Analytics Modeling – General Motors

In this project, I have detailed the different analytics models and data that would be required to facilitate analytical decision making for a large auto manufacturer – General Motors (GM).

https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Industry-Profiles/General-Motors

Though the report is mainly based on the information for General Motors, I have also taken inspiration from the success stories of other auto manufacturers as below:

https://www.sas.com/en_us/customers/kia-motors-america.html

https://www.sas.com/en_us/customers/american-honda.html

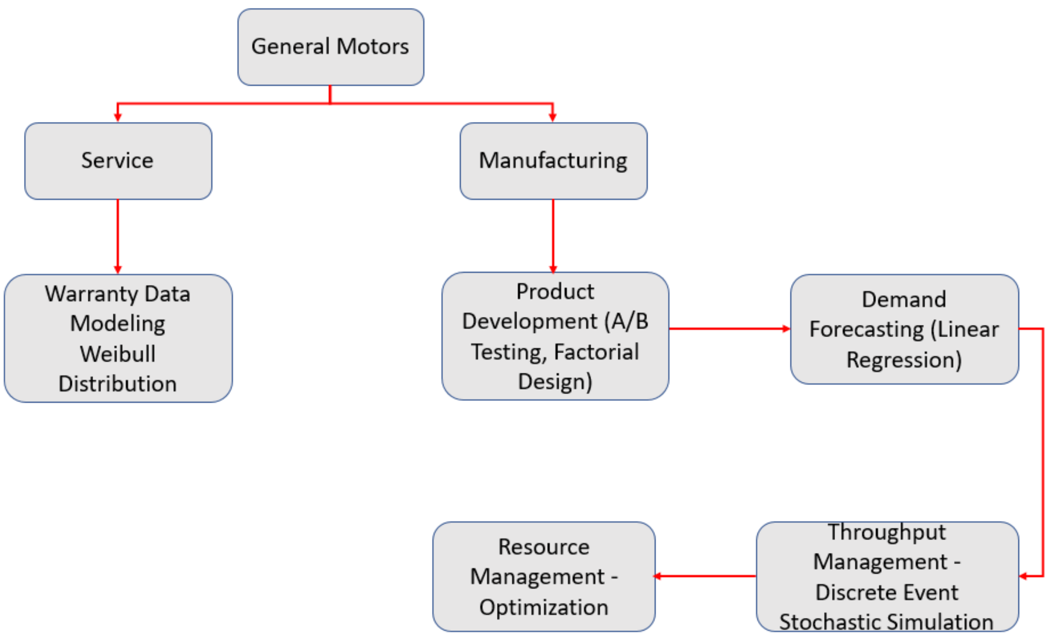

Figure 1

As shown in figure 1, I would like to approach this by looking into two broad areas of operations for General Motors. The flow diagram indicates that the output of one activity/model could form the basis of input into the next one.

Manufacturing:

First, we start with product development and the key here is to understand what features in an automobile are important to the customer. For example, let us consider the design of the dashboard – where should the start/ignition switch be, what colors should be used, how should navigating different options such as Audio, Settings, Phone and so on be structured. These are some of the questions that would come up during the design phase.

A/B testing would be a good option to answer such questions where the company could test two different dashboards to better gauge which one resonates with the customers the most. As discussed in the lectures, there are three items that GM must consider as given below:

- GM should be able to collect data quickly
- Data must be representative of the entire population – GM should select a diverse audience to test these options
- Amount of data collected should be small when compared to the entire population

Once the company finds out which option is better, they can perform a partial factorial design to understand why the customers preferred that option. Specifically, to find out what are the different features in that option that seemed very important to the customers and why. Overall, I am confident that a company as big as General Motors undertakes the above-mentioned tools as an integral part of the product development process.

Demand Forecasting:

The next step after product development is demand forecasting. General Motors makes diverse automobiles in many countries, so it is good to understand the demand by model and region. Below is the list of data that would be helpful in predicting the demand for any model type in a specific region/location.

- Model name
- Model type
- No. of doors
- Fuel type
- Body style
- Drive type
- City mpg
- Highway mpg
- Engine size
- Fuel capacity
- Model launch year
- Energy star rating
- Region
- Country
- State
- City
- Zip code
- Month
- Year
- Day
- Horsepower
- Average price
- Population
- Unemployment rate

- Auto loan lowest interest rate
- S&P 500 30-day returns
- Median income
- Crude oil price
- GDP
- Sales/Demand (units)

Before we fit any model to the above data, it would be good to perform appropriate variable selection methods such as stepwise, Lasso regression to determine the most important variables. Also, as shown above there are a few categorical variables, so it would be good to use one-hot encoding or similar techniques to get the data ready for Linear Regression.

The outcome of the regression model is units of automobiles that would indicate the demand.

Simulation:

Let us say that the regression model predicts that the demand for Chevy Bolt's over the next 1 year is 500,000 units. Now challenge for General Motors is to see if they can meet the demand. Do they have the resources in place to manufacture that many units? Also, if this question is extrapolated to many models across many regions, the problem becomes complex. The answer to these questions that I can infer from the article is Discrete Event Stochastic Simulation. It has to be noted that GM can use simulation in two situations: - 1) there is an existing manufacturing plant, model the processes in that plant to see if it can cater to the demand 2) there is no plant in that region, so prior to constructing a new manufacturing plant GM can model the designed processes to see if it is worth to do so.

We understand an Automobile has numerous parts some of which could be directly manufactured, and some could be procured from other vendors. So, let us assume that GM is trying to understand if they have the capacity to manufacture 500,000 Chevy Bolt engines in a year. To create a simulation model, the following data points are needed. Also, I would like to clarify that data can be collected by looking at historical process, if no such process exists then we can look at the design specifications or a process that seems to be very similar to the one we are considering modeling.

- Process flow – how many machines/operations are there. For example, operation 1 could be a grinding operation, operation 2 could be a milling operation, operation 3 could be to attach nuts and bolts and so on. In other words, we must know the sequence of operations required to manufacture an engine.

- The cycle time/processing time at each operation. We will use a probability distribution for the data rather than using a deterministic value. For example, rather than saying an operation takes 5 minutes always, we will say it is based on a Triangular distribution with a minimum 3 minutes, mode 5 minutes and maximum 7 minutes. We will use such distributions for all operations and can analyze historical data to come up with the appropriate distributions.

- In addition, we must know if these operations require other resources – for example labor, or other tools. Some operations are automatic, and some could require labor. If they require labor, then we must know about the processing rules, i.e., what happens if the labor is not available, if there are multiple resources do we always prefer the most qualified resource or all resources equally preferred. These are some of the questions that must be answered prior to building the model.

- There are fast operations and slow operations. It is obvious that queues will build up in front of slow operations. If so, what are the maximum queue sizes that are allowed in the model - how many engines can be kept in a queue in front of each operation. Once an engine enters a queue, how do they leave – a common rule is First in First Out (FIFO) but there are others.

- Machines breakdown and they need to be setup/serviced often. For example, a key data point that we need to know is how often machines breakdown. A machine that performs a drilling operation can breakdown approximately every 3 weeks, or after drilling 5,000 units. As explained above for processing times, these times should be modeled as a probability distribution – maybe an exponential distribution for the frequency of failures and a triangular or Weibull distribution for the repair time. The same logic can be applied for modeling setup time. For example, a cutting machine could require its blade be replaced every 1,000 operations. This also should be based on a probability distribution as sometimes the blade could last 1,000 operations and sometimes 1,200.

- Usually, quality checks are performed in a manufacturing process. For example, after a certain number of operations the engine will be checked to see if it adheres to the required standards. To model this we must analyze historical data and come up with the probability distribution to model the failure rate. For example, history can tell us that 5% of the engines fail to pass the quality test, out of which 95% are reworked and 5% are bad so scrapped. Some distributions that can be used to model failure rates are Log normal, Weibull and even Uniform distribution.

- We also need to collect information on other sequencing/routing rules. For example, what if there are 5 machines that perform the same operation, where does the engine get routed first – all machines could be equally preferred or maybe there could be a newer machine that is more preferred.

- People work in shifts and they need to take breaks. So, we must know when does the shift start and end and what happens during the breaks. This required to capture reality as much as possible.

- We also have to undertake the necessary steps to verify and validate the model. Verification is needed to make sure that the model behaves the way we have modeled, and Validation is needed to make sure that the model represents reality as much as possible.

- Once we have captured all the relevant details in the model, we must decide on the number of replications to run. In addition, we must know how long we should run the model for. If GM is trying to see if they can meet a yearly demand, then the model should be run for at least one year of manufacturing time after it achieves a steady state. It is also recommended that multiple replications are run, and the average of all replications be considered.

- In terms of the results, specific things that I am looking for are as follows:
    1) Throughput
    2) Total cycle time (start to finish) to manufacture an engine
    3) Bottleneck operation (what is the slowest operation)
    4) Machine and resource utilizations (are there machines or resources that are busy or idle more than the norm)
    5) Queues wait times
    6) Queue utilizations
    7) Impact of decisions in my model – decision to scrap parts rather than rework them
    8) Do any random events cause the manufacturing process to breakdown or get into a grid lock – for example if a machine breaks down frequently and for long times how does that affect my process.
    9) Cost incurred to rework or scrap parts.

- After analyzing the results mentioned above, we can change some of the parameters such as add more machines, labor, increase queue capacity etc. and re-run the model to observe the changes.

- Overall, the simulation model will not be useful if it does not represent reality, hence the above-mentioned details are required to make sure that the model is indeed a good representation of the real-life process. In reading the article about GM using Operations Research and analytics to manage their manufacturing process, I can envision all the above steps as an integral part of the study.

Optimization:

The next step is to optimize the results obtained from simulation. Let us assume that General Motors found that it needs to add 10 new machines, hire 25 more labor, and increase the queue capacity by 50% to achieve a high throughput. The question is what items General Motors should implement because everything will cost money and there is a budget that the company should stay within. This is where Optimization can help – specifically the variables will be the list of items such as number of machines, labor, size of queue and anything that affects the throughput. We can also have binary variables that would indicate if a machine of a particular type should be added or not. The constraints will be to keep the total cost to a budget, restrict the number of machines to be added to less than or equal to a number and so on. The objective function will be to maximize the throughput. Overall, the results of optimization will give GM a list of things that they should do to maximize the throughput by keeping their total costs within a budget.

Service:

So far, we discussed the data and models required from a manufacturing perspective. The same logic can be applied to the service segment of the company. For example, GM can take all the data that is collected every time a customer gets their car serviced. Specific items that could be of interest are:

- Model
- Year

- Miles driven
- Number of times the vehicle has been serviced so far
- Part that is broken/replaced/serviced
- Time taken to repair/replace that part
- Experience level of the technician performing the task
- Recommended service/replacement frequency for that part
- Where is the replacement part located – in terms of miles to the nearest location
- Manufacture date for the part
- Cost to manufacture/procure that part

The company can then use the above data to better understand when exactly some parts fail and why. They can then leverage that finding to better stock and service such parts. Specifically, GM can fit a Weibull distribution to the data and make decisions depending upon the failure rate obtained.

<u>Conclusion:</u>

Overall, we have considered 5 different analytical models that GM could be using to get insights into their operations. We also indicated that the steps listed under manufacturing are sequential in the sense that output of one study could form the input for the next one.