**ALY6060 Module 5 – Assignment**

**Northeastern University**

A picture containing logo

Description automatically generated

**By**

**Kotha Apuroop – 002967351 (**[**kotha.a@northeastern.edu**](mailto:kotha.a@northeastern.edu)**)**

**28th October 2023**

**Driving Data-Driven Innovation: Honda's Journey in the Age of Industry 4.0**

**Introduction:**

Honda, a globally renowned automobile company with a rich history dating back to 1946, has long been a leader in the automotive industry. Known for its commitment to innovation and technological advancement, Honda has consistently produced high-quality vehicles and motorcycles that have captured the imaginations of consumers worldwide. As we delve into the role of data analytics and artificial intelligence (AI) in Honda's operations, it becomes clear that the company's commitment to innovation extends to its embrace of data-driven decision-making and cutting-edge technology.

**Industry Overview:**

The automobile industry is one of the most dynamic and fiercely competitive sectors globally. It's characterized by rapid technological advancements, shifting consumer preferences, and intense competition. As the automotive industry evolves, companies like Honda must remain agile to adapt to new trends, including electric vehicles (EVs), autonomous driving, and sustainability. Success in this industry requires not only staying ahead of the curve but also shaping the future of transportation, all of which relies heavily on harnessing data and AI to drive innovation and maintain a competitive edge.

**How Honda Uses AI in Their Business:**

Honda is actively incorporating AI in various facets of its business operations:

1. **Predictive Maintenance**: Honda employs AI algorithms to predict maintenance requirements for its vehicles. This approach minimizes vehicle downtime for customers and enhances the overall ownership experience. Drivers can have their cars serviced before a major breakdown occurs, resulting in safer and more reliable vehicles.
2. **Autonomous Vehicles**: Honda is investing in research and development related to autonomous vehicle technologies, where AI plays a pivotal role in enabling safe and efficient self-driving capabilities. This represents the future of mobility, making transportation more accessible and safer for everyone.
3. **Manufacturing Process Optimization**: Honda integrates AI into its manufacturing processes to improve quality control, reduce defects, and enhance production efficiency. This streamlines the manufacturing process, leading to higher-quality vehicles and increased competitiveness.
4. **Customer Insights**: AI is crucial for analyzing customer data and feedback. This deep dive into customer behavior allows Honda to gain a more comprehensive understanding of market trends and customer preferences, ultimately shaping future product designs and marketing strategies.

**Data Analytics Culture at Honda:**

A robust data analytics culture is cultivated at Honda, emphasizing data-driven decision-making and innovation:

1. The company invests in training its workforce, equipping employees with the skills required to extract meaningful insights from data. This empowerment ensures that everyone, from engineers to marketing teams, contributes to the company's data-driven approach.
2. A culture of analytical thinking and data-driven decision-making is fostered, where employees are encouraged to rely on evidence-based insights to make informed choices.

**Using AI and Data Analytics for Decision-Making and Product Offerings:**

Honda's commitment to AI and data analytics extends to key business operations and product development:

1. **Product Development**: Data analytics enables Honda to identify market trends and consumer preferences. These insights drive the design and development of vehicles that meet customer expectations and stay ahead of competitors.
2. **Quality Assurance**: AI-driven quality control processes are pivotal in ensuring that Honda vehicles consistently meet stringent quality standards. This, in turn, enhances customer satisfaction and upholds the company's reputation for producing high-quality vehicles.
3. **Supply Chain Optimization**: The company employs data analytics to optimize its supply chain. This results in cost reductions and improved production efficiency, making Honda more competitive in the global market.
4. **Personalized Marketing**: Honda uses AI to personalize marketing campaigns. By tailoring promotions and offerings to individual customer preferences, the company can better engage its target audience, leading to increased sales and customer loyalty.

**Benefits of BI and Data Analytics:**

The utilization of business intelligence (BI) and data analytics brings several advantages to Honda:

1. **Data-Backed Insights**: Data analytics provides a deeper understanding of customer behavior, market trends, and internal operations, allowing Honda to make informed decisions.
2. **Real-time Information:** The real-time data offered by AI and analytics tools enables Honda to respond more rapidly to changing market conditions and customer demands.
3. **Efficiency Gains**: AI-driven processes improve operational efficiency, reducing costs and enhancing overall productivity. This directly impacts the bottom line, ensuring Honda remains competitive.

**Preparation for Industry 4.0:**

As Industry 4.0 continues to transform the automotive sector, Honda must focus on several key areas:

1. **Advanced Technologies**: Investing in cutting-edge technologies such as 5G, the Internet of Things (IoT), and blockchain is essential for maintaining a competitive edge and adapting to the rapidly evolving automotive landscape.
2. **Cybersecurity**: With increasing connectivity, Honda must strengthen its cybersecurity measures to protect data and customer information in the age of connected vehicles. Cyber threats are a growing concern in the industry.
3. **Sustainability:** As global environmental concerns grow, Honda must continue to embrace sustainability practices. This includes the development of electric vehicles and eco-friendly manufacturing processes, aligning with global goals for a greener future.

**Global Impact and Future Success:**

Honda's commitment to data analytics and AI extends beyond its own business interests. The company is addressing critical industry and societal needs, such as sustainable transportation and data security. Honda's proactive approach positions it for continued success and a leadership role in the automotive industry as Industry 4.0 progresses and reshapes the future of transportation. Honda's dedication to innovation, data analytics, and AI demonstrates its commitment to meeting the evolving needs of customers and the world, and the company is well poised to remain a global leader in the automotive industry.

**Dashboard**

The Dataset used for this project is Honda Car sales data and the consumer reviews about Honda.

Dataset: - *Honda Cars Data*. (2023, March 28). Kaggle. <https://www.kaggle.com/datasets/omartorres25/honda-data?select=honda_sell_data.csv>

The other dataset used for this project is consumer reviews: - *Edmunds-Consumer Car Ratings and Reviews*. (2018, October 1). Kaggle. <https://www.kaggle.com/datasets/ankkur13/edmundsconsumer-car-ratings-and-reviews>

The sample of the Dataset is   
  
A screenshot of a computer

Description automatically generated

Python has been implemented to remove the unwanted columns and the rows from the dataset and the preprocessing tasks were performed by implementing Python.

In Tableau, the different models and the version of the Honda series were grouped together under the names Accord series, Civis series, HR-V series, Odyssey series, Passport series, pilot series, and Ridgeline series.

A screenshot of a computer screen

Description automatically generated

Fig 2:- the above image describes the exterior ratings given by the consumers to the Honda, here the series filtered is Accord series. Among the Accord series Accord sport 1.5T has the best ratings for the exterior design.

A screenshot of a computer

Description automatically generated

Fig 3:- the above image describes the distribution of Model price of Honda, here the series filtered is Accord series. Among the Accord series, the average price of accord series is $30,010.

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated A screenshot of a survey

Description automatically generated

Fig 4: - Dashboard. The dashboard has 5 charts, the consumer rating, distribution of model price, exterior rating, price vs engine vs model rating from the consumers and the pie chart describing the engine type.

The filter option has been applied on the model group, when the specific model has been selected the charts change as per the model series selected.

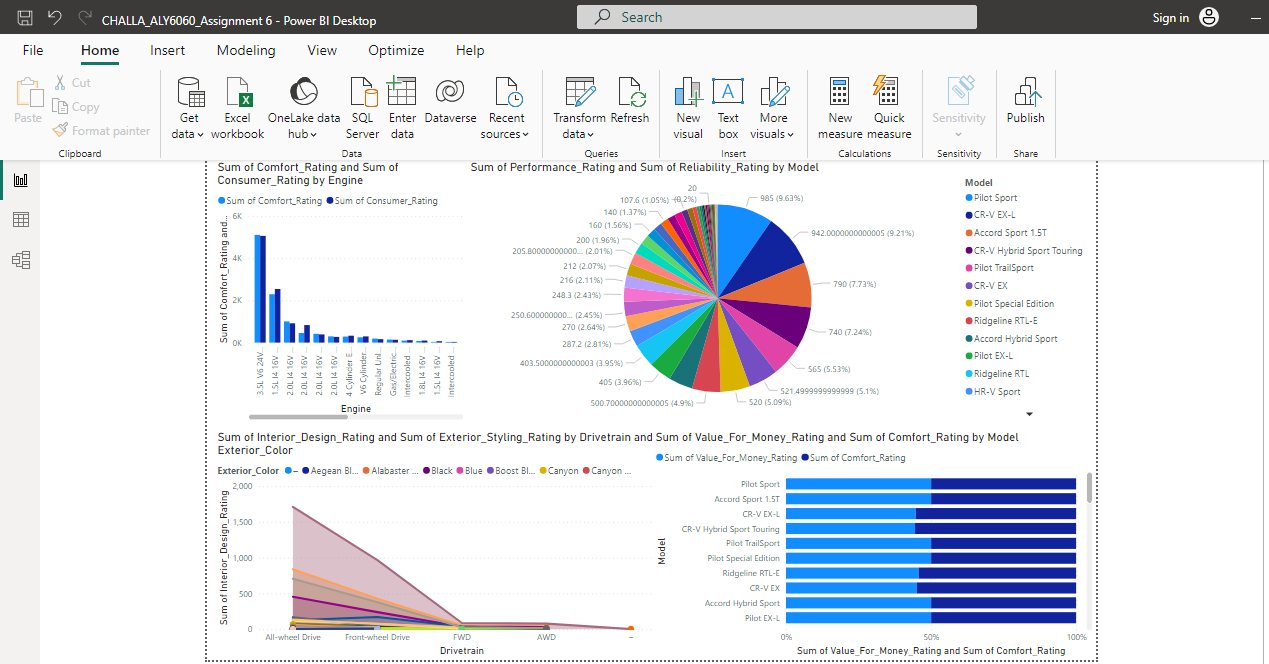


Fig 5: - Dashboard 2. Pilot Sport had the highest Sum of Performance\_Rating at 985, followed by CR-V EX-L and Accord Sport 1.5T. CR-V Hybrid EX-L had the lowest Sum of Performance\_Rating at 4.50. Pilot Sport accounted for 9.63% of Sum of Performance\_Rating. Sum of Comfort\_Rating and total Sum of Consumer\_Rating are positively correlated with each other. 3.5L V6 24V GDI SOHC accounted for 47.65% of Sum of Comfort\_Rating. Sum of Consumer\_Rating and Sum of Comfort\_Rating diverged the most when the Engine was 2.0L I4 16V GDI DOHC, when Sum of Consumer\_Rating were 378.20 higher than Sum of Comfort\_Rating.

Platinum White Pearl had the highest total Sum of Interior\_Design\_Rating at 2850, followed by Crystal Black Pearl and Sonic Gray Pearl. Meteorite Gray had the lowest total Sum of Interior\_Design\_Rating at 6.Platinum White Pearl had the highest average Sum of Interior\_Design\_Rating at 570, followed by Crystal Black Pearl and Sonic Gray Pearl. Meteorite Gray had the lowest average Sum of Interior\_Design\_Rating at 3. At 985, Pilot Sport had the highest Sum of Value\_For\_Money\_Rating and was 21,313.04% higher than CR-V Hybrid EX-L, which had the lowest Sum of Value\_For\_Money\_Rating at 4.60.

Sum of Value\_For\_Money\_Rating and total Sum of Comfort\_Rating are positively correlated with each other. Pilot Sport accounted for 9.91% of Sum of Value\_For\_Money\_Rating. Sum of Comfort\_Rating and Sum of Value\_For\_Money\_Rating diverged the most when the Model was CR-V Hybrid Sport Touring, when Sum of Comfort\_Rating were 185 higher than Sum of Value\_For\_Money\_Rating.

Later on, by using the dataset of the reviews provided by the consumers, and by using python and NLTK toolkit, We have performed Natural language performance and performed the sentiment analysis on the analysis.

**Sentiment analysis,** also known as opinion mining, is a fascinating field of natural language processing (NLP) that revolves around deciphering the sentiments and emotions expressed within textual content. It is a powerful tool for understanding and extracting valuable insights from the vast amounts of text data available today, ranging from social media posts and customer reviews to news articles and survey responses.

**Key Components of Sentiment Analysis:**

* Text Data: Sentiment analysis begins with text data, which can be in the form of short messages (e.g., tweets), long documents, or any text-based content.
* Preprocessing: The text is preprocessed to remove noise and irrelevant information, including punctuation, special characters, and stopwords.
* Feature Extraction: Sentiment analysis often involves converting text into numerical representations. Techniques like word embeddings (e.g., Word2Vec or GloVe) are used to capture word semantics.
* Sentiment Classification: At the core of sentiment analysis is sentiment classification. It categorizes text into predefined sentiment classes, typically "positive," "negative," and "neutral." Some advanced analyses may involve more nuanced sentiments.

The sample of the dataset is as below.

A screenshot of a car review

Description automatically generated

By basing on the review’s column, by using the preprocessing, stop words and using the text processing and calculating the sentiment scores, the reviews were categorized into positive and negative reviews.

A screenshot of a computer

Description automatically generated

A red and blue pie chart

Description automatically generated

Fig 6:- this pie chart describes the distribution of the reviews on the car engine. The engine types referred in the dataset are as follows,

A screenshot of a computer

Description automatically generated

Using the tokenization and count words, extracted the most common words used under positive reviews and negative reviews.

Tokenization is a fundamental text preprocessing step in natural language processing (NLP). It involves breaking down a text document or sentence into smaller units called tokens. Tokens can be words, phrases, sentences, or even individual characters, depending on the specific task and requirements. In most cases, tokenization refers to splitting text into words or sub word units (sub tokens).

Word Count, on the other hand, is a straightforward operation that, as the name suggests, counts the number of words in a text document or a piece of text. It is a basic and common text analysis technique that provides insights into the length and complexity of a document.

A screenshot of a computer

Description automatically generated

And to understand the context of the common words and the most frequently used words, bi grams and trigrams were used to extract the context of the reviews.

A screenshot of a computer program

Description automatically generated

The sample describing the counts of most commonly used positive words,

A screenshot of a cell phone

Description automatically generated

The sample data describing the counts of most commonly used negative words,

A screenshot of a computer

Description automatically generated

By visualizing the positive and negative words in tableau, below are the charts showing the data,

A green squares with black text

Description automatically generated

Fig 7:- this Heat map describes the most frequently used positive words in the reviews these are the bigrams. Most of the consumers were discussing about the gas and mileage.

A screen shot of a computer

Description automatically generated

Fig 8:- this Heat map describes the most frequently used negative words in the reviews these are the trigrams. Here the most frequently used negative reviews were buy another Honda, this might be because the consumers were referring to a specific model, so they might have reviewed as buy another Honda.

**Conclusion: -**

**Leveraging Sentiment Analysis for Honda Car Reviews**

Analyzing customer sentiment in reviews of Honda cars, by employing tokenization, word count, and sentiment scoring with the NLTK toolkit in Python, offers valuable insights to both the automotive industry and the manufacturing sector as a whole.

The process of sentiment analysis begins with breaking down the text in the reviews into tokens, typically individual words. Word count provides a basic but essential metric, indicating the length and complexity of the reviews. However, the real power of sentiment analysis lies in understanding the emotional tone and context conveyed in the text.

By categorizing reviews into positive and negative sentiments, manufacturers gain the ability to systematically gauge customer satisfaction and identify areas for improvement. Positive reviews can highlight features and aspects of Honda cars that resonate well with customers, while negative reviews pinpoint pain points, defects, or shortcomings.

This data-driven approach enables manufacturers to make informed decisions about product updates, enhancements, and even future vehicle design. They can prioritize areas that receive positive feedback for further development and address issues raised in negative reviews. For instance, if a significant number of negative reviews mention a recurring problem with a specific car model, the manufacturer can take proactive steps to rectify the issue, improving customer satisfaction and brand loyalty.

Additionally, sentiment analysis helps the automotive industry understand changing market trends and consumer preferences. This insight can inform marketing strategies, product positioning, and even research and development efforts. It enables manufacturers to stay competitive by adapting to evolving consumer demands and staying ahead of the competition.

Sentiment analysis, when applied to Honda car reviews and similar consumer feedback, provides a data-driven foundation for enhancing product quality, customer satisfaction, and overall market performance. By embracing the insights offered by sentiment analysis, the manufacturing industry can continue to evolve its products to meet the ever-changing needs and expectations of its customer base, ensuring long-term success and brand loyalty in a highly competitive market.

References: -

* Medhat, W., Hassan, A., & Korashy, H. (2014, December 1). *Sentiment analysis algorithms and applications: A survey*. Ain Shams Engineering Journal; Elsevier BV. <https://doi.org/10.1016/j.asej.2014.04.011>
* *Sentiment analysis of product reviews: A review*. (2017, March 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/7975207?casa_token=tioL8JgeyZEAAAAA:JRPAahZX9klnbxb9MX_SoDGbSAFkdUXCBrSNs4rgAQ1MmwbQ1URSr7XzqSoFNrWmmqIkz5SWVNk>
* *Twitter sentiment analysis: A case study in the automotive industry*. (2015, November 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/7360594?casa_token=Z-XI-bA5QYkAAAAA:cWKyaJCjR_mD8n4YBDSii3Ro3wZ4pPYkE0p4exXjfgHsfD0GVJVNpVxxHUH97_Rl-TUA3UH2-Gw>
* *Honda Cars Data*. (2023, March 28). Kaggle. <https://www.kaggle.com/datasets/omartorres25/honda-data?select=honda_sell_data.csv>
* Leung, K. (2021, September 23). *How Tableau Helps Your Organization Achieve Greater Data Insights*. <https://www.datacamp.com/blog/how-tableau-helps-your-organization-achieve-greater-data-insights>
* Avasthi, S., Chauhan, R., & Acharjya, D. P. (2021, January 1). *Processing Large Text Corpus Using N-Gram Language Modeling and Smoothing*. Lecture Notes in Networks and Systems. <https://doi.org/10.1007/978-981-15-9689-6_3>