Tutorial on Feature Importance

Mitodru Niyogi,
M.Sc., B.Tech.

Interdisciplinary Center for Scientific Computing,
Heidelberg University, Heidelberg, Germany
niyogi@cl.uni-heidelberg.de

June 16, 2023

Outline

Impurity-based Feature Importance

Impurity-based feature importance refers to a method of evaluating the importance of features in a decision tree or random forest model based on the decrease in impurity or entropy that each feature provides when making splits in the tree.

Limitations of Impurity-based Feature Importance

This problem stems from two limitations of impurity-based feature importance:

- Impurity-based importance is biased towards high cardinality features.
- Impurity-based feature importance can inflate the importance of numerical features due to allowing for more possible split points because of continuous values and hence more decrease in impurity.
- Impurity-based importance is computed on training set statistics and therefore do not reflect the ability of a feature to be useful in making predictions that generalize to the test set (when the model has enough capacity).

Permutation Importance Algorithm

Algorithm 1: Estimating Feature Importance

Input: Model fitted to the training set

Output: Feature importance scores

Estimate baseline performance on an independent dataset;

foreach *feature j* **do**

  Randomly permute feature column *j* in the original dataset;

  Measure the performance of the model on the permuted dataset;

  Compute the feature importance as the difference between baseline performance and performance on the permuted dataset;

end

Repeat the above steps exhaustively or a large number of times;

Compute the feature importance as the average difference;

Permutation Importance Pros and Cons

+ Model agnostic
+ Based on metric of choice
+ Easy to understand
± Feature importance is specific to the particular model and may vary for another model
+ Unlike impurity-based random forest importance, it does not suffer from "overfitting" since an independent dataset is used
- Like impurity-based random forest importance, the importance is undervalued if two features are highly correlated

SHAP:SHapley Additive exPlanations

- SHAP calculates Shapley values, representing each feature's contribution to the prediction.
- Shapley values quantify how much a feature influences the prediction by comparing scores with and without the feature.
- Removing features is equivalent to calculating the expectation value of the prediction across all possible removed feature values.
- SHAP deconstructs predictions into contributions from each input variable, providing insights into their individual effects.
- A machine learning model's prediction, $f(x)$, can be represented as the sum of its computed SHAP values, plus a fixed base value, such that:
  $f(x) = \text{base value} + \sum \text{SHAP values}.$

Pros and Cons of SHAP

Pros:

- Interpretable: Provides clear and intuitive interpretation of feature contributions.
- Model Agnostic: Can be applied to various machine learning models.
- Global and Local Interpretability: Offers insights at both global and local levels.
- Handles Feature Interactions: Detects and quantifies interactions between features.

Cons:

- Computational Complexity: Can be computationally expensive for complex models and large datasets.
- High-Dimensional Data: Interpretability challenges with a large number of features.
- Correlated Features: Influence of correlated features can affect interpretability.

SHAP Summary Plot

The summary plot combines feature importance with feature effects. Interpretation:

1. Each point represents a Shapley value for a feature and an instance.
2. The y-axis position corresponds to the feature.
3. The x-axis position corresponds to the Shapley value.
4. The color indicates the feature value (from low to high).
5. Overlapping points are jittered in the y-axis direction. giving a sense of the distribution of the Shapley values per feature
6. Features are ordered by importance.

SHAP Dependence Plot

A dependence plot is a scatter plot that shows the effect a single feature has on the predictions made by the model.
Key features of SHAP dependence plots:

- Show interaction effects between features unlike traditional partial dependence plots which show the average model output when changing a feature's value.
- Provide insights into the distribution of effects.
- Determine if the effect of a certain value is constant or varies based on other feature values.

Interpretation of SHAP dependence plot

- Each dot represents a single prediction (row) from the dataset.
- The x-axis represents the value of the feature (from the X matrix).
- The y-axis represents the SHAP value for that feature, indicating how much knowing that feature's value changes the output of the model for that sample's prediction.
- The color corresponds to a second feature that may have an interaction effect with the feature being plotted.
- A distinct vertical pattern of coloring indicates an interaction effect between the two features.

Model Performance Metrics

Table: Random Forest Regressor Model Performance

Metric	Training	Validation	Test
MSE	0.430 (0.014)	3.072 (0.204)	2.441
R^2	0.887 (0.003)	0.190 (0.045)	0.284095

Table: GBR Model Performance

Metric	Training	Validation	Test
MSE	2.112 (0.041)	2.904 (0.149)	2.490
R^2	0.444 (0.010)	0.234 (0.022)	0.269748

Results

Table: Decision Tree Regressor Model Performance

Metric	**Training**	**Validation**	**Test**
MSE	0.000 (0.000)	5.663 (0.406)	4.728
R^2	1.000 (0.000)	-0.493 (0.098)	-0.386619

PDP Plot of GBR

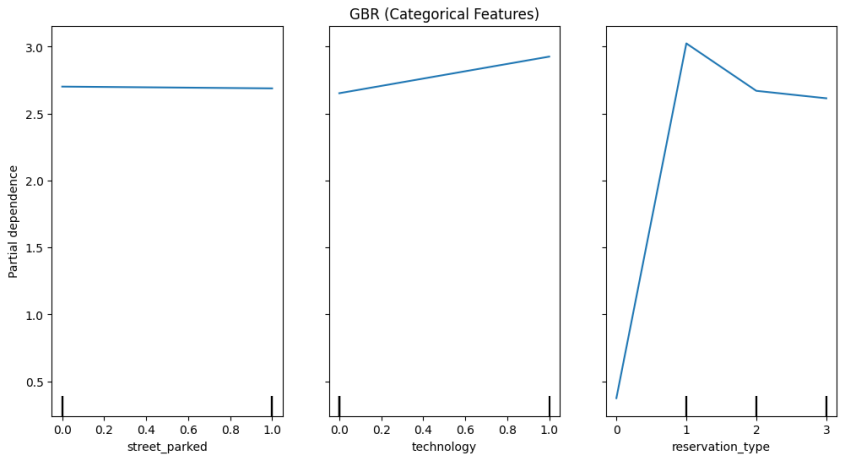
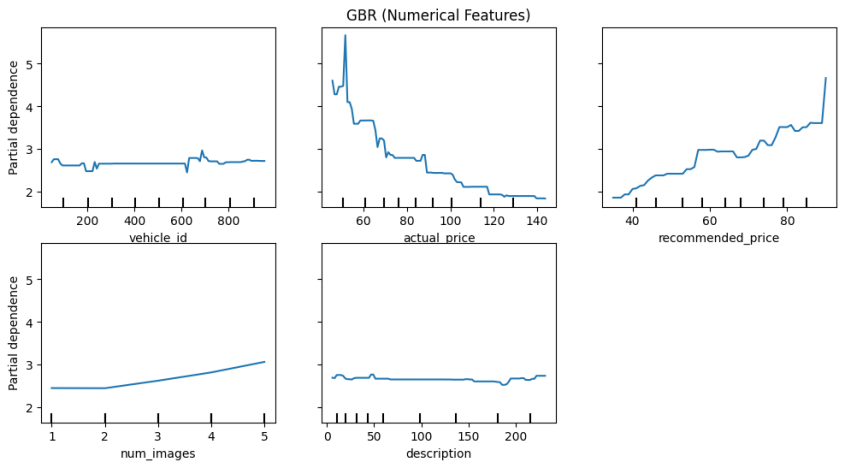
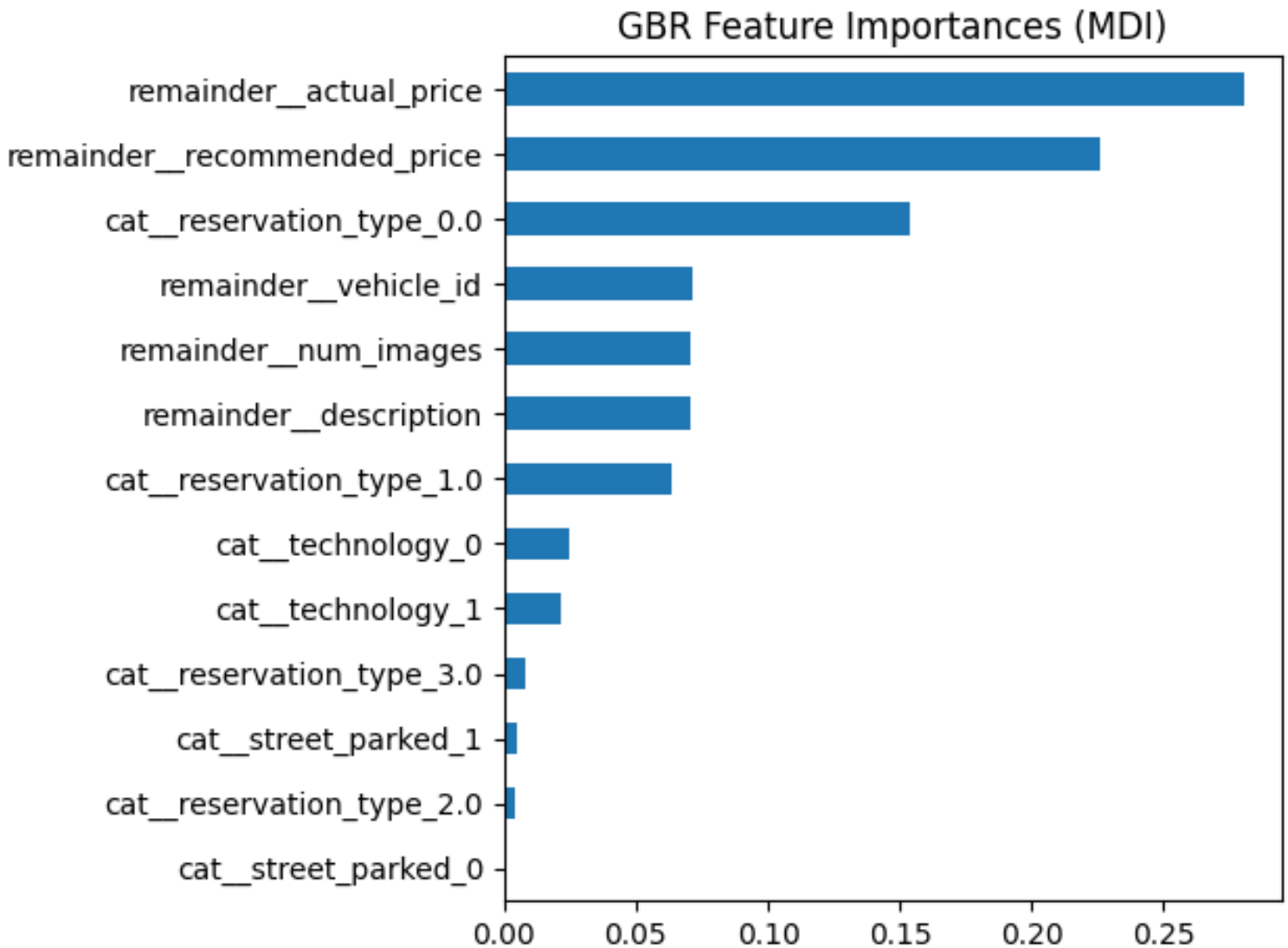


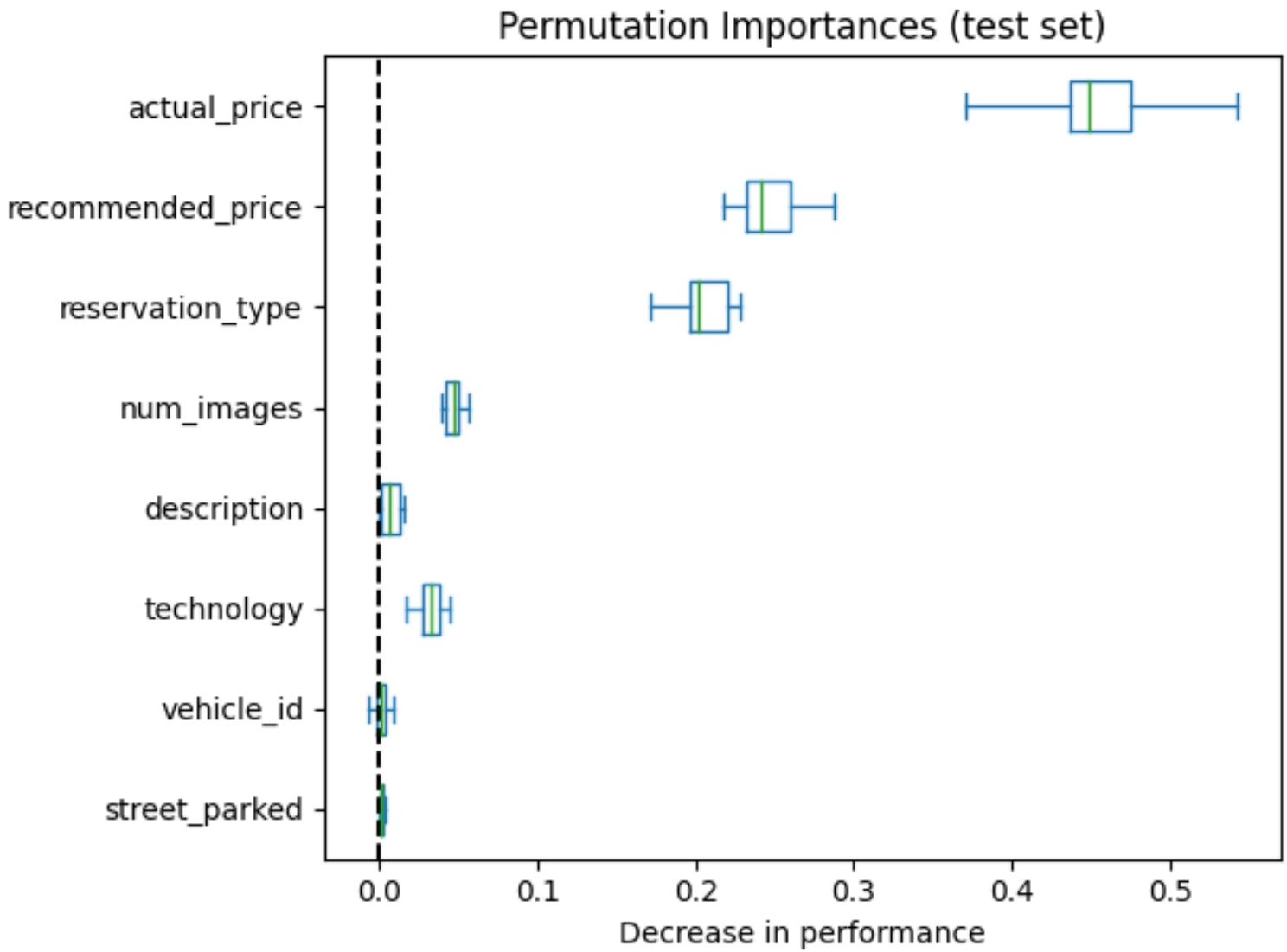
Figure: PDP Plot of GBR



Feature Importance (MDI)



Permutation Importance of GBR features on Test Set



SHAP Feature Importance Plot for GBR

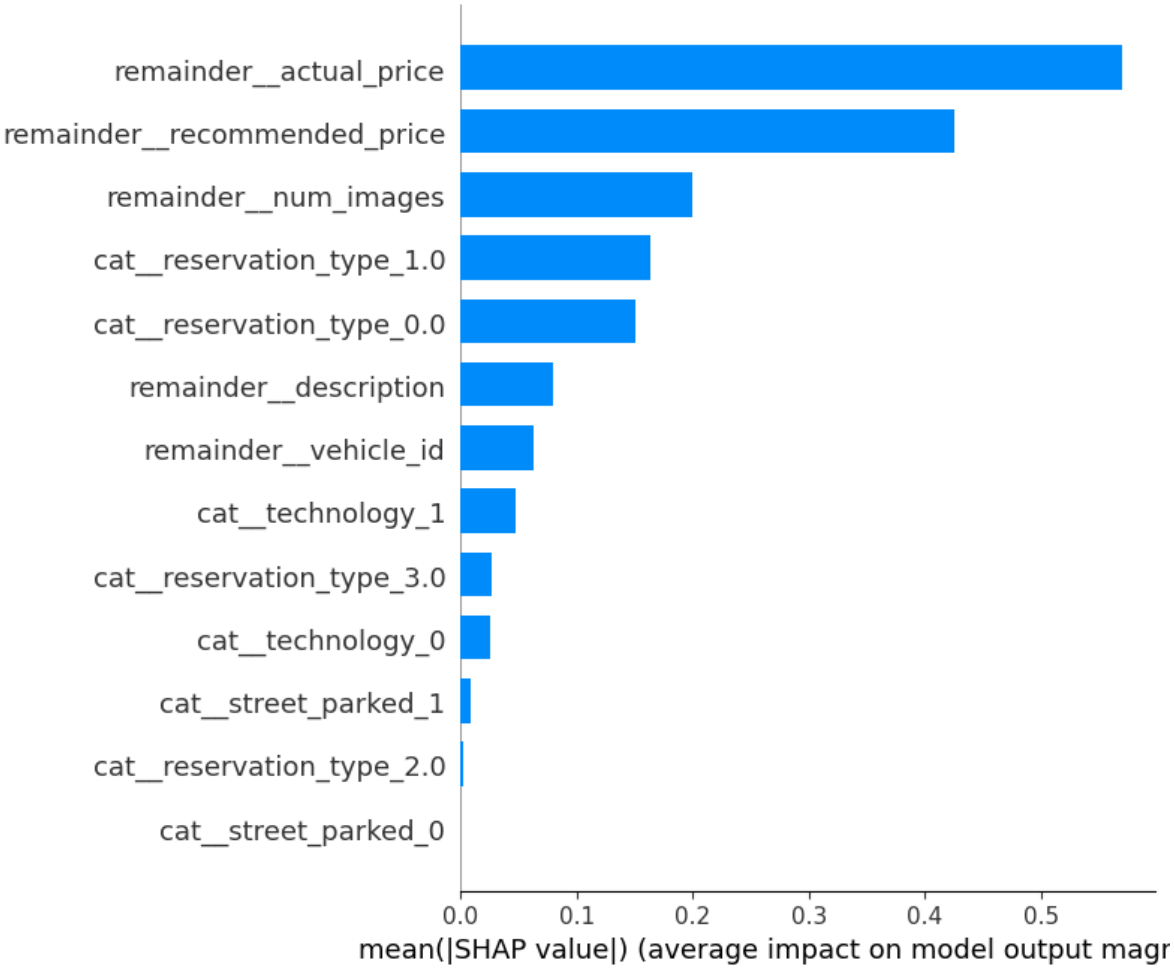
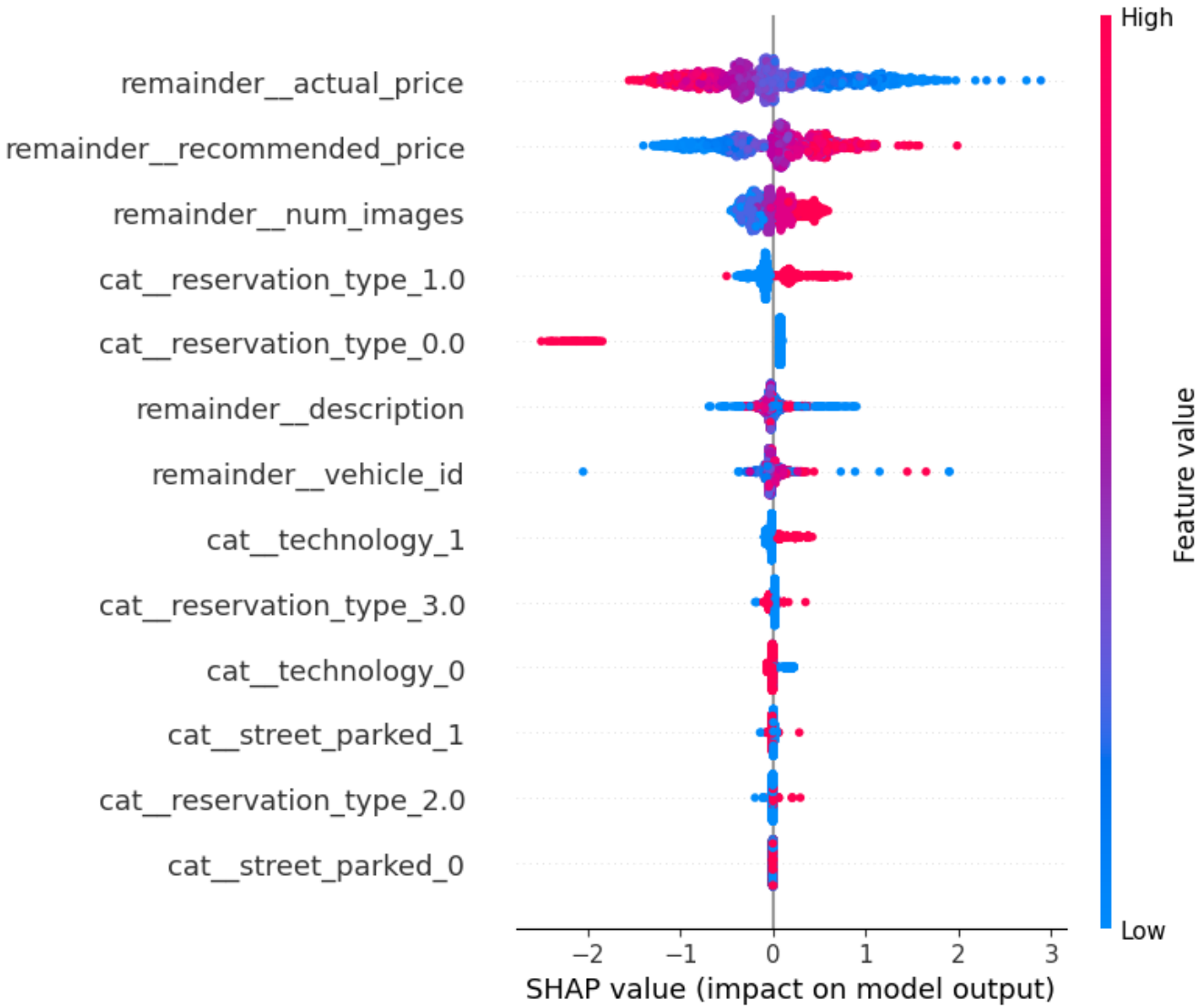
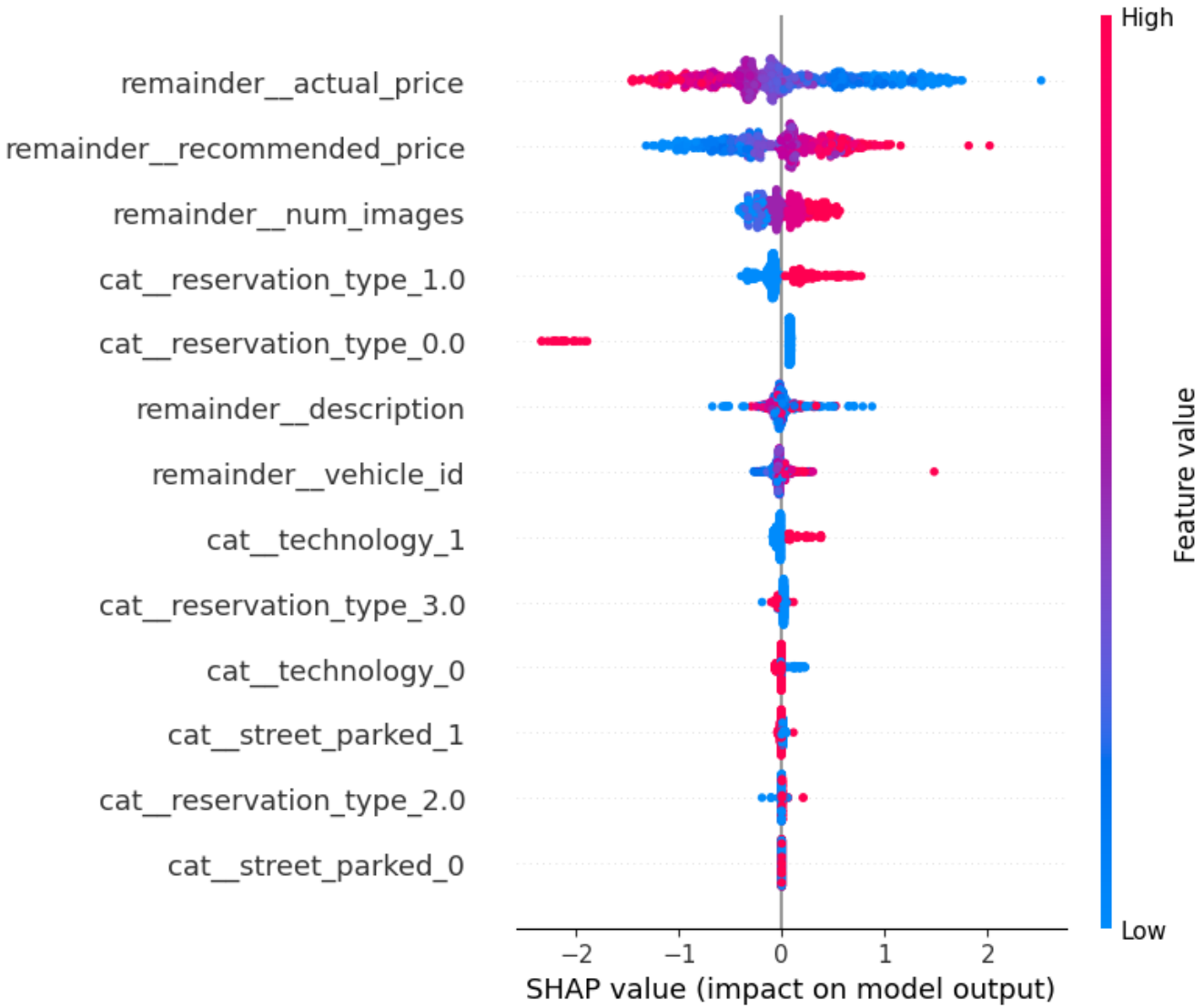


Figure: GBR Feature Importance Plot

SHAP Summary Plot



SHAP Summary Plot: Test Set



SHAP PDP of GBR

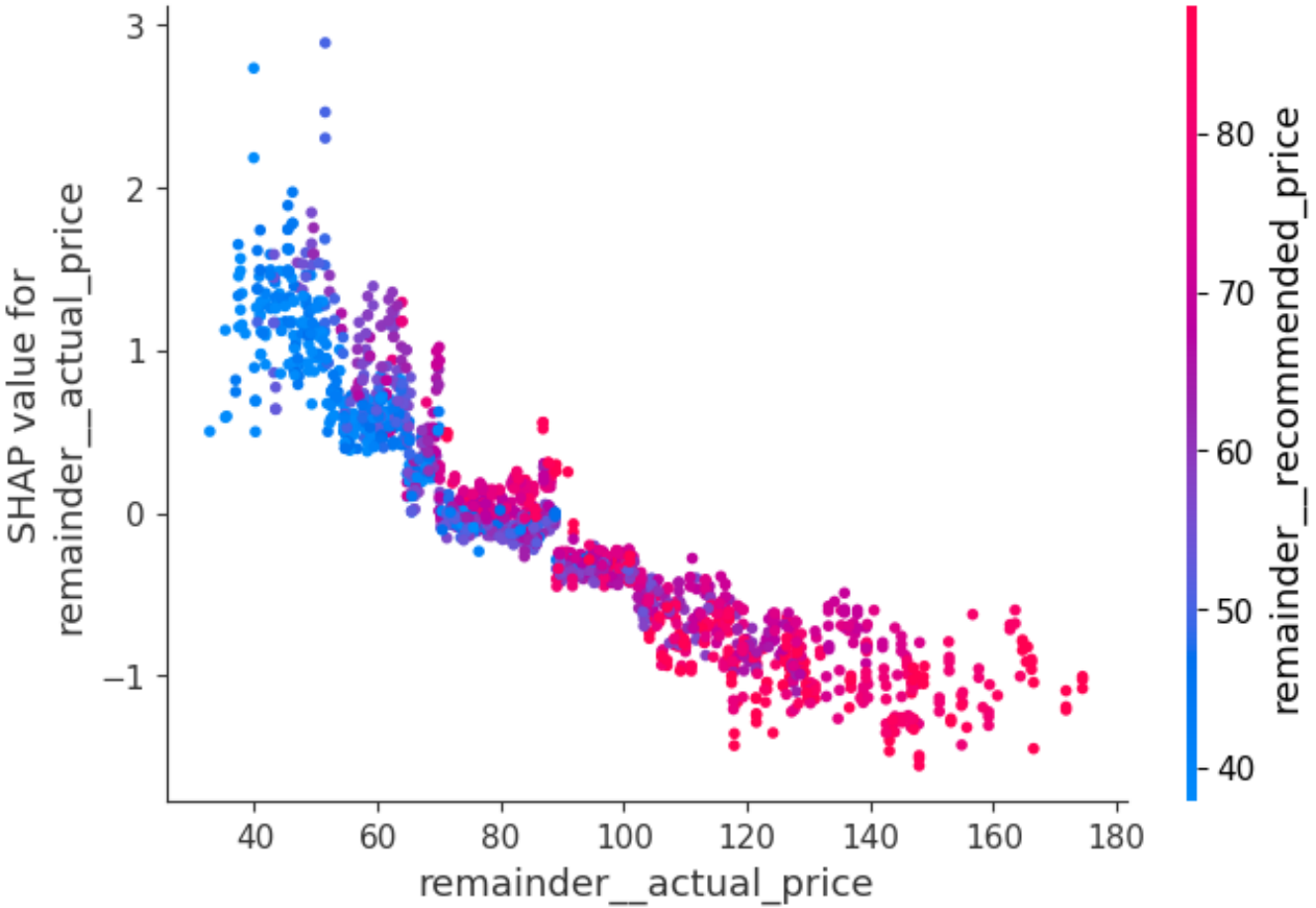


Figure: SHAP PDP (Rank1)

SHAP PDP of GBR

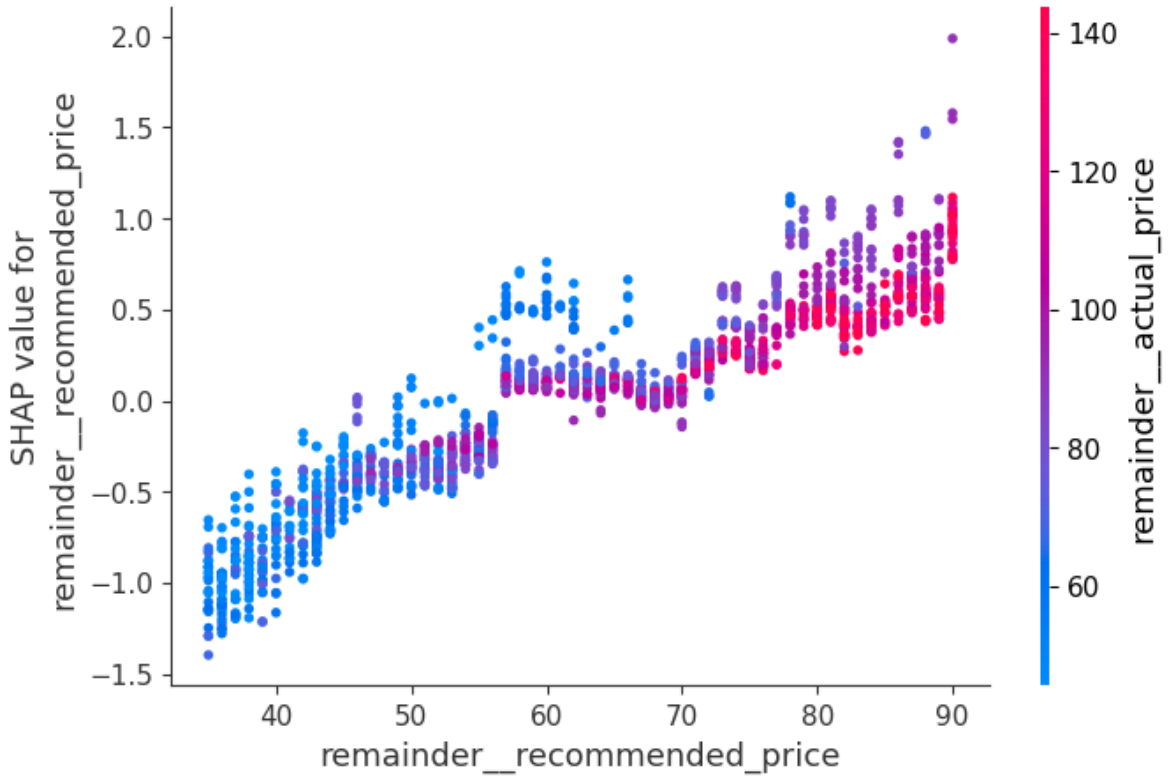


Figure: SHAP PDF for Rank 2

SHAP PDP of GBR

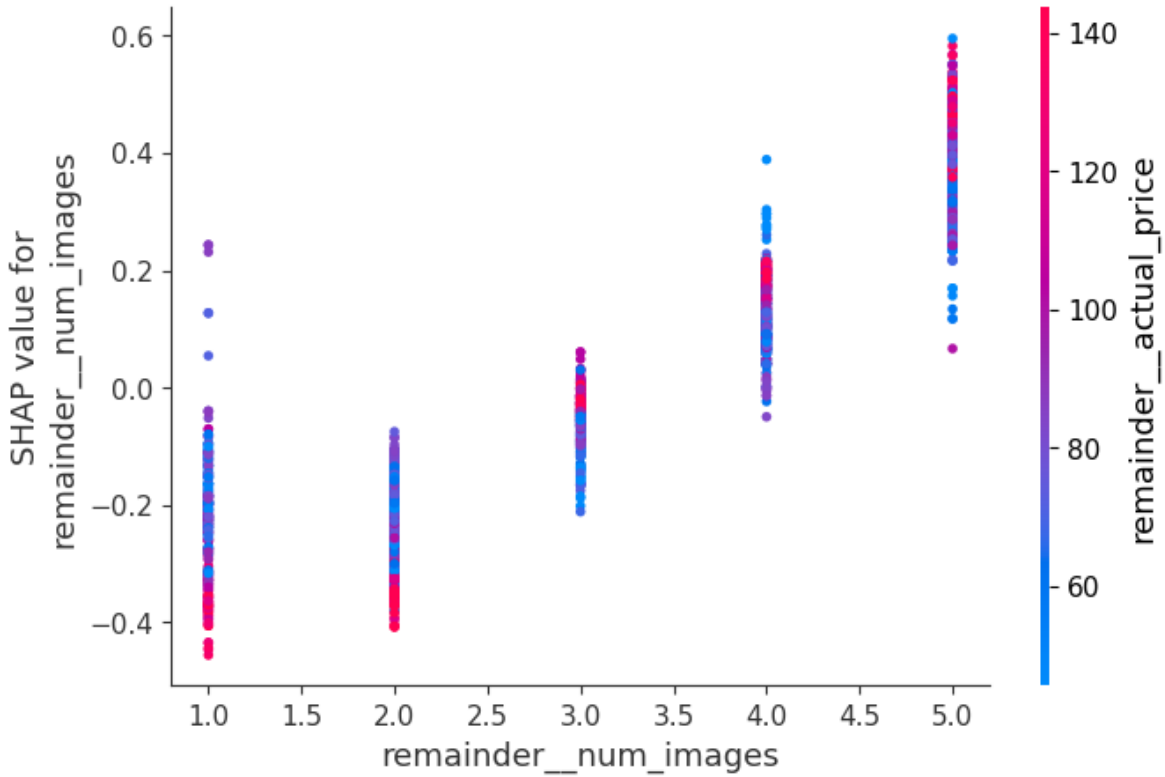


Figure: SHAP PDP Rank 3

SHAP PDP of GBR

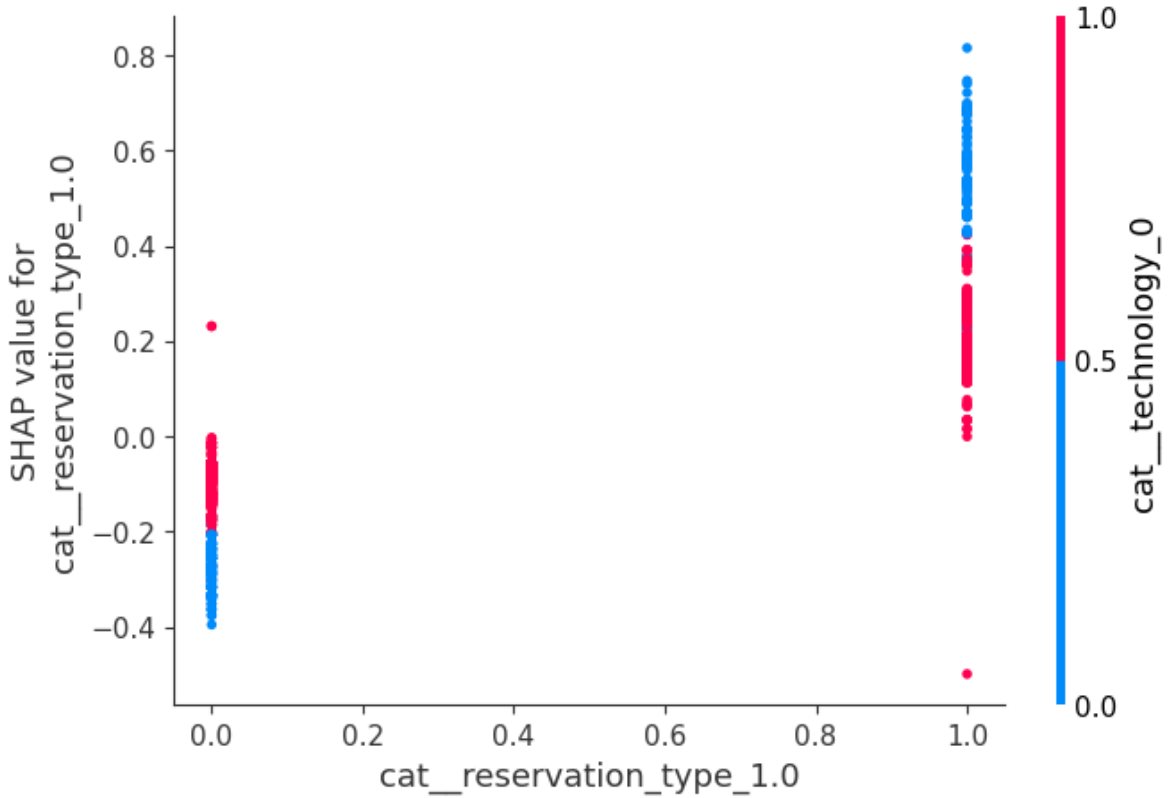


Figure: SHAP PDP Rank 4

SHAP PDP of GBR

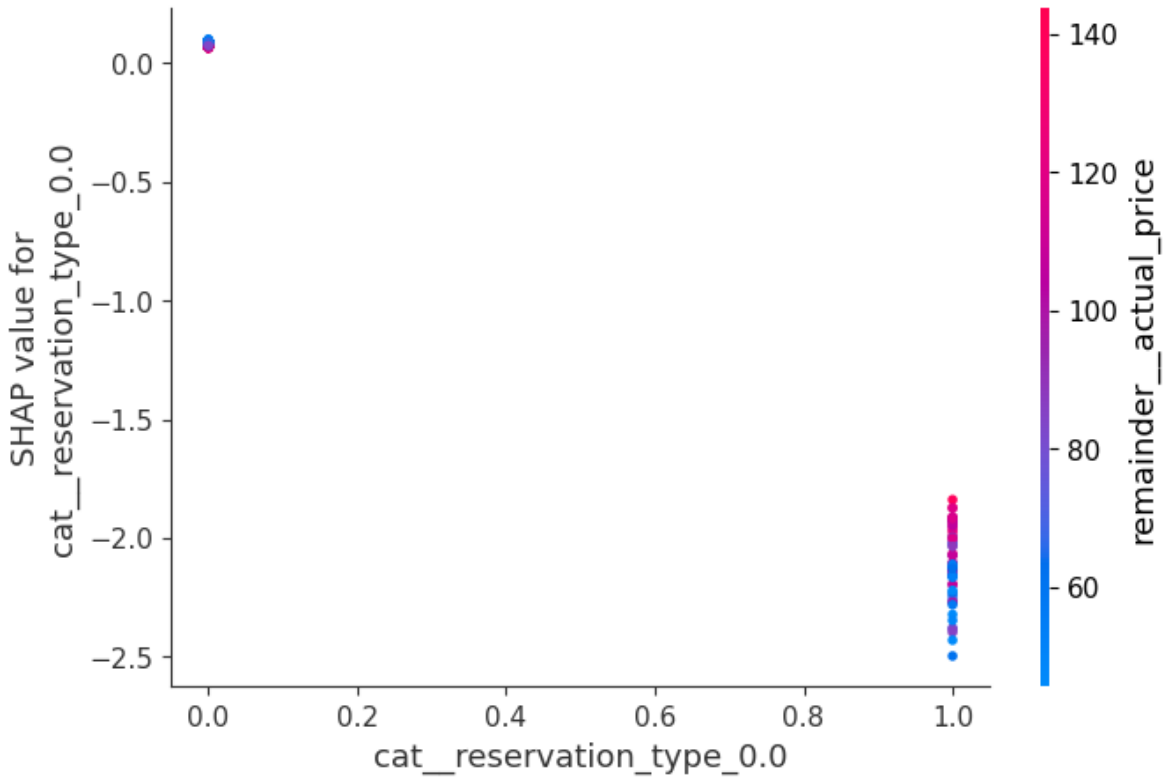


Figure: SHAP PDP Rank 5