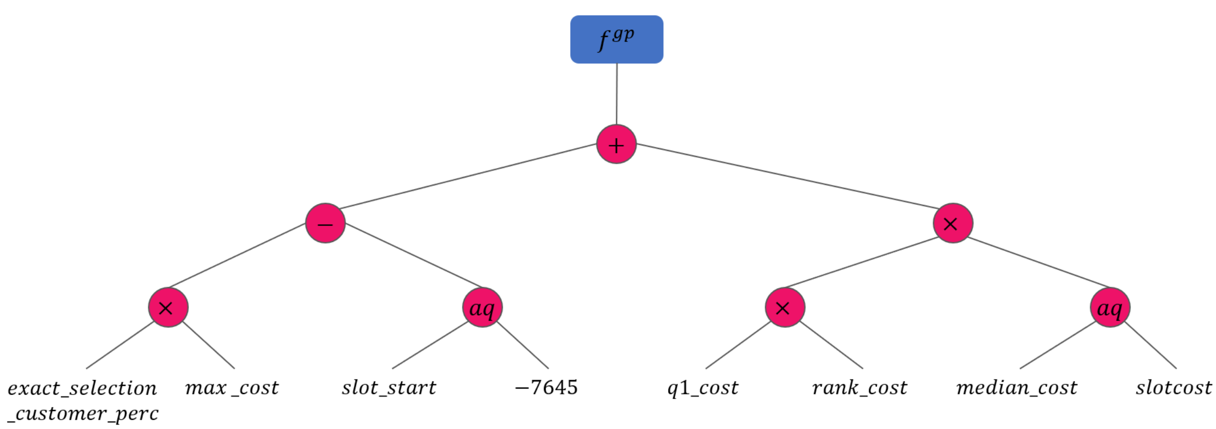
The aim of this document is to collect biases and constraints for better explanation generation and evaluation. Diabetes dataset is used as an illustrative example of what kind of metadata we are looking for. It is ok to leave some parts empty. We are looking forward to getting a causal graph based on your field knowledge, but it is ok if the first versions of it are not precise.
Diabetes example as a reference can be found here:
https://docs.google.com/document/d/1eSGbvntG0Thy6vIkw2-qcxmQXUOCTAHlu5DUp7T3yLA/edit?usp=sharing
Our presentation in Porto: :
https://docs.google.com/presentation/d/1mmJQx_mab9wAApbyezfdxrcmLiOIuvaI/edit?usp=sharing&ouid=103661675820702400870&rtpof=true&sd=true

Data description

For now we can focus on the WTP problem, since the data for the CTS is not final. But if you're ready to share information for both, you are welcome.
Data shape (number of instances x number of features): 3529 x 12
Feature list: slotcost, slot_start, exact_selection_customer_perc, rank_cost, median_cost, partial_selection_customer_perc, expanding_avg_days_to_delivery, days_since_first_purchase, q1_cost, max_cost, min_cost, slot_width

Output shape: 3529 x 1
Output values: 1, if the time slot is chosen by the customer; 0, otherwise.

Train/val/test
Observations: Train: 3529 samples / Test: 52072 samples (same as the one for the GBM)
Score (log-loss): Train: 0.408 / Test: 0.287 / On a balanced Test Dataset: 0.582



Feature meta information

Feel free to share this kind of data in any form or format, the table is a proposal.
*Actionable means if counterfactual explanation for this feature makes sense

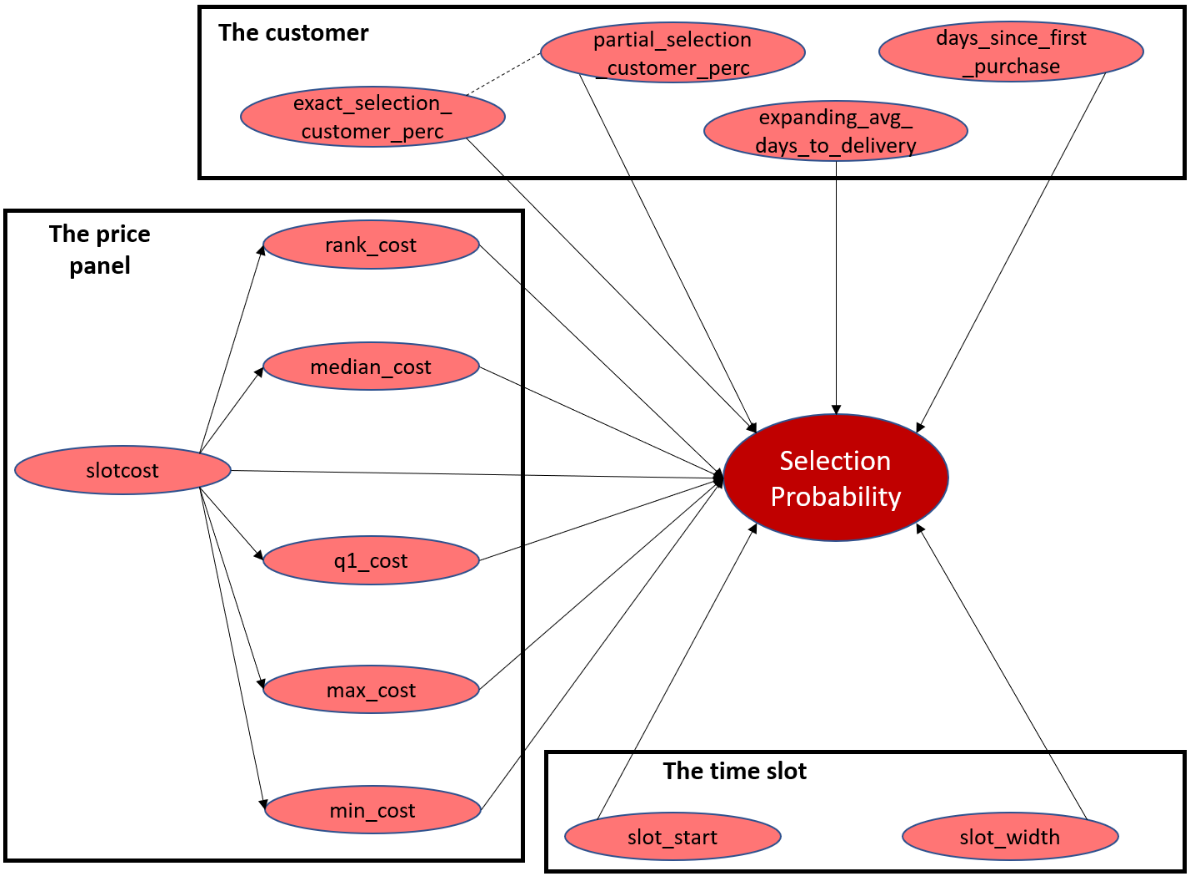
Feature_name	Meaning	Type	Actionable*	Range	Mean, std	Constraints
slotcost	The displayed time slot price	cont.	yes	[4.59; 10.18]	(6.21, 0.59)	Decision-maker could establish that min_price <= slotcost <= max_price
slot_start	Number of minutes since the order instant until the opening time of the slot	cont.	yes	[840; 22800]	(9873.17, 6149.87)	slot_start >= 0
exact_selection_customer_perc	The historical percentage of times that the customer selected a given time slot	cont.	yes	[0; 1]	(0.065, 0.142)	0 <= exact_selection_customer_perc <= 1
rank_cost	Considering all time slots presented to the customer, gives the percentile in which the time slot lies with respect to price, e.g., rank_cost = 12% means that the time slot is in the 12% cheapest slots offered	cont.	no (consequence of slotcost and remaining time slot offers)	[0; 1]	(0.529, 0.248)	0 <= rank_cost <= 1
median_cost	The median price of all time slots shown to the customer	cont.	no (consequence of slotcost and remaining time slot offers)	[5.75; 7.38]	(6.51, 0.28)	
partial_selection_customer_perc	The historical percentage of times that the customer selected a time slot that intersects the time slot under analysis with respect to time, e.g., If the customer had selected time slot ranging from 1pm to 3pm and the time slot under analysis was from 2pm to 4pm, the previous case would be taken into account for the percentage.	cont.	yes	[0; 1]	(0.116, 0.176)	0 <= partial_selection_customer_perc <= 1
expanding_avg_days_to_delivery	The average number of days between the moment of the order and the start of the time slot for previous time slot selections	cont.	yes	[0; 6]	(1.46, 0.64)	Expanding_avg_days_to_delivery >= 0 (and expanding_avg_days_to_delivery <= 6 for our model, since we discarded time slots 7 days ahead of the moment of the order)
days_since_first_purchase	The number of days since the customer last made a purchase using the retailer attended home delivery service	cont.	yes	[1; 363]	(162.99, 100.31)	Days_since_first_purchase >= 0
q1_cost	Considering the price distribution of the time slots displayed to the customer, this feature provides the first quartile.	cont.	no	[5.41; 6.90]	(5.99, 0.31)	

max_cost	The highest price among the prices displayed to the customer	cont.	yes	[6.31; 13.46]	(7.37, 0.94)	When varying max_cost, the price distribution of the time slot price panel presented to the customer will change. To test variations in max_cost, price features need to be updated
min_cost	The lowest price among the prices displayed to the customer	cont.	yes	[4.42; 6.32]	(5.28, 0.30)	When varying min_cost, the price distribution of the time slot price panel presented to the customer will change. To test variations in min_cost, price features need to be updated
slot_width	The range of a time slot in minutes, e.g., a slot that starts at 1pm and ends at 3pm as a slot_width of 2h or 120 min.	cont	yes	[120; 450]	(147.22, 33.57)	Decision-maker could establish that min_width <= slot_width <= max_width

Causal graph

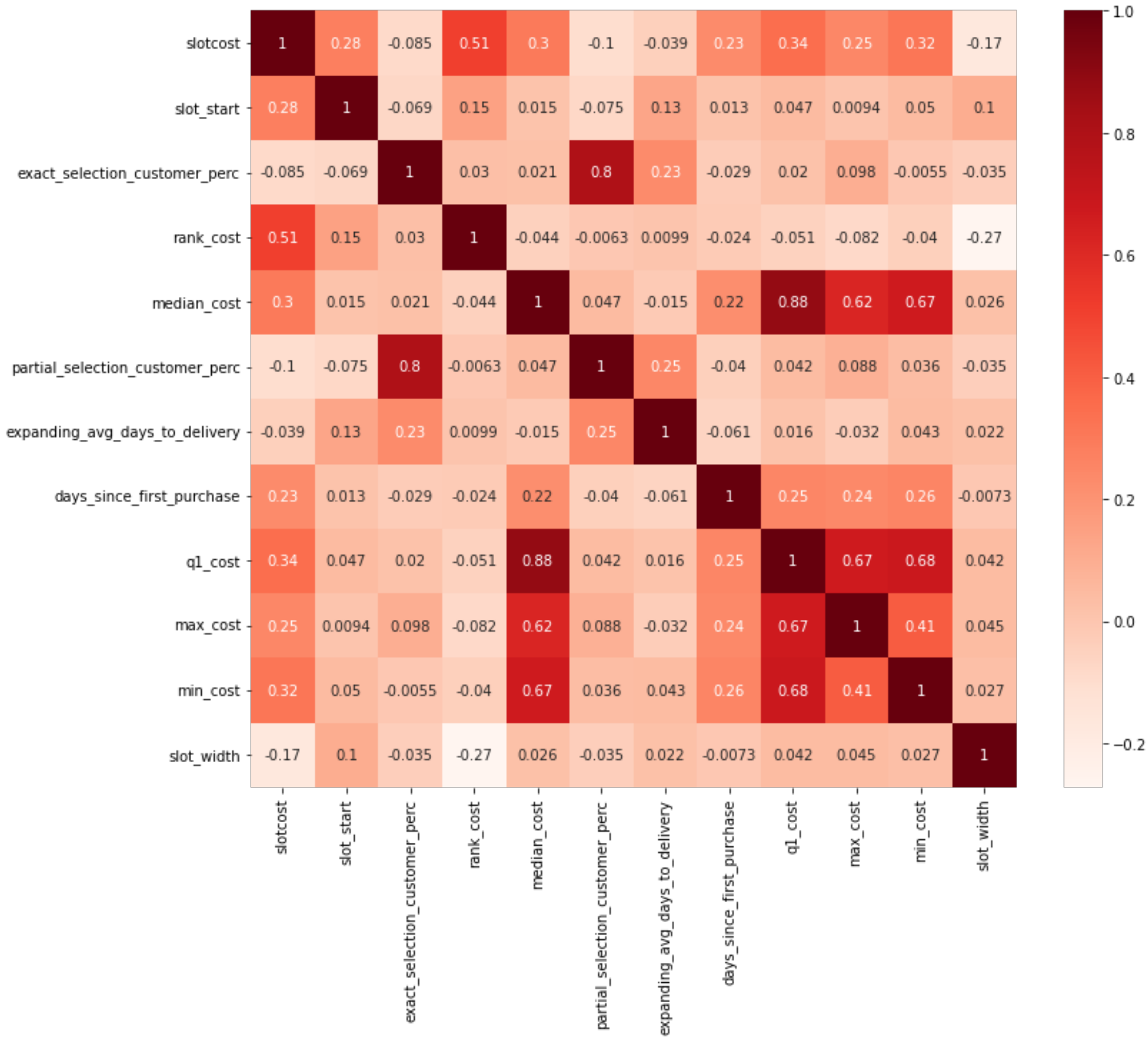
Please draw a causal graph for your use case. The causal graph should be approximate at this stage. You can comment about some relations if you find some more or less trustworthy.

a. A short explanation and two examples of causal graphs: https://docs.google.com/document/d/1oinBe8nKoT3XoKdCWz5qxOq1hbPRQA67RajFZBBGycU/edit?usp=sharing



The customer
partial_selection
_customer_perc
days_since_first
_purchase
exact_selection_
customer_perc
expanding_avg_
days_to_delivery
The price
panel
rank_cost
median_cost
slotcost
Selection
Probability
q1_cost
max_cost
min_cost
The time slot
slot_start
slot_width

Feature correlation



Rules

Can you think of any rules in your data, is there common knowledge that one feature increase leads to another feature increase?

In the literature (paper) on customer preferences for attended home delivery services, the following three features are commonly known to be determining:

- Lead Time: customers tend to prefer slots that are closer in time (where the difference between slot start and the order time is lower)
- Availability: when the number of available options to choose from, customers are willing to pay more

- Range: Customers prefer narrower slots as these allow them to know more accurately the moment of delivery, e.g., choosing a time slot ranging from 1pm to 2pm provides a more accurate estimate of when I need to be home to receive the order, while a time slot lasting for the whole afternoon does not.

Regarding price, generally, when the price of a time slot is increased its selection probability decreases. However, there could be an encapsulated effect due to the retailer's current pricing policies. For those slots that are more popular, the retailer introduces mark-ups to collect higher revenues. Therefore, in the data, we could find the pattern that time slots with a higher price actually tend to be more preferred.

We treat the problem as a classification problem where we want to answer the following question:

- "Given a combination of customer and time slot, will the customer select the slot? And with which probability?

However, in the context of solving the retail use case, we want a model capable of determining such probability while taking into account the characteristics of competing time slots. Therefore, we introduce several statistical measures to characterize the price distribution of the offer presented to the customer. Therefore, we know that features rank_cost, median_cost, q1_cost, max_cost and min_cost all depend on slotcost.

For problem features:

↑ slotcost → ↓ selection probability

↑ slot_start → ↓ selection probability

↑ exact_selection_customer_perc → ↑ selection probability

↑ slotcost → ↑ rank_cost (potentially)

↑ rank_cost → ↓ selection probability

↑ slotcost → ↑ median_cost (potentially)

↑ median_cost → ↑ selection probability (assuming the slot under analysis remains with the same price)

↑ partial_selection_customer_perc → ↑ selection probability

↑ expanding_avg_days_to_delivery (could indicate that customer does not value service speed and is more price sensitive)

↑ days_since_first_purchase (could indicate customer generally presents higher time slot choice probabilities as he/she is loyal to the retailer and its service)

↑ slotcost → ↑ q1_cost (potentially)

↑ q1_cost → ↑ selection probability (assuming the slot under analysis remains with the same price)

↑ slotcost → ↑ max_cost (potentially)

↑ max_cost → ↑ selection probability (assuming the slot under analysis remains with the same price)

↑ slotcost → ↑ min_cost (potentially)

↑ min_cost → ↑ selection probability (assuming the slot under analysis remains with the same price)

↓ slot_width → ↑ selection probability

Applicable Human biases

Do you think that some biases are more applicable to your use-case than others?

- **Overgeneralization -** after seeing one explanation generalize over other similar cases (X)
  - There could be the case that aforementioned rules are accepted blindly. Higher slot prices have a negative impact on customer selection but they could also increase the customer's perception of quality.
- **Backgrounding -** pushing possible cause to background because it is not relevant anymore
- **Abnormality -** tendency to explain using abnormal causes
- **Recency -** more recent events have more weights in explanations (X)
  - I do not think that this observation matches completely the meaning

behind this bias type, but I think that a relevant aspect of this problem is model drift. In case the retailer decides to put our approach into action, prices will start to be presented based on our selection probabilities estimates. Therefore, the environment in which the model is deployed will be changed. After some time, the customer could potentially understand when to book an order to catch cheaper slots and, therefore, customer behavior would change dramatically.