DATA 1201 – DATA COLLECTION

ASSIGNMENT 1: DATA COLLECTION AND MANAGEMENT MAP

PROFESSOR: MINA LABIB

PREPARED BY: AROWOLO BANKOLE

100833830

1ST OCTOBER, 2021

1. What is the name of your organization?

The name of my organization is **Applecart Nigeria**.



Applecart Nigeria is an online grocery which is located in Lagos, a state in Nigeria. Applecart Nigeria is a Nigeria-based company launched in 2014. Ojuolape Arojo, founder and CEO of her own business, she started the business mainly because shopping for groceries has never been easy for her. She understood that there are other people faces the same problem like hers, so with no background knowledge or professional skill, she set up an e-commerce website with payment features all on her own.

The whole idea of this business is to:

* save customers the time they would have used to drive to the market, find a car park and negotiating price with seller before driving back home to prepare their meal.
* Bring the market to ones smartphone so one can easily go through the items and prices in less time and received a ready-to prepare goods to ones doorstep.
* Provide customers with a healthy meal for their families.

She also faced the major problem most startups faces which is “discouragement”. With her good degree and background in school, many thought she should pursue a salary earned job instead of running a business.

My **main** reason for choosing this company is that applecart has an e-commerce website which makes the data entities more which makes thing easier for analysis.

1. What business problem is your organization trying to solve?

Problem solving is the act of understanding the customer’s interest, identifying the problem, finding the cause of the problem, looking for different way to solve the problem and carrying out the plan to solve the problem.

1. **Background of the problem:**

A survey was carried out and this was identified as the **critical challenge** for the business. Due to the COVID-19, online shopping has increased drastically and the **customer’s volume expectations** have also **increased**. Unfortunately, it also exposes the **Inconsistency** in the database as the **quantities** from online orders **outrun** the **quantities of the grocery items** available at the store. The consequence of this is an inability to **fulfill customer orders**.

1. **Definition of the problem:**

Applecart Nigeria is trying to solve the problem:

* **Availability of Products**
* **Customer Satisfaction**.
* **Freshness of Products**

1. **Root of the problem:**

Cause of the problem will be listed in this section.

* Just a few months back, COVID-19 stopped a lot of retail grocery shopping and the pandemic pushed online groceries to the top. Therefore, the competition in the online grocery business increased and with the **Inconsistency** in the **quantities of product available** in the company, it has lost even loyal customers.
* **Fresh products** have a short shelf-life. In online grocery, there is no means to know the **quality of a product** at the time of placing an order. Therefore, storing and delivering stale products causes **dissatisfaction among buyers**.

1. **Developing Solution for the existing problem:**

To improve these problems the company has tried the following:

* **Product availability** does not really mean the product is available 100% of the time but rather the product is made available when the **customer needs/orders** it. The company checked with the warehouse and also asks for the **database** to be **consistent** with the **quantities** **available** in the store.
* The company has made effort to communicate through the **use of alert notifications** to give an update on any **challenges** with the **customers** when **deliveries** are not likely to be on time so that the **customer** is kept waiting.
* For the challenge of **fresh products**, the use of **refrigerated** containers for **chilled** and **frozen items**, after they had been picked has improved.

1. **Conclusion:**

Applecart Nigeria has tried to find a solution for the existing problem but with the amount of data that has been accumulated over the years since the business started, they are going to need a professional data scientist to work on the data and adjust the database.

1. DATA STEPS:
2. Data Generation: **Where is the data coming from? What is it focusing on? Why were the data generated?** These are the question one asks no matter whether the data are found or designed. Applecart may generate data from the following methods:

* Customer Interviewing
* Online or Paper-based Questionnaire study
* Number of customers visiting page
* Online shopping records
* Online shopping Payment method. E.g. PayPal, Google Pay, Debit/Credit card etc
* Number of membership card purchase
* Effect of promos on customers
* Number of orders before/after covid-19
* Review and Rating on online products

In order to **generate** and **capture** the data in this situation; acts such as **interpreting**, **selecting** and **recording** the data are **important** to produce data that are as **accurate** indications as possible.

1. Data Collection & Transmission: This phase answers the question **what overall process will you use to collect the data?** **What format will the data come in?**

The data can be obtained from all possible data sources. Different data sources can be reliable in generating relevant data.

* It could be data from the **customers** through **interviews or survey** in which you can directly ask the **customers** for **relevant information**. It can both be conducted **online** or over **email** or **in person**.
* **Online Marketing campaigns** are also another way of collecting valuable data. Either it is run on **webpages** or elsewhere, the software used will provide feedback on the interactions on the **ads**, how many times it was clicked and more.
* **Social Media platform** such as **Instagram**, **Facebook**, **Twitter**, etc. has proven to be one of the most important sources of **customer data**. It can be used to determine the **popularity** of a product by the number of **engagement** or **feedback** it gets.

**Collecting** high-quality data is foundational for the business’ success. Therefore, different format will be used to **gather** the **information** the company need, as well as **organize** and **analyze** it.

* Text, Documentation, Scripts XML, PDF/A, HTML, Excel format
* Images: TIFF, JPEG 2000, PNG, JPEG/JFIF, BMP, GIF
* XML, CSV, TAB

1. Data Pre-Processing: In this step, the **raw data** collected from different data sources may be **messy** and is then **pre-processed** it into a **format** that can be **understood** and **analyzed**. The data can be pre-processed with the steps below:
2. **Integration:**

* This first step is an important step for this company because the data will be **collected** from different data sources.
* These data may come in different formats and will be **converted** to the **standard format**.

1. **Cleaning**:

* This involves the adding of **missing** data and **correcting** or removing **irrelevant** data. It is the most important step of preprocessing because it will ensure that your data is ready to go.
* Though we use our companies personal records which is as good as **clean** since the main source is the company’s own records, still, **Data cleaning** will **correct** all of the **inconsistent** data **collected** from web sources which contain **more errors**, **unwanted columns**, a huge number of **missing values**, etc.
* There are a number of ways to **correct** for **missing data**, few of them are:
* The first step to data cleaning is **removing unwanted observations** from our dataset. This includes duplicate or irrelevant observations. We are collecting data from **multiple sources**; it is inevitable to have irrelevant data.
* **Structural errors** such as typographical error and unnecessary capitalization can be checked.
* Outliers can have a huge impact on data analysis results. They should be **identified and filtered.**
* Since we will also be using **3rd party sources**, take a look for **missing data** **fields**, or **unanswered** survey questions. This arises from human error or incomplete data.

1. **Elimination of Redundant Data:**

* In this stage, data is reduced.
* The redundant data needs to be removed and make the data more understandable.
* Filtering and compression will be used to reduce the data.

1. **Data Storage:**

* Looking at the data that needs to be collected, the volume will be large and the velocity will be high. The company will need to choose a system that can provide a reliable storage space and store the data well.
* Hence, Applecart will use cloud-based virtual data storage.
* We will have the following ACID (Atomicity, Consistency, Isolation, and Durability) features in our data transition, especially for the online sales:
* Atomicity: ‘all occur, or nothing occurs.’ That is either all the operations (insert, update, delete) inside a transaction take place or none. For example, any waste merchandise should be recorded simultaneously so that there will be no gap in the inventory.
* Consistency: Data transition will be consistent. That is, for example, if a customer cancels the online purchase order after confirming it, it should be seen immediately both his/her application and in ours. Besides, the payment transaction should be trackable from both sides’ accounts.
* Isolation: Isolation determines how transaction integrity is visible to other users and systems. The company needs the isolation property to perform uninterrupted online sales. For example, when an online purchase order is created, from the center to all branches, it should be able to be seen. Besides, every branch must have its own order and sale details regarding the location.
* Durability: Durability is the property that guarantees that committed transactions will survive permanently. Durability is needed again especially in online sales. For instance, once the customers has filled the cart and then accidentally close the application by confirming the order, all the changes will be saved and seen in the next visit of the customer.

1. **Data Analysis:**

* The company needs solutions to its problems. To solve these problems, a comprehensive data analysis will be done. In particular:
* **Trending product analysis**: to determine which products are demanded most and least in online.
* **Social media analysis**: to see how well the opinion of the public towards the company’s online store among social media users.
* **Online sales analysis**: to observe how fast the demand for online sales is increasing.
* **Precaution analysis**: to determine and decide what kind of extra measures can be taken to protect the staff, customer, and public health during the pandemic.

For data analysis, our team as data management specialists needs the following tools:

* **R** will be used for doing the statistical analysis. For instance, profit, mean, standard deviation, etc. Some diagrams can be generated to increase the understanding level.
* **SQL** will be used for adding queries to and deleting queries from the dataset together with joining the tables.
* Tableau will be used for visualization.
* We will use **graphs, charts, maps, and diagrams** for the best visualization.

1. **References**
2. Applecart Nigeria Webpage: <https://www.applecartng.com/>
3. DATA 1201- Data Collection course materials (Assignment#1-Example)