

Company X Project

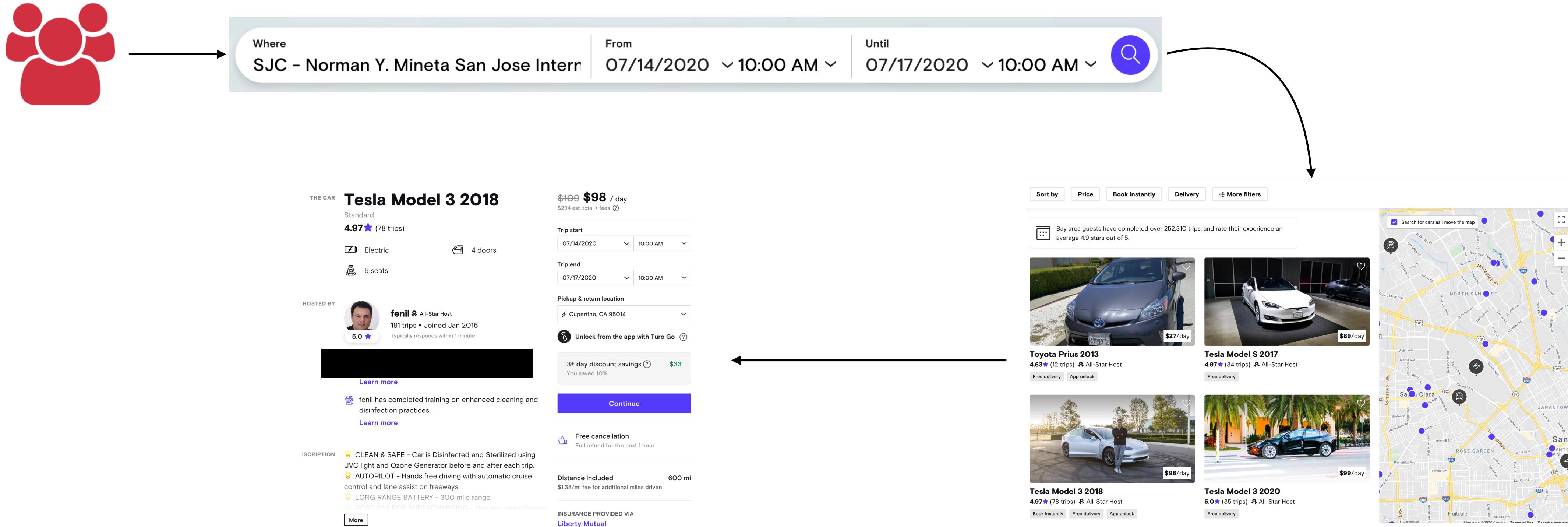
Ankur Manikandan

Objective

- Using the two attached datasets, vehicles and reservations, determine:
 - 1. Which factors seem to be most important in driving total # of reservations for our vehicles? (use a model to drive your insights)**
 - 2. How does the car's technology type seem to affect reservations?**
- But what is the purpose of the above two questions? How does it affect the business? To understand this, we first need to understand the user journey a guest has to follow to complete the reservation process.

User journey

Step 1: Guest provides location along with from and to date, time of reservation



Step 3: Guest selects vehicle of choice and completes the payment and booking process

Step 2: Guest is provided with a list of vehicles to choose from

How does identifying the important factors help?

- In Step 2 of the user journey, the guest is provided with a list of vehicles to choose from.
- This list of vehicles is most likely generated by a search rank engine.
- Ideally, we want to provide a list of vehicles to the guest that would increase the chances of the guest to complete the reservation process.
- But how do we identify which vehicles to list?
- In addition to the inputs provided by the guest in Step 1, we have a database of vehicle attributes. Moreover, Company X has data on historical reservations.
- Therefore, we can list vehicles based on the factors that increase the chances of the guest completing the reservation process.

So what if we are increasing the chances of the guest to complete the reservation process?

- On every reservation, Company X generates revenue by charging,
 - trip fee (variable percentage of trip price),
 - protection fee (based on protection plan chosen by guest), and
 - young driver fee (ages 18-25).
- Therefore, higher the number of reservations, greater the revenue generated by Company X.

Trip fee

This fee is a variable percentage of the Trip price. We show you the exact amount at checkout before you submit your request. Guests in the US pay a minimum of \$15 per trip. This fee covers the costs of our 24/7 customer support, platform security and upgrades, and vendor fees.

Protection

This is the cost of the [protection plan](#) you choose. Plan availability vary by country. Each plan costs a different amount and offers different levels of coverage. If you book a car with a [Commercial Host](#), they'll

Young Driver Fee

We charge a [Young driver fee](#) to guests ages 18-25 who are booking a vehicle in the US. Outside of the US, we charge this fee to drivers who are ages 18-24. The fee for US drivers ages 18-20 is \$50*/day or higher, depending on the vehicle. For US drivers ages 21-25 the fee is \$30*/day or higher, depending on the vehicle. In Canada, the fee is \$15*/day or higher, depending on your age and the vehicle you choose. In the UK, the fee is \$25*/day or higher depending on your age and the vehicle you choose.

How do we maximize the total number of reservations?

- In the reservations dataset, we have three reservation types: hourly rentals, daily rentals and weekly rentals.
- Identifying the factors that positively influences the number of reservations for each reservation type will help with maximizing the total number of reservations of Company X's vehicles.
- We have the vehicle attributes and we know how many times each vehicle has been rented hourly, daily and weekly.
- What do we have here, a regression problem!

Set-up of the regression problem

- Let us begin with a simple model to set a baseline, that is, linear regression.
- The predictor variables are: technology, actual_price, recommended_price, num_images, street_parked, description.
- The target variable: number of reservations of each reservation type corresponding to a vehicle.

technology	actual_price	recommended_price	num_images	street_parked	description	reservation_type_1	
0	1	67.85	59	5	0	7	1.0
1	0	100.70	53	5	0	224	4.0
2	0	74.00	74	4	1	21	1.0
3	0	135.00	75	1	0	184	1.0
5	0	82.15	53	5	1	44	4.0

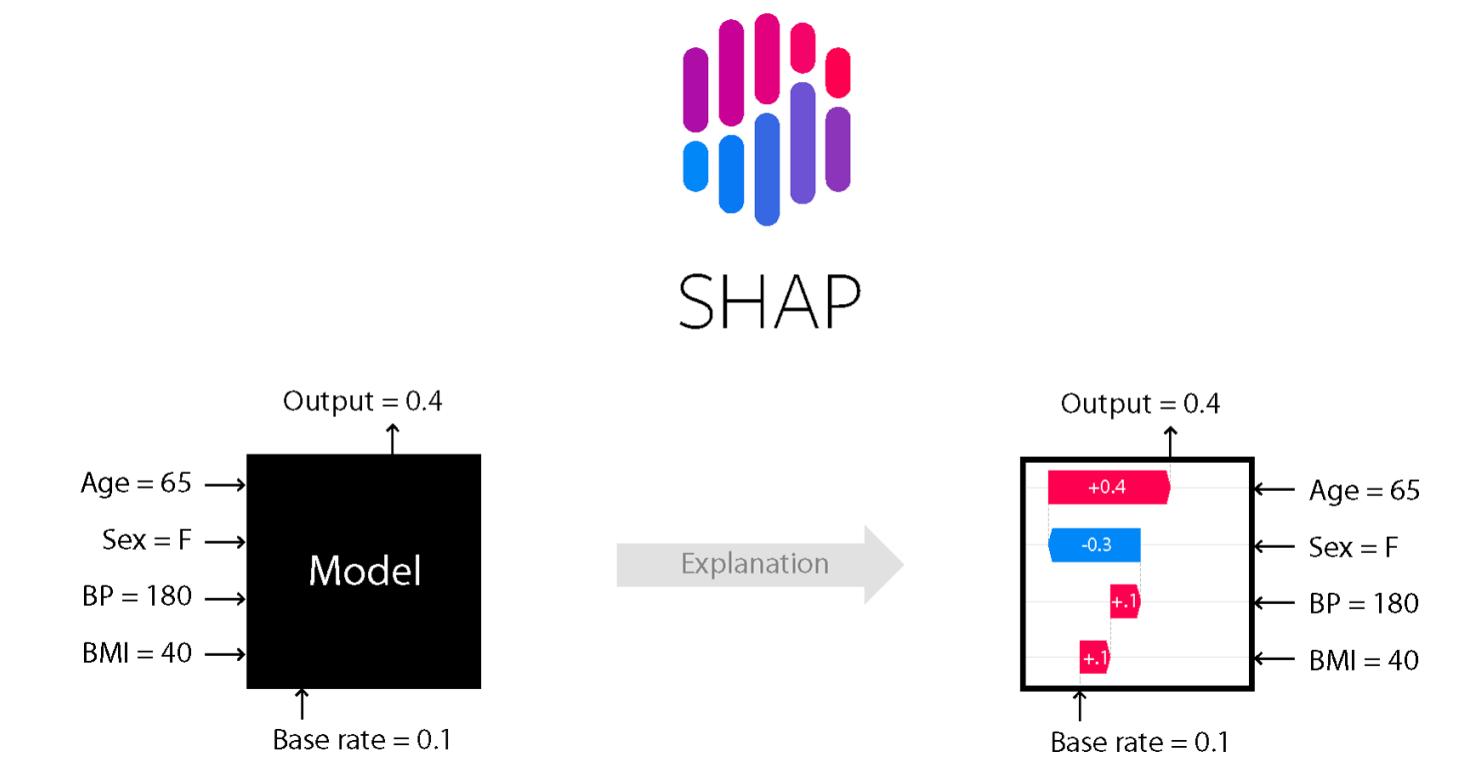
Sample dataset for regression

But, how do we identify the most important factors?

- We turn to SHAP (SHapley Additive exPlanations)!

- **What is SHAP??**

SHAP^[1] is a tool that leverages coalitional game theory concepts to help explain the feature contributions of any machine learning model.

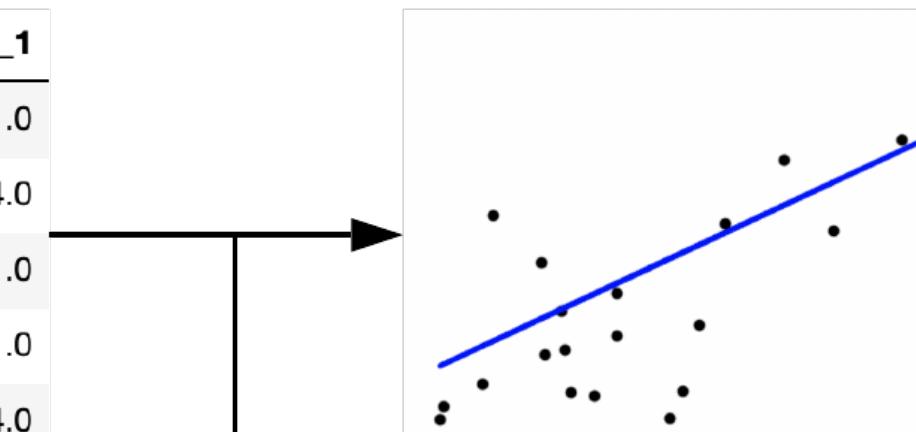


^[1] A Unified Approach to Interpreting Model Predictions, Scott M. Lundberg, Su-In Lee, Advances in Neural Information Processing Systems 30 (NeurIPS 2017)

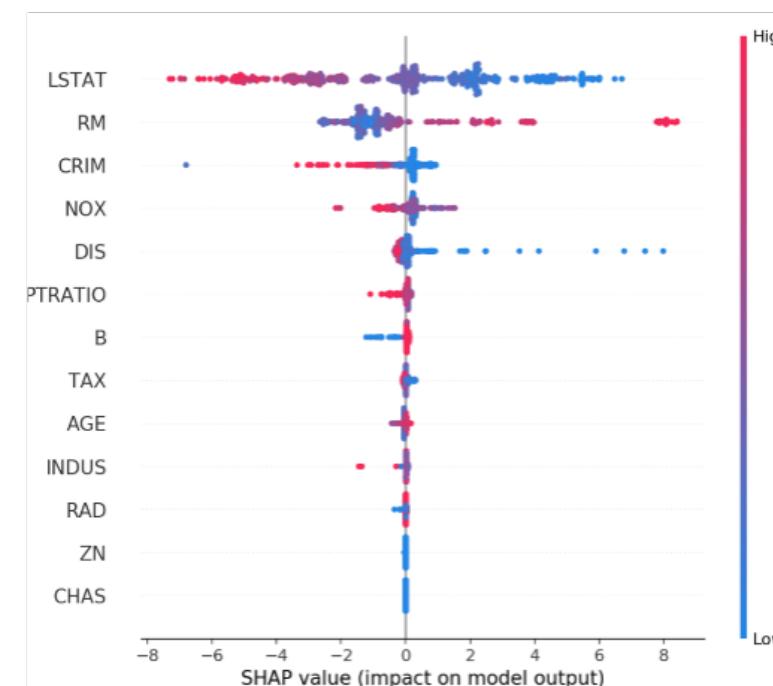
End-to-end process flow

Data to train the linear regression model

	technology	actual_price	recommended_price	num_images	street_parked	description	reservation_type_1	
0	1	67.85		59	5	0	7	1.0
1	0	100.70		53	5	0	224	4.0
2	0	74.00		74	4	1	21	1.0
3	0	135.00		75	1	0	184	1.0
5	0	82.15		53	5	1	44	4.0



Train linear regression model



generate explanations

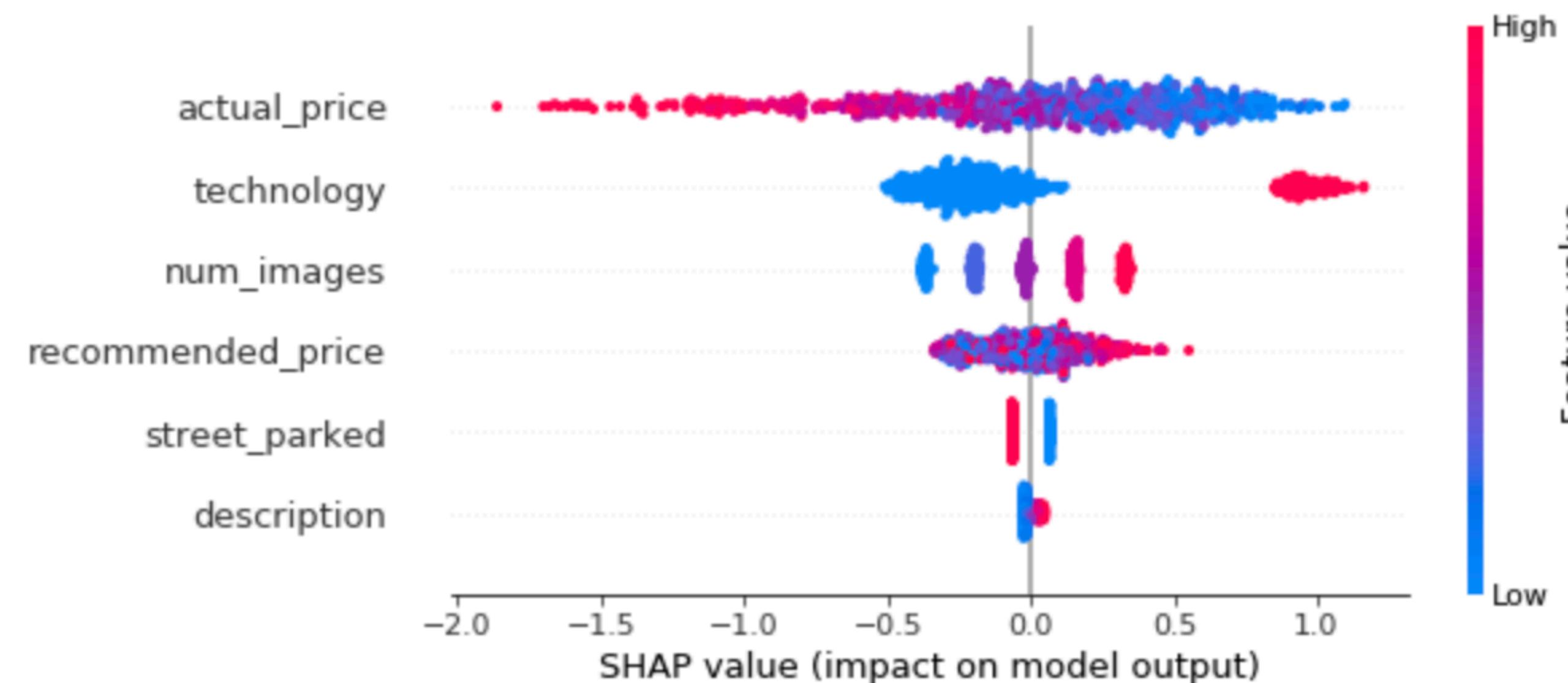


Provide SHAP with the trained model and the data

Explanations

Which factors seem to be most important in driving total # of reservations for our vehicles?

Most important factors: Hourly reservations



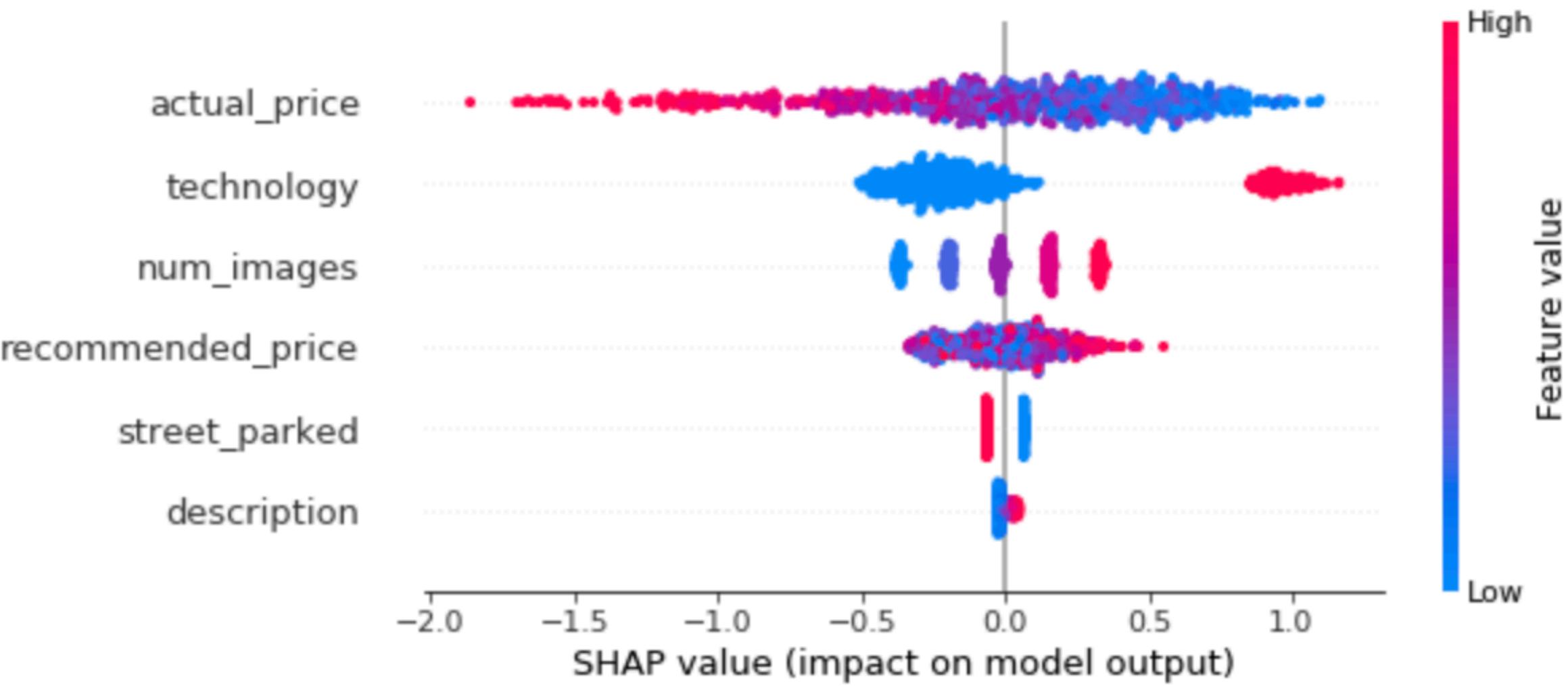
How do we read the above plot?

- We have a ranked list of features on the right-hand side of the plot.
- The dots on the plot represent each data point in our dataset (for hourly reservations).
- The color represents the magnitude of the feature value.
- At the bottom of the plot is the range of SHAP values. It is the effect that each feature value of each data point has on the model's output.

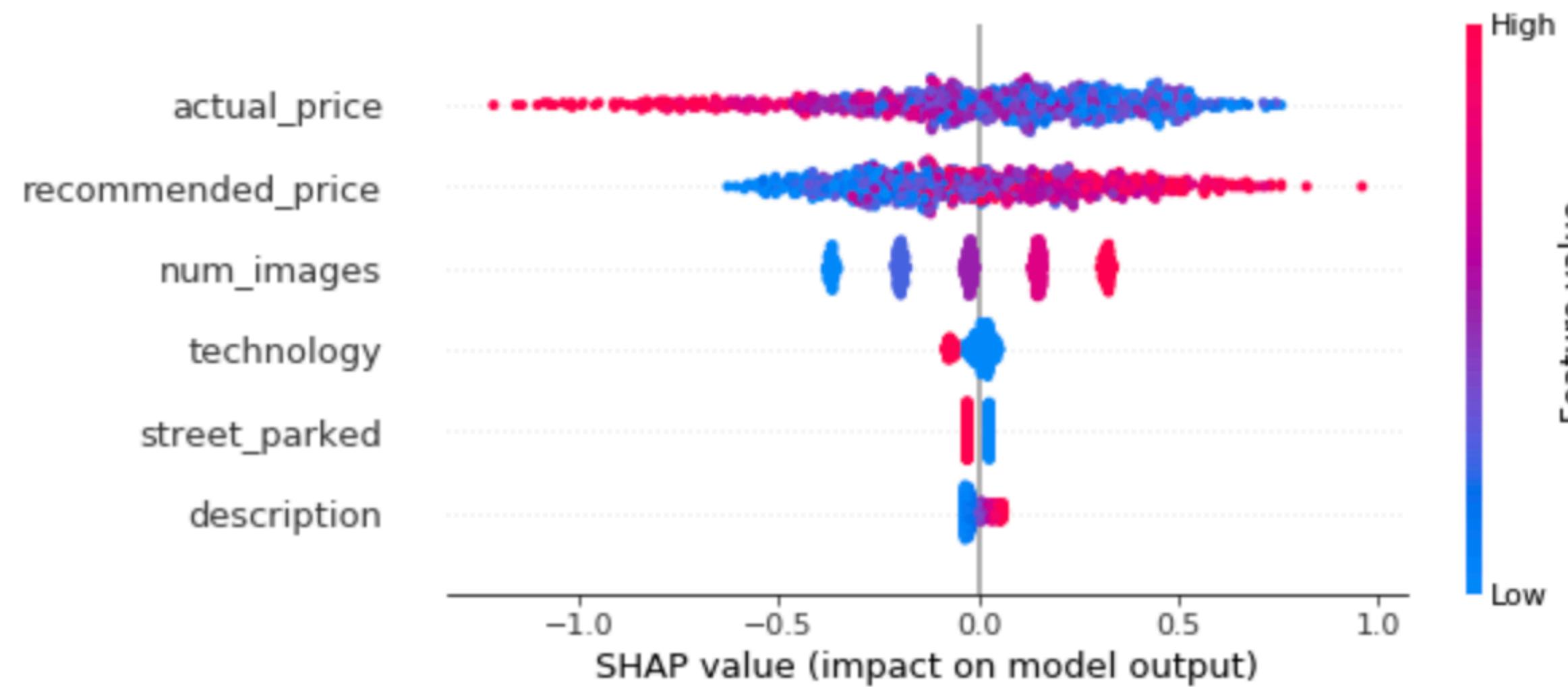
Most important factors: Hourly reservations contd.

The ranked (from most important to least important) list of factors are as follows:

1. **actual_price**: inversely proportional to number of hourly reservations
2. **technology**: presence of technology in the vehicle is proportional to number of hourly reservations
3. **num_images**: presence of 4 or 5 images is proportional to number of hourly reservations.
4. **recommended_price**: proportional to number of hourly reservations.
5. **street_parked**: if the vehicle does not have to be street parked it increases (marginal contributions) the number of hourly reservations.
6. **description**: if a host uses more characters to describe the vehicle it increases (marginal contributions) the number of hourly reservations.



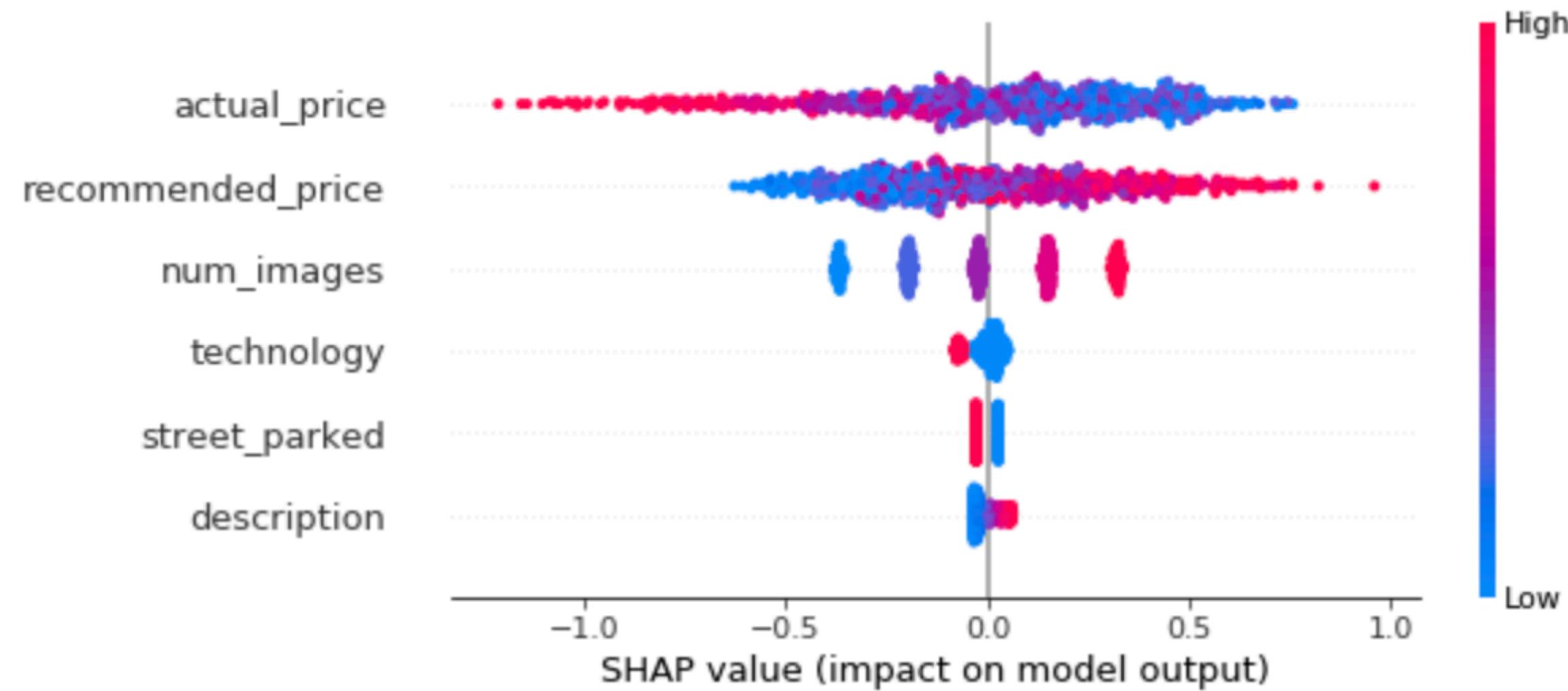
Most important factors: Daily reservations



The ranked (from most important to least important) list of factors are as follows:

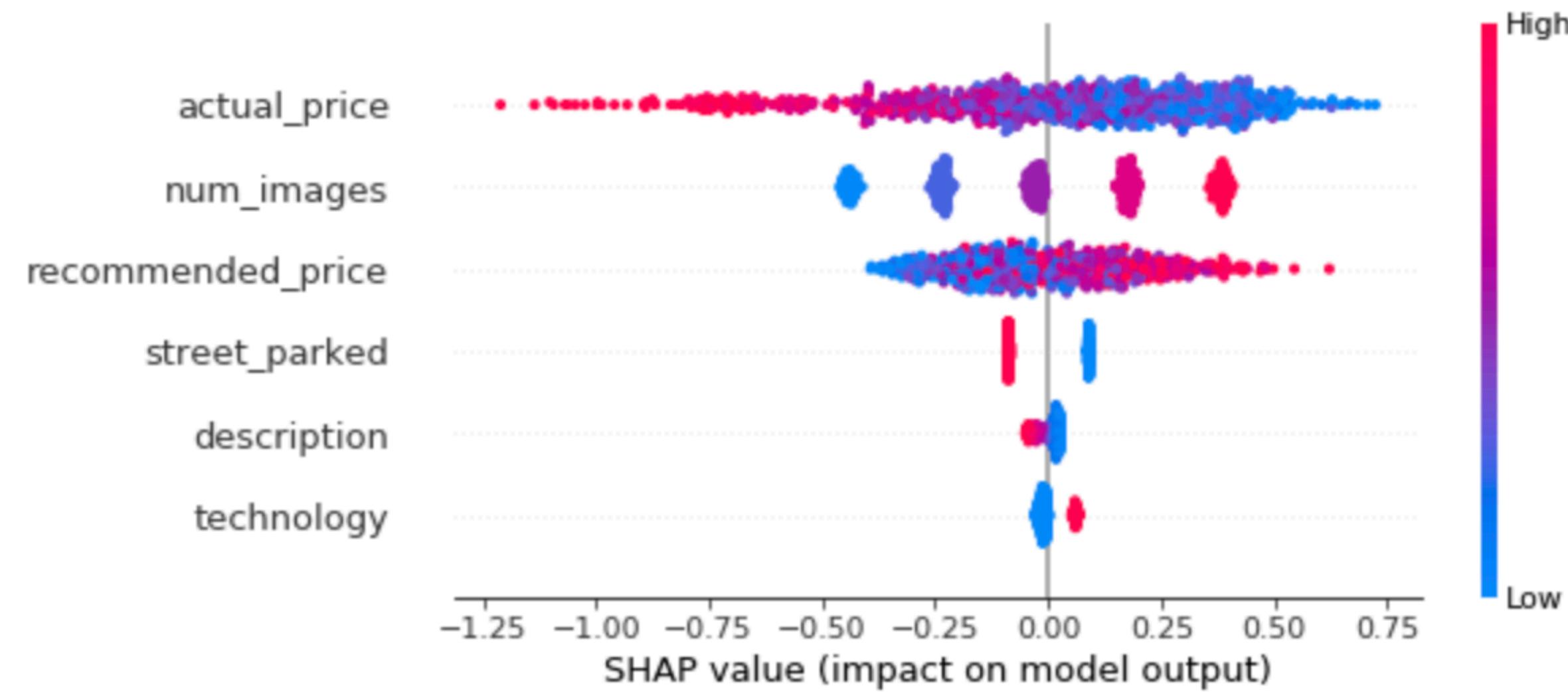
1. **actual_price**: inversely proportional to number of daily reservations
2. **recommended_price**: proportional to number of daily reservations.
3. **num_images**: presence of 4 or 5 images is proportional to number of daily reservations. 5 images has the highest impact.

Most important factors: Daily reservations contd.



4. **technology**: presence of technology in the vehicle is inversely proportional to number of daily reservations.
5. **street_parked**: if the vehicle does not have to be street parked it increases (marginal contributions) the number of daily reservations.
6. **description**: if a host uses more characters to describe the vehicle it increases (marginal contributions) the number of daily reservations.

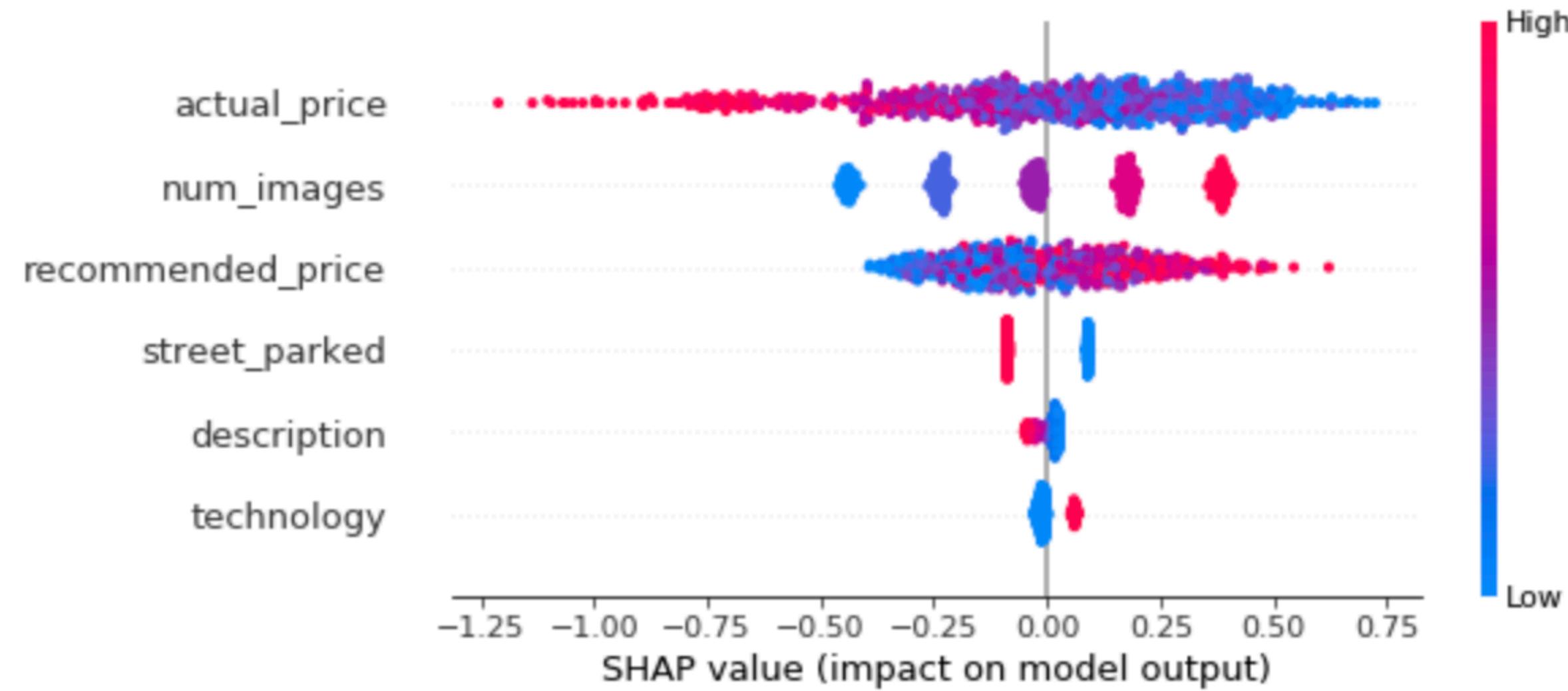
Most important factors: Weekly reservations



The ranked (from most important to least important) list of factors are as follows:

1. **actual_price**: inversely proportional to number of weekly reservations. The effect of actual daily price on weekly reservations is lower than the effect of the same on hourly and daily reservations.
2. **num_images**: presence of 4 or 5 images is proportional to number of weekly reservations. 5 images has the highest impact.
3. **recommended_price**: proportional to number of weekly reservations.

Most important factors: Weekly reservations contd.

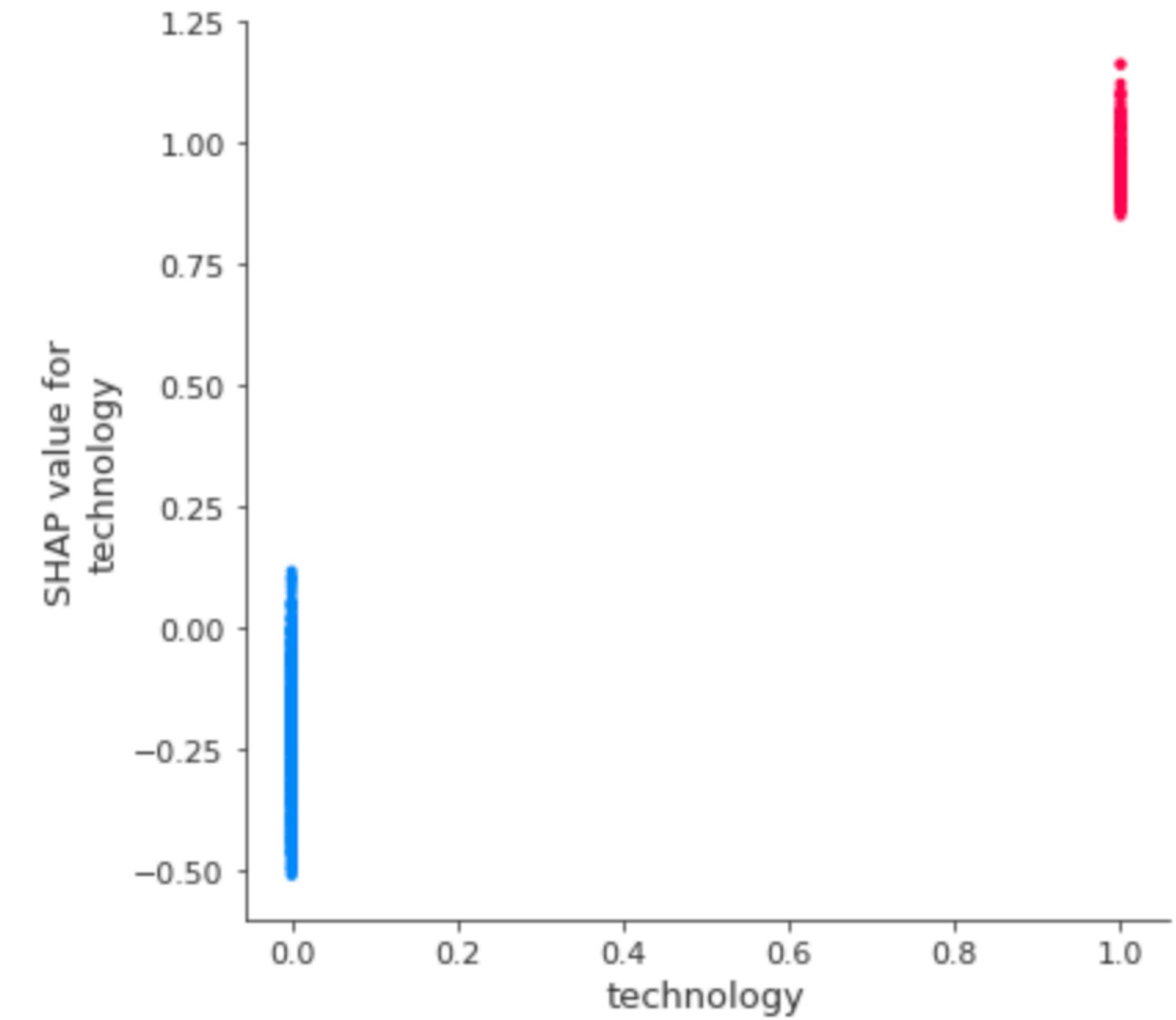


4. **street_parked**: if the vehicle does not have to be street parked it increases (marginal contributions) the number of weekly reservations.
5. **description**: if a host uses more characters to describe the vehicle it decreases (marginal contributions) the number of weekly reservations.
6. **technology**: presence of technology in the vehicle is proportional to number of weekly reservations.

How does the car's technology type seem to affect reservations?

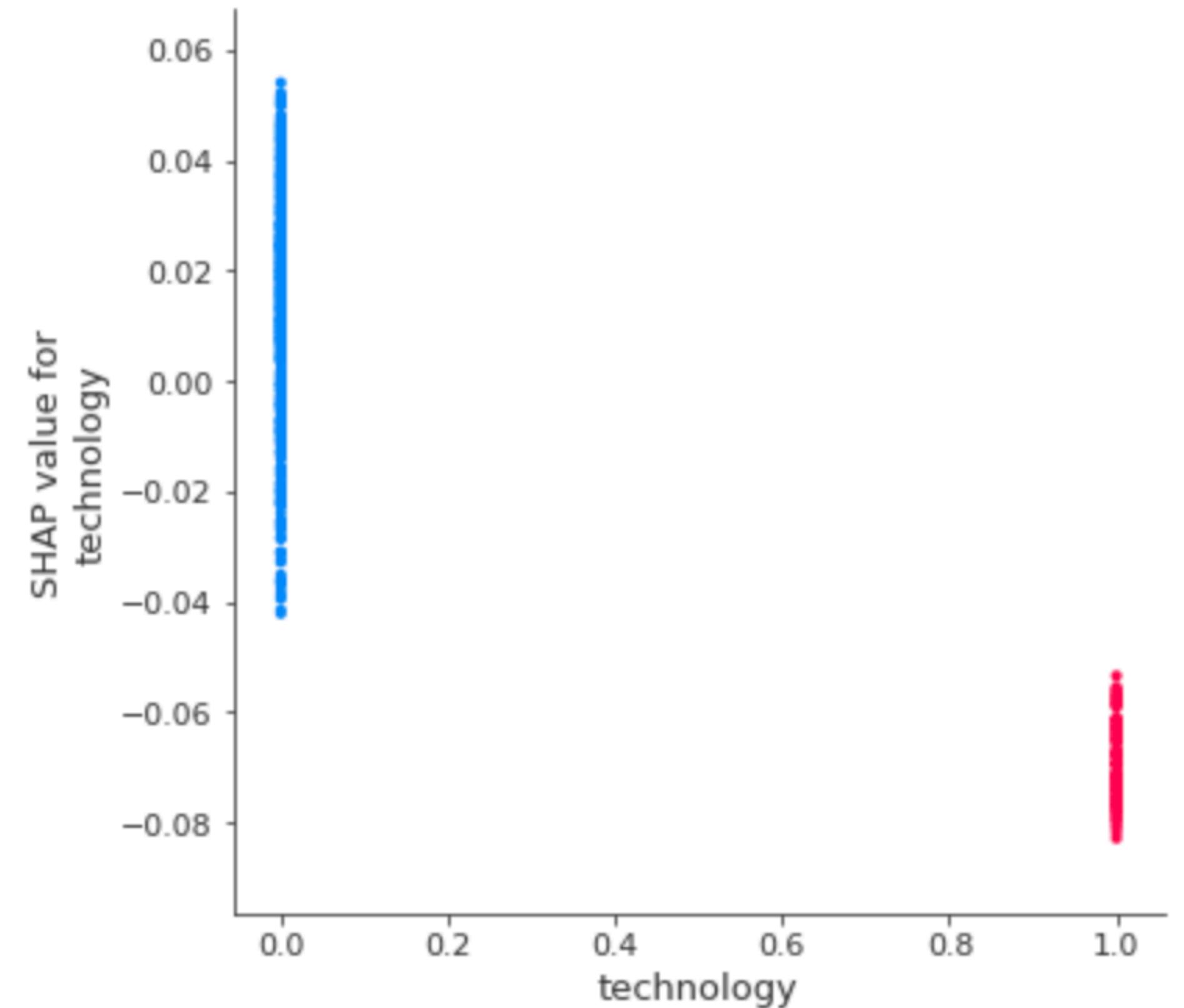
Company X Go's affect on hourly reservations

- Presence of technology in the vehicle is proportional to number of hourly reservations. The effect of technology on the number of hourly vehicles is stronger than the effect of technology on daily and weekly reservations.
- **Why this affect on hourly reservations? My assumption.** More than the presence of technology in the vehicle, I think there is another factor, most likely location, that leads to the inversely proportional relationship between technology and hourly reservations. Maybe the location of vehicles with Company X Go is primarily at airports. Company X Go is currently available in LA, SF, San Diego and London. In these locations, the airports are some distance away from majority of the residential localities. Guests who require hourly rentals would prefer to have quick access to a vehicle, preferably closer to their homes or offices.



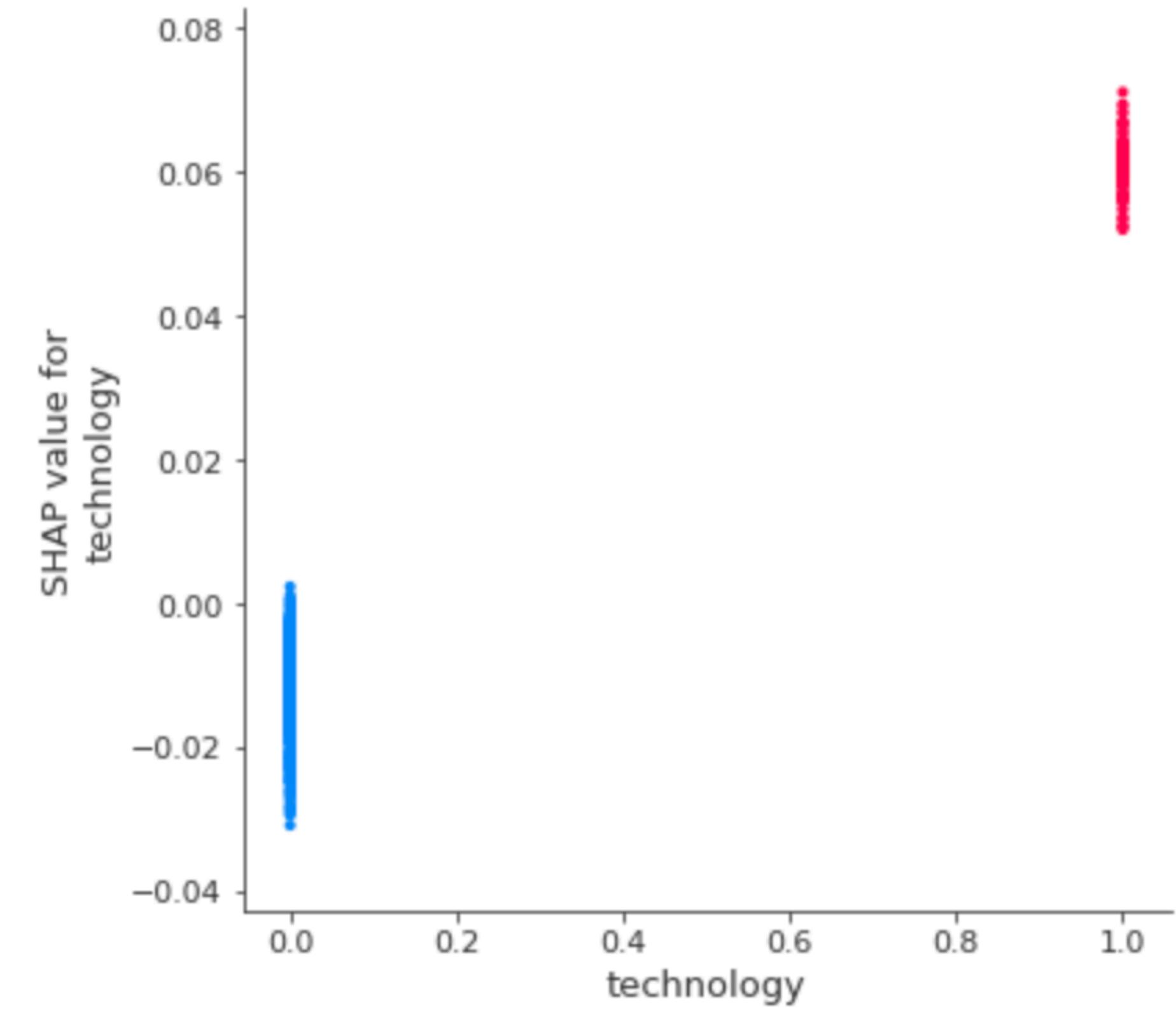
Company X Go's affect on daily reservations

- As presence of technology in the vehicle is inversely proportional to number of daily reservations. The effect of technology on the number of daily reservations is lesser than the effect of technology on hourly reservations but, greater than the effect of technology on weekly reservations.
- **Why this affect on daily reservations? My assumption.**
Similar to hourly reservations, I think there is another factor, most likely location, that leads to the inversely proportional relationship between technology and daily reservations. Maybe the location of vehicles with Company X Go is primarily at airports. Company X Go is currently available in LA, SF, San Diego and London. In these locations, the airports are some distance away from majority of the residential localities.
Guests who require hourly rentals would prefer to have quick access to a vehicle, preferably closer to their homes or offices.



Company X Go's affect on weekly reservations

- Presence of technology in the vehicle is proportional to number of weekly reservations. The effect of technology on the number of weekly reservations is the least, relative to the effect of technology on hourly and daily reservations.
- **My assumption on why technology is proportional to weekly reservations**, unlike in hourly and daily reservations, is as follows - weekly rentals are primarily used by guests who travel. If majority of these guests land at airports they would prefer using a vehicle that has Company X Go. This allows the guest to instantly book the vehicle and pick it up without having to meet with the host.
- **My next assumption on why technology has the weakest effect on weekly reservations is**, when a guest is looking to reserve a vehicle for a week, price and images are going to be the most important factors. As we have seen in the SHAP summary plot for weekly reservations, the remaining factors, technology, street parking and description have weak effects on the number of weekly reservations. These three factors have similar ranges of SHAP values.



Next Steps

If I had more time to work on the assignment, my next steps would be:

1. Investigate the SHAP dependence plots between the factors for hourly, daily and weekly reservations.
2. Engineer a few features, like - ratio of actual price to recommended price and ratio of description to images and study their effects on the total number of reservations.
3. Identify range of values for each factor that can help maximize the hourly, daily and weekly reservations.
4. Request new factors, such as - vehicle pickup location, type of vehicle, number of seats, number of passengers associated to reservation, pickup time, and drop-off time.

Thank you!