|  |
| --- |
| detail of persons hands with scissors, markers, workingPROJECT REPORT: MARKET BASKET ANALYSIS |

|  |  |  |
| --- | --- | --- |
| Team 5 |  |  |
| |  | | --- | | Jyoti Lnu | | Bhavnil Patel | | Dhairya Sheth | | |  | | --- | | Ansshita Varshney | | Ruofan Wu | | Chongli Zhao | |  |

TABLE OF CONTENTS

[MARKET BASKET ANALYSIS 3](#_Toc14979178)

[Strategic Highlights / Business Case 3](#_Toc14979179)

[Objective 3](#_Toc14979180)

[Data Collection 3](#_Toc14979181)

[Data Preparation 3](#_Toc14979182)

[Analysis / Model 5](#_Toc14979183)

[Conclusions 8](#_Toc14979184)

[FINANCIAL STATEMENTS 9](#_Toc14979185)

[Statement of Financial Position 9](#_Toc14979186)

[Statement of Comprehensive Income (Profits and Losses) 9](#_Toc14979187)

[Statement of Changes in Equity 9](#_Toc14979188)

[Statement of Cash Flows 9](#_Toc14979189)

[References 10](#_Toc14979190)

# MARKET BASKET ANALYSIS

## Strategic Highlights / Business Case

Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items.

End goal is to design a suitable marketing strategy for the target customers.

## Objective

1. How to achieve customer segmentation using machine learning algorithm (KMeans Clustering) in Python in simplest way.
2. Who are your target customers with whom you can start marketing strategy [easy to converse]
3. How the marketing strategy works in real world

## Data Collection

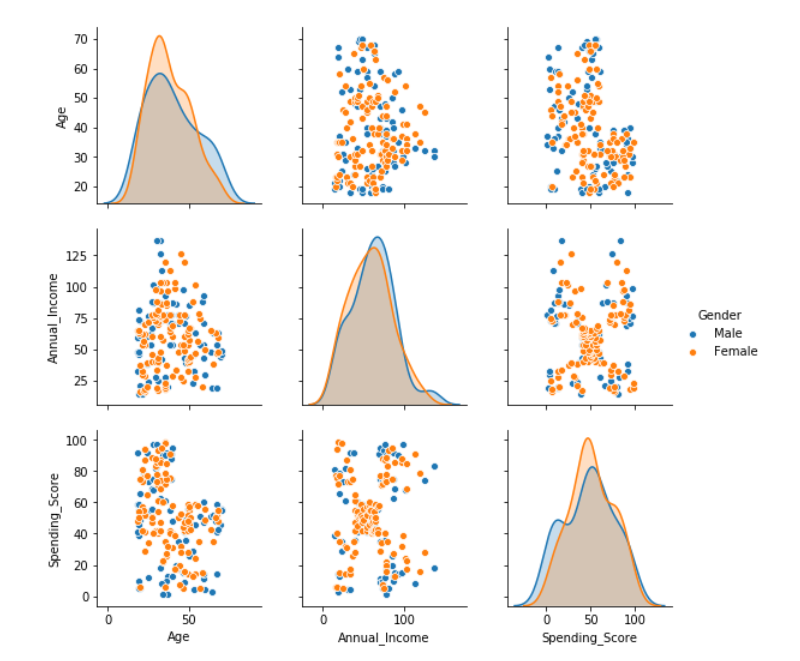
Through membership cards, some basic data about customers like Customer ID, age, gender, annual income and spending score are gathered for a shopping mall’s customers.

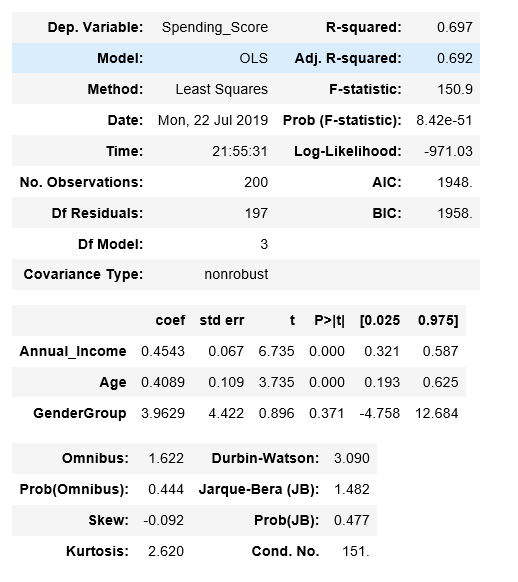
***Spending score is a score assigned to the customer based on certain defined parameters like customer behavior and purchasing data.***

|  |
| --- |
| ***Problem Statement: Understand the likelihood of buying or predicting the customers who are more likely to converge [Target Customers] so that marketing strategies could be designed to cater to those target customers.*** |

## Data Preparation

Summary observation on original dataset:





1. No clear linear relationship between the spending \_score and independent variables like annual\_income and age.

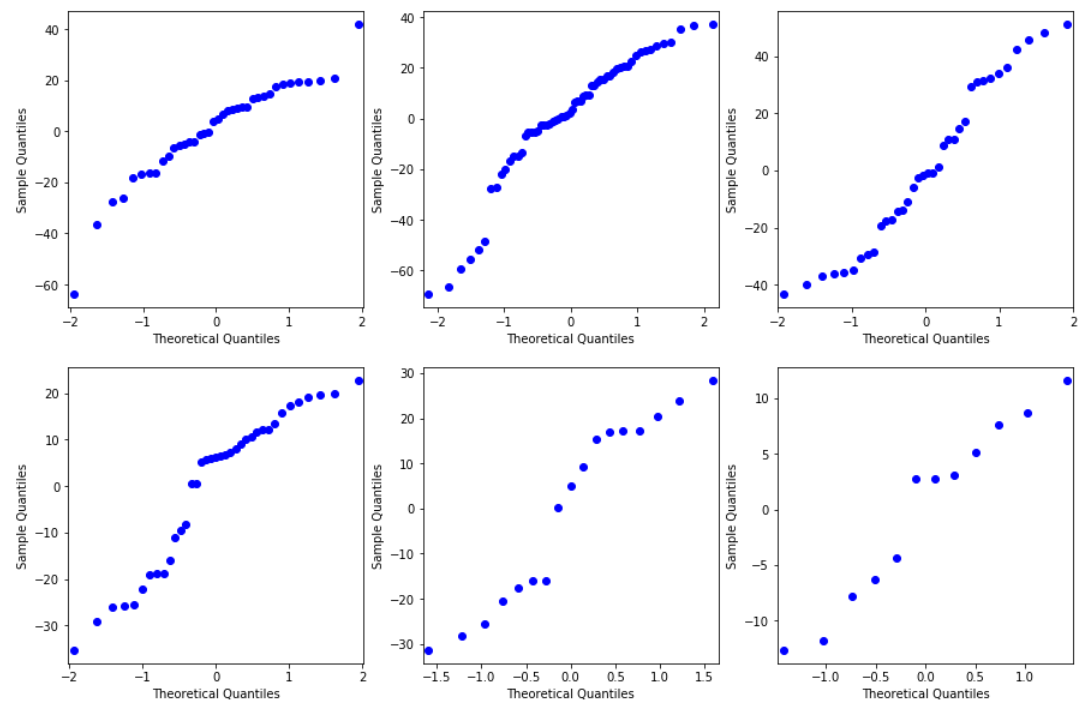
2. A single linear multivariate regression that models the relationship between the dependent variable and the independent variables returns a low R-squared at around 70%

3. An intercept should be excluded as no annual income suggests no spending power/score.

## Analysis / Model

Apply a forward elimination on the linear multivariate regressions on each age group to select the combinations of independent variables that yield highest R-squared.

|  |  |  |
| --- | --- | --- |
| Age Group | Regression | R-squared |
| Below 25 | Spending\_Score = 4.0296 \* Age - 0.6769 \* Annual\_Income  (T-value:0) (T-value:0) | 89% |
| 26 - 35 | Spending\_Score = 1.6762 \* Age + 0.1718 \* Annual\_Income  (T-value:0) (T-value:0.11) | 86% |
| 36 - 45 | Spending\_Score = 0.5828 \* Age + 0.2917 \* Annual\_Income  (T-value:0.13) (T-value:0.15) | 70% |
| 46 - 55 | Spending\_Score = 0.7401 \* Age  (T-value:0) | 83% |
| 56 - 65 | Spending\_Score = 0.5358 \* Age  (T-value:0) | 71% |
| Above 66 | Spending\_Score = 0.2173 \* Age + 0.6344 \* Annual\_Income  (T-value:0.01) (T-value:0.19) | 97% |



Observation:

1. R-squared for regression on each age group improve significantly.

2. the residuals are normally distributed.

3. Some independent variables have T-value higher than 5%, suggesting that they are not a significant determinant for the spending\_score. However, we argue that including those variables give us the best predicting power as combing them with other variables yield highest R-squared.

4. Annual\_Income is negatively correlated with the spending\_score for customers below 25.

5. Annual\_Income does not seem to have any relationship with the spending\_score for customers between 46 years old to 65 year old.

6. there are not many data for the age group above 56 (56-65 and above 66). Therefore, the regression results in these two groups could change significantly with more data collection.

# Conclusions

Use this section to give a brief summary of your financials, highlighting important points. Some of the sample text in this document indicates the name of the style applied, so that you can easily apply the same formatting again.

For example, this is the List Bullet style.

Here is another sentence formatted in List Bullet style.

You can find easy-to-use tools on the Insert tab, such as to add a hyperlink, insert a comment, or add automatic page numbering.

|  |
| --- |
| people at a table working |

View and edit this document in Word on your computer, tablet, or phone. You can edit text; easily insert content such as pictures, shapes, and tables; and seamlessly save the document to the cloud from Word on your Windows, Mac, Android, or iOS device.

# FINANCIAL STATEMENTS

## Statement of Financial Position

* Liabilities
* Statement of Financial Position
* Ownership Equity

## Statement of Comprehensive Income (Profits and Losses)

* Income
* Expenses
* Profits

## Statement of Changes in Equity

Well, it wouldn’t be an annual report without a lot of numbers, right? This section is the place for all those financial tables.

To get started with a table that looks just like the sample here, on the Insert tab, tap Table.

|  |  |  |  |
| --- | --- | --- | --- |
| DESCRIPTION | REVENUE | EXPENSES | EARNINGS |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Statement of Cash Flows

* Operating
* Investing
* Financing

# References

* <https://www.albionresearch.com/data_mining/market_basket.php>