



Mahidol University

การจัดกลุ่มลูกค้าด้วยเทคนิค K-mean clustering

ปราศรัย วัชรพิสิฐศิริ

Wisdom of the Land

Introduction

- เป็นข้อมูลทางการตลาด ที่เก็บจากข้อมูลทั่วไปของลูกค้า เช่น รายได้ ระดับการศึกษา หรือ สถานภาพการสมรส
- เป้าหมายของงานการจัดกลุ่มลูกค้า เพื่อช่วยให้ทราบถึงลักษณะและพฤติกรรมของลูกค้าที่คล้ายกัน เพื่อการส่งข้อเสนอพิเศษหรือโปรโมชั่นที่เหมาะสมกับลูกค้าแต่ละกลุ่ม

Variable

ลำดับ	ชื่อตัวแปร	ความหมาย
1	ID	รหัสลูกค้า
2	Year_Birth	ปีเกิดของลูกค้า
3	Education	ระดับการศึกษาของลูกค้า
4	Marital	สถานภาพสมรสของลูกค้า
5	Income	รายได้ประจำปีของครัวเรือนของลูกค้า
6	Kidhome	จำนวนเด็กในครัวเรือนของลูกค้า
7	Teenhome	จำนวนวัยรุ่นในครัวเรือนของลูกค้า
8	DtCustomer	วันที่ลูกค้าลงทะเบียนกับบริษัท
9	Recency	จำนวนวันตั้งแต่การซื้อของครั้งสุดท้าย
10	MntWines	จำนวนเงินที่ใช้ในการซื้อไวน์ในช่วง 2 ปีที่ผ่านมา
11	MntFruits	จำนวนเงินที่ใช้ในการซื้อผลไม้ในช่วง 2 ปีที่ผ่านมา
12	MntMeatProducts	จำนวนเงินที่ใช้ในการซื้อผลิตภัณฑ์จากเนื้อสัตว์ในช่วง 2 ปีที่ผ่านมา
13	MntFishProducts	จำนวนเงินที่ใช้ในการซื้อผลิตภัณฑ์จากปลาในช่วง 2 ปีที่ผ่านมา
14	MntSweetProducts	จำนวนเงินที่ใช้ในการซื้อผลิตภัณฑ์ของหวานในช่วง 2 ปีที่ผ่านมา
15	MntGoldProds	จำนวนเงินที่ใช้ในการซื้อผลิตภัณฑ์ทองในช่วง 2 ปีที่ผ่านมา
16	Response	1 หากลูกค้ายอมรับข้อเสนอในแคมเปญล่าสุด, 0 หากไม่ยอมรับ



Process



Understanding

Data.info()

Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	ID	2240 non-null	int64
1	Year_Birth	2240 non-null	int64
2	Education	2240 non-null	object
3	Marital_Status	2240 non-null	object
4	Income	2240 non-null	int64
5	Kidhome	2240 non-null	int64
6	Teenhome	2240 non-null	int64
7	Dt_Customer	2240 non-null	object
8	Recency	2240 non-null	int64
9	MntWines	2240 non-null	int64
10	MntFruits	2240 non-null	int64
11	MntMeatProducts	2240 non-null	int64
12	MntFishProducts	2240 non-null	int64
13	MntSweetProducts	2240 non-null	int64
14	MntGoldProds	2240 non-null	int64
15	Response	2216 non-null	float64

dtypes: float64(1), int64(12), object(3)

data.isna().sum()

ID	0
Year_Birth	0
Education	0
Marital_Status	0
Income	0
Kidhome	0
Teenhome	0
Dt_Customer	0
Recency	0
MntWines	0
MntFruits	0
MntMeatProducts	0
MntFishProducts	0
MntSweetProducts	0
MntGoldProds	0
Response	24

dtype: int64

Data Cleaning

ปรับรูปแบบของคอลัมน์วันเกิด

0	4/9/2012
1	8/3/2014
2	21/8/2013
3	10/2/2014
4	19/1/2014
...	
2235	13/6/2013
2236	10/6/2014
2237	25/1/2014
2238	24/1/2014
2239	15/10/2012

Before

0	2012-04-09
1	2014-08-03
2	2013-08-21
3	2014-10-02
4	2014-01-19
...	
2235	2013-06-13
2236	2014-10-06
2237	2014-01-25
2238	2014-01-24
2239	2012-10-15

After

กรอกค่าที่หายไปของข้อมูล

```
data['response'].bfill(inplace=True)
```

0	id	2240 non-null	int64
1	year_birth	2240 non-null	int64
2	education	2240 non-null	object
3	marital_status	2240 non-null	object
4	income	2240 non-null	int64
5	kidhome	2240 non-null	int64
6	teenhome	2240 non-null	int64
7	dt_customer	2240 non-null	datetime64[ns]
8	recency	2240 non-null	int64
9	mntwines	2240 non-null	int64
10	mntfruits	2240 non-null	int64
11	mntmeatproducts	2240 non-null	int64
12	mntfishproducts	2240 non-null	int64
13	mntsweetproducts	2240 non-null	int64
14	mntgoldprods	2240 non-null	int64
15	response	2240 non-null	float64

Data Cleaning

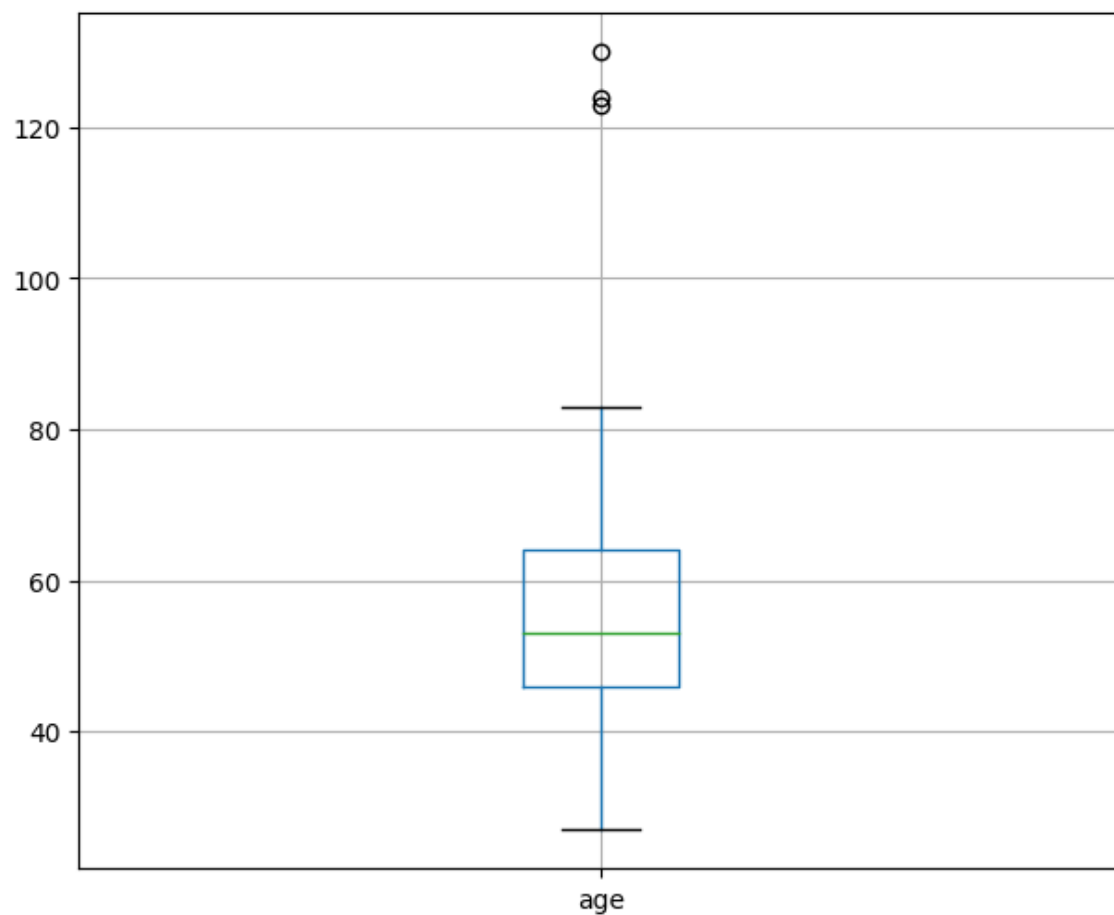
เพิ่ม Features ที่เหมาะสำหรับการวิเคราะห์ข้อมูล

- **อายุ** สามารถคำนวณได้จาก วันเกิดของลูกค้า เทียบกับ ปี 2023
- **ระดับรายได้** แบ่งออกเป็น 3 ระดับ คือ รายได้ต่ำ มีค่าน้อยกว่า 50,000 รายได้ปานกลาง มีค่าอยู่ระหว่าง 50,001 – 150,000 และ รายได้สูง มีค่ามากกว่า 150,000
- **จำนวนเด็กรวมในครัวเรือน** คำนวณจากจำนวนเด็กรวมกับจำนวน วัยรุ่นในครัวเรือน
- **ยอดซื้อรวม** เป็นผลรวมของจำนวนเงินในการซื้อสินค้าต่าง ๆ ได้แก่ ไวน์ ผลไม้ เนื้อสัตว์ ปลา ของหวาน และทองคำ

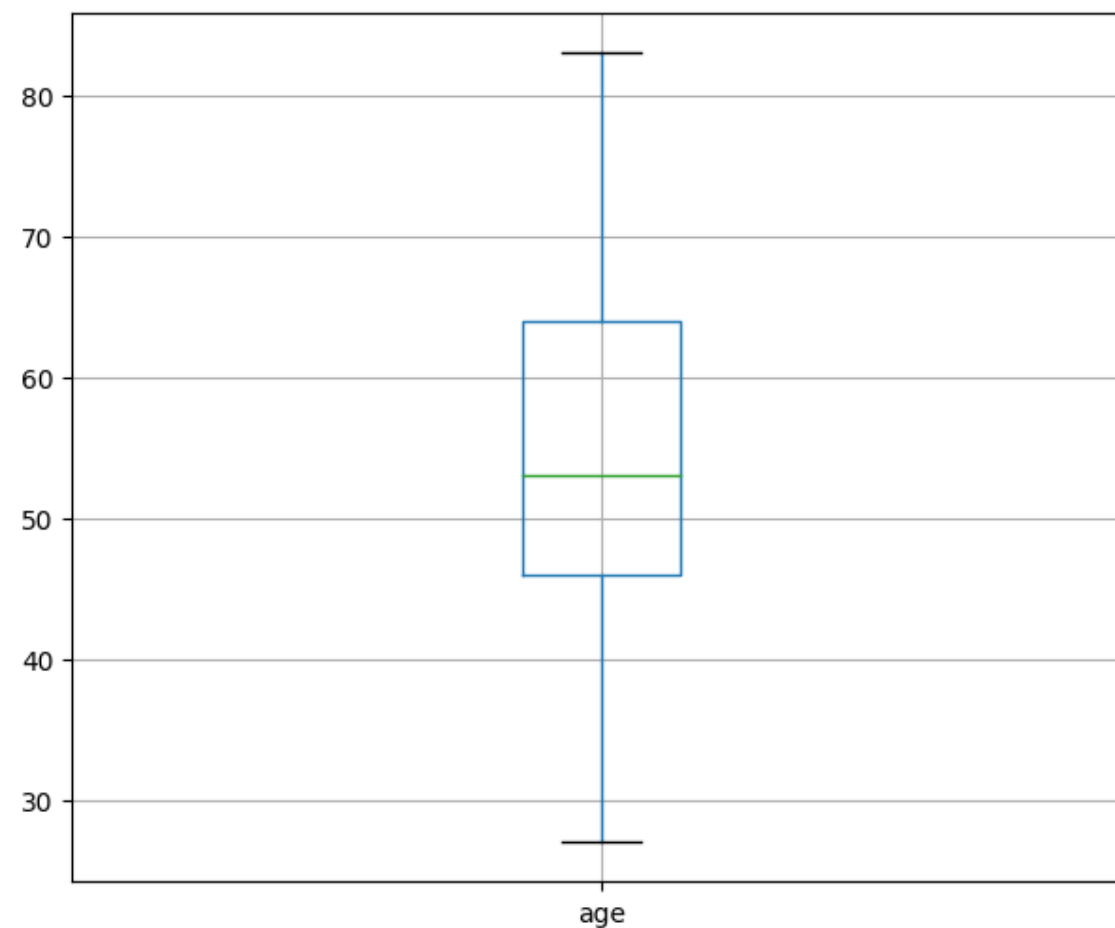
	age	income_level	total_children	total_purchase_amt
0	66	Medium	0	1617
1	69	Low	2	27
2	58	Medium	0	776
3	39	Low	1	53
4	42	Medium	1	422
...
2235	56	Medium	1	1341
2236	77	Medium	3	444
2237	42	Medium	0	1241
2238	67	Medium	1	843
2239	69	Medium	2	172

Distribution

Age outlier



Upper Limit : 91.0
Lower Limit : 19.0

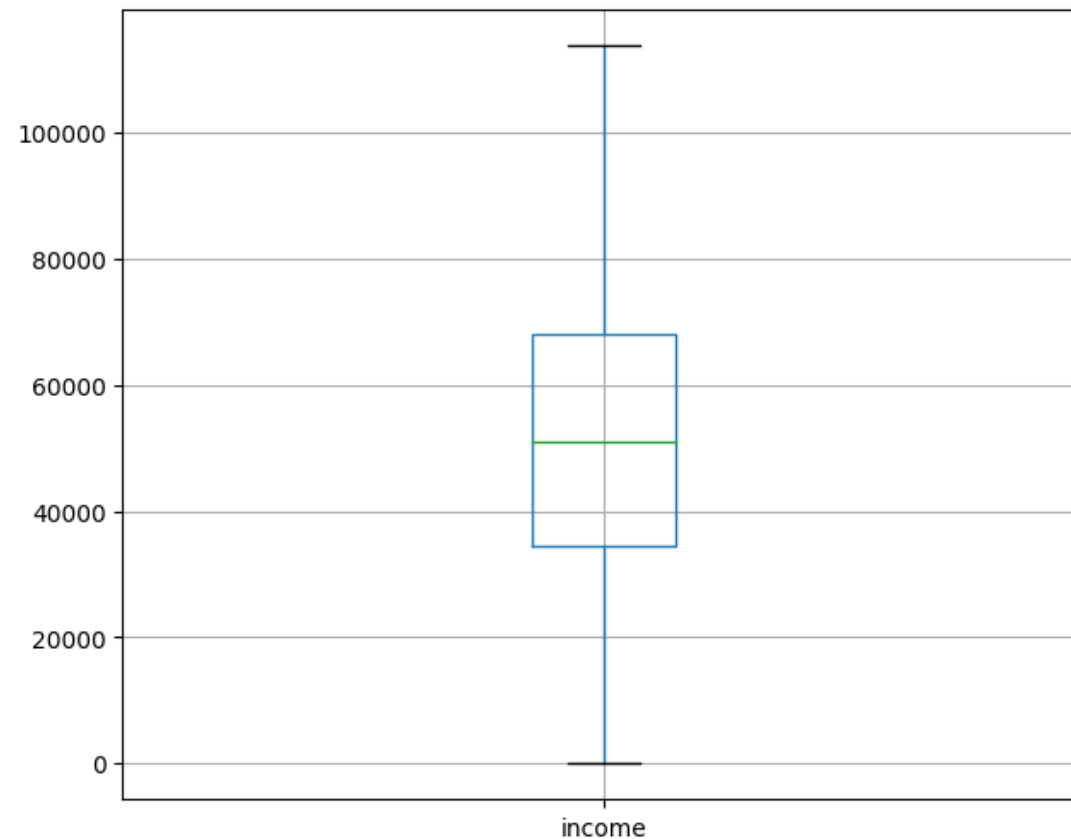
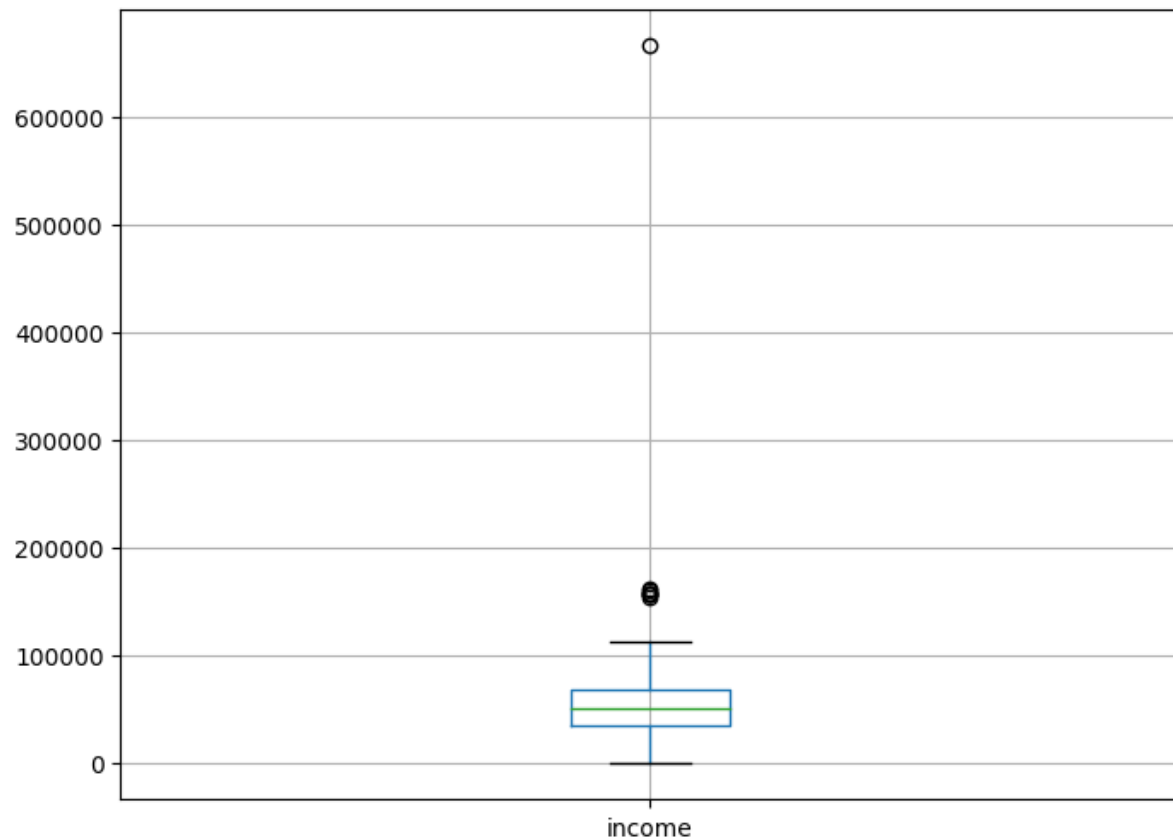


Distribution

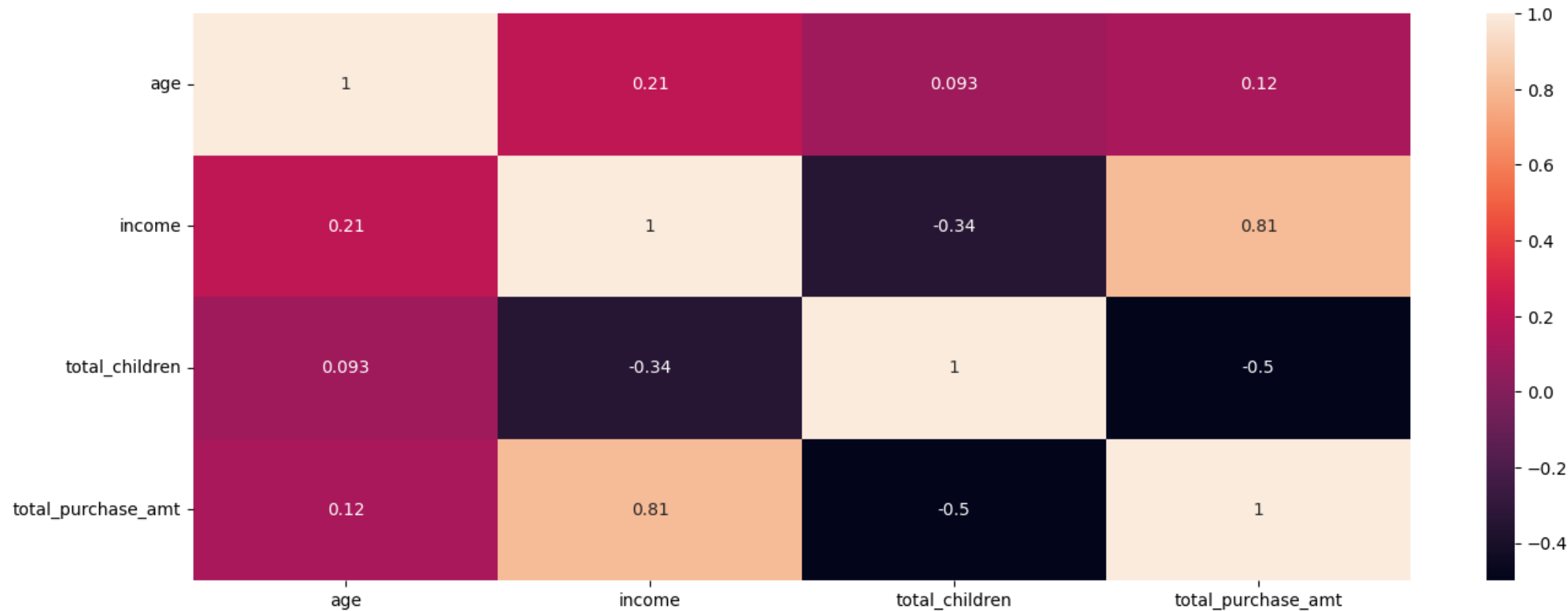
Income outlier

Upper Limit : 118,646

Lower Limit : -15,662



Correlation



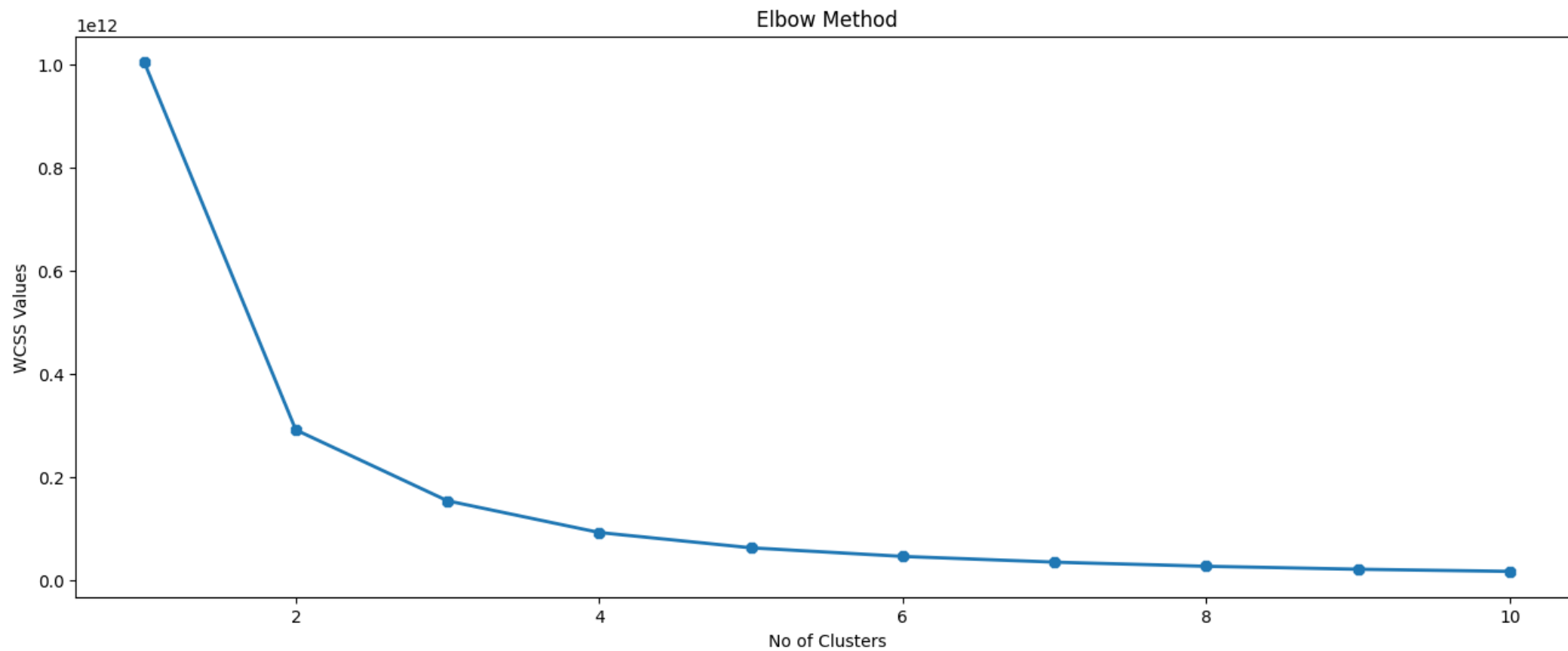
Chisquare Test of Association

Hypothesis ; $\alpha = 0.05$

- $H_0 =$ There is no association between total_purchase_amt and Income
- $H_1 =$ There is an association between total_purchase_amt and Income

Statistics : 1223420.6234825351 P-Value : 0.0
--

Model Development





Conclusion



Thank you