

Department (Computers And Data Science )

1st Year

Introduction to Data Science

Final Project with R

|  |  |  |
| --- | --- | --- |
| Name | ID | Group |
| Ali Mohamed Sayed Ahmed | 20221449583 | G 1 |
| Zeyad Ashraf Ibrahim Taher | 20221369225 | G 1 |
| Ahmed Khalaf Elsayed Ramadan | 10 | G 1 |
| Abdelwahab | 21 | G 2 |
| Abdallah | 45 | G 1 |

* **What will the program do?**
* The program performs operations on the data:

**(1) Data Visualization** to display data and display (Compare cash and credit totals) and (Compare each age and sum of total spending) and (Show each city total spending and arrange it by total descending) and (Display the distribution of total spending)

**(2) put all The plots in one dashboard** )It is a large table that includes a set of charts and tables(

**(3) Split the customers to (n)groups by using k\_means** To link each data to its approximate central points.

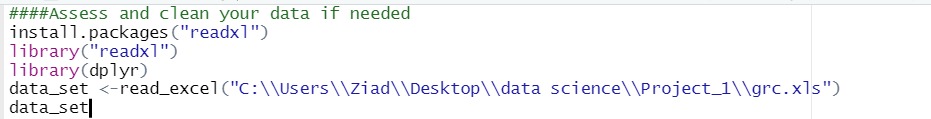
**(4) Generate association rules between item** association rules are used to find correlations and co-occurrences between data sets

* **What the input to the program will be:**
* Data Visualization inputs (paymentType col –age col – city col – total col )
* The dashboard **(** 4 plots )
* K\_means inputs are 3 columns ( age, total, custmor) and number of clusters from the user
* Association rules “apriori\_algorithm “ inputs (min\_support and min\_ confidence from the user – items column )
* **What the output from the program will be.**

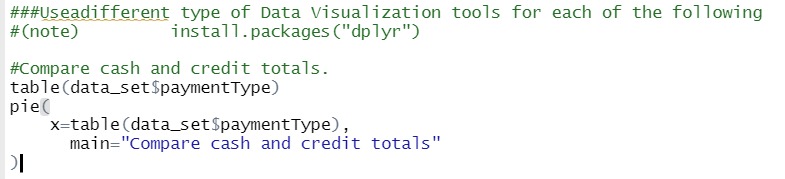
1. One dashboard with four plots (2 Barblot – 1 Pie – 1 Boxplot)
2. Split customers into number of cluster (depend on the user input) by applying k\_means method
3. Generate association rules between items by applying apriori\_algorithm

* **The full description of your dataset:**
* Our Dataset consists of 9836 Records and 8 columns.
* It contains different items and the count of them.
* It contains different customers who bought items, their ages and their cities.
* It contains payment methods customers used during purchasing their items (cash-credit).
* It contains RND column.
* **Screen shots from your Project steps.**

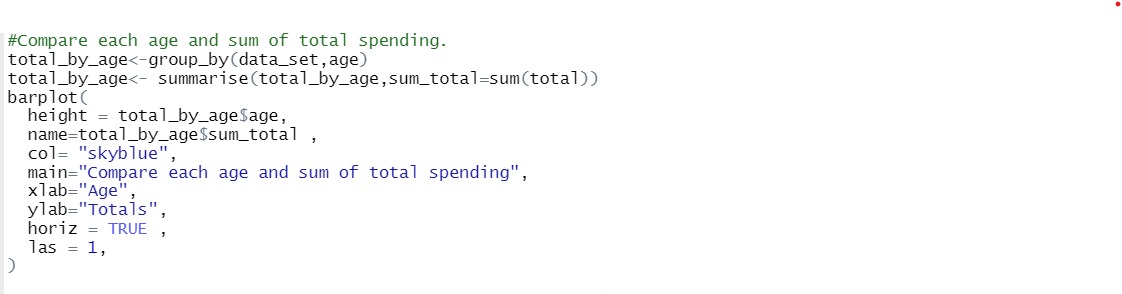
**(1)**



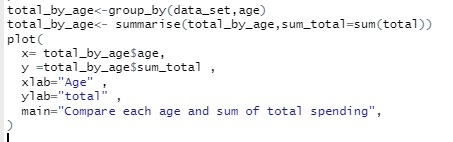
**(2)**

****

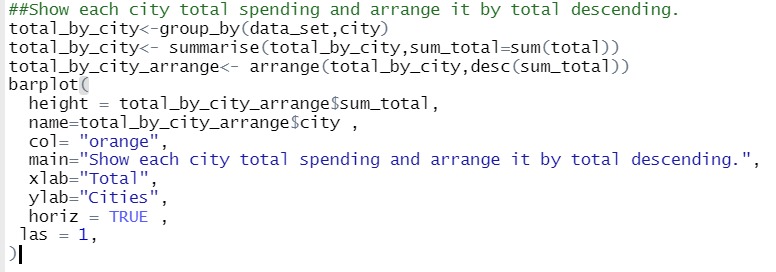
**(3)**

****

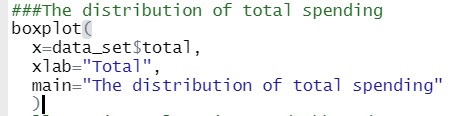
**Another(3)**

****

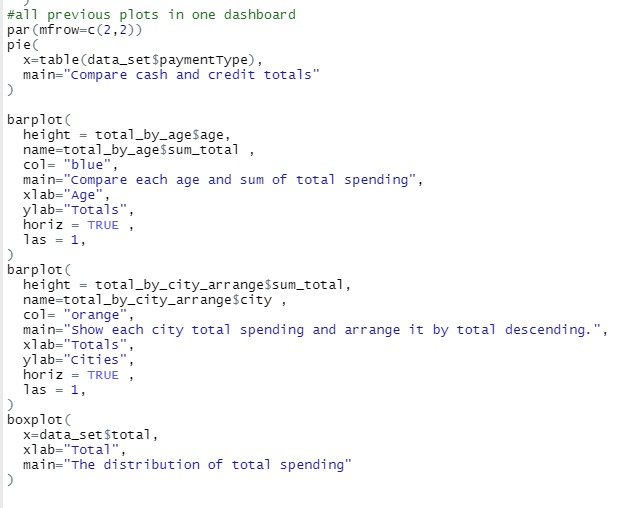
**(4)**



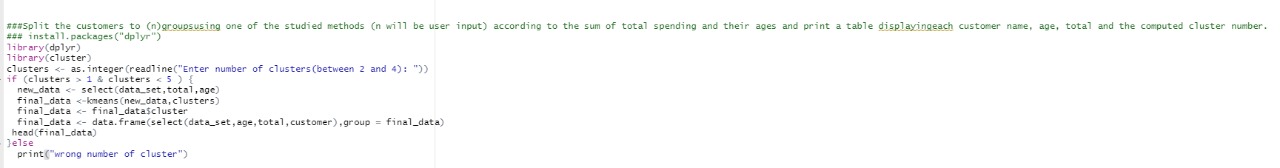
**(5)**

****

**(6)**

****

(7)



(8)

