

Predicting Public Procurement Irregularities in the COVID-19 Response of Local Government Units (LGUs) in the Philippines

John Raymond B. Barajas
Nico O. Aspra
Pee Jay N. Gealone
Arpon T. Lucero Jr.
Oliver M. Padua
Marben S. Ramos

CONTENTS

1	Executive Summary	1
1.1	Highlights	1
1.2	Short Summary	1
2	Motivation	2
3	Data Source and Dataset Description	5
3.1	Philippine Government Electronic Procurement System Procurement Data (2020)	5
3.2	Commission on Audit Annual Audit Reports for Local Government Units (2020)	7
4	Methodology	8
4.1	Data Cleaning and Pre-Processing	9
4.1.1	Data Cleaning and Pre-Processing of PhilGEPS 2020 Procurement Data	9
4.1.2	Extraction of Local Government Unit Contracts	11
4.1.3	Data Cleaning and Pre-Processing of Commission on Audit 2020 Annual Audit Reports	11
4.2	Identification of Irregular Contracts	12
4.3	Dimensionality Reduction	13
4.4	Hierarchical Clustering	13
4.5	Training and Validation of Logistic Regression Model	13
5	Exploratory Data Analysis	14
5.1	Distribution of LGU Tenders According to Notice Status	14
5.2	Approved Budget Contract According to Notice Status	15
5.3	Contract Duration According to Notice Status	16
5.4	Distribution of LGU Tenders According to Classification	17

5.5	Exploration of Approved Budget of the Contracts According to Classification and Business Category	18
5.6	WordCloud Visualization of PhilGEPS Dataset Containing Text Data	21
5.7	WordCloud Visualization of Reason for Award Feature	22
6	Exploration of Local Government Unit COVID-19 Contracts	23
6.1	Distribution of COVID-19 Contracts Posted by LGUs According to Notice Status	23
6.2	Approved Budget Contract of COVID-19 Tenders According to Notice Status	24
6.3	Budget Allocated on COVID-19 Contracts Posted by LGUs According to Business Category	25
7	Exploration of Potentially Irregular Contracts Posted by Local Government Units	28
7.1	Top Local Government Units Identified to have Potentially Irregular Contracts	29
7.2	Budget Allocated on Potentially Irregular Contracts Posted by LGUs According to Business Category	30
8	Generalization of Audit Findings through Hierarchical Clustering	33
8.1	Generalization of Audit Observations through Hierarchical Clustering	33
8.2	Generalization of Audit Recommendations through Hierarchical Clustering	37
8.3	Practical Implication of the Identified Audit Topics	42
9	Training and Validation of Logistic Regression Model	43
9.1	Reduction of Features through Dimensionality Reduction	43
9.2	Training and Validation of Logistic Regression Model	45
9.3	Limitations of the Present Work	47
9.3.1	Limited Data on COVID-19 Related Tenders	47
9.3.2	Limitation of the Developed Prediction Model	48

10 Suggested Future Work	49
11 Policy Recommendations	50
12 Acknowledgements	52
References	53

EXECUTIVE SUMMARY

Highlights

- On average, **only 2 out of 10 LGU contracts** have been awarded in 2020.
- **For every Php1 spent, approximately Php1 remains unspent** in the procurement of goods and services made by LGUs.
- A total of Php481 billion were distributed across all LGUs in the country for 2020 but only **10% of this budget was allocated for the procurement of drugs and medicines**. 40% of this budget went to construction projects.
- Excluding LGU contracts not posted in the PhilGEPS website, **only Php10 billion (2.16% of the total LGU budget) was allocated for COVID-19 related contracts**.
- An equivalent amount of **Php720 million was potentially lost** from 786 LGU contracts flagged as irregular.
- Audit findings for LGUs were primarily centered on directing accountable officers to **comply to documentary requirements** mandated by existing circulars, memorandums, and Philippine laws.
- **A logistic regression model with an accuracy of 91.29% was developed** to identify contracts that are potentially irregular.

Short Summary

From examination of 296,220 local government unit contracts, this project was able to develop a logistic regression model capable of predicting potentially irregular LGU contracts posted in the PhilGEPS website for the fiscal year

2020 at an accuracy of 91.29%. Validation of the model using metrics derived from the confusion matrix revealed that the developed model had a recall score of 1.0 and a precision score of 0.029. While precision of the model may be low, the high recall score is deemed more important in this use-case since it would be more costly for an LGU to miss out on irregular contracts. Overall, the developed prediction model is seen to be highly beneficial as a decision support tool for LGUs since this could potentially narrow down the number of awarded LGU contracts to be legally reviewed resulting to faster turnover of review cycles conducted within a given fiscal year.

MOTIVATION

Almost a year after the unprecedented worldwide spread of the SARS-CoV-2 virus, governments across the globe have already moved to implement various economic policies with COVID-19 vaccine procurement identified as one of the primary stimuli for ensuring a pathway towards a sustainable economic recovery [1]–[4]. While it is a known fact that a global roll-out of COVID-19 vaccines would be a “game-changer” in restarting the global economy, securing economic sustainability during and after the COVID-19 pandemic strongly demands for an equitable distribution of COVID-19 vaccines to any procuring country regardless of its economic status [5]–[9]. This notion of “equitability” in vaccine distribution, however, is extremely challenged by corrupt practices deeply rooted in political interests that promulgate public procurement inefficiencies [10]. Instances of such which resulted to unwarranted delays in COVID-19 related procurement activities have already been reported and these have been proven to be specifically true in the context of low-income countries (e.g. Philippines) [11]–[13].

Corrupt practices which include the use of corroborated activities that aim to cover-up procurement inefficiencies, discourage a unilateral call for accountabil-

ity, and minimize transparency for existing COVID-19 response efforts have already propagated numerous irreversible negative economic impacts that highly affected low-income countries [10]–[14]. As the last country in Asia perceived to bounce back from economic losses induced by the COVID-19 pandemic [15], the devastating effects of these unwanted corrupt activities have been clearly illustrated in the context of the Philippine setting [16]–[28]. According to a report published by Open Contracting Partnership [28], only 47% of the posted COVID-19 related tenders have been awarded by the Philippine government even with various issuances that eased contract requirements for COVID-19 related procurements. This relatively low awarding rate is highly indicative of inefficiencies in the Philippine public procurement system [28]. Moreover, plagued by a series of controversies surrounding the procurement of COVID-19 vaccines [16]–[22], the Philippines was also reported to have missed its chance to secure COVID-19 vaccines in January of 2021 due to the health secretary not being able to sign crucial procurement documents (e.g. confidentiality disclosure agreement) on time [16]–[18]. This act was cited to be tantamount to “gross and inexcusable negligence” under the existing Philippine anti-graft law since negotiations to secure the first 10 million doses of COVID-19 vaccines for the Filipino people started as early as July of 2020 [16]. However, no charges were filed and the presidential office still expressed its trust and confidence citing that no major lapse was found in the actions of the health secretary [23]. This controversy was further magnified by the fact that the Philippine government highly favored the procurement of 25 million doses of a COVID-19 vaccine that have a lesser efficacy but is more costly than existing antivirals in the market [18]–[22]. Due to the public uproar expressed in various social media platforms, an official statement was publicly made to clarify this issue where it was announced that the price per dose of this target COVID-19 vaccine cannot be disclosed since this was the present “industrial practice” [24], [25]. As a direct consequence of these events, it was validated through the survey conducted by Pulse Asia that only one (1) out of three (3) Filipinos have expressed willingness to be inoculated with the COVID-19 vaccine if it is made available in

their respective municipalities [26], [27]. If this unwillingness to be vaccinated against the SARS-CoV-2 virus is not reversed, this would ultimately push the contraction of the Philippine economy beyond the reported shrinkage of -9.5% as described in [29]. Indeed, these cited instances have extremely demonstrated the unfortunate consequences of corrupt practices and these scenarios strongly calls for government transparency and data-driven solutions designed to mitigate instances of such practices.

In an attempt to provide a data-driven solution which promotes integrity and transparency in the Philippine public procurement system especially in times health emergencies (i.e. COVID-19 crisis), this paper developed a prediction model that is capable of identifying irregularities in tenders posted in the Philippine Government Electronic Procurement System (PhilGEPS). In addition, the practical implications of the developed model were also analyzed and interpreted in the context of the official annual audit reports published by the Commission on Audit (COA) of the Philippines. While initial literature review conducted suggests the adoption of validated predictors described in [30]–[34], these predictors of irregularities still need to be tested and screened for consistency with the identified predictors unique to the Philippine public procurement system (e.g. index reported in [28]). Nevertheless, the accuracy of the developed prediction model was also tested through the use of test data. By collating then these predictors into a prediction model, tenders posted in PhilGEPS could now be systematically tagged for potential irregularities in each transaction made. Deployed in the field, applications of this developed prediction model are then seen to be extended as an objective and scientific approach to raise public awareness for potentially corrupt procurement practices and promote accountability for validated procurement irregularities.

DATA SOURCE AND DATASET DESCRIPTION

Philippine Government Electronic Procurement System Procurement Data (2020)

To achieve the overall objective this work, the first dataset that was used in this project was the 2020 PhilGEPS Bids Notices and Award Notices Open Data taken from the PhilGEPS official website (see [Figure 1](#)) [35]. This dataset is divided into four chunks as PhilGEPS releases open procurement data on a quarterly basis. Specifically for this work, the Jan-Mar 2020, Apr-Jun 2020, Jul-Sep 2020, and Oct-Dec 2020 open data were merged into a single dataframe yielding a total of 810863 contracts. This dataset has forty (40) features and these features could be summarized into five major categories:

- **Information about procuring agency:** Organization Name, Reference ID, and Solicitation No. are the features of the dataset that provide basic information about a procuring agency.
- **Information about mode of procurement:** Funding Source, Funding Instrument, Procurement Mode, and Trade Agreement are the features of the dataset that provide basic information about the mode of procurement implemented by a procuring agency.
- **Information about bid dates:** Publish Date, PreBid Date, and Closing Date are the features of the dataset that enumerate the important dates relative to the tender posted by a procuring agency.
- **Information about posted tenders:** Notice Title, Classification, Notice Type, Business Category, Approved Budget of the Contract (ABC), Area of Delivery, Contract Duration, Line Item No., Item Name, Item Desc, Quantity, Item Budget, Notice Status, UNSPSC Code, and UNSPSC Description are the features of the dataset that provide comprehensive details about a tender posted by a procuring agency. The approved budget, the

specific goods or service, and the quantity being procured are detailed in these features.

- **Details about awarded tenders:** Award No., Award Title, Award Type, Contract Amount, Contract No., Publish Date (Award), Award Date, Notice to Proceed Date, Contract Effectivity Date, Contract End Date, Reason for Award, and Award Status are the features of the dataset that provide comprehensive details on tenders that have been awarded to a winning bidder. The name of the winning bidder, the reason for awarding the tender, the contract amount, and when the contract is to be implemented and agreed date to be completed are detailed in these features.

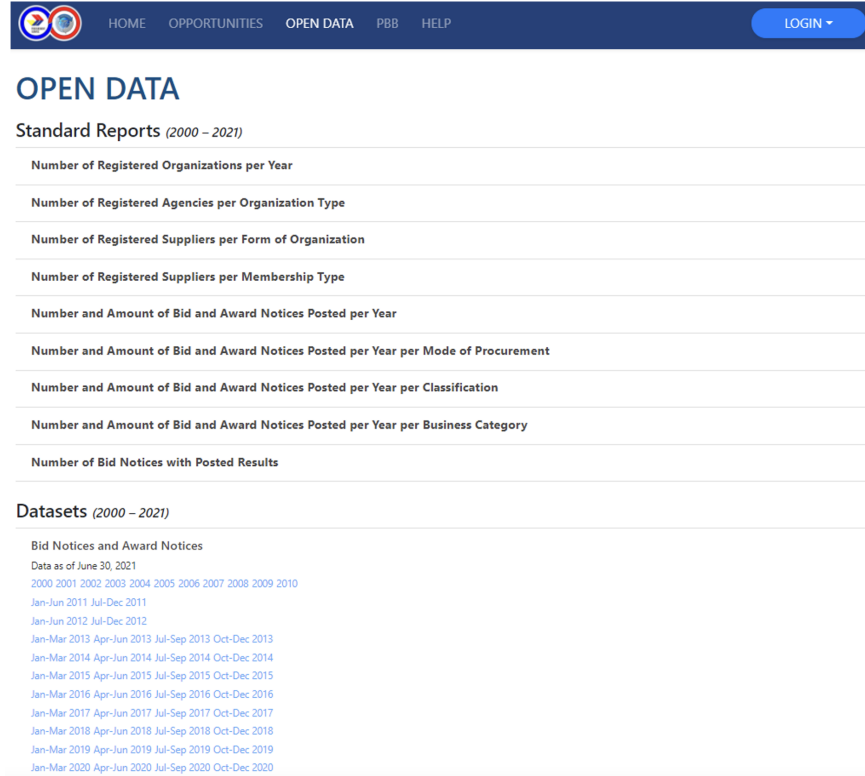


Figure 1: Open Philippine procurement data posted PhilGEPS official website. To consider the effect of the COVID-19 pandemic on observed procurement behavior, only the 2020 procurement data was considered in this work. This data can be accessed through this [website](#).

Commission on Audit Annual Audit Reports for Local Government Units (2020)

The second dataset used in this work to supplement the analysis done for the 2020 PhilGEPS Bids Notices and Award Notices Open Data was the local government unit (LGU) annual audit reports posted by Commission on Audit (COA) for the fiscal year 2020 (see [Figure 2](#)) [36]. While the first dataset is divided into quarters, this second dataset is divided by the number of regions in the Philippines. Merging of the annual audit reports from across Philippine regions yielded about 6254 documents containing audit observations with their corresponding audit recommendations. This dataset has twelve features but only seven (7) features were found to be related to the first dataset. The features that have been used from this dataset are grouped into three (3) categories and these are summarized below:

- **Information about local government unit:** Region Number, Region Name, Province Name, LGU Name, and LGU Type are the features of the dataset that provide basic information on the local government unit audited.
- **Audit Observations:** This feature contains the relevant audit findings observed for a local government unit. Audit observations found usually revolve around discrepancies in financial statements, findings on inventory management, and violations made relative to issued memorandums and/or circulars.
- **Audit Recommendations:** This feature contains the relevant steps suggested to prevent further occurrence relative to a particular set of audit observations. Audit recommendations found usually focused on compliance to documentary requirements.



Figure 2: Open annual audit reports for Philippine local government units posted by the Commission on Audit. Downloaded annual audit reports were only as of July 31, 2021. This data can be accessed through this [website](#).

METHODOLOGY

With the underlying research questions this work is trying to solve, the general process shown in [Figure 3](#) was implemented in this project. Briefly, the PhilGEPS 2020 procurement data and the COA annual LGU audit reports for the fiscal year 2020 were collected. The PhilGEPS 2020 procurement data did not anymore require any pre-processing work since it is made available as a csv file; the downloaded COA annual LGU audit reports, however, were manually typed and converted to doc files since these documents were uploaded as images rather than text files. Once these data have been structured as two separate dataframes, data cleaning and pre-processing were implemented. After this step, potentially irregular contracts were then identified by looking at the bid dates for each posted contract. Dimensionality reduction was also done in order to reduce the features of the PhilGEPS 2020 procurement data and manage the training time of the employed logistic regression analysis. Parallel to this step,

hierarchical clustering was also employed to the COA data to further supplement the results of the logistic regression employed. Finally, logistic regression was implemented and recommendations anchored to the results of the analysis were formulated. Details of the employed methodology are further discussed in the succeeding sections.

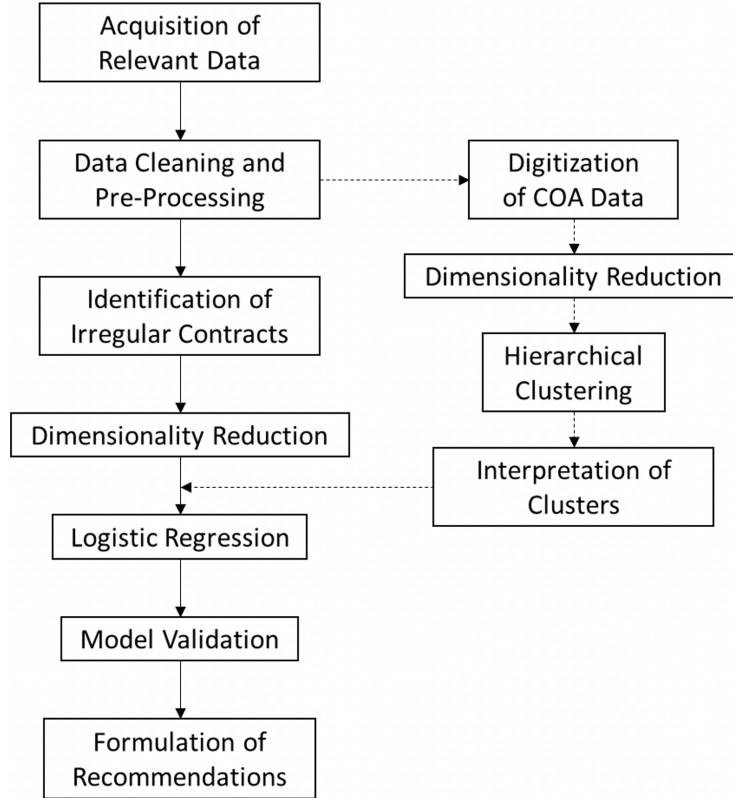


Figure 3: High-level methodology implemented in the project. Hierarchical clustering of audit observations and audit recommendations were done in parallel to the logistic regression pipeline in order to supplement the results of this employed analysis.

Data Cleaning and Pre-Processing

Data Cleaning and Pre-Processing of PhilGEPS 2020 Procurement Data

Initial exploration of the PhilGEPS 2020 procurement data revealed that the number of missing data was not a significant issue since the salient features of

the dataset did not contain null values. Further examination of the dataset confirmed, however, that a number of features in the dataset – specifically Publish Date (Award), Award Date, and Notice to Proceed Date columns – had 90.51% of its values as either missing or null. While this is undesirable, careful inspection of the Notice Status column revealed that this number of missing values for these columns is to be expected since majority of the posted contracts for 2020 have not yet been awarded. Imputation of data, hence, was not anymore pursued.

After validating that the PhilGEPS 2020 procurement data was clean, this data was then subjected to pre-processing. Initially, the features of this dataset were categorized into three major groups namely numeric, categoric, and text columns. The following were then implemented in order to ensure this pre-processed data could be fed as training and test data:

- **Numeric Features:** Null values under the column Contract Amount were changed to “0” since these rows corresponded to contracts that were not awarded to any bidder. No further pre-processing was done since the numeric features were already clean.
- **Categoric Features:** One-hot encoding was performed to these features in order to incorporate these columns into the logistic regression model. The Area of Delivery column had about 7.31% of its values as null values. No imputation was done but these null values were changed to “Others” to still capture these rows as training and test data.
- **Text Features:** The columns Notice Title, Item Name, Item Desc, and Award Title were first combined as a single feature since these features are strongly correlated to each other. For these features, only words (a sequence of alphabet characters) were extracted as text data. After extraction, stopwords were removed and the extracted text was subsequently lemmatized. Finally, the term frequency – inverse document frequency (TF-IDF) of these data were calculated. These same set of steps were also applied to the Reason for Award column.

Lastly, from the results of these pre-processing steps, a single dataframe comprising of 810863 rows and 934 columns were created. This cleaned and pre-processed data was used for further analysis.

Extraction of Local Government Unit Contracts

Once the PhilGEPS 2020 procurement data had been cleaned and pre-processed, contracts posted by LGUs were extracted. To extract these data, “CITY OF”, “MUNICIPALITY OF”, and “PROVINCE OF” were used as keywords to search for contracts made LGUs only. Inspection of the extracted LGU data revealed that various Philippine government agencies were also included in the performed search. These were then manually removed from the dataset in order to cover only contracts specifically made by LGUs. This extracted dataset (296220 rows) is the final training and test data fed to the logistic regression pipeline described in this work.

Data Cleaning and Pre-Processing of Commission on Audit 2020 Annual Audit Reports

A significant number of COA annual audit reports made for LGUs in 2020 were made available as images contained in PDF files under the official COA website. These PDF files were converted into machine readable text files by manually typing their contents into separate doc files. From these contents, audit observations for each LGU, together with their audit recommendations, were extracted and structured into a single dataframe saved as a csv file. The following features were thus created from these data:

- **LGU Information:** The corresponding Region Number, Region Name, Province Name, LGU Name, and LGU Type were manually recorded for each examined audit findings contained in an LGU annual audit report.
- **Audit Findings:** The audit observations matched with their correspond-

ing audit recommendations were recorded under separate columns for each LGU.

From this dataframe, each text column (i.e. Audit Observations and Audit Recommendations) were cleaned and pre-processed by removing stopwords and text lemmatization. This cleaned and pre-processed dataset (6254 rows and 1000 columns) was then used for cluster analysis.

Identification of Irregular Contracts

Initial exploration of the PhilGEPS 2020 procurement data revealed that potential irregularities for each contract were not obvious within the dataset. However, following closely the cycle in which LGU contracts are awarded – to investigate whether or not the prescribed procurement process is being observed – negative differences in procurement dates posted were found. Specifically, the addition of the feature Award-NTP (Days) – the difference in days between the Notice to Proceed date and Award date of an LGU contract – implied (for 786 LGU awarded tenders in particular) that Notice to Proceed orders for such contracts were issued first prior to awarding each contract to a winning bidder. This was considered a red flag since this is in clear violation of the Revised Implementing Rules and Regulations of Republic Act 9184 [37]. While it is highly probable that this observation may be indicative of errors in the PhilGEPS data, further verification done through the careful examination of COA annual audit reports indeed confirm that such a practice is being observed in various municipalities in the Philippines. Hence, from these validations steps implemented, these contracts were labeled as “irregular” while the remaining were labeled as “not irregular” and this was finally collated as the target feature for the cleaned and pre-processed PhilGEPS 2020 procurement data.

Dimensionality Reduction

In order to significantly reduce runtime for training the logistic regression model implemented in this work, dimensionality reduction was performed separately for both the cleaned and pre-processed PhilGEPS 2020 procurement and the 2020 COA LGU annual audit report data. Specifically, truncated singular value decomposition (SVD) was implemented in order to reduce the features of both datasets into a more manageable size at a cumulative explained variance ratio sum of 80% [38], [39].

Hierarchical Clustering

Hierarchical clustering of the cleaned and pre-processed 2020 COA LGU annual audit reports were conducted separately for the audit observations and audit recommendations. To perform hierarchical clustering on these two separate columns, the Ward’s method with the terminal criterion Δ was employed [40]. The formed clusters that had the highest merging cost (i.e. Δ) were used as the cut-off to determine the optimal number of clusters for these data.

Training and Validation of Logistic Regression Model

A logistic regression analysis as detailed in [41] and [42] was adopted to develop a prediction model capable of identifying “irregular” and “not irregular” LGU contracts based on a given set of features similar to what is made available in the PhilGEPS 2020 procurement data. Since the formulated problem in this work is a classification problem, two clusters were identified as the target feature (i.e. “irregular” or “not irregular”) for the training and validation of the logistic regression model. It is worth noting that the cleaned and pre-processed PhilGEPS 2020 procurement data is highly imbalanced with only about 0.27% of the data representing “irregular” contracts. To deal with this imbalanced classification

problem, a Stratified K-fold Cross-Validation (10-fold) and Synthetic Minority Over-Sampling Technique (SMOTE) was employed to the splitting of the data [43], [44]. Furthermore, to identify which of the features of contracts (i.e. Award Date, Notice to Proceed Date, Approved Budget of the Contract, Notice Status, Item Name, Item Description, Award Title, Award-NTP (Days)) have the most significant impact to the target feature, L2 regularization (i.e. Ridge Regression) was also implemented [45]. Lastly, to validate the model, the test accuracy was determined at the optimal hyperparameter C and a confusion matrix was drawn from this best estimator in order to identify the precision and recall score of the trained prediction model. Recommendations were then formulated according to the observed results.

EXPLORATORY DATA ANALYSIS

Prior to the training and validation of the logistic regression model for identifying irregular LGU contracts, an exploratory data analysis was initially conducted in order to extract insights on the PhilGEPS procurement and COA annual audit report data. The succeeding sections discusses these insights further.

Distribution of LGU Tenders According to Notice Status

In this time of a health-crisis, it is crucial that LGUs in the Philippines come equipped with all that is necessary in order to mitigate the consequences of the COVID-19 pandemic. This also includes an efficient, fast, but transparent public procurement process which has been recently identified as a key to effectively curbing the effects of the COVID-19 pandemic [46]. While this instance is ideal since the public procurement process is expected to slow down due to mobility restrictions that have been imposed to control the spread of the SARS-CoV-2

virus, the public procurement process implemented in the Philippines is indeed in bad shape with only about 2 out of 10 LGU contracts awarded in 2020 (see [Figure 4](#)). This number definitely calls for an intervention that would increase the awarding rate of LGUs.

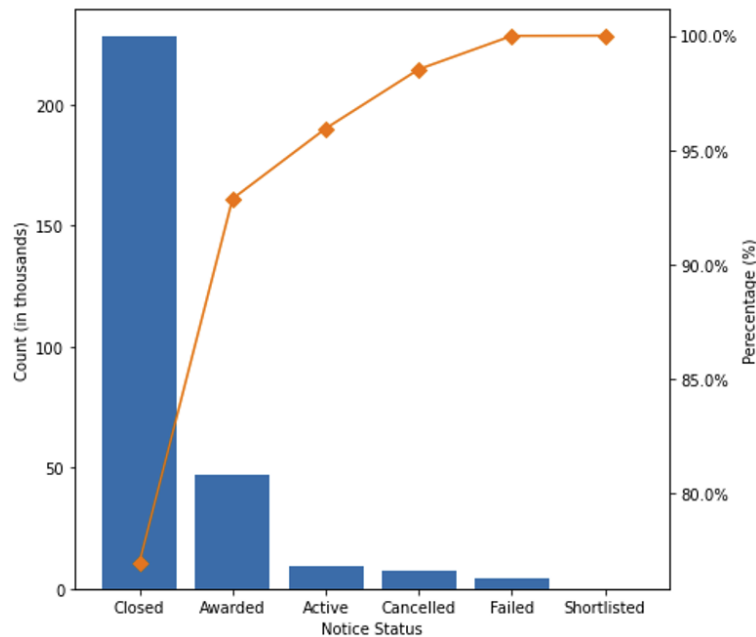


Figure 4: Pareto chart describing the distribution of Notice Status of posted LGU contracts. LGU contracts that have Notice Status of “closed” – posted LGU tenders that have been closed but not awarded – had the highest count.

Approved Budget Contract According to Notice Status

Looking at [Figure 4](#) in another perspective, the observed mean ABCs for Awarded and Unawarded LGU contracts have been found to be significantly close with each other. As shown in [Figure 5](#), the mean ABCs for Unawarded and Awarded LGU contracts were near the Php250,000 mark ($p\text{-value} = 0.04046$) with Cancelled, Shortlisted, and Active categories having mean ABCs greater than that of the Awarded LGU contracts. This finding further validates the fact that the LGUs have a one to one correspondence when awarding contracts when viewed in the perspective of ABCs. This essentially means that for every

Php1 spent, approximately Php1 remains unspent. This unspent public funds definitely is indeed indicative of inefficiencies in the public procurement process implemented by the LGUs [47], [48].

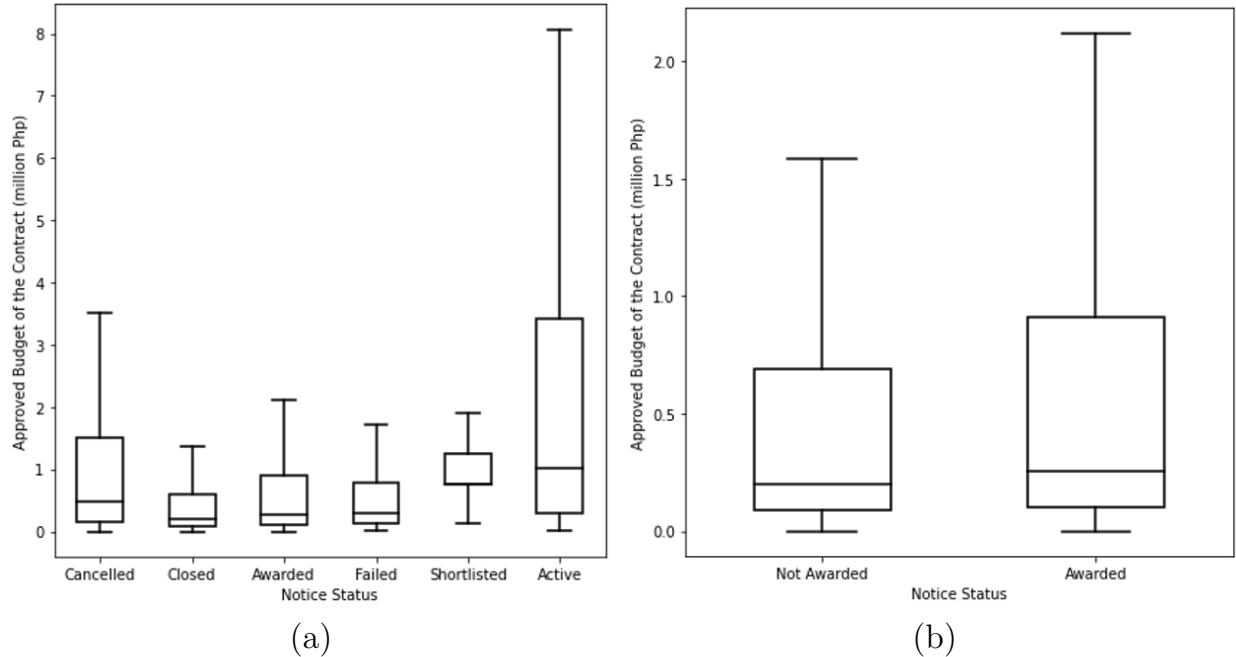


Figure 5: Boxplots of the Approved Budget Contracts (ABCs) of posted LGU contracts according to Notice Status. The average ABCs for Unawarded Contracts were found to be equivalent to that of Awarded contracts.

Contract Duration According to Notice Status

In terms of the Contract Duration (Days), the observed means for Awarded and Unawarded contracts were found to be almost equivalent at 10 days. While statistical examination of the means of these groups using Welch’s test strongly suggested that the observed Contract Duration (Days) for these LGU contracts are significantly different from each other, it is perceived that these relatively short means are representative of the spike in tenders related to the procurement of goods and services needed to alleviate the consequences of the COVID-19 pandemic. Streamlining these posted contracts further according to their Notice Status as shown in Figure 6, “active” contracts had higher means than all the

other contracts. As to be discussed further in this report, it is perceived that this spike could be attributed to substantial rise in construction projects (e.g. establishment of quarantine facilities) related to the control and mitigation of the COVID-19 pandemic in each of the LGUs. While no standard contract duration is to be followed, it is generally suggested to have lower contract durations so that the goods or services procured be delivered earlier to the procuring agency especially in this time of a health crisis [48].

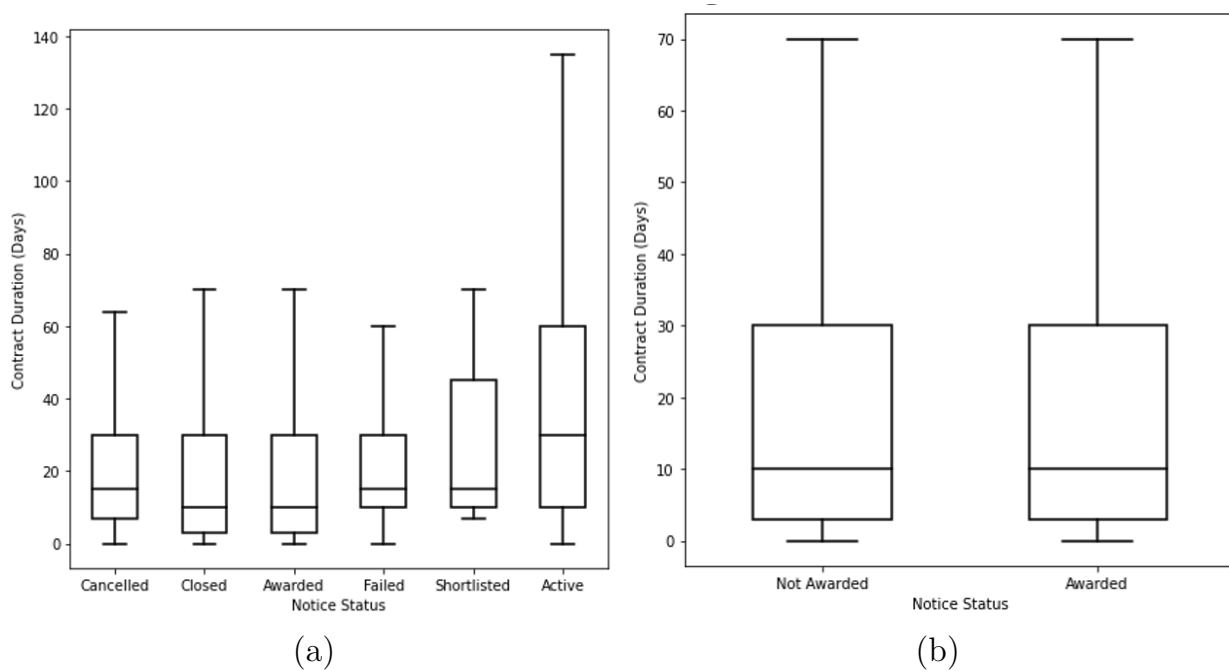


Figure 6: Boxplots of the Contract Duration (Days) of posted LGU contracts according to Notice Status. While seen to be almost equal to each other, the average Contract Duration (Days) for Unawarded and Awarded contracts were found to be statistically different ($p\text{-value} < 0.0001$).

Distribution of LGU Tenders According to Classification

As a support to the findings discussed in the previous section, the Classification of the tenders posted by each LGU was also explored. As shown in Figure 7, it was found that “Goods” were among the Classification being prominently procured by the LGUs. This is generally a good sign in terms of the procurement

behavior of LGUs since replenishment of consumables (e.g. consumables such as supplies, drugs and medicines) are critical in this time of a pandemic.

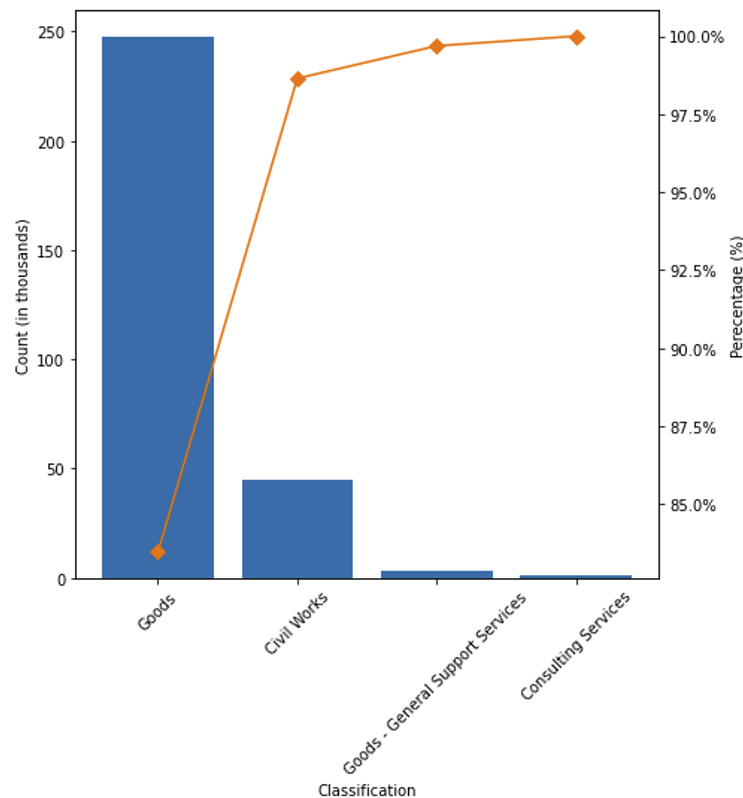


Figure 7: Pareto chart describing the distribution of Classification of posted LGU contracts. LGU contracts that have Classification of “goods” had the highest count.

Exploration of Approved Budget of the Contracts According to Classification and Business Category

While [Figure 7](#) might be indicative that LGUs posted numerous contracts for the replenishment of goods, further exploration of the PhilGEPS data strongly suggested that infrastructure projects were the ones with contracts having the largest budget allocation among all business categories. As shown in [Figure 8](#), LGU contracts classified under “Civil Works” garnered the highest average ABCs (contracts averaging Php1 million) while LGU contracts classified as “Goods” had a substantially low mean ABC. Moreover, [Figure 9](#) further val-

idates the fact that “construction projects” had the highest budget allocation corresponding to at least 40% of the total budget distributed for all LGUs in the country. While “drugs and medicines” was ranked second in terms of the total budget allocated by LGUs, this only represents at most 25% (or about Php50 billion) of the total investment made by LGUs into infrastructure projects. While it could be argued that infrastructure projects related to the establishment of more hospitals, COVID-19 isolation and quarantine facilities usually require larger approved budgets due to the inherent nature of such contracts (i.e. capital intensive), larger allocations for different types of goods and services necessary to mitigate and control the spread of the SARS-CoV-2 virus should also be considered. In the context of Philippine LGUs where budget allocated to these agencies is usually deliberated and approved by congress a year prior to implementation of such tenders, it would also be beneficial in the succeeding rounds of budget planning and deliberation to increase budget allocation for the purchase of COVID-19 vaccines in preparation for mass vaccination against the SARS-CoV-2 virus. As seen in the succeeding exploration of the PhilGEPS data (e.g. [Figure 15](#)), the word “vaccine” was not found in any of the details in the tenders posted by these LGUs.

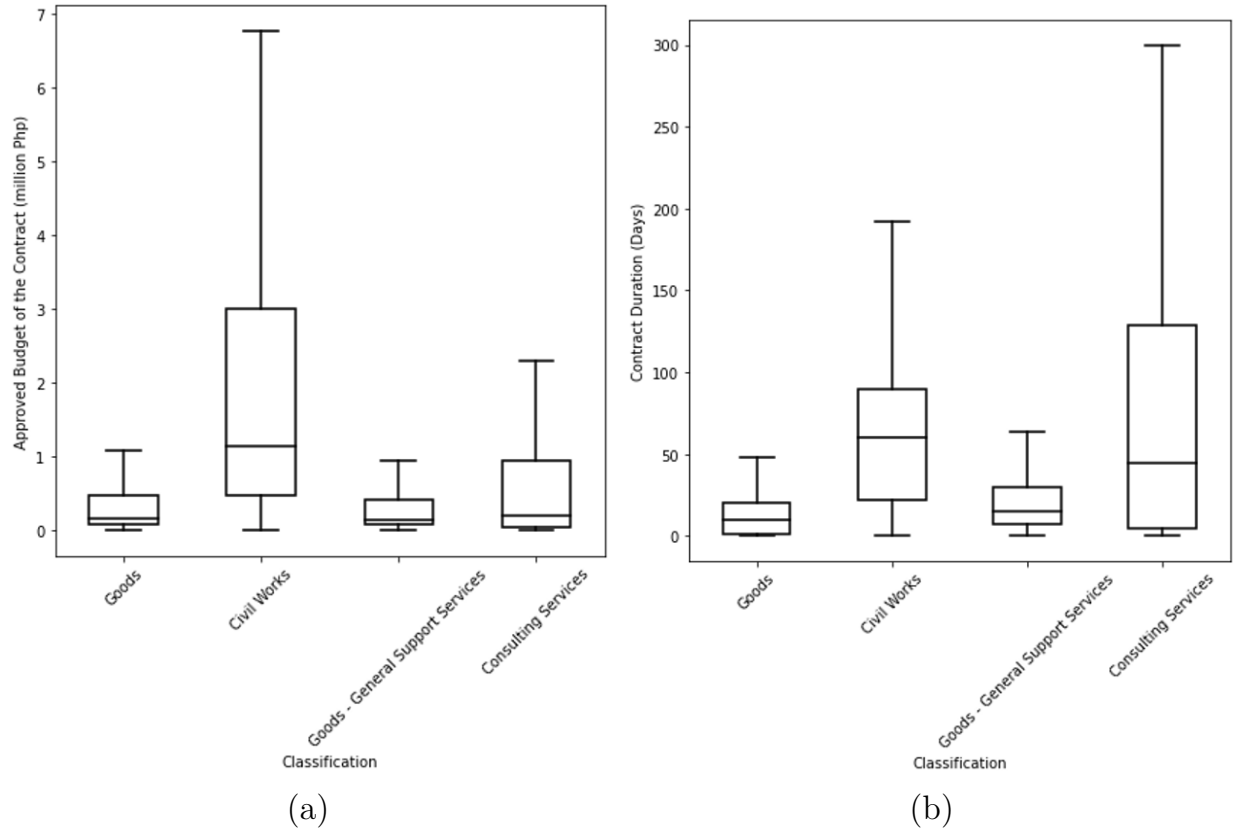


Figure 8: Boxplots of the (A) Approved Budget Contracts (ABCs) and (B) Contract Duration (Days) according to Classification. LGU contracts classified as Civil Works were found to have the highest mean for both ABCs and Contract Duration (Days).

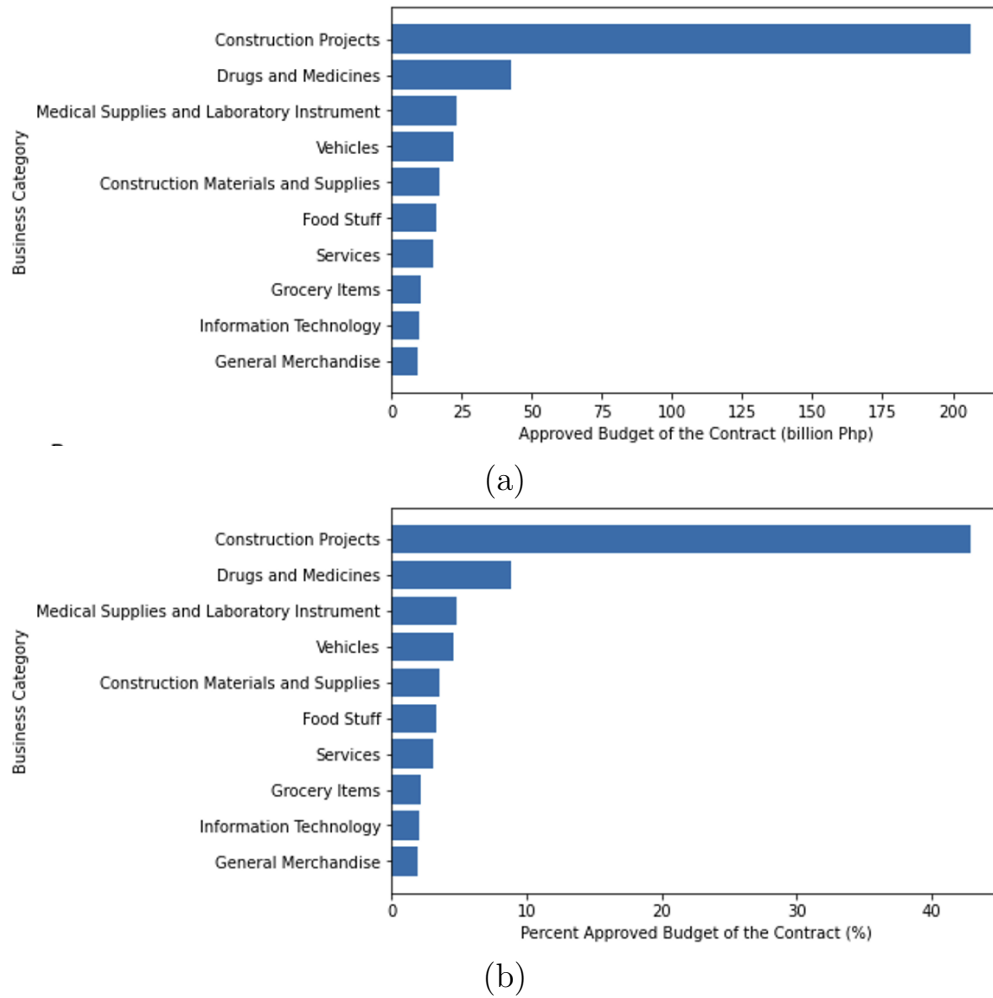


Figure 9: Barplots of the top 10 Business Categories of LGU contracts ranked according to the total Approved Budget Contracts (ABCs) in (A) billion Php and (B) total percentage. Construction projects had the largest budget allocation among all Business Categories.

WordCloud Visualization of PhilGEPS Dataset Containing Text Data

In parallel to what have been observed in Figure 7, Figure 8, and Figure 9, it could be generalized as shown in Figure 10 that office supplies, construction materials, medical supply, and drug and medicines were among the most common contracts posted by LGUs. Unfortunately, as previously discussed, bigrams (e.g. multi purpose, material construction, construction supply) related to infrastructure projects were among the most frequent words found in the

LGU posted contracts.

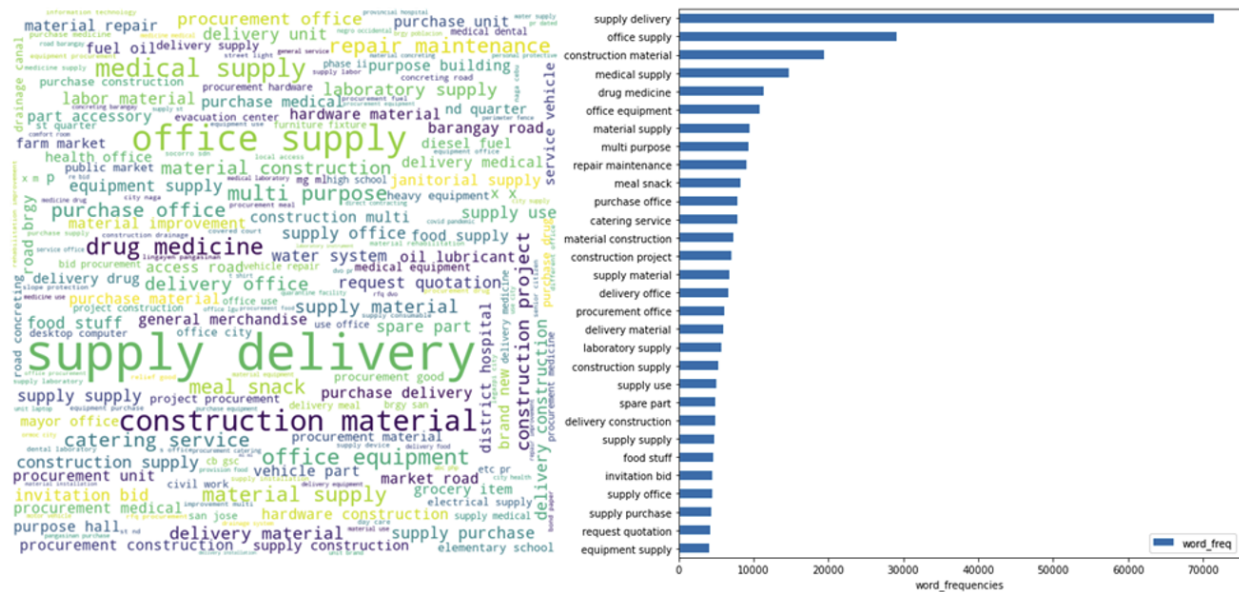


Figure 10: Wordcloud Visualization of Notice Title, Item Name, Item Desc, and Award Title features of the PhilGEPS dataset. LGU contracts related to office supply, construction material, and medical supply were among the bi-grams that had the highest frequency counts.

WordCloud Visualization of Reason for Award Feature

As part of the exploration done in this work, the text data under Reason for Award feature were also visualized. As validated from Figure 11, the prominent reason on why LGU contracts are awarded is based on the lowest calculated bid. This was found to be in-line with the Revised Implementing Rules and Regulations of Republic Act 9184 [37].

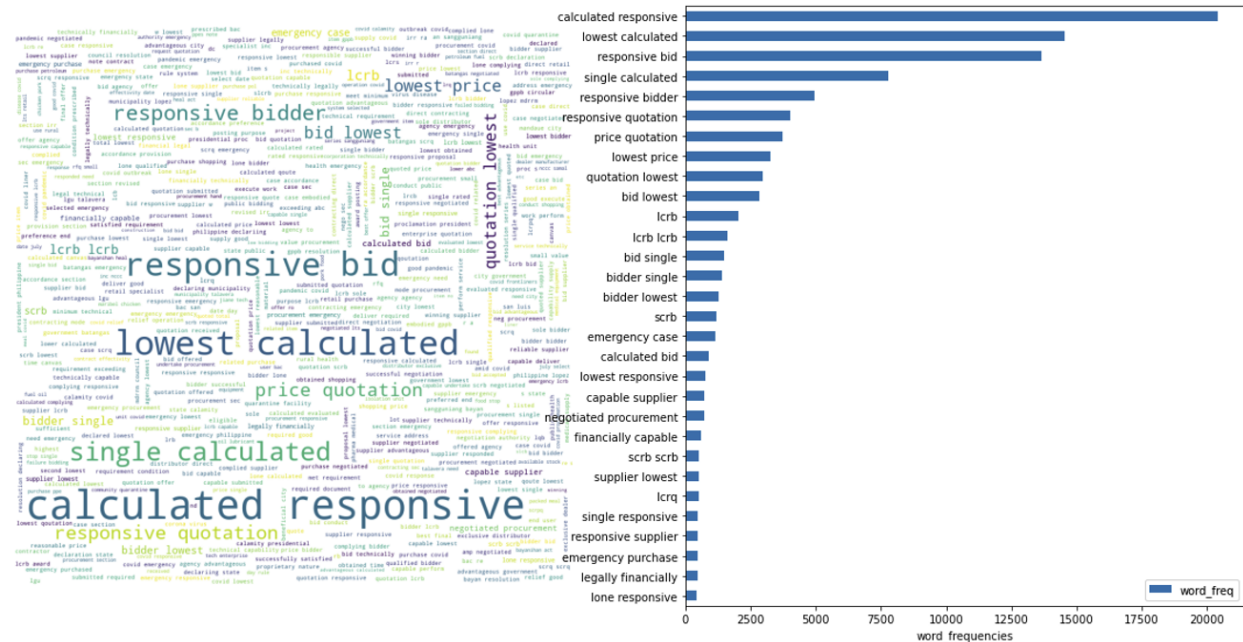


Figure 11: Wordcloud Visualization of Reason for Award feature of the PhilGEPS dataset. The primary reason for awarding LGU contracts were observed to be the bidder that offered the lowest calculated bid.

EXPLORATION OF LOCAL GOVERNMENT UNIT COVID-19 CONTRACTS

Further exploration of the PhilGEPS data revealed that about 8047 COVID-19 related tenders posted by LGUs were found for the fiscal year of 2020. The results of the exploration of these tenders are discussed in the succeeding sections.

Distribution of COVID-19 Contracts Posted by LGUs According to Notice Status

While LGU COVID-19 contracts categorized as “closed” had the highest frequency count, the percentage of awarded contracts were significantly higher

when compared to the results observed in Figure 4. It is still worth noting, however, that for every 1 COVID-19 contract awarded by an LGU, there is also at least 1 COVID-19 contract that remains unawarded. Intervention beyond that of the Bayanihan to Heal as One Act are still needed to increase the ratio between the number of awarded and unawarded contracts.

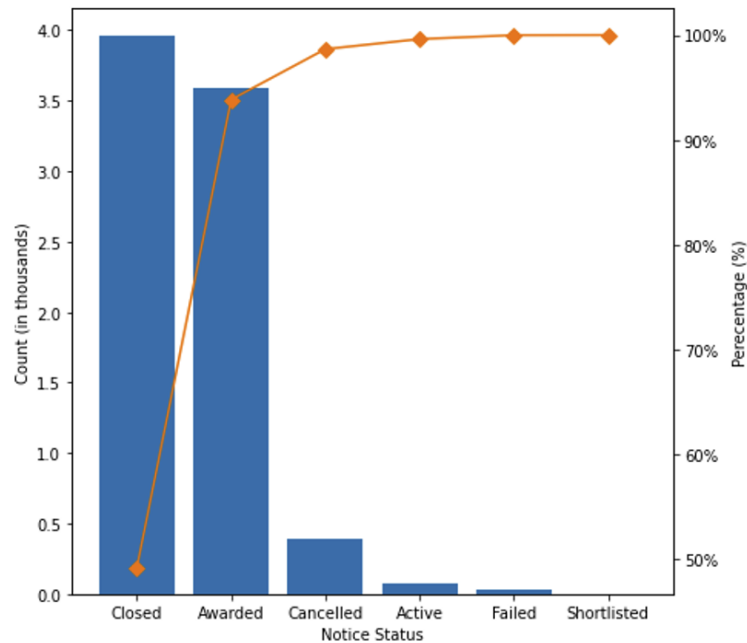


Figure 12: Pareto chart describing the distribution of Notice Status of posted LGU COVID-19 contracts. LGU COVID-19 contracts that have Notice Status of “closed” had the highest count.

Approved Budget Contract of COVID-19 Tenders According to Notice Status

Similar to what was done in Figure 5, the mean ABCs of the COVID-19 contracts posted by LGUs were also explored. Parallel to what was observed in Figure 12, Figure 13 further verified that the implementation of the Bayanihan to Heal as One Act was potentially effective since the mean ABCs of the unawarded COVID-19 contracts were significantly lower than that of the awarded COVID-19 contracts. It is still worth noting, however, that the mean ABCs for

both Active and Cancelled COVID-19 contracts were still higher than that of the Awarded contracts. This essentially means that the interventions intended to hasten the public procurement of goods and services related to the mitigation of the COVID-19 pandemic might not be fully maximized.

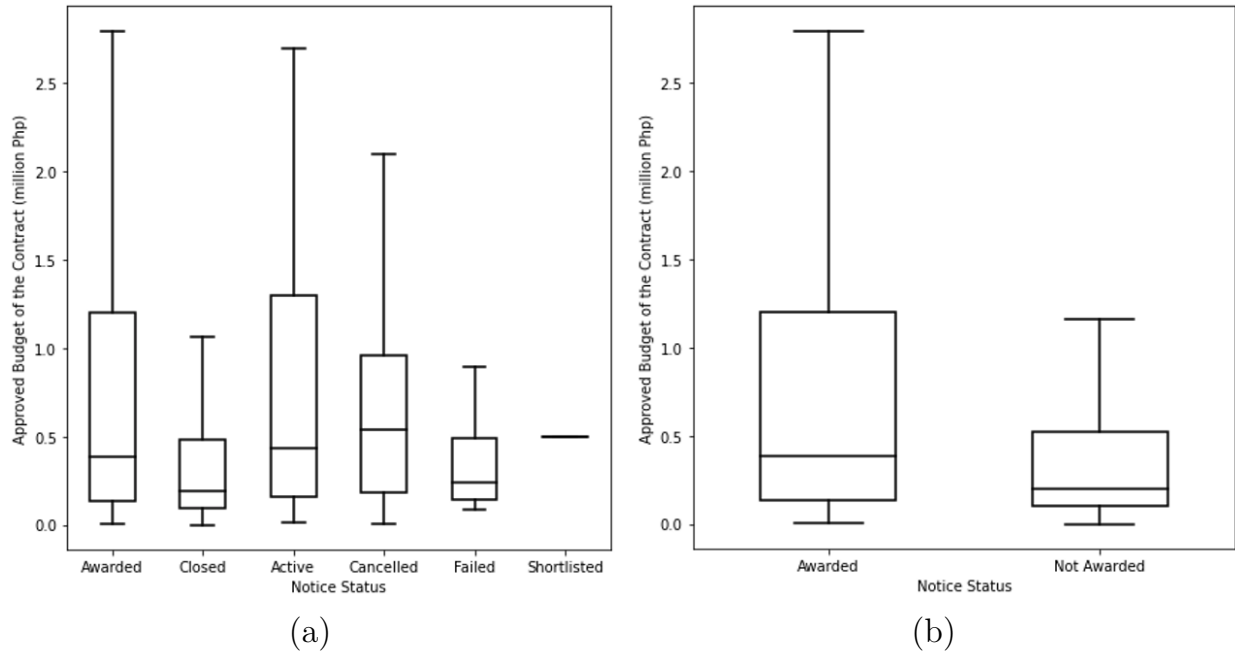
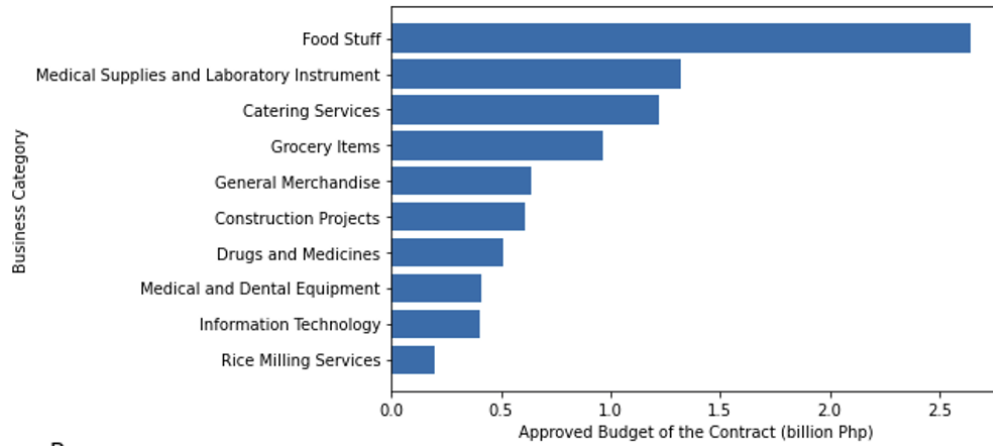


Figure 13: Boxplots of the Approved Budget Contracts (ABCs) of posted LGU COVID-19 contracts according to Notice Status. The average ABCs for Awarded contracts were found to be significantly higher to that of Unawarded contracts (p -value < 0.0001).

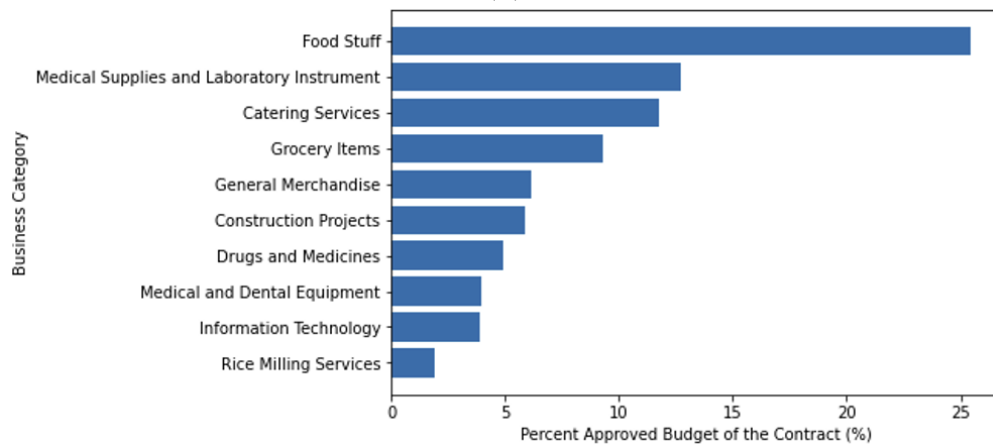
Budget Allocated on COVID-19 Contracts Posted by LGUs According to Business Category

In contrast to Figure 9, it could be seen from Figure 14 that construction projects were not the priority of the COVID-19 contracts posted by LGUs. Procurement of consumables related to relief efforts was found to be prominent in these contracts since nationwide lockdowns were first implemented in the first quarter of 2020. With about 25% of the COVID-19 budget allocated to “food stuff”, these contracts clearly represent the efforts made by all LGUs to pro-

vide basic necessities to all families affected by control measures implemented to lessen the human-to-human transmission of the SARS-CoV-2 virus (see [Figure 15](#)). While these efforts by the LGUs are timely and desirable, it is very unfortunate that these COVID-19 contracts (both awarded and unawarded) only denotes about 2.16% of the total budget allocated to all LGUs. This essentially meant (with the exclusion of negotiated COVID-19 contracts and COVID-19 contracts not posted in the PhilGEPS website) that only Php10 billion out of Php481 billion was appropriated for efforts to mitigate the consequences of the COVID-19 pandemic. At a population of 100 million, this appropriated budget implied that every Filipino would be given an allocation of Php100 (about \$2 US Dollars) at the minimum. While it could be argued that the allocated budget was already passed as a law in the previous fiscal year (i.e. 2019), this allocation for every Filipino should be significantly increased to at least Php2000 (about \$40 US Dollars) to cover vaccination costs in 2021.



(a)



(b)

Figure 14: Barplots of the top 10 Business Categories of LGU COVID-19 contracts ranked according to the total Approved Budget Contracts (ABCs) in (A) billion Php and (B) total percentage. Food Stuff had the largest budget allocation among all Business Categories.

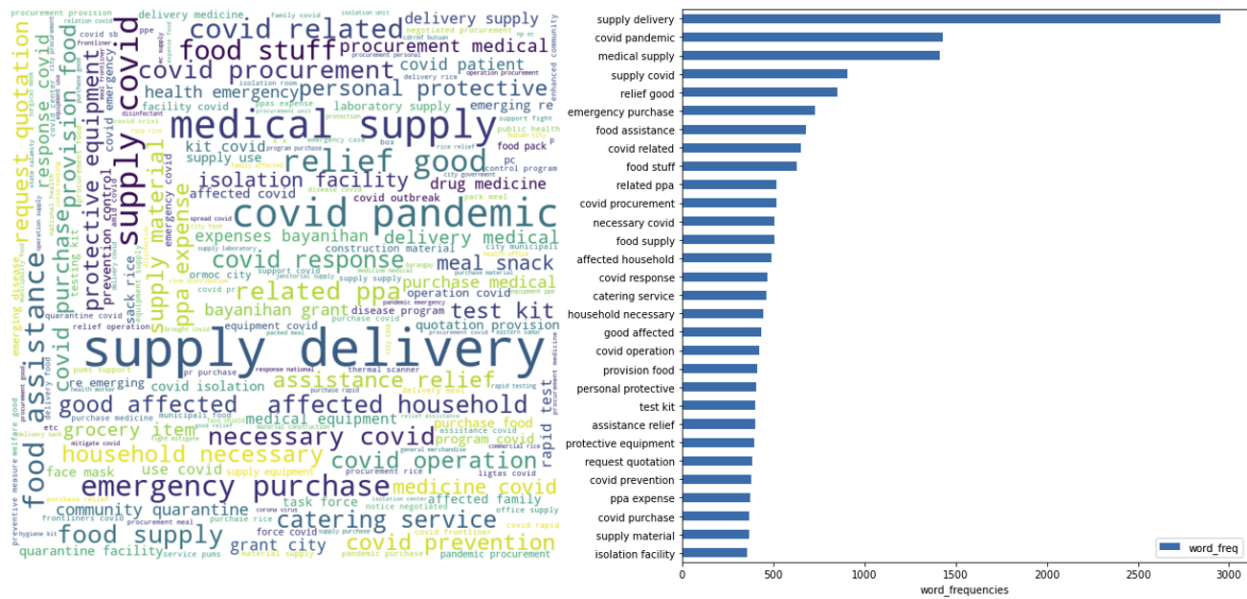


Figure 15: Wordcloud Visualization of Notice Title, Item Name, Item Desc, and Award Title features of the PhilGEPS dataset. LGU COVID-19 contracts related to the procurement of relief goods (e.g. medical supply, relief good, food assistance) were among the bi-grams that had the highest frequency counts.

About 786 contracts were flagged as potentially irregular since the difference between the Notice to Proceed and Award dates were found to be negative. As previously mentioned, this is in clear violation of the Revised Implementing Rules and Regulations of Republic Act 9184 [37]. Details on these contracts were further explored in order to determine the amount of public funds potentially lost from these transactions. These details are discussed in the succeeding sections.

Top Local Government Units Identified to have Potentially Irregular Contracts

Exploration of the LGU contracts flagged as potentially irregular revealed that the City of Cebu (Province of Cebu) and the Municipality of Tuba (Province of Benguet) had the greatest number of contracts flagged as potentially irregular and had the most negative difference of Award-NTP (Days), respectively. This observation is validated in [Figure 16](#). About 1 out of 7 potentially irregular contracts could be traced from the City of Cebu (Province of Cebu) and City of Malaybalay (Province of Bukidnon) while the contracts having the most negative difference in the Award-NTP (Days) feature were observed for Municipality of Tuba (Province of Benguet) and City of Cebu (Province of Cebu). These findings strongly call for a deeper investigation by authorities on why these have been observed for these LGUs.

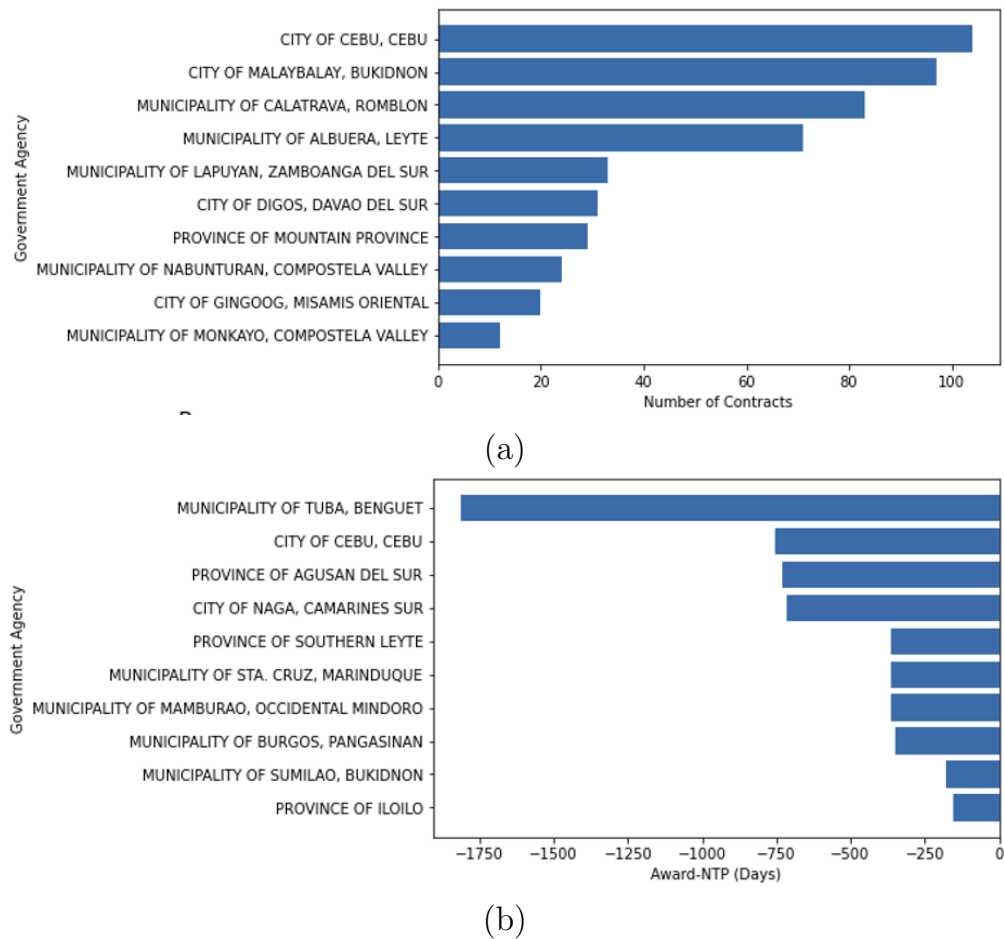


Figure 16: Barplots of the top 10 LGUs having the (A) greatest number of potentially irregular contracts and (B) most negative difference in Award-NTP (Days) feature.

Budget Allocated on Potentially Irregular Contracts Posted by LGUs According to Business Category

Further examination of the ABCs of the contracts classified into Business Category revealed that a total of Php720 million were allocated to potentially irregular contracts. About 24% of Php720 million were allotted to Construction Projects while Food Stuff and Catering Services combined represented about 25% of this total amount. While the deliverables of these contracts might have already been completed at the present time, the irregularities observed are potentially indicative of collusion between LGU personnel and the winning bidder.

For instance, citing an audit observation below made for City of Cebu in connection to the issuance of Notice to Proceed date strongly support the fact that contracts having negative Award-NTP (Days) are potentially indicative of connivance between LGU personnel and a winning bidder (note that the names of the persons and the company involved were not included in the verbatim citation of this audit observation) [36]:

The Construction Agreement between the City of Cebu and the winning bidder was not entered into and the Notice to Proceed (NTP) was not issued before the start of the construction of the Quarantine Building at Block 27, North Reclamation Area, Cebu City for Php58,851,245.67, contrary to Government Procurement Policy Board (GPPB) Circular 01-2020 dated April 6, 2020, thus casting doubt on the legality of the procurement.

Following this audit observation closely [36], it was revealed through an unnumbered memorandum that the end-user of this contract was directed to negotiate directly with the winning bidder. While negotiated procurements were allowed under the Bayanihan to Heal as One Act [49], adherence to documentary requirements, however, prior to award of contract is still enforced (e.g. issuance of Notice to Proceed order was made prior to issuance of Construction Agreement) as directed in the Government Procurement Policy Board (GPPB) Circular 01-2020 [50]. As disclosed in this full audit report, construction of the Quarantine Building was already on-going even prior to the compliance of the winning bidder to documentary requirements and Notice to Proceed issuance to be issued by the LGU. Moreover, funds were already disbursed in advance even with the unqualified acceptance of the contractor for this awarded project. Note that the LGU and the contractor did not yet enter into a Construction Agreement at this time rendering the disbursement of funds highly irregular. This observed violation is substantial evidence that the legality of this procurement is indeed questionable as this insinuates prejudice against the interest of the LGU and imply that such violations strongly suggest collusion between both parties since funds cannot be disbursed in the absence of documentary requirements. Collectively, such contracts potentially represent a significant portion public

funds being lost or misused due to corrupt practices in the public procurement process.

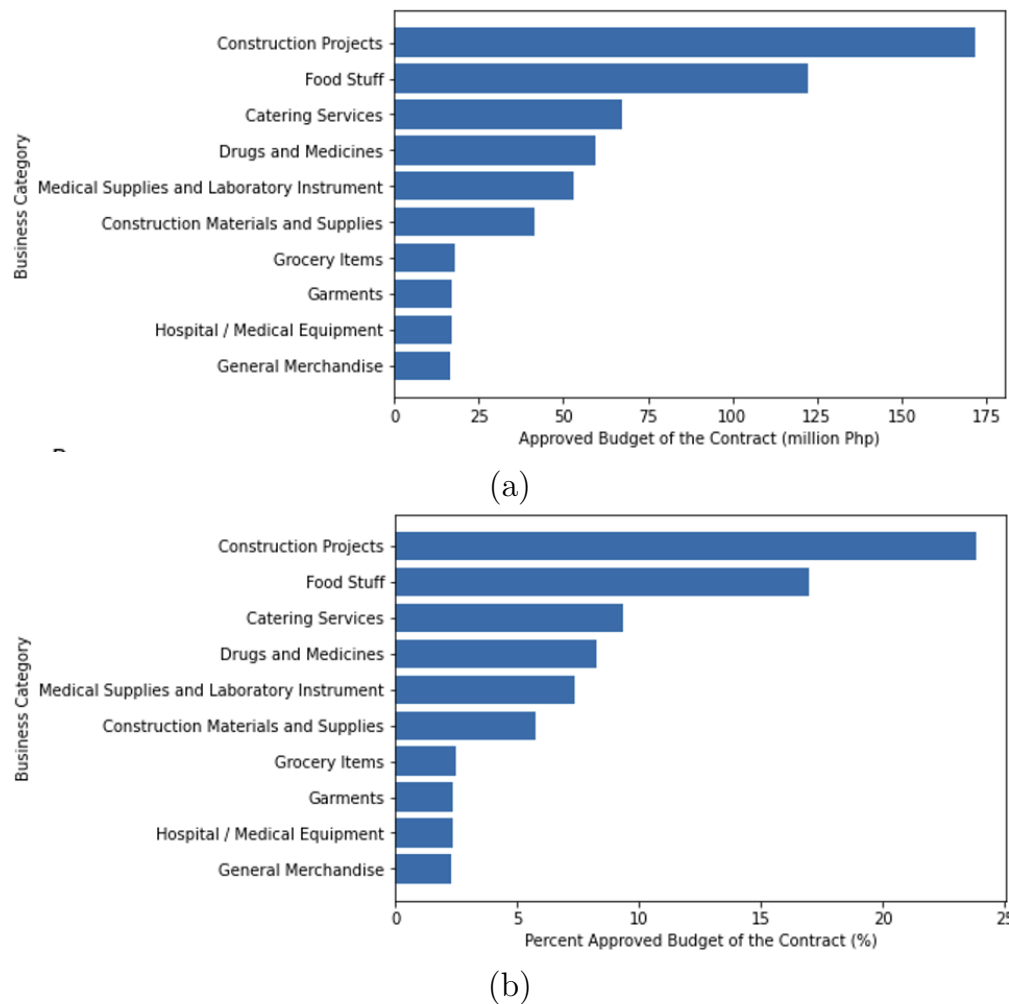


Figure 17: Barplots of the top 10 Business Categories of LGU contracts flagged as potentially irregular ranked according to the total Approved Budget Contracts (ABCs) in (A) million Php and (B) total percentage. Construction Projects and Food Stuff were identified to had the largest budget allocation among all Business Categories.

GENERALIZATION OF AUDIT FINDINGS THROUGH HIERARCHICAL CLUSTERING

Following the discussion made for identified irregular contracts, it would also be beneficial for LGUs to determine the major audit findings issued to them. In this regard, identification of audit topics through clustering would help these LGUs attend to usual problems encountered in their procurement process. Discussion of the clusters formed from both Audit Observations and Audit Recommendations are discussed in the succeeding sections.

Generalization of Audit Observations through Hierarchical Clustering

As shown in [Figure 18](#) and [Figure 19](#), two clusters were observed to be formed from the results of the implemented hierarchical clustering approach through the Ward's Method. Audit observations from 1126 annual audit reports were clustered into Cluster 1 while the remaining audit observations from 5128 annual audit reports were clustered into Cluster 2. Interpretation of the formed clusters are as follows:

- **Financial Statements and Inventory Management Audit Observations (Cluster 1):** The most prominent bi-grams observed for this cluster were “financial statement”, “property equipment”, “plant equipment”, “coa circular“, ”government accounting”, “ppe account”, and “physical count” (see [Figure 20](#)). It then could be generalized from these bi-grams that these cluster contains audit observations that pertain to discrepancies to financial statement caused by potential violations of existing COA circulars for inventory management or disposal. Sampling an audit observation from a document in this cluster further validated this interpretation. The sampled audit observations included in this cluster are given below:

- Document 6242: The existence, completeness and condition of all property, plant and equipment (PPE) accounts of the Municipality with net book value of Php36.630 million for all funds cannot be ascertained due to partial conduct of physical count on some properties totaling Php10.046 million book value by the Inventory Committee in compliance with Section 124, Volume I of New Government Accounting System (NGAS) Manual for Local Government Units (LGUs) and Item No. 5.1 of COA Circular No. 2020-006, thus, cast doubt on the fair presentation of the said accounts in the financial statements.
- Document 6243: Inventory taking by the Inventory Committee to PPE items totaling of Php10,046,602.00 may have been done inefficiently and ineffectively due to non-preparation of Physical Inventory Plan (PIP) prior to its partial conduct of physical count in contrary to Item No. 5.9 of COA Circular No. 2020-006.
- **Disaster Risk Reduction and Management, and COVID-19 Related Procurement Audit Observations (Cluster 2):** The most prominent bi-grams observed for this cluster were “coa circular”, “disaster risk”, “risk reduction”, “date april”, “reduction management”, “republic act”, and “memorandum circular” (see [Figure 21](#)). It then could be generalized from these bi-grams that these cluster contains audit observations that pertain to discrepancies or potential violations to existing circulars, memorandums, and Philippine laws made in procuring goods or services in times of the COVID-19 pandemic. Sampling an audit observation from a document in this cluster further validated this interpretation. The sampled audit observations included in this cluster are given below:
 - Document 23: Payments of COVID-19 Hazard Pay to the regular, COS and JOs of the City Government amounting to Php10,006,200.00 covering the period March 17, 2020 to May 15, 2020 were not supported with adequate documents to establish the validity of claims which is not in accordance with Section 2(b) of Administrative Order (AO) No. 26 dated March 23, 2020, Section 7 of Department of Budget and Man-

agement (DBM) Budget Circular No. 2020-1 and Section 4(6) of PD No. 1445, thus the propriety of the payment of the Hazard Pay could not be established.

- Document 203: Disbursements amounting to Php1,108,625.49 charged against the Bayanihan Grant to Cities and Municipalities (BGCM) were not among those programs, projects or activities (PPAs) specifically authorized under Local Budget Circular No. 125, which may have undermined the Municipality’s COVID-19 measures and responses.

Collectively, these identified validate the fact that irregularities were indeed present in the contracts posted by LGUs in times of the COVID-19 pandemic.

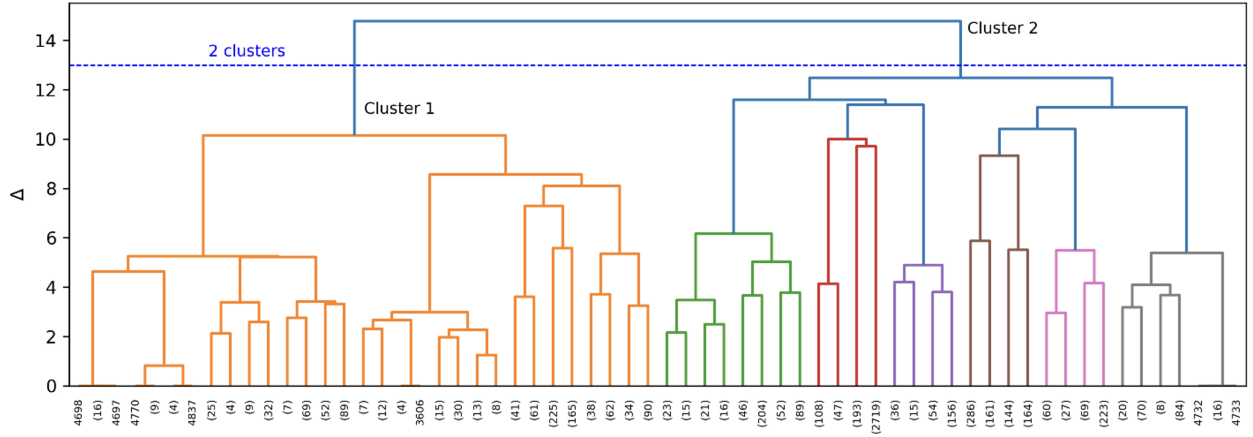


Figure 18: Results of hierarchical clustering on the pooled audit observations from 6254 annual audit reports. Two big clusters have been observed to be formed at a Δ value of 13.

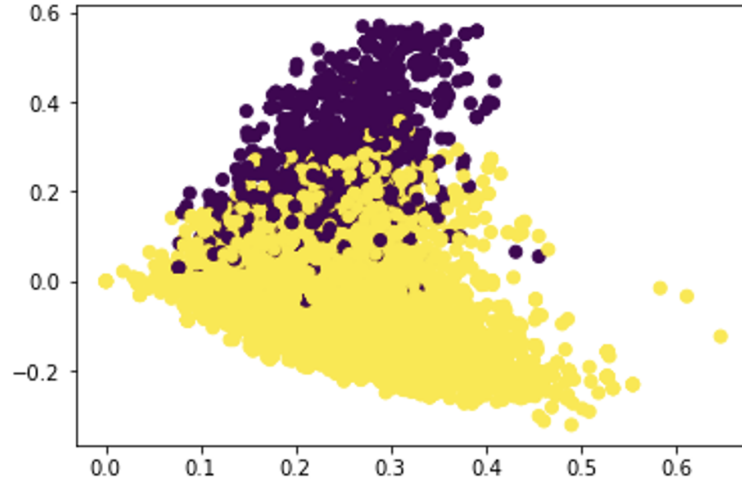


Figure 19: Visualization of the clusters formed from the pooled audit observations through hierarchical clustering. The two different colors indicate the presence of two clusters.

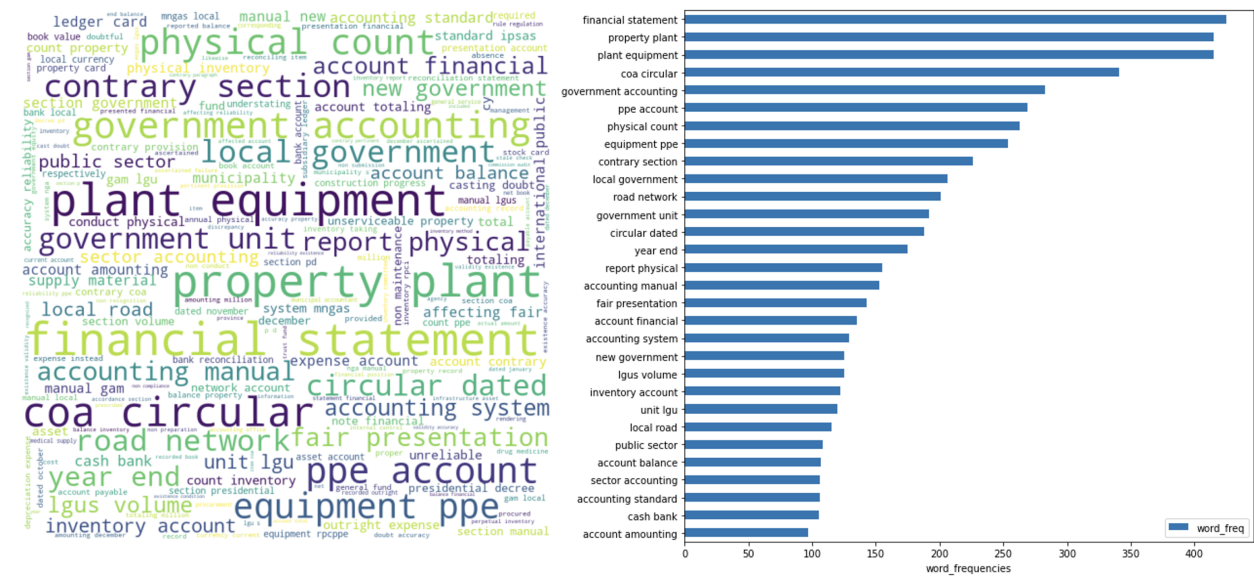


Figure 20: Wordcloud Visualization of Pooled Audit Observations for Cluster 1. The bi-grams “financial statement”, “property plant”, “plant equipment”, and “coa circular” were found to be the most prominent for this cluster.

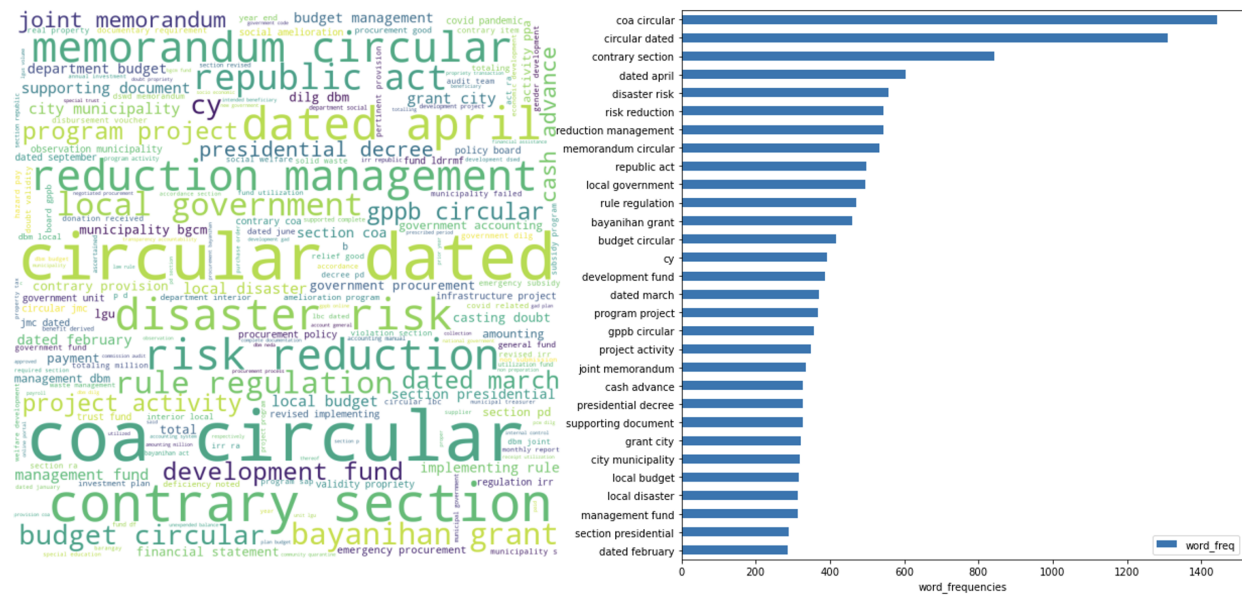


Figure 21: Wordcloud Visualization of Pooled Audit Observations for Cluster 2. The bi-grams “coa circular”, “disaster risk”, “risk reduction”, “memorandum circular”, and “republic act” were found to be the most prominent for this cluster.

Generalization of Audit Recommendations through Hierarchical Clustering

As shown in Figure 22 and Figure 23, two clusters were observed to be formed from the results of the implemented hierarchical clustering approach through the Ward’s Method. Audit recommendations from 688 annual audit reports were clustered into Cluster 1 while the remaining audit observations from 5566 annual audit reports were clustered into Cluster 2. Interpretation of the formed clusters are as follows:

- **Audit Recommendations to Financial Statement and Inventory Management Audit Observations (Cluster 1):** The most relevant bi-grams observed for this cluster were “municipal accountant”, “physical count”, and “inventory committee” (see Figure 24). It could then be generalized from these bi-grams that this cluster contains audit recommendations

that pertain to auditor suggested solutions to the observed discrepancies in LGU financial statements caused by potential violations of existing COA circulars for inventory management or disposal. As clearly seen from the visualization made, the municipal accountant and the inventory committee were identified to be the key persons in ensuring that the financial statements and inventory of an LGU are put in order. Sampling an audit observation from a document in this cluster further validated this interpretation. The sampled audit observations included in this cluster are given below:

- Document 215: We recommended that Management require the Municipal Accountant to maintain the Supplies Ledger Cards per stock number, PPE Ledger Cards for each category of assets and Real Property Ledger Cards for land and for the Municipal Treasurer to maintain the Stock Cards per stock number and Property Cards per category of PPE which are necessary to support the accuracy and reliability of the balances of Inventories and PPE accounts in the financial statements.
- Document 215: We also recommended that Management require the conduct of physical count of all its PPE through the Inventory Committee composed of adequate number of members and to reconcile the inventory count with property and accounting records through the Property/Treasury and Accounting Unit consistent with the guidelines and procedures provided in COA Circular No. 2020-006 dated January 31, 2020.
- **Audit Recommendations to Discrepancies in Disaster Risk Reduction or COVID-19 Related Procurement (Cluster 2):** The most prominent bi-grams observed for this cluster were “municipal accountant”, “coa circular”, “chief executive”, “local chief”, “municipal treasurer”, “municipal mayor”, “supporting document”, “require municipal”, and “direct municipal” (see [Figure 25](#)). It could then be generalized from these bi-grams that these cluster contains audit recommendations that pertain to auditor suggested solutions to the observed discrepancies or potential violations of

LGUs to existing circulars, memorandums, and Philippine laws made in procuring goods or services in times of the COVID-19 pandemic. Similar to Cluster 1, the municipal accountant was also identified to be a key person in keeping the LGU procurement process in check. It is worth noting, however, that the end-users (e.g. chief executive, municipal treasurer, municipal mayor) of the posted LGU contracts were also identified to be key persons in keeping the LGU procurement process in check. Moreover, it then could be inferred from this cluster that compliance to documentary requirements (or document submission) is consistently being echoed as the primary solution to the audit findings made by COA auditors for 2020. Sampling an audit observation from a document in this cluster further validated this interpretation. The sampled audit observations included in this cluster are given below:

- Document 3553: Post all procurement of goods and supplies related to COVID-19 relative to the Procurement Projects undertaken in the GPPB Online Portal the necessary data as enumerated in the GPPB Circular issuances. These data or information shall be used by the GPPB and its Technical Support Office to comply with the reportorial requirements under the Bayanihan Act.
- Document 4982: We recommended that Management prepare an Alternative Work Arrangement (AWAs) per department to be approved by the head of the agency, as prescribed by CSC Memorandum Circular No. 10, series of 2020, hereafter, furnish the Office of the Auditor copies of AWAs and DTRs to support the claim for COVID-19 hazard pay.

Collectively, identification of accountable officers and as well as the compliance to documentary requirements were essentially the main themes of the identified clusters.

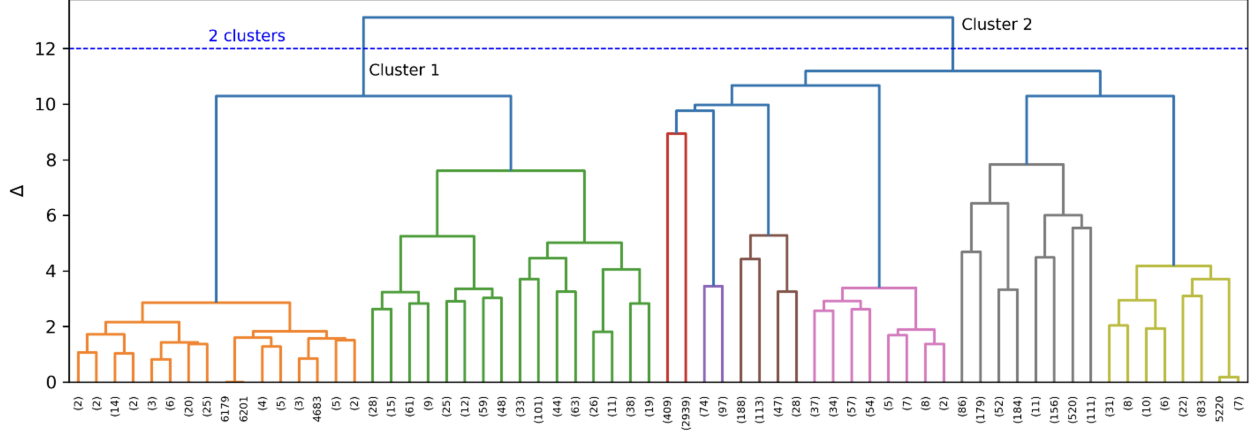


Figure 22: Results of hierarchical clustering on the pooled audit recommendations from 6254 annual audit reports. Two big clusters have been observed to be formed at a Δ value of 12.

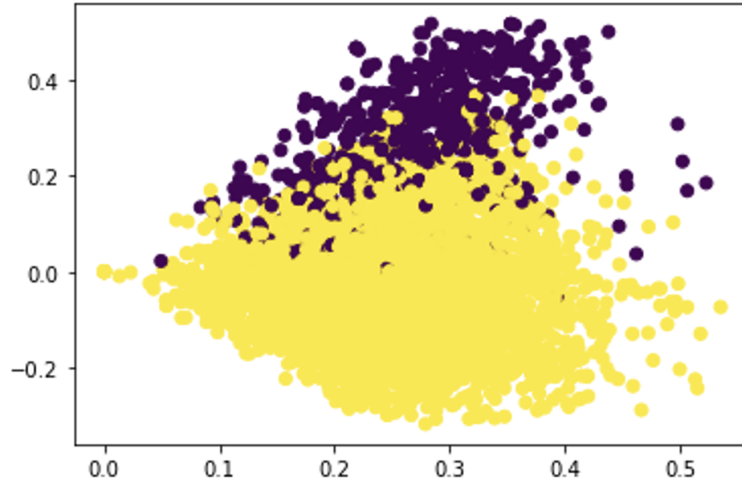


Figure 23: Visualization of the clusters formed from the pooled audit recommendations through hierarchical clustering. The two different colors indicate the presence of two clusters.

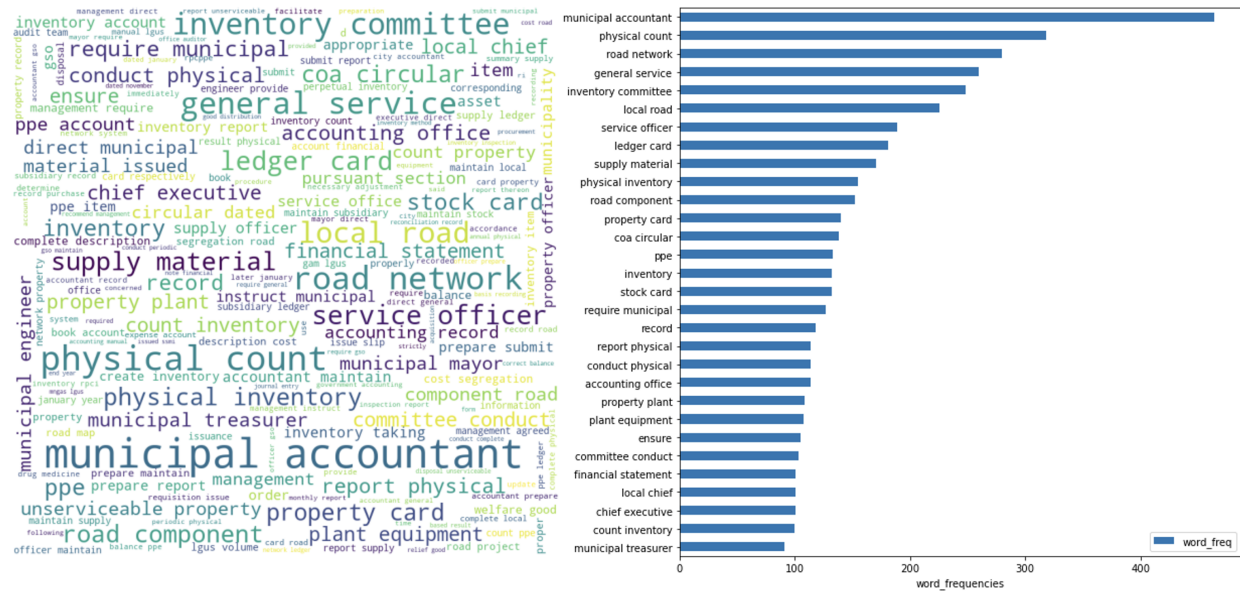


Figure 24: Wordcloud Visualization of Pooled Audit Recommendations for Cluster 1. The bi-grams “municipal accountant”, “physical count”, “inventory committee” were found to be the most relevant for this cluster.

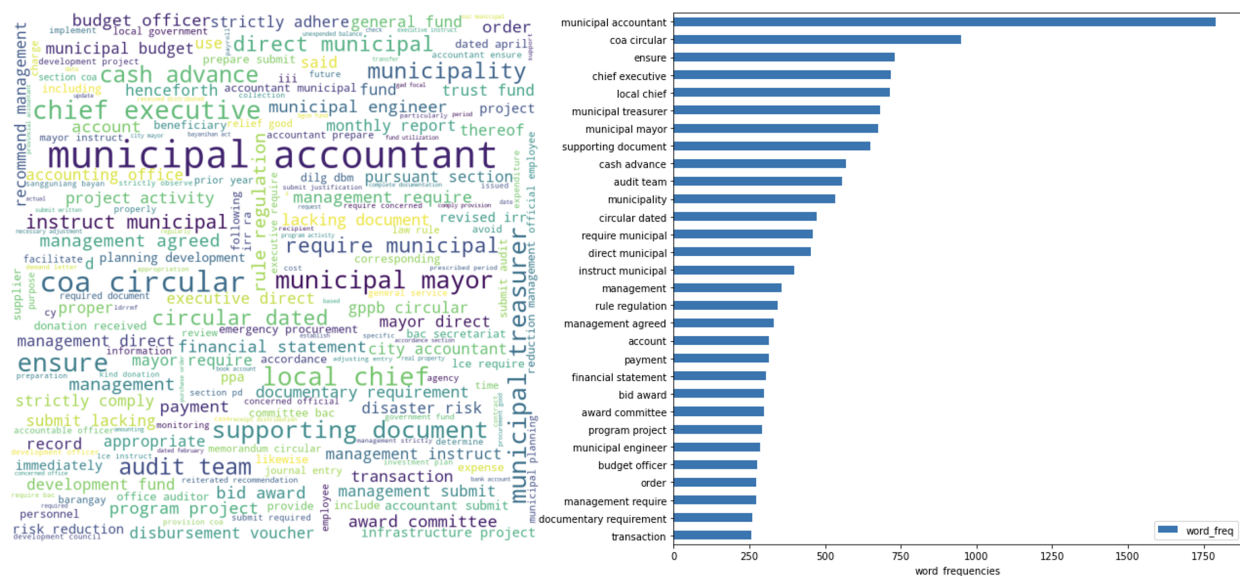


Figure 25: Wordcloud Visualization of Pooled Audit Recommendations for Cluster 2. The bi-grams “municipal accountant”, “coa circular”, “chief executive”, “local chief”, “municipal treasurer”, “municipal mayor”, and “supporting document” were found to be the most prominent for this cluster.

Practical Implication of the Identified Audit Topics

Now that it has been established that two main topics encompass each of the audit observations and audit recommendations investigated (total of four topics), it is worth noting that these clustered audit findings present an overview of what LGUs typically violate in the implemented procurement process and what typically needs to be done in order to prevent such violations. As seen from the results, it could be inferred that violations primary revolve around compliance to documentary requirements and unauthorized cash advances as signaled by the bi-grams “supporting document” and “cash advance” and further elaborated in the examined annual audit reports (e.g. Document 215). Parallel to this, submission of documentary requirements and immediate return of the cash advances were found to be generally recommended.

Aside from a general overview of both audit observations and audit recommendations, the sub-clusters of audit findings formed could also provide an avenue of reducing the time it would need to audit awarded contracts. For example, audits are usually done on procurement activities conducted on the previous fiscal year since audits usually involve time-intensive activities such as data collection and site inspection. While this may seem to be the usual practice, the increase in government transactions over the past decade have essentially doubled thereby making auditing of awarded contracts more tedious than before due to the significant increase in volume. Hence, it might be more efficient and appropriate if audits are actually conducted on the previous quarter of the present fiscal year. But then again, this is not possible due to the time constraints in the conduct of the audit. By utilizing then the sub-clusters identified in this work, it would now be possible to narrow down activities and cut the time required for site visits and inspection of auditors since they would now have a grasp on what an LGU or group of LGUs typically violate in the public procurement process. In essence, auditors could use these grouped audit findings to help them create a customized template of their audit plans tailor-fitted

to the past and present procurement practices of an LGU (or even a group of LGUs). While this work was not anymore able to provide a proof of concept for this insight due to the limitation on computational resources, these formed clusters are seen to be substantial foundations for the development of a recommender system that could help partially automate the creation of these audit plans.

TRAINING AND VALIDATION OF LOGISTIC REGRESSION MODEL

To collate all of the findings discussed in this work as an actionable solution, a prediction model through a logistic regression model capable of identifying irregular contracts based on a set of features equivalent to that of what is made available in the PhilGEPS website was implemented. Development and assessment of this prediction model is discussed further in the succeeding sections.

Reduction of Features through Dimensionality Reduction

Prior to training the logistic regression model for predicting “irregular” and “not irregular” LGU contracts, dimensionality reduction using truncated singular value decomposition was initially employed. At a cumulative explained variance ratio sum of 80%, the 934 features of the PhilGEPS 2020 procurement data were successfully reduced to 86 singular vectors. The three (3) most relevant singular vectors resulting from this dimensionality reduction are enumerated in [Figure 26](#). From these results, the following interpretation of the singular vectors could then be drawn:

- **Features Relative to Contracts Classified as “Goods” (SV1):** This is the most relevant singular vector from the 86 singular vectors formed. This is to be expected since majority of the contracts fall under the clas-

sification “Goods” that are unawarded. It could be seen that top 10 listed features were found to be positively correlated to this singular vector.

- **Features Relative to Contracts Classified as “Civil Works” (SV2):** This singular vector represents contracts classified under “Civil Works”. It could be seen that the procurement mode of these contracts was primarily made through public bidding as indicated by the positively correlated features to this singular vector.
- **Features Relative to Awarded Contracts (SV3):** This singular vector collates all relevant information that pertain to features that influence how a contract is awarded to a winning bidder. It could be inferred that contracts are awarded to the bidder with lowest calculated bid and is usually awarded at the 3rd quarter of the fiscal year.

Dimensionality reduction was performed in order to capture all relevant features that affect the target feature and as well as manage the time needed to train the logistic regression model.

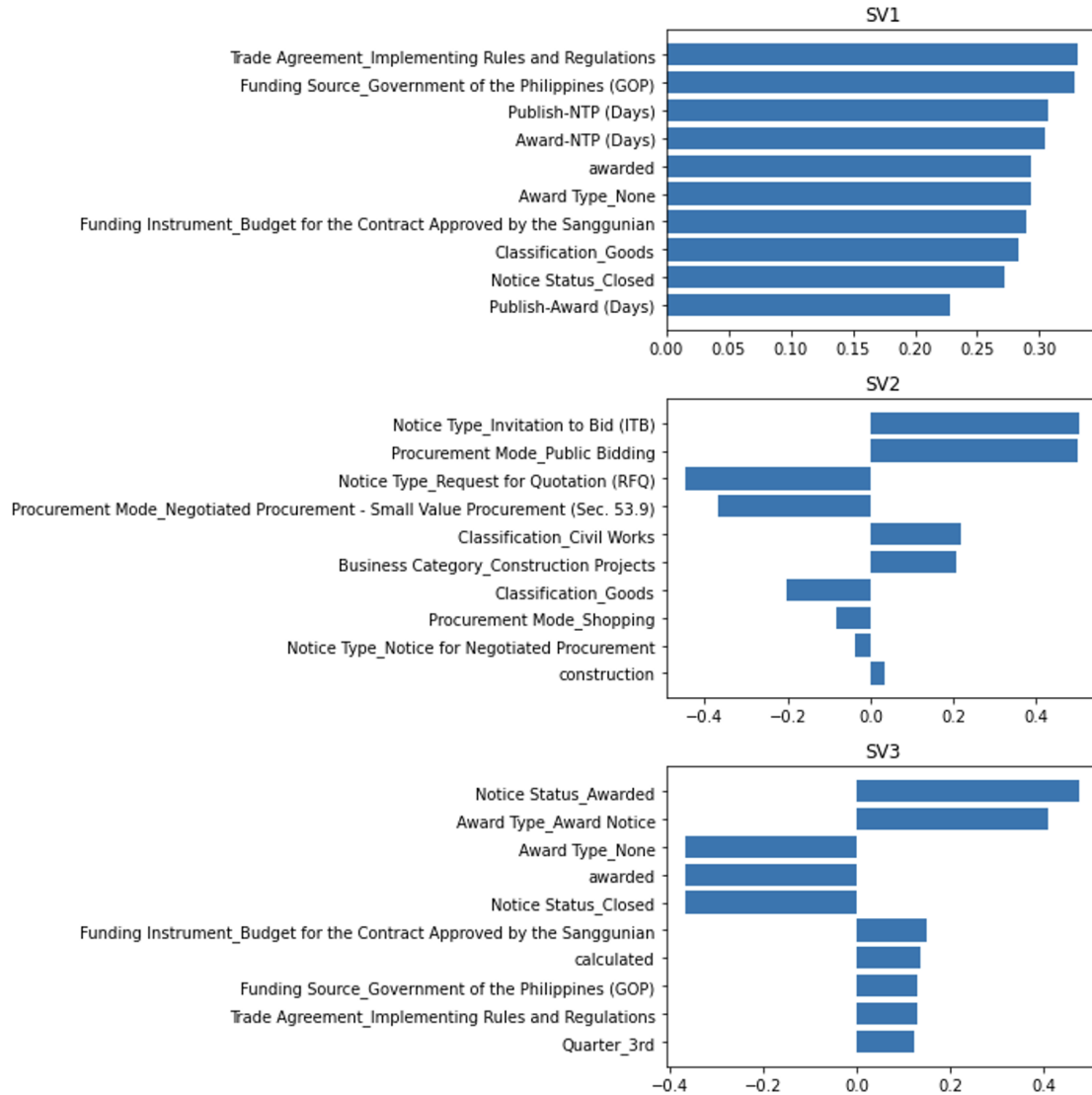


Figure 26: The three (3) most relevant singular vectors of the 934 feature PhilGEPS 2020 Procurement Data.

Training and Validation of Logistic Regression Model

As shown in the visualization under Figure 27, it is further confirmed that the classification problem at hand comprised of a highly imbalanced dataset. Hence, as previously discussed, using the dimensionally reduced data, stratified k-fold and SMOTE were implemented in order to deal with this observed imbalance in the distribution of the target feature.

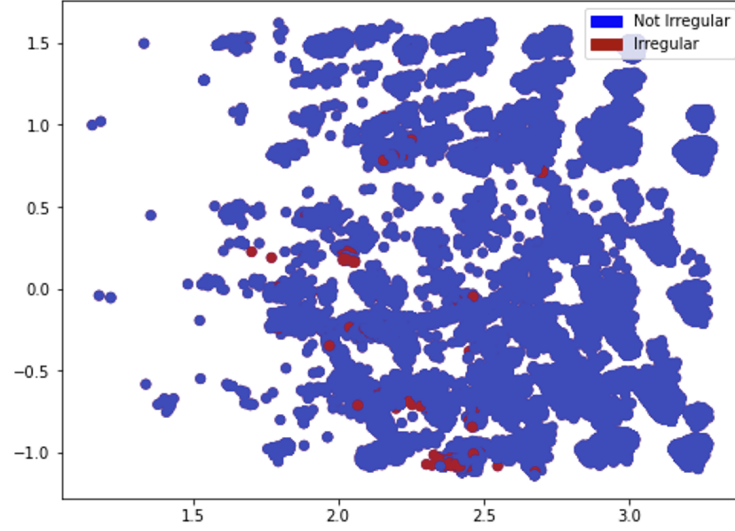


Figure 27: Visualization of the distribution of irregular and not irregular LGU contracts. It could be seen from this visualization that the distribution of the target feature of the PhilGEPS 2020 procurement data is highly imbalanced.

Training of the logistic regression through this approach resulted to a test accuracy of about 91.29% at an optimal hyperparameter value of $C = 0.001$. This accuracy is relatively high and this essentially means that the model is viable for deployment. However, since the dataset is highly imbalanced, it is also necessary to inspect whether the developed model is correctly classifying potentially “irregular” LGU contracts. Through the use of a confusion matrix shown in [Figure 28](#), it could be seen that the developed model has a precision score of 0.029 (the measure of the number of correct predictions made by an algorithm) and a recall score of 1.0 (the measure of the number of right predictions made by an algorithm relative to the total positive predictions made). It is worth noting however that about 9.6% of the “not irregular” contracts were also classified as potentially “irregular” contracts. While it would be expected that a significant number of contracts would be flagged as potentially “irregular”, it could be argued that retrieving all of these potentially “irregular” contracts is more important in this classification problem since it would be more costly for an LGU to not retrieve these potentially “irregular” contracts in its cyclical reviews of awarded contracts. Hence, with the recall metric be-

ing more substantial than precision in this use-case, the trained and validated logistic model, based from these results, is seen to be suitable for deployment as a decision support tool of LGUs to be used in its cyclical reviews of awarded contracts.

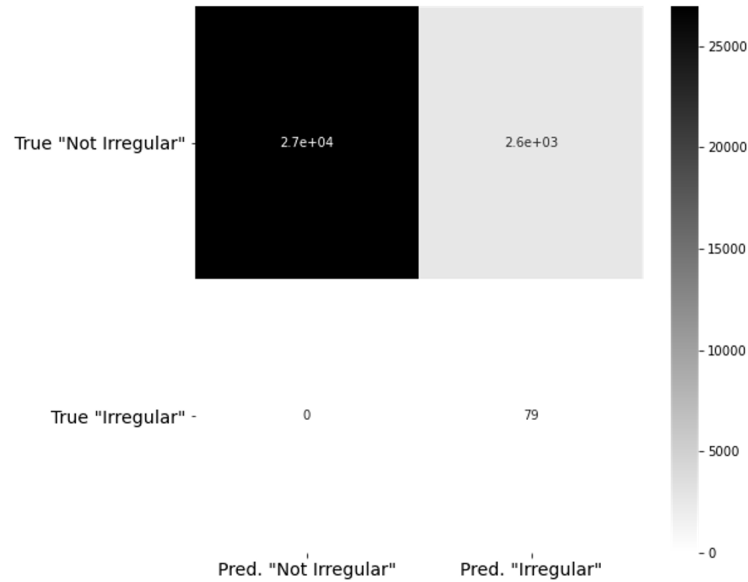


Figure 28: Confusion matrix for the trained logistic regression model. All potentially “irregular” contracts were correctly identified by developed logistic regression model.

Limitations of the Present Work

Limited Data on COVID-19 Related Tenders

It is worth noting that the number of COVID-19 related tenders uploaded on the PhilGEPS website for the fiscal year of 2020 might be underreported. As provided under Section 4 of the Bayanihan to Heal as One Act [49], COVID-19 related procurements are exempt from the provisions provided under the revised implementing rules and regulations of Republic Act 9184 (i.e. posting of COVID-19 related tenders in the PhilGEPS website). Due to the unavailability of this data, this work was not able to include these unposted contracts in the implemented analysis. Moving forward, however, it is suggested that future

COA annual audit reports (e.g. 2021 and 2022 COA annual audit reports) once made available be further considered as these reports could contain detailed information for these unposted contracts.

Limitation of the Developed Prediction Model

Examination of the developed logistic regression model also revealed that SV3, as shown in [Figure 29](#), had the highest influence in correctly classifying the target feature. As previously discussed, SV3 is the singular vector that represents collated features of the dataset relative to how a contract is usually awarded to a winning bidder. While this was found to be consistent with the findings observed in this work, this observation does present a limitation of the developed logistic regression model. This observation then implied that the developed prediction model is only capable of identifying potentially “irregular” LGU contracts that have already been awarded. This developed model would then not be capable of identifying potentially “irregular” contracts that have not yet been awarded by an LGU to a winning bidder since the most relevant basis used in this work was the difference between the observed Notice to Proceed and Award dates. In this context, the developed prediction model could only be used to narrow down review of contracts that have been awarded. While no features of unawarded contracts were found to be key identifiers of the nature of these contracts (i.e. whether “not irregular” or potentially “irregular”), exploring the differences among the Publish, Prebid, and Closing dates, however, revealed that the public procurement process of unawarded contracts is not usually in order (e.g. the posted Prebid date is found to be before the Publish date). This could also perhaps be suggested as a potential avenue for investigation by both the LGU and COA. Nevertheless, this model is still a viable approach to predicting erroneous contracts posted in the PhilGEPS website.

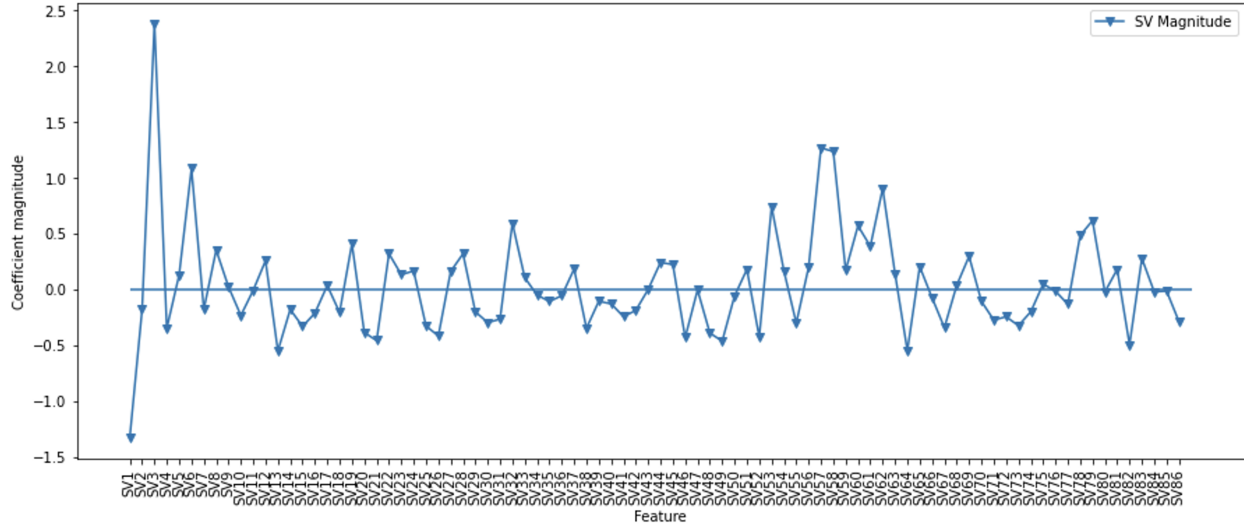


Figure 29: Coefficient magnitudes of the 86 singular vectors that comprised the developed logistic regression model. SV3 was found to be the most influential to the target feature investigated.

SUGGESTED FUTURE WORK

Now that a prediction model capable of tagging contracts as potentially “irregular” have been developed, the next suggested steps to be undertaken is to improve the precision of the model and possibly pilot-test its deployment after it is optimized.

As revealed from the confusion matrix, the precision of the developed model is very low – about 2.9% only. Though in this use-case recall is of more importance, it would still be beneficial for the LGUs to be given a more precise model as this would cut down the number of contracts classified as false positives. In this context, it is suggested that other resampling techniques be explored – Adaptive Synthetic Sampling Method (ADASYN) and Rapidly Converging Gibbs (RACOG) for highly imbalanced data. From the validation implemented in [51], it could be observed that models tested using these oversampling techniques had better metrics than with models pre-processed with SMOTE. Higher

precision and accuracies were achieved using these resampling techniques. It is also possible that the logistic regression model, in general, is not fully capable of achieving a high precision score. With the availability of computational resources, other prediction models could also be explored. Gradient boosting machines, reported to be one of the best set of models achieving high accuracies and precisions, also offer a wide range of machine learning models that is also applicable to this use-case [52].

After model tuning, it would now be ready for deployment in the field. For model deployment, there are two pathways being considered for this next step: (a) serve the developed model as a mobile app, and (b) serve the developed model in an API client integrated in a website. These two pathways are the two basic frameworks to be followed when a prediction model is to be pilot-tested. It is suggested that the prediction be first served offline in a periodic cycle in order to trouble shoot problems to be encountered upon deployment of the developed model. Results from these initial pilot-testing could also serve as a baseline for developing an effective architecture that will support an efficient deployment of the model. Once these problems are sorted out, transition to an online model service should also be explored. This service could also be potentially monetized by PhilGEPS as a premium service that could be integrated in the present electronic procurement system.

Collectively, these are just some of the enumerated possibilities that could be done as an extension to this work. Indeed, as supported by the findings in this work, the developed logistic regression model is still very promising.

POLICY RECOMMENDATIONS

It is substantially clear from the findings of this project that irregularities are indeed present in the LGU contracts posted in the PhilGEPS website. As validated from the audit findings posted in the COA website, discrepancies in the

financial statements, inventory management, violation of circulars, memorandums, and Philippine laws were among the most common audit observations given to LGUs procuring goods and services in this time of a pandemic. While the Bayanihan to Heal as One Act hastened the public procurement process in this time of an emergency [49], it is still imperative that the implemented public procurement process adhere to imposed guidelines in order to safe guard public funds and the interest of the general public. Based from the findings of this work, the following policy recommendations are hence derived:

- LGUs should consider allocating budget for the procurement of COVID-19 vaccines in preparation for mass vaccination against the SARS-CoV-2 virus [3], [10]. In the preparation of the work financial plans (WFP) for fiscal years beyond 2021, procurement of COVID-19 vaccines should also be included as one of the agenda in the budget planning and deliberations of LGUs. For instance, LGUs may consider realigning their unused allocated budget in infrastructure projects if additional funds coming from the National Government of the Philippines would not be sufficient for the vaccination of all of their constituents.
- Compliance to documentary requirements was seen to be the most critical audit finding common to the LGUs considered in this work; hence, to avoid violations in the public procurement process, it is suggested that LGUs should mandate and direct key personnel (e.g. municipal accountants, municipal treasurers) to check the completeness and legality of submitted documents prior to awarding of contract and issuance of Notice to Proceed orders.
- LGUs should spend public funds. The ratio of spent and unspent funds should be increased to at least 2:1 especially in this time of a health crisis [3], [10]. As determined from the COA annual audit reports, compliance to documentary requirements has been identified as one of the bottlenecks which significantly delays LGU procurements [36]. Hence, it is suggested that procuring LGUs ensure compliance to documentary requirements ahead of

the prescribed deadlines to minimize these delays [36].

- It is suggested that an Award-NTP (Days) feature be included in the standard open PhilGEPS dataset in order to monitor potentially “irregular” contracts.
- With the unavailability of comprehensive audit findings for each LGU tender posted in the PhilGEPS website, it is also suggested that the relevant audit findings and audit recommendations by COA be integrated in the open standard dataset of bids and awards notices for each of the contracts posted. Integration of these audit findings in the dataset as additional features for each contract posted could pave the way for an automated contract irregularity detection system that could be embedded within the PhilGEPS website.

ACKNOWLEDGEMENTS

The proponents of this project would like to extend their utmost appreciation to Open Contracting Partnership for the funding provided to this work. The proponents of this project would also like to thank Camila Salazar and the rest of the Open Contracting Partnership team who provided technical support for the completion of this project. The proponents of this work would also like to give their utmost appreciation to Mimay Lu Rodrigo for proof reading this work.

REFERENCES

- [1] Aspen Institute, *Promoting economic recovery after COVID-19*, Jun. 16, 2020. [Online]. Available: <https://www.economicstrategygroup.org/wp-content/uploads/2020/11/Promoting-Economic-Recovery-After-COVID-0615-FINAL.pdf> (visited on 01/03/2021).
- [2] McKinsey & Company, *Saving our livelihoods from COVID-19: Toward an economic recovery*, Apr. 2020. [Online]. Available: <https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20Sector/Our%20Insights/Saving%20our%20livelihoods%20from%20COVID%2019%20Toward%20an%20economic%20recovery/Saving-our-livelihoods-from-COVID-19-Toward-an-economic-recovery.pdf> (visited on 01/03/2021).
- [3] C. A. Tisdell, “Economic, social and political issues raised by the COVID-19 pandemic,” *Economic Analysis and Policy*, vol. 68, pp. 17–28, Dec. 2020. DOI: 10.1016/j.eap.2020.08.002.
- [4] S. Gössling, D. Scott, and C. M. Hall, “Pandemics, tourism and global change: A rapid assessment of COVID-19,” *Journal of Sustainable Tourism*, vol. 29, no. 1, pp. 1–20, Apr. 2020. DOI: 10.1080/09669582.2020.1758708.
- [5] S. M. C. Abo and S. R. Smith, “Is a COVID-19 vaccine likely to make things worse?” *Vaccines*, vol. 8, no. 4, p. 761, Dec. 2020. DOI: 10.3390/vaccines8040761.
- [6] L. O. Gostin, S. A. Karim, and B. M. Meier, “Facilitating access to a COVID-19 vaccine through global health law,” *Journal of Law, Medicine & Ethics*, vol. 48, no. 3, pp. 622–626, 2020. DOI: 10.1177/1073110520958892.
- [7] H. Meyer, “After a COVID-19 vaccine: Collaboration or competition?” *Health Affairs*, vol. 39, no. 11, pp. 1856–1860, Nov. 2020. DOI: 10.1377/hlthaff.2020.01732.

-
- [8] Y. Liu, S. Salwi, and B. C. Drolet, “Multivalued ethical framework for fair global allocation of a COVID-19 vaccine,” *Journal of Medical Ethics*, vol. 46, no. 8, pp. 499–501, Jun. 2020. DOI: 10.1136/medethics-2020-106516.
 - [9] J. H. Kim, F. Marks, and J. D. Clemens, “Looking beyond COVID-19 vaccine phase 3 trials,” *Nature Medicine*, vol. 27, no. 2, pp. 205–211, Jan. 2021. DOI: 10.1038/s41591-021-01230-y.
 - [10] A. M. Viens and O. Eyawo, “COVID-19: The rude awakening for the political elite in low- and middle-income countries,” *BMJ Global Health*, vol. 5, no. 5, e002807, May 2020. DOI: 10.1136/bmjgh-2020-002807.
 - [11] A. E. Yamen, “Tax evasion, corruption and COVID-19 health risk exposure: A cross country analysis,” *Journal of Financial Crime*, vol. ahead-of-print, no. ahead-of-print, Jan. 2021. DOI: 10.1108/jfc-10-2020-0220.
 - [12] C. C. Ezeibe, C. Ilo, E. N. Ezeibe, C. N. Oguonu, N. A. Nwankwo, C. K. Ajaero, and N. Osadebe, “Political distrust and the spread of COVID-19 in Nigeria,” *Global Public Health*, vol. 15, no. 12, pp. 1753–1766, Oct. 2020. DOI: 10.1080/17441692.2020.1828987.
 - [13] G. Lasco, “Medical populism and the COVID-19 pandemic,” *Global Public Health*, vol. 15, no. 10, pp. 1417–1429, Aug. 2020. DOI: 10.1080/17441692.2020.1807581.
 - [14] G. Lasco and N. Curato, “Medical populism,” *Social Science & Medicine*, vol. 221, pp. 1–8, Jan. 2019. DOI: 10.1016/j.socscimed.2018.12.006.
 - [15] R. Mendiola, *Moody’s analytics: Philippines last to recover in Asia*, Jan. 8, 2021. [Online]. Available: <https://www.asianjournal.com/philippines/across-the-islands/moody-analytics-philippines-last-to-recover-in-asia/> (visited on 01/03/2021).
 - [16] J. C. Gotinga, “PH missed chance to get Pfizer vaccines in January because of Duque – Lacson,” *Rappler*, Dec. 17, 2020. [Online]. Available: <https://www.rappler.com/philippines/health/coronavirus/2020/12/17/philippines-missed-chance-to-get-pfizer-vaccines-in-january-because-of-duque-lacson/>

- [//www.rappler.com/nation/lacson-says-philippines-missed-chance-get-pfizer-vaccines-january-2021-because-duque](https://www.rappler.com/nation/lacson-says-philippines-missed-chance-get-pfizer-vaccines-january-2021-because-duque) (visited on 01/03/2021).
- [17] CNN Philippines Staff, “Locsin confirms Duque ‘dropped the ball’ in Pfizer vaccine deal, but urges end to blame game,” *CNN Philippines*, Dec. 21, 2020. [Online]. Available: https://cnnphilippines.com/news/2020/12/21/Locsin-Duque-dropped-the-ball-Pfizer-vaccine.html?fbclid=IwAR2dkxmcCx7jW6Mzm_57yfa16bMUdLO64kGRc5T5RcLHYSDI-negKMEw2Jo (visited on 01/03/2021).
- [18] —, “PH eyes to seal deal for Sinovac’s 25 million vaccine doses this week,” *CNN Philippines*, Dec. 14, 2020. [Online]. Available: https://cnnphilippines.com/news/2020/12/14/Sinovac-COVID-vaccine-Philippines-supply.html?fbclid=IwAR1iP17cfGDpNdFzrx4VYR-nKRT8BL2gx0_Jq9jfEQxZDz3VHTizVHP1BII (visited on 01/03/2021).
- [19] S. Tomacruz, “Philippines targets arrival of China vaccine by March 2021,” *Rappler*, Dec. 14, 2020. [Online]. Available: <https://www.rappler.com/nation/philippines-targets-arrival-sinovac-china-covid-19-vaccine-march-2021> (visited on 01/03/2021).
- [20] M. Cruz, “Solons question government preference for Sinovac vaccine,” *Manila Standard*, Dec. 26, 2020. [Online]. Available: <https://manilastandard.net/mobile/article/342967> (visited on 01/03/2021).
- [21] R. Mendoza, “DoH: Pfizer deal was for data sharing, not supply,” *The Manila Times*, Dec. 19, 2020. [Online]. Available: <https://www.manilatimes.net/2020/12/19/news/national/doh-pfizer-deal-was-for-data-sharing-not-supply/813460/> (visited on 01/03/2021).
- [22] H. Torregoza, “Solon wonders why IATF prefers Sinovac vaccine of China,” *Manila Bulletin*, Dec. 25, 2020. [Online]. Available: <https://mb.com.ph/2020/12/25/solon-wonders-why-iatf-prefers-sinovac-vaccine-of-china/> (visited on 01/03/2021).

-
- [23] V. Barcelo, W. Casas, and M. Ramos-Araneta, *Palace vouches for Duque, sees 'no major laps'*, Dec. 18, 2020. [Online]. Available: <https://manilastandard.net/mobile/article/342344> (visited on 01/03/2021).
 - [24] C. M. Ramos, *COVID-19 deals may be jeopardized if prices are publicized, vaccine czar warns*, Jan. 18, 2020. [Online]. Available: <https://newsinfo.inquirer.net/1384910/galvez-on-revealing-sinovac-price-during-senate-hearing> (visited on 01/20/2021).
 - [25] D. J. Esguerra, *'an industry practice,' duterte says in defense of undisclosed vaccine price*, Jan. 19, 2020. [Online]. Available: <https://newsinfo.inquirer.net/1385315/an-industry-practice-duterte-defends-non-disclosure-of-sinovac-vaccine-price> (visited on 01/20/2021).
 - [26] A. Calonzo, "Nearly half in Philippines don't want COVID vaccine, survey says," *Bloomberg*, Dec. 7, 2021. (visited on 01/03/2021).
 - [27] CNN Philippines Staff, "Nearly half of Filipinos opt out of COVID-19 vaccination, poll finds," *CNN Philippines*, Jan. 7, 2020. [Online]. Available: <https://www.cnn.ph/news/2021/1/7/COVID-19-vaccine-survey-Philippines.html> (visited on 01/03/2021).
 - [28] J. R. Barajas, "Contextualizing procurement capacity of Philippine local government units (LGU) in response to the COVID-19 pandemic: A multi-criterion decision analysis perspective," 2020. [Online]. Available: <https://drive.google.com/file/d/1OitWnFHqb3l5LHSMtONqoBiCsdVgxY2Z/view>.
 - [29] B. M. Laforga, *Philippine gdp shrinks by record 9.5% in 2020*, Jan. 29, 2021. [Online]. Available: <https://www.bworldonline.com/philippine-gdp-shrinks-by-record-9-5-in-2020/> (visited on 01/30/2021).
 - [30] C. Salazar, J. Pane, and R. Fernandez, *A duration analysis of public procurement processes*, 2020. [Online]. Available: <https://docs.google.com/document/d/1kw7u0Cif674zyztpEuLQb8n0LFJhd7qmzCFSQEH4DM/edit#> (visited on 01/03/2021).

-
- [31] J. Ferwerda, I. Deleanu, and B. Unger, “Corruption in public procurement: Finding the right indicators,” *European Journal on Criminal Policy and Research*, vol. 23, no. 2, pp. 245–267, May 2016. DOI: 10.1007/s10610-016-9312-3.
 - [32] E. Dávid-Barrett and M. Fazekas, “Grand corruption and government change: An analysis of partisan favoritism in public procurement,” *European Journal on Criminal Policy and Research*, vol. 26, no. 4, pp. 411–430, Jun. 2019. DOI: 10.1007/s10610-019-09416-4.
 - [33] T. Tátrai and A. Németh, “Improving red flag instruments for public procurement,” *ERA Forum*, vol. 19, no. 2, pp. 267–285, May 2018. DOI: 10.1007/s12027-018-0513-8.
 - [34] M. Fazekas, I. J. Tóth, and L. P. King, “An objective corruption risk index using public procurement data,” *European Journal on Criminal Policy and Research*, vol. 22, no. 3, pp. 369–397, Apr. 2016. DOI: 10.1007/s10610-016-9308-z.
 - [35] Philippine Government Electronic Procurement System, *Modernized government e-procurement system*. [Online]. Available: https://www.philgeps.gov.ph/CmsHomePages/open_data_grid (visited on 01/03/2021).
 - [36] Commission on Audit, *Annual audit reports*. [Online]. Available: <https://www.coa.gov.ph/index.php/reports/annual-audit-report> (visited on 07/31/2021).
 - [37] Government Procurement Policy Board, *The 2016 revised implementing rules and regulations of republic act no. 9184*, 2016. [Online]. Available: <https://www.gppb.gov.ph/laws/laws/RevisedIRR.RA9184.pdf>.
 - [38] P. C. Hansen, “The truncatedSVD as a method for regularization,” *BIT*, vol. 27, no. 4, pp. 534–553, Dec. 1987. DOI: 10.1007/bf01937276.
 - [39] A. Rea and W. Rea, *How many components should be retained from a multivariate time series PCA?* 2016. arXiv: 1610.03588.

-
- [40] J. H. Ward, “Hierarchical grouping to optimize an objective function,” *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, Mar. 1963. DOI: 10.1080/01621459.1963.10500845.
 - [41] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, “Type 2 diabetes mellitus prediction model based on data mining,” *Informatics in Medicine Unlocked*, vol. 10, pp. 100–107, 2018. DOI: 10.1016/j.imu.2017.12.006.
 - [42] C. Zhu, C. U. Idemudia, and W. Feng, “Improved logistic regression model for diabetes prediction by integrating PCA and k-means techniques,” *Informatics in Medicine Unlocked*, vol. 17, p. 100 179, 2019. DOI: 10.1016/j.imu.2019.100179.
 - [43] N. Nnamoko and I. Korkontzelos, “Efficient treatment of outliers and class imbalance for diabetes prediction,” *Artificial Intelligence in Medicine*, vol. 104, p. 101 815, Apr. 2020. DOI: 10.1016/j.artmed.2020.101815.
 - [44] T. Sasada, Z. Liu, T. Baba, K. Hatano, and Y. Kimura, “A resampling method for imbalanced datasets considering noise and overlap,” *Procedia Computer Science*, vol. 176, pp. 420–429, 2020. DOI: 10.1016/j.procs.2020.08.043.
 - [45] T. Hastie, “Ridge regularization: An essential concept in data science,” *Technometrics*, vol. 62, no. 4, pp. 426–433, Aug. 2020. DOI: 10.1080/00401706.2020.1791959.
 - [46] O. Balaeva, Y. Rodionova, A. Yakovlev, and A. Tkachenko, “Public procurement efficiency as perceived by market participants: The case of russia,” *International Journal of Public Administration*, pp. 1–12, Aug. 2021. DOI: 10.1080/01900692.2021.1964083.
 - [47] B. Hoekman, A. Shingal, V. Eknath, and V. Ereshchenko, “COVID-19, public procurement regimes and trade policy,” *The World Economy*, Mar. 2021. DOI: 10.1111/twec.13118.

- [48] L. Folliot Lalliot and C. R. Yukins, “COVID-19: Lessons learned in public procurement. time for a new normal?” *Time for a New Normal*, pp. 46–58, Sep. 3, 2020. [Online]. Available: <https://ssrn.com/abstract=3685860>.
- [49] Official Gazette of the Philippines, *Bayanihan to heal as one act*, 2020. [Online]. Available: <https://www.officialgazette.gov.ph/downloads/2020/09sep/20200911-RA-11494-RRD.pdf> (visited on 09/20/2021).
- [50] Government Procurement Policy Board, *Guidelines for emergency procurement under republic act no. 11469 or the bayanihan to heal as one act*, Apr. 6, 2020. [Online]. Available: <https://www.gppb.gov.ph/issuances/Circulars/GPPB%20Circular%20No.%2001-2020.pdf> (visited on 09/21/2021).
- [51] J. Kong, T. Rios, W. Kowalczyk, S. Menzel, and T. Bäck, “On the performance of oversampling techniques for class imbalance problems,” in Springer International Publishing, 2020, pp. 84–96. DOI: 10.1007/978-3-030-47436-2_7.
- [52] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” vol. 7, 2013. DOI: 10.3389/fnbot.2013.00021.