Analysis of Brand Preferences Among Top 2W Companies in Indian Auto

Vaibhav Bodhe

3/4/24

Big Data Analytics: Analysis of Text

Table of Contents

Introduction:

India, home to one of the world's largest populations, comprised **17.76%** of the global populace in 2023. Government statistics indicate a modest annual population growth rate ranging from **0.8% to 0.9%.** Projections suggest a rise of **1.4 million individuals**, equating to a **7.19% increase by 2028**. This demographic surge is poised to amplify demand across sectors like automotive, particularly for transport options such as public and private transport, driven by stringent governmental regulations. By the close of 2023, urban residents were anticipated to make up **approximately 36%** of India's population. This burgeoning populace fuels the need for enhanced transportation services. **Urbanization and migration** have fueled the expansion of the transportation sector, spurring demand for various vehicles including motorcycles, cars, and bikes nationwide.

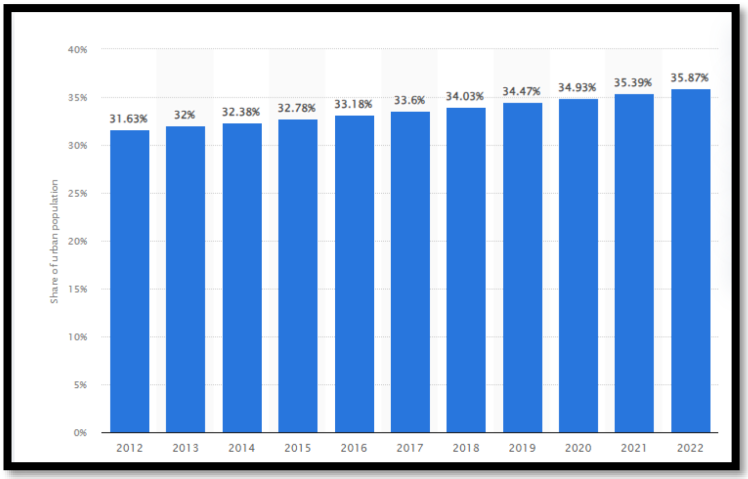
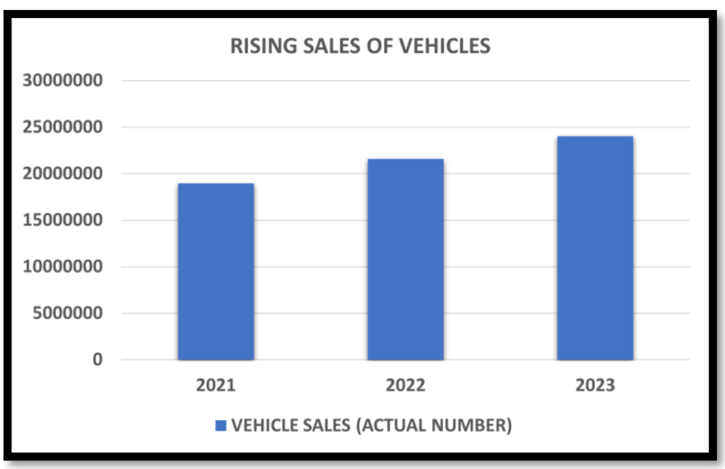


Fig. Degree of urbanization from 2012 to 2022

Source: https://www.statista.com/statistics/271312/urbanization-in-india/

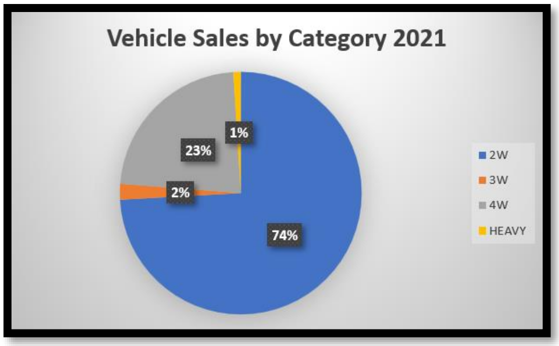
As vehicle sales surge, the streets of India are bustling with an increasing number of vehicles. Below, we observe the upward trend in vehicle sales.

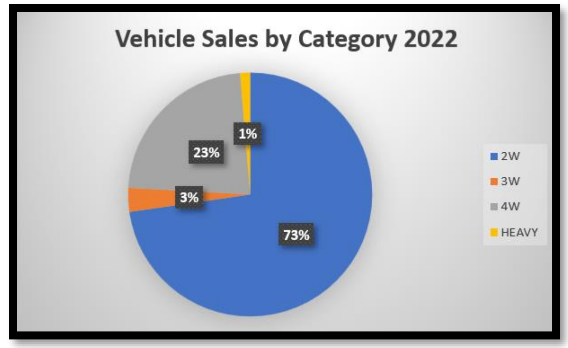


Source: (https://parivahan.gov.in/parivahan//en/content/vehicle-related-services)

Indian two-wheeler Automobile Market:

With the sale numbers surging we can clearly see the demand for 2 wheelers on Indian roads, with approximately **74% share** in overall sales of automobile.





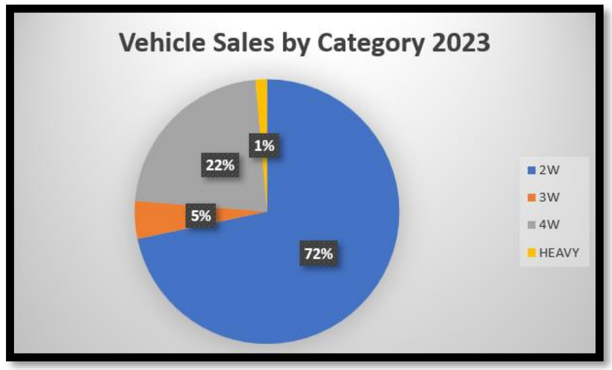


Fig. The Graph above shows the consistent demand for 2W on Indian.

Source: (https://parivahan.gov.in/parivahan//en/content/vehicle-related-services)

In addition to the traditional gasoline-powered two-wheelers, there is a notable surge in the popularity of electric vehicles within this segment. The burgeoning interest in electric mobility has ushered in a new era, marked by a significant rise in sales of electric two-wheelers. This shift towards cleaner and more sustainable transportation options not only reflects evolving consumer preferences but also underscores the growing importance of environmental consciousness in the automotive industry. **In 2021, 1.12%** of two-wheelers sold were electric, showing modest adoption. By **2022, this ratio surged to 4.05%,** indicating growing acceptance fueled by environmental awareness and technology. In **2023, it further rose to 5.03%,** signaling a notable shift towards electric vehicles as consumer preference evolves. Electric two-wheeler industry rebounds **post-subsidy cuts, gearing up for low-cost model launches**. High-speed e-bike **sales up 20% in September to 63,716 units**, while **ICE bikes see 22% growth amid festive season**. Despite initial dip to 46,000 units post-subsidy reduction, Elara Capital analysis shows steady growth with 66,600 monthly registrations in H1 2023-24, surpassing 2022-23's average of 60,500 units.

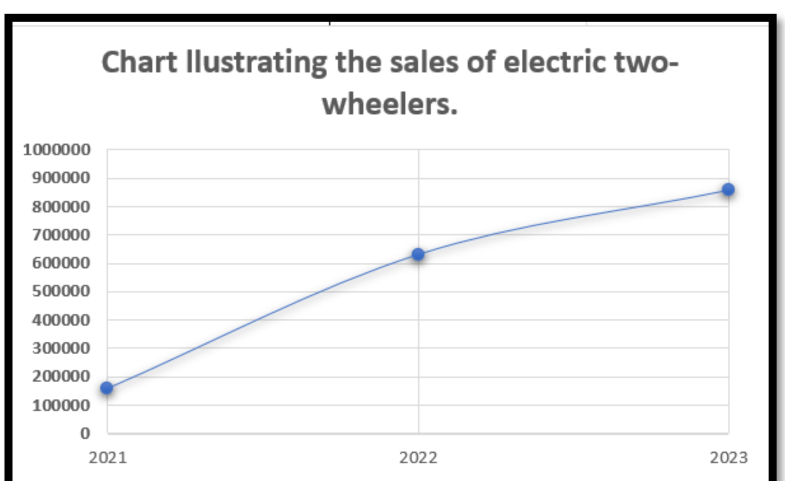


Fig : Increasing trend of electric 2W sales.

Source : https://parivahan.gov.in/parivahan//en/content/vehicle-related-services

Current Major Players:

The significant increase in electric vehicle sales percentages from 2021 to 2023 has sparked a keen interest in exploring the dynamics of the electric two-wheeler segment. Motivated by these compelling figures, **a research initiative is being undertaken to delve into brand preferences** within this burgeoning market. Focusing on three major players — **OLA Electric, Bajaj, and TVS** the study aims to conduct a thorough brand preference analysis. **TVS is selected as the focal firm for in-depth examination**, while **OLA Electric and Bajaj are considered competing firms**. Through this research, insights into consumer preferences, brand loyalty, and market positioning of these prominent players are being sought, shedding light on the evolving landscape of the electric two-wheeler industry.

In the domain of electric scooters, **OLA** announced a significant 74% sales surge in December 2023 compared to the previous year, along with a 68% growth in the quarter ending December 2023, totalling 83,963 units. The EV manufacturer reported selling 2.65 lakh electric scooters in 2023.

Offerings by OLA Electric:

OLA Electric Offerings		
Sr No.	**Models**	**Price**
1	OLA S1 Pro 2nd Gen	Rs.1,29,999
2	OLA S1 air	Rs.1,04,999
3	OLA S1 X	Rs.79,999

Bajaj Auto's Chetak also has exhibited notable advancement, nearly tripling its market share within just five months. Current Vahan data indicates a market share nearing 15%, compared to its previous standing of approximately 5% a few months ago. Chetak's performance has been exceptionally strong for the company. It would have registered higher sales numbers since the Chetak's launch but for the fact that the Chetak is being sold in far fewer markets than its rival. But now there are aggressive plans to expand the Chetak's exclusive retail network.

Offerings by Bajaj Chetak :

Bajaj Chetak Offerings		
Sr No.	**Models**	**Price**
1	Bajaj Chetak Urbane – Standard	Rs.1,15,001
2	Bajaj Chetak Urbane – Tecpac	Rs.1,23,001
3	Bajaj Chetak Premium – Standard	Rs.1,35,463
4	Bajaj Chetak Premium – Tecpac	Rs.1,44,463

TVS's I Qube, our focal product, experienced a remarkable sales increase in FY2023, with 96,654 units sold, up by a significant 797% from FY2022. This surge was propelled by the introduction of the refreshed iQube in May 2022, boasting enhanced range and features. However, despite the sales growth, **there's been a decline in the TVS brand** due to numerous customer complaints about its services. It's crucial for the brand to carefully analyze customer feedback to drive operational and product improvements and sustain sales momentum.

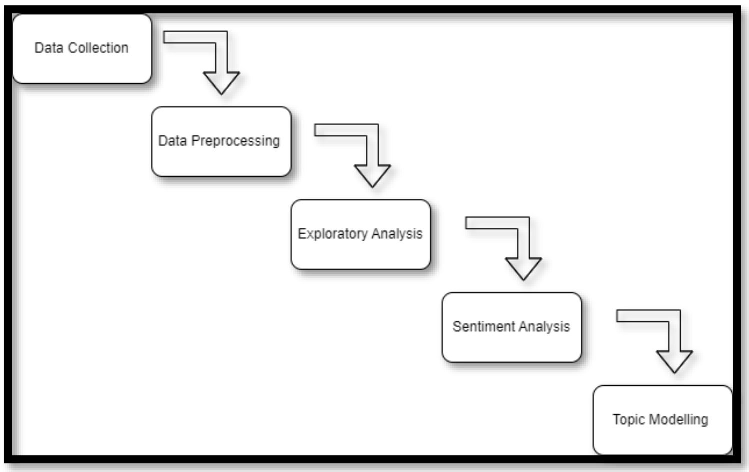
Offerings by TVS :

TVS Offerings		
Sr No.	**Models**	**Price**
1	TVS iQube	Rs.1,55,553
2	TVS iQube S	Rs.1,62,090
3	TVS iQube ST	Rs.1,25,000

Having thoroughly examined the Indian automobile market and conducted a brief assessment of our competitors' performance, we are now poised to delve into a comprehensive analysis of customer feedback pertaining to these brands. Our objective is to gain insights into the aspects of products and services that are favored or disliked by customers post-purchase. **By meticulously scrutinizing customer reviews across various platforms, we aim to discern prevailing market sentiments, identify key triggers for customers, and pinpoint areas for enhancement within our focal firm, TVS.**

Analysis Framework Adopted:

To formulate recommendations for the management of the focal firm, we have implemented a series of coherent steps that adhere to industry standards, aimed at acquiring actionable insights. Each outlined step is thoroughly elucidated, and the outcomes are provided in the accompanying documentation.



1.Data Collection

Step a. Extraction

1. The report consolidates feedback data comprising 250 reviews per company sourced from various platforms, including ZigWheels, BikeWale, BikeDekho, AutoCar, as well as user-generated content platforms like Mouthshut. Additionally, reviews from YouTube were included.
2. Data extraction from websites was facilitated by a web-scraper script, while comments from YouTube were obtained using a dedicated API account through a Python script.

Step b: Data Integration

1. The extracted results from both website sources and YouTube comments were merged into into single Excel.

Note : The compiled Excel containing the aggregated data for each company are provided as attachments accompanying this report submission.

2.Data Preprocessing:

Below outlines the steps executed to prepare the extracted reviews from various sources for further processing.

Step 1: <u>Define Regular Expression Patterns</u>

- A regular expression pattern is established to identify characters that are **not alphanumeric** or punctuation.

- Another regular expression pattern is formulated to identify **non-ASCII characters**.

Step 2: <u>Text Preprocessing</u>

- **Tokenize the text** into words using the **'word_tokenize()' function** from the NLTK library.

- Utilize a predefined set **of English stopwords** obtained from the NLTK library to eliminate stopwords from the tokenized text and **concatenate the preprocessed tokens** into a single.

3.Exploratory Data Analysis:

To initiate the exploratory data analysis, we first examine the structure of our dataset. It comprises two columns: "Company" and "Review," with each row representing an individual's feedback. Additionally, we have extracted 250 reviews for each company, resulting in a dataset size of **750 rows and 2 columns**.

Value Counts	
OLA	251
TVS	251
BAJAJ	250

Given our focus on textual data, it's pertinent to examine the distribution of review lengths submitted by users. This provides insight into the depth of emotional or practical engagement exhibited by customers in their feedback.

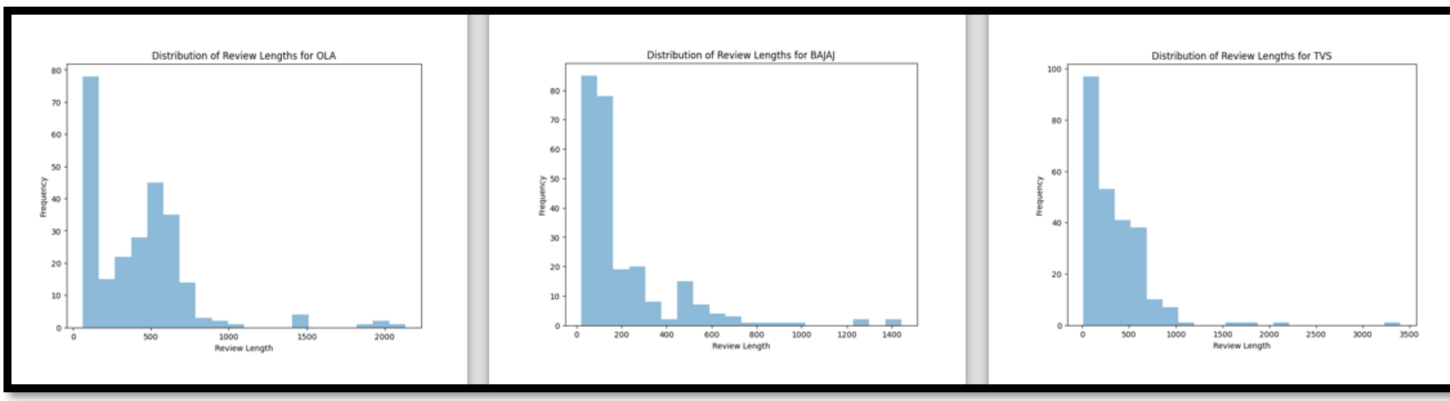
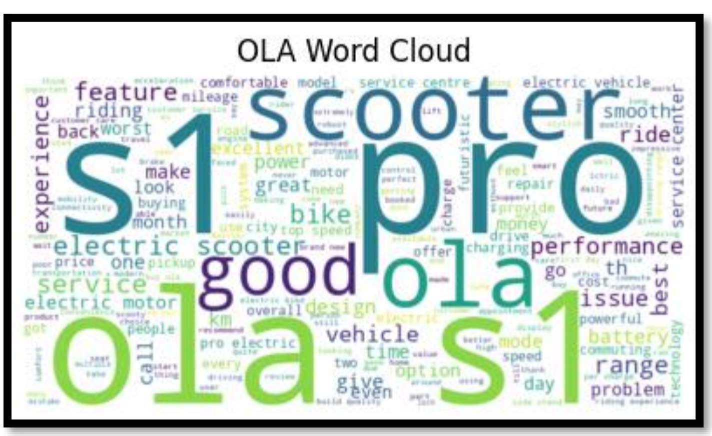
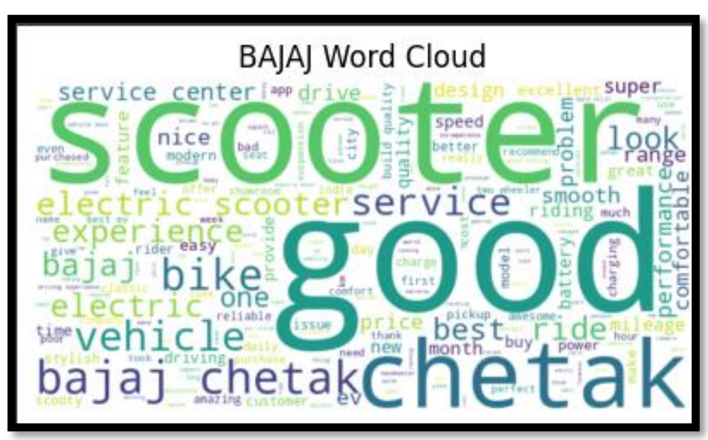


Fig. Company wise Distribution of review lengths posted by users.

Upon close examination of the distributions, one clear inference emerges: **users typically express themselves within 1000 characters**, as evidenced by the left-skewed distribution displayed above. To quickly overview the reviews, we generate word clouds for each company, enabling a rapid analysis of the review content at a glance.





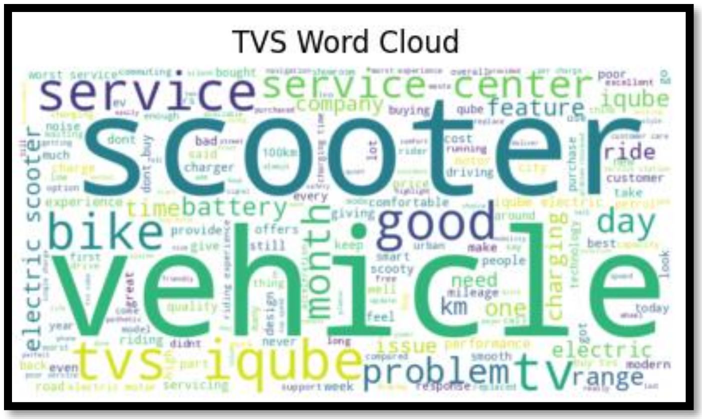


Fig. Company Wise Word Clouds of review text.

Although word clouds visually represent the most frequent words in a corpus, facilitating the identification of prevalent themes, pinpointing exact themes in a large corpus requires specialized modes of analysis to gain a clearer understanding of customer interaction. We will be going through some techniques in detail for this analysis

4.Sentiment Analysis:

Sentiment analysis, also referred to as opinion mining, constitutes a fundamental **natural language processing** methodology employed to ascertain the sentiment conveyed within textual content, encompassing **positive, negative, or neutral expressions**. This technique serves as a valuable tool in **comprehending public sentiment or customer feedback**. Within our context, it holds paramount significance to assess each review based on its sentiment score, facilitating the **quantification of user sentiments** conveyed through the reviews collected.

For Sentiment Analysis we have used **Roberta Model** to quantify the sentiments of the review collected, **Roberta, an advanced Transformer-based model in NLP**, excels in various tasks like text classification and sentiment analysis. With its pre-trained knowledge **and fine-tuned parameters**, it's widely used for **processing and analyzing textual data**, delivering state-of-the-art results in diverse applications.

The computing begins by loading a pre-trained sentiment analysis model named "**cardiffnlp/twitter-roberta-base-sentiment**". Subsequently, the **associated tokenizer is initialized to convert raw text into numerical inputs that the model can comprehend**. Once the tokenizer is set up, the sentiment analysis model is initialized, **specifically tailored for analyzing the sentiment of text inputs**. With both the tokenizer and model prepared, **the next step involves tokenizing each review's text**, breaking it down into individual tokens or words. Finally, the tokenized text is passed through the sentiment analysis **model to compute sentiment scores for each review,** thus completing the sentiment analysis process.These scores represent the model's prediction of the sentiment of each review. The scores are then aggregated on company-level to check the overall brand positioning

Results:

Company	Avg Negative Score	Avg Neutral Score	Avg Positive Score
BAJAJ	0.201	0.206	0.591
OLA	0.196	0.192	0.61
TVS	0.401	0.297	0.3

BAJAJ:

- BAJAJ's higher average positive sentiment score (0.591) suggests overall satisfaction among customers, outweighing negative (0.201) and neutral (0.206) scores.

OLA:

- OLA's elevated average positive sentiment score (0.61) reflects contentment with its services, surpassing negative (0.196) and neutral (0.192) scores.

TVS:

- TVS displays a mixed sentiment, with its average positive score **(0.3)** lower than negative (0.401) and neutral (0.297) scores.

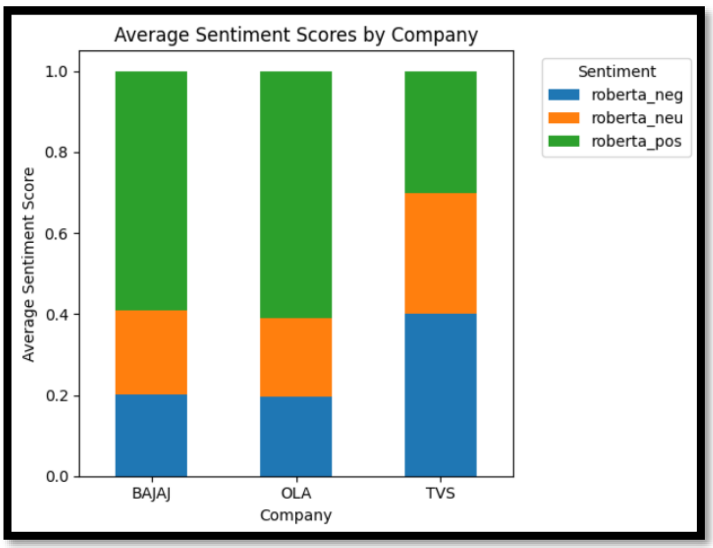


Fig. Average Sentiment Score by Company

Given that the general sentiment towards TVS leans towards negativity, it would be beneficial to initially examine the distribution of negative sentiments to gauge the extent of dissatisfaction expressed by customers**.**

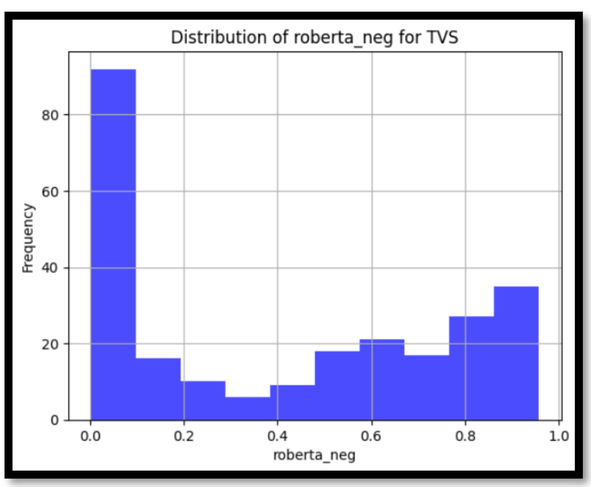


Fig. Distribution of Negative reviews sentiment score for TVS

Here we can clearly observe that sentiment score equal to and above 0.4 have a significant amount of occurrence now it becomes of utmost importance to identify the the key elements that is causing this satisfaction among customers.

To introspect the elements that are causing this negative connotions among customers we tried to create a word cloud from the reviews/comments from customers having negative sentiment score of 0.4 or above.

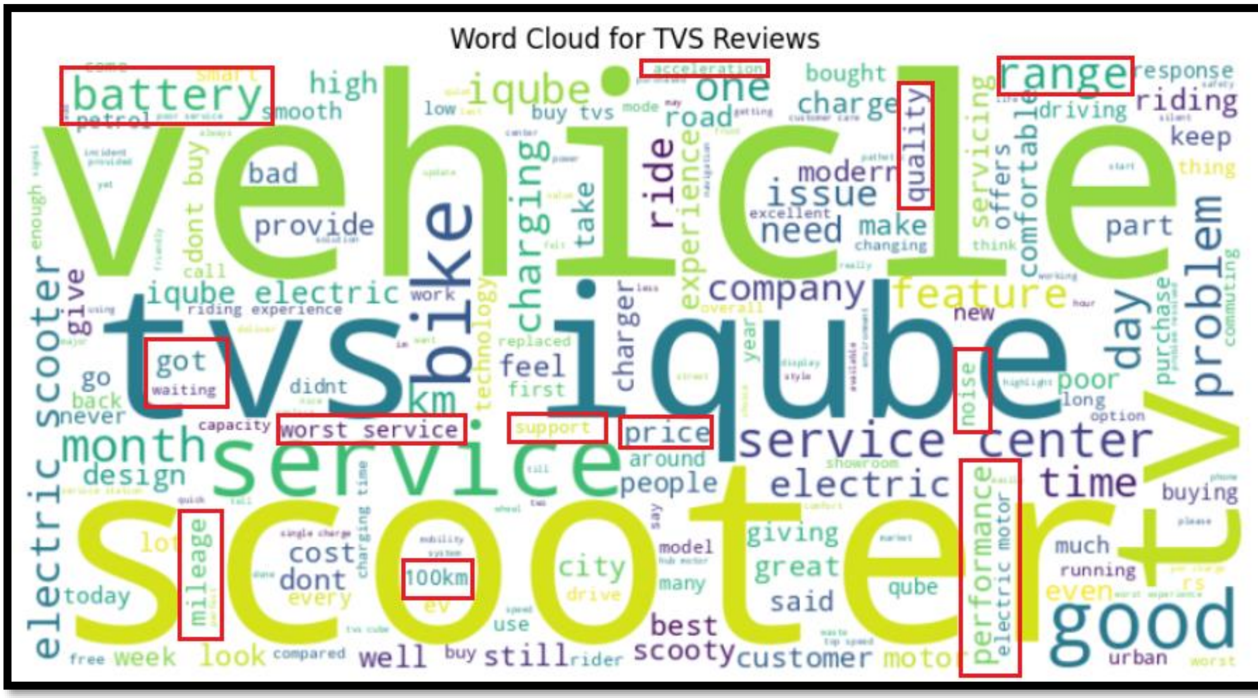


Fig. Word Cloud for reviews with negative sentiment score >=0.4

Upon careful analysis of the word cloud presented above, several noteworthy observations can be made regarding the concerns voiced by customers.

Primarily, terms such as **"worst service," "waiting," and "service center**" likely signify instances of subpar servicing experiences and inadequate support from the service team.

Additionally, words like **"battery," "range," and "mileage"** suggest potential issues with the power components of the scooter, including discrepancies in mileage performance or rapid battery depletion.

Furthermore, terms such as **"quality," "noise," "performance," and "acceleration"** hint at possible product-level defects causing inconvenience to customers. These observations underscore the importance of addressing these concerns to enhance customer satisfaction and product quality.

This initial analysis serves as a foundation for further exploration. Subsequent steps involve delving deeper into the reviews to extract and analyze specific topics.

5.Topic Modelling:

Topic modelling in our context serves as a valuable technique to uncover underlying themes and patterns within customer reviews. By analyzing the distribution of words and phrases, topic 9odelling algorithms such as Latent Dirichlet Allocation (LDA) help identify distinct topics, enabling a deeper understanding of customer sentiments, preferences, and concerns across the dataset.

We go through below set of steps to perform Topic modelling :

Step 1: TF-IDF Vectorization

Import the TfidfVectorizer class from scikit-learn to transform company reviews into TF-IDF features. Set parameters like max_df, min_df, and stop_words to filter out common and irrelevant words, then apply the fit_transform() method to generate a TF-IDF matrix.

Step 2: LDA Model Training

Import the LatentDirichletAllocation class from scikit-learn, and use it to apply the LDA algorithm for topic extraction from the TF-IDF matrix. Specify the number of topics with the n_components parameter, and fit the LDA model to the TF-IDF matrix using the fit_transform() method to obtain topic distributions.

Step 3: Extract Top Keywords for Each Topic

Define a function, get_top_keywords, to extract top keywords for each topic by:

- Looping over each topic

- Retrieving the indices of the top words

- Utilizing the TF-IDF vectorizer's vocabulary to obtain the actual words corresponding to these indices

- Returning a list of lists containing the top keywords for each topic

Step 4: Perform TF-IDF Analysis and LDA for Each Company and Print Topics and Keywords

Print the topics and their keywords.For each topic, print its number along with keywords.Iterate over each company's data.Print the company's name followed by "Analysis:".Call the tfidf_and_lda function to analyze the data using TF-IDF and LDA.

Results:

```
BAJAJ Analysis:
LDA Topics and Keywords:
Topic 1: service, scooty, better, good, scooter, chetak, best, problem, worst, customer
Topic 2: good, scooter, chetak, electric, bajaj, nice, drive, ride, range, best
Topic 3: vehicle, service, center, took, incompetence, properly, weeks, ages, zero, avoid
Topic 4: bike, amazing, driving, bajaj, scooter, cost, buy, experience, chetak, model
Topic 5: awesome, experience, smooth, chetak, purchased, body, bajaj, fixed, love, till

OLA Analysis:
LDA Topics and Keywords:
Topic 1: service, bike, issues, vehicle, ola, scooter, good, day, issue, year
Topic 2: scooter, good, s1, electric, pro, ola, features, experience, range, like
Topic 3: best, nice, scooty, india, new, pickup, electric, power, speed, bike
Topic 4: product, smooth, vehicle, nice, ola, better, awesome, scooty, road, bike
Topic 5: service, th, ola, centre, customer, scooter, ev, worst, lctric, dont

TVS Analysis:
LDA Topics and Keywords:
Topic 1: problem, tvs, service, buy, dont, vehicle, scooty, purchase, bike, center
Topic 2: good, poor, road, service, didnt, want, quality, noise, weeks, bike
Topic 3: service, battery, vehicle, company, worst, charging, servicing, time, scooter, best
Topic 4: electric, scooter, iqube, riding, features, ride, tvs, design, good, smooth
Topic 5: charges, mileage, high, mode, iqube, simply, hub, tvs, charging, use
```

Parameter Tuning: In the process of parameter tuning using **coherence score**, we begin by creating a Gensim dictionary from the tokenized text data, which maps unique **words to numerical IDs**.

Following this, we convert the TF-IDF matrix into a **Gensim corpus format**, representing documents as vectors of word frequencies.

Then, we evaluate the performance of the Latent Dirichlet Allocation (LDA) model with varying numbers of topics. Iterating over a range of topic numbers, we train an LDA model for each number of topics and compute the coherence score using the 'c_v' coherence metric, which measures topic interpretability.

These coherence scores are stored for each number of topics and plotted against the number of topics to visualize coherence score trends. Finally, the optimal number of topics is determined by selecting the number that corresponds to the highest coherence score, providing the most coherent and interpretable topics.

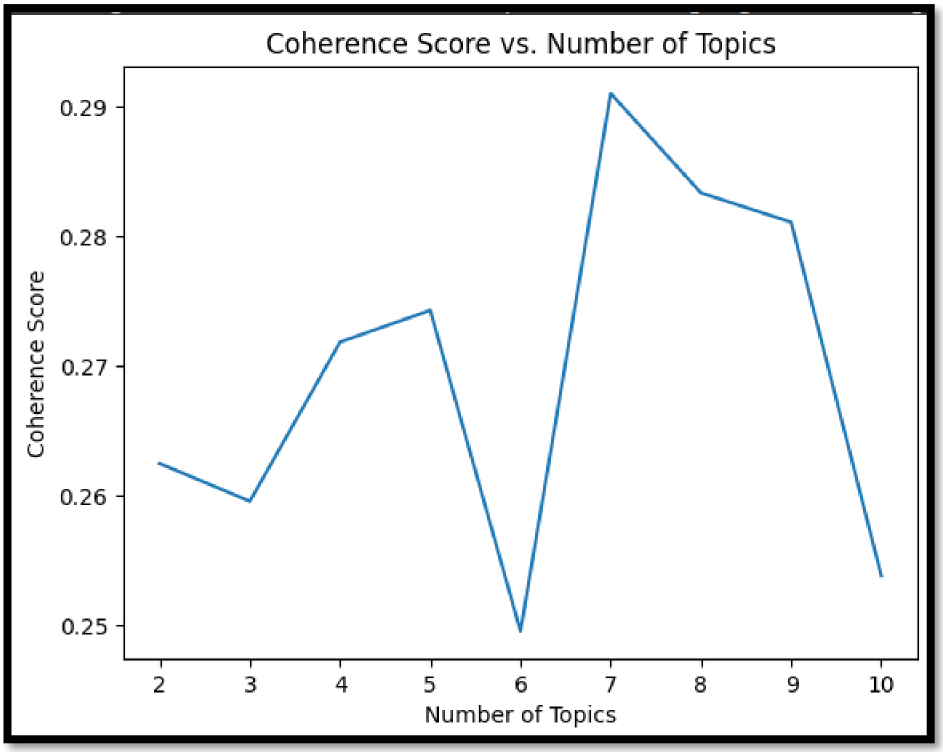


Fig. Parameter tuning plot for no. of topics

The optimal number of topics come out to be 7.

We train the LDA model again with optimal number of topics, the optimality result is below.

```
BAJAJ Analysis:
LDA Topics and Keywords:
Topic 1: service, chetak, zero, staff, problem, new, worst, customer, days, old
Topic 2: good, bike, nice, super, best, charging, scooter, drive, driving, chetak
Topic 3: service, vehicle, center, took, scooty, incompetence, properly, weeks, ages, issue
Topic 4: electric, scooter, bajaj, chetak, best, india, like, design, space, price
Topic 5: awesome, smooth, experience, built, body, driving, good, till, class, easy
Topic 6: buy, looking, bikes, ev, bad, vehicle, range, available, dont, poor
Topic 7: chetak, scooter, good, comfortable, drive, better, bajaj, ride, absolutely, best

OLA Analysis:
LDA Topics and Keywords:
Topic 1: bike, vehicle, ola, service, issues, worst, day, scooter, customer, center
Topic 2: good, scooter, electric, s1, pro, ola, best, range, features, experience
Topic 3: new, highly, multiple, breakdown, know, plan, 80, cost, used, worst
Topic 4: scooty, goes, repair, accident, warranty, ola, dont, centre, old, awesome
Topic 5: service, ev, year, ola, till, scooter, suggest, dont, centre, buy
Topic 6: product, service, customer, care, peace, complaints, running, buy, ola, issue
Topic 7: th, available, lctric, urban, dsign, scootr, smooth, faturs, driving, powrful

TVS Analysis:
LDA Topics and Keywords:
Topic 1: service, tvs, problem, center, vehicle, dont, purchase, buy, motor, iqube
Topic 2: iqube, electric, ride, scooty, smooth, design, tvs, riding, features, fast
Topic 3: battery, scooter, issues, electric, problems, service, ev, drive, best, replacement
Topic 4: provides, modern, acceleration, features, electric, scooters, market, smartphone, scooter, iqubes
Topic 5: poor, service, quality, charging, high, getting, mode, charges, actually, use
Topic 6: good, scooter, range, service, tvs, vehicle, mileage, bike, time, like
Topic 7: servicing, worst, kms, company, customers, extremely, speed, cost, pay, care
```

Inference:

The inferred topics and associated keywords provide insights into the main themes discussed in the reviews of each company.

BAJAJ Analysis:

Topics	Inference
Topic 1	Issues related to service, customer problems, and dissatisfaction.
Topic 2	Positive sentiments regarding the performance and features of Bajaj bikes and scooters.
Topic 3	More issues related to service and incompetence in handling problems.
Topic 4	Positive sentiments about Bajaj's electric scooter, design, and pricing.
Topic 5	Positive experiences with smooth driving and built quality.
Topic 6	Negative sentiments about purchasing decisions, vehicle availability, and range.
Topic 7	Positive reviews about the comfort and driving experience of Bajaj Chetak scooters.

OLA Analysis:

Topics	Inference
Topic 1	Complaints about vehicle issues, service problems, and customer dissatisfaction.
Topic 2	Positive reviews about Ola's scooters, electric features, and riding experience.
Topic 3	Reports of breakdowns, costs, and dissatisfaction with service.
Topic 4	Issues related to repairs, accidents, and warranty concerns.
Topic 5	Suggestions against purchasing Ola scooters due to service and reliability issues.
Topic 6	Complaints about product quality, customer service, and ongoing issues.
Topic 7	Positive feedback on available features, design, and smooth driving experience.

TVS Analysis:

Topics	Inference
Topic 1	Complaints about service problems, vehicle issues, and purchasing decisions.
Topic 2	Positive sentiments about TVS iQube electric scooters, ride quality, and design.
Topic 3	Issues related to battery problems, service, and electric drive performance.
Topic 4	Positive reviews about modern features and advancements in TVS electric scooters.
Topic 5	Concerns about poor service quality, high charging costs, and usage issues.
Topic 6	Positive feedback on TVS scooters' range, mileage, and overall performance.
Topic 7	Complaints about servicing experiences, costs, and customer care dissatisfaction.

After conducting topic 12odelling, we find that users not only express negative sentiments but also highlight strengths of focal firms, making them stronger competitors in the future. The focal firm's management should focus on these strengths and customer preferences identified in topic 12odelling, while also learning from competitors' strengths to better penetrate the market.

Recommendations to Management of Focal Firm:

Based on sentiment analysis findings, TVS management should capitalize on their strengths while leveraging positive aspects of competitors.

<u>Comparative Recommendations:</u>

1. ***<u>Service Quality and Customer Support:</u>***

- **Comparison**: Ola and Bajaj have received positive feedback regarding service quality and customer support.
- **Recommendation**: TVS can learn from Ola and Bajaj's successful service strategies by investing in training programs for service personnel, implementing efficient service processes, and enhancing customer support channels. By improving service quality and responsiveness, TVS can increase customer satisfaction and loyalty.

2. ***<u>Electric Vehicle Performance and Innovation:</u>***

- **Comparison**: Both Ola and Bajaj have received praise for their electric scooter performance and innovation.
- **Recommendation**: TVS should focus on innovation and technology advancements in their electric vehicle lineup to match the performance and features offered by Ola and Bajaj. This includes investing in research and development to improve battery efficiency, range, and charging infrastructure. By enhancing electric vehicle performance, TVS can attract environmentally conscious consumers and compete effectively in the electric scooter market.

3. ***<u>Product Design and Aesthetics:</u>***

- **Comparison**: Bajaj has been commended for its attractive product design and aesthetics.
- **Recommendation**: TVS can prioritize product design enhancements to align with consumer preferences and market trends. By investing in sleek and modern designs, TVS can enhance brand appeal and attract a wider audience. Collaborating with designers and leveraging customer feedback can help TVS develop visually appealing and functional products that stand out in the market.

4. ***<u>Marketing and Branding Strategies:</u>***

- **Comparison**: Ola has effectively marketed its scooters, emphasizing their electric features and environmental benefits.
- **Recommendation**: TVS can enhance its marketing and branding efforts to emphasize the unique features and benefits of its products. By adopting innovative marketing strategies, such as social media campaigns, influencer partnerships, and experiential marketing events, TVS can increase brand visibility and attract new customers. Highlighting the sustainability and performance advantages of TVS scooters can differentiate the brand and drive consumer interest.

<u>Overall Recommendations:</u>

- ***<u>Enhance Product Features</u>*****: Invest in product innovation** and feature enhancements to stay competitive and meet evolving customer preferences.
- ***<u>Integration of Power Components</u>***: Address **potential issues related to power components** such as battery, range, and mileage. Focus on improving battery technology and optimizing mileage performance to ensure customer satisfaction and product reliability.
- ***<u>Product Quality Improvement</u>***: **Prioritize addressing concerns related to product quality, noise, performance**, and acceleration. Implement measures to rectify any defects or discrepancies, enhancing overall product quality and customer experience.
- ***<u>Improve Service Quality</u>***: Prioritize improving service quality and customer support to address issues and enhance overall customer experience.
- ***<u>Efficient Issue Resolution</u>***: Streamline customer support processes to efficiently address concerns related to service problems and product defects. Ensure prompt resolution of customer issues to foster trust and loyalty.
- ***<u>Monitor and Adapt</u>***: Continuously monitor customer feedback and market trends, adapting strategies accordingly to maintain a competitive edge**.**

References:

1. **Author: SESHADRI TIRUNILLAI and GERARD J. TELLIS* Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation**
2. **https://www.statista.com/statistics/271312/urbanization-in-india/**
3. **https://parivahan.gov.in/parivahan//en/content/vehicle-related-services**
4. **https://www.autocarindia.com/**
5. **https://www.zigwheels.com/**
6. **https://www.bikewale.com/tvs-bikes/iqube/reviews/page/3/**
7. **https://www.mouthshut.com/**
8. **https://www.bikedekho.com/**
9. **https://youtu.be/kyVPs7-LKkE?si=vVGW9MqtUEMlhOuR**
10. **https://youtu.be/ewSiObxMwb4?si=Hgo-yV8e9Cy0_my0**
11. **https://youtu.be/dau4XKAKUlI?si=jt8lPzdzCIYYDcEj**
12. **https://youtu.be/e10lm6yqyMg?si=LUstQyzCYW3IBVX3**
13. **https://youtu.be/1QNg9YOBzMI?si=buy-M7SF6GICIMQ6**
14. **https://youtu.be/5uKhkE0AuJ0?si=recHDKX-ZbxwGBmvx**
15. **https://www.olaelectric.com/**
16. **https://www.tvsmotor.com/electric-vehicle/tvs-iqube**
17. **https://www.chetak.com/**