Case Study: Predicting Customer Churn at RetailGenius

AI Project Methodology

SHAP Report

Report Submitted

by

Sanjaya Deshapriya Gunawardena

EPITA FALL 2023

SHAP Report

First imported the required packages and libraries.

```
import sys
sys.path.append('..')

import shap
import pandas as pd
import numpy as np
import joblib
from Churn_predict.inference import make_predictions
from Churn_predict.preprocess import preprocess_data_with_objects
from Churn_predict import selected_features
import matplotlib.pyplot as plt
import xgboost as xgb
from shap import Explainer, Explanation
from shap import waterfall_plot
```

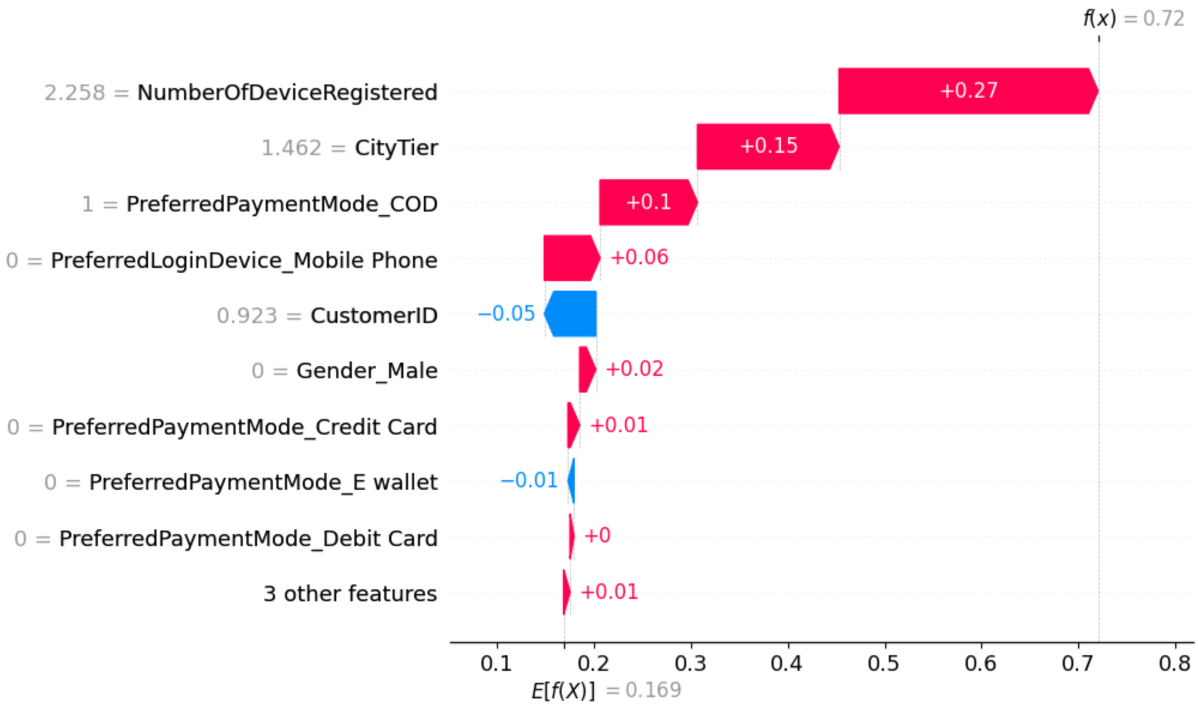
Then initialize the SHAP JavaScript library with shap.initjs(), you enable the rendering of SHAP plots directly within the notebook environment.

Let's go through the steps in the SHAP.ipynb:

1. **Load the Model**: Load a pre-trained machine learning model from a file.
2. **Read Data**: Read a test dataset.
3. **Select Relevant Data**: Select specific features from the test dataset that are used as input for the model.
4. **Preprocess Data**: Preprocess the selected data, likely handling missing values, encoding categorical variables, and scaling numerical features. (using the function preprocess_data_with_objects).
5. **Create Explainer Object**: Initialize an explainer object using the SHAP library, passing the pre-trained model.
6. **Compute SHAP Values**: Calculate SHAP values for the processed data. These values help in understanding the impact of each feature on the model's predictions.

7. **Create Explanation Object**: Create an Explanation object containing SHAP values, base values, data, and feature names. This object is used to generate visualizations and interpret the model's predictions.
8. **Select Data Point**: Choose a specific data point from the dataset for which you want to explain the model's prediction. Used the first data point.
9. **Generate Waterfall Plot**: Create a waterfall plot to visualize how each feature contributes to the final prediction for the selected data point. This plot helps in understanding the factors driving the model's decision for that particular instance.

<u>**Waterfall Plot**</u>



- **Efx value (0.169)**: This represents the expected or average effect of the feature on the model's output across all data points. In other words, on average, this feature contributes 0.169 to the model's prediction.
- **Fx value (0.72)**: This represents the observed or actual value of the feature for the specific data point being examined. In this case, for this data point, the value of the feature is 0.72.

Comparing these values can help you understand how the specific value of the feature for this data point deviates from the expected average effect across all data points. If Fx > Efx, it indicates that the feature value for this data point has a stronger positive impact on the prediction compared to the average effect.

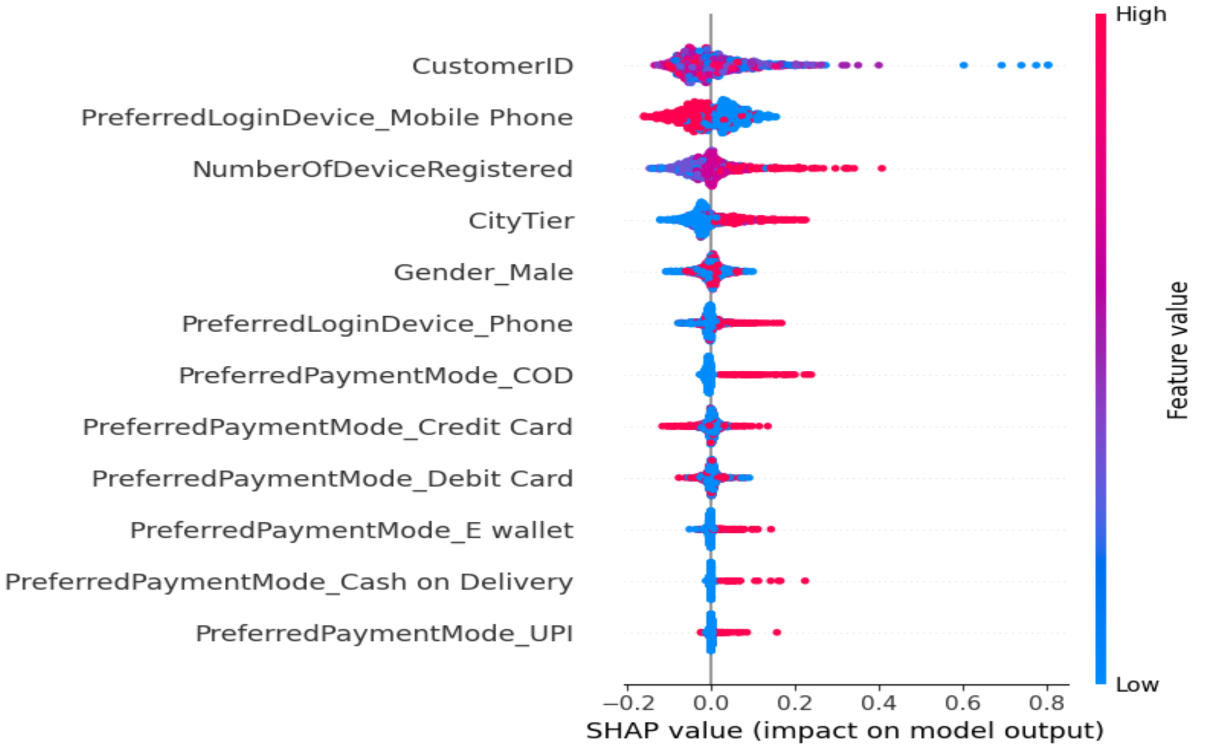
Based on feature contribution.

- **NumberOfDeviceRegistered (2.258)**: This feature has the highest positive contribution to the model's prediction. A higher number of devices registered is associated with a significantly higher predicted outcome.
- **CityTier (1.462)**: The CityTier feature also positively contributes to the prediction but to a lesser extent compared to NumberOfDeviceRegistered. Being in a higher city tier is associated with a moderately higher predicted outcome.
- **PreferredPaymentMode_COD (1)**: The PreferredPaymentMode_COD feature has a positive contribution, indicating that customers who prefer Cash on Delivery tend to have a higher predicted outcome compared to those who don't prefer it.
- **PreferredLoginDevice_Mobile Phone (0.1)**: While still positive, the contribution of this feature is relatively small compared to the previous ones. Customers who prefer to log in using a mobile phone have a slightly higher predicted outcome.
- **CustomerID (0.923)**: The CustomerID feature also has a positive contribution, suggesting that certain customer IDs are associated with a higher predicted outcome.
- **Gender_Male (0)**: Gender_Male has no contribution to the prediction, suggesting that gender does not significantly affect the predicted outcome in this model.
- **PreferredPaymentMode_Credit Card, PreferredPaymentMode_E wallet, PreferredPaymentMode_Debit Card**: These features have no contribution to the prediction, suggesting that the choice of payment mode (Credit Card, E wallet, Debit Card) does not significantly impact the predicted outcome.
- **Other features**: There are three other features that are not specified but have some contribution to the prediction.

Overall, the model's prediction is primarily influenced by the number of devices registered, city tier, and preference for Cash on Delivery, while other factors such as login device and customer ID also play a role, to a lesser extent. Gender and choice of payment mode seem to have minimal impact on the predicted outcome according to the model.

Summary Plot

Visualize explanation for all points of the dataset at once.

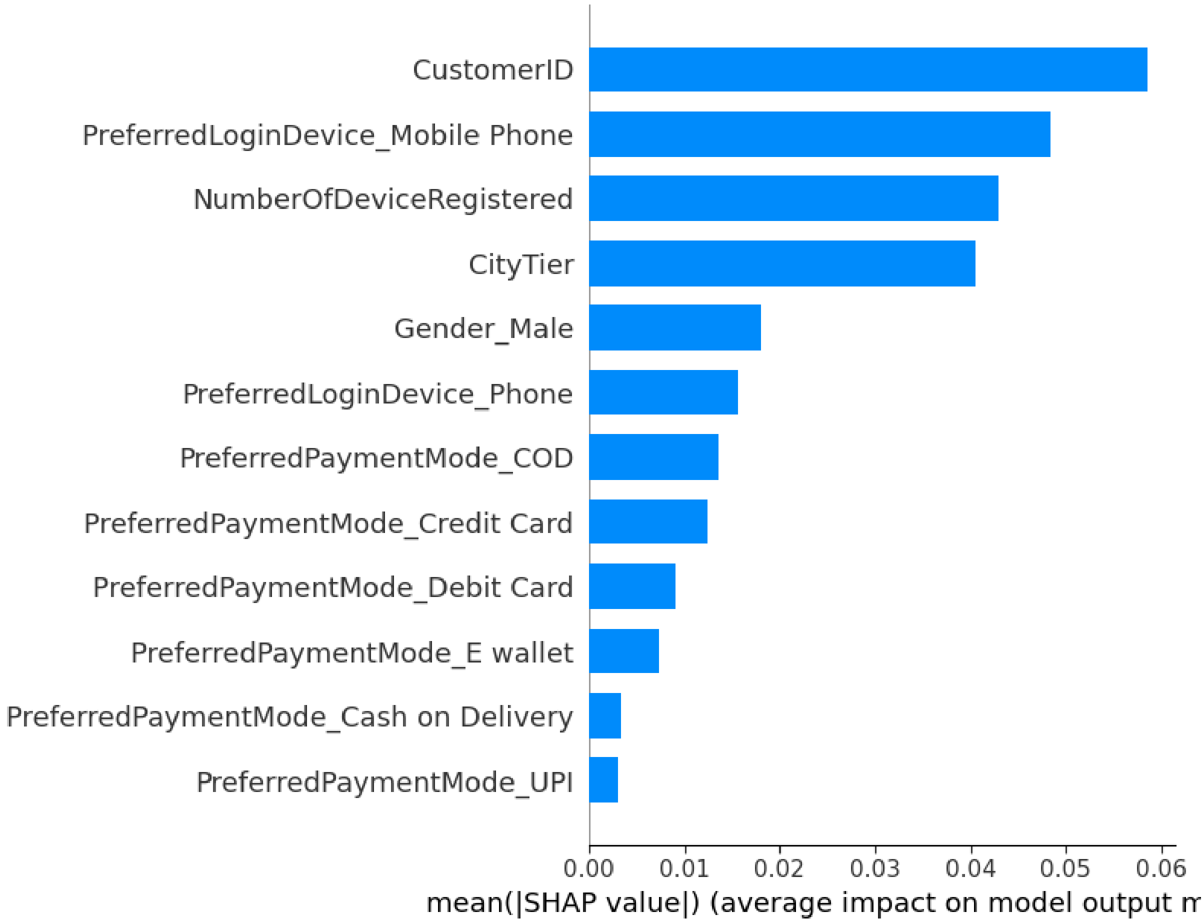


1. **Plot Description**:
    - The plot illustrates the impact of various features on the output of a machine learning model.
    - Each point represents a **SHAP value** associated with a specific feature and an individual prediction.
2. **Features Displayed**:
    - The x-axis represents the **SHAP value**, indicating the effect on the model's output. Positive values increase the probability of a certain class, while negative values decrease it.
    - Each row on the y-axis corresponds to a different feature that influences the model's predictions.

- The color of each point ranges from pink to blue:
  - Pink indicates high feature values.
  - Blue indicates low feature values.

3. **Interpretation**:
   - Most features exhibit both positive and negative impacts, as evidenced by points scattered on both sides of the zero line.
   - The SHAP values help us understand how each feature contributes to the model's predictions.

Visualize a summary plot for each feature.



This plot helps us understand how different **features** impact the output of a machine learning model.

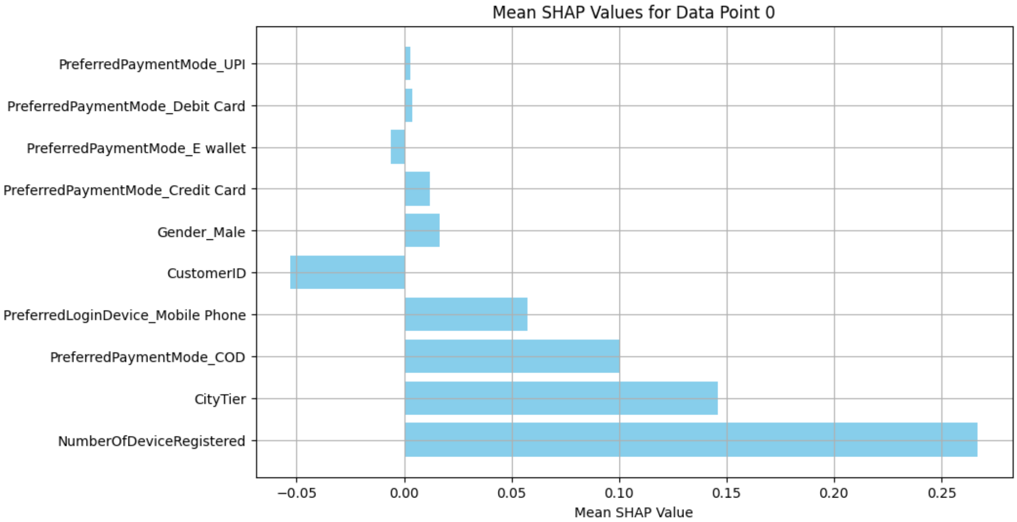
1. **Features and Their Impact**:
    - Each feature (like **CustomerID**, **PreferredLoginDevice**, etc.) has a bar.
    - The longer the bar, the **more influential** that feature is in the model's predictions.
    - Features can have **positive** or **negative** impact (shown on both sides of the zero line).
2. **Interpretations**:
    - **CustomerID** has the highest impact.
    - **PreferredLoginDevice (Mobile Phone)** and **NumberOfDeviceRegistered** also matter.
    - Payment modes and gender have smaller impacts.

Force Plot

The plot shows how different **features** influence the model's decision for a specific prediction.

1. **Features and Their Impact**:
    - Each feature (like **City Tier**, **NumberOfDeviceRegistered**, etc.) has a bar.
    - Longer bars mean the feature has a **bigger impact** on the prediction.
    - Some features push the prediction **higher** (to the right), while others pull it **lower** (to the left).
2. **Interpretation**:
    - Look at the bars:
        - **City Tier** (1.462) and **NumberOfDeviceRegistered** (2.258) push the prediction higher (red).
        - **CustomerID** (0.9228) has less impact (blue).
    - The final prediction score is **0.72** (far right).

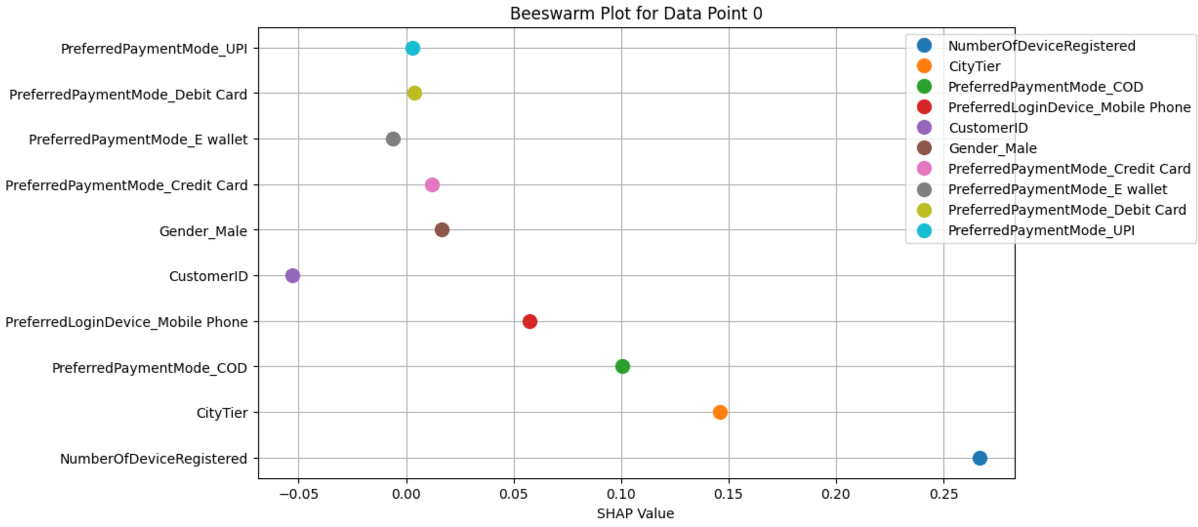
Mean SHAP values for data point 0



Magnitude of Impact: The length of each bar indicates the magnitude of the feature's impact on the model's prediction for data point 0. Longer bars suggest features that have a stronger influence on the prediction, either positively or negatively.

Interpretation: By examining the bar plot, we can identify which features are most influential in determining the model's prediction for data point 0. Features with longer bars contribute more to the prediction, while features with shorter bars have less impact.

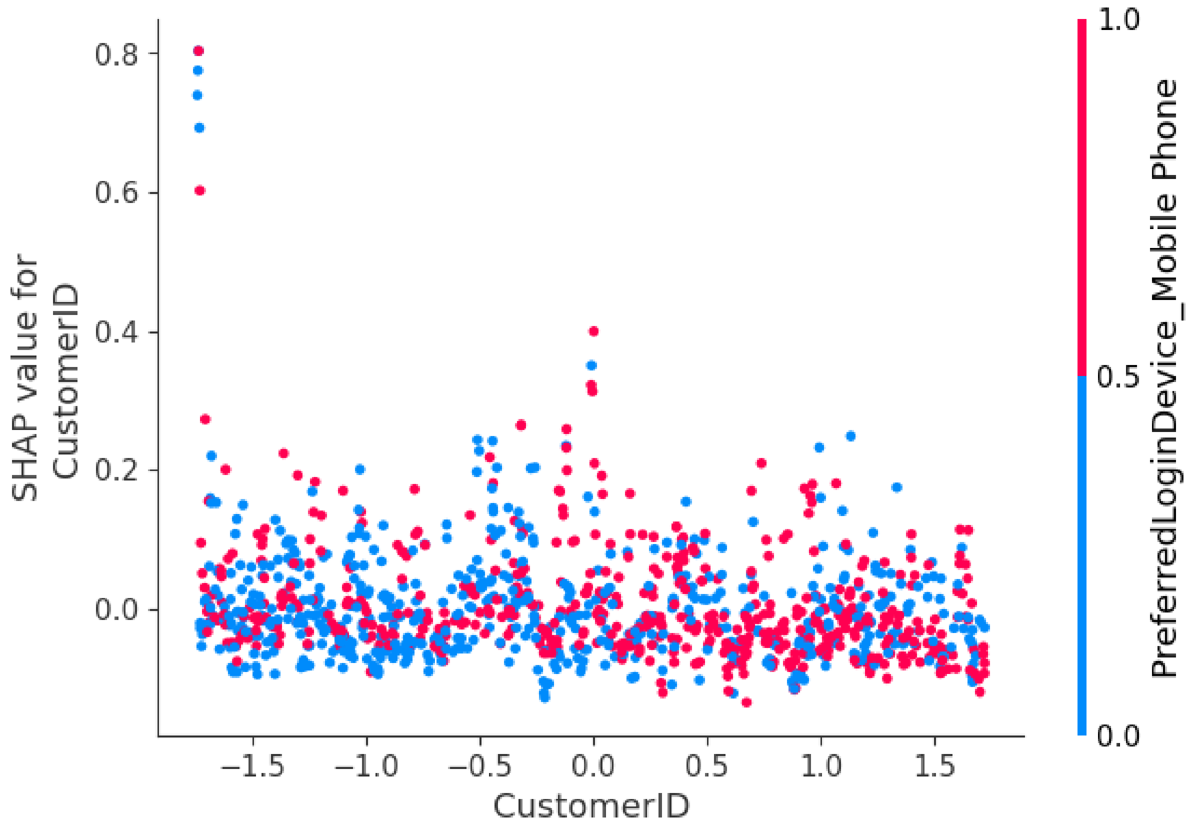
Beeswarm Plot



The position of the dot along the x-axis represents the SHAP value for that instance, indicating how much that feature contributed to the model's prediction. A higher SHAP value means that the feature had a greater impact on the model's prediction.

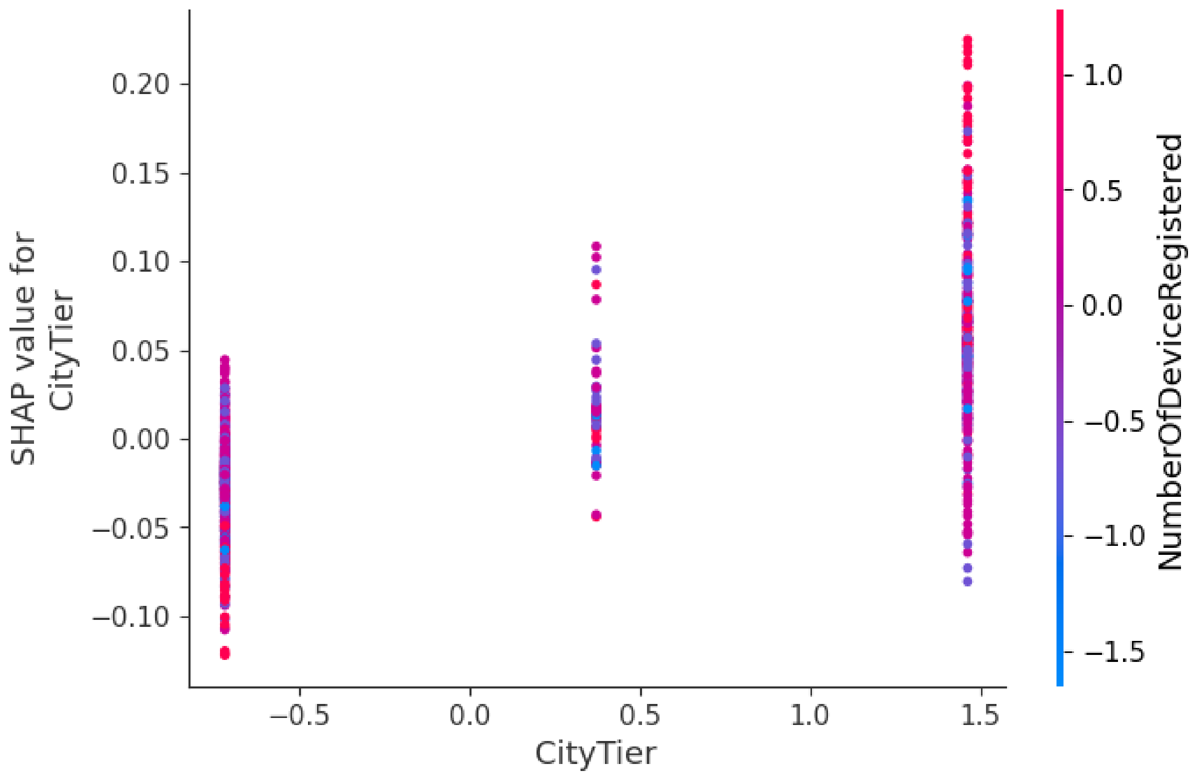
Dependence Plots

- Dependence Plot between CustomerID and PreferredLoginDevice_MobilePhone:



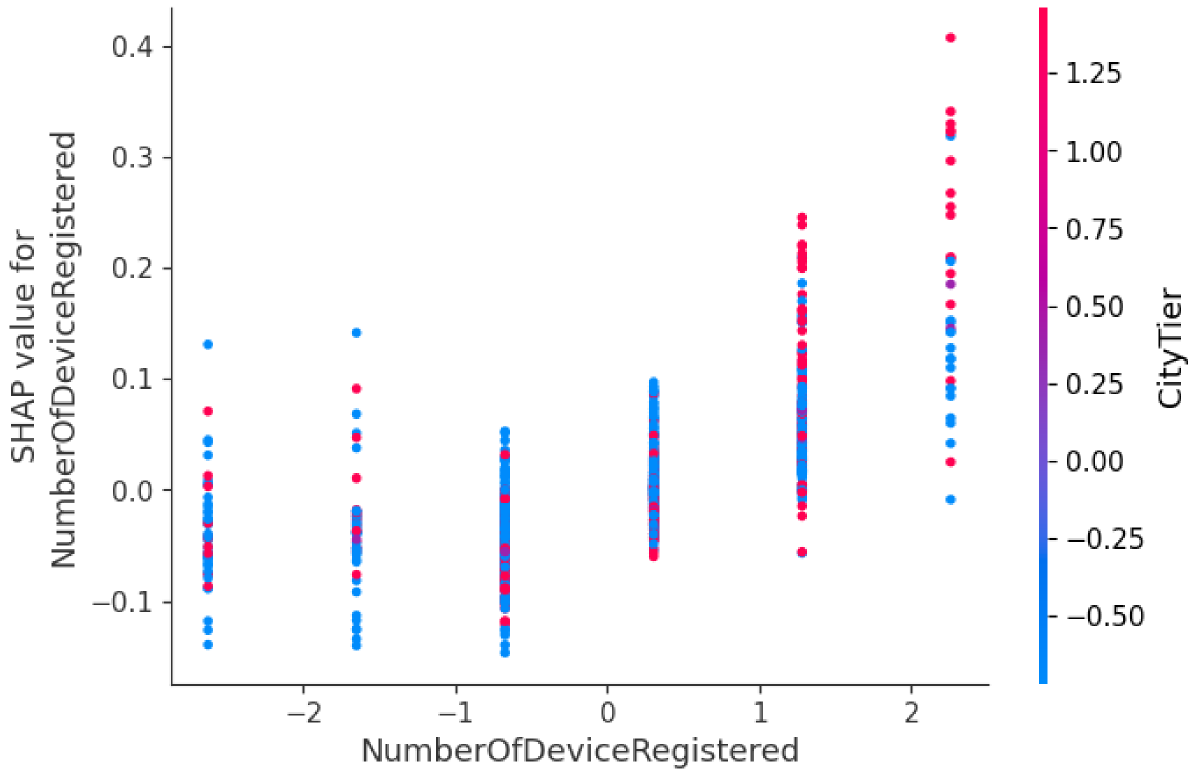
  - The y-axis shows the **SHAP value for CustomerID** (ranging from 0 to 0.8).
  - Each dot on the plot represents a data point.
  - The color of each dot indicates the **preferred login device**:
    - Red dots: High preference for **mobile phones** (value = 1).
    - Blue dots: Low preference for mobile phones (value = 0).
  - Most data points cluster around **low SHAP values** (between 0 and 0.2).
  - Some outliers have **higher SHAP values** (up to 0.8).
  - The plot helps us understand how **CustomerID** influences predictions, especially when considering preferred login devices.

- Dependence plot between CityTier and NumberOfDeviceRegistered



  - The x-axis represents **City Tier** values (ranging from -0.5 to 1.5).
  - The y-axis shows the **SHAP value for City Tier** (ranging from -0.10 to 0.20).
  - Each dot on the plot represents a data point.
  - The color of each dot indicates the **preferred login device**:
    - Red dots: High preference for **mobile phones** (value = 1).
    - Blue dots: Low preference for mobile phones (value = 0).
  - Most data points cluster around **low SHAP values** (between 0 and 0.2).
  - Some outliers have **higher SHAP values** (up to 0.8).
  - The plot helps us understand how **City Tier** influences predictions, especially when considering preferred login devices.

- Dependence plot between NumberOfDeviceRegistered and CityTier



The dependence plot shows how a specific feature (in this case, **NumberOfDeviceRegistered** influences the model's output.

- The x-axis represents **NumberOfDeviceRegistered** values (ranging from -2 to 2).
- The y-axis shows the **SHAP value for NumberOfDeviceRegistered** (ranging from -0.1 to 0.4).
- Each dot on the plot represents a data point.
- The color of each dot indicates the **city tier**:
  - Red dots: High city tier preference.
  - Blue dots: Low city tier preference.
- Most data points cluster around **low SHAP values** (between 0 and 0.2).
- Some outliers have **higher SHAP values** (up to 0.8).

- The plot helps us understand how **NumberOfDeviceRegistered** influences predictions, especially considering different city tiers