Thi Phan

INFO 5082.002

Professor: Dr. Junhua Ding

Sep 26, 2021

Project Report

The Covid-19 Impact of Small Business – Shipping company

1. Introduction

The COVID-19 pandemic has turned the world upside down and caused a devastating loss of life. It has disrupted lives, pushed the hospital system to its capacity, and created a global economic slowdown. The pandemic created a financial shock, a demand shock, and a supply shock. It leads to swings in the shipping industry and causes disruption of the supply chain including the movement of essential goods and foodstuffs. I used to work at a shipping company and worked with many trucking companies. I knew they have faced many difficulties in the pandemic. In this study, I will do the research to see how the COVID-19 impact small businesses, especially shipping companies.

In this study, I used data from ABC company and these data describe the daily sales of all agents in the ABC system. I would like to explore the sales distribution through the peak of the COVID-19 pandemic and sales forecast for the next several months. The calculations were conducted in the Python environment using the main packages pandas, numpy, sklearn, keras, matplotlib, seaborn, nltk, and Jupiter Notebook to conduct the analysis.

The main analysis in this study is sales prediction. Sales prediction is very important to organizations, especially in COVID-19. Through this project, I have learned a lot about sales prediction and sales forecasting machine learning algorithms. To compare with before, I must state the problems and sold them.

2. Statement of the Problem

First, I need to know how many agents, services, and destinations are in the dataset. I would like to see the sales report to compare the revenue before and after (controlled) COVID-19. From the trend of sales over the months, I will build the model for prediction for the last 3 months in 2021.

3. Literature review

The impact of COVID-19 on small business owners: Evidence from the first three months after widespread social-distancing restrictions

The early effects of COVID-19 on small businesses and entrepreneurs are not well known because of the lack of timely business-level data released by the government. Most major industries faced large drops in the number of active business owners in April with the only exception being agriculture. Construction, restaurants, hotels, transportation, and personal/laundry services all faced large declines in the number of active business owners due to COVID-19. Simulations reveal that the concentrations of female, black, Latin, and Asian businesses in industries hit hard by the pandemic contributed to why losses in business activity were higher for these groups than the national average loss in April. May and June brought a partial rebound for most industries. These first estimates of impacts of COVID-19 on small businesses from the April 2020 CPS indicate that losses were spread across demographic groups and types of business—no group was immune to negative impacts of social-distancing policy mandates and demand shifts. However, they also reveal a partial bounce back for all groups. Although there is no way to

know at this time if these business closures will be permanent each month of inactivity has an impact on the revenues, profits, and employees of these businesses. They use 2 survey methods to do the research – current population survey and survey timing and social-distancing restrictions. According to the paper result, African-American business owners were hit the hardest by COVID-19. Latinx businesses were also hit hard by COVID-19 losing 32% of active business owners in April, 19% in May, and 10% in June. Asian business owners experienced a 26% decline in business activity over the critical 2-month window and continued losses of activity of 21% in May and 10% in June.

COVID-19 Impact on the Sharing Economy Post-Pandemic

The outbreak of the COVID-19 pandemic has disrupted the global economy completely, with its unprecedented effect across the globe. The effect of this pandemic is seen across all the economic domains, both formal and informal sections (Santos, 2019). The pandemic has disrupted the informal economy including the rental market – peer-to-peer rental market. The world economy is progressing slowly due to the pandemic and there are drastic changes in the business (Ince, 2017). The outbreak also resulted in job losses, and the unicorns had to lay off their employees on a large scale. With the sharing economy, companies were able to introduce the gig economy concept but due to the pandemic, the sustenance of this huge workforce was not possible. India saw a major hit in the automobile industry in the year 2019 and it was further aggravated during the pandemic season (Santos, 2019). In this study, the researchers focus on the definition of sharing economy and study its progress in data. From the analysis, it is inferred that people are highly likely to continue using the ride-sharing options with precautionary additions. The technological advancement would also help in opening new areas in the sharing economy rather than restricting to hospitality, ride-sharing or co-work spaces (Bardhi, 2012). The shared economy can be expanded to fun-related activities by providing sports facilities, gaming arenas and equipment for

rent which can increase the fun and motivate people to meet and collaborate with people across the community.

Online Book Shopping in Vietnam: The Impact of the COVID-19 Pandemic Situation

The coronavirus disease 2019 (COVID-19) pandemic will have large effects on key stakeholders in the publishing industry. Given that physical bookstores may be forced to close temporarily and that consumers may not want to travel to such brick-and-mortar stores because of health concerns, the demand for online book shopping will expectedly rise quickly in the short term. This study aims to contribute to the extant literature by investigating the impact of the COVID-19 pandemic situation and consumers' motivations on their intentions to buy books online. It also focuses on Vietnam which represents a potential market opportunity for online retailers including e-bookstores [3, 10, 11]. They collected the data by using an online survey. The algorithms in this paper were descriptive statistics, reliability analysis, Bivariate Correlation, and Multiple Regression Analysis. Finally, this study's findings have important practical implications for publishers and online booksellers. Given the significant impact of the COVID-19 pandemic situation on purchase intention, online booksellers should make every effort to their product portfolios on their websites as well as provide consumers with more sales promotions

Machine-Learning Models for Sales Time Series Forecasting

In this study, the authors considered different machine-learning approaches for time series forecasting. Sales prediction is rather a regression problem than a time series problem. The use of regression approaches for sales forecasting can often give us better results compared to time series methods. One of the main assumptions of regression methods is that the patterns in the historical data will be repeated in the future. The accuracy of the validation set is an important indicator for choosing an optimal number of iterations of machine-learning algorithms. The effect of machine-learning generalization consists in fact of capturing the patterns in the whole set of data. This effect can be used to make sales predictions when

there is a small number of historical data for specific sales time series in the case when a new product or store is launched. In the stacking approach, the results of multiple model predictions on the validation set are treated as input regressors for the next-level models. As the next level model, Lasso regression can be used. Using stacking makes it possible to consider the differences in the results for multiple models with different sets of parameters and improve accuracy on the validation and on the out-of-sample data sets. They conducted descriptive analytics to do data visualization. Then, they built different forecasting models to compare.

Prediction Analysis for Business To Business (B2B) Sales of Telecommunication Services using Machine Learning Techniques

Sales forecast analysis requires intelligent data mining techniques with accurate prediction models and high reliability. Sales estimates provide data on how a company should manage its sales team, products, and also budgeting flows. Accurate estimates enable organizations to increase in accordance with the growth of the market in the possible maximum level of income. This research will focus on the comparison of prediction analysis and B2B sales reliability sales using machine learning techniques. This study employed 4 (four) algorithms of machine learning and tested the model to see the accuracy performance in which the model with the best accuracy was selected (Generalized Linear Model, Decision Tree, Gradient Boosted Trees, Random Forest, and Forecast Estimation, Evaluation & Transformation). Prediction performance mainly deals with accuracy. Confusion matrix, showing the total prediction indicated by RMSE, MSE, and absolute errors, and accuracy of each class are measured.

4. Research Design and Methodology

4.1. Data collection

Data was collected from customers who have used the company services to ship packages to Vietnam (the main marketplace in Vietnam) from the end of 2019 to Sep 2021 and were used in this study to predict future sales.

4.1. Exploratory data analysis

The dataset has over 21,894 observations and includes 15 columns. The dataset needs to be pre-processed to identify outliers, missing values and analyzed further to determine the features relevant to the business problem statement. After doing cleaning data, the dataset remains 19,951 observations with no null value.

	id	sender_id	sender_zipCode	receiver_id	quantity	actual_weight	rate	buy_in_store	extra_charge
count	19951.000000	19951.000000	19951.000000	19951.000000	19951.000000	19951.000000	19951.000000	19951.000000	19951.000000
mean	14632.965165	4136.043707	58969.626685	7544.727633	1.056689	20.659347	4.385973	14.524478	13.365527
std	6947.853448	2518.844245	24148.596822	3912.769104	0.553634	23.065253	1.913032	52.255303	60.621173
min	2206.000000	2.000000	2151.000000	1.000000	1.000000	0.000000	1.500000	-410.000000	-3475.000000
25%	8730.500000	2052.000000	22180.000000	4662.500000	1.000000	9.000000	3.500000	0.000000	0.000000
50%	14482.000000	4208.000000	75040.000000	7465.000000	1.000000	15.000000	4.500000	0.000000	0.000000
75%	20451.500000	6082.500000	75074.000000	10565.500000	1.000000	25.925000	4.500000	0.000000	10.000000
max	26838.000000	9285.000000	99762.000000	15267.000000	47.000000	1392.000000	130.000000	2595.000000	4176.000000

Figure 1 Description the dataset's features

4.2. Data analysis

Data analytics algorithms and research method

- This stage is all about identifying relevant attributes which are related with the response variable. A descriptive qualitative design and reliability analysis will be used for this study.
- Lastly, I use the training samples to fit the model using various algorithms and validate the model against the test data. Some models I may use in this study: Linear Regression, Support Vector Machines (SVMs), Decision Tree, Random Forest.

The basic ideas for data analysis:

- The total weight they shipped each month,
- The main destination (Vietnam has 61 provinces and cities; however, they separate the destination by 2 main sections: SAIGON and TINH),
- The most of commodities that customers sent,
- Which month they shipped the most,
- How many agents do they have?
- Compare the revenue before and after (controlled) COVID-19,
- Predict for the last 3 months in 2021

4.3. Data visualization

	Agent_ID	No_of_Transactions
0	ABC	12943
1	QUY	5715
2	NEW	746
3	LTM	304
4	NGO	218
5	VNC	24
6	ABCVN	1

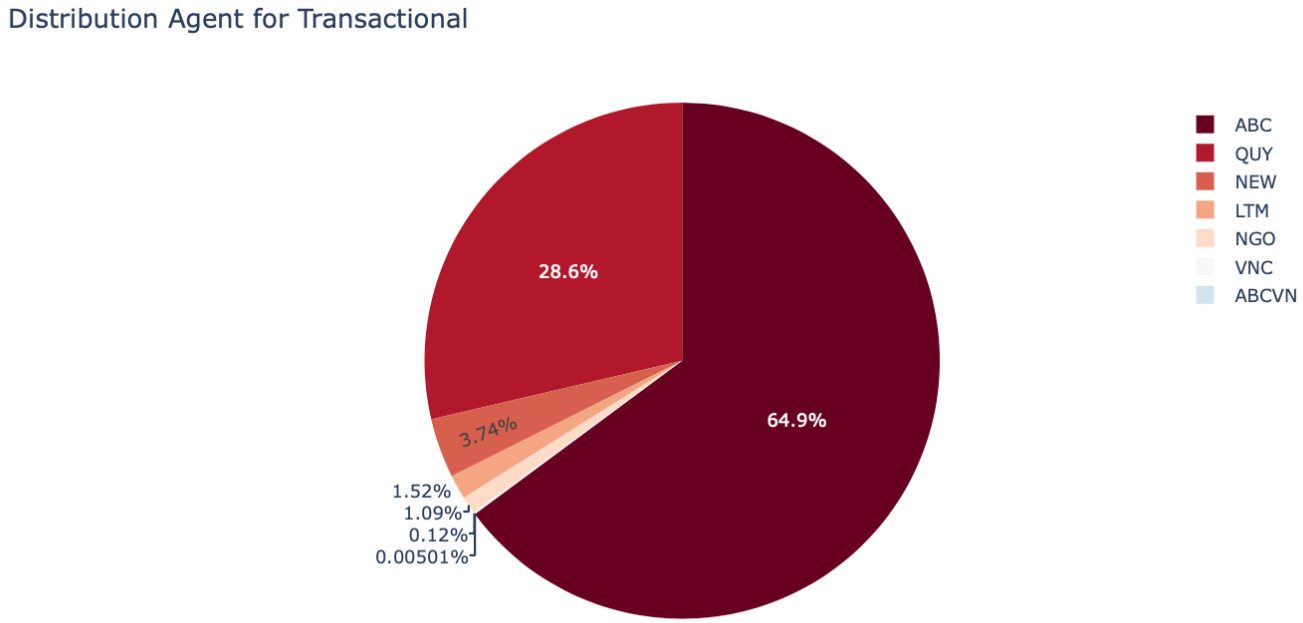


Figure 2 - Number of Agents / Number of Transactions per Agent

	Service	No_of_Transactions
0	TINH	9858
1	SAIGON	9712
2	DOM	144
3	INT	93
4	EXPRESS	63
5	CAMBODIA	51
6	CHERRY	29
7	FEDEX	1

Figure 3 - Number of Services

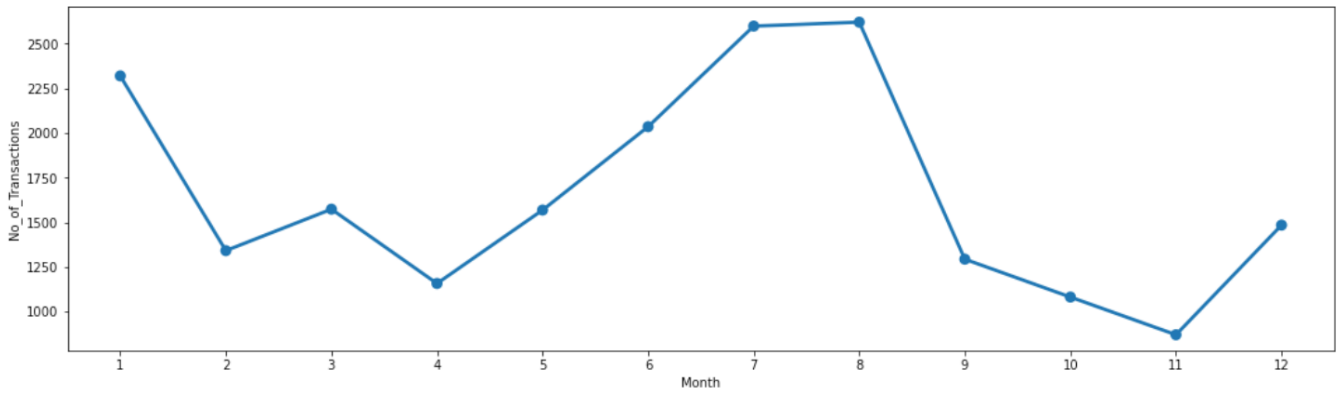


Figure 4 - Distribution of Transactions by Month

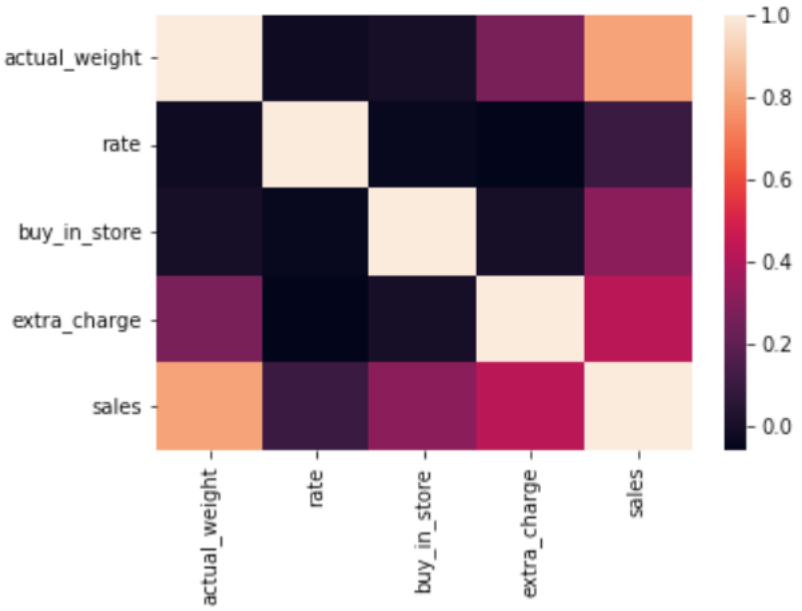


Figure 5 - Heatmap of sale feature

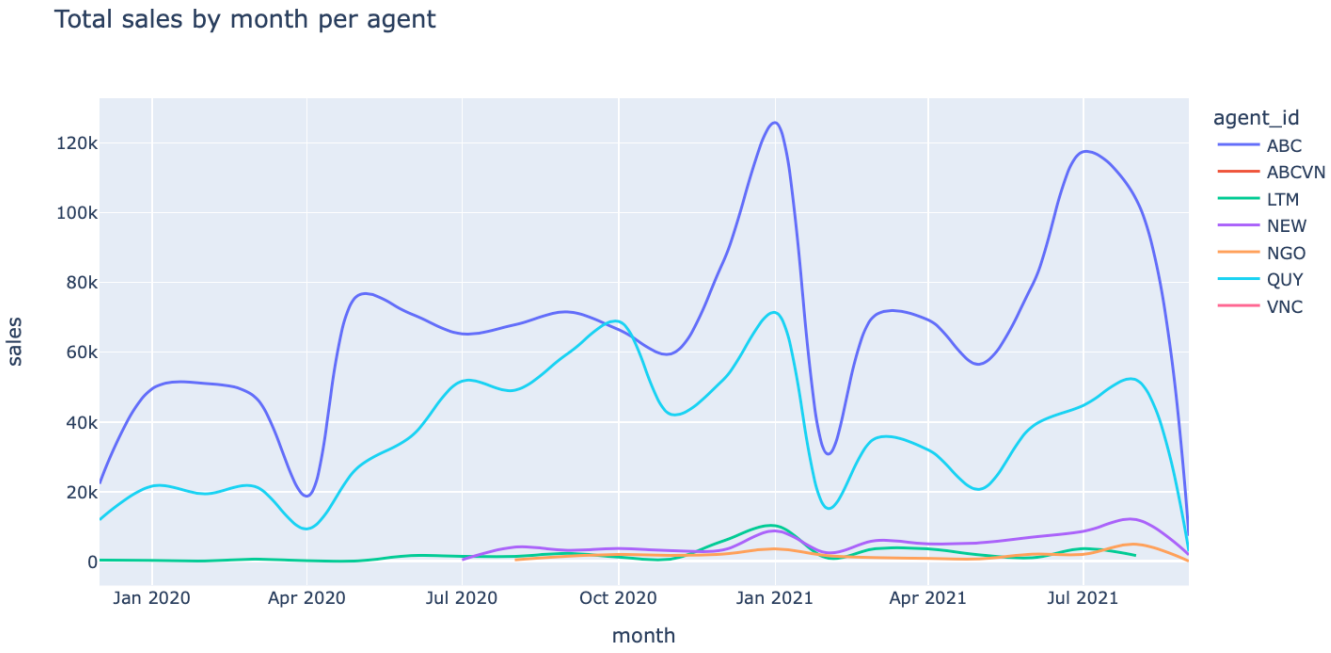
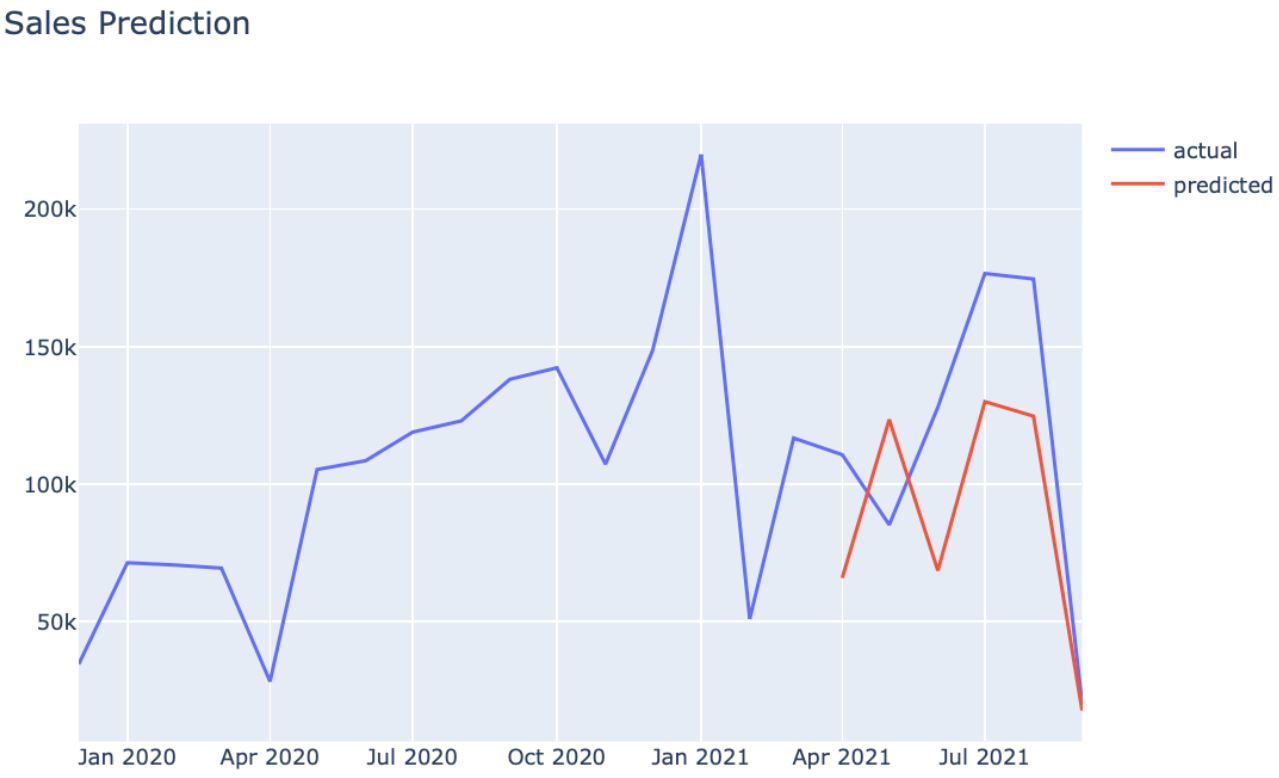
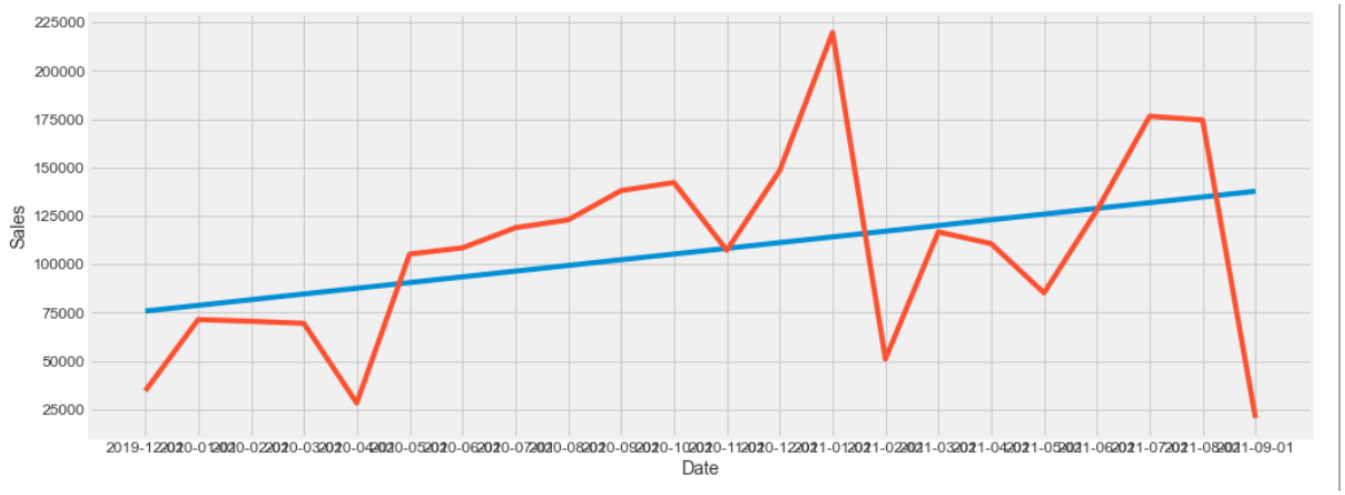


Figure 6 - Total sales by Month per Agent




```
RMSE: 44086.685590335284
MAE: 40291.642250000004
R2 Score: 0.32695066751096946
```

Figure 7 - Sale Prediction using LSTM model



5. Conclusion and Result

This study is a part of its kind that has investigated the impacts of the COVID-19 pandemic on the shipping industry, especially small shipping companies. The big issue that the small business concerns are the profit and revenue, and this is also my interesting thing. The study showed that the COVID-19 is affecting on ABC company and small businesses in general.

Through the study, July and August are the months which have the greatest numbers of transactions. However, January is the month that has the greatest total sales. That means in the holiday season, they will send a lot of packages. April 2020 is the time that the pandemic started, and the sale was going down several months after that. According to the trend chart, it will improve at the end of the year. After building linear regression, we have good news that the prediction for the next 3 or 6 month and the sale is going up.

Almost transactions of ABC company belong to ABC agent. So, the business needs to support other agents if they want to develop their business, such as promotions, coupons on the upcoming holidays (New Year, Lunar New Year).

The range of this dataset is a little bit short, and it only covered a total of 22 months. Therefore, I cannot train some models that require more than 22 observations. I wish I have time to collect more to bring out a better result.

Reference

Hasan, Syed; Allen, Klaiber H; Sheldon. The impact of science parks on small- and medium-sized enterprises' productivity distributions: the case of Taiwan and South Korean. In *Small Business Economics; Dordrecht Vol. 54, Iss. 1* (Jan 2020): p. 135-153. https://doi.org/10.1007/s12109-020-09732-2

Sube Singh, Ramesh Kumar, Rohit Panchal & Manoj Kumar Tiwari. Impact of COVID-19 on logistics systems and disruptions in food supply chain. In *International Journal if Production Research, Vol. 59, 2021, Iss. 7* (Jul 2020)

Wei Wang, Wenjing Huan, Xiaoxue Liu, Dwight A. Hennessy. Psychological impact of mandatory COVID-19 quarantine on small business owners and self-employed in China. In *Current Psychology* (Jun 2021). https://doi.org/10.1007/s12144-021-01983-2

Robert Fairlie. The impact of COVID-19 on small business owners: Evidence from the first three months after widespread social-distancing. In *Journal of Economics & Management Strategy, Vol. 29, Iss. 4 (Aug 2020) p. 727-740*. https://doi-org.libproxy.library.unt.edu/10.1111/jems.12400

Hoang Viet Nguyen, Hiep Xuan Tran, Le Van Huy, Xuan Nhi nguyen, Minh Thanh Do, Ninh Nguyen. Online Book Shopping in Vietnam: The Impact of the COVID-19 Pandemic Situation. In *Publishing Research Quarterly (2020) 39: 437-445*. https://doi.org/10.1007/s12109-020-09732-2

Bylen, Slawoir. Market of Logistics Services During the Covid-19 Pandemic. In *European Research Studies, suppl. Special issue 3; Anixis Vol. 23 (2020): p. 47-61*

Elinor Aviv-Sharon, Asaph Aharoni. Generalized logistic growth modeling of the COVID-19 pandemic in Asia. In *Infect Dis Model. 2020; 5: 502-509*. doi: 10.1016/j.idm.2020.07.003

Umaña-Hermosilla, Benito; de la Fuente-Mella, Hanns; Elórtegui-Gómez, Claudio; Fonseca-Fuentes, Marisela. Multinomial Logistic Regression to Estimate and Predict the Perceptions of Individuals and Companies in the Face of the COVID-19 Pandemic in the Ñuble Region, Chile**.** *In Sustainability, Basel Vol. 12, Iss. 22 (2020): 9553***.** DOI:10.3390/su12229553

Paul P. Vinod and Dipasha Sharma, COVID-19 Impact on the Sharing Economy Post-Pandemic. *In JEL classification: G14, AABFJ, Volume 15, No. 1, 2021*

Wisesa Oryza, Adriansyah Andi; Khalaf Osamah Ibrahim, Prediction Analysis for Business To Business (B2B) Sales of Telecommunication Services using Machine Learning Techniques, *In Majlesi Journal of Electrical Engineering; Isfahan, Vol. 14, Iss.. 4, (Dec 2020): 145-153.DOI:10.29252/mjee.14.4.145*

Grigorios Tsoumakas, A survey of machine learning techniques for food sales prediction, *In The Artificial Intelligence Review**;** Dordrecht Vol. 52, Iss. 1, (Jun 2019): 441-447.DOI:10.1007/s10462-018-9637-z*

Liu X, Ichise R (2017), Food sales prediction with meteorological data A case study of a Japanese chain supermarket. *In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10387 LNCS, pp 93–104*

Ching-She, W., Patil, P., and Gunaseelan, S., "Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data," *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 10.1109/ICSESS.2018.8663760, 2018.*

Emilio Soria Olivas [and others], Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, *Hershey, Pa. : IGI Global (701 E. Chocolate Avenue, Hershey, Pennsylvania, 17033, USA), [2010]*

Kumar Anil, Kabra Gaurav, Mussada Eswara Krishna, Dash Manoj Kumar, Rena Prashant Singh, Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention**.** In *Neural Computing & Applications*, *Feb2019 Supplement 2, Vol. 31 Issue 2, p877-890. 14p. DOI: 10.1007/s00521-017-3047-z*