



**SEMESTER ONE 2024/2025 ACADEMIC YEAR**

**SCHOOL COMPUTING AND INFORMATICS TECHNOLOGY**

**DEPARTMENT OF COMPUTER SCIENCE**

**MASTER OF SCIENCE IN COMPUTER SCIENCE**

**MCS-7103 - MACHINE LEARNING**

**EXPLORATORY DATA ANALYSIS ASSIGNMENT 1**

**PATRICK ANAKU**

**2024/HD05/21916U**

**2400721916**

## Table of Contents

INTRODUCTION .....	3
DATA COLLECTION.....	3
DATA WRANGLING .....	3
EXPLORATORY DATA ANALYSIS .....	4
CONCLUSION .....	4

## INTRODUCTION

Customer satisfaction is an important growth component for any business and in this task, my main objective was to analyze the different touch points between the customer and the different departments of SONIC Company LTD. Such engagements include presales activities/data, technical assistance data, customer care data, and billing queries to find patterns and relationships for improving overall customer services and satisfaction.

- What can be done to improve customer satisfaction?
- What kind of information do I need to achieve this?
- What processes currently exist?
- How do customers reach out to us?

## DATA COLLECTION

The dataset was extracted from the customer management system of SONIC Co. LTD. This includes all logged interactions (queries or complaints) between the customer and sales team, customer and customer care team, customer and billing team, customer and the field support team. The period under consideration is January 2020 to December 2023. Below are the data sources in the dataset.

- Presales data – captured presales data include customer requirements
- Technical support data – interactions between customers and technical team.
- Customer care team – interactions between customer support team members while following up on issues.
- Billing queries - records relating to customer billing issues.

After collecting the data, below are the questions that I asked myself to help me in the data analysis

- What does the data represent.
- What insight can I learn from the data and use it effectively to improve customer support?

## DATA WRANGLING

Described in this section are the data pre-processing approach that I used to clean and organize the data. Python was chosen as the data analysis tool.

Important Python libraries: The first step in the data wrangling involved importing the python libraries Pandas, Matplotlib, Seaborn. Pandas is an open-source python data analysis and manipulation tool, Matplotlib is used to perform data visualization in python and lastly seaborn is also a data visualization tool (based on Matplotlib) and is closely integrated with the Pandas data structures making it easier to use, with beautiful statistical graphs.

Reading the Dataset: After extracting the raw customer support dataset, which was extracted in the CSV file format, I loaded the dataset into a Pandas Dataframe. The dataset was downloaded and named as a csv file and then loaded into Pandas data frame for cleaning and exploratory analysis.

Assessing the Dataset: the purpose of this phase was to mainly dig into the dataset and understand the structure, content and check if there are problems in the dataset. The detailed process involved looking at the features attributes in the dataset for instance the Pandas functions head() and dtypes() were useful in understanding the meaning of and datatype of each column in my dataset.

Dataset Cleaning: After understanding the structure and content of the dataset I realized that it was necessary to clean the data as there were some missing values and some attributes that were not important for my purpose. The process involved writing some python code that checked for any null value values and removed them, renamed certain fields as this dataset contains sensitive information.

## **EXPLORATORY DATA ANALYSIS**

After performing the data assessment part, I was able to identify certain features that were imported for my purpose.

Some of the important features I identified from the dataset were, the complaint or query category, the time was logged and resolved, service plan of customer, incident id, the customer account with SONIC, whether query came via email, phone or self-logged by user

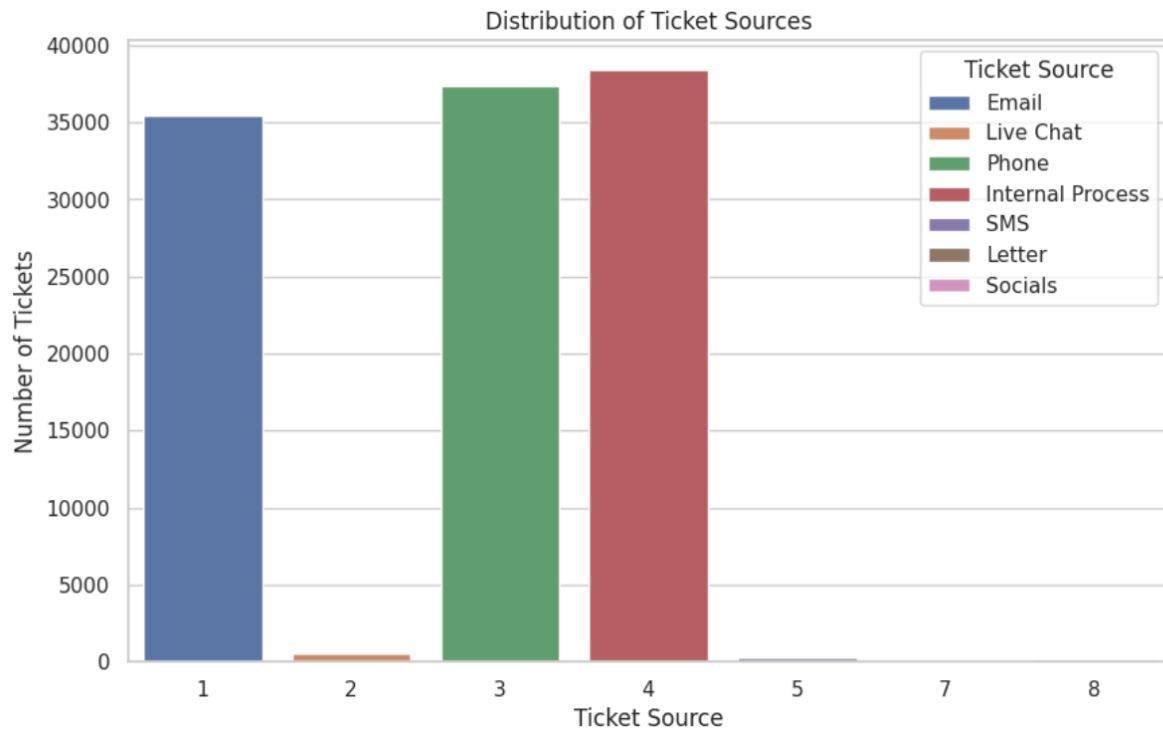
The features identified are subjected to bivariate analysis to understand more about the relationship between the variables.

## **CONCLUSION**

Below are my findings from the analysis of the customer support dataset

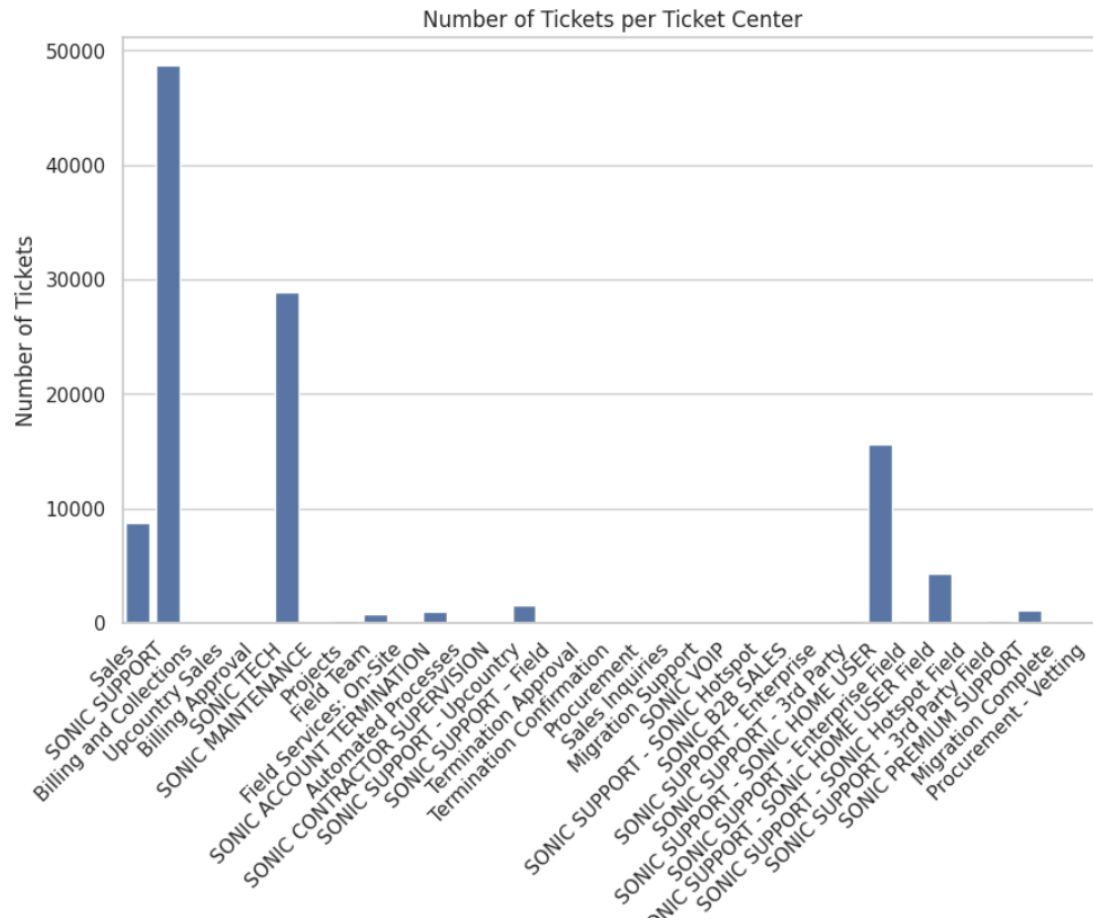
1. There are three main avenues through which customers interact with SONIC various SONIC departments. See appendix A
2. The trend is that the customer support teams generally respond to tickets based on ticket priority. Tickets with normal priority receive the least attention. See appendix B
3. Customer queries that come via a Letter are the most important and have least resolution time. Queries generated internally within the various SONIC departments take the longest to resolve. See appendix C
4. The most tickets are handled by the support team, SONIC Support. This mostly likely is the department that interacts with customers directly i.e. first level support.
5. There are negligible correlations between source of ticket, date it was logged, allocated time and time to close.

## Appendix A



## Appendix B





## Appendix E

