MCDA5580 Assignment 3

Team Member

* Hemalatha Srinivasan A00452621
* Ajay Jain A00455849
* Kin Wa, Chan A00467755

**Table of Content**

[Executive Summary 2](#_Toc129011619)

[Objectives 2](#_Toc129011620)

[Data Analysis 3](#_Toc129011621)

[Design/ Methodology/ Approach 4](#_Toc129011622)

[Overview 4](#_Toc129011623)

[Customer Level Analysis 4](#_Toc129011624)

[Session Level Analysis 4](#_Toc129011625)

[Conclusion 7](#_Toc129011626)

[Appendix 8](#_Toc129011627)

[R Script 8](#_Toc129011628)

[Customer Level Analysis 8](#_Toc129011629)

[Session Level Analysis 8](#_Toc129011630)

[Reference/ Citation 10](#_Toc129011631)

# Executive Summary

# Objectives

SimplyCast is a company provides software solutions for marketing communication automations (e.g. email automation, Twitter automation). The company website “<https://www.simplycast.com/>” played a important role in the business, it serves many functions including the presentation of production information, show use cases/ demos, providing customer service and support. It is important for the company to promote its products and maintain its services to customers.

In this report a study will perform to analyze the associations between functions from the browsing history of the website to achieve these goals:

* Reveal user behavior, adjust marketing strategy. E.g. Track associations of the products browsed, develop packages to attract customers.
* Improve user experience in UI. E.g. Discover frequent used functions and make them conspicuous on UI,
* Discover potential UI design problem E.g. Users have difficulties get access to a function due to wrong descriptions or over complexity on UI.

# Data Analysis

# Design/ Methodology/ Approach

## Overview

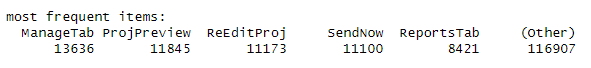
In order to analyze the associations between functions in the websites, Apriori algorithm is used. It is because the data is not labeled so an unsupervised learning method is needed. In addition, this method could be able to list out all association rules with specific support and confidence which are intuitive and easy to communicate with end users.

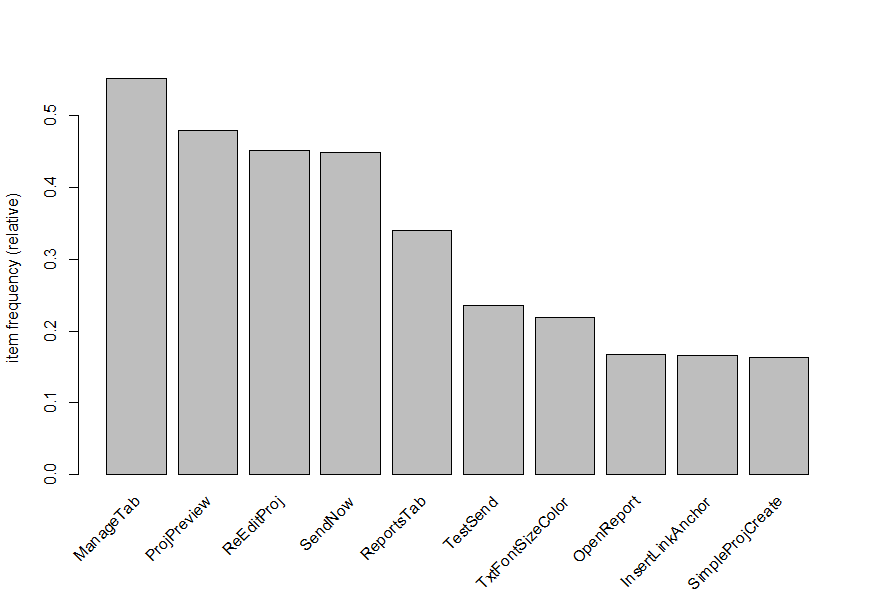
There are two types of browsing history extracted from the website: Customer level browsing history and Session level browsing history. Analysis by Apriori algorithm will be carried out separately on these two sets of data and final summary will be made based on the results.

## Customer Level Analysis

## Session Level Analysis

Session level browsing history is gathered from website to proceed the analysis. There is the summary of the most frequent user accessed functions:



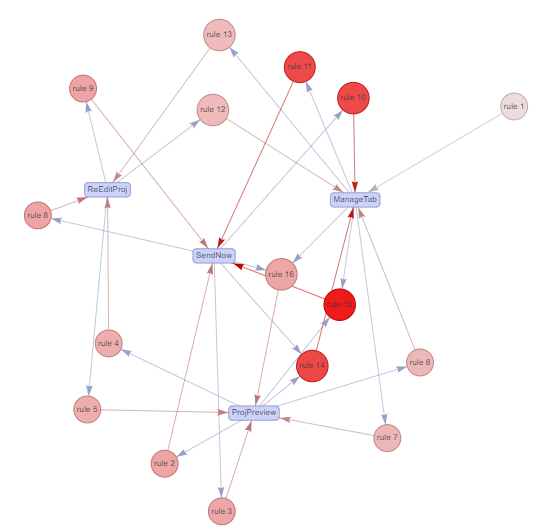


Apriori algorithm is used to analyze the session level data. Here is the list of top 16 rules (support =0.25, confidence =0.5) ordered by lift.

一張含有 文字 的圖片

自動產生的描述

Plot for the rules



Most frequent item set by support

一張含有 文字 的圖片

自動產生的描述

Analysis

The top rule is {ManageTab, ProjPreview} => {SendNow} have both high confidence (~0.8) and lift (~1.77) which means user highly likely to access {SendNow} when {ManageTab, ProjPreview} are accessed. Moreover, the support (~0.25) of this rule is high, it implies a quarter of records contains the {ManageTab, ProjPreview} item set. These all make this rule very meaningful.

In the second rule {ProjPreview, SendNow} => {ManageTab}, the items listed are the same as the first rule and also with high value in confidence (~0.9), lift (~1.64) and support (~0.25). Therefore, we can see there is close relationship between the items {ManageTab, ProjPreview}, and {SendNow}.

The rules 3 and 4 are {SendNow} => {ManageTab} and {ManageTab}=> {SendNow} respectively. They have very high support value (~0.4) and similar confidence (~0.73 to 0.9) and lift (~1.63) compared with rules 1 and 2. It shows the associations appear extremely frequent and relationship between {SendNow}, {ManageTab} is strong.

The rules 5 and 6 are {ReEditProj} => {SendNow} and {SendNow} => {ReEditProj}, the patterns are similar to rules 3 & 4 ({ManageTab} is replaced by {ReEditProj}) with high support (~0.27) but lower confidence (~0.59 to 0.6) and lift (~1.32). It shows the association between {ReEditProj} and {SendNow} are strong that should not be ignored.

The remaining rules are somehow the combinations of the most frequent items ({ManageTab}, {ProjPreview}, {ReEditProj}, {SendNow}) . It shows the importance of the items on the website and also verify the consistency of our analysis.

# Conclusion

# Appendix

## R Script

### Customer Level Analysis

### Session Level Analysis

library(plyr)

library(arules)

library(arulesViz)

df\_session = read.csv("D:/#Spring 2023/5580 - Text Mining/Assignment3/sessionDistinctMilestoneDec15 (2).csv")

df\_session$id = paste(as.character(df\_session$user\_id) , as.POSIXct(df\_session$date, format="%Y-%m-%d"))

# Transpose source

df\_session= ddply(df\_session,c("id"),function(dfl)paste(dfl$milestone\_name, collapse=","))

# Remove id field

df\_session$id = NULL

df\_session$date = NULL

df\_session$user\_id = NULL

# Write to temp file

write.table(df\_session,"D:/#Spring 2023/5580 - Text Mining/Assignment3/session2.csv",quote=FALSE,row.names=FALSE,col.names=FALSE)

# Read temp file

tr = read.transactions("D:/#Spring 2023/5580 - Text Mining/Assignment3/session2.csv",format="basket",sep=",")

# Plot Frequent items

#summary(tr)

itemFrequencyPlot(tr, topN=10)

# List the rules

rules<- apriori(tr, parameter= list(supp=0.25, conf=0.5))

inspect(sort(rules,by='lift'))

# Visualize the rules

plot(rules,

method="graph",

engine="visNetwork")

# List item set

itemsets=unique(generatingItemsets(rules))

inspect(sort(itemsets,by='support'))

## Reference/ Citation

Comparison of Association Algorithm

<https://www.educba.com/association-rules-in-data-mining/>

<https://www.upgrad.com/blog/association-rule-mining-an-overview-and-its-applications/>

Apriori Algorithm

<https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/>

<https://www.engati.com/glossary/apriori-algorithm>

<https://www.javatpoint.com/apriori-algorithm>