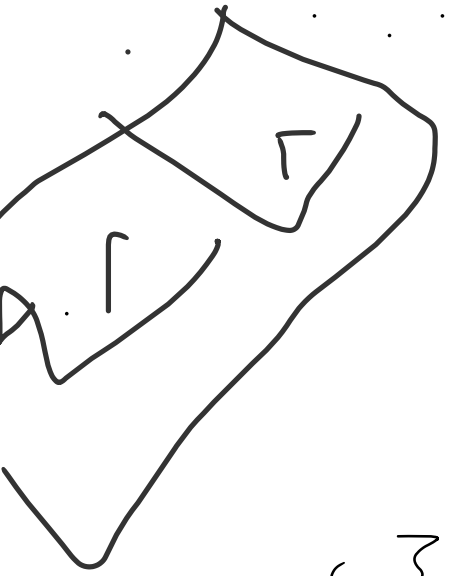
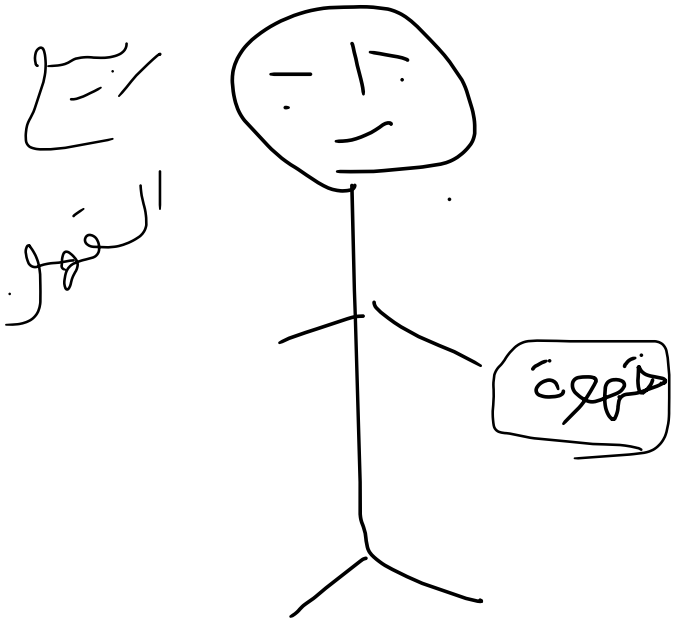
Data Analysis Project

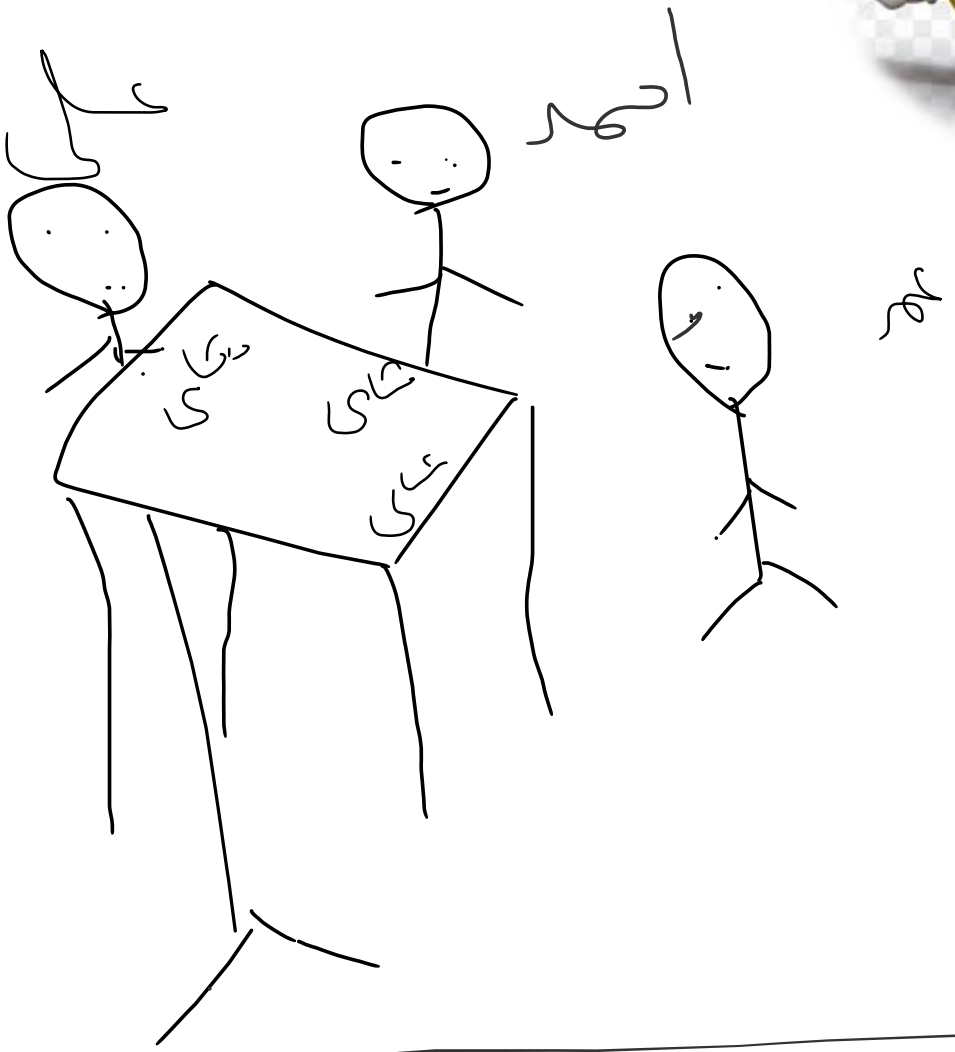
Final project of Engineer Mustafa Othman's course Data Analysis using Python

CAFE REWARDS

Funny story







Goal

The task was to perform as a Senior Marketing Analyst at Maven Cafe, who just run a test by sending different combinations of promotional offers to existing rewards members. Now that the 30-day period for the test has concluded, the task is

- to identify key customer segments
- and develop a data-driven strategy for future promotional messaging & targeting.

The Objective

My mission as Sr. Marketing Analyst was two-fold:

- Identify Key Customer Segments: Determine which customer groups were most (and least) engaged with the different promotional offers.
- Develop a Data-Driven Strategy: Use insights gathered from the test to recommend a strategic approach for future promotional messaging and offer targeting, ensuring increased customer engagement.

About the Data Set

The data consists of a table with 17'000 members

The analysis is based on data simulating customer behavior over a 30-day period, capturing transactions and interactions with different promotional offers. The dataset is split into three key tables:

- Offers table: Information on the different promotional offers, including type, difficulty, and rewards.
- Customers table: Age, gender, income, and membership details of the cafe’s rewards members.
- Events table: Capturing how customers engaged with offers—whether they received, viewed, or completed an offer, and whether they made a transaction related to it.
- For a transaction to be tied to an offer, it must occur at the same time the offer was "completed" by the customer. This nuanced behavior offers an important insight into how customers interacted with the promotions.

df1

customers

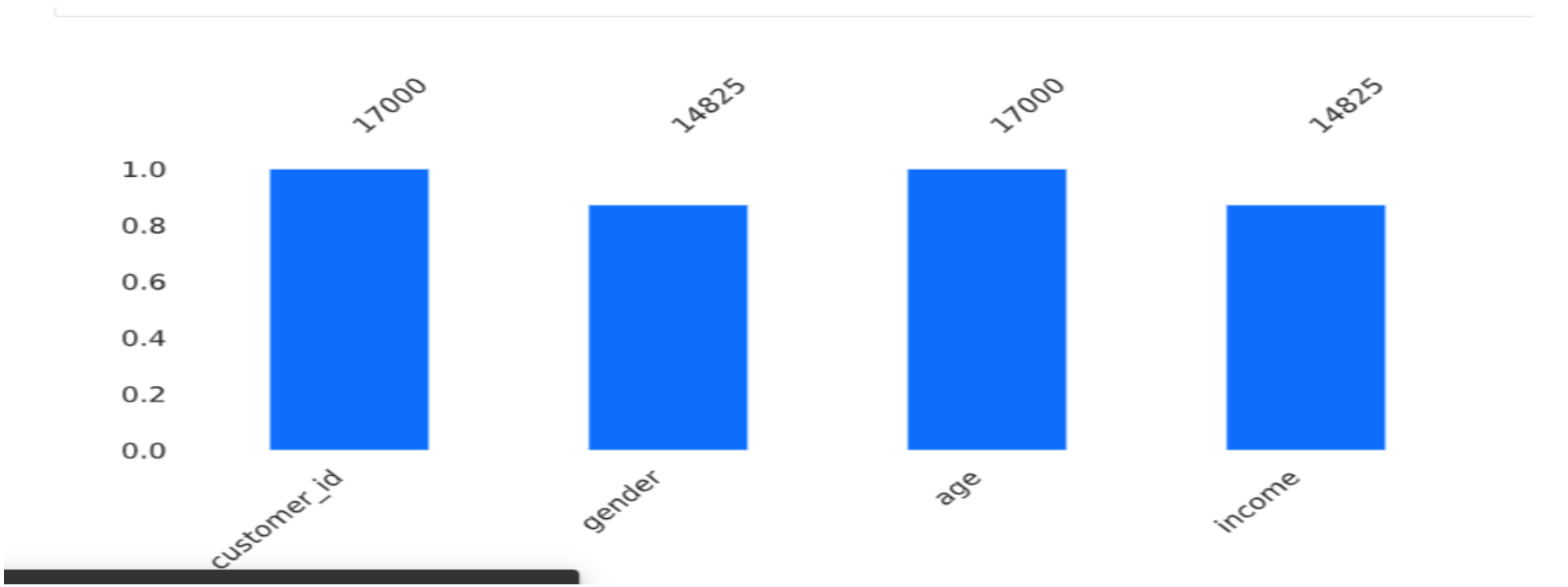
became_member_on	customer_id	gender	age	income
2017-02-12	68be06ca386d4c31939f3a4f0e3dd783	NaN	118	NaN
2017-07-15	0610b486422d4921ae7d2bf64640c50b	F	55	112000.0
2018-07-12	38fe809add3b4fcf9315a9694bb96ff5	NaN	118	NaN
2017-05-09	78afa995795e4d85b5d9ceeca43f5fef	F	75	100000.0
2017-08-04	a03223e636434f42ac4c3df47e8bac43	NaN	118	NaN

- The first thing I knew from the data consists of 5 columns and 17,000 rows

- After I called the data, the second thing I noticed was that I had a problem with the data and it was clear to everyone that in the income column there were empty cells.

- Thirdly, I made a quick report to explain everything about the data through a library.

YDATA_PROFILING



I knew that my data has a problem in the number of rows is not equal in all columns and in the difference from 14225 to 17000

Overview | Alerts 3 | Reproduction

Dataset statistics

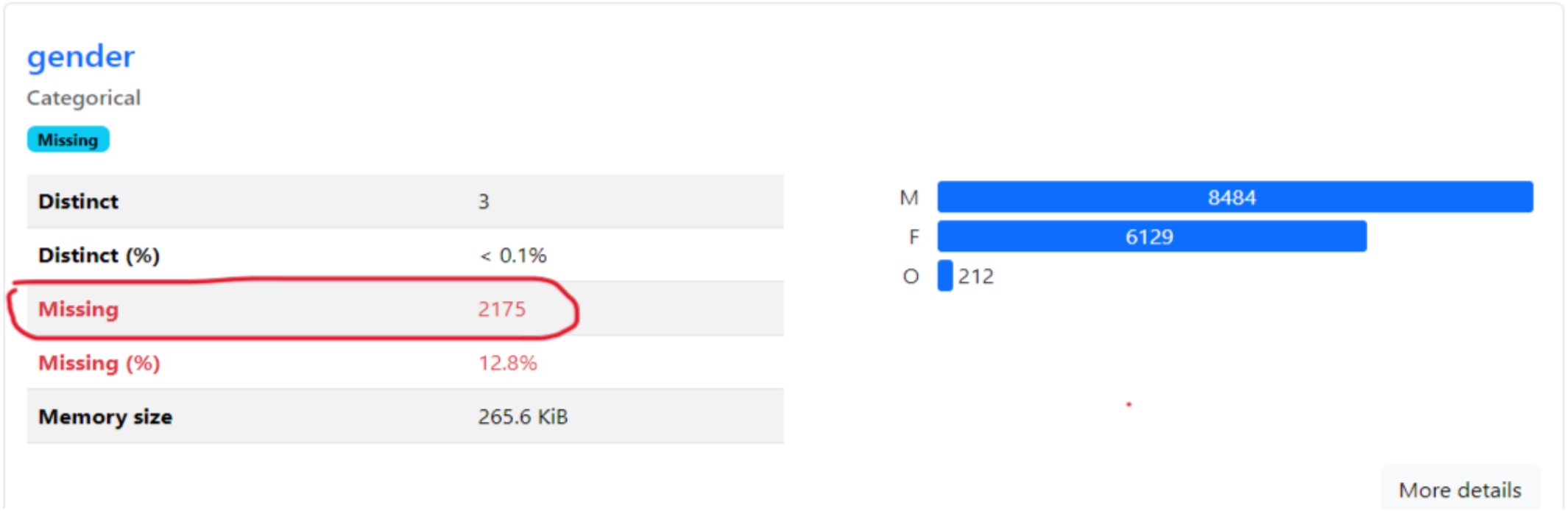
Number of variables	4
Number of observations	17000
Missing cells	4350
Missing cells (%)	6.4%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	664.1 KiB
Average record size in memory	40.0 B

Variable types

Text	1
Categorical	1
Numeric	2

I also knew that I have 4350 cells missing by 6.4%

In the gender column :



- I noticed that I have 2175 missing in the gender column

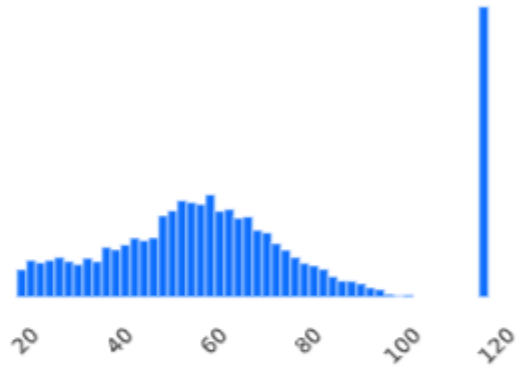
- and that there are more male customers than female customers

In the age column :

age

Real number (ℝ)

Distinct	85	**Minimum**	18
Distinct (%)	0.5%	**Maximum**	118
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	62.531412	**Memory size**	265.6 KiB



- I noticed that the ages of our customers range from 18 to 118
- I also learned that the average age of our customers is 62

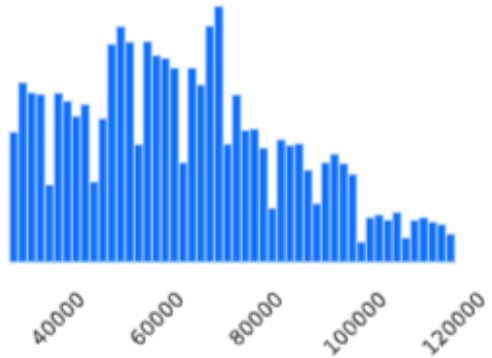
In the income column :

income

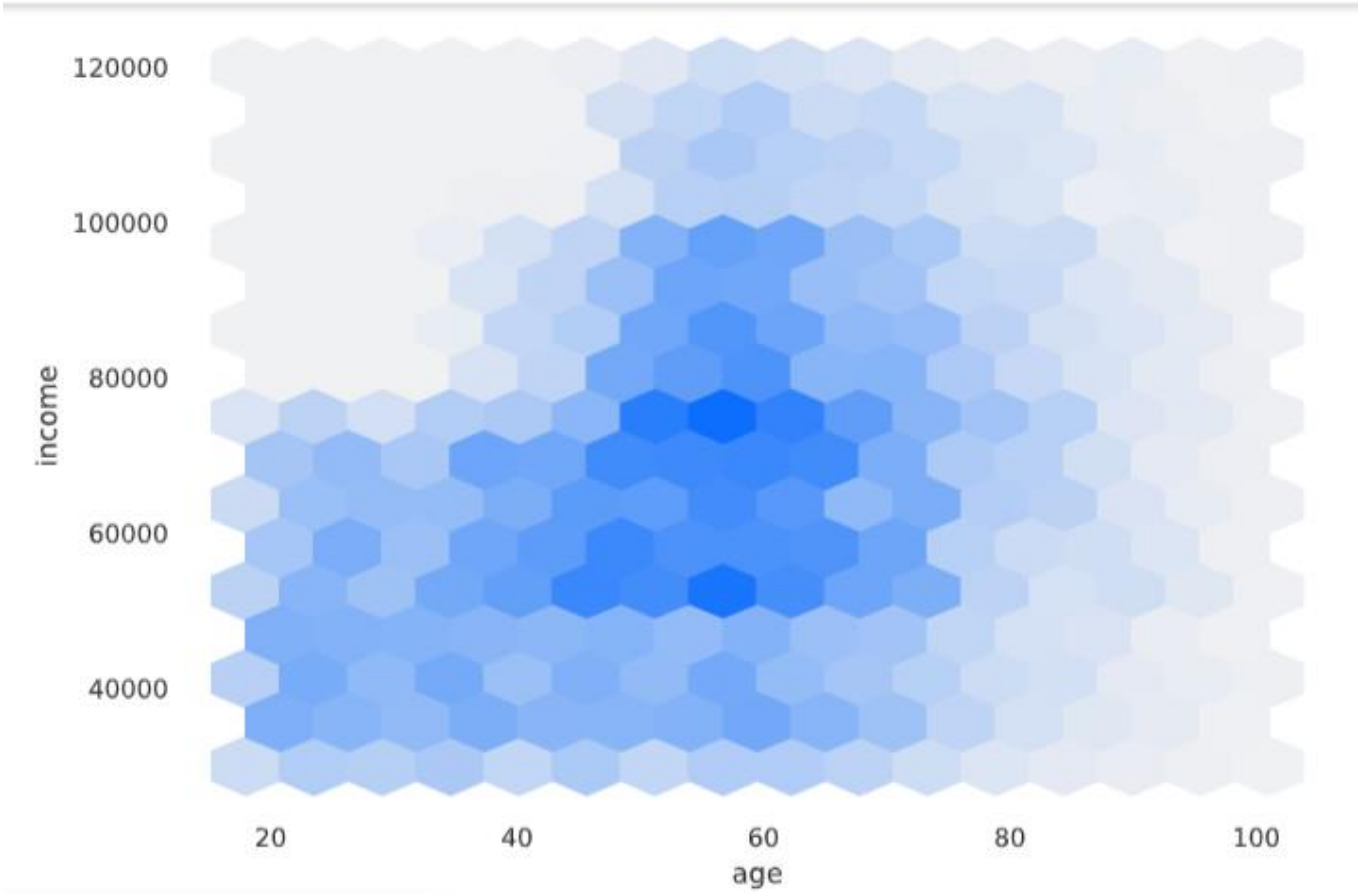
Real number (ℝ)

Missing

Distinct	91	**Minimum**	30000
Distinct (%)	0.6%	**Maximum**	120000
Missing	2175	**Zeros**	0
Missing (%)	12.8%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	65404.992	**Memory size**	265.6 KiB



- Here again there were missing 2175 cells
- and that the income of customers starts from 30,000 to 120,000 $ with an average of 65,404 $



This was a chart showing the relationship between income and age.

- We notice that the number of customers is concentrated at the age of 60 and the income is about 70,000.

df2

events

1	customer_id	event	value	time
2	78afa995795e4d85b5d9ceeca43f5fef	offer received	{'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}	
3	a03223e636434f42ac4c3df47e8bac43	offer received	{'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}	
4	e2127556f4f64592b11af22de27a7932	offer received	{'offer id': '2906b810c7d4411798c6938adc9daaa5'}	
5	8ec6ce2a7e7949b1bf142def7d0e0586	offer received	{'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}	
6	68617ca6246f4fbc85e91a2a49552598	offer received	{'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}	
7	389bc3fa690240e798340f5a15918d5c	offer received	{'offer id': 'f19421c1d4aa40978ebb69ca19b0e20d'}	
8	c4863c7985cf408faee930f111475da3	offer received	{'offer id': '2298d6c36e964ae4a3e7e9706d1fb8c2'}	
9	2eeac8d8feae4a8cad5a6af0499a211d	offer received	{'offer id': '3f207df678b143eea3cee63160fa8bed'}	
10	aa4862eba776480b8bb9c68455b8c2e1	offer received	{'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}	
11	31dda685af34476cad5bc968bdb01c53	offer received	{'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}	
12	744d603ef08c4f33af5a61c8c7628d1c	offer received	{'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}	

Dataset statistics

Number of variables	4
Number of observations	306534
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	396
Duplicate rows (%)	0.1%
Total size in memory	9.4 MiB
Average record size in memory	32.0 B

Variable types

Text	2
Categorical	1
Numeric	1

I made a quick report to explain everything about the data through a library … YDATA_PROFILING

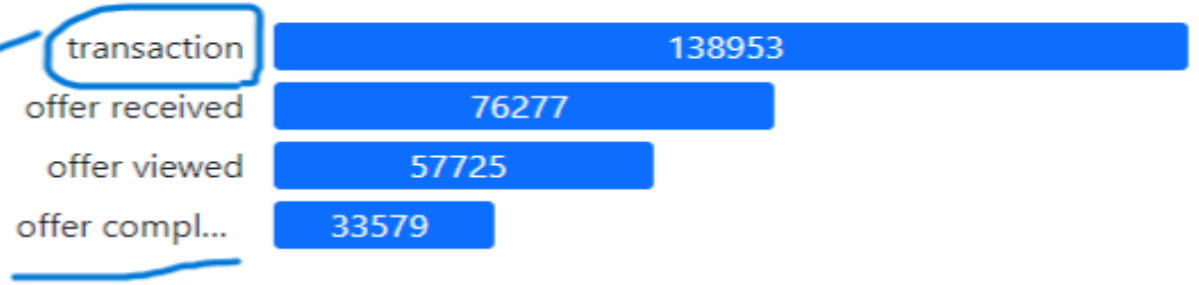
- **I discovered that the data consists of 306534 rows and 4 columns**
- **I noticed that I have 396 duplicate rows, which is 0.1%.**

In the event column :

event

Categorical

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	2.3 MiB



- I noticed that I have transaction 1389953 completed
- and 33579 offer completed only.

df3

offers

10*6

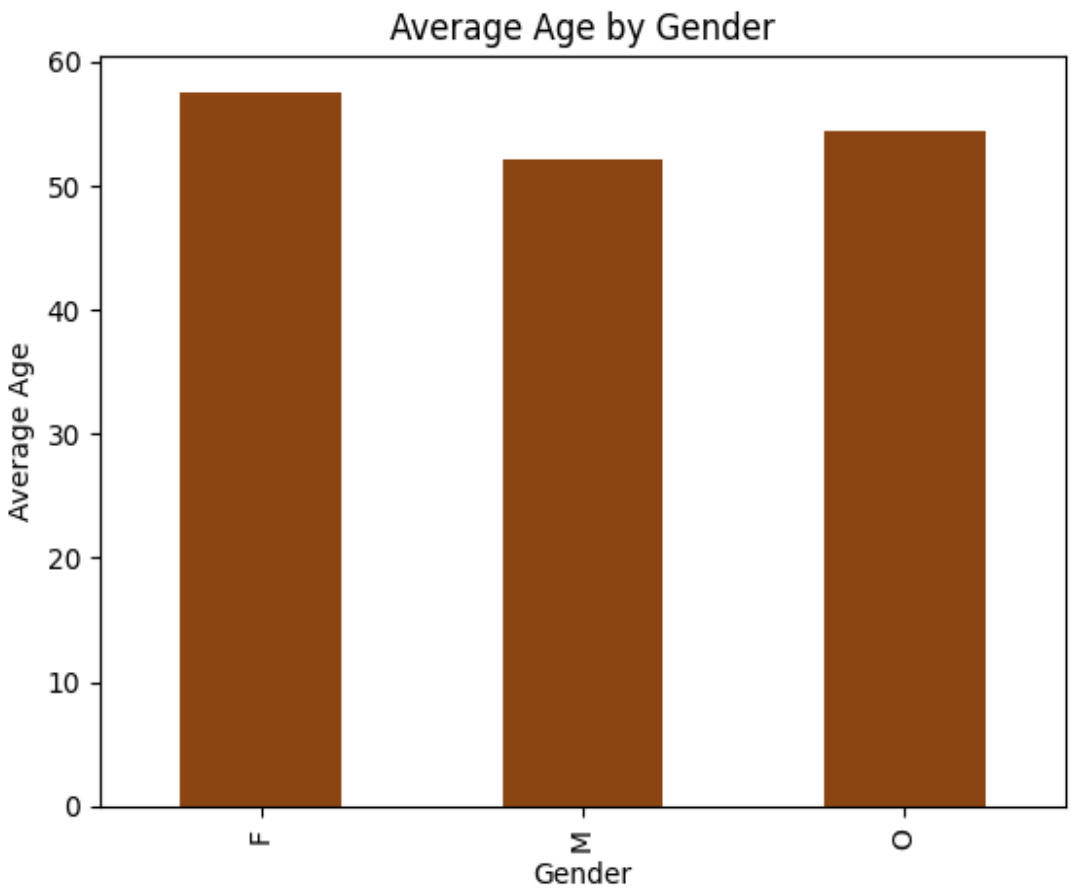
	offer_id	offer_type	difficulty	reward	duration	channels
2	ae264e3637204a6fb9bb56bc8210ddfd	bogo	10	10	7	['email', 'mobile', 'social']
3	4d5c57ea9a6940dd891ad53e9dbe8da0	bogo	10	10	5	['web', 'email', 'mobile', 'social']
4	3f207df678b143eea3cee63160fa8bed	informational	0	0	4	['web', 'email', 'mobile']
5	9b98b8c7a33c4b65b9aebfe6a799e6d9	bogo	5	5	7	['web', 'email', 'mobile']
6	0b1e1539f2cc45b7b9fa7c272da2e1d7	discount	20	5	10	['web', 'email']
7	2298d6c36e964ae4a3e7e9706d1fb8c2	discount	7	3	7	['web', 'email', 'mobile', 'social']
8	fafdcd668e3743c1bb461111dcafc2a4	discount	10	2	10	['web', 'email', 'mobile', 'social']
9	5a8bc65990b245e5a138643cd4eb9837	informational	0	0	3	['email', 'mobile', 'social']
10	f19421c1d4aa40978ebb69ca19b0e20d	bogo	5	5	5	['web', 'email', 'mobile', 'social']
11	2906b810c7d4411798c6938adc9daaa5	discount	10	2	7	['web', 'email', 'mobile']

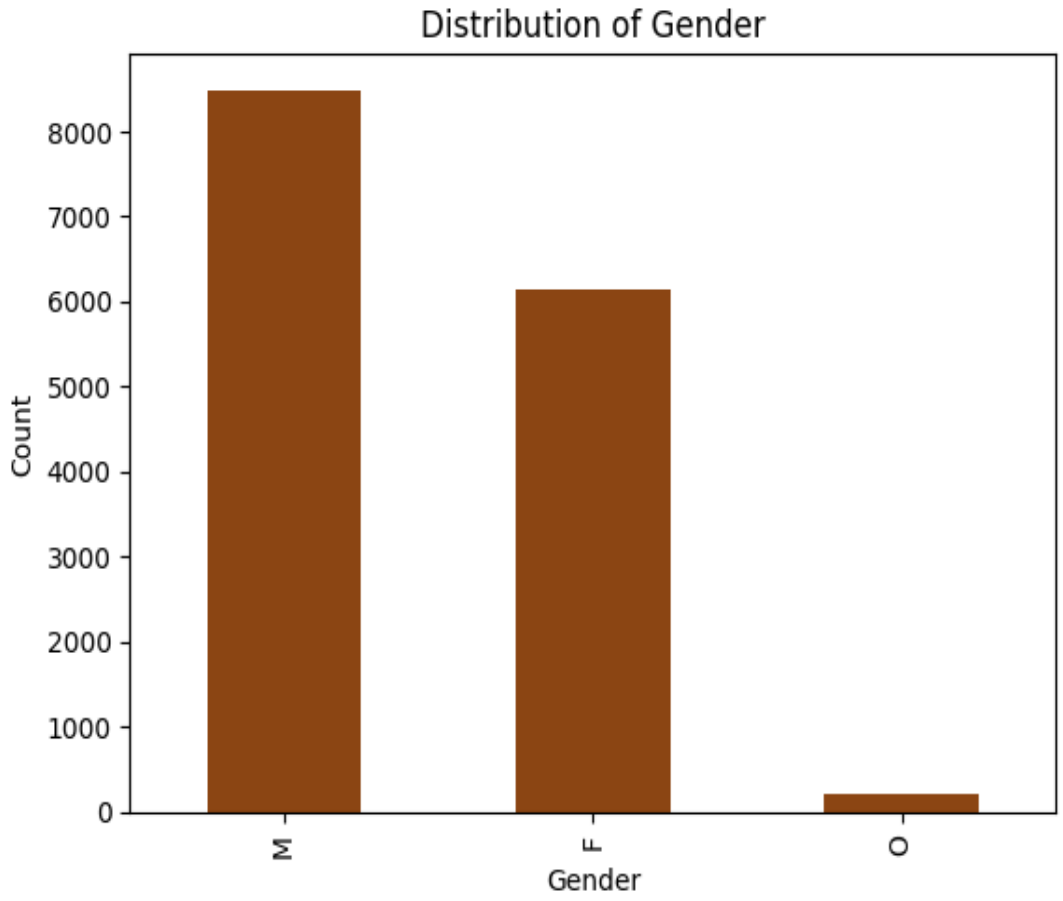
Daata Cleaning :

After solving the missing problem in df1 and duplicating in f2,

- df1 now consists of **5*14825**
- and df2 consists of **4*306137**

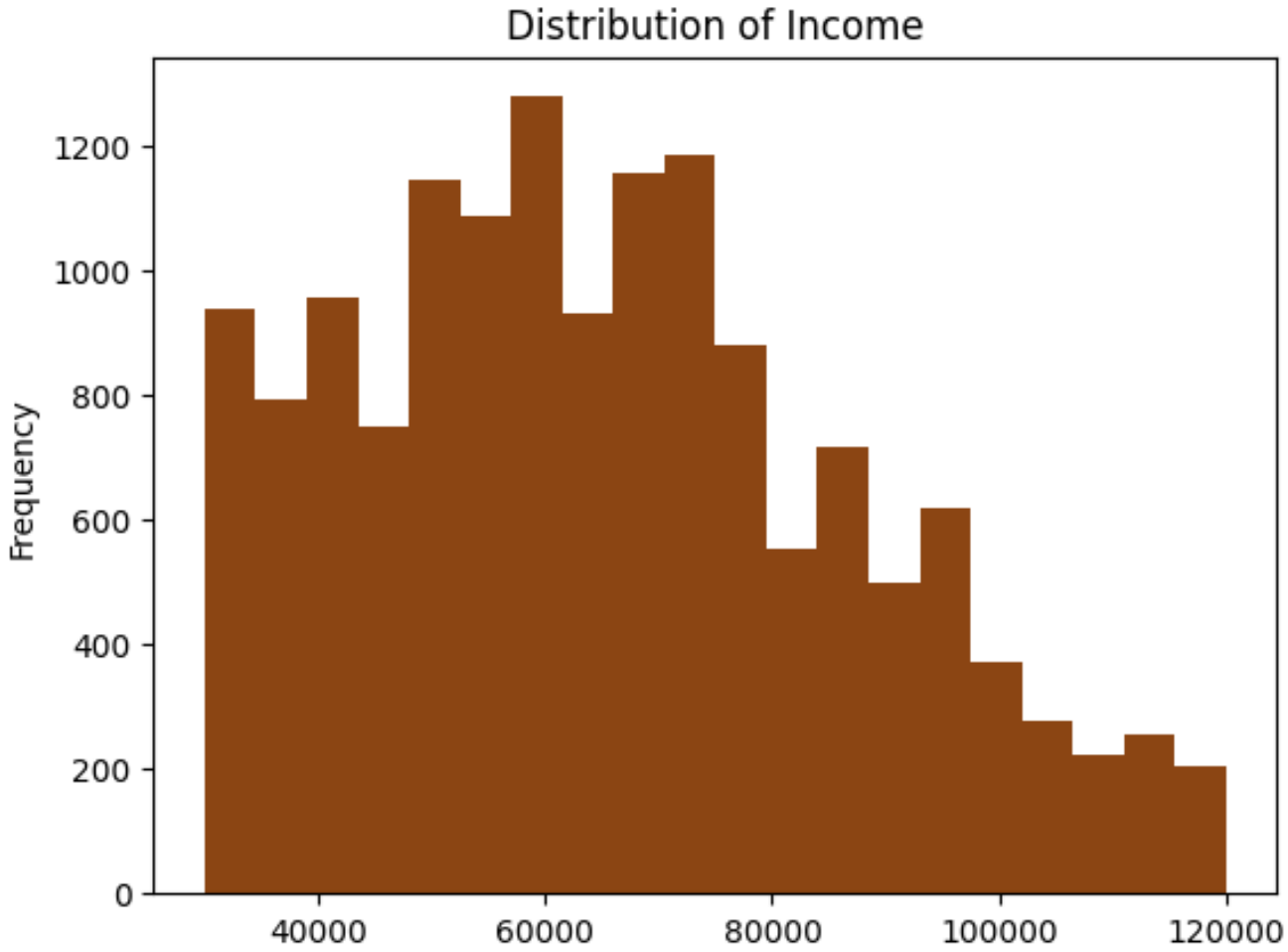
Exploratory Data Analysis

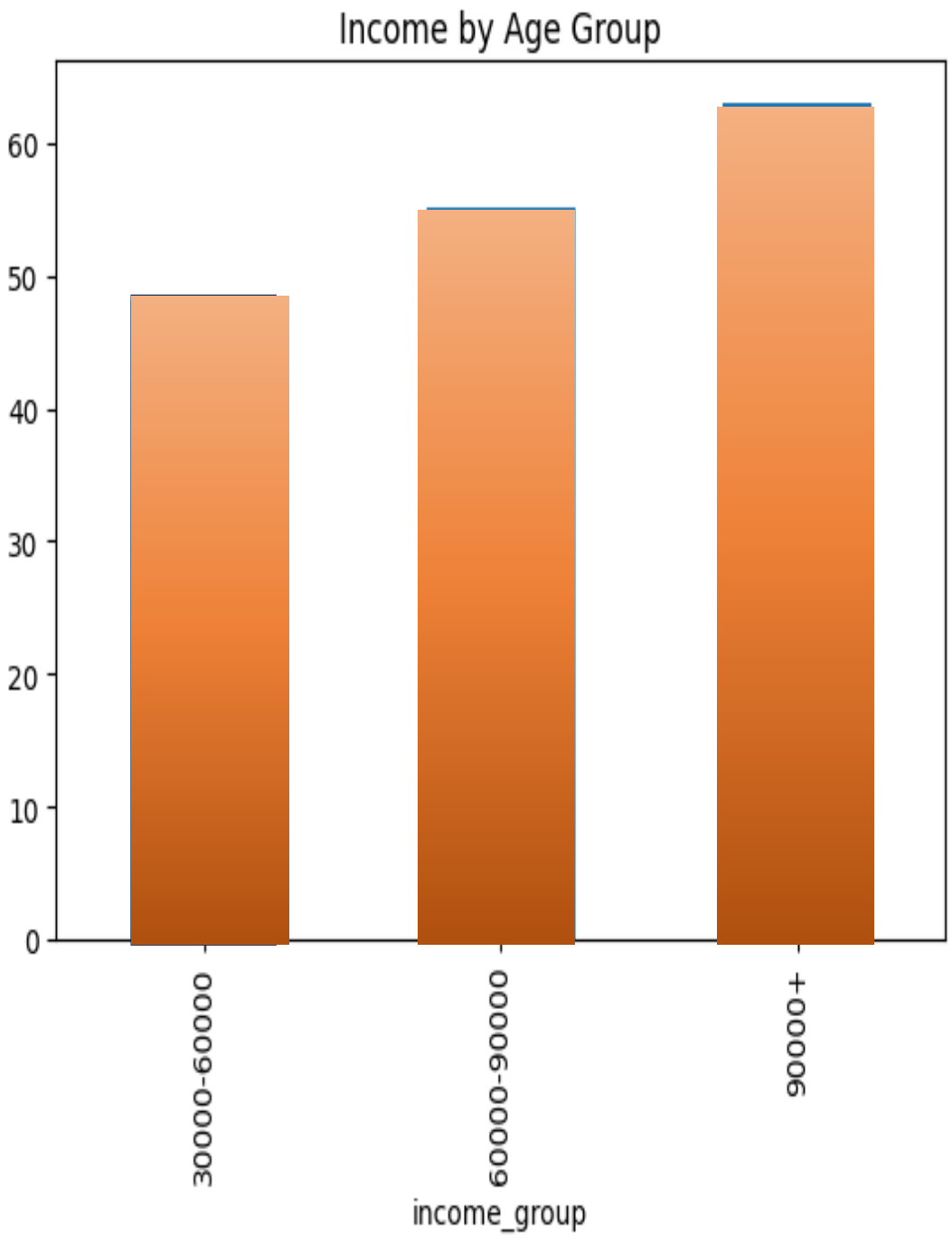




- It shows the average age of men and women and the slight difference for women, with the average age of women being **57.5** and the average age of men being **52.1**.

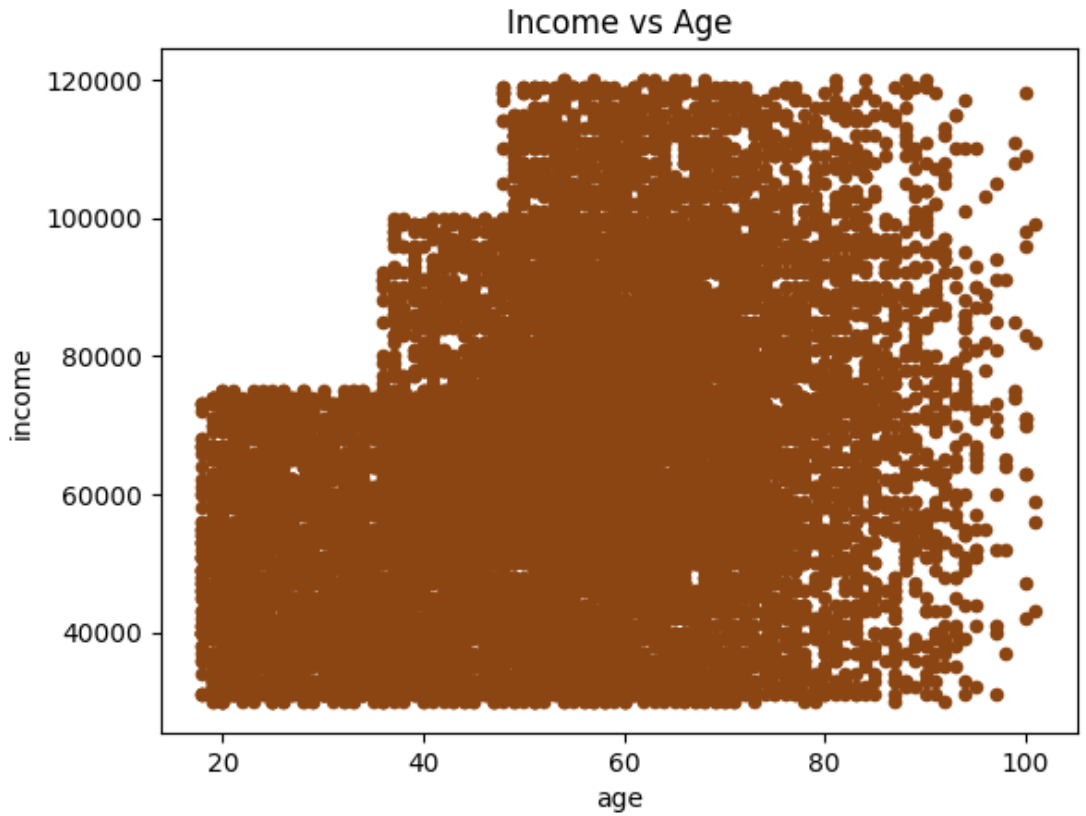
- It shows that the number of women is more than that men, as their number is **8484** and the number of men is **6129**.



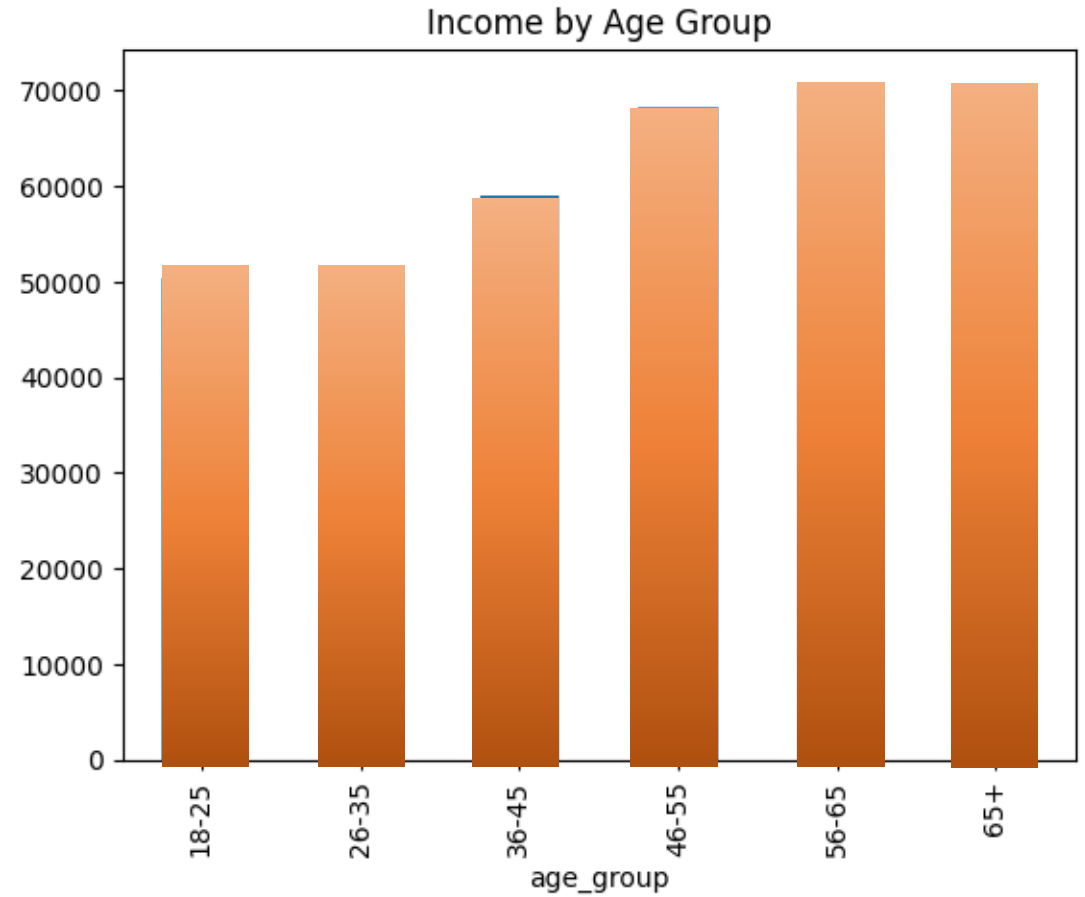


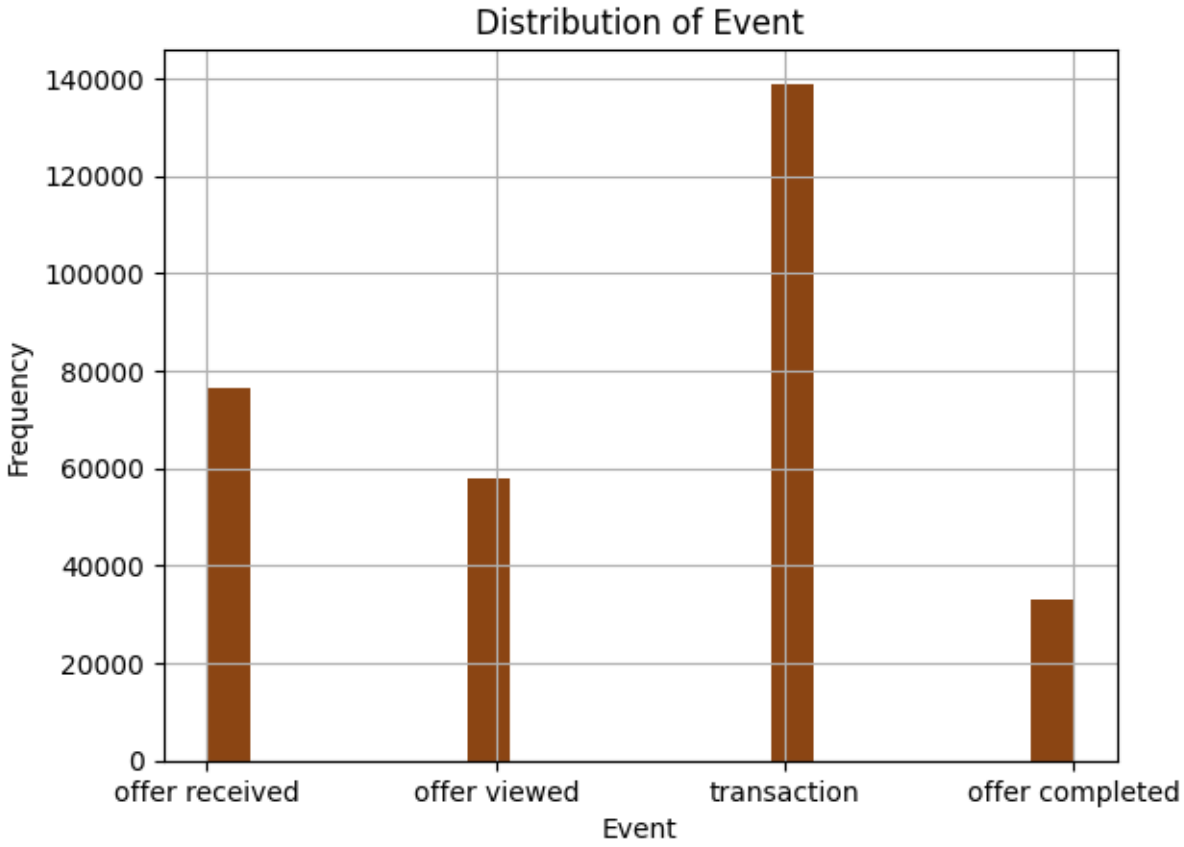
- It shows that the average income is **65404**

- You may also notice that the income of males is higher than that of females at older ages, and that the income of females is higher over a longer period.



- We notice that the largest number of customers is around **50-70** years old.



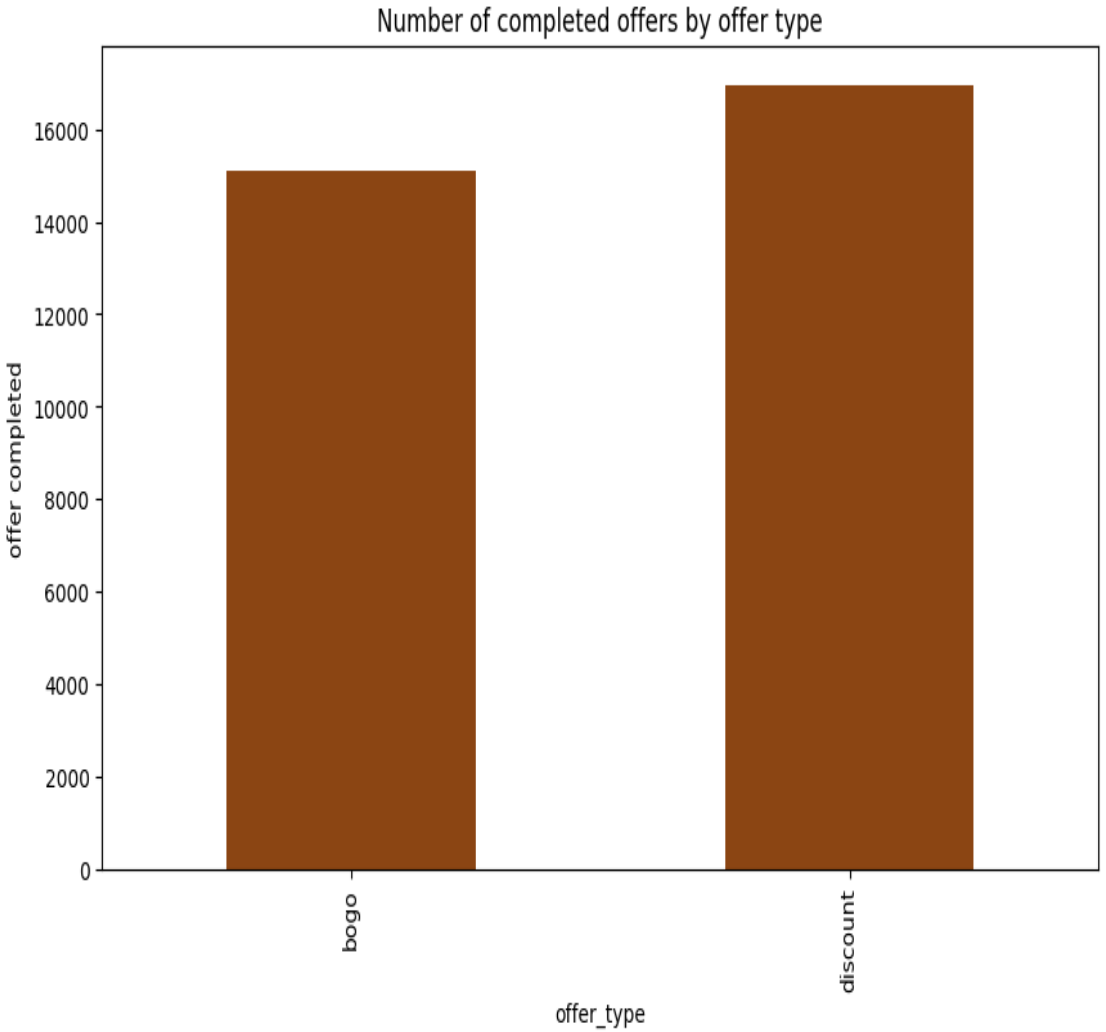


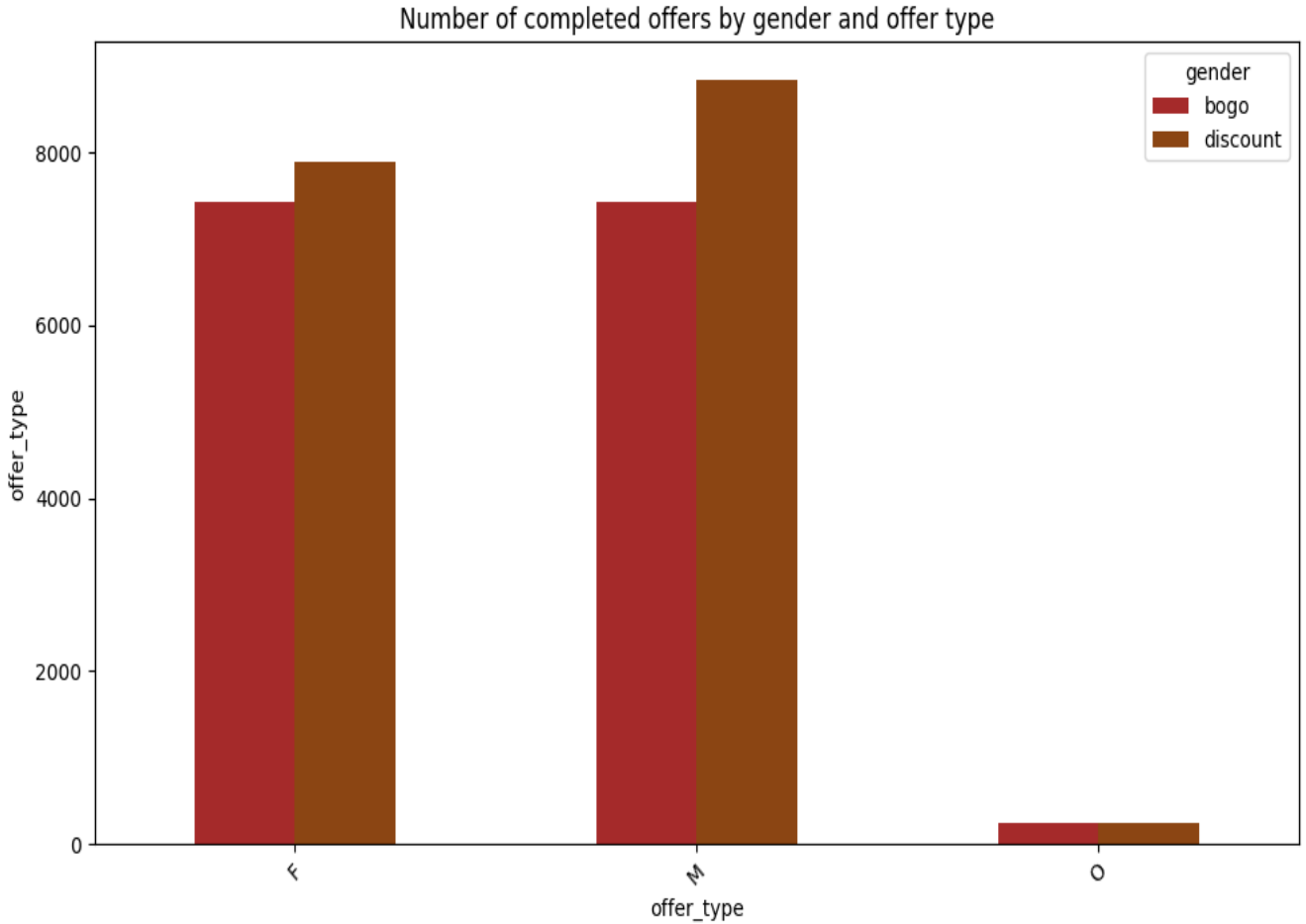
- 18-25 (Young Adults), 26-35 (Early Career Professionals), 36-45 (Young Families), 45655 (Mature Professionals), 56-65 (Pre-Retirement), 65+ (Retirees).

- transaction **138953**
- offer received **76277**
- offer viewed **57725**
- offer completed **33182**

Analytical Questions

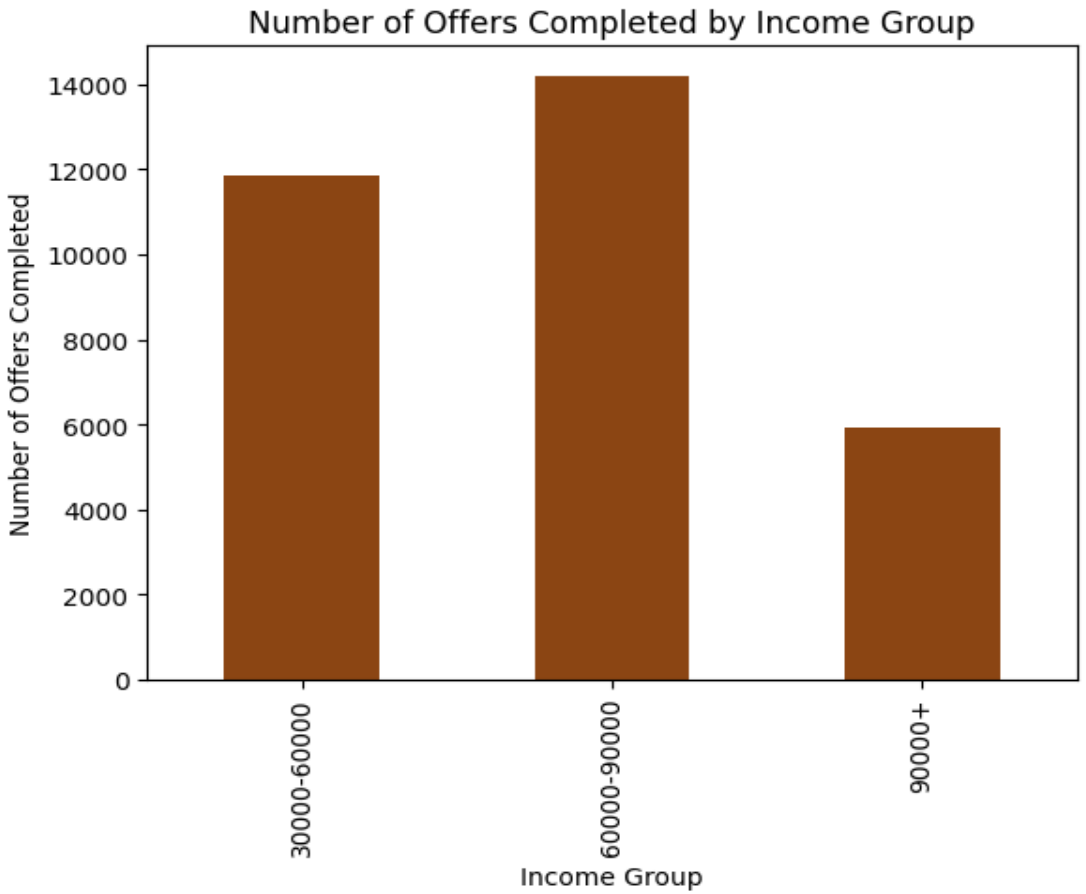
- What is the age distribution for each offer type?
- How does gender affect the success of offers (e.g., number of offers completed for each gender)?
- What are the most popular offers based on number of offers completed?
- Is there a relationship between income and the success of offers (offers completed)?
- What channels are most effective for promoting offers?

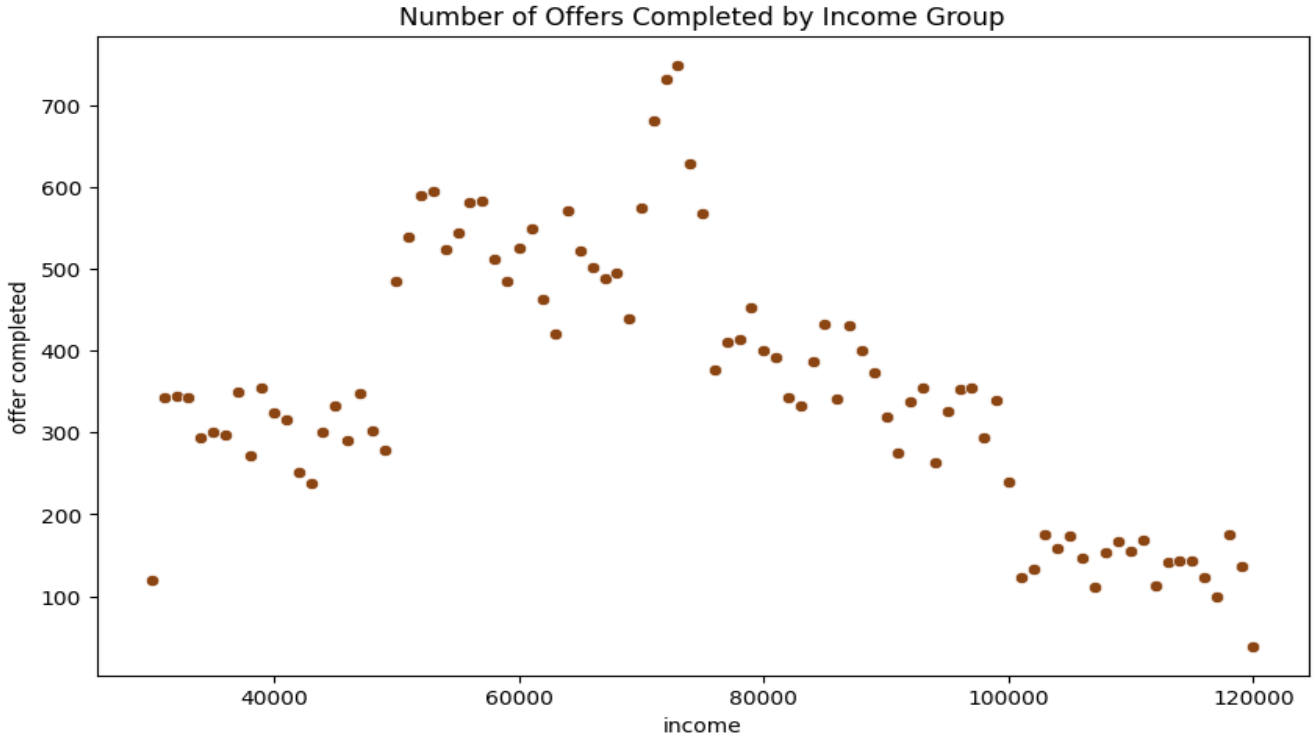




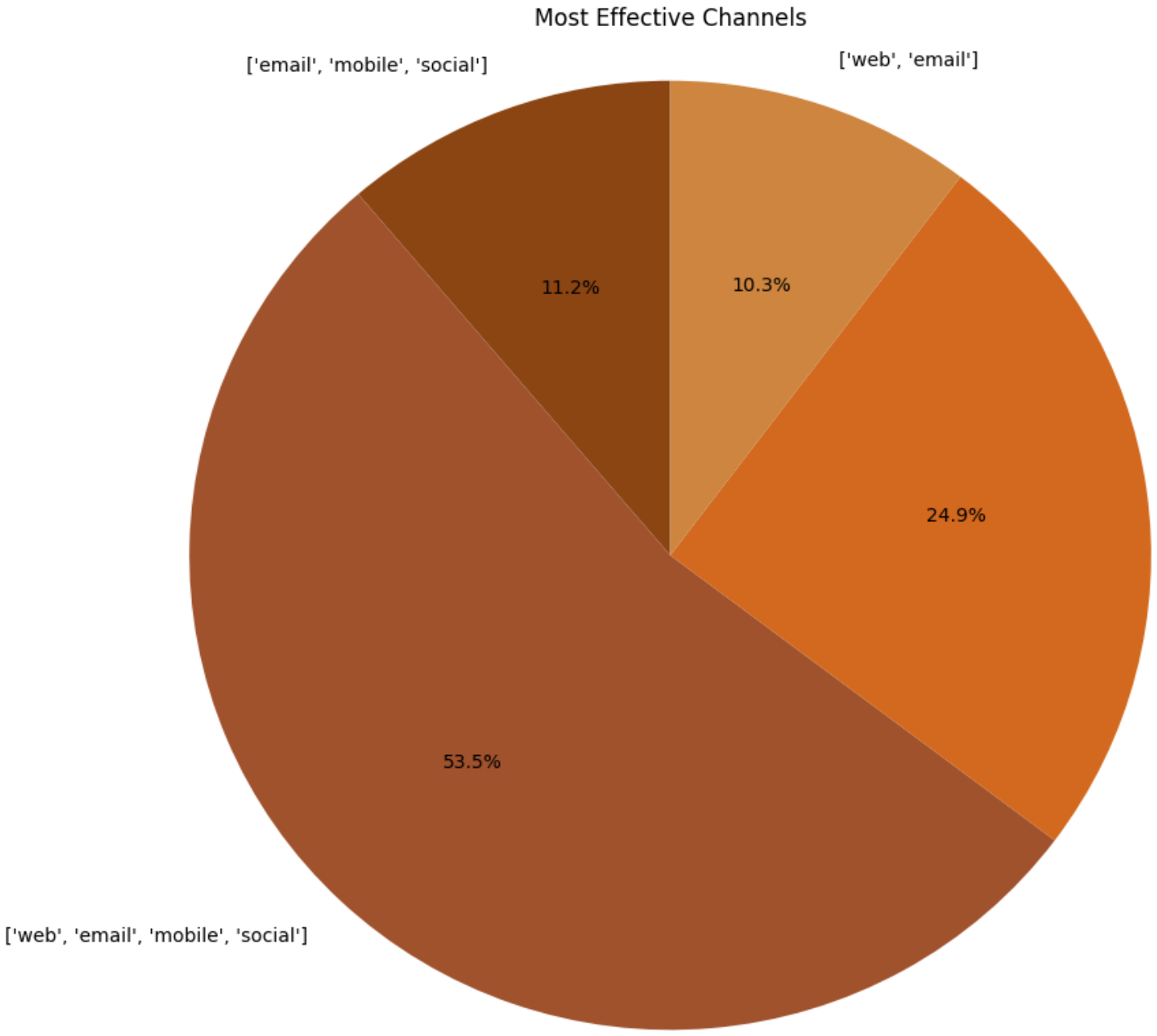
- Offer (bogo) used 15100
- Offer (discount) used 16970

- Male and Female used discount more than bogo.





- We note that those with a low income of 30,000-60,000 have a good number of completed offers **11,838**
- We also note that those with a high income of 90,000+ have a low number of completed offers **5,924**
- But the largest percentage of completed offers was in the medium income category of 60,000-90,000 and they have **14,188** completed offers



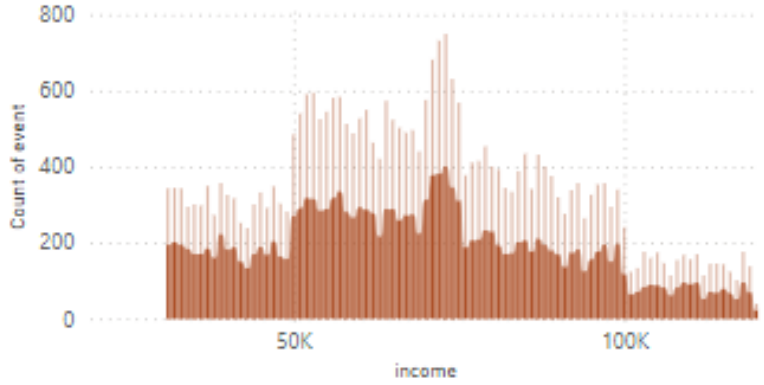
['email', 'mobile', 'social']	3604
['web', 'email', 'mobile', 'social']	17159
['web', 'email', 'mobile']	8001
['web', 'email']	3306


DASHBOARD

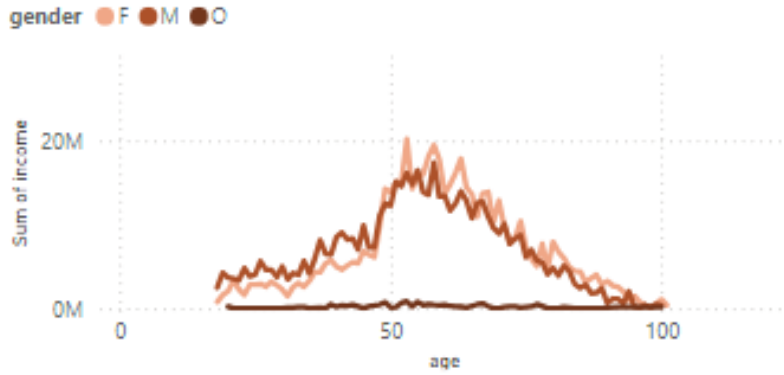
Promotions sent to Maven Cafe Customers

Sum Of offers completed
17.08K

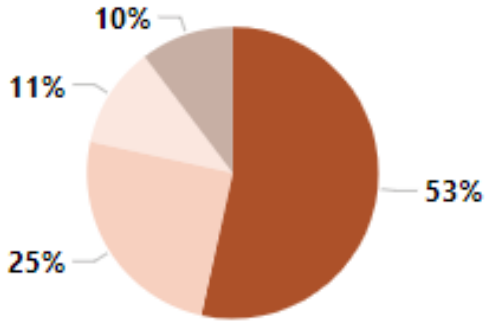
Count of Offers Completed by income



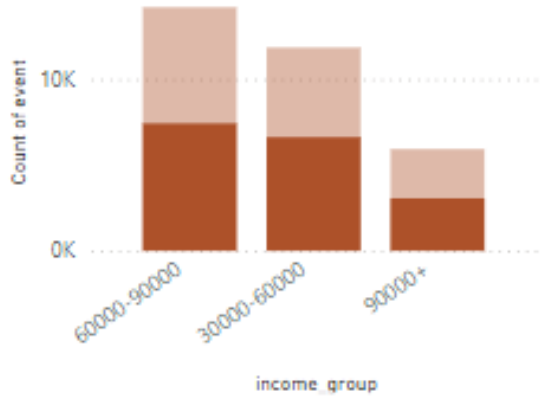
Sum of income by age and gender



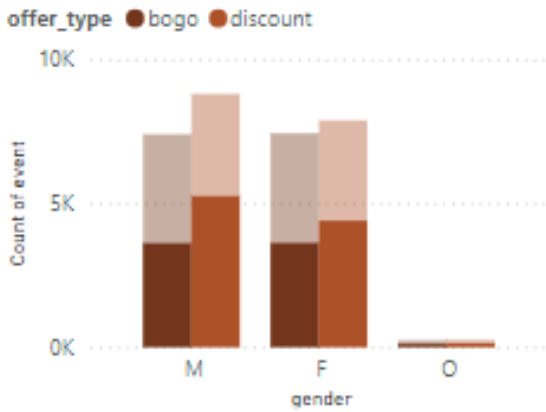
Count of Offers Completed by channels



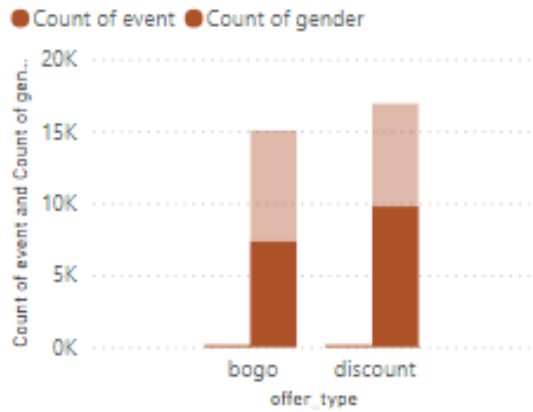
Count of Offers Completed by income_group



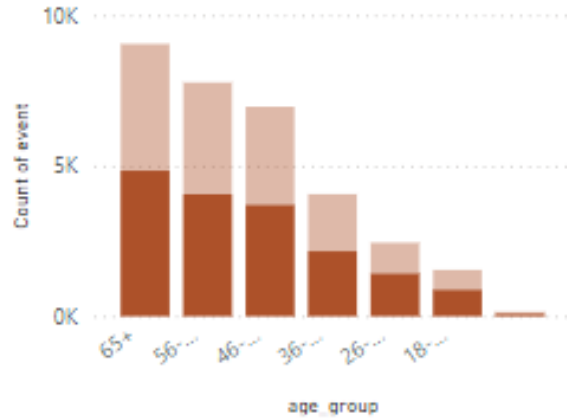
Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



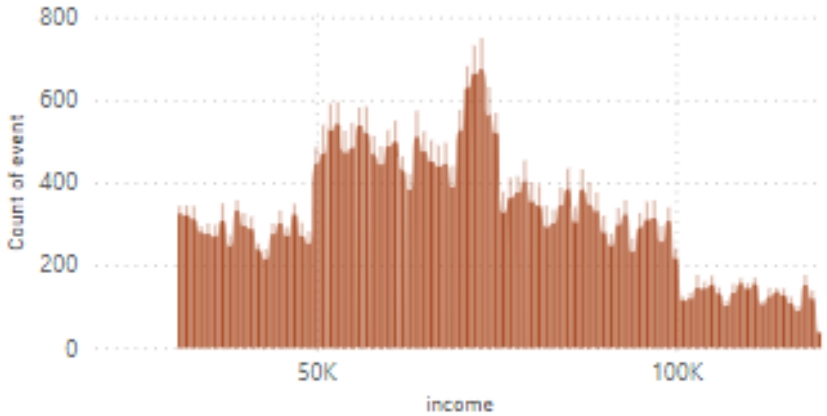
Count of offers completed by age_group



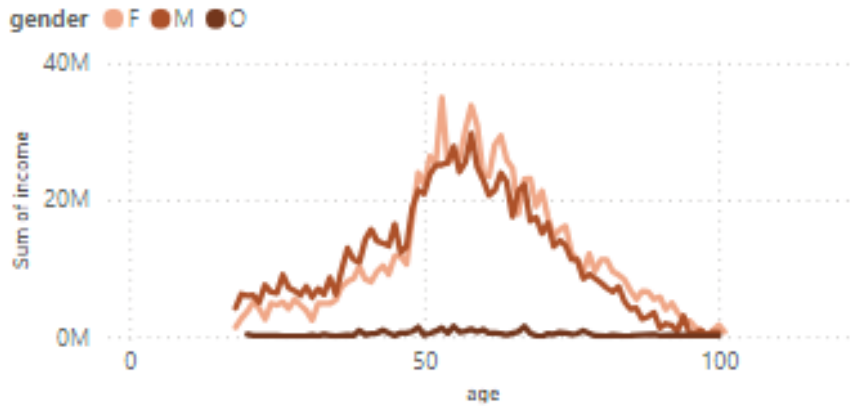
Promotions sent to Maven Cafe Customers

Sum Of offers completed
28.65K

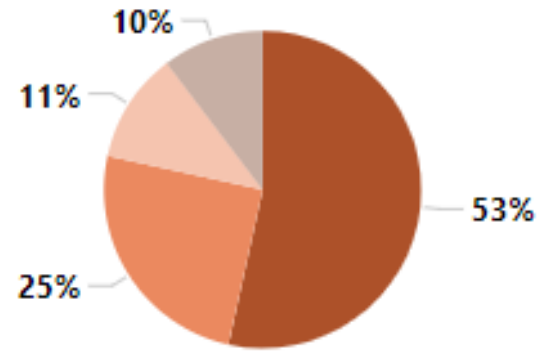
Count of Offers Completed by income



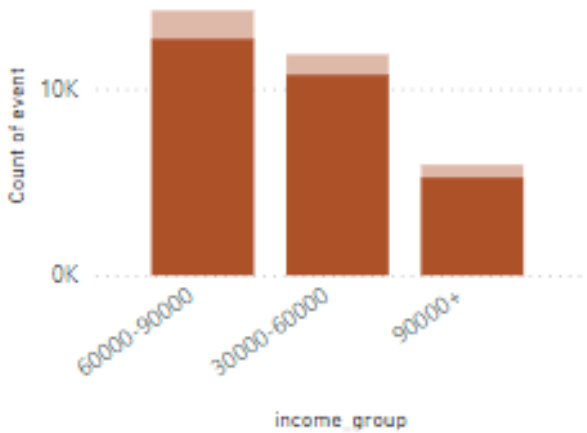
Sum of income by age and gender



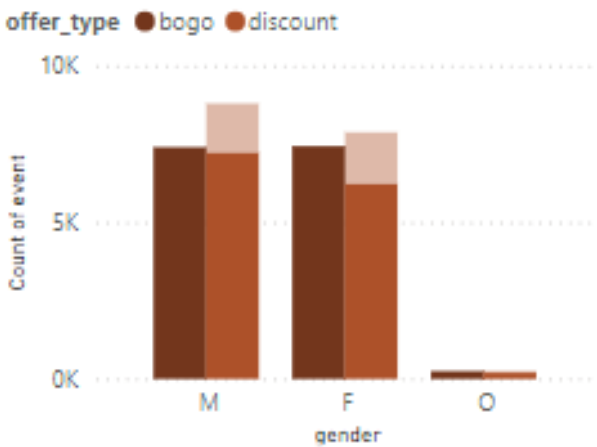
Count of Offers Completed by channels



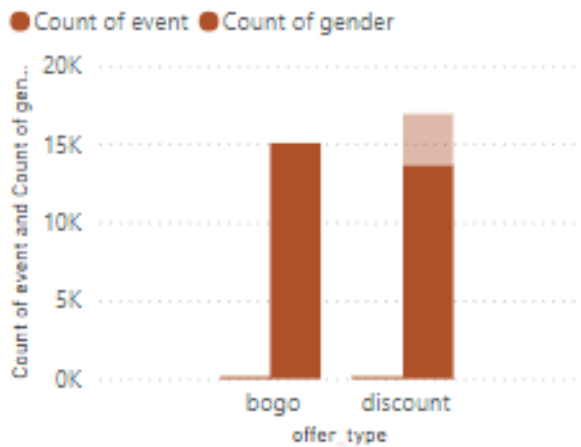
Count of Offers Completed by income_group



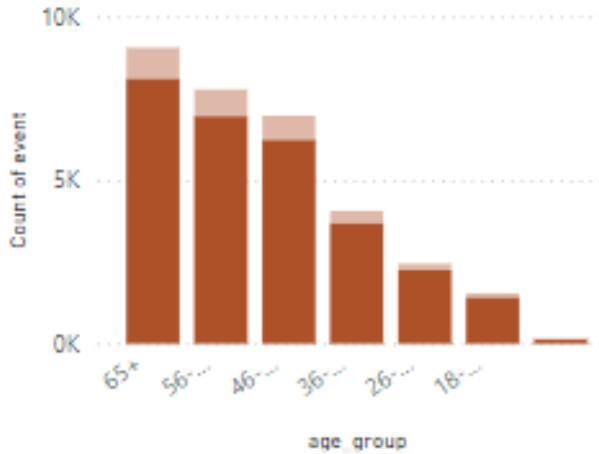
Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



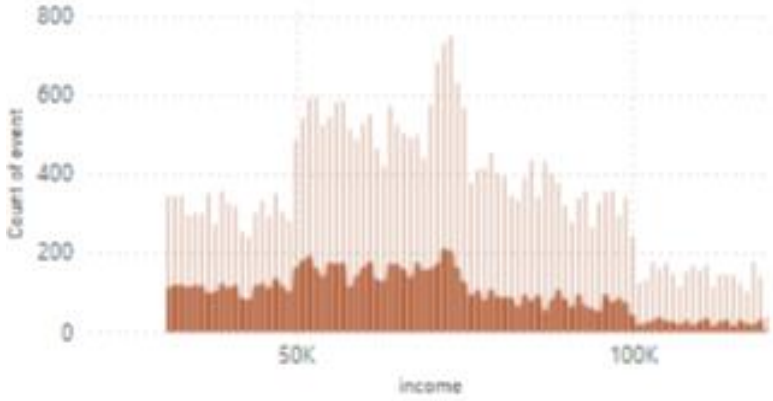
Count of offers completed by age_group



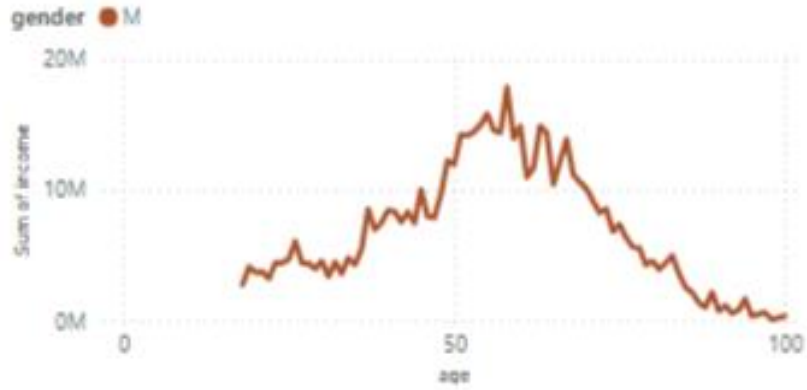
Promotions sent to Maven Cafe Customers

Sum Of offers completed
8788

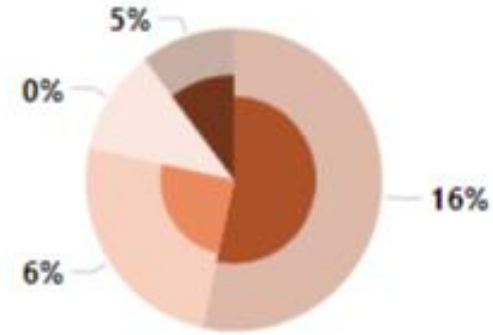
Count of Offers Completed by income



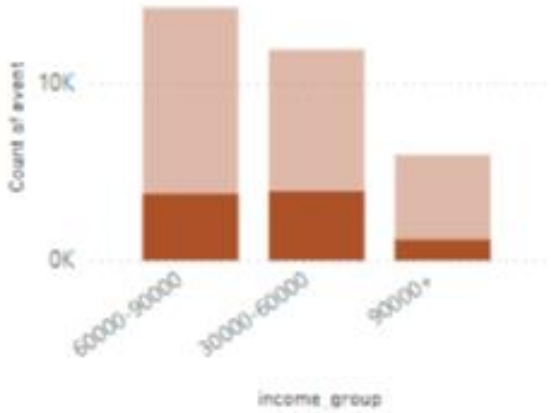
Sum of income by age and gender



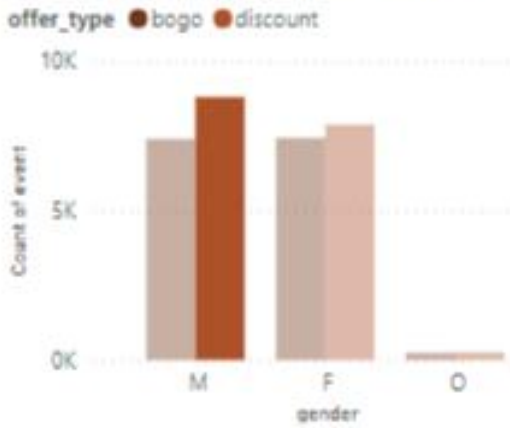
Count of Offers Completed by channels



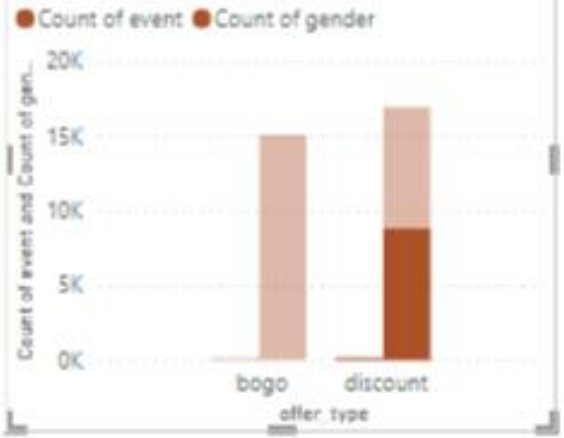
Count of Offers Completed by income_group



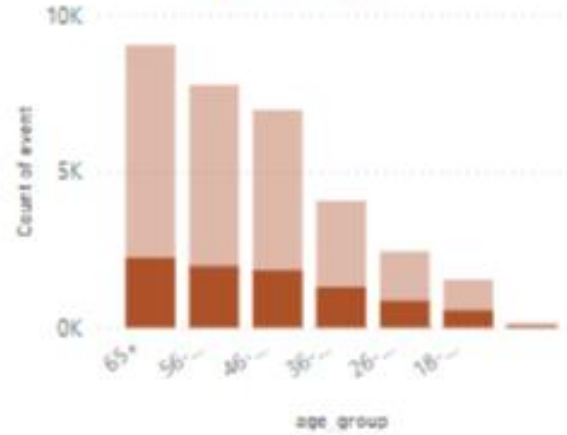
Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



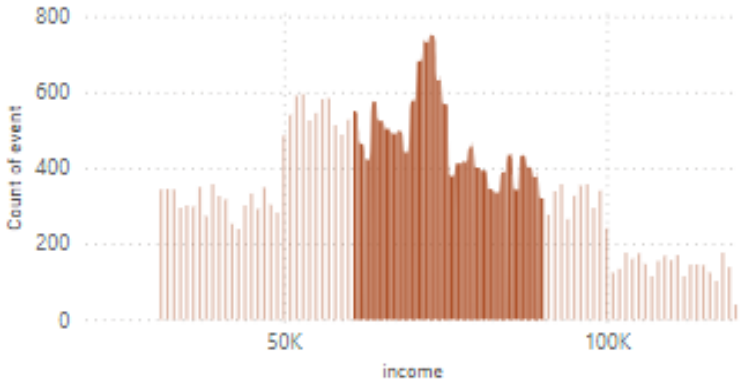
Count of offers completed by age_group



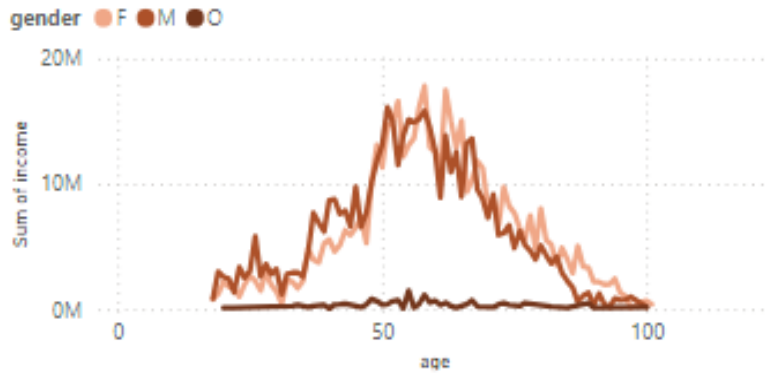
Promotions sent to Maven Cafe Customers

Sum Of offers completed
14.19K

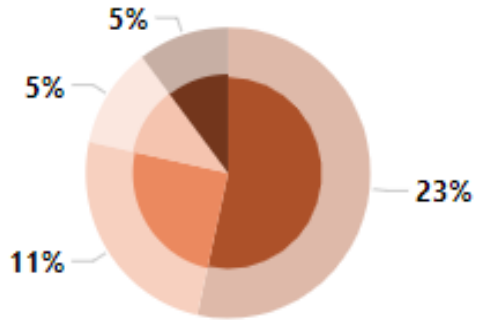
Count of Offers Completed by income



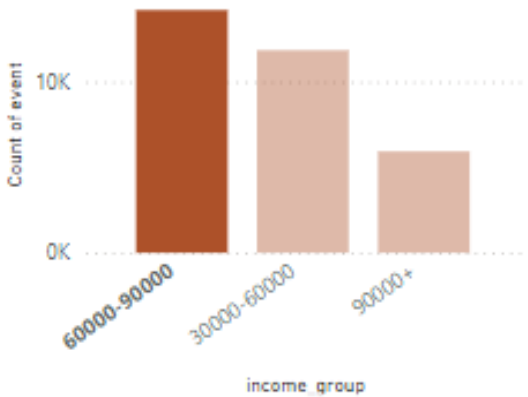
Sum of income by age and gender



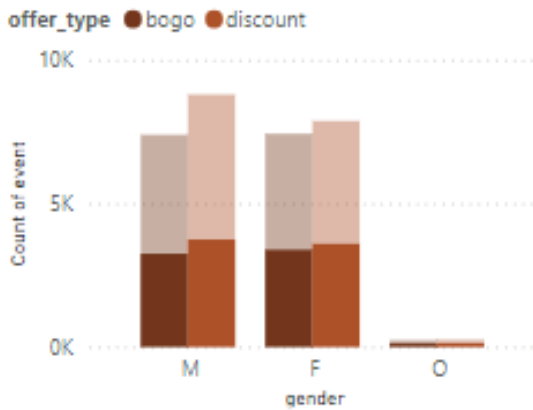
Count of Offers Completed by channels



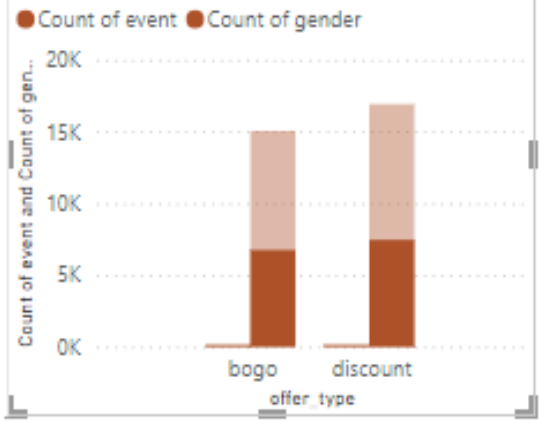
Count of Offers Completed by income_group



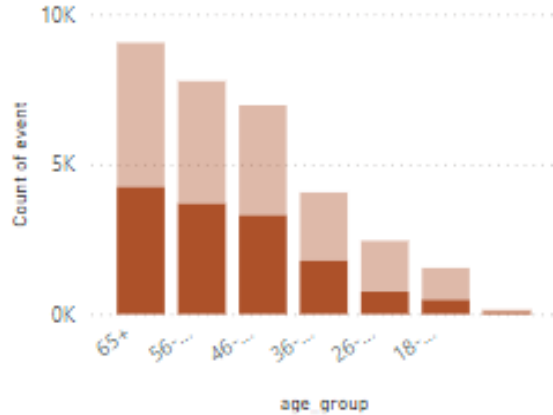
Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



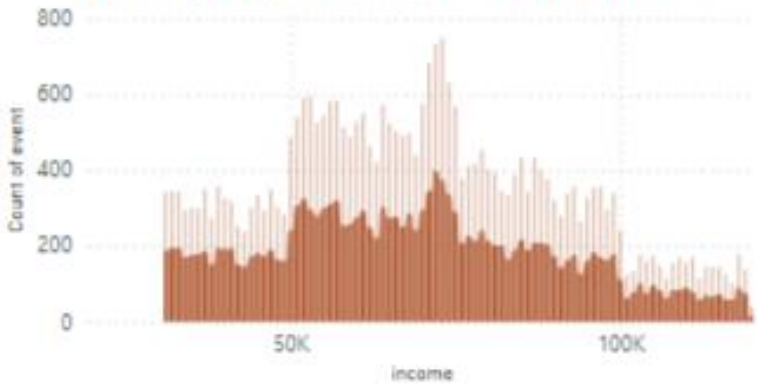
Count of offers completed by age_group



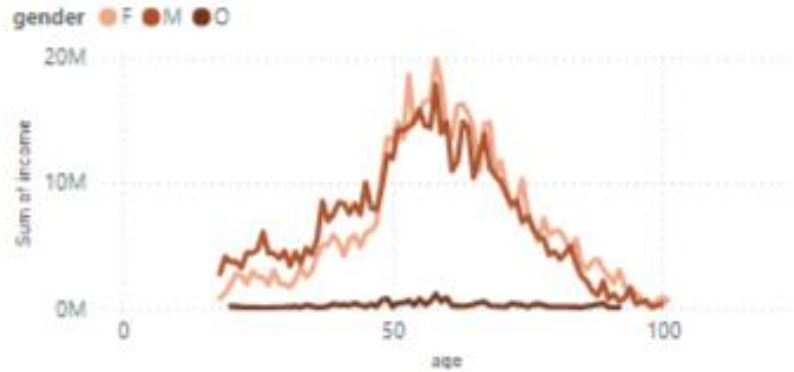
Promotions sent to Maven Cafe Customers

Sum Of offers completed
16.91K

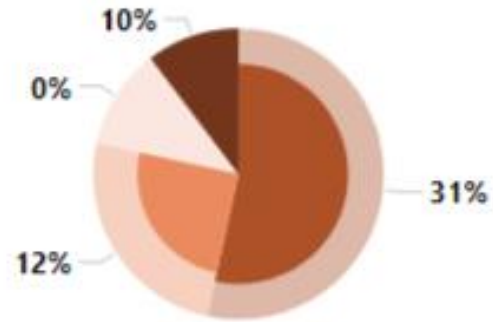
Count of Offers Completed by income



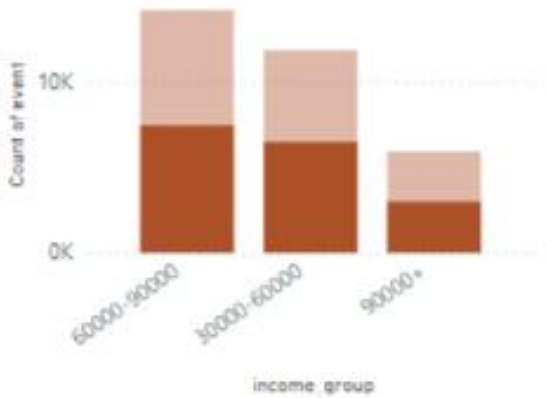
Sum of income by age and gender



Count of Offers Completed by channels

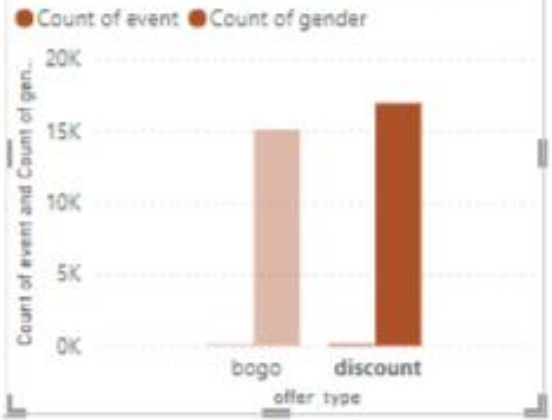


Count of Offers Completed by income_group

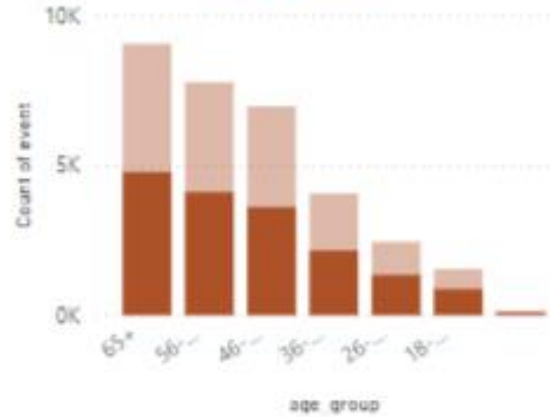


Count of offers completed by gender and offer_type

Count of offers completed and Count of gender by offer_type



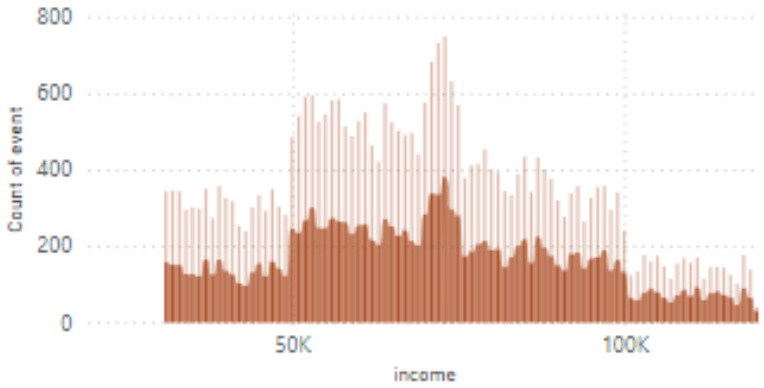
Count of offers completed by age_group



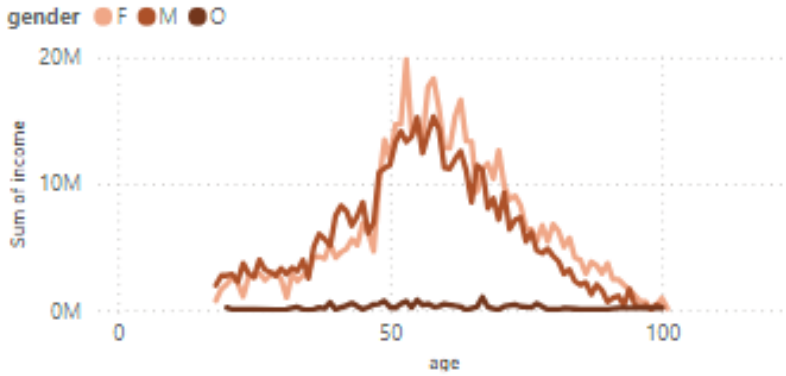
Promotions sent to Maven Cafe Customers

Sum Of offers completed
15.04K

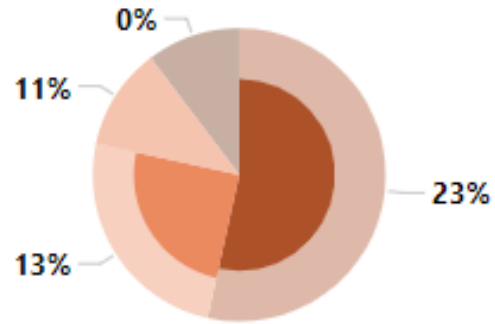
Count of Offers Completed by income



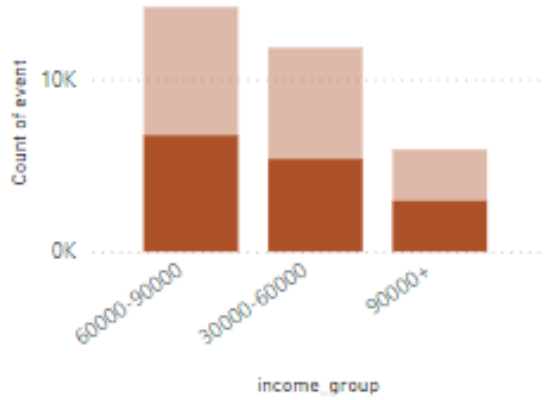
Sum of income by age and gender



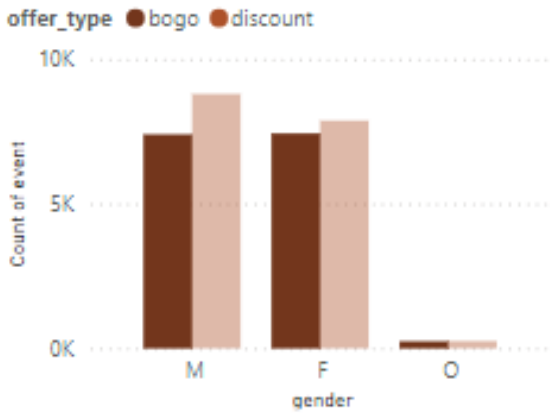
Count of Offers Completed by channels



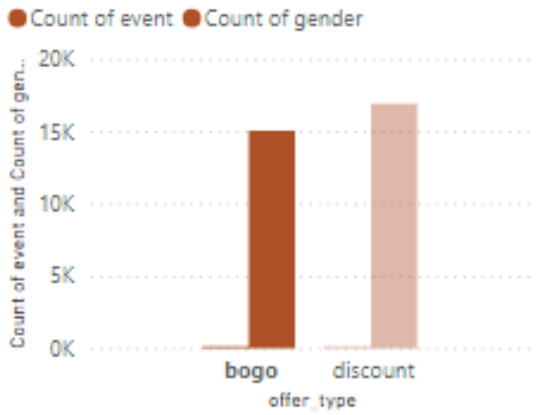
Count of Offers Completed by income_group



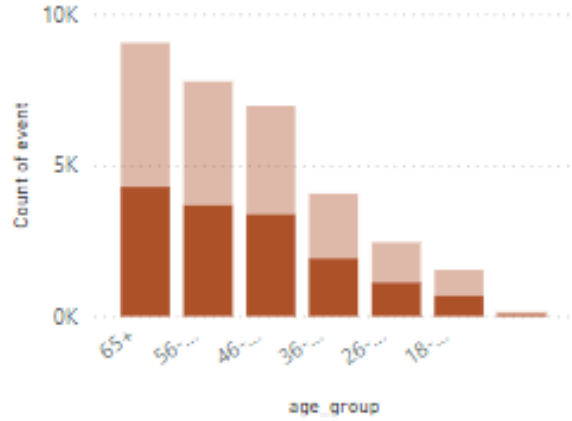
Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



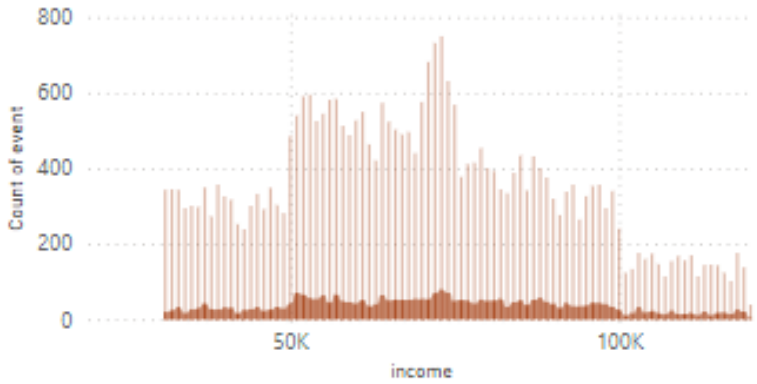
Count of offers completed by age_group



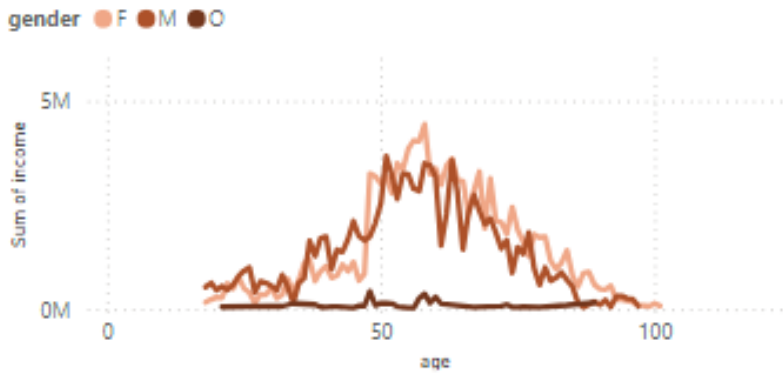
Promotions sent to Maven Cafe Customers

Sum Of offers completed
3300

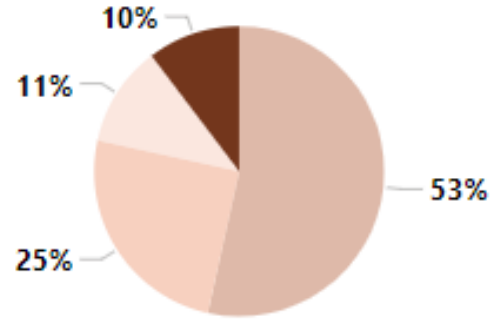
Count of Offers Completed by income



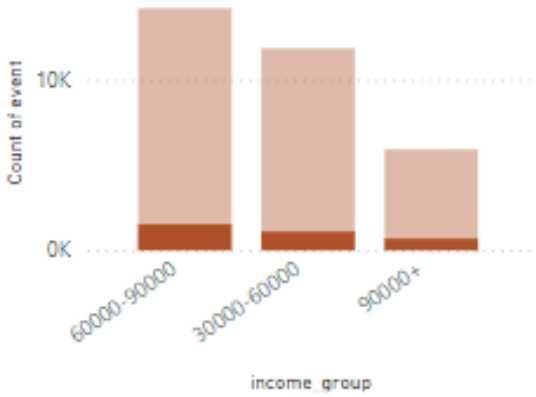
Sum of income by age and gender



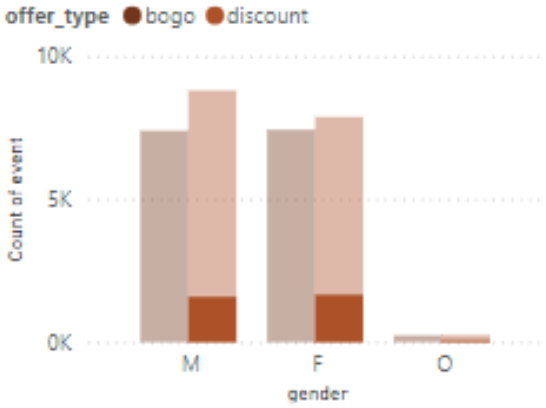
Count of Offers Completed by channels



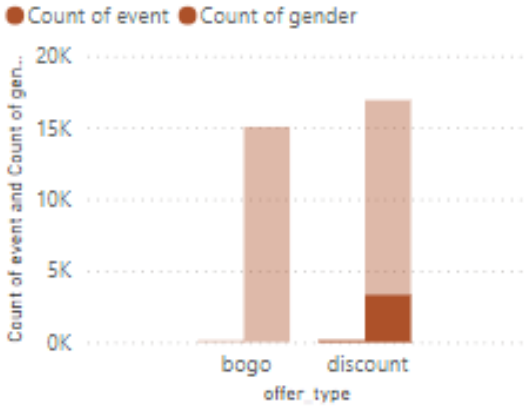
Count of Offers Completed by income_group



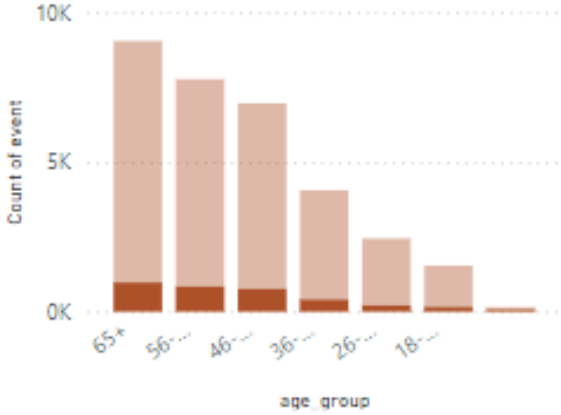
Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



Count of offers completed by age_group

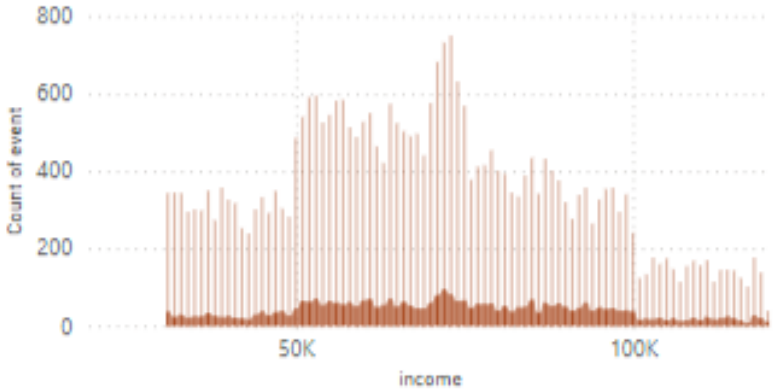


Promotions sent to Maven Cafe Customers

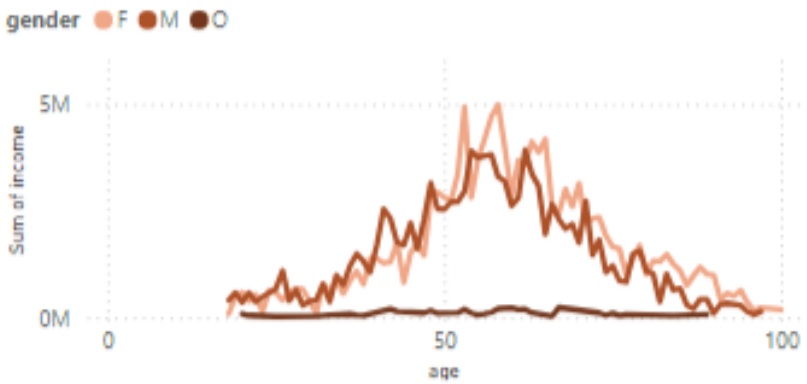
Sum Of offers completed

3596

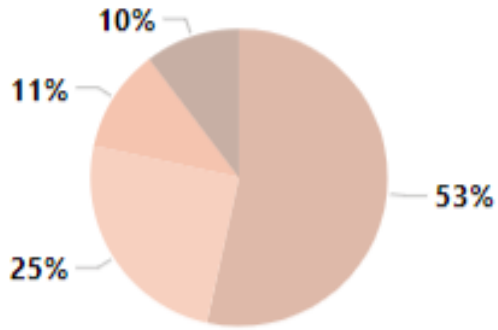
Count of Offers Completed by income



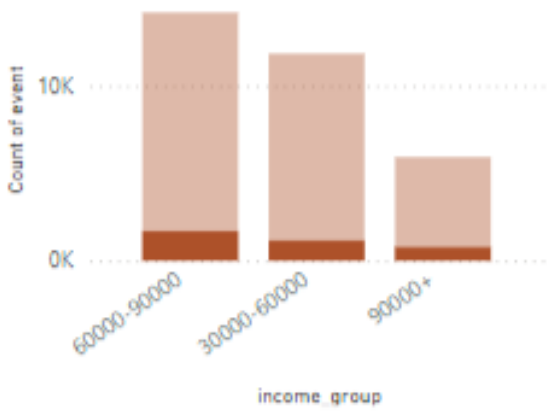
Sum of income by age and gender



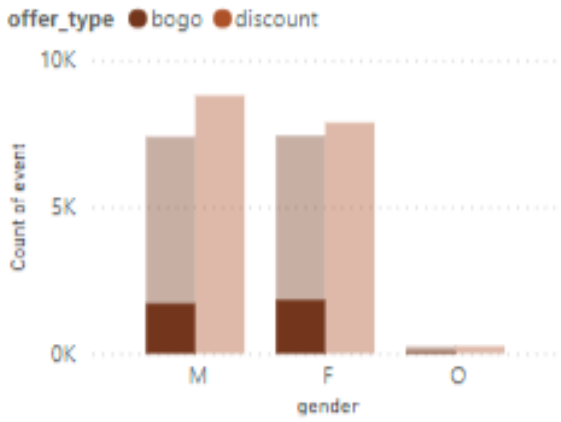
Count of Offers Completed by channels



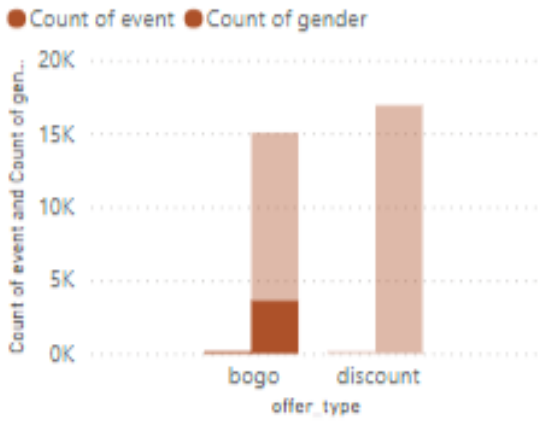
Count of Offers Completed by income_group



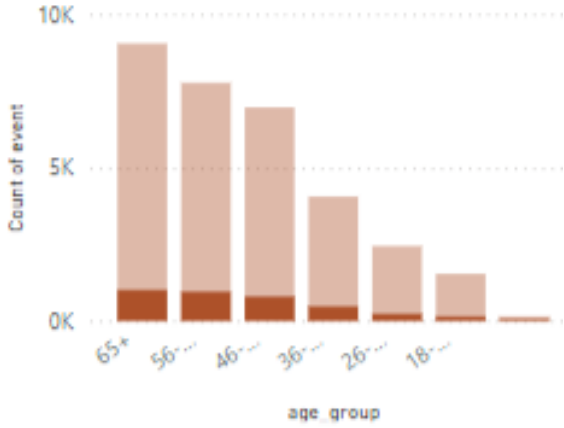
Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



Count of offers completed by age_group

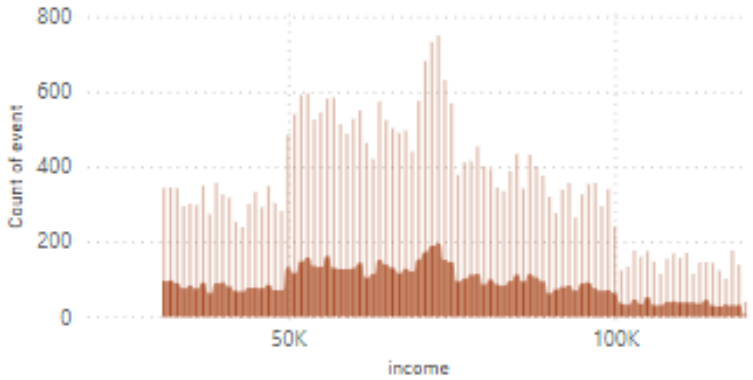


Promotions sent to Maven Cafe Customers

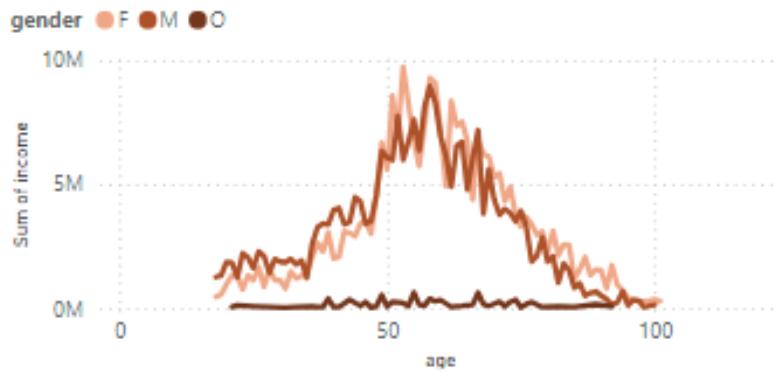
Sum Of offers completed

7971

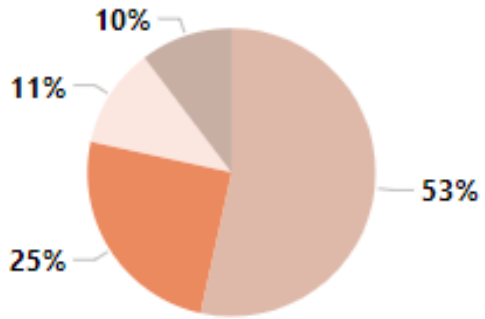
Count of Offers Completed by income



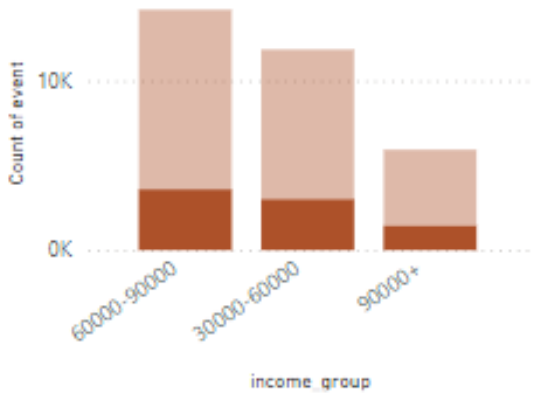
Sum of income by age and gender



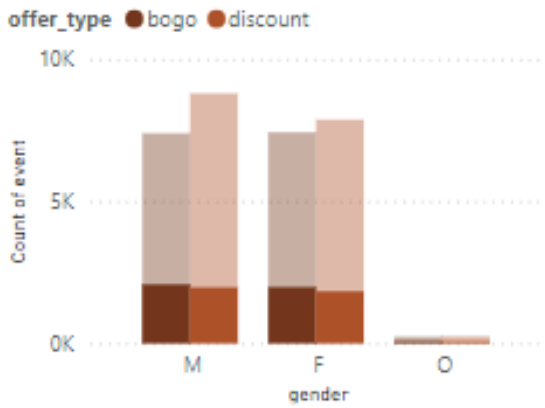
Count of Offers Completed by channels



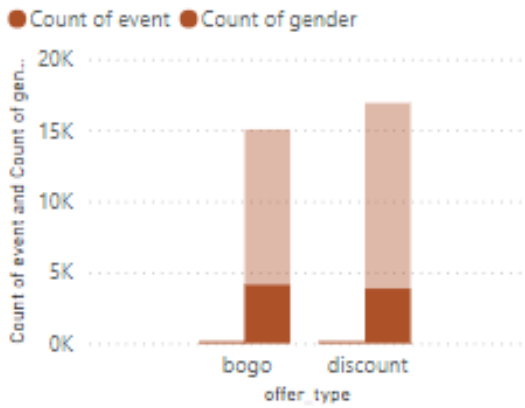
Count of Offers Completed by income_group



Count of offers completed by gender and offer_type



Count of offers completed and Count of gender by offer_type



Count of offers completed by age_group