How many Red coloured top Honda SUV cars do you think will be sold in India in 2022?

What factors do you think will impact sales?

- A Honda SUV approximately costs around 25 to 35 lakhs, which most of them can't afford. Only the higher middle class and Rich can afford a car worth 25 to 35 lakhs.
- Price of the SUV - Higher the price, lesser the individuals.

  28% of India's population is middle class -- of which 14% is lower middle class and about 3% is upper middle class. The top 1 per cent of Indians now own 33% of the country's wealth compared to 31.7 per cent previously. Hence, we observe comparatively 5% individuals will tend to afford an SUV worth . But SUV's, are bought more by travel agencies across the country.

- Brand
- Fuel Economy and Performance

  Travel agencies tend to buy more SUV's based upon the fuel economy and performance. Also, if individuals look to buy they would like to compare every brand's car performance.

- Dealership experience and competition

  A decade ago there were only a handful of car brands for buyers to choose from. Now the options are a lot more. Our cities have also grown bigger, and manufacturers now have more than one dealer in one city. As a result, competition between dealers has increased. Prospective buyers are aware of this and are going for brands which are making them feel their money's worth.

- Family requirements

  A family of 4 will not look to buy and SUV, if they are rich enough, they will look to buy a sedan from a better brands than SUV.

- After-sales and Service costs

  Honda has higher service costs more than other brands

- Resale value

  Considering all these factors, re-sale value will be decided.

Assume you have all the data you need, what statistical methodology or algorithm will you use to make this sales forecast?

- Associate rule Learning –

First up is association rule learning, also known as market basket analysis. This technique can uncover interesting relationships between variables in large data sets, and helps recognize hidden patterns that can be used to identify point-of-sales data. These patterns can be used for product or service recommendations based on what others have purchased before, as well as based on similar products/services bought together.

- Classification Analysis –

A systematic process of obtaining valuable and relevant information about data, as well as metadata, classification analysis plays a key role in identifying the proper categories data belongs to. Here, the data is classified into various sets in order to make an accurate analysis and/or forecast. From the sales point of view, classification is used for grouping and categorizing, where proper classification allows for a more precise description and qualification for companies that seek to leverage sales intelligence.

- Clustering Analysis –

Through clustering analysis, data is grouped in clusters based on its similarity to other data objects in some way. This technique allows companies to understand both the similarities and differences within the data. Sales-wise, the result of a clustering analysis is the buyer persona, a fictional representation of a targeted customer based on specific demographics, behavior, and other data. A practical example would be the report feature found in sales insights platforms, whereas a detailed overview of an account, a group, or an entire category can help in reducing customer attrition, as well as significantly improve customer retention

- Regression Analysis –

Also called value estimation because it revolves around estimating a value for a given variable, regression is used for predicting sales, prices, revenues, and other numerical values for a specific data set. There is a response variable (dependent) and one or more predictor variables (independent). Changing the value of predictor can alter the value of responsive variables (but not vice-versa, mind you). As such, it is often used as a forecasting technique. For instance, a sales forecast for sales per customer per month, based on this type of data mining can help set achievable, realistic goals and motivate salespeople to be more effective. By putting regression to work in order to estimate sales goals, sales leaders can provide their sales team with attainable targets that can drive their performance.

How would you evaluate your model or determine its accuracy?

To evaluate the model:

- How well is my model doing? Is it a useful model?
- Will training my model on more data improve its performance?
- Do I need to include more features?

The most important thing you can do to properly evaluate your model is to not train the model on the entire dataset.

It's important to use new data when evaluating our model to prevent the likelihood of overfitting to the training set. However, sometimes it's useful to evaluate our model as we're building it to find that best parameters of a model - but we can't use the test set for this evaluation or else we'll end up selecting the parameters that perform best on the test data but maybe not the parameters that generalize best. To evaluate the model while still building and tuning the model, we create a third subset of the data known as the validation set. A typical train/test/validation split would be to use 60% of the data for training, 20% of the data for validation, and 20% of the data for testing.

The three main metrics used to evaluate a classification model are accuracy, precision, and recall.

Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

Accuracy = correct_predictions/all_predictions

Precision is defined as the fraction of relevant examples (true positives) among all of the examples which were predicted to belong in a certain class.

Precision = true_positives/(true_positives+false_positives)

Recall is defined as the fraction of examples which were predicted to belong to a class with respect to all of the examples that truly belong in the class.

Recall = true_positives/(truepositives+falsenegatives)