INF20016 FINAL PROJECT: BIG DATA MANAGEMENT

HUY NGOC NGUYEN – 103582239
HAI NGUYEN PHAM - 102953559

April 1, 2021

1. EXECUTIVE SUMMARY

This report is an analysis of data provided by Huskie Motor Corporation (HMC). The main purpose of this report is to obtain a thorough understanding of the company's performance through the data analysis process and data visualization. Some key findings related to the company's performance can be found below:

a. Findings:

- 64 vehicle transactions were eliminated during the data analysis process.
- North America, specifically the USA, generates the most sales and profit.
- Brand-wise, Apachete is the most profitable and Jackson is the least profitable in HMC.
- Model-wise, Advantage, Bloom, and Chare are the most profitable while Jespie and Mortimer are at the other end of the spectrum.
- Chare and Summet have the highest average contribution margin of all models.
- The top seller of add-on options is premium radio.
- The most sold package option is the technology package.
- Chare has the highest average variable cost and the highest variability in variable cost.
- Fleet operations generate the most sales of all sales channels, followed by retail and dealer operations.
- Island spends the most days on lot before sale, while Summet spends the least days on lot.
- The forecast of sales volume and contribution margin in 2017 shows little activity before drastically dropping in the last quarter.

b. Recommendations:

- Jespie, Mortimer, Rambler, and Crux models should be removed to increase profitability.
- Employee/partner programs and government accounts should be put under consideration as they generate negative profit and should be closed.
- An investigation should be made on why several car models have higher average days on lot than others.

2. INTRODUCTION

Huskie Motor Corporation (HMC) is an automobile manufacturing company with production and sales in 15 countries across Europe, South America, and North America. Although HMC is a new contender in this competitive industry, the company has generated an efficient number of sales, especially in North America, and has received positive feedback from its customers. With the rapid development of Big Data in the automobile industry, HMC has initiated a strategy focusing on recording a vast number of transactions to make better business decisions. This aims to get a better understanding of their customer base, and allows them to get an in-depth analysis of their data. To further extrapolate from this, HMC has asked D & A Consulting to help manage the massive quantity of data and create meaningful data visualizations from it. This report describes the application of the data analysis process as well as providing several visualizations of HMC's data to assess the company's performance and deliver possible recommendations to address any specific challenges.

3. OVERVIEW OF HMC AND THE AUTOMOBILE INDUSTRY

Huskie Motor Corporation (HMC) is an automobile manufacturing company that has recently come to the playing field. A profound understanding of customers, markets, and costs is extremely necessary to have a spot in the automotive industry, a highly competitive industry that currently has 15 companies with a global reach that provide up to 88% of the total number of vehicles around the world. There are 15 countries in North America, South America, and Europe where the HMC is currently residing and performing sales operations. Several key activities within HMC involve sales management, forecast analysis, product and package design, and data management.

- HMC employs different pricing options, discounts to better negotiate with their customers, and thus, reinforcing their customer relationship to create better customer satisfaction.
- HMC aims to predict its sales volume and revenue based on existing models to have a better sense of judgment towards their current operations, helping them to achieve or strive for a higher target.
- HMC's product and package design ranges from series, colors, seat type, add-on options, to package bundles. Altogether, the company looks to manufacture a vehicle comfortably under a customer's demand.
- With the rise of big data, HMC aims to create better services and products as well as figure out their valuable customers. The insight possibly gained from the influx of data will provide good momentum for the company to survive and have a chance to compete in this industry.

As for the automotive industry, Big Data has a crucial position in improving the customer's experience and safety by improving vehicles to become more reliable and convenient as well as providing better services in general (Shah 2020). The existence of this industry is dominated by only a handful number of vehicle manufacturers such as Toyota and Hyundai (Lee 2011). Hence, this market is considered oligopolistic. For instance, Toyota's revision of its strategies based on the utilization of big data allowed them to thrive in the automobile world. By focusing on providing customized products and prioritizing quality through diversified technologies, the company was able to rebrand its image to a higher standard (Nkomo 2013). Assessing their current performance and competition through the lens of Porter's Five Forces, intensity-wise, the Threat of Entry, Supplier Power, and Threat of Substitutes are very weak, the Power of Consumers is moderately strong, and the Competitive Rivalry is strong (Nkomo 2013, Porter 2008). Not to mention, Toyota's profitability in the automotive industry has always been affected due to the competition of other giant car manufacturers. Yet the Japanese automobile powerhouse demonstrated adaptability in its production in accordance to demand trends (Holweg 2008).

4. CHALLENGES DURING THE DATA ANALYSIS PROCESS

Before performing a preliminary analysis of the data, it is necessary to approach the data carefully by utilizing the data analysis process to simplify the decisions needed to overcome the challenges.

The challenge in approaching sample data is to grasp what the given data is, how each of the data components looks like, and the relationship between them. These questions can be answered by using the first step of the process, which involves defining and understanding the questions related to the company's goals. In this case study, HMC's representatives have outlined a few critical analyses to approach. The overall performance analytics, for example, assess the performance of the company and its brands, sales channels, and models. Other analytics regarding the finance and operations of HMC touch on variable costs and contribution margin of certain aspects of the company. Furthermore, the forecast analytics map out the potential trajectory of sales volume and contribution margin of the company. Each of these analytics serves as a starting point to get a better understanding of the current situation at the company.

The second step revolves around collecting the data. The sample data given in the report are represented inside an Excel spreadsheet that includes 25% of HMC's overall data. More than 2,000 car transactions were sold between the years 2015 and 2016. Each transaction contains several attributes, including a unique identifier, the brand and model, its variable and fixed costs, its sales and profits, as well as the country of sale.

The third step is to clean unnecessary data. With the possibility of errors and inconsistencies within the database, this challenge can be approached with the help of Excel and Tableau Prep. This helps to determine the quality of the data to make better decisions and predictions. Several errors in the sample data can be categorized as:

- Duplicate values: The VINs are the unique identifiers for each vehicle transaction. While these values can be simply renamed, it is better to remove them for clarity and safety measures.
- Null values: Although null values can be assumed based on the average or median of a specific attribute, it is too risky because some attributes are dependent on one another. Hence, rows with null values were removed.
- Non-existent values: In this case, there were a few brands, models, series, and packages that do not exist within the database offered by the company. Hence, these records were deleted.
- Incorrect attribute alignment: Some attributes were assigned incorrectly to its component. For example, several records showed the USA being allocated to Europe and Canada allocated to South America
- Incorrect calculation: The calculation of variable costs, fixed costs, and tariffs of several transactions was incorrect. This issue could be solved by assigning a formula in an Excel sheet, but specifically for this case, these records were removed.

The fourth step of the process deals with analyzing and manipulating the data. For this case study, Tableau will be used to generate graphs and illustrations to better visualize the data and understand the trend of each model. Further discussions of the company's analytics will appear in the next section.

The final step of the data analysis process concerns the interpretation of the above analysis. In other words, the end product of the case study will determine what actual values HMC can gain through the above steps. While considering the trends, forecasts, and recommendations the company could gain from the study, it is also necessary to think about any challenges that might occur in future data analysis projects

5. INTERPRETATION AND RESULTS

a. Overall Performance Analysis

HMC made the majority of its sales and profits in North America, aggregating nearly $43 million in sales and $4.9 million in revenue. The region also topped the chart of sales volume, with the USA covering 46% of the volume, followed by its regional counterparts Canada and Mexico, covering 7.5% and 7% respectively. The sales and profits gained by each region are labeled in Table 1. Although the company has documented half of its sales volume in the USA, the country only produced around 29% of the overall net revenue (Figure 1).

Region	Gross Sales	Net Revenue	After-tax
North America	42,899,500	4,891,711	3,349,886
South America	19,374,092	2,717,236	1,271,841
Europe	16,342,046	2,177,485	1,136,896

Table 1: Gross sales, net revenue, and profit of each region

HMC's after-tax income has risen from $2.8 million in 2015 to $3 million in 2016. Notable changes appeared in North America, where the regional income was reduced by $700,000 due to a decrease in sales volume. On the other hand, Europe and South America experienced an increase in sales volume and after-tax profit, with Europe gaining more than $700,000 and South America gaining nearly $150,000 (Table 2). In general, the company looks to be moving in a positive trajectory shown by the overall sales and profits. While the sales in North America have decreased in 2016, sales in other regions have displayed a promising improvement. This signifies HMC's resolution to expand their market globally.

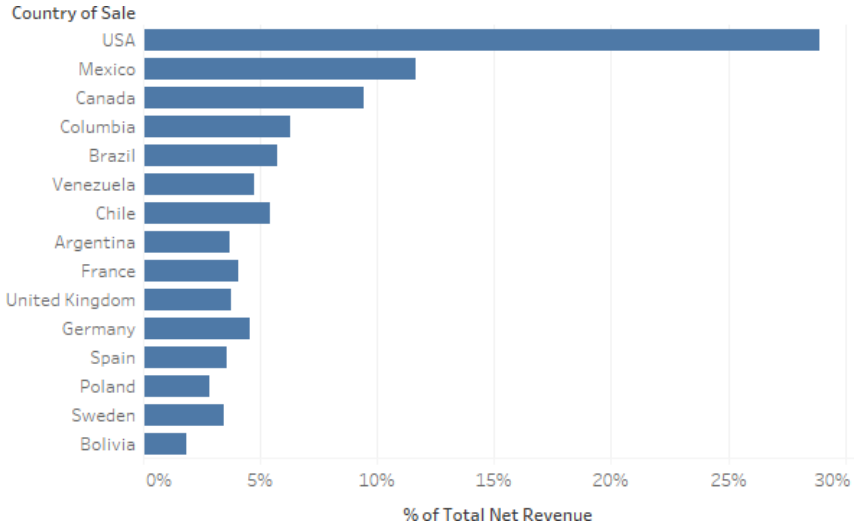

Figure 1: Percentage of net revenue shown in each country

	Sold Date			
	After-tax		Sales Volume	
Region	2015	2016	2015	2016
North America	2,028,075	1,321,811	901	695
South America	559,163	712,678	219	316
Europe	206,147	930,749	168	323
Grand Total	2,793,385	2,965,238	1,288	1,334

Table 2: After-tax income and sales volume in each region (2015-2016)

HMC currently offers three brands: Apechete, Jackson, and Tatra, with each supplying different types of models. In terms of gross sales, net revenue, and after-tax profit, Apechete has the highest indices in all three aspects out of the other two brands, with its gross sales over $31.3 million, net revenue at around $5.1 million, and profit at $3.1 million. As indicated in Figure 2, Tatra's gross sales are roughly $1 million less than Apachete's, yet the brand has surprisingly the lowest profit out of all brands. Jackson records the lowest gross sales in the company, but comes very close with Tatra's revenue and after-tax income. Looking through the years, all three brands have shown an increase in every aspect mentioned above.

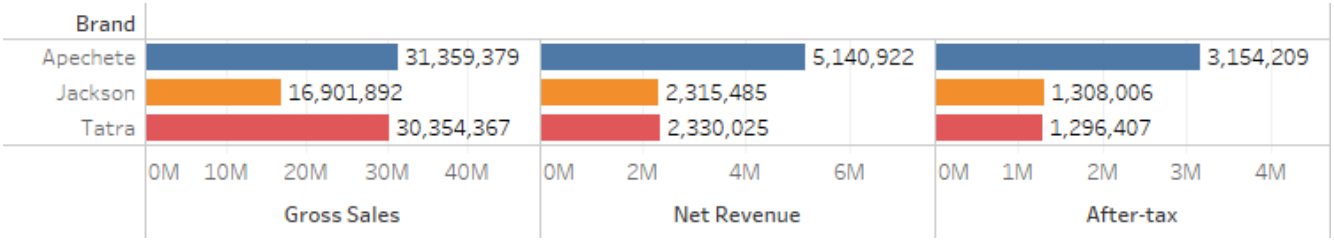

Figure 2: Gross sales, net revenue, and profit of each brand

HMC sells its automobiles through three sales channels:

- Sales channel L1 tells whether the transaction is made through a dealer or self-registration, fleet, or retail operations.
- Sales channel L2 identifies the type of customer account, which includes rental, commercial, employee/partner, government, and government accounts.
- Sales channel L3 refers to the transaction method, which includes cash, financing, or lease.

Out of all channels in the L1 level, fleet operations contributed the most with 37% of the transactions going towards gross sales, which is 0.5% more than the transactions made through retail operations. The sales volume made through the fleet was also the highest of all operations, standing at nearly 42% of the total volume. Dealer operations produced the lowest amount in both gross sales (26%) and sales volume (21%).

In the L2 channel, rental accounts contributed the most in gross sales, amassing $20.7 million in total (Table 3). Employee/partner and non-employee accounts were both placed behind the top-selling account with $16 million each. The sales volume was fairly distributed amongst every type of account, with the highest being employee/partner programs (25%) and the lowest being commercial accounts (16%). Throughout the years, both gross sales and sales volume in the L2 channel showed a marginal increase in all accounts. Especially in 2016, employee/partner programs and commercial accounts demonstrated a similar trend in the L1 channel as both accounts grew in relevance to their fleet operations.

In the L3 channel, cash and financing methods both increased their gross sales while the leasing method experienced a slight drop. The financing method experienced the biggest change out of all three methods, with an increase of 3.5% in gross sales and 2.5% in sales volume. By 2016, every transaction method was more or less valued at $13 million.

		Sales Channel L3		
Sales Channel L1	Sales Channel L2	Cash	Financing	Leasing
Dealer/Self Registration	Rental	7,359,979	6,056,860	7,333,980
Fleet	Employee/Partner Progra..	5,216,312	4,703,295	6,148,294
	Commercial Accounts	4,900,669	3,545,147	4,657,429
Retail	Non-Employee	4,820,729	6,185,586	5,773,341
	Government	3,889,368	3,746,015	4,278,634

Table 3: Gross sales of the sales channels in HMC

To find out the most and least profitable models in HMC, a list of after-tax profits in respect to each of the brand's models is displayed in Table 4. As shown, Advantage, Bloom, Chare, and Summet are the most profitable models, of which the first two models belong to Tatra and the latter ones belong to Apachete. Surprisingly, Jespie and Mortimer models are found at the other end of the spectrum, in which both of them are produced by Tatra. It is important to note that not every model is sold equally in each country or region, meaning that some models might have made a profit in North America but not in South America, for example.

Model	
Advantage	2,963,532
Bloom	1,386,050
Chare	1,174,036
Summet	1,166,117
Wood	962,200
Pebble	792,017
Rebel	396,574
Fiddle	192,805
Brutus	169,177
Robin	129,737
Island	-107,698
Rambler	-111,651
Crux	-412,749
Mortimer	-1,271,875
Jespie	-1,669,649

Table 4: After-tax profit of each model

b. Financial Analytics

The current contribution margin (CM) per model is illustrated in Figure 3. While Chare and Summit models top the chart, Mortimer is the only model that has a negative CM. Apachete models have the highest average contribution margin out of all brands, with the lowest model being Island, accumulating slightly below $10,000.

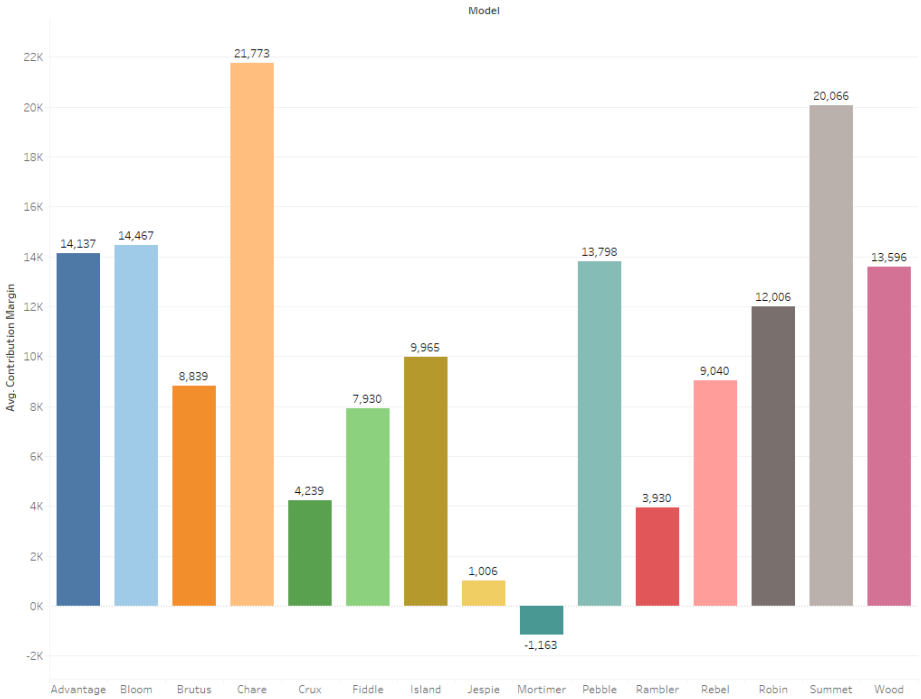

Figure 3: Average contribution margin of HMC's models

Figure 4 shows the average variable cost of each model in quarters, in which Chare is the model with the highest average variable cost, averaging around $28,700 through the quarters. Other models that sit above the average line are all Apachete models, which show an increasing trend over two years. Hence, every other model managed by Tatra and Jackson is located below the average line, with the lowest being Mortimer at $13,400 and the highest being Bloom at $17,500. Even though most models display shifts through the quarters, their progressions show little to no activity from Q1 2015 to Q4 2016.

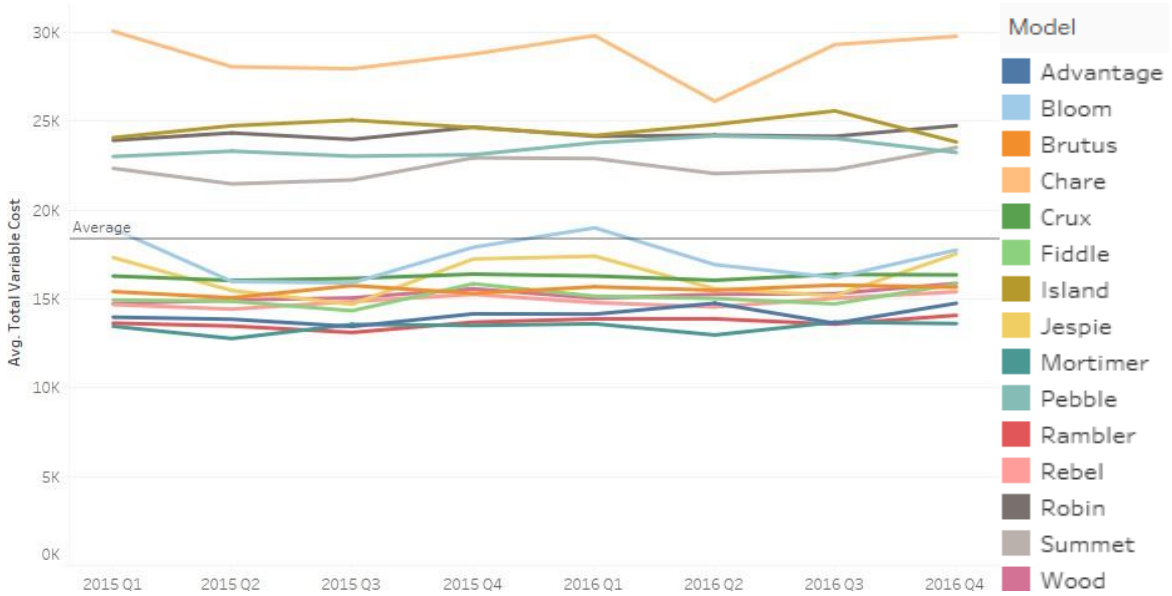

Figure 4: Average variable cost of HMC's models (in quarters)

By dividing the standard deviation of total variable cost by its average, the coefficient of variation is the measurement for the variability of variable cost. As a result, the model with the most variability is Jespie, with a standard deviation of $4.380 and a coefficient of 27.23% (Table 5). Although Chare has the highest standard deviation, its average variable cost keeps its coefficient down at 25.66%. Robin has the lowest variability in average variable cost at 3.6%, implying that the model has the most gradual and smooth development over time.

Model	
Chare	7,381
Jespie	4,380
Island	4,109
Pebble	3,703
Bloom	3,493
Advantage	3,154
Crux	2,569
Rebel	1,400
Summet	1,289
Wood	1,072
Rambler	1,036
Fiddle	1,032
Mortimer	898
Robin	891
Brutus	782

Table 5: Standard Deviation of Total Variable Cost of HMC's models

Table 6 displays the current CM per sales channel. In the L1 channel, 37.6% of the sales are made through retailers. In 2015, 77.8% of the sales went through either dealers or retailers. However, a massive surge in fleet operations saw that number fell to 63.8% in the following year. While retail operations experienced a small increase, dealer and self-registration sales went down by 12%.

In the L2 channel, rental and non-employee accounts accounted for 61% of the total sales while sales made through government accounts were only 9.54%. Throughout both years, rental and non-employee accounts encountered a reduction in sales, yet both remained the top-selling accounts in the company while other accounts attracted more sales.

In the L3 channel, sales through the leasing method were the most out of all methods, though the other methods also made a fair amount of CM. From 2015 to 2016, cash and financing methods experienced an increase in sales while the CM of leasing slightly decreased. Overall, all three methods achieved a similar amount of CM by the end of 2016.

Sales Channel L1	Sales Channel L2	Cash	Financing	Leasing
Dealer/Self Registration	Rental	2,807,175	2,415,435	2,835,755
Fleet	Commercial Accounts	1,457,520	1,024,050	1,216,765
	Employee/Partner Progra..	974,795	1,085,041	1,457,786
Retail	Non-Employee	1,922,538	2,552,730	2,413,713
	Government	869,177	679,902	789,527

Table 6: Current CM of HMC's sales channels

c. Operations Analytics

Figures 5 and 6 show the sales volume of the add-on options. Premium radio is the top-selling option, while parking assist is the bottom seller. Additionally, most vehicles are purchased with the Technology package (29.3%), followed by the Exterior Protection package (15.33%).

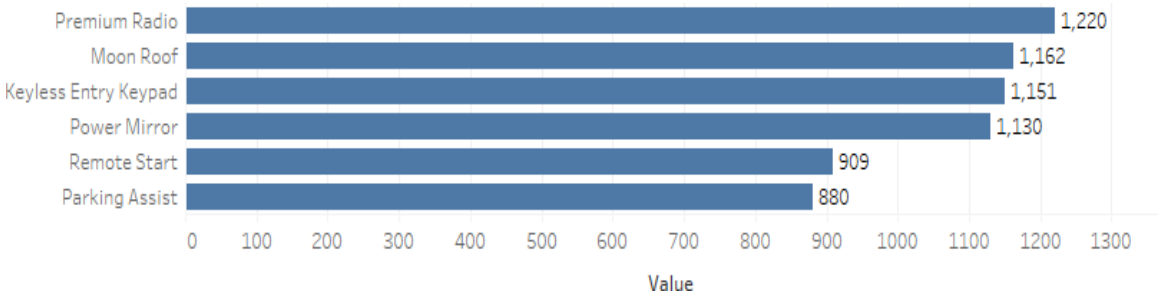

Figure 5: Sales volume of the add-on options

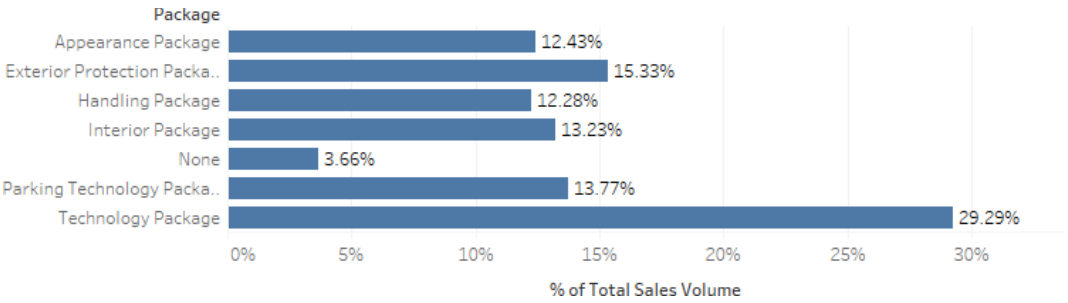

Figure 6: Percentage of the sales volume of packages

Figure 7 shows the average number of days a car spends on the lot before it is sold based on the model. While the Summet spends roughly 3 months on average on the lot before it is sold, the Island model takes more than 4 months to generate a transaction. However, this visualization only provides the average value of each model in all countries. There could be situations where a car model sells faster in a certain country but not in other regions.

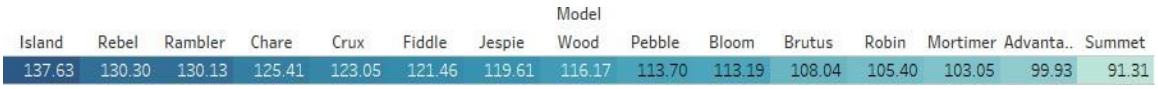

Figure 7: Average days on lot of HMC's models

d. Forecast Analytics

Based on the existing data on sales volume in 2015 and 2016, the forecast predicts a steady decrease in the first three quarters before a sharp plummet occurs in the last quarter of 2017 (Figure 8). In other words, the sales volume in 2017 looks to slightly decrease with no signs of improvement throughout the year.

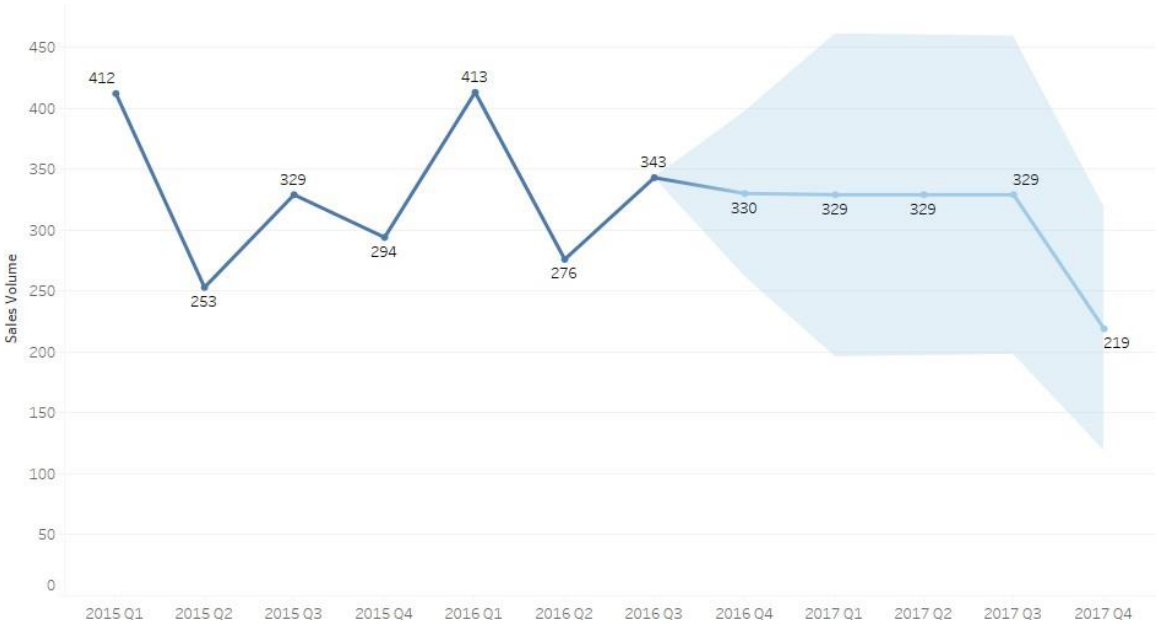

Figure 8: Forecast of HMC's sales volume (next 4 quarters)

Figure 9: Forecast of HMC's contribution margin (next 4 quarters)

The forecast of HMC's contribution margin in Figure 9 displays the same trend as the forecast of the sales volume. According to the Forecast function in Tableau, the CM does not exceed $3.13 million for the predicted first 3 quarters of 2017 and then drops down to $2 million in the last quarter. Unlike the sales volume forecast, the CM model shows a slight increase before stabilizing in the next three quarters.

In general, both forecasts do not portray a good representation based on existing data. Although, if the numbers in the predicted quarters are aggregated, they might be equal to the amount of any of the existing two years. If there were more data points over a larger span of years, the forecasts would be more accurate and show signs of shifting in the models.

6. PROPOSED APPORACHES

The visualization in Table 7 shows HMC's after-tax profit of HMC's brands and models in each region and country. HMC's first course of action should be eliminating models that have created a substantial loss and keep developing models with a consistent profit in every country. For example:

- Jespie, Mortimer, and Rambler models sold in the USA have contributed to a total loss of $3 million, while Chare, Summet, and Wood models have produced profits in every country. Thus, the first three models should be eliminated, and the latter models should be retained.
- The Crux model manufacture should be ceased from further production because it has made a loss of at least $100,000 in every region.
- The Island and Bloom models production in Europe and South America should be put into question as to their after-tax profits in most countries are negative.
- Advantage and Bloom models sold in the USA yield the highest profits and should be considered for more development to increase future sales.
- Tatra is the only brand that makes a negative profit in Europe and South America. Once the Bloom model is removed, the Jespie model should be closely reevaluated for the upcoming year in those two regions.

		Brand / Model														
		Apechete					Jackson					Tatra				
Region	Country of Sale	Chare	Island	Pebble	Robin	Summet	Brutus	Crux	Fiddle	Rebel	Wood	Advantage	Bloom	Jespie	Mortimer	Rambler
Europe	France	47,174	-4,434	40,381	3,264	33,716	26,620	991	26,207	-21,603	53,533	12,212	-10,109	1,841		
	Germany	63,195	-4,184	65,449	11,069	74,535	7,305	-28,169	21,489	25,692	88,488	5,270	-15,679	-7,719		
	Poland	42,215	-13,194	17,678	4,464	19,686	7,915	-63,622	-6,105	28,985	61,672	2,858	-9,484	16,014		
	Spain	26,005	-10,147	19,714	7,901	53,925	668	-15,328	7,324	8,780	64,437	388	-14,810	7,500		
	Sweden	55,206	6,345	33,726	11,419	32,388	7,518	-30,974	20,437	88,116	16,220	15,634	-2,371	-12,519		
	United Kingdom	41,353	-12,414	21,056	-804	23,993	14,998	-17,151	17,573	-19,418	60,390	-138	-3,863	-11,795		
North America	Canada	185,917	23,204	99,721	13,565	126,520	14,907	-31,219	29,911	99,196	83,745	18,114	-3,221	19,403		
	Mexico	135,562	-5,783	87,316	8,068	202,964	39,879	-64,042	28,463	93,474	103,534	73,718	-26,489	-5,324		
	USA	182,098	29,080	153,695	52,499	91,405	12,159	-49,385	-6,003	149,593	88,384	2,806,647	1,529,961	-1,657,822	-1,271,875	-111,651
South America	Argentina	24,812	-1,240	-10,862	-13,153	68,720	4,651	-38,408	-1,692	-6,207	62,016	2,327	-36,251	-8,960		
	Bolivia	23,465	-18,650	7,165	2,579	38,408	2,874	-11,418	540		17,474					
	Brazil	89,553	-64,548	6,752	-16,485	89,966	3,799	-37,163	6,094	-2,020	36,410	1,764	-15,731	8,692		
	Chile	101,282	6,575	109,150	21,480	132,592	9,849	22	18,586	-34,005	55,534	18,920	616	549		
	Columbia	70,123	-3,218	117,843	15,387	120,012	18,011	-367	16,976	-23,288	92,231	2,904	-1,082	-9,982		
	Venezuela	86,076	-35,090	23,233	8,483	57,288	-1,976	-26,518	13,003	9,280	78,133	2,914	-5,435	-9,527		

Table 7: After-tax profit of HMC's models in each country

HMC should also pay attention to employee/partner programs and government accounts as they are making negative profits through every sales method (Table 8). The company should consider eliminating these accounts immediately to maximize profit. At the same time, there is a risk of losing potential valuable customers if these accounts are removed.

The average days on lot of Island, Rebel, and Rambler models are quite high, so it might be worth it for HMC to investigate that aspect. Also, it is understandable that HMC is expanding its global market, but their after-tax profit in North America has decreased sharply in 2016. Thus, the company needs to reevaluate its business strategies in that region.

		Sales Channel L3		
Sales Channel L1	Sales Channel L2	Cash	Financing	Leasing
Dealer/Self Registration	Rental	1,111,310	1,012,252	1,125,469
Fleet	Commercial Accounts	227,747	128,225	69,245
	Employee/Partner Progra..	-360,665	-72,526	-29,810
Retail	Non-Employee	846,279	1,195,734	1,183,088
	Government	-95,976	-265,580	-316,170

Table 8: After-tax profit of HMC's sales channels

If Jespie, Mortimer, Rambler, Crux models along with transactions of Island and Bloom models in South America and Europe are all removed, then the overall profit would increase by 40%. Specifically, Europe and South America would both receive a 20% boost in after-tax profit, and North America would double their profit. The same trend can be observed when removing accounts with negative margins in sales channels L2.

The elimination of models with a negative profit would lead to fewer problems in liability and investment management as well as creating more opportunities to specialize in the more profitable models. However, as stated above, this procedure might result in a loss of valuable customers and repel the company's presence in the global market.

REFERENCES

Holweg, M, 2008, "The Evolution of Competition in the Automotive Industry", *Build to Order*, pp. 13-34.
Lee, C, 2011, "The Rise of Korean Automobile Industry: Analysis and Suggestions", *International Journal of Multidisciplinary Research*, vol. 1, no. 6, pp. 428-439. Nkomo, T, 2013, "Analysis of Toyota Motor Corporation", *Harvard Scholar*, analysis_of_toyota.pdf (harvard.edu)
Porter, E. M, 2008, "The Five Competitive Forces That Shape Strategy", *Harvard Business Review*, The Five Competitive Forces That Shape Strategy (hbr.org)
Shah, B, 2020, "How Does Big Data Impact the Automobile Industry?", *LHP Engineering Solutions*, How Does Big Data Impact the Automotive Industry? (lhpes.com)