AN ANALYSIS OF RIDESHARING QUALITY OF SERVICE PRE- AND AMID-COVID-19 PANDEMIC USING ADVANCED TEXT MINING TECHNIQUES—USA VS INDIA CASE STUDY ACROSS DIFFERENT GENDER AND AGE GROUPS

AmirMahdi Khabooshani, Research Assistant, School of Computer and Information Technology, Engineering, Sadjad University of Technology, 64 Jalal Al Ahmad St, Mashhad, Iran, (+98)939-858-6881, a.mkh000@gmail.com

Anae Sobhani, Ph.D., Assistant Professor, Barney School of Business, Harford University, Beatrice Fox Auerbach Hall, 200 Bloomfield Ave, West Hartford, CT 06117, United States, (+1) 617-580-1734, sobhani@hartford.edu

Wenxiang Xu, Graduate Research Assistant, Hangzhou Innovation Institute, Beihang University 18, Chuanghui Street, Binjiang District, Hangzhou, Zhejiang, 310052, China, (+86)198-2183-4125, xwxtom.163.com

Ting Fu, Ph.D., Corresponding Author, Research Professor, School of Transportation Engineering, Tongji University, 4800 Cao'an Highway, Shanghai, 201804, China, (+86)138-1681-0642, Fax: (+86)021-69585717, tingfu@tongji.edu.cn

Zahra Bour, Research Assistant, Barney School of Business, Harford University Beatrice Fox Auerbach Hall, 200 Bloomfield Ave, West Hartford, CT 06117, United States, (+98)910-511-2483, sarasb1010@gmail.com

Behnaz Raouf, Transportation Planner, Fairfax County, Washington DC-Baltimore Area, +15713989355, behnaz.rauofhassanzadeh@fairfaxcounty.gov

ABSTRACT

One of the factors influencing passengers' transport selection in ridesharing services is the quality of services. After COVID-19 hit, ridesharing services underwent huge challenges. To examine which factors influenced riders' opinion changes on the quality of ridesharing services, 73,528 tweets from January 1st, 2019 to April 30th, 2022 were collected. The Latent Dirichlet allocation (LDA) model is used for topic modeling and filters out topics related to ridesharing quality of service (QoS). The Bidirectional Encoder Representations from Transformers (BERT) model is conducted to determine riders' sentiment. The topic modeling results reveal that topics have a remarkable difference between the pre- and amid-pandemic eras and between the USA and India. Sentiment analysis results show that during the pandemic period, sentiment is more positive. The results of this study can contribute to delivering a higher QoS in the ridesharing industry.

Keywords: ridesharing, quality of service, COVID-19, sentiment analysis, topic modeling, Twitter data

INTRODUCTION

Overview

Sharing economy is defined as networks of people exchanging products and services for less money than they would via corporations (Berg & Fawn, 2016). The rising use of mobile internet and smartphones among customers has led to the development of numerous sharing economy-based business models (Eckhardt, Houston, Jiang, Lamberton, Rindfleisch, & Zervas, 2019; Zhou, Liu, Ryan, Wang, & Zhang, 2020). Industries such as vacation rentals (e.g., Airbnb, HomeAway), mobile food delivery applications (e.g., Talabat, DoorDash), health services (e.g., Doctor On Demand), online shopping (e.g., e-bay, Amazon), freelancing services (e.g., Fiverr), and transportation (e.g., Uber, Lyft) have evolved through the perspective of sharing economy (Laukkanen & Tura, 2020). Ridesharing is defined as the sharing of an automobile trip by two or more people traveling to a particular destination. Ridesharing is operated and arranged via different mobile applications such as Uber and Lyft. Uber, however, currently holds a dominant market share (Shaheen, Totte, & Stocker, 2018) and in many cities, it has both gained control over the ridesharing industry and inspired innovative initiatives for indigenous tech-based transportation companies (Mourdoukoutas, 2017). Providers of transportation services require great tools such as creativity and extensive information and analysis in evaluating the quality of service (QoS) in order to respond to the customers` needs in this crisis (SERAFIMOVA, 2020). Thus, numerous studies are being conducted on the COVID-19 pandemic`s implications on issues pertaining to ridesharing as a consequence of its arrival. There are several studies available that focus on evaluating service quality and customer satisfaction with ridesharing, especially comparing pre-pandemic and amid-pandemic conditions for different age, gender, and income groups with the help of online questionnaires. However, few studies have focused on comparing different nations to provide an understanding of how social factors like culture or the development level of the nation can affect service quality for ridesharing services during the pandemic (Barbieri, Lou, Passavanti, Hui, Lessa, Maharaj, Banerjee, Wang, Chang, Naik, Yu, Liu, Sikka, Tucker, Foroutan Mirhosseini, Naseri, Qiao, Gupta, Abbas, Fang, Ghasemi, Peprah, Goswami, Hessami, Agarwal, Lam, & Adomako, 2020). Hence, some new research is focusing on different forms of collection data other than online questionnaires to expand the information accessible to them (Vidya, Fanany, & Budi, 2015; Kumar & Zymbler, 2019; Habib & Anik, 2021; Morshed, Khan, Tanvir, & Nur, 2021). Social networking websites such as Facebook and Twitter are perfect examples of platforms with millions of active users posting information on various topics that contain some form of emotional (Kumar & Zymbler, 2019; Wisnu, Afif, & Ruldevyani, 2020; Habib & Anik, 2021). In this study, we are using two techniques of Topic modeling and Sentiment analysis for analyzing users` tweets. Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information (n.d.-a). Analyzing information posted on social media platforms with these text mining techniques can guide us through gaining a deeper understanding of topics and reaching relatively more clear results that were formerly far-fetched. (Lyu, Han, Luli, & Lyu, 2021) have shown that using such techniques on social media text can bring us great insight on users` opinions.

Research context

The goal of this study is to look at the opinions of a sizable community of ridesharing users in the Twitter application for the United States and India. The objective of this study is to compare user opinions on the quality of services pre- and amid-pandemic in the United States and India while considering users` age and gender. According to the data gathered, there were 27727 tweets from the United States and 2227 from India. The methods we chose for data analysis are Bidirectional Encoder Representations from Transformers model (BERT) and the Valence Aware Dictionary and sEntiment Reasoner (VADER) for sentiment analysis and Latent Dirichlet Allocation (LDA) for topic modeling. (Pournarakis, Sotiropoulos, Systems, & 2017, n.d.) for instance, used LDA to perform topic modeling for transportation services. This study developed and applied a Genetic Algorithm based on LDA to the task of categorizing tweets into the various topics, which enhanced the K-means clustering approach. (Sun, Huang, & Qiu, 2019) constructed an auxiliary sentence to transform TABSA (targeted aspect-based sentiment analysis) from a single sentence classification task to a sentence pair classification task based on BERT. The outcome demonstrates that BERT-pair significantly outperforms other models on the SentiHood dataset for aspect detection and sentiment analysis.

LITERATURE REVIEW

Various studies have been undertaken to help clarify how the QoS affects customer satisfaction and therefore people`s use of ridesharing services. Some studies have been done aimed at anticipating customers` intention to use ridesharing applications with some being conducted during the COVID-19 outbreak (Lavieri & Bhat, 2019; Brown & Williams, 2021; Garaus & Garaus, 2021; Rasheed Gaber &

Elsamadicy, 2021). While some other studies have mainly focused on how and what affects customer satisfaction and dissatisfaction (Maziriri, Mapuranga, Mushwana, & Madinga, 2020) with the use of online surveys and categorizing results based on age, gender, frequency of usage, and income (Balachandran & Hamzah, n.d.; Maziriri et al., 2020). Rasheed Gaber et al., propose that four key factors (performance expectancy, economic benefits, facilitating conditions, and social influence) determine customers` further use of ridesharing applications (Rasheed Gaber & Elsamadicy, 2021). Moreover, they indicated that effort expectancy, perceived infectability, and fear of COVID-19 do not influence customers` intention to use these applications in Egypt (Rasheed Gaber & Elsamadicy, 2021). Even though the effect of social influence and effort expectancy on intention to use the Uber application was not different between genders, both younger users below 24 and females showed a high correlation with performance expectancy which shows their desire for potential benefits or rewards (Lee, Ruane, Lim, Zhang, & Shin, 2021). In order to assess the quality of services, some studies used the five dimensions from the SERVQUAL model (Kuswanto, Sundari, Harmadi, & Hariyanti, 2020; Shetu & Hamid, 2021) which are: responsiveness, assurance, tangibility, empathy, and reliability (Parasuraman, A., Ziethaml, & V., & Berry, 1985). Kaysher Hamid proposes that in the case of ridesharing services in Dhaka city reliability, assurance, and empathy are significant constructs where tangibility and responsiveness are found to be statistically insignificant (Shetu & Hamid, 2021). For Indonesian ridesharing services trust has a strongly positive effect on satisfaction and loyalty, while satisfaction positively influences loyalty (50). Users freely express their opinions, concerns, thoughts, and emotions on social networking platforms and that through comparative sentiment and emotion analysis they can provide insights into the understanding of the impact of COVID-19 on transportation (Al-Otaibi, Alnassar, Alshahrani, Al-Mubarak, Albugami, Almutiri, & Albugami, 2018). While many researchers have analyzed the elements and changes in service quality that affect customer satisfaction with ridesharing services in the COVID-19 era, detecting the changes in different age groups and genders while comparing two very different nations with the help of a relatively new type of data collection and analysis of service users` behavioral patterns using sentiment analysis enable us to determine emotions toward specific aspects such as tangibility and responsiveness, is a crucial move forward for understanding how ridesharing businesses can increase the satisfaction level of their customers and therefore sustainability of shared mobility. This study aims to provide a deeper understanding of how the COVID-19 outbreak affected users' perceptions of the quality of ridesharing services in India and the United States.

METHODOLOGY FRAMEWORK

Figure 1 illustrates the flowchart of the methodology. The framework consists of three parts: I) Data collection. The tweets related to ridesharing and their users' information are collected in this part. II) Ridesharing QoS-related topic modeling. The QoS-related keywords are extracted in this part based on the word frequency sort and correlation analysis using the flash-text method (Singh & Co, 2017) with the help of the SERVQUAL model (Parasuraman, A. et al., 1985). Then the dataset of tweets related to QoS is built based on the keywords with the help of topic modeling. To investigate the influences of users' characteristics and also the COVID-19 pandemic on the topics discussed in the tweets, the topics' differences among each gender, age, and country groups pre-and post-pandemic are compared. III) Sentiment analysis of ridesharing QoS. In this part we use VADER and BERT models to identify the sentiment of each tweet, and based on model performance, a more suitable model is chosen for further analysis. The changes in sentiments of tweets for each gender and age group in the USA and India from pre-pandemic to amid-pandemic stages are then compared, using sentiment and significant analysis.

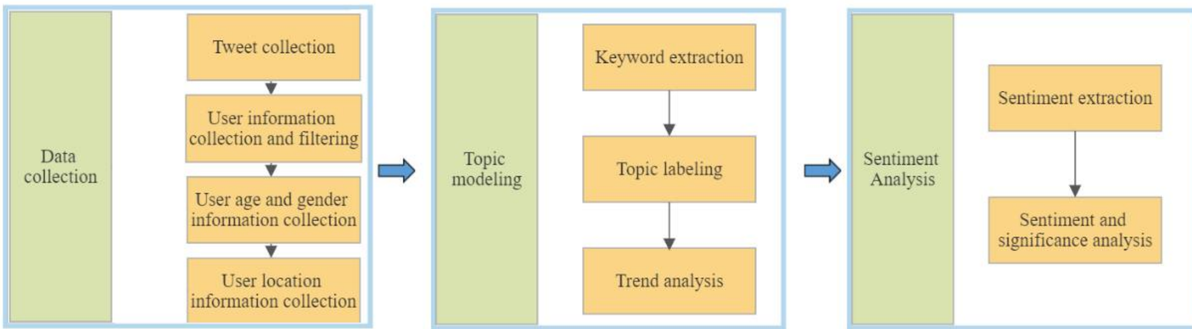


FIGURE 1 FLOWCHART OF THE METHODOLOGY FRAMEWORK

Data Collection

User information collection and filtering: The information about every user of the gathered tweets, including their profile pictures` URLs, location, and the number of followers is collected using a Python library for accessing the Twitter API called Tweepy (n.d.-b). Users` profile picture URLs are collected in order to gather age and gender information in part III. User age and gender information collection: Users` age and gender information are gathered based on their profile pictures with the help of the Microsoft Azure FACE service (n.d.-c). In this article, FACE identification was run on each user`s profile picture and collected age and gender for each one. If the profile picture contained no faces or more than one face, its user was flagged as a user with no age and gender information.

TABLE 1 DESCRIPTION OF THE STRUCTURE OF THE GATHERED TWEETS

Data	Description
username	Twitter username of the user.
date	The exact date and time of the tweet submission.
tweet	Text of the tweet.
location	User-specified location.
avatar	The user’s profile picture URL.
followers	The number of user’s followers.
age	The age gathered from user’s profile picture.
gender	The gender gathered from user’s profile picture.
country	The user’s country gathered from user’s location.

Topic Modeling of Ridesharing Service

To extract the related tweets, the words with the highest frequency in the original dataset are gathered and then sorted based on the flash-text method in Python. After collecting the tweets, 18,440 tweets related to ridesharing QoS based on 16,088 Twitter users are kept for analysis. This study chooses the age of 35 (only using this threshold, the data shows the significance) to divide the users into younger adults (between 18 to 35 years old) and older adults (36+ years old) categories. To obtain the underlying structure of latent topics in our dataset based on LDA, Python`s Gensim library (Řehůřek & Sojka, 2010) is used. This paper concentrates on exploring the topic difference between the country, gender, and age groups pre- and post-pandemic, the data is divided into 7 groups (all tweets, USA-based tweets, India-based tweets, male user tweets, female user tweets, younger adult user tweets, and older adult user tweets). Then, 14 documents, which include 7 groups, pre-, and post-pandemic from January 2019 to April 2022, are used for topic modeling based on the LDA model (Blei, Ng, & Jordan, 2003). We ran the LDA algorithm on the data by varying the topic number from 2 through 20. Some studies considered the use of several clustering techniques to group keywords into predefined topics, such as the Moreno (Moreno & Iglesias, 2021) method, which integrates the topics based on the LDA results using the K-means clustering algorithm, a genetic algorithm, and a local convergence algorithm. Yet, grouping the topics using such methods suffers from the same drawback; researchers still need to manually label the clustering result's topics. The requirement to pre-define the number of topics is the main flaw of the supervised method since it decreases the informational value of text and the underlying variables of topics. So, we chose the tweets related to these clusters as the QoS tweets.

Sentiment Analysis

In the present study, the VADER and BERT models are used for identifying the sentiment of each tweet. VADER Model: VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) that are generally labeled according to their semantic orientation as either positive or negative. VADER not only tells about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is. The correlation coefficient shows that VADER ($r = 0.881$) performs as well as individual human raters ($r = 0.888$) at matching ground truth (aggregated group mean from 20 human raters for sentiment intensity of each tweet). The model is composed of one or more input sequences, added with an initial token “CLS” and a token “SEP” to separate segments.
In this paper, the BERT base model employs $L = 12$, $A = 12$ and $H = 768$. Normally, the BERT takes an input of a sequence of no more than 512 tokens (which is lowered here to 128 due to the small length of tweets).

RESULTS

Topic Modeling Performances and Results

Keywords distribution and results: We downloaded 73,528 tweets containing the ridesharing keywords which are mentioned in the methodology section posted from 1/1/2019 to 4/5/2022. After cleaning, we got a total of 69,802 tweets from 56,780 unique users. Then, each tweet related to ridesharing QoS is chosen from the original dataset with the help of topic modeling. After collecting and constructing the ridesharing QoS tweet dataset, we calculated the frequency of tweets for each month, which is demonstrated in **Figure 2**. The results show that the number of tweets decreased significantly from January 2020 to April 2020. This decrease is most likely due to the decrease in users of ridesharing services due to the fear of catching the COVID-19 virus and later on due to companies suspending the option of shared rides concerning prevention of the spread of the virus (n.d.-d). The trend then shows a stable wave. This is probably due to the fact that ridesharing companies did not bring back shared rides until the end of the data collection time and riders still preferred using ridesharing less frequently (n.d.-e; n.d.-f; n.d.-d).

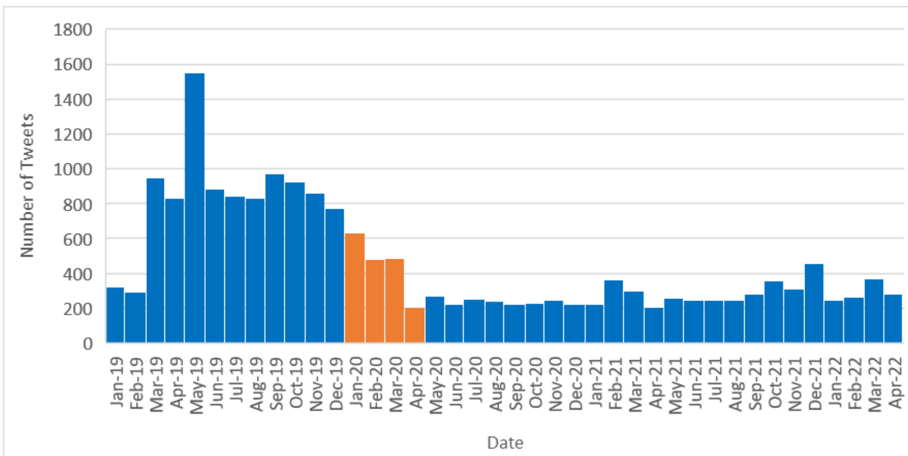


FIGURE 2 MONTHLY NUMBERS OF RIDESHARING QOS–RELATED TWEETS FROM JAN-19 TO APR-22

The frequency of keywords in each of the three categories (country, gender, and age) is also collected for checking how the most prevalent and important issues surface on the platform. As can be seen in **Figure 3**, word clouds show each word in the dataset proportionally scaled according to its frequency. The most frequent words in all groups before and amid the pandemic were words like "price" and "rider". This suggests that the users were mainly concerned with the price of the rides and the behavior of other riders. It appears that people in the USA are more concerned about the price of the services than those in India. But Indian people are more concerned about shared rides being suspended amid the pandemic than the American people. Before the pandemic, male users tend to think about the price of the rides much more than female users, but it seems that amid the pandemic, female users have a higher concern about the price matter. Comparing the age groups, we can see that older users are more concerned than younger ones about the vehicles that they are riding. Also, younger users seem to care more about the people that they are sharing the car with.

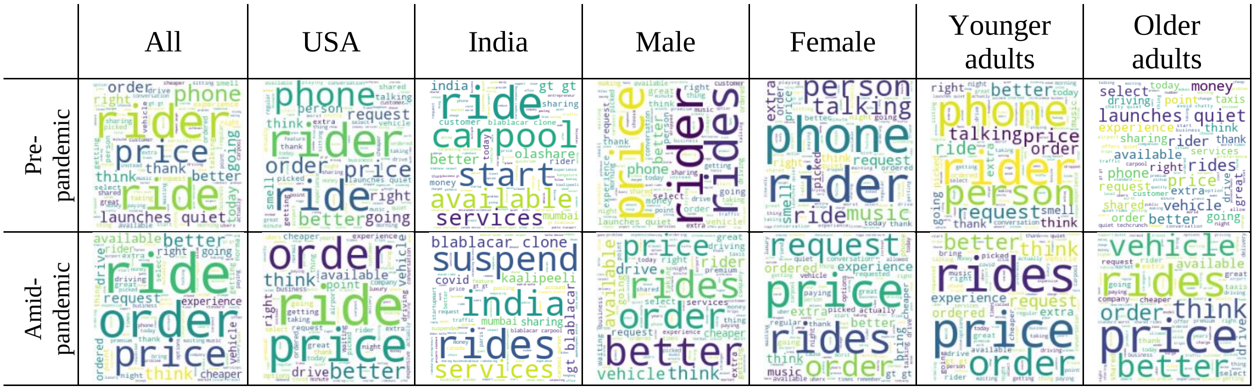


FIGURE 3 WORDCLOUDS FOR EACH GROUP PRE- AND AMID-PANDEMIC

Topic modeling and development trend analysis: As mentioned in the methodology section, we collected keywords related to ridesharing QoS based on the flash-text method with the help of the SERVQUAL model (Parasuraman, A. et al., 1985). We chose 4 dimensions from the SERVQUAL model (Tangibles, Responsiveness, Assurance, and Empathy) because we saw more terms associated with them in our QoS tweets. To perform the task of extracting the most discussed topics in each group pre-and post-pandemic, the clusters of seven groups (all-USA-India-male-female-younger-older) are extracted based on the LDA model. Then, the clusters are labeled based on the proposed method. As a result, we have 4 topics labeled. The details of the topics and their content can be seen in **Table 2**.

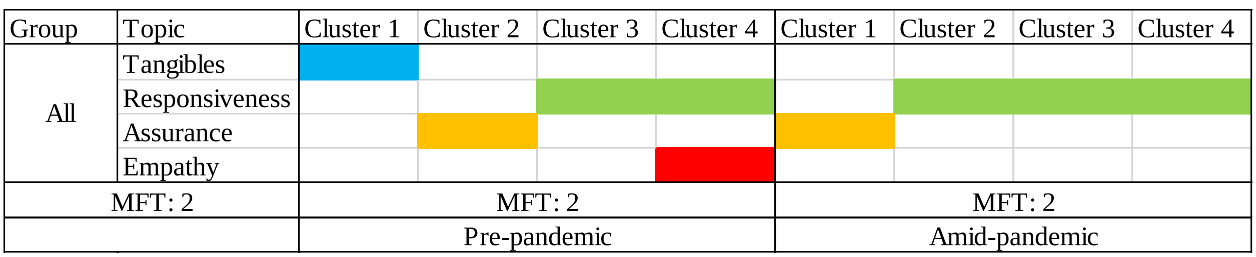
TABLE 2 THE LABELED TOPICS AND THEIR DESCRIPTION

Item	Number	Topic label	Description
Ridesharing QoS	1	Tangibles	physical facilities (Haddad, Nagy, & Popp, n.d.)
	2	Responsiveness	caring, paying personal attention, and providing services to customers (Haddad et al., n.d.)
	3	Assurance	employees' courtesy and knowledge, and their capacity to transfer confidence and trust in customers (Haddad et al., n.d.)
	4	Empathy	Customers need to feel that they are made a priority by the organization providing services (Haddad et al., n.d.)

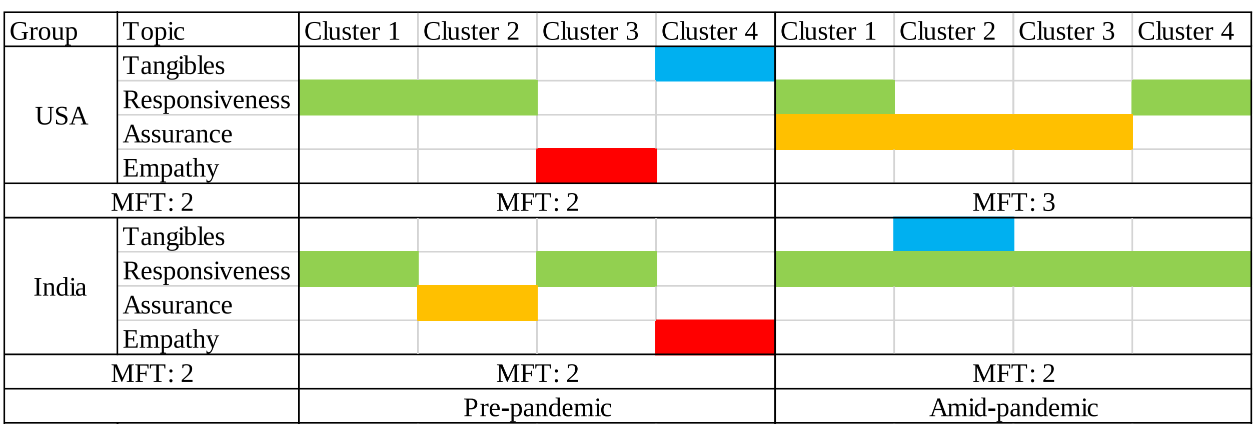
By adapting (Haddad et al., n.d.) descriptions of 4 dimensions of QoS.

The topic distribution of the four clusters is indicated in the trend graphs in **Figure 4**. In these graphs, each cluster is assigned one or more topics by filling out the corresponding topic cell below it. The left side of

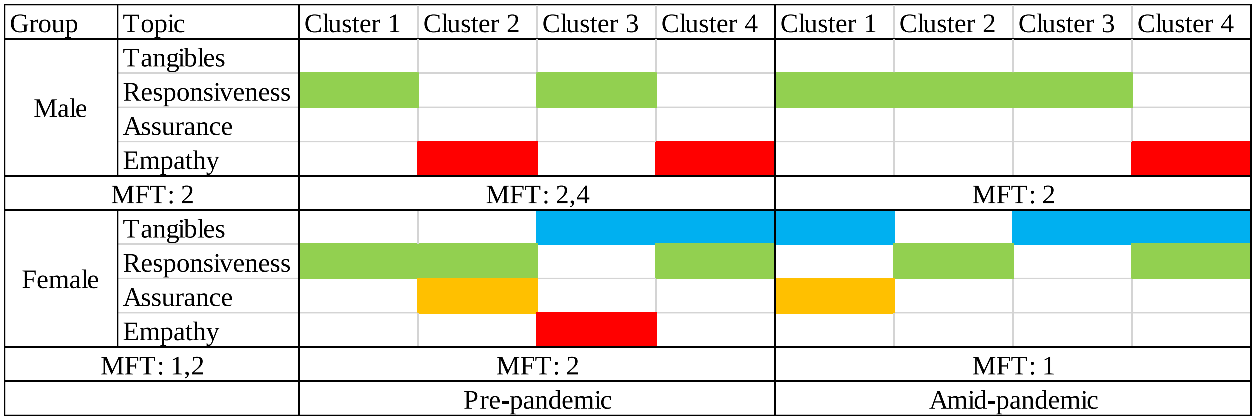
the graph depicts the groups and the ridesharing QoS labels. The center of the graph illustrates the distribution of the topics in each group. The MFT in each graph indicates the most frequent topic in each group and also for the pre-pandemic and amid-pandemic periods. As demonstrated by the graphs, the second topic (Responsiveness) is discussed more in all the groups during both pre- and amid-pandemic periods. The important finding in the understanding of the data in this figure is that in almost every group, the second topic (responsiveness) is more often discussed amid-pandemic than in the pre-pandemic period. Comparing the USA and India, it is clear that Americans are more worried about topic 3 (assurance) during the pandemic than Indians, who are more concerned with responsiveness during the time. Indian people tend to talk more about tangibles during the pandemic, however people in the USA talked about it more before the pandemic.
The gender groups' graph reveals that female users are more willing to discuss tangibility than male users who tweeted more about empathy. Females also reflect on assurance, whereas males don't seem to be as concerned about the topic.
In both the younger and older adult groups, we see that topic 2 (responsiveness) has gained more attention amid-pandemic. But the older users seem to have more interest in the topic, whereas younger ones are still concerned about topics 1 (tangibles) and 4 (empathy).



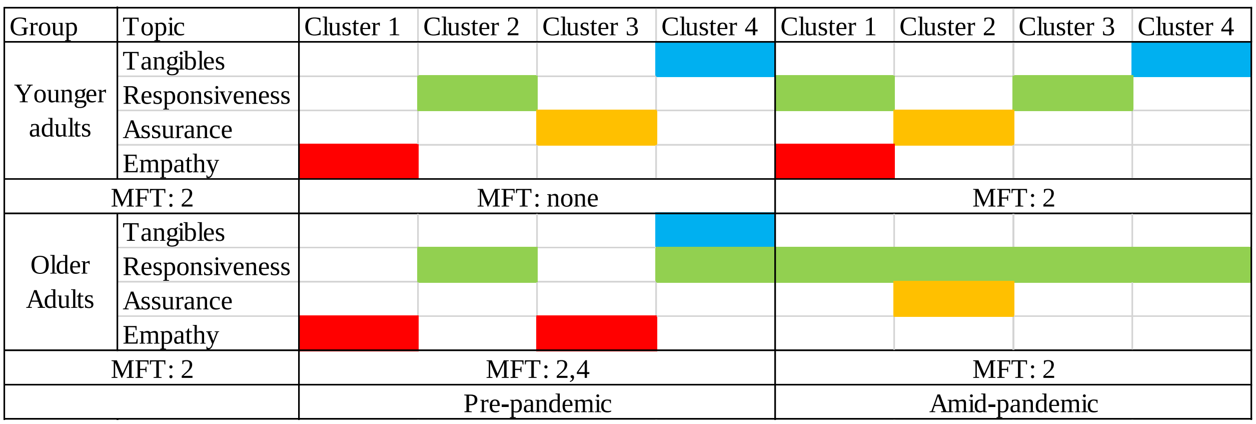
a) trend graph of topics related to all groups



b) trend graph of topics related to USA and India groups



c) trend graph of topics related to male and female groups



b) trend graph of topics related to younger and older adult groups

FIGURE 4 TOPIC MODELING CLUSTERS AND THEIR LABELS

Sentiment and significance analysis of the tweets

The VADER and BERT Sentiment analysis models have been conducted to examine the sentiment in the tweets. The models have been implemented on a computer with an Intel (R) Xeon (R) Silver 4110 CPU @ 2.10 GHz processor and 64 GBs of RAM, as well as an IDE disk under the Centos 7.6 operating system. For developing the algorithms in Python, the Anaconda 2021.03 open-source software has been used. To fine-tune the models, NVIDIA V100 GPUs are used. The accuracy of the BERT model, proved higher than the VADER model (BERT = 87%, VADER = 61%). To further test the applicability of the models, a number of 400 tweet samples' sentiment is manually labeled. The results show that the BERT model still proved to be more accurate compared to the VADER model (BERT accuracy = 72.3%, BERT Mean Absolute Error = 0.15, VADER accuracy = 52%, VADER Mean Absolute Error = 0.32). The BERT model proved more effective at handling the sentiment analysis problem of tweets. In light of this, the BERT model is chosen as a basis for further analysis of the sentiment of the tweets.

Figure 5 shows the results obtained from running the BERT sentiment analysis model on the QoS tweets for each month. This result shows that negative expressions are more prevalent than positive ones in tweets relating to complaints and problems about the quality of services. The chart shows that the total percentage of negative tweets has subtly decreased and that positive tweets have seen a slight increase (mean negative tweet percentage pre-pandemic = 46.15%, amid-pandemic = 43.82%, mean positive tweet percentage pre-pandemic = 21.81%, amid-pandemic = 24.90%). Having inspected several sample tweets, it can be concluded that this phenomenon may be due to the fact that in the absence of shared rides, users start to realize that they are missing out on those services and that they miss them amid the pandemic.

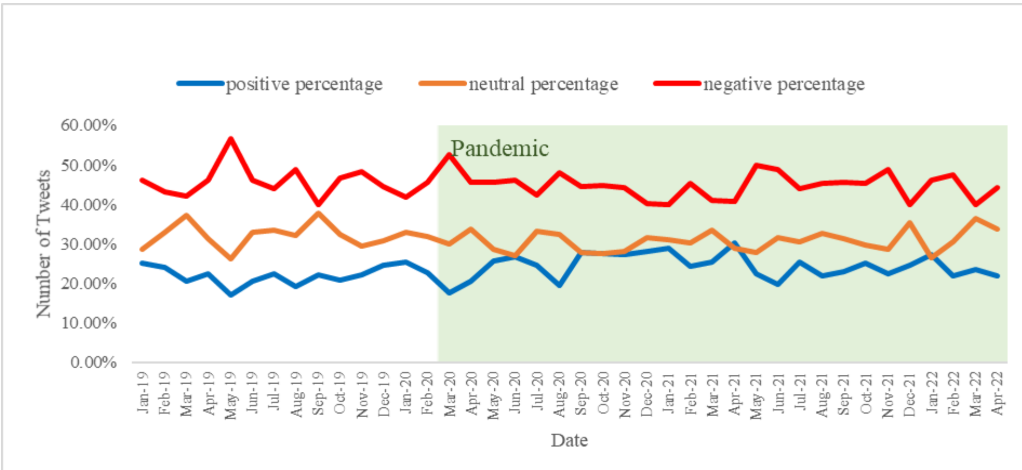


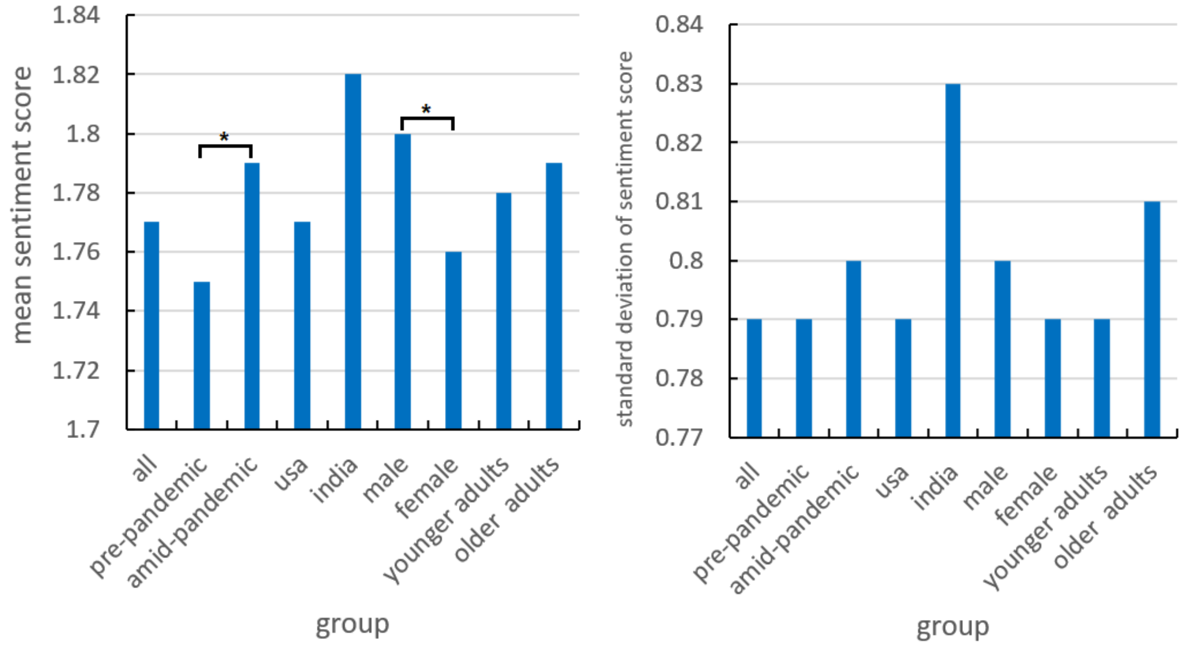
FIGURE 5 MONTHLY SENTIMENT PERCENTAGES FROM JANUARY 2019, TO APRIL 2022.

To deeper analyze the sentiment, positive enhancement is conducted, where negative, neutral, and positive sentiments are assigned values of 1, 2, and 3 respectively.

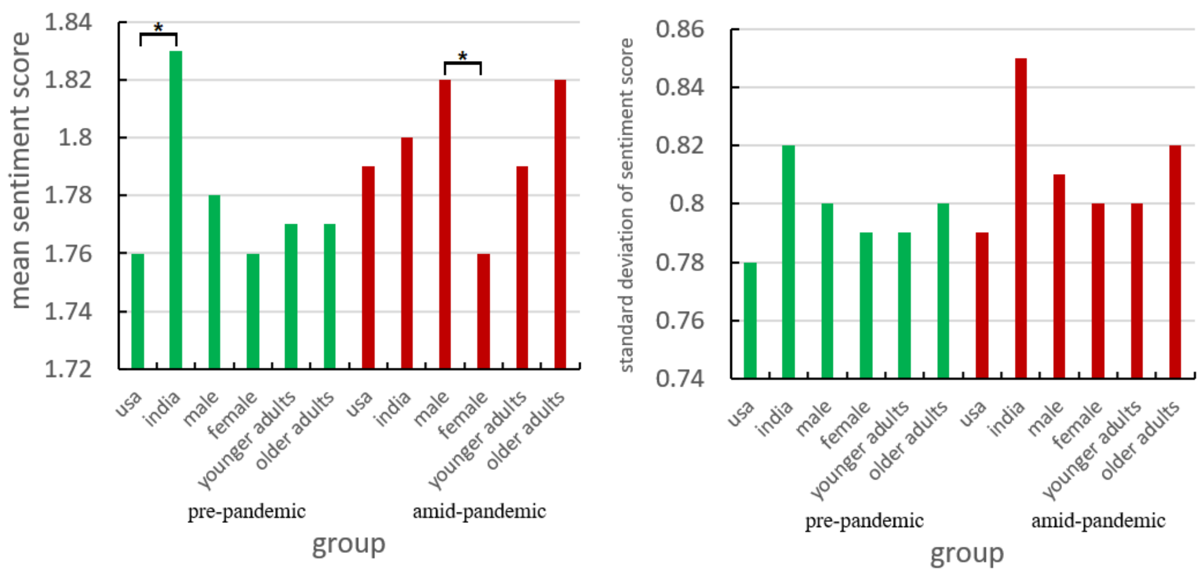
The mean and standard deviation (S.D.) of each group are compared to determine the extent to which there are differences in sentiment between groups. As shown in **Figure 6 (a)**, the sentiment between pre- and amid-pandemic periods, countries, and genders were found to vary significantly. The chart indicates that people have a more positive view of ridesharing QoS amid-pandemic as was also shown in **Figure 5**. We can see in the chart that Indians had a much more positive opinion of ridesharing QoS compared to Americans. Also, when it comes to gender, it is revealed that male users tend to be more positive towards QoS than females.

From the data in the S.D. chart, it is apparent that during the pre-pandemic era, users maintained a more stable sentiment. Also, the female, USA, and younger adult groups show a more stable attitude towards ridesharing QoS.

The differences between sentiment in each group pre- and amid-pandemic are then analyzed. As can be seen in **Figure 6 (b)**, there are significant differences in sentiment between the United States and India before the pandemic, but the difference becomes slight during the pandemic period. Conversely, for the gender groups, we see little difference between males and females in the pre-pandemic period, but for the amid-pandemic period, the difference grows significantly. We also see a growth in the difference between the two age groups' sentiment in the amid-pandemic era. The two groups see eye to eye pre-pandemic, where after the beginning of the pandemic, older users show a more positive attitude towards the ridesharing QoS compared to younger ones. Sentiment in all of the groups became more positive amid the pandemic except for the Indian group, which became less positive, and the female group, whose sentiment remained the same. As can be seen in the S.D. chart, all of the groups' attitude toward ridesharing QoS happens to be more stable pre-pandemic as S.D. grows in all of the groups amid-pandemic. The statistically significant differences are shown in the chart as * (p-value < 0.05).



a) The difference in sentiment between different groups



b) The difference in sentiment between different groups pre- and amid-pandemic

FIGURE 6 THE MEAN AND S.D. OF SENTIMENT IN EACH GROUP

CONCLUSIONS

As ridesharing has become increasingly crucial to people's everyday lives, it is important to investigate the possible effects that a global pandemic can have on the opinions of users regarding the quality of services. By employing topic modeling and sentiment analysis on users' tweets, this study has found out what dimensions of ridesharing QoS people tend to care more about during the period of the pandemic and how their opinions have changed after the pandemic hit the societies. The distribution of topics gathered by using topic modeling revealed that people are more concerned about the responsiveness of ridesharing companies in the pandemic era. This may be due to the fact that companies experienced many difficulties managing the services and striking a balance between customers' and drivers' health and gaining profit during the pandemic era, which resulted in users demanding a response from the companies about the services. By comparing the topic distribution, it is discovered that female users are more concerned about the tangibles of ridesharing in the amid-pandemic period while males are more attentive to the responsiveness factor at the same time. As for Americans and Indians, we see that Americans are talking more about the assurance they get from the ridesharing companies, while Indians put a higher priority on the responsiveness of the companies. Older adults also discuss responsiveness more than younger ones. The investigation of users' sentiments also showed a remarkable difference. The majority of users are more positive in their tweets in the amid-pandemic period. Indians have a more positive opinion on the quality of ridesharing services than American people. While in the gender groups, males seem to have higher positivity on the matter compared to females. We see that during the pandemic, older adults are more positive in contrast to younger adults. This study provides a rather new way of understanding the effects of the COVID-19 pandemic on people's opinions, using advanced text mining methods and using trend analysis on topics, and analyzing sentiment in the proposed manner, using Twitter data. The results of the present study could be used by ridesharing companies to understand how they can enhance users' experiences of the quality of their services and what their users' expectations could be of them.

This study was limited by the small number of tweets especially tweets from Indian people. Future studies should be carried out using more tweets to ensure the validity of the approach. Additionally, topic modeling should be combined with sentiment analysis. The correlation between sentiment and different variables such as age and gender could also be analyzed in order to understand which factors influence sentiment the most. Emotion analysis can also be employed to gather a deeper understanding of users' opinions.

REFERENCES

Bibliography

Al-Otaibi, S., Alnassar, A., Alshahrani, A., Al-Mubarak, A., Albugami, S., Almutiri, N., & Albugami, A.

(2018). Customer satisfaction measurement using sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *9*(2), 106–117.

Balachandran, I., & Hamzah, I. Bin. (n.d.). THE INFLUENCE OF CUSTOMER SATISFACTION ON RIDE-SHARING SERVICES IN MALAYSIA.

Barbieri, D. M., Lou, B., Passavanti, M., Hui, C., Lessa, D. A., Maharaj, B., Banerjee, A., Wang, F., Chang, K., Naik, B., Yu, L., Liu, Z., Sikka, G., Tucker, A., Foroutan Mirhosseini, A., Naseri, S., Qiao, Y., Gupta, A., Abbas, M., Fang, K., Ghasemi, N., Peprah, P., Goswami, S., Hessami, A., Agarwal, N., Lam, L., & Adomako, S. (2020). A survey dataset to evaluate the changes in mobility and transportation due to COVID-19 travel restrictions in Australia, Brazil, China, Ghana, India, Iran, Italy, Norway, South Africa, United States. *Data in Brief*, *33*, 106459.

Berg, C., & Fawn, F. (2016). How Brands Can Take Advantage of the Sharing Economy, (97), 1–9.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation, *3*, 993–1022.

Brown, A., & Williams, R. (2021). Equity Implications of Ride-Hail Travel during COVID-19 in California: *https://doi.org/10.1177/03611981211037246*, 036119812110372.

Eckhardt, G. M., Houston, M. B., Jiang, B., Lamberton, C., Rindfleisch, A., & Zervas, G. (2019). Marketing in the Sharing Economy. *Journal of Marketing*, *83*(5), 5–27.

Garaus, M., & Garaus, C. (2021). The Impact of the Covid-19 Pandemic on Consumers' Intention to Use Shared-Mobility Services in German Cities. *Frontiers in Psychology*, *12*, 367.

Habib, M. A., & Anik, M. A. H. (2021). Impacts of COVID-19 on Transport Modes and Mobility Behavior: Analysis of Public Discourse in Twitter. *Transportation Research Record: Journal of the Transportation Research Board*, 036119812110299.

Haddad, H., Nagy, J., & Popp, J. (n.d.). The Service Quality Dimensions that Affect Customer Satisfaction in the Jordanian Banking Sector, 1–24.

Kumar, S., & Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data*, *6*(1).

Kuswanto, A., Sundari, S., Harmadi, A., & Hariyanti, D. A. (2020). The determinants of customer loyalty in the Indonesian ride-sharing services: offline vs online. *Innovation and Management Review*, *17*(1), 75–85.

Laukkanen, M., & Tura, N. (2020). The potential of sharing economy business models for sustainable value creation. *Journal of Cleaner Production*, *253*, 120004.

Lavieri, P. S., & Bhat, C. R. (2019). Investigating objective and subjective factors influencing the adoption, frequency, and characteristics of ride-hailing trips. *Transportation Research Part C: Emerging Technologies*, *105*, 100–125.

Lee, C. C., Ruane, S., Lim, H. S., Zhang, R., & Shin, H. (2021). Exploring the Behavioral Intention to Use Collaborative Commerce: A Case of Uber. *Journal of International Technology and Information Management*, *30*(5).

Lyu, J. C., Han, E. Le, Luli, G. K., & Lyu, J. C. (2021). COVID-19 Vaccine – Related Discussion on Twitter : Topic Modeling and Sentiment Analysis Corresponding Author :, *23*, 1–12.

Maziriri, E. T., Mapuranga, M., Mushwana, J., & Madinga, N. W. (2020). Antecedents that influence the intention to use the uber mobile application: Customer perspectives in South Africa. *International Journal of Interactive Mobile Technologies*, *14*(8), 76–96.

Moreno, A., & Iglesias, C. A. (2021). Understanding Customers' Transport Services with Topic Clustering and Sentiment Analysis. *Applied Sciences*, *11*(21).

Morshed, S. A., Khan, S. S., Tanvir, R. B., & Nur, S. (2021). Impact of COVID-19 pandemic on ride-hailing services based on large-scale Twitter data analysis. *Journal of Urban Management*, *10*(2), 155–165.

Mourdoukoutas, E. (2017). Africa's app-based taxis battle Uber over local market share. *Africa Renewal*, *31*(2), 20–21.

Parasuraman, A., Ziethaml, & V., & Berry, L. (1985). SERVQUAL: A multiple item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *62*(1), 12–40.

Pournarakis, D., Sotiropoulos, D., Systems, G. G.-D. S., & 2017, undefined. (n.d.). A computational model for mining consumer perceptions in social media. *Elsevier*.

Rasheed Gaber, H., & Elsamadicy, A. M. (2021). What drives customers to continue using ride-sharing apps during the COVID-19 pandemic? The case of Uber in Egypt. *Cogent Business and Management*, *8*(1).

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA, 45–50.

SERAFIMOVA, T. (2020). COVID-19 : an opportunity to redesign mobility towards greater sustainability and resilience?

Shaheen, S. P., Totte, H., & Stocker, A. (2018). UC Berkeley UCCONNECT Title Future of Mobility White Paper Permalink https://escholarship.org/uc/item/68g2h1qv.

Shetu, S. N., & Hamid, K. (2021). SERVQUAL Model Dimensions and Customer Satisfaction towards Ride-Sharing Services in Dhaka City. *Journal of Innovation in Business Studies*, *1*(1).

Singh, V., & Co, B. (2017). Replace or Retrieve Keywords In Documents at Scale.

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 380–385.

Vidya, N. A., Fanany, M. I., & Budi, I. (2015). Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers. *Procedia Computer Science*, *72*, 519–526.

Wisnu, H., Afif, M., & Ruldevyani, Y. (2020). Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes. *Journal of Physics: Conference Series*, *1444*(1).

Zhou, B., Liu, T., Ryan, C., Wang, L. en, & Zhang, D. (2020). The satisfaction of tourists using bicycle sharing: a structural equation model - the case of Hangzhou, China. *Journal of Sustainable Tourism*, *28*(7), 1063–1082.

(N.d.-a). Sentiment analysis - Wikipedia, accessed October 7, 2022, available at https://en.wikipedia.org/wiki/Sentiment_analysis.

(N.d.-b). tweepy/tweepy: Twitter for Python!, accessed July 30, 2022, available at https://github.com/tweepy/tweepy.

(N.d.-c). Facial Recognition | Microsoft Azure, accessed July 30, 2022, available at https://azure.microsoft.com/en-us/services/cognitive-services/face/.

(N.d.-d). Uber And Lyft Have Suspended Uber Pool And Shared Rides Features Due To The Coronavirus, accessed July 31, 2022, available at https://www.buzzfeednews.com/article/ryanhatesthis/uber-suspends-uber-pool-due-to-coronavirus.

(N.d.-e). Uber is doing 70 percent fewer trips in cities hit hard by coronavirus - The Verge, accessed July

25, 2022, available at https://www.theverge.com/2020/3/19/21186865/uber-rides-decline-coronavirus-seattle-sf-la-nyc.

(N.d.-f). Lyft - Statistics & Facts | Statista, accessed July 25, 2022, available at https://www.statista.com/topics/4919/lyft/#dossierContents__outerWrapper.