**Case**

The URL link of the case given below is the case on which we would be working on.

[**https://www.sas.com/en\_us/customers/american-honda.html**](https://www.sas.com/en_us/customers/american-honda.html)

**Objective**

The objective of this project is to help American Honda’s dealerships and its subsidiary Acura’s dealership achieve the goals given below with the help analytics modelling.

The goals are

1. Identify the suspicious claims accurately to reduce warranty expense.
2. Have sufficient service technicians to perform repair procedure depending upon the volume of requests.
3. Forecast future needs of vehicle parts so that all the required vehicle parts are available when customer comes up with a service request.
4. Predict and prescribe future needs pertaining to Design, Engineering, Manufacturing etc. for the next version of vehicles based on customer feedback.

**Approach**

Let’s do a deep dive into each of the goals.

**GOAL 1**

**Help Identify suspicious** **claims accurately and reduce warranty expense.**

**SITUATION**

1. Honda has around 12000 dealerships nationwide that provide services to the vehicle owners in case of any issues in the vehicles. Different vehicle owners have enrolled for different warranties which in turn are eligible for different time frames.
2. Honda being a big company collects all the customer information while the vehicle is purchased and has a good data infrastructure (Datawarehouse) that stores all the relevant data pertaining to purchase, parts and the customers personal information within the permission given by the law.
3. Honda has a training program to cover all the warranty guidelines.

**COMPLICATION**

1. When a customer approaches to the dealership and claims his/her warranty to repair a vehicle part, there is a possibility of the claim being a false claim and not being eligible for warranty.
2. It takes a dealership examiner time to identify a false claim and yet it is very difficult be accurate. The examiner ends up identifying only 35% false claims.

**QUESTION**

1. Of all the warranty claims requested every month how many of them are false warranty claim and how many True.?

**SOLUTION (For each question)**

1. The solution for the question can be can be achieved using **Logistic Regression**. We can use last 5 years of data and divide it into Training, Testing data set. Use the training dataset to fit the model and then test it on the testing data set. Here the purpose of Logistic regression is to predict if the claim is a non-compliant claim or a true claim.
2. The company can collect all the new claims from its dealership throughout the country and apply the tested model on it and predict the validity of the claim and provide it back to the dealership

**Analytics Modelling Method**: Logistic Regression, Principle Component Analysis

**Purpose**: Predict if the claim is suspicious or not.

**Predictors:**

|  |  |  |
| --- | --- | --- |
| **Data Elements** | **Description** | **Examples** |
| Claim Narrative Category | Category of the warranty claim | Bumper to Bumper, Engine Warranty etc |
| Vehicle Build Year | The year the vehicle was built | 2019, 2010 |
| Vehicle in Service Month | The month the vehicle was serviced | January, Febuary etc |
| Vehicle in Service Week | The week of the month vehicle was serviced | 1st, 2nd , 3rd ,4th 5th |
| VIN | Vehicle Identification Number | 1H19370987732, 1H9237478879 |
| Customer Warranty Coverage | Coverage Period | 2 years, 3 years etc |
| Mileage Range | Rage of the mileage in bands | 0 - 25,000, 25,000 - 75,000 etc |
| Claim ID | The unique identification of the claim | 16523, 42324 |
| Make | Honda or Acura |  |
| Model | Model of the car | Civic, Accord, Crosstour etc |
| Failure Date | Date the failure occurred |  |
| Repair Date | Date the repair was requested |  |
| Failed Part | Part failed | Engine, Break Pads |
| Replacement parts | Parts that need replacement apart from pasts failed. |  |
| Dealership Location | Zipcode of the dealership | 28277, 75206 |

**Response:**

* Valid Claim (Yes or No)

To improve the accuracy of the model and select the correct variables we can us **Principle Component Analysis** method and chose the first few principle components.

**INFERENCE**

1. Dealership examiners will now save time in identifying the suspicious claims because they already have that info provided by Honda.
2. The success rate in identification of the suspicious claim before the model creation was only 35%. But by building a good model this success rate can be doubled.

**GOAL 2**

**Have sufficient service technicians to perform repair procedure depending upon the volume of requests.**

**SITUATION**

1. As mentioned in situation of Goal 1 Honda has around 12,000 dealerships nationwide that provide services to the vehicle owners in case of any issues in the vehicles.
2. Honda at this point has identified all the claims that me be a potential false claim using the solution of Goal 1.
3. Technicians are trained to handle different auto repairs.
4. Honda being a big company collects all the customer information while the vehicle is purchased and has a good data infrastructure (Datawarehouse) that stores all the relevant data pertaining to purchase, parts and the customers personal information within the permission given by the law.

**COMPLICATION**

1. Each dealership has limited number of service technicians that are assumed to work 5 days a week.
2. The number of service requests that can come in future changes depending on the month of the year and week of the month.

**QUESTION**

1. How many service requests are expected per month for a year?
2. How many technicians are needed to handle the services with minimum cost of paying the technicians?

**SOLUTION (For each question)**

1. The number of service requests expected per month can be forecasted using linear regression. Honda can take last 5 years of its data of service requests to fit a linear regression model. The data can be divided into Training dataset and test dataset. The model can be fitted on the train dataset and tested on the test dataset. The model can be run at every month and adjust the predictions accordingly

**Analytics Modelling Method**: Linear Regression

**Purpose**: Predict total number of service requests per month expected per dealership

**Predictors**:

* Month of the claim (March, April, June etc.)
* Warranty Period Pending in Month
* Is Warranty Eligible (Yes or No)
* Make
* Model
* Dealership

**Response:**

* Total number of claims

1. In order to reduce the cost of paying for technicians our aim is to have correct number of technicians to minimize the cost. Here we can use Optimization model to minimize the cost. At a time the number of technicians available must be between minimum and maximum allowed.

**Analytics Modelling Method**: Optimization

**Purpose**: To determine the number of technicians needed per month to handle all service requests.

n = Number of service requests per day

m = Minimum number of technicians allowed per day

M = Maximum number of technicians allowed per day

C = Cost per technician per hour

**Variable**:

t = number of technicians required per day

**Constraints**

t >=m => The total number of Technicians should be more than minimum technicians allowed

t <=M => The total number of technicians should be less than maximum technicians allowed.

**Objective Function**

Minimize ∑ C\*T => Minimize total cost per hour spent on technicians.

**INFERENCE**

1. The dealership can now know how many service requests to expect per month with the help of linear regression method
2. The dealership also now knows how many technicians to have ready depending on the number of claims

**GOAL 3**

**Forecast future needs of vehicle parts so that all vehicle parts are available when customer comes up with a service request.**

**SITUATION**

1. We now have a forecast of total number of claims per month and all the claims made in history.
2. We also have a model to identify if a claim is non-compliant or not.
3. We have a model to service all the technicians ready to handle a project.

**COMPLICATION**

1. Honda has limited number of service parts which needs to be distributed throughout 1200 dealership.
2. Honda also needs decide how many parts they need to distribute at each dealership depending on the nature of the service request.

**QUESTION**

1. How many engine parts may be needed per month per dealership for next year?

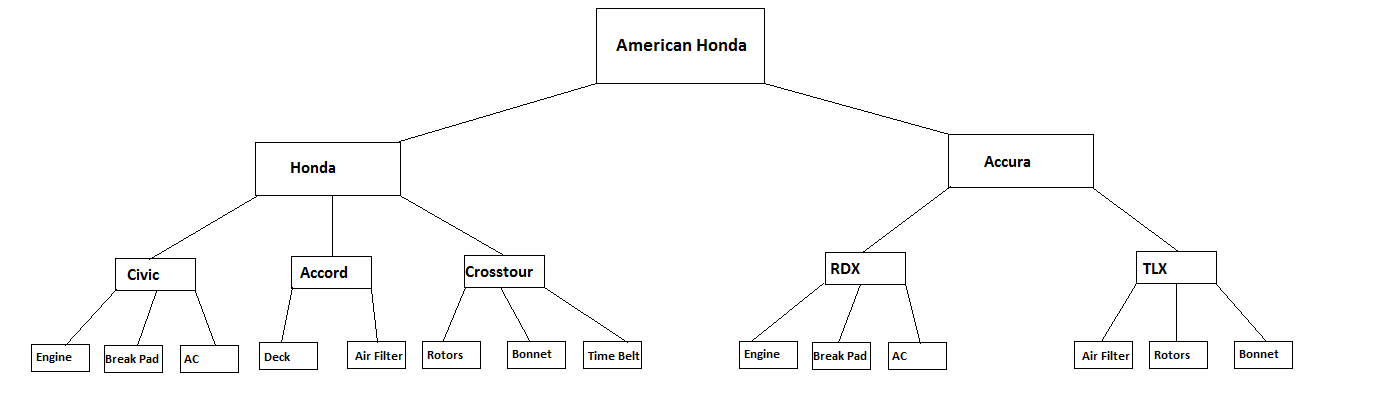
**SOLUTION (For each question)**

1. Here we can use a decision tree method and split the data first on vehicle model and then on the service part. Then apply Linear regression to each service parts to find out total number of service parts needed with the help of last 5 years of data. Once we have a prediction of total parts needed they can be pre ordered or pre-manufactured send it to the dealership.

**Analytics Modelling Method**: CART, Linear Regression, Exponential Smoothing

**Purpose**: To predict how many engine parts are needed per month for next year.

The picture below may not be a complete picture but a subset of the entire picture. The classification is done based on company, model and service part of the vehicle parts and then we do the regression of each vehicle part to predict how many vehicle parts are needed per month**.**

****

**Predictors for each service parts**:

* Month of purchase
* Month of manufacturing
* Year of purchase
* Make
* Mileage range
* Servicing frequency

**Response**

1. Quantity of the part needed

In order to validate the model used we can also use **exponential smoothing** each service part to verify the cyclical nature of the requests made every each for each part every year and get a pattern out of it. We can then change the variables of our linear regression model to be in sync with our time series analysis.

**INFERENCE**

1. Honda now has a plan for number of service part needed for each dealership in future and can make its parts ready before dealership approaches them and orders the parts.

**GOAL 4**

**Predict and Prescribe future needs pertaining to Design, Engineering, Manufacturing etc. for the next version of vehicles based on customer feedback.**

**SITUATION**

1. Honda (and Acura) releases a new model every year.
2. Honda has all the customer records across all of America with customer’s vital geographical information like zip code, city, state etc.
3. Honda has all the customer records across all of America with customer’s information like credit score, income, immigration status etc.
4. Honda has around 12,000 dealerships which can be used as means to connect to the customer physically and get feedback.
5. Honda also has historical records of customer feedback of all the surveys made in past.
6. Based on models built in Goal 1, Goal 2 and Goal 3, we have total number of service request a customer is going to make, which also means that we have the number of customers that are going to visit the dealership.

**COMPLICATION**

1. Every time a customer visits a dealership for making a service request he/she is made to fill a fairly long form which for some people may be frustrating. Making them fill a survey adds to complications
2. Not every customer is educated or has a background where he can give a feedback on technical aspects of a vehicle.

**QUESTION**

1. What questions to ask in the survey based on the customer that is approaching to the dealership?
2. How to make a right choice based on the outcomes of the survey results?

**SOLUTION (For each question)**

1. Since Honda releases a new version of each car model every year, we need to choose our survey questions which helps in getting a good comparison of current version of the vehicle with the previous version of the vehicle. It should also help us getting a feedback from a customer on features where customer had strong positive experience or strong negative experience. The questions asked must also help Honda get answers pertaining to parameters like Design, Engineering, comfort etc.The sample survey done must be a representative of all kinds of population. Hence we would classify our population using K-Nearest Neighbor Method using the Honda’s historical customer database.

**Analytics Modelling Method**: K-Nearest Neighbor Algorithm

**Purpose**: To classify the customers into 3 classes as shown below. Once this classification is done, survey questions will be made ready for each class.

Based on my understanding of the data, I would classify customers into 3 classes which covers most of the population.

**Class 1: GRAZERS.**

* Education is low may be have just completed school.
* Age is moderate to high compared to other clusters.
* Income is Moderate due to profession or they are retired.
* As the income is moderate their expenditure is moderate.
* House hold size is relatively high as they have a close-knit family.

**Class 2: BARGAIN HUNTERS.**

* Education is moderate as they are either in College or Graduating.
* Age is lowest among all clusters.
* Income is low as they are either doing an small job or doing an Internship.
* Expenditure is low as their income is low; they always hunt for Best Deals.
* House hold size is lowest as they are teenagers.

**Class 3: HIGH ROLLERS.**

* Education is highest among all clusters.
* Age is quite moderate.
* Income is highest among all clusters which can be derived from their profession.
* Expenditure is highest as their income is very high.
* House hold size is relatively high as they have a close-knit family.

1. Based on the outcomes of the survey made by American Honda, it can evaluate the changes to the design that can be made and come up with different designs for the next version with an appropriate choice of engineering to suit the customer needs for each car model of its vehicle. Once multiple designs are created proposed, we can use Multi-Armed Bandits method to choose the best car-model out of all the car-models proposed.

**Analytics Modelling Method**: Multi-Armed Bandits

**Pupose**: Multiple designs are proposed per car-model based on the outcome of the survey report. We need to choose the best design.

We can explore this model to a point where we can choose the best outcome.

**INFERENCE**

1. Now Honda has a method of classifying its population.
2. Based on the population Classified it can set up its survey questions
3. Based on the outcome of its survey Honda can now come up with many designs of its car-models for the next year.
4. Using Multi-Armed Bandit approach it can select the best design to be released.

**Summary:**

The table below gives a summary of models used to help Honda achieve its goals.

|  |  |  |  |
| --- | --- | --- | --- |
| **Goal No.** | **Goal** | **Modelling Method** | **Purpose** |
| 1 | Identify the suspicious claims accurately to reduce warranty expense. | Logistic Regression, Principle Component Analysis | Predict if the claim is suspicious or not. |
| 2 | Have sufficient service technicians that help depending upon the volume of requests. | Linear Regression | Predict total number of service requests per month expected per dealership |
| Optimization | To determine the number of technicians needed per month to handle all service requests. |
| 3 | Forecast future needs of vehicle parts so that all vehicle parts are available when customer comes up with a service request. | CART, Linear Regression, Exponential Smoothing | To predict how many engine parts are needed per month for next year. |
| 4 | Predict and Prescribe future needs pertaining to Design, Engineering, Manufacturing etc. for the next version of vehicles based on customer feedback. | K-Nearest Neighbor Algorithm | To classify the customers into 3 classes as shown below. Once this classification is done, survey questions will be made ready for each class. |
| Multi-Armed Bandits | Multiple designs are proposed per car-model based on the outcome of the survey report. We need to choose the best design. |