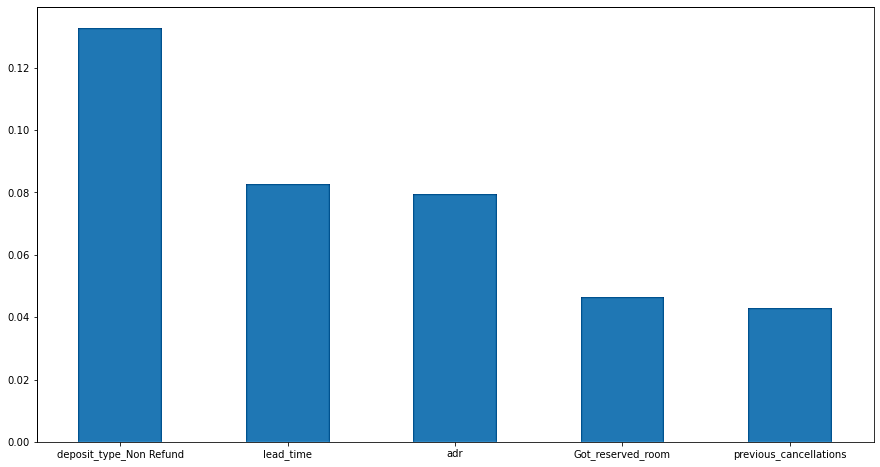
**Feature Importance.**



Feature selection allows for the identification of the most important or influencing factors on the response (or dependent) variable. In this example, feature selection techniques are used to predict the most important influencing factors on whether a customer chooses to cancel their hotel booking or not.

**Feature Selection Tools:**

The feature selection tool used to analyse this dataset is:

* **Feature selection based on Information Gain:** Information gain calculates the reduction in entropy or surprise from transforming a dataset in some way. Information gain can also be used for feature selection, by evaluating the gain of each variable in the context of the target variable. In this slightly different usage, the calculation is referred to as mutual information between the two random variables.

**Important Features:**

1. **“deposit\_type\_Non\_Refund”:** This variable is affecting the hotel booking severely. It means when the deposited amount is not refunded to the customer then there is a high chance that the customer will cancel the booking. This specifies that customers are very keen on their refunds.
2. **“lead\_time”:** The lead time specifies the number of days that passed between the entering date of the booking into the Property Management System (PMS) and the arrival date. There is a direct relationship between the lead time and the booking cancellation. If the lead time is more then there are always high chances that the booking will be cancelled.
3. **“adr”:** The ADR (Average Daily Rate) is defined by dividing the sum of all lodging transactions by the total number of staying nights. The values range from 0 to 5400.
4. **“Got\_reserved\_room”:** This variable specifies whether the customer got the room that he/she has reserved at the time of booking online. This variable plays an important part while predicting the hotel booking cancellation because if the customer is unable to secure the reserved room, then there is a good chance that the customer may develop an ambiguous notion about the hotel management system.
5. **“previous\_cancellation”:** The variable specifies the number of previous bookings that were cancelled by the customer before the current booking. This variable is important to carry out a further analysis because customer history always plays a crucial role in predicting the future of an organization. We’ll have to consider the frequency of the cancellation for every customer to categorize them into a particular segment.