|  |  |
| --- | --- |
| **Assignment Case** |  |
| COMP6140 | COMP6140001 | COMP6140049  Data Mining |
| **Computer Science** | **Example Case** |
| ***Valid on*** *-* | **Revision 00** |

## Soal

*Case*

**The Convenient Store**

**The Convenient Store** is a conveniently placed store where many shoppers go to buy their daily goods**. The Convenient Store** has chosen you to help them **analyze** and **visualize** all their transaction data so that the shop manager knows what needs improvements.

**The Convenient Store** requested you to research the data that is available. The data is in csv format as the following:

* **orders**.**csv**

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Data Type** | **Description** |
| order\_id | Integer | The id of the transaction |
| product\_id | Integer | The id of the product |

* **products**.**csv**

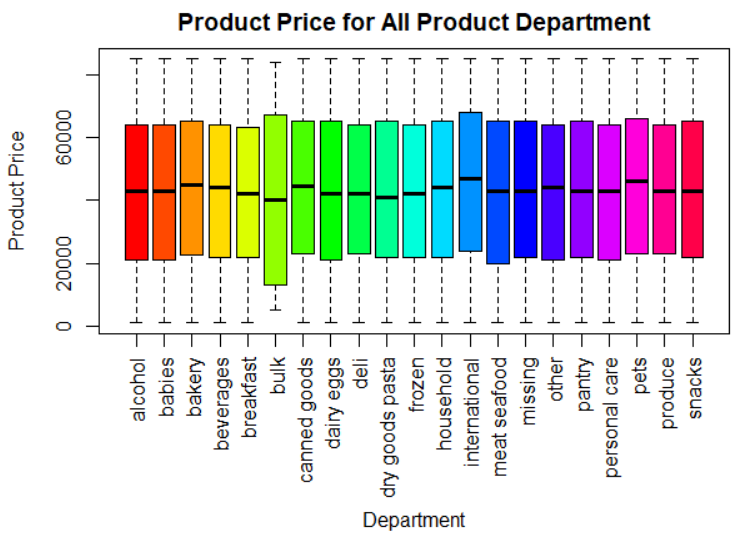
|  |  |  |
| --- | --- | --- |
| **Attribute** | **Data Type** | **Description** |
| product\_id | Integer | The id of the product |
| product\_name | Character | The name of the product |
| aisle | Character | The aisle where the product is located |
| department | Character | The department where the product is located |
| product\_price | Integer | The price of the product |

You are asked to help him analyze and visualize the data based on specification below:

1. **Data Visualization**

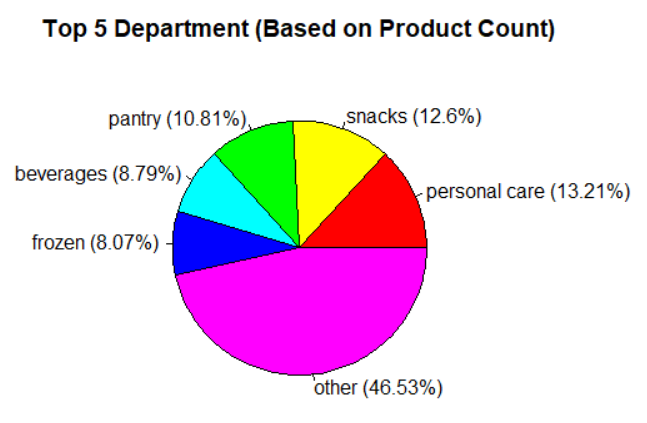
To describe the data better, you are asked to visualize the data in graph form. Some data that needed to be visualized are:

* 1. Show the **Product Price** for **all Product Department.**



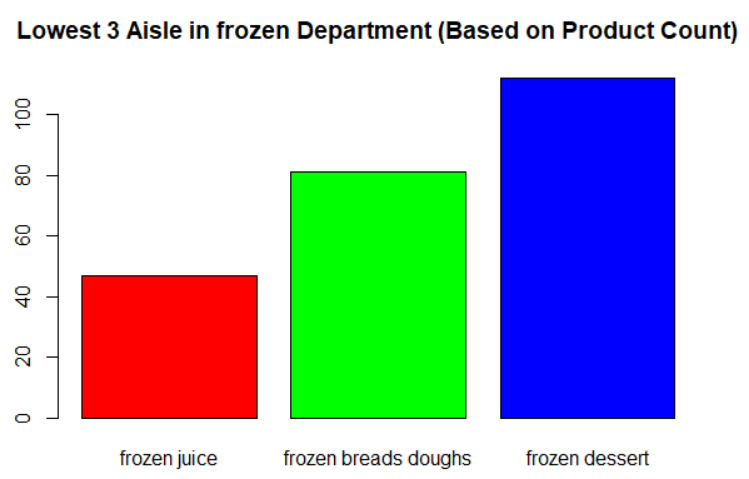
**Figure 1. Product Price for all Product Department Chart**

* 1. Show **top 5** **department** basedonits **product count**. Product which is **not** in the **top 5 department** will go into “**other**” categories. For each area, put **label** to show **percentage of the department** mentioned along with **department name**.



***Figure 2. Top 5 Department based on Product Count Chart***

* 1. Show the **lowest 3 aisle** based on its **product count**. Take only data in which the **department** is **frozen**.



**Figure 3. Lowest 3 Aisle based on Product Count Chart**

1. **Frequent Pattern Analysis**

You are asked to do frequent pattern analysis to know the **frequent product** that the people bought. To get the frequent product, use “**orders.csv**” and “**products.csv**” and follow all steps below:

1. **Data Preprocessing**

In thisphase, some data can't be used for further analysis. Do the following task to **cleanse** the data:

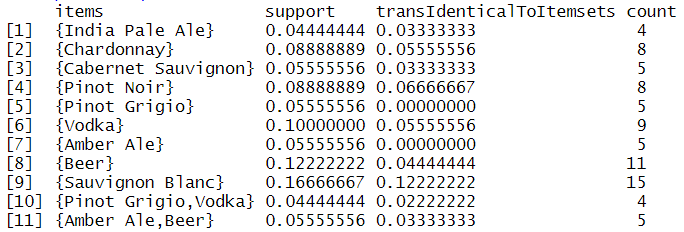
* Removeall **product** which department **is** **not** **alcohol**
* Removeall **product** which aisle is **Specialty wines champagnes**
* Removeall **duplicated** data for the analysis

1. **Data Transformation**

In this phase, you need to change the data, so it is suitable to be used in the **Apriori** analysis. Prepare the product data in terms of the **product’s name**.

1. **Data Mining**

* Show **frequent product** using **Apriori** algorithm with **minimum support**: **0**.**04** based on the data that have already pre-processed



**Figure 4. Frequent Product Result using Apriori**

* Show the **association** **rules** using **minimum confidence**: **0**.**5** based on the **frequent product** that resulted from step above.



**Figure 5. Association Rules Result**

**References**:

https://www.kaggle.com/c/instacart-market-basket-analysis/data

**If you do not understand, please ask your assistant!**