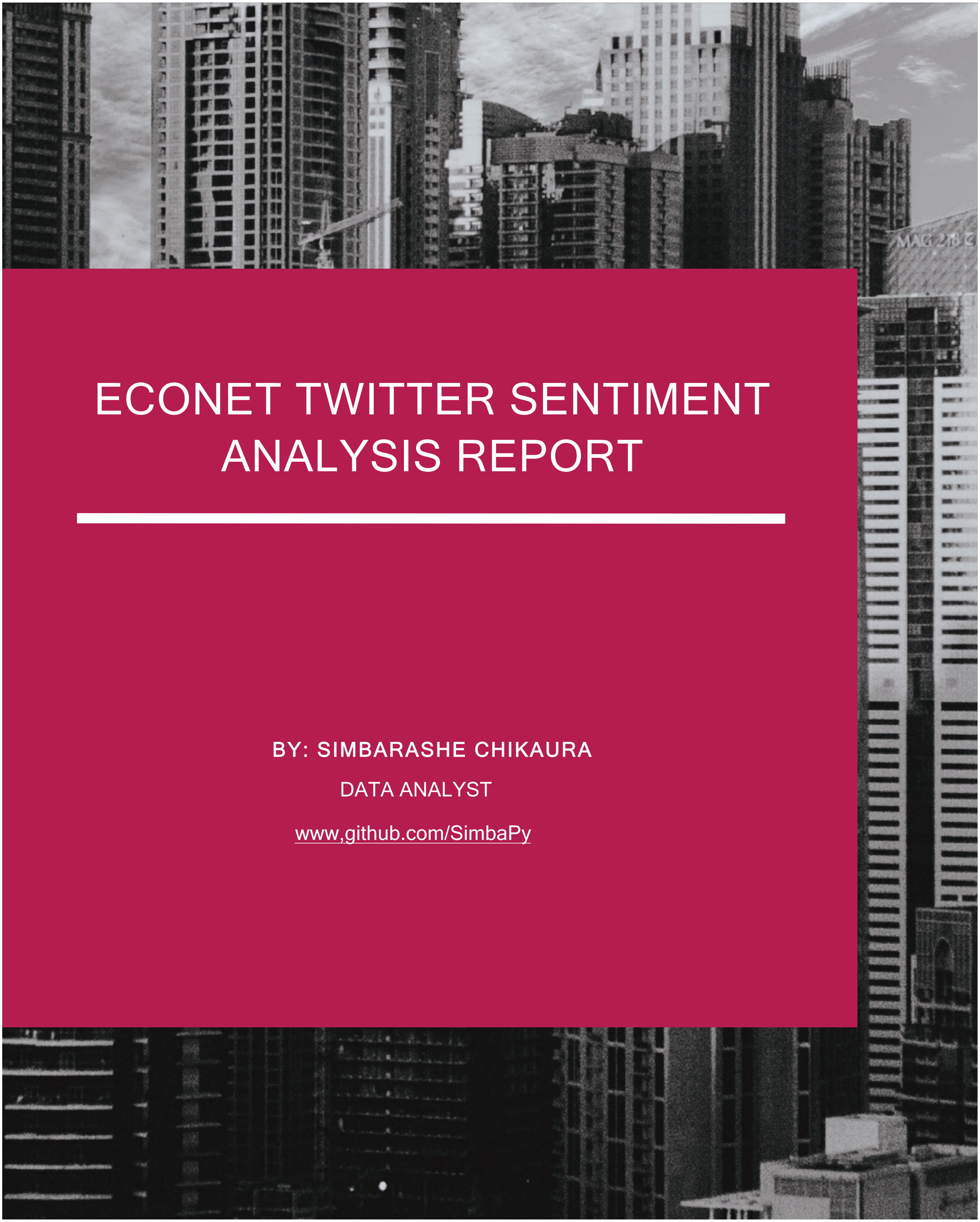

ECONET TWITTER SENTIMENT ANALYSIS REPORT

BY: SIMBARASHE CHIKAURA

DATA ANALYST

www,github.com/SimbaPy

TABLE OF CONTENTS

ABSTRACT

Econet Wireless Zimbabwe is one of the biggest and most successful companies in Zimbabwe. The Mobile Network Operator (MNO) is the dominant player in the mobile telecommunications sector, commanding a market share of 66.3% (total mobile subscribers) as of the fourth quarter of 2018. The entity has diversified since its inception in 1998, extending its arms to sectors such as banking (mobile & traditional), insurance, digital education and even solar energy. Arguably, one of the reasons why Zimbabweans have embraced Econet is because of the goodwill it gained from the public amid its contentious founding and the battle it had with the government for the right to operate. However, that sentiment has gradually eroded over the years, increasingly becoming more and more negative over the last few years.

The purpose of this project was to investigate the sentiment of Econet subscribers towards the brand(s) of the MNO and how that has changed over the last 7 years in particular. Econet's customer service strategy includes social media, where a dedicated account is responsible for handling customer complaints. The data source for this project was subscriber interactions with this account, as well as any other tweets that mentioned Econet or any of its brands. The exercise was essentially a natural language processing (NLP) task, analyzing the sentiment of tweets progressively from 2012 (the year the first official Econet account joined twitter) to early 2019.

The project's objectives were outlined as follows:

- A time series analysis of how the frequencies of complaints have changed over the last 9 years. How does Econet's customer complaint count compare to its competitors?
- How do complaint categories change over time?
- Which month does Econet experience the most complaints, on average?
- Which type of complaint is most prevalent?
- Is the amount of complaints Econet has gotten over the last 9 years correlated to its increase in market share and product range in the same time period?

After iteratively answering the questions raised by the objectives, data, Ecocash and customer service are the recurring issues customers raise. Also, sentiment against Econet has become increasingly negative over time, exponentially especially over the last 2 years. Surprisingly, there seems to be no clear relationship between market share, and/or product range, and the frequencies of complaints.

Overall, the data suggests that users believe Econet is out of touch with the needs of the average customer. The ultimate conclusion was that the sentiment against Econet is currently at its poorest since the organization's inception. Many now view it as extortionate and inefficient, a far cry from its mission.

METHODOLOGY

DATA PREPARATION

Figure 1: Data Prep Workflow

To ensure a valid source of data that can produce results free of negligible error and noise, I first prepared the data using the steps depicted in the figure to the left.

The data used in this project was mined, processed and analyzed using the Python programming language. A detailed step-by-step guide of how this project was done, complete with the original dataset, can be found on my github page (referenced at the end of this document).

1. DATA COLLECTION

The data was mined using web scraping from twitter. Each tweet was saved as an entry in a table. Tweets were mined from Feb 2012 to March 2019. The reason for this was because the first Econet account (@econetzimbabwe) joined twitter in Feb 2012, so the presumption is this is the period Econet subscribers would've begun interacting with the entity on twitter. The end date was simply the date the research was carried out.

The sample of tweets in the 7 year time period was taken from the top tweets section. According to Twitter's own FAQs, "*Top Tweets are the most relevant Tweets for your search. We determine relevance based on the popularity of a Tweet (e.g., when a lot of people are interacting with or sharing via Retweets and replies), the keywords it contains, and many other factors.*" This filtered the results to the most relevant from the search query "*econet OR netone OR telecel OR stewardbank since:2012-02-01*"

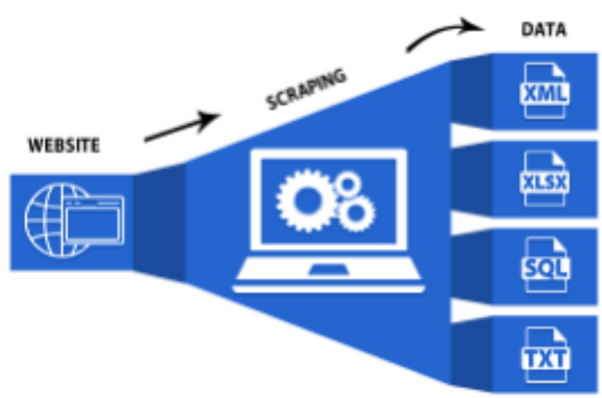

Figure 2: Web Scraping

2. STRUCTURED DATA

The python application used in step 1 (web scraper), stored each search result (tweet) as a row in a table with the following columns:

- created_at - date the tweet was posted,
- likes - number of times the tweet got liked,
- retweets - number times the tweet was retweeted,
- text - the text content of the tweet,
- user_screen_name - the handle (ID) of the user responsible for the tweet.

Figure 3: Example Tweet

Table 1: Example row in dataset

created_at	likes	retweets	text	user_screen_name
2019-03-26 02:08:01	74	9	Econet needs to stop sending me 4 SEPARATE promo messages each time I buy airtime/data!	TinoNyandoro

After the app finished scraping twitter for the relevant data, the table had 4553 rows and 6 columns, a table with a shape of [4553, 6].

3. DATA CLEANING

Although I tried to be as specific as possible with my search query, as well as chose the ‘top’ results of that search query, it is extremely probable that the data still contained noise. Hence, the data needed to be cleaned.

When cleaning data, it is best to specify certain criteria data has to adhere to for it to remain in the dataset. If it doesn’t follow this criteria, it is considered noise and discarded. Another way to achieve the same task is to specify criteria the data shouldn’t follow and discard it on that basis.

Table 2: Decision Criteria

Keep tweets:	Discard tweets:
addressed to @econet_support. These are automatically enquiries or complaints since that account solely exists for handling customer complaints,	from official econet accounts, e.g., @econet_support, @econetzimbabwe, @EcocashZW, etc. These tweets are not enquiries from clients. The analysis is based on customer sentiment.
that contain the pattern 'econet, ecocash, steward bank or ecosure' that have a negative sentiment.***	about econet from media accounts, e.g., @techzim, @theherald, @FingazLive, etc.
that contain both econet and a competitor (netone and/or telecel). This tweet is most likely an econet customer threatening to switch to another service provider.	that involve econet's philanthropic/corporate responsibility activities, e.g., cyclone idai, cholera, ebola, etc.
	that contain more than 50% of non-english words. These confuse the sentiment algorithm and will just end up being noise.***
	that mention econet but have a positive sentiment - if the content of a tweet is positive, it probably isn't an enquiry or complaint. ***

****this last criteria can only be investigated after the tweets have been classified according to sentiment later on in the project.*

After cleaning, the dataset decreased from [4553, 6] to [3508, 6].

4. EXPLORATORY DATA ANALYSIS

Data can never be too clean. For it to answer the questions of this project, it has to be scrutinized as much as possible. After cleaning according to the criteria, I explored the content of the text of the dataset for anything that spiked curiosity. This was done in the form a wordcloud, shown to the right.

The wordcloud brought up more criteria under which irrelevant data might be discarded. For instance, tweets that contain 'strive masiyiwa' or 'fayaz king' are not likely

to be enquiries, but corporate related.

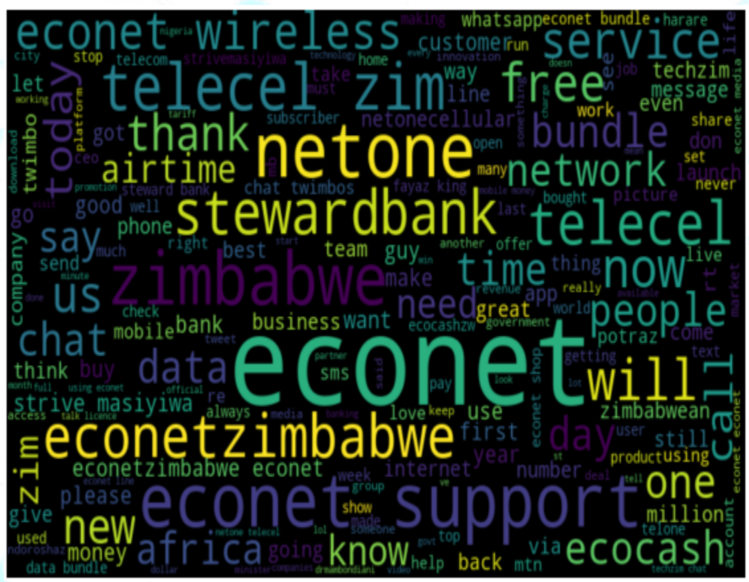

Figure 4: WordCloud 1

After iteratively generating the wordcloud and removing rows which contained data that was irrelevant, the shape of the dataset reduced to [3126, 6]. The wordcloud below shows a leaner dataset containing more relevant data.

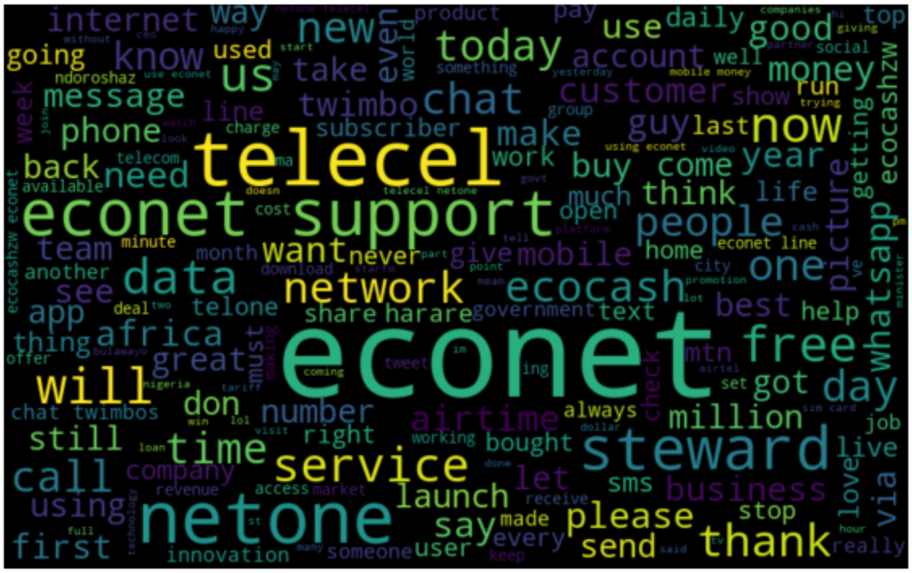

Figure 5: WordCloud 2

At this point, the data was clean enough to apply a Sentiment Analyzer onto it, and classify tweets as positive (pos), negative (neg) or neutral. The algorithm used to classify the tweets was built on an API from http://text-processing.com/demo/. Test data was tested on the API and it produced the correct result 92% of the times, making it sufficient for the analysis.

Of the 3,126 tweets:

neutral: 1735
neg: 742
pos: 649

The objective at this point was to have a dataset ready for analysis that had only enquiries and/or complaints. If an entry (tweet) in the dataset was negative, it automatically qualified as a valid row in the dataset. Therefore, all 742 tweets with the ‘neg’ label remained. On the other hand if a tweet was neutral or positive, it apparently doesn’t seem like a complaint. However, an exploration of these tweets showed that some of them were enquiries that simply did not contain aggressive language, therefore the sentiment analyzer would’ve mislabeled them. To go around this, if a tweet contained the handles: ‘@econet_support’, ‘@netonesupport’, ‘@EcocashZW’ or ‘@stewardbank’, it’d be retained. This is because those would most likely be polite enquiries to the accounts responsible for such enquiries.

After following these criteria, I remained with a dataset with 2,610 rows/tweets. I then moved on to the analysis part of my project with the security that my data was sound.

ANALYSIS

The analysis part of this project was the basis of answering the objectives listed below:

- A time series analysis of how the frequencies of complaints have changed over the last 9 years. How does Econet's customer complaint count compare to its competitors?
- How do complaint categories change over time?
- Which month does Econet experience the most complaints, on average?
- Which type of complaint is most prevalent?
- Is the amount of complaints Econet has gotten over the last 9 years correlated to its increase in market share and product range in the same time period?

1. 9 YEAR TIME SERIES

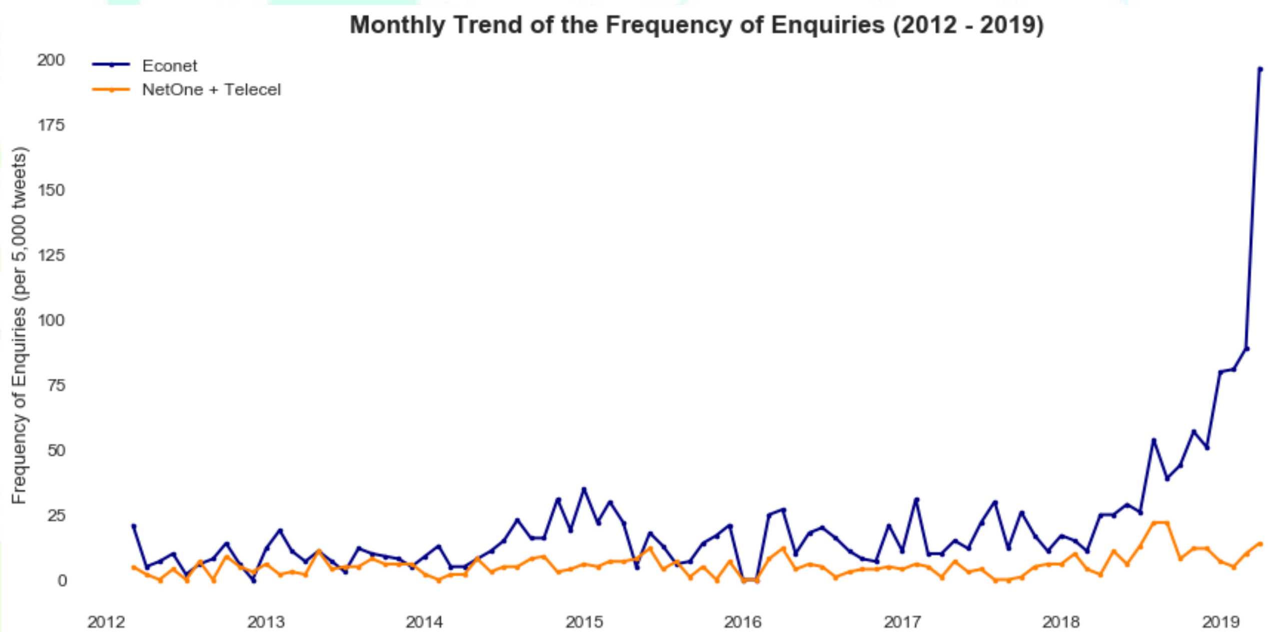

Figure 6: Monthly Time Series by Brand

- The trend shows that, per every 5,000 mentions of Econet on twitter, tweets related to enquiries oscillated between 0 & 25 in the period from 2012 to 2017. However, from 2018 onwards, complaints rise exponentially, peaking at close to 200 in the month of March, 2019.
- Econet's MNO competitors, NetOne and Telecel, generate less than half the enquiries of Econet combined. Their frequency relatively remains the same over the years.

2. CATEGORICAL TIME SERIES

To gain insights into the reasons behind the different frequencies year-on-year, it was best to aggregate the enquiries by category, and visualize them.

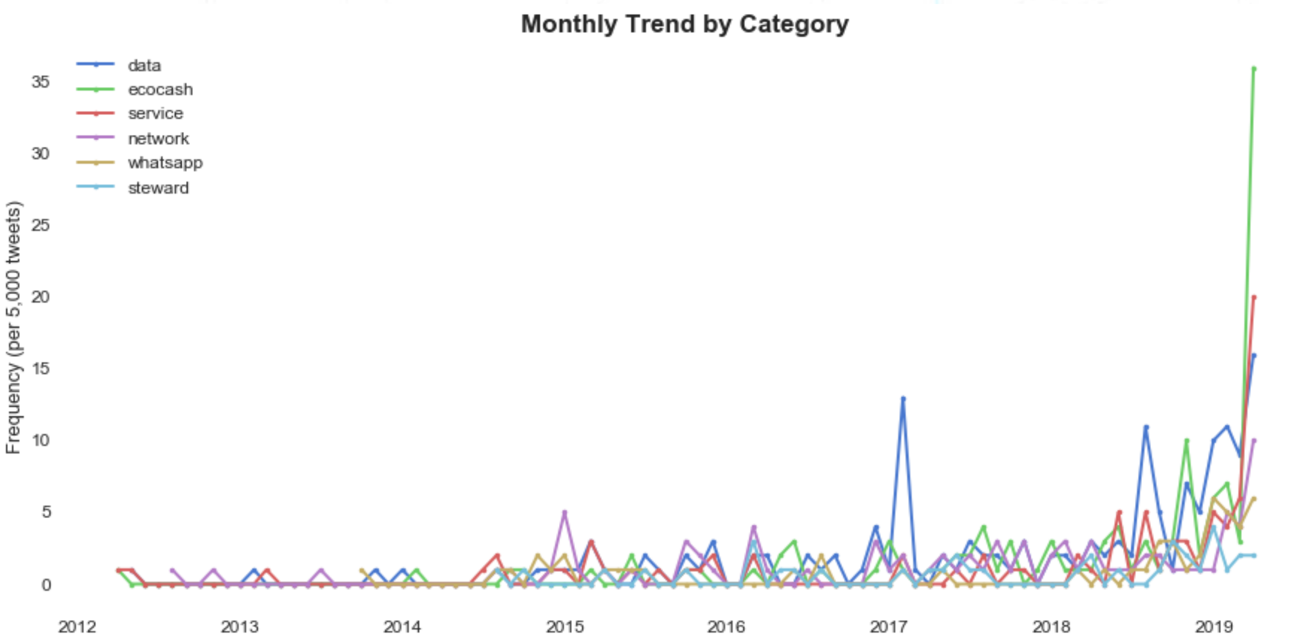

Figure 7: Monthly Trend Series by Category

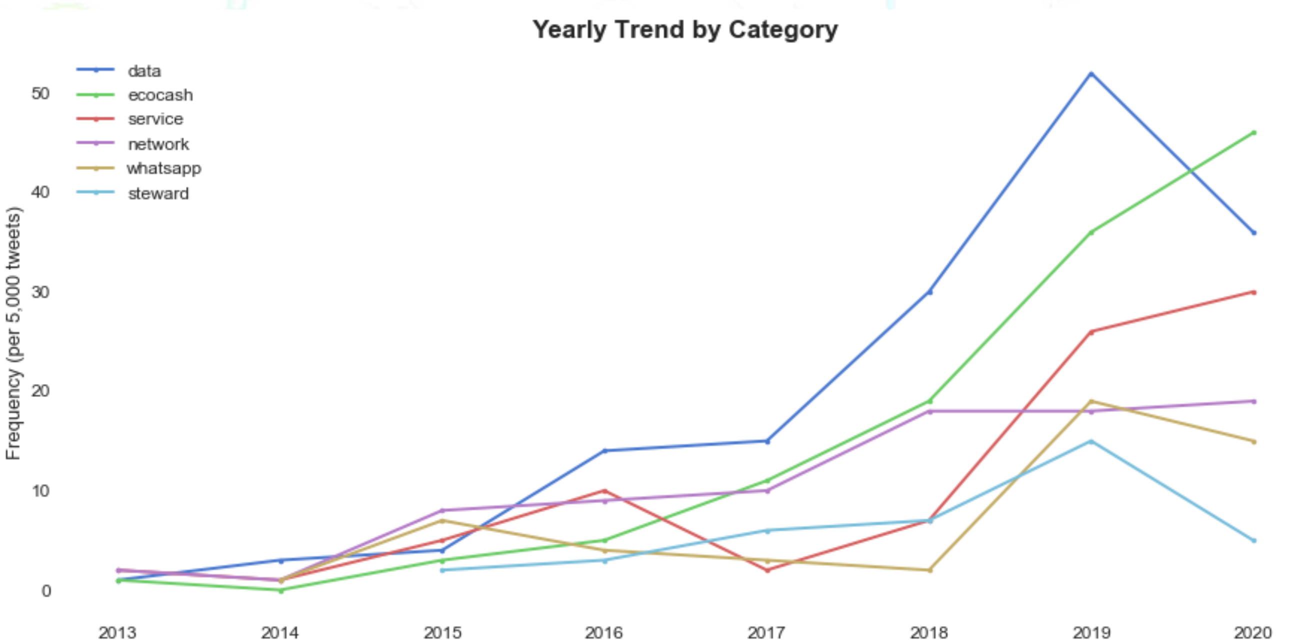

Figure 8: Yearly Trend Series by Category

- The figures above provide some valuable insights behind the enquiries Econet gets over time.
- Using a combination of publicly available data (e.g. Potraz annual reports), the tweets themselves and Econet's own annual reports, I was able to produce some meaningful presumptions of what could be the drivers of these complaints.

Table 3: Complaint Drivers

Year	Enquiry Occurrence	Possible Driver/Reason
2014	• From the yearly time series, we can see that the most prevalent complaint at the close of 2014 was network. This was mostly attributed to the last month of 2014, December. This is where network problems were at their peak in that year.	• According to the 2014 Econet Integrated Annual Report, Econet introduced new products into the market: Whatsapp bundles & 4G LTE. • Since this was the maiden entry of these products/services, it can be expected that there be issues with both and hence the corresponding enquiries on twitter.
2015	• Network doesn't change significantly the following year but there are major upshots in the service and data categories. Service at least doubles in quantity while data quadruples at the close of 2015. Unlike network in 2014, service and data complaints are more spread out throughout the year.	▪ To start of with service, Econet introduced 4 new products in 2015. The common criticism to Econet has always been that they have way too many products and this compromises their service delivery. As far as the data enquiries go, Potraz notes in its reports that there was a 12% increase in mobile data subscriptions from 2014 to 2015. And with Econet making up almost 80% of mobile data usage, it can be argued that most of those subscriptions belonged to them. Even more interesting is the granular increase of mobile data subscriptions in the form of LTE. LTE increased a mammoth 39,322% from 2014 to 2015. What this means is that there was a large new customer base using Econet's new data service. This could explain the quadrupling of the data enquiries.
2016	▪ There's a drop in service complaints and whatsapp as well at the end of 2016. The latter has been on a steady decrease since 2014. There are no changes in data and network, but a sharp increase in ecocash is observed. Ecocash complaints peak in May and December of 2016.	▪ There was a sharp dive in service complaints from 2015 to 2016. This is in line with Econet's 2016 report where they document an increase of 16% (from 32% to 48%) when it comes to their Net Promoter Score, a metric presumably used to measure quality of service. However, I cannot find information that explains the rise in ecocash enquiries.
2017	• Whatsapp problems continue on their steady descent at the end of 2017, but there are sharp increases in all other categories (except steward bank, it remains the same). Data leads, with ecocash, network and service following in that order. Data's lead could be attributed to its very large spike in January 2017. Ecocash and network fluctuate throughout the year.	• #datamustfall was one of the leading hashtags of tweets directed to Econet in 2017. This explains the spike in data complaints, • Most tweets which lament service delivery are targeted 2 Econet products: Steward Bank and KweseTV.
2018	• 2018 seems to be the year where Econet got most of its complaints. Apart from network, which mostly remained the same, every other category shot sharply. Data, ecocash, service, whatsapp and steward bank all had sharp rises.	• The 'data' peak in July is actually attributed to user information rather than internet units. Econet subscribers were accusing the MNO of giving political parties their information for campaign purposes without consent. • Service complaints encapsulate network, data and steward bank.
2019	• The current year (2019) could be the most interesting of the bunch. It has the highest peaks of all the years, and this data only ends in March. All categories seem to have been multiplied by a factor of 4. Interestingly, ecocash now leads, with data in second.	• ecosure deductions that were made on customers' accounts without their conscent, • challenges with buying airtime and data using the service, • challenges buying zesa with ecocash

3. AVERAGE NUMBER OF COMPLAINTS

It might be interesting to know which month (on average) Econet receives the most complaints.

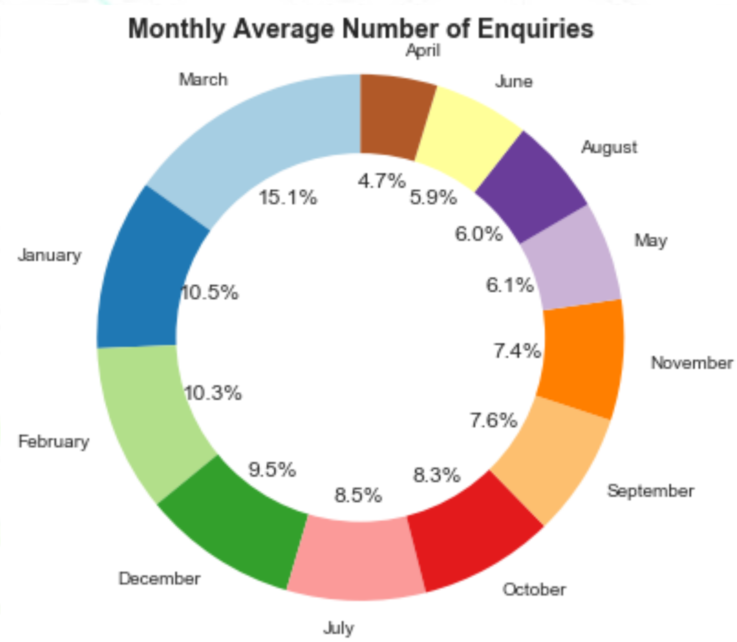

Figure 9: Monthly Enquiries Pie Chart

March seems to be the month with the most enquiries/complaints, historically.

However, it is important to take into account the abnormally high numbers of 2019. Collating the pie chart to the left with the line chart in the previous section, we can see that Jan, Feb and Mar of 2019 have some of the highest enquiries in the entire time series. It is of no surprise then that Mar, Jan and Feb are the 3 highest months on the pie chart to the left.

To get visualization without the last 3 months' bias, a pie chart ending in 2018 was also plotted.

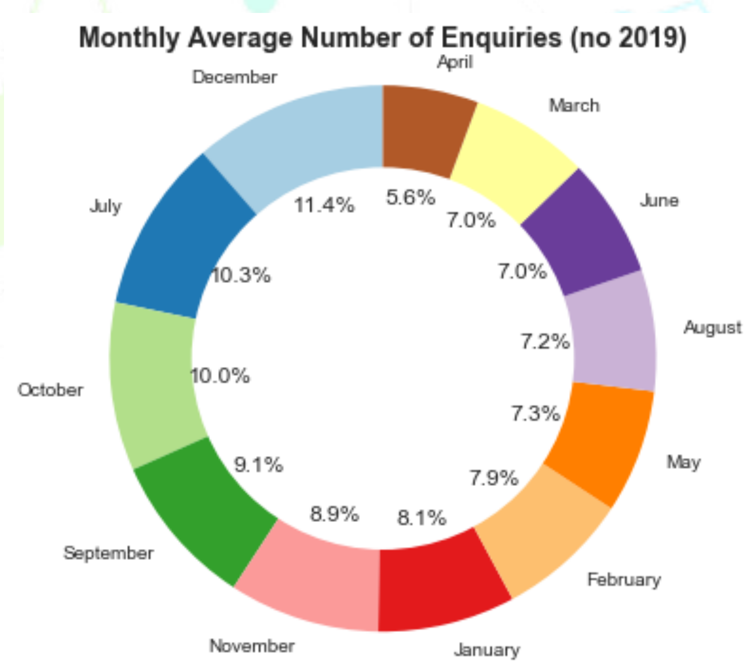

Figure 10: Monthly Enquiries Pie Chart (no 2019)

After excluding the 2019 data, December becomes the month with the largest portion of enquiries.

4. ENQUIRY PREVALENCE BY CATEGORY

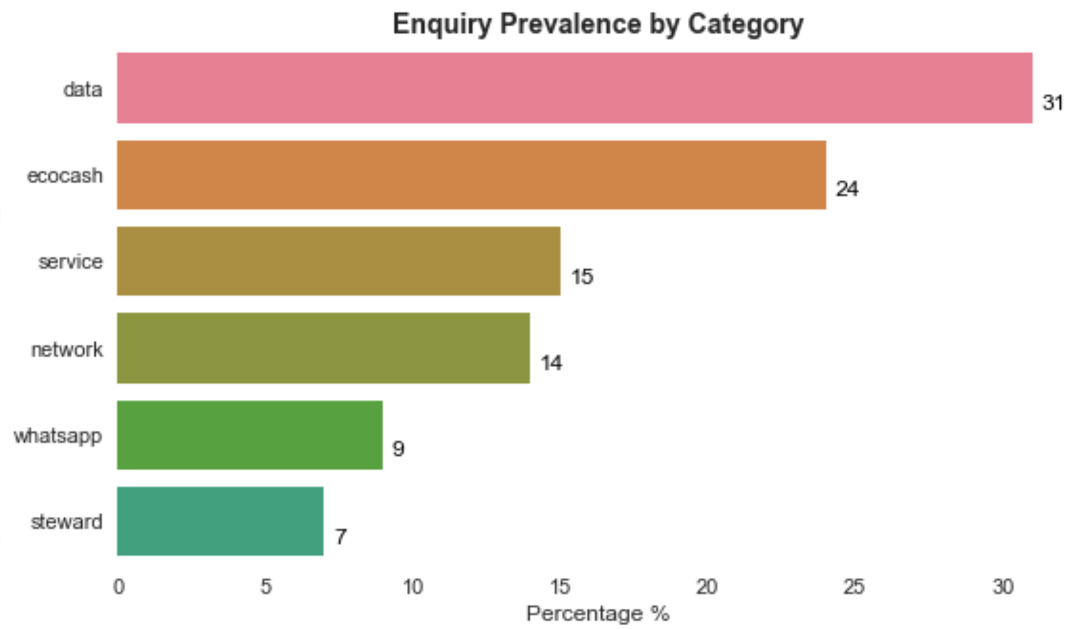

Figure 11: Enquiry Prevalence by Category Bar Chart

According to the dataset, customers complain about data more than any other problem. Ecocash follows closely behind with another significant share.

5. ENQUIRIES VS. MARKET SHARE AND NUMBER OF PRODUCTS/SERVICES

Table 4

year	enquiry_count	market share	cumulative number of products/services
2012	91	63.5	4
2013	111	64.0	7
2014	197	54.7	9
2015	175	52.5	13
2016	174	49.4	17
2017	213	53.1	18
2018	456	66.3	19

The table above shows the aggregated number (sum) of enquiries by year in the dataset. The columns 'market share' and 'cumulative number of products' were added to the table for the corresponding years. These were sourced from Potraz and Econet annual reports, respectively.

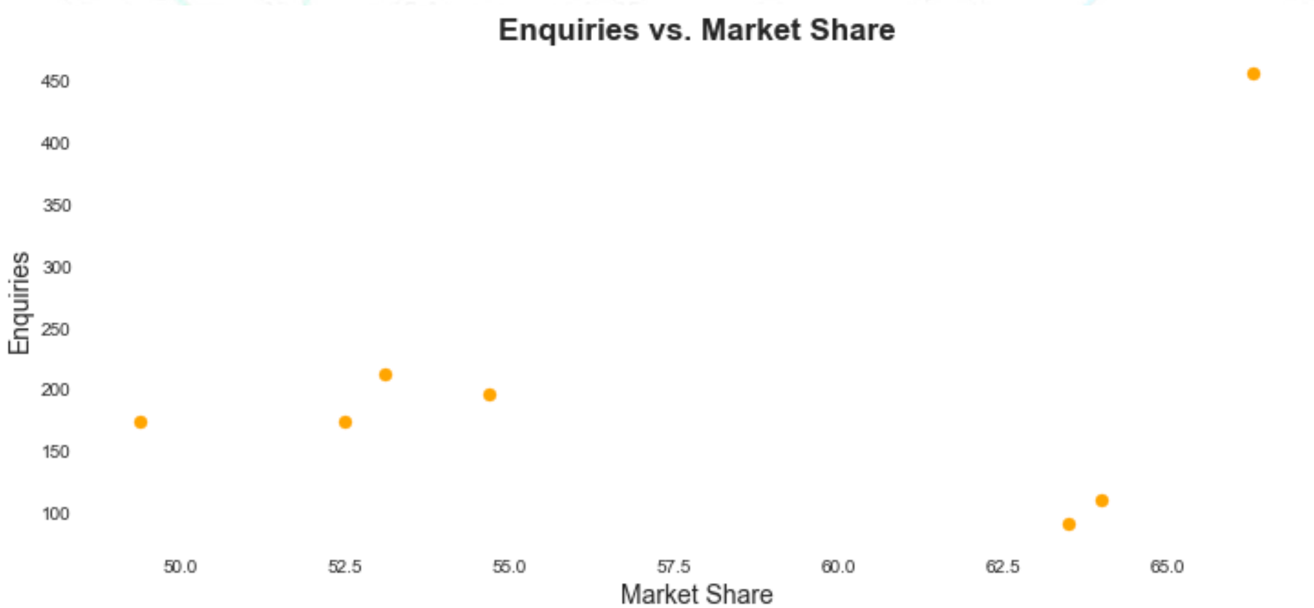

Figure 12: Enquiries vs. Market Share Scatter Plot

- The scatter plot above of enquiries vs. market share does not demonstrate a linear relationship at all.
- Enquiries do not seem to be dependent in any way on market share.

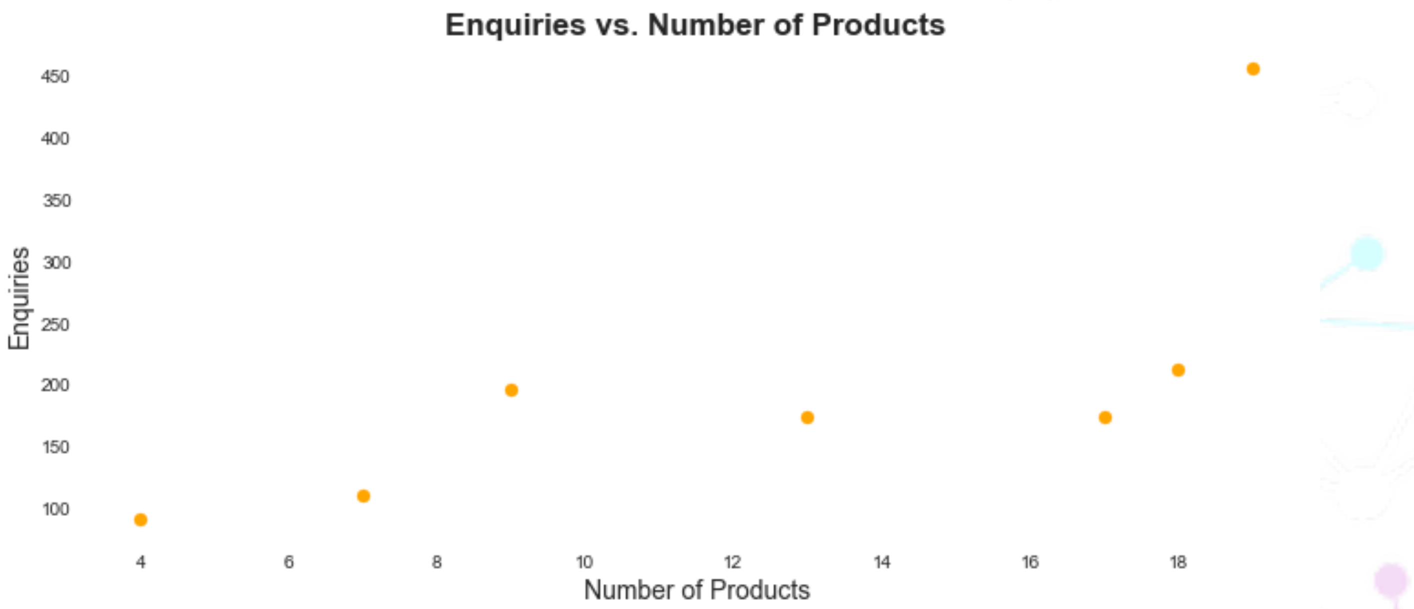

Figure 13: Enquiries vs. Number of Products Scatter Plot

- The scatter plot for enquiries vs. number of products also does not show a definitive relationship between the 2 variables.

An interesting point to note down is that both scatter plots have an outlier point, which for both is attributed to the spike in 2018. Without this outlier, both lines of best fits on the 2 scatter plots would be almost straight, indicating no correlation.

LIMITATIONS

No study is immune to barriers and limitations, and this one was no exception. The following is a list of inconveniences that could potentially compromise the validity of the study

- Twitter search engine – the search engine on twitter does not always return the same data every time a search query is run. This is because the 'top' setting is designed to return tweets with the most engagement. This can change over time. Therefore, it is possible that the results of this project could be different if it was to be reproduced in its entirety.
- Cleaning text data – text data is naturally unstructured, & it requires a lot of cleaning and exploration for it to take a desired format. It contains a lot of noise, which might compromise results. Although every possible effort was taken to remove this noise, it is possible some may still remain because of the nature of the dataset.
- Lack of confidential Econet data – to come up with solid indisputable conclusions for this project, the results, which were produced by the analysis, need to be cross referenced with data that only Econet itself can provide. This may lead to some of the conclusions being presumptuous at best.
- Lack of prior research studies on the topic – no information is publicly available to cross reference this study on, based on similarity.

CONCLUSIONS

The main focus of this project was to investigate the recent negative sentiment against Econet, as well as what drove it. The social media application Twitter is where the MNO interacts with a portion of its customers, and that's where the data that supports the aforementioned statement was mined. Text data was scrapped from Feb 2012 to Mar 2019 for tweets that were associated with Econet and its 2 MNO competitors (NetOne and Telecel). The following questions were posed to the data:

Que 1: **A time series analysis of how the frequencies of complaints have changed over the last 9 years. How does Econet's customer complaint count compare to its competitors?**

Ans 1: The time series shows that enquiries exponentially rose from 2018 going onwards. As a matter of fact, 2018 and 2019 combined make 45% of the enquiries in the dataset. If twitter data is anything to go by, then Econet needs to introspect as to what caused these dramatic changes in the last 2 years.

Que 2: **How do complaint categories change over time?**

Ans 2: Data, Ecocash and service are the major complaints that have a marked increase, particularly post 2015. Data's main peak was in Jan 2017, which is when Econet drastically increased its data without any justification. The customers responded with the *#datamustfall* hashtag, which in part contributed to Econet reverting back to its old data prices. Ecocash peaked more than any other category in March 2019. In the entire dataset, this is the month with the most complaints. The complaints directed at Ecocash were mostly of how EcoSure had deducted premiums from Ecocash accounts without any authorization from the holders of those accounts. The Zesa token purchase system was also compromised during this period, and more complaints were directed towards Econet because of this. The results of the analysis show that 2018 and early 2019 is the period Econet subscribers finally lost patience with the MNO and began to voice their concerns with increasing regularity against most (if not all) of their products/services.

Que 3: **Which month does Econet experience the most complaints, on average?**

Ans 3: December seems to be the month when Econet experiences the most complaints on twitter. This could be because this is when traffic is most prevalent. People are on holiday and the diaspora comes back and uses the network, congesting it.

Que 4: **Which type of complaint is most prevalent?**

Ans 4: Data is the most prevalent complaint on twitter, followed by Ecocash. They make up 31% and 24% of all the complaint investigated, respectively. The content of the complaints seems to point at how data is too expensive and Ecocash doesn't work most of the time.

Que 5: **Is the amount of complaints Econet has gotten over the last 9 years correlated to its increase in market share and product range in the same time period?**

Ans 5: Enquiries posed to Econet on twitter do not seem to be correlated to either market share or Econet's number of products. This data then does not support the common hypothesis that the reason why Econet continues to get an increased number of complaints is because of its ever-increasing umbrella of brands/services. However, it should be noted that this conclusion was reached after only using 7 data points. It's a valid assumption that should be tested if more data was to be made available.

Recommendation: The data issue does not seem to be going away anytime soon because of the business environment we find ourselves in. I'd recommend creative packages tailored around how subscribers use their data such that they get the best value for money. As for Ecocash, some refactoring of the entire system might be necessary to minimize the inefficiencies that lead to customer inconveniences. If this isn't possible, the tech team might need to develop an automated system that processes and validates wrongful deductions of accounts such that customers might be refunded without much a hassle.

In conclusion, the data reveals that subscriber sentiment against Econet has become increasingly negative year-on-year, exponentially especially over the last 2 years. The Econet brand is fast being associated with extortion, inefficiency and out of touch with its customer base.

REFERENCES

1. Chikaura S 2019, *SimbaPy/EconetSentimentAnalysis*, Chikaura S, started 21 May 2019, <https://github.com/SimbaPy/EconetSentimentAnalysis>
2. Potraz.gov.zw 2019, *Reports - Postal and Telecommunications Regulatory Authority of Zimbabwe*, downloaded 24 March 2019, <http://www.potraz.gov.zw/?page_id=527>
3. Africanfinancials.com 2019, *ECONET WIRELESS ZIMBABWE LIMITED (ECO.zw) - AfricanFinancials,* downloaded 24 March 2019, <https://africanfinancials.com/company/zw-eco/>
4. Chauhan A 2018, Selenium | Tweepy to Scrap Tweets from Twitter and Analyzing Sentiment, visited 21 March 2019, <https://towardsdatascience.com/selenium-tweepy-to-scrap-tweets-from-tweeter-and-analysing-sentiments-1804db3478ac>