

# DATA COLLECTION DATA 1201-02

Assignment 3 - Data Technology Plan

# Submitted by: AROWOLO BANKOLE (100833830)

# Professor: Dr. Mina Labib

## Review:

* Name of the organization: **Applecart Nigeria.**

## **Applecart Nigeria** is a standard and licensed online grocery business that is located in one of the largest and busiest cities and state in **Nigeria** – **Lagos**. It offer an **alternative** to market and restaurants during tough economic times, as buying groceries for home cooking is one alternative. Buying prepared foods/drinks, which are new **higher-margin** offerings at grocery stores, is an additional alternative to restaurants.

* The **mission** of Applecart Nigeria is to **establish** an online grocery business that will make available a **wide range of goods and products** from top production at **affordable prices** to the residence of Lagos and **reduce the stress** of having to take a trip to the market place.

## Data Technology Plan:

We believe that it is our responsibility to ensure our **products** to be **good for consumption** and **achieve better customer satisfaction**, through the effective and appropriate use of technology.

In order for technology use to meet customer needs and align with the business mission and vision, a technology plan must be developed.

A **technology plan** can save money by helping you to make **good decisions**, not spend more than you need to meet present and future needs, and make **better use of resources**. It works to match what you have to do for making **storage**, **preprocessing** and **visualization** decisions and with the tools you need to do it.

In this Data Technology Plan, we will be using two criteria**: Technical Criteria** and **Business Criteria** while also selecting the best technology.

The Company will be using the following tools for its Data Technology plan:

* **Hadoop & SQL**
* **Tableau**

## **R Studio**

## **Python**

## Here are the Measure Names and their Technology Plans:

### Order Management & Order Tracking

|  |  |  |  |
| --- | --- | --- | --- |
| Data Phase | Key Technical Criteria | Key Business Criteria | Technology Solution |
| Storage | This data has been collected from **in-store sources** with the help of the Admin and the data collecting team.As previously mentioned in the data collecting plan, the **order managements and tracking** are collected by the time of order placed.The Admin will upload the data on cloud data storage by using **SaaS**. The data team will store the conveyed data in **SQL and Hadoop.** | The Admin who is responsible for **conveying the data** to the data collection team is **suitable** at **handling the data**.On the collection frequency which was mentioned in the data collection plan, the Admin will forward the data to the collection team at the end of every week.The data collection team is experienced and will handle this the way they have been handling past records. | After the collecting and conveying of data to the data team, we will be using **Hadoop** together with **SQL**.**SQL** is used because of its **high performance, productivity, maintainability and security**.**Hadoop** is used because it is very **reliable** and **flexible** to store data. It is also able to **handle big data** when needed easily. |
| Pre-processing | **Thousands of data** are collected are collected from users through the **online & offline surveys, interviews.****Customer feedbacks** are also collected. | Some of the **limitations** are:Data can be **incomplete**.**Inaccurate data****Unnecessary data**Data format may vary from different sources.Recommendations:**Data cleaning** – removal of irrelevant data.Mandatory fields/question when giving feedbacks. | From the **Technical** and **Business criteria**, we will be using both **R studio** and **Python** for the **data pre-processing.** **Python** been a **versatile** programming language, we will use it for our **data cleaning** and **removing** of the **unnecessary data.****R studio** is known to have an **extensive library of tools** for data and database manipulation. We will be using it for preprocessing the big data, developing **improved statistical tests** and **models** and can also be used to create some visualization. |
| Visualization | **Data Visualization** is an important process. It helps to tells stories by **presenting data** into a form that is **easier to understand** and make **better decisions**.As mentioned in Data collection and Management map, we will be using **Tableau** for our visualization. We will be using **graphs, charts, maps,** and **diagrams** for user friendly.We will be displaying all these in a **tableau dashboard** to introduce to the stakeholders in a presentation format and final outcome for decision making. | The **business stakeholders** will observe our technology to carry out their findings by making use of the analysis report that the team presents to them to make their decisions.Public sources data like, **feedbacks** or **surveys** may contain **inaccurate** or **unnecessary** data. This can be overcome by using a **sample of data** for the visualization. | For the **visualization**, we will be making use of the **Tableau**.We can use different **charts** for different reasons. The **bar chart** can be used to compare **multiple products** and give visualization on the bestselling products over the years, **line chart** can be used for the **sales trends** etc.**Tableau** is very **creative** and **interactive** software.It can handle **big data** with ease and different format.The dashboard in tableau is very **responsive** and **easy to navigate**. |

### Product Rating, Incomplete Product Claim, Expired products at the warehouse

|  |  |  |  |
| --- | --- | --- | --- |
| Data Phase | Key Technical Criteria | Key Business Criteria | Technology Solution |
| Storage | This data has been collected from **in-store sources** with the help of the Admin and the data collecting team.As previously mentioned in the data collecting plan, the **order managements and tracking** are collected by the time of order placed.The Admin will upload the data on cloud data storage by using **SaaS**. The data team will store the conveyed data in **SQL and Hadoop.** | The Admin who is responsible for **conveying the data** to the data collection team is **suitable** at **handling the data**.On the collection frequency which was mentioned in the data collection plan, the Admin will forward the data to the collection team at the end of every week.The data collection team is experienced and will handle this the way they have been handling past records. | After the collecting and conveying of data to the data team, we will be using **Hadoop** together with **SQL**.**SQL** is used because of its **high performance, productivity, maintainability and security**.**Hadoop** is used because it is very **reliable** and **flexible** to store data. It is also able to **handle big data** when needed easily. |
| Pre-processing | **Thousands of data** are collected are collected from users through the **online & offline surveys, interviews.****Customer feedbacks** are also collected.Count **every** warehouse expired products. | Some of the **limitations** are:Data can be **incomplete**.**Inaccurate data****Unnecessary data**Data format may vary from different sources.Recommendations:**Data cleaning** – removal of irrelevant data.Mandatory fields/question when giving feedbacks. | From the **Technical** and **Business criteria**, we will be using both **R studio** and **Python** for the **data pre-processing.** **Python** been a **versatile** programming language, we will use it for our **data cleaning** and **removing** of the **unnecessary data.****R studio** is known to have an **extensive library of tools** for data and database manipulation. We will be using it for preprocessing the big data, developing **improved statistical tests** and **models** and can also be used to create some visualization.**Python** will be used for **preprocessing** every expired product in the warehouse. |
| Visualization | **Data Visualization** is an important process. It helps to tells stories by **presenting data** into a form that is **easier to understand** and make **better decisions**.As mentioned in Data collection and Management map, we will be using **Tableau** for our visualization. We will be using **graphs, charts, maps,** and **diagrams** for user friendly.We will be displaying all these in a **tableau dashboard** to introduce to the stakeholders in a presentation format and final outcome for decision making. | The **business stakeholders** will observe our technology to carry out their findings by making use of the analysis report that the team presents to them to make their decisions.Public sources data like, **feedbacks** or **surveys** may contain **inaccurate** or **unnecessary** data. This can be overcome by using a **sample of data** for the visualization. | For the **visualization**, we will be making use of the **Tableau**.We can use different **charts** for different reasons. The **bar chart** can be used to compare **multiple products** and give visualization on the bestselling products over the years, **line chart** can be used for the **sales trends** etc.**Tableau** is very **creative** and **interactive** software.It can handle **big data** with ease and different format.The dashboard in tableau is very **responsive** and **easy to navigate**. |

### Web Data

|  |  |  |  |
| --- | --- | --- | --- |
| Data Phase | Key Technical Criteria | Key Business Criteria | Technology Solution |
| Storage | This data has been collected from **in-store sources** with the help of the Admin and the data collecting team.As previously mentioned in the data collecting plan, the **order managements and tracking** are collected by the time of order placed.The Admin will upload the data on cloud data storage by using **SaaS**. The data team will store the conveyed data in **SQL and Hadoop.** | The Admin who is responsible for **conveying the data** to the data collection team is **suitable** at **handling the data**.On the collection frequency which was mentioned in the data collection plan, the Admin will forward the data to the collection team at the end of every week.The data collection team is experienced and will handle this the way they have been handling past records. | After the collecting and conveying of data to the data team, we will be using **Hadoop** together with **SQL**.**SQL** is used because of its **high performance, productivity, maintainability and security**.**Hadoop** is used because it is very **reliable** and **flexible** to store data. It is also able to **handle big data** when needed easily. |
| Pre-processing | **Analysis** ofevery data from the public survey**Thousands of data** are collected are collected from users through the **online & offline surveys, interviews.****Customer feedbacks** are also collected. | Some of the **limitations** are:Data can be **incomplete**.Some **un-structured data** not been able to be converted to **structured data****Inaccurate data****Unnecessary data**Data format may vary from different sources.Recommendations:**Data cleaning** – removal of irrelevant data.Mandatory fields/question when giving feedbacks. | From the **Technical** and **Business criteria**, we will be using both **R studio** and **Python** for the **data pre-processing.** **Python** been a **versatile** programming language, we will use it for our **data cleaning** and **removing** of the **unnecessary data.****R studio** is known to have an **extensive library of tools** for data and database manipulation. We will be using it for preprocessing the big data, developing **improved statistical tests** and **models** and can also be used to create some visualization.Both **R Studio** and **Python** will be used for pre-processing of every public survey carried out. |
| Visualization | **Data Visualization** is an important process. It helps to tells stories by **presenting data** into a form that is **easier to understand** and make **better decisions**.As mentioned in Data collection and Management map, we will be using **Tableau** for our visualization. We will be using **graphs, charts, maps,** and **diagrams** for user friendly.We will be displaying all these in a **tableau dashboard** to introduce to the stakeholders in a presentation format and final outcome for decision making. | The **business stakeholders** will observe our technology to carry out their findings by making use of the analysis report that the team presents to them to make their decisions.Public sources data like, **feedbacks** or **surveys** may contain **inaccurate** or **unnecessary** data. This can be overcome by using a **sample of data** for the visualization. | For the **visualization**, we will be making use of the **Tableau**.We can use different **charts** for different reasons. The **bar chart** can be used to compare **multiple products** and give visualization on the bestselling products over the years, **line chart** can be used for the **sales trends** etc.**Tableau** is very **creative** and **interactive** software.It can handle **big data** with ease and different format.The dashboard in tableau is very **responsive** and **easy to navigate**. |

REFERENCES:

1. Applecart ng web page: <https://www.applecartng.com/>

## DATA 1201 Assessment 1: Data Collection and Management Map

## DATA 1201 Assessment 2: Data Collection Plan

## DATA 1201 Assignment 3 Examples