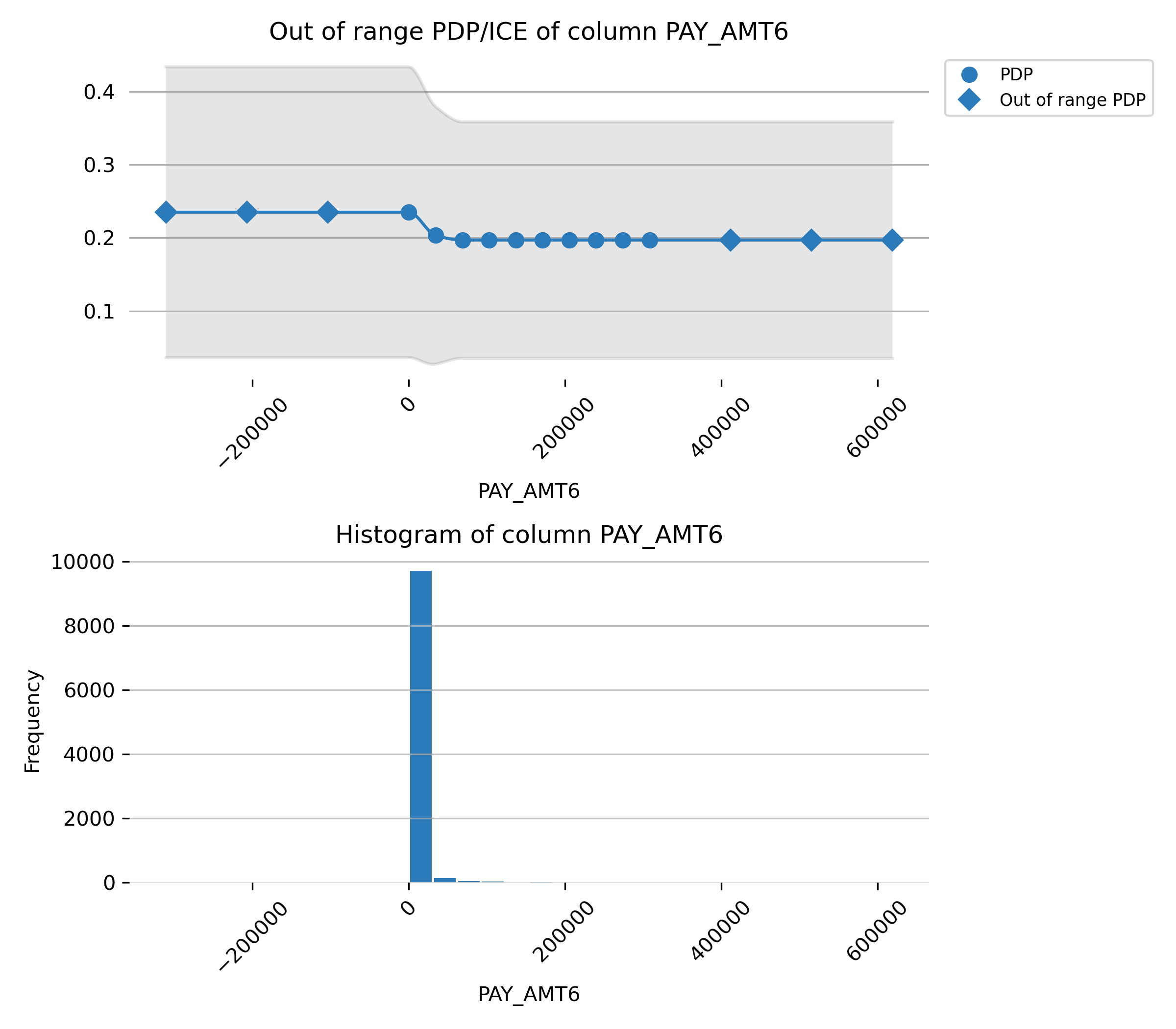
Feature **PAY\_6** 

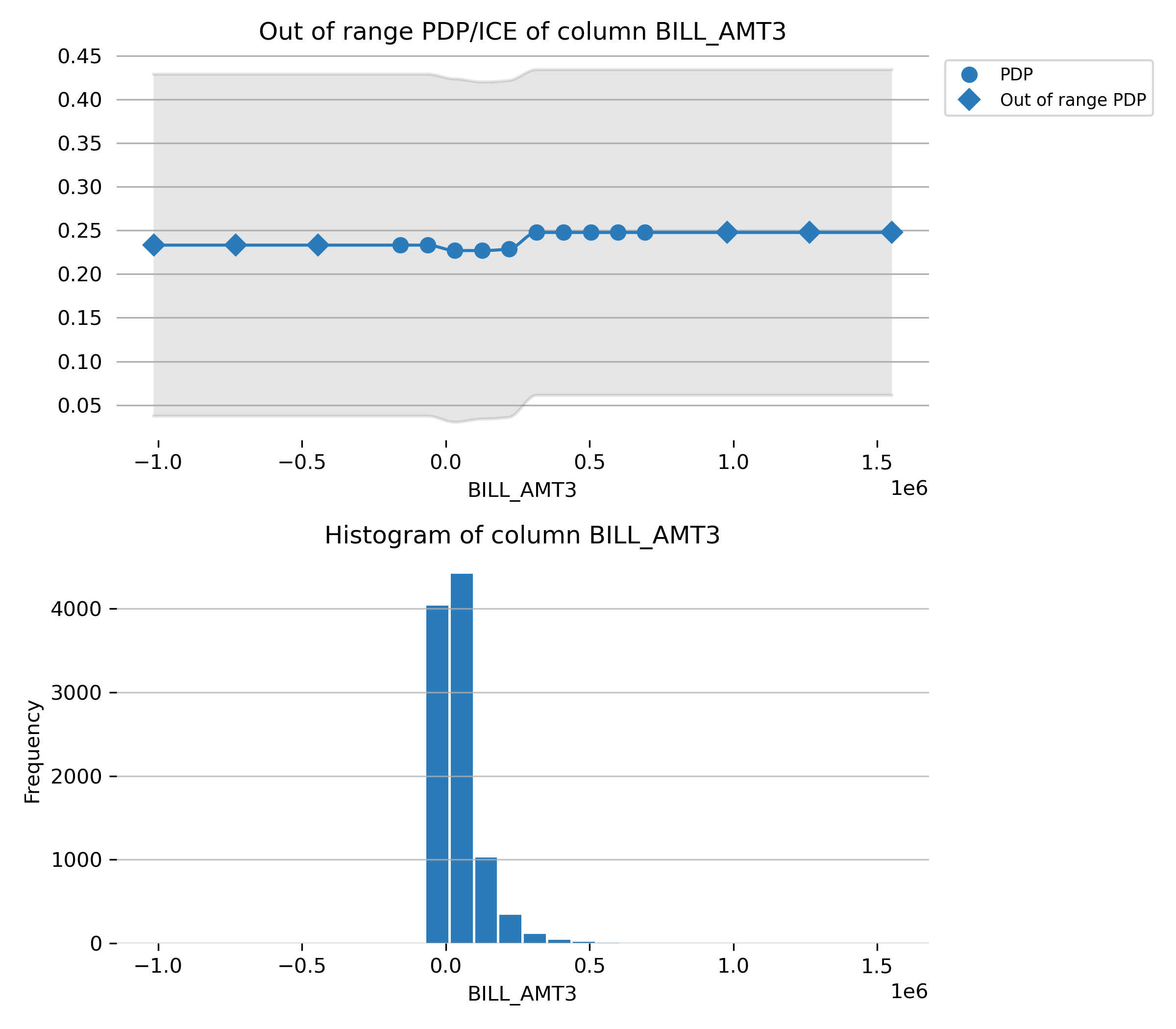
Feature **PAY\_AMT2** 

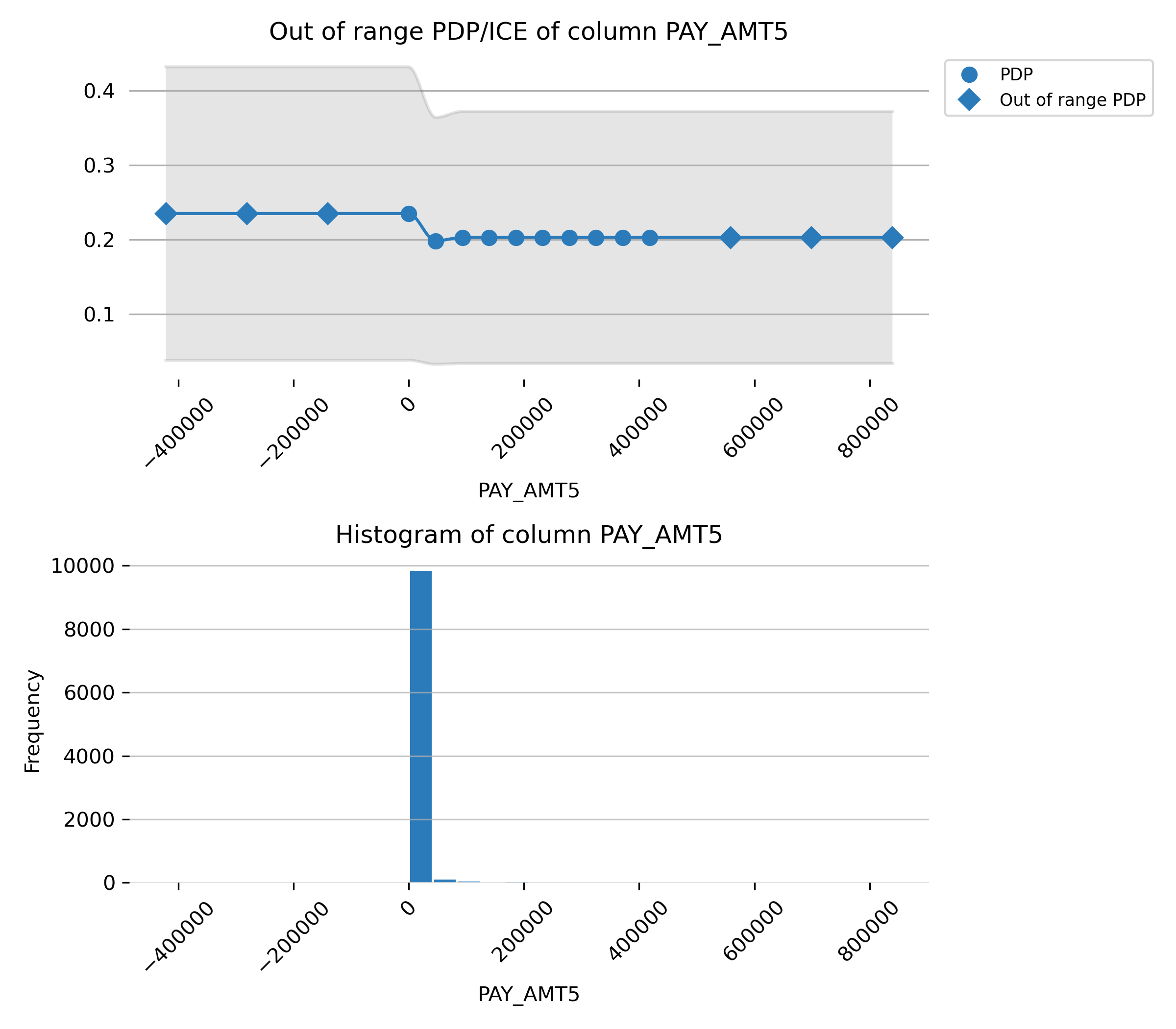
Feature **BILL\_AMT1** 

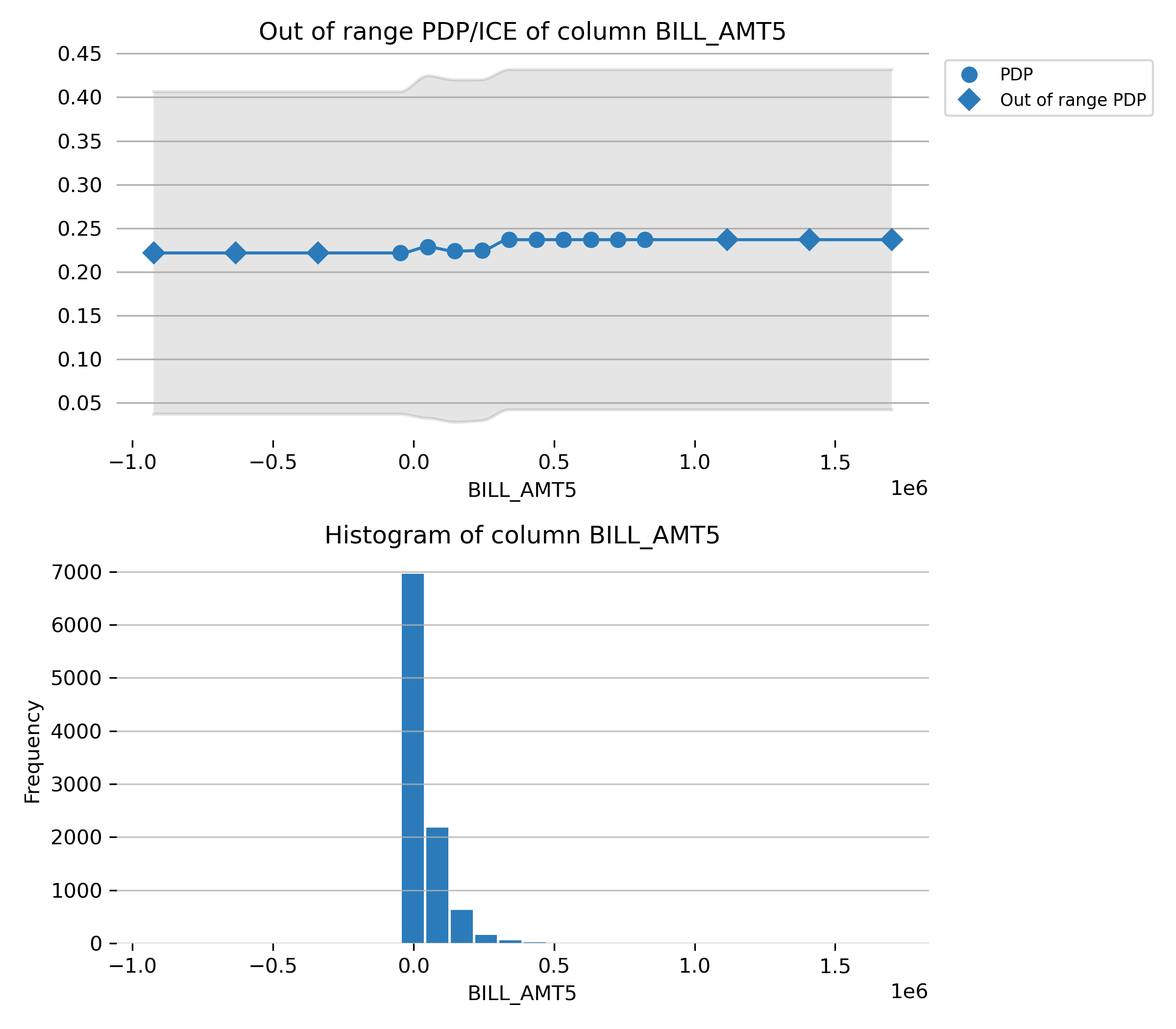
Feature **PAY\_AMT4** 

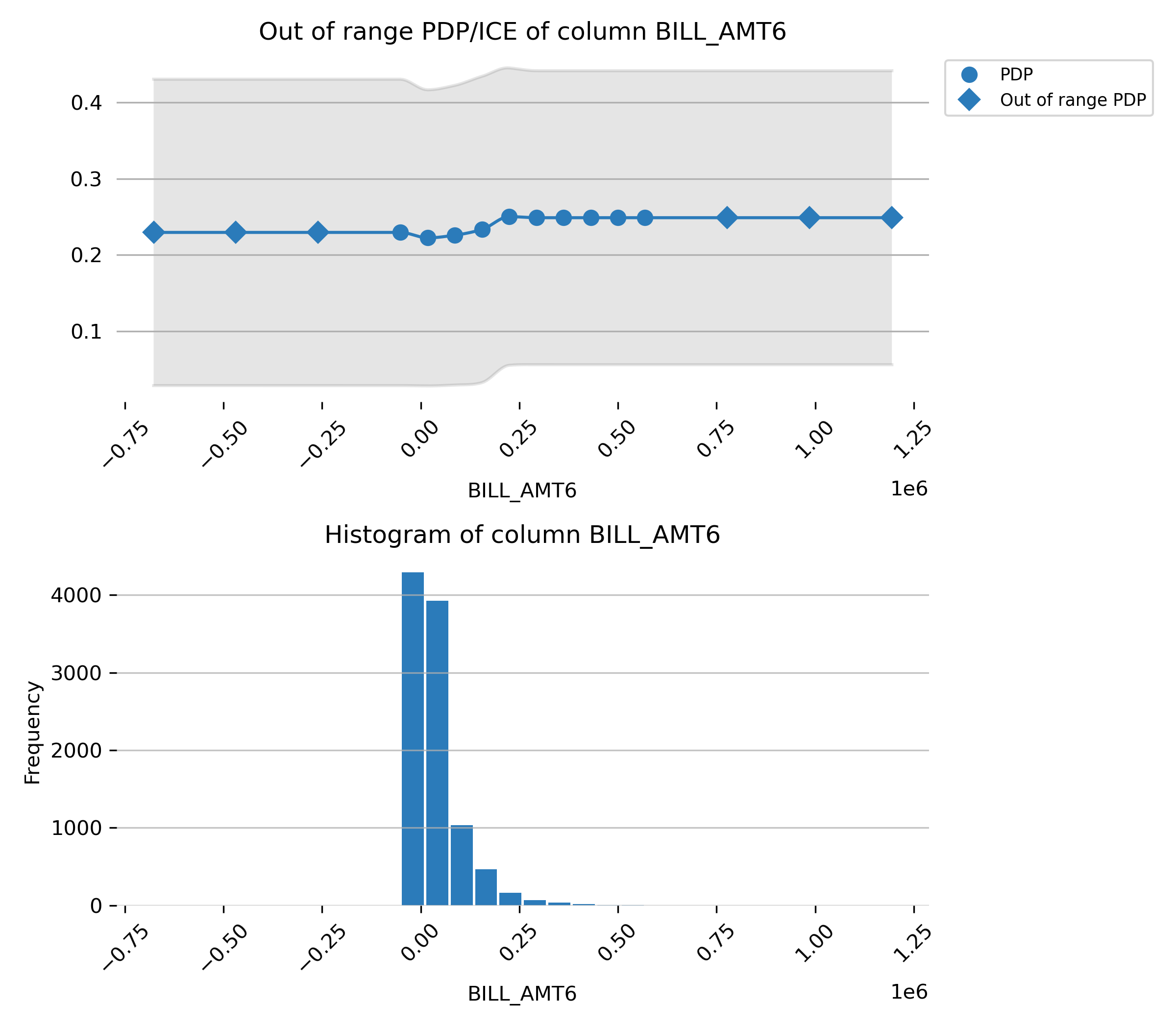
Feature **BILL\_AMT2** 

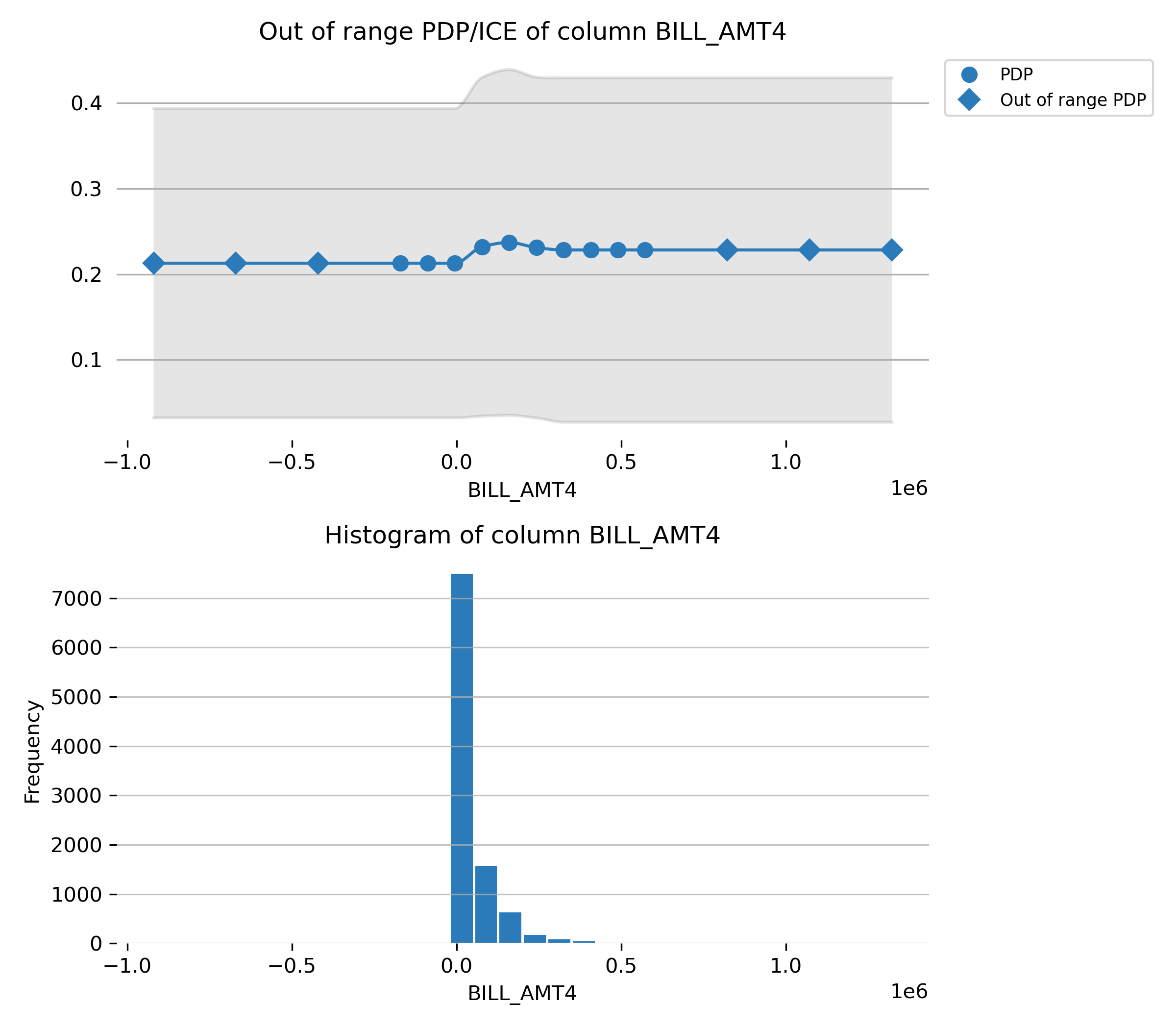
Feature **PAY\_AMT6** 

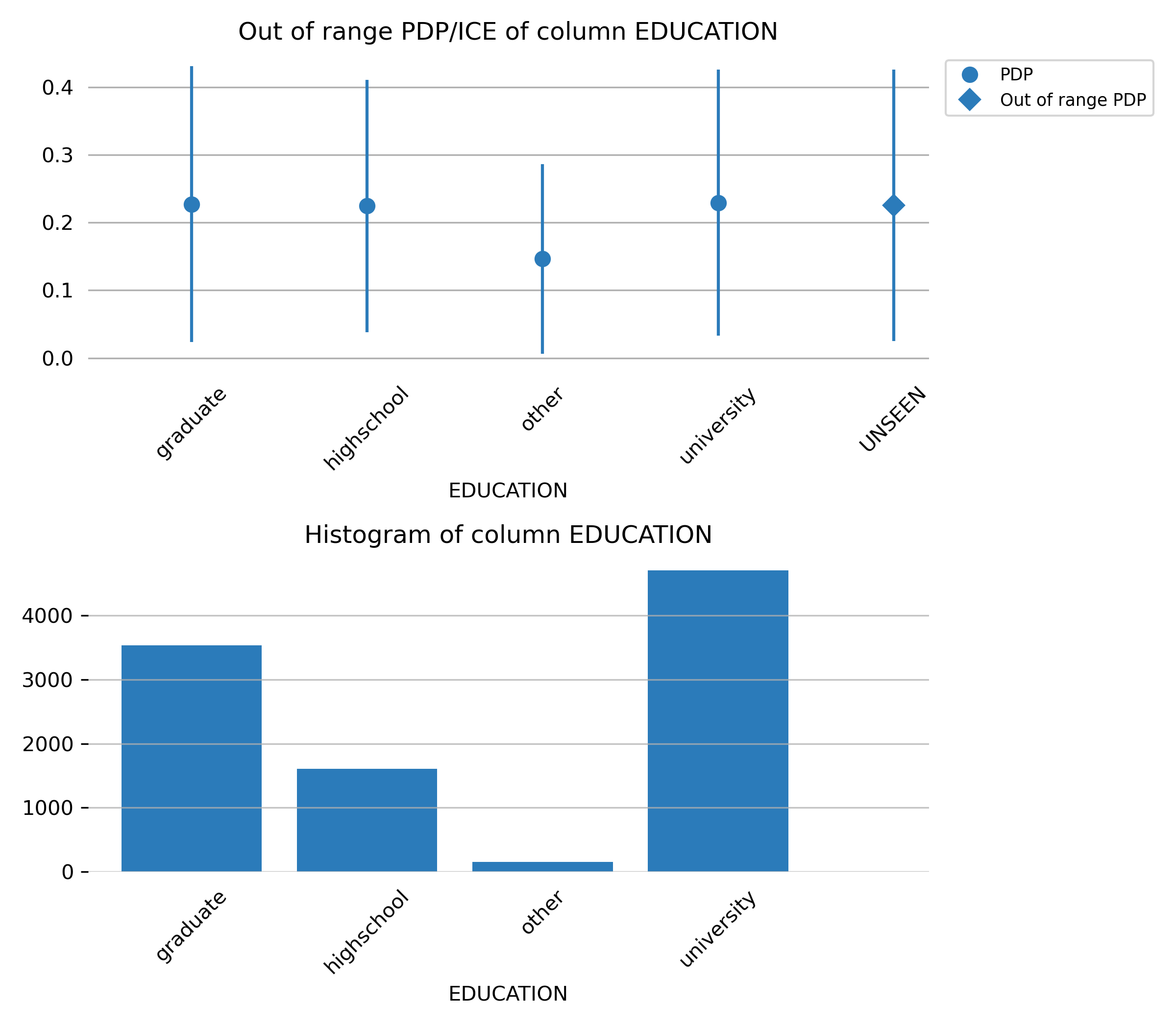
Feature **BILL\_AMT3** 

Feature **PAY\_AMT5** 

Feature **BILL\_AMT5** 

Feature **BILL\_AMT6** 

Feature **BILL\_AMT4** 

Feature **EDUCATION** 

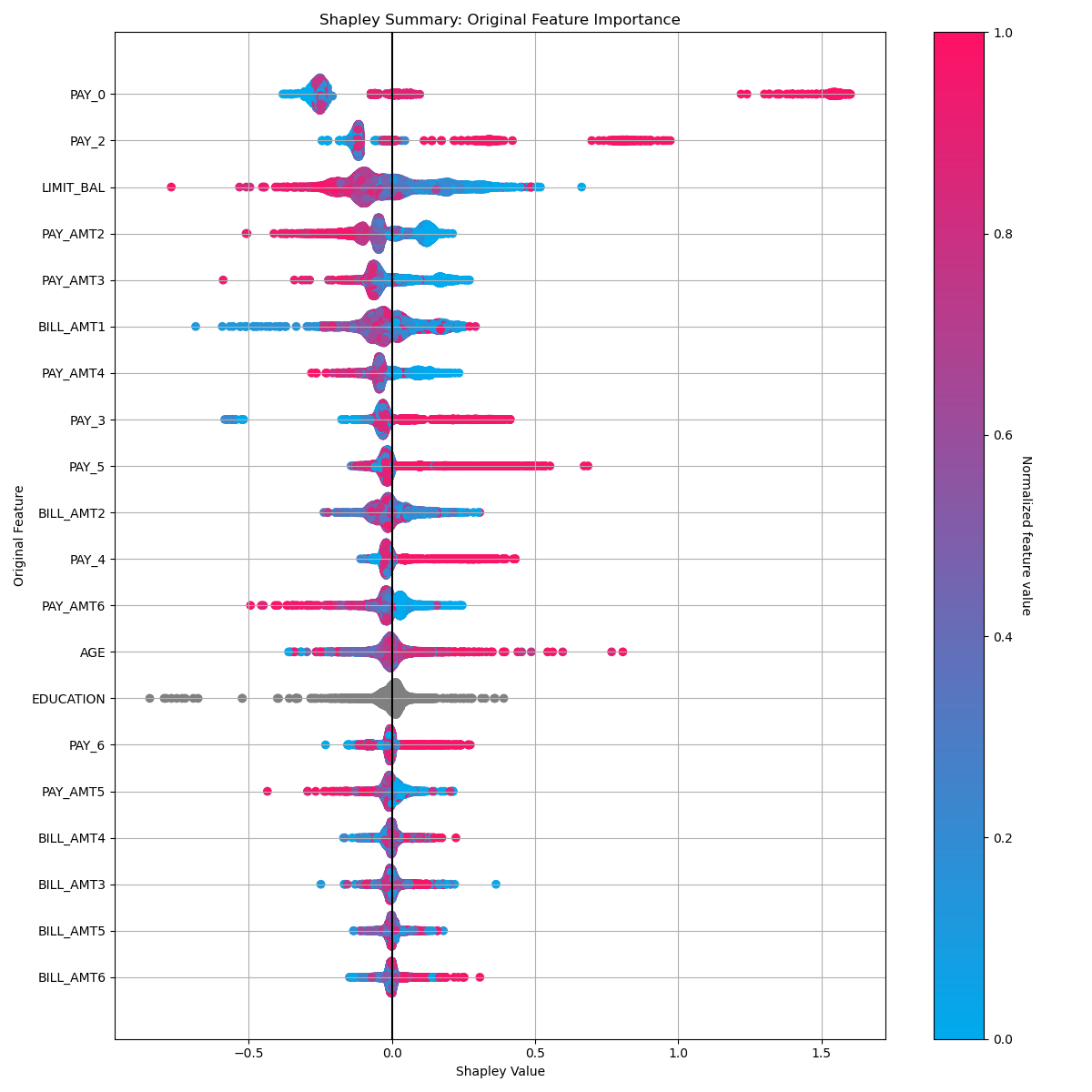
## Shapley Values

##### Shapley Contributions on the Test Dataset

Shapley explanations are a technique with credible theoretical support that presents consistent global and local feature contributions. For regression problems, local Shapley feature contributions plus the bias term sum to the final model's prediction. For classification problems, they sum to the prediction before applying the link function.

This section uses Driverless AI's Naive Shapley method to calculate local Shapley explanations for original features. These explanations are approximation for the original features and are based on how often the features are used in transformed features, and how important those transformed features are to the final model. The importance of each transformed feature is distributed equally to all original features that helped create it. This is then summed for each original feature.

The following Shapley summary plot is created from a random sample of 10000 rows (the autodoc\_pd\_max\_rows configuration controls random sampling for this plot).



## Appendix

### Final Model Details

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Index** | **Type** | **Model Weight** | **Num Folds** | **Fitted features** | **Target Transformer** |
| 0 | XGBoostGBMModel | 1 | 3 | 22 | LabelEncoder |

**Model Index: Final Model - Single Model**

|  |  |
| --- | --- |
| **parameter** | **value** |
| IS\_FINAL | True |
| \_build\_info | {'commit': '8bd2dc9', 'version': '1.10.1'} |
| accuracy | 1 |
| base\_score | 0.5 |
| booster | gbtree |
| colsample\_bytree | 0.8 |
| debug\_verbose | 0 |
| disable\_gpus | False |
| early\_stopping\_rounds |  |
| enable\_early\_stopping\_rounds | False |
| encoder |  |
| ensemble\_level | 1 |
| eval\_metric | auc |
| experiment\_description | 7.Test |
| gamma | 0.0 |
| gpu\_id | 0 |
| grow\_policy | depthwise |
| importance\_type | gain |
| interpretability | 1 |
| label\_counts | [18630, 5369] |
| labels | [0, 1] |
| layer | 0 |
| learning\_rate | 0.05 |
| lossguide | False |
| max\_bin | 64 |
| max\_delta\_step | 0.0 |
| max\_depth | 6 |
| max\_leaves | 64 |
| min\_child\_weight | 1 |
| model\_class\_name | XGBoostGBMModel |
| model\_id | Final Model - Single Model |
| model\_origin | SEQUENCE |
| monotonicity\_constraints | False |
| n\_estimators | 95 |
| n\_gpus | 1 |
| n\_jobs | 8 |
| ngenes | 22 |
| ngenes\_max | 10000000 |
| nthread | 8 |
| num\_class | 1 |
| num\_classes | 2 |
| num\_layers | 1 |
| objective | binary:logistic |
| optuna\_trial |  |
| optuna\_trial\_state | unused |
| outer\_trial | [0, 2] |
| pred\_gap |  |
| pred\_periods |  |
| random\_state | 383676314 |
| reg\_alpha | 0.0 |
| reg\_lambda | 0.0 |
| resumed\_experiment\_id | 37769638-3740-11ec-9ec3-ac1f6b6b49a6 |
| scale\_pos\_weight | 1.0 |
| score\_f\_name | AUC |
| seed | 383676314 |
| silent | 1 |
| str\_uuid | ret\_ab544eb5-c50c-492d-8abc-c31c8933d4b8 |
| subsample | 0.7 |
| target |  |
| test\_data\_name | [Test] |
| tgc |  |
| time\_column |  |
| time\_tolerance | 1 |
| train\_dataset\_name | CreditCard\_Cat-train.csv |
| train\_shape | [23999, 25] |
| tree\_method | gpu\_hist |
| tsp |  |
| ufapt |  |
| uses\_gpu | True |
| valid\_data\_name | [Valid] |
| valid\_shape | [1, 1] |
| nfolds | 1 |

### Config Overrides

The Config Overrides represent the fine-control parameters.

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| missing\_values | ['', '?', 'None', 'nan', 'NA', 'N/A', 'unknown', 'inf', '-inf', '1.7976931348623157e+308', '-1.7976931348623157e+308'] |
| monotonicity\_constraints\_dict | {} |
| optuna\_pruner\_kwargs | {'n\_startup\_trials': 5, 'n\_warmup\_steps': 20, 'interval\_steps': 20, 'percentile': 25.0, 'min\_resource': 'auto', 'reduction\_factor': 4, 'min\_early\_stopping\_rate': 0, 'n\_brackets': 4, 'min\_early\_stopping\_rate\_low': 0, 'upper': 1.0, 'lower': 0.0} |
| optuna\_sampler\_kwargs | {} |
| params\_lightgbm | {} |
| params\_xgboost | {} |
| params\_dart | {} |
| params\_tensorflow | {} |
| params\_gblinear | {} |
| params\_decision\_tree | {} |
| params\_rulefit | {} |
| params\_ftrl | {} |
| params\_grownet | {} |
| params\_tune\_lightgbm | {} |
| params\_tune\_xgboost | {} |
| params\_tune\_dart | {} |
| params\_tune\_tensorflow | {} |
| params\_tune\_gblinear | {} |
| params\_tune\_rulefit | {} |
| params\_tune\_ftrl | {} |
| params\_tune\_grownet | {} |
| autodoc\_github\_pat | ghp\_iYwSDQDRtVKEA97Y0LDULwk09GBMLK09nUfs |
| h2o\_recipes\_kwargs | {} |
| recipe\_dict | {} |
| enabled\_file\_systems | ['upload', 'file', 'hdfs', 's3', 'recipe\_file', 'recipe\_url'] |
| supported\_file\_types | ['csv', 'tsv', 'txt', 'dat', 'tgz', 'gz', 'bz2', 'zip', 'xz', 'xls', 'xlsx', 'jay', 'feather', 'bin', 'arff', 'parquet', 'pkl', 'orc', 'avro'] |
| recipe\_supported\_file\_types | ['py', 'pyc', 'zip'] |
| dask\_cuda\_cluster\_kwargs | {'scheduler\_port': 0, 'dashboard\_address': ':0', 'protocol': 'tcp'} |
| dask\_cluster\_kwargs | {'n\_workers': 1, 'processes': True, 'threads\_per\_worker': 1, 'scheduler\_port': 0, 'dashboard\_address': ':0', 'protocol': 'tcp'} |
| dask\_scheduler\_env | {} |
| dask\_cuda\_scheduler\_env | {} |
| dask\_worker\_env | {'NCCL\_P2P\_DISABLE': '1', 'NCCL\_DEBUG': 'WARN'} |
| dask\_cuda\_worker\_env | {} |
| ts\_target\_trafo\_epidemic\_params\_dict | {} |
| application\_id | dai\_320892 |
| included\_transformers | ['AutovizRecommendationsTransformer', 'BERTTransformer', 'CVCatNumEncodeTransformer', 'CVTECUMLTransformer', 'CVTargetEncodeTransformer', 'CatOriginalTransformer', 'CatTransformer', 'ClusterDistCUMLDaskTransformer', 'ClusterDistCUMLTransformer', 'ClusterDistTransformer', 'ClusterIdAllNumTransformer', 'ClusterTETransformer', 'DBSCANCUMLDaskTransformer', 'DBSCANCUMLTransformer', 'DateOriginalTransformer', 'DateTimeDiffTransformer', 'DateTimeOriginalTransformer', 'DatesTransformer', 'EwmaLagsTransformer', 'FrequentTransformer', 'ImageOriginalTransformer', 'ImageVectorizerTransformer', 'InteractionsTransformer', 'IsHolidayTransformer', 'IsolationForestAnomalyAllNumericTransformer', 'IsolationForestAnomalyNumCatAllColsTransformer', 'IsolationForestAnomalyNumCatTransformer', 'IsolationForestAnomalyNumericTransformer', 'LagsAggregatesTransformer', 'LagsInteractionTransformer', 'LagsTransformer', 'LexiLabelEncoderTransformer', 'MeanTargetTransformer', 'NumCatTETransformer', 'NumToCatTETransformer', 'NumToCatWoEMonotonicTransformer', 'NumToCatWoETransformer', 'OneHotEncodingTransformer', 'OriginalTransformer', 'StandardScalerTransformer', 'StringConcatTransformer', 'TSNECUMLTransformer', 'TextBiGRUTransformer', 'TextCNNTransformer', 'TextCharCNNTransformer', 'TextLinModelTransformer', 'TextOriginalTransformer', 'TextTransformer', 'TimeSeriesTargetEncTransformer', 'TruncSVDAllNumTransformer', 'TruncSVDCUMLDaskTransformer', 'TruncSVDCUMLTransformer', 'TruncSVDNumTransformer', 'UMAPCUMLDaskTransformer', 'UMAPCUMLTransformer', 'WeightOfEvidenceTransformer'] |
| included\_models | ['Constant', 'DecisionTree', 'FTRL', 'GLM', 'ImageAuto', 'ImbalancedLightGBM', 'ImbalancedXGBoostGBM', 'LightGBM', 'LightGBMDask', 'RFCUML', 'RFCUMLDask', 'RuleFit', 'TensorFlow', 'TextALBERT', 'TextBERT', 'TextCamemBERT', 'TextDistilBERT', 'TextMultilingualBERT', 'TextRoBERTa', 'TextXLM', 'TextXLMRoberta', 'TextXLNET', 'TorchGrowNet', 'XGBoostDart', 'XGBoostDartDask', 'XGBoostGBM', 'XGBoostGBMDask', 'XGBoostRF', 'XGBoostRFDask'] |
| included\_scorers | ['ACCURACY', 'AUC', 'AUCPR', 'F05', 'F1', 'F2', 'FDR', 'FNR', 'FOR', 'FPR', 'GINI', 'LOGLOSS', 'MACROAUC', 'MACROF1', 'MACROMCC', 'MCC', 'NPV', 'PRECISION', 'RECALL', 'TNR'] |
| autodoc\_report\_name | https://github.com/Mathanraj-Sharma/python\_boilerplate/tree/dev/tests/report-dai.docx |
| last\_recipe | auto |
| recipe\_activation | {'transformers': [], 'models': [], 'scorers': [], 'data': []} |
| last\_exclusive\_mode | safe |
| prob\_lag\_non\_targets | 0.1 |
| prob\_default\_lags | 0.2 |
| prob\_lagsinteraction | 0.2 |
| prob\_lagsaggregates | 0.2 |