CAPSTONE PROJECT
"MSG FRESH MARKET CUSTOMER DEMOGRAPHICS ANALYTICS"

KAREN GUTIERREZ - C0867263
LUCIA MELCHOR - C0871591
RICARDO MIGUEL GRAU - C0875718
SHREYAN VAIDYA- C0877811

Submitted to Dr. Shivani Anand

Business Intelligence Tools- 2024 Class

March 29th, 2023

Table of contents

01

Introduction

02

Purpose and Objectives

03

Project Strategy & Analysis Questions

04

Data Analysis and steps

- Multivariate correlation
- K-means clustering
- Demographic Analysis
- Behavior Analysis

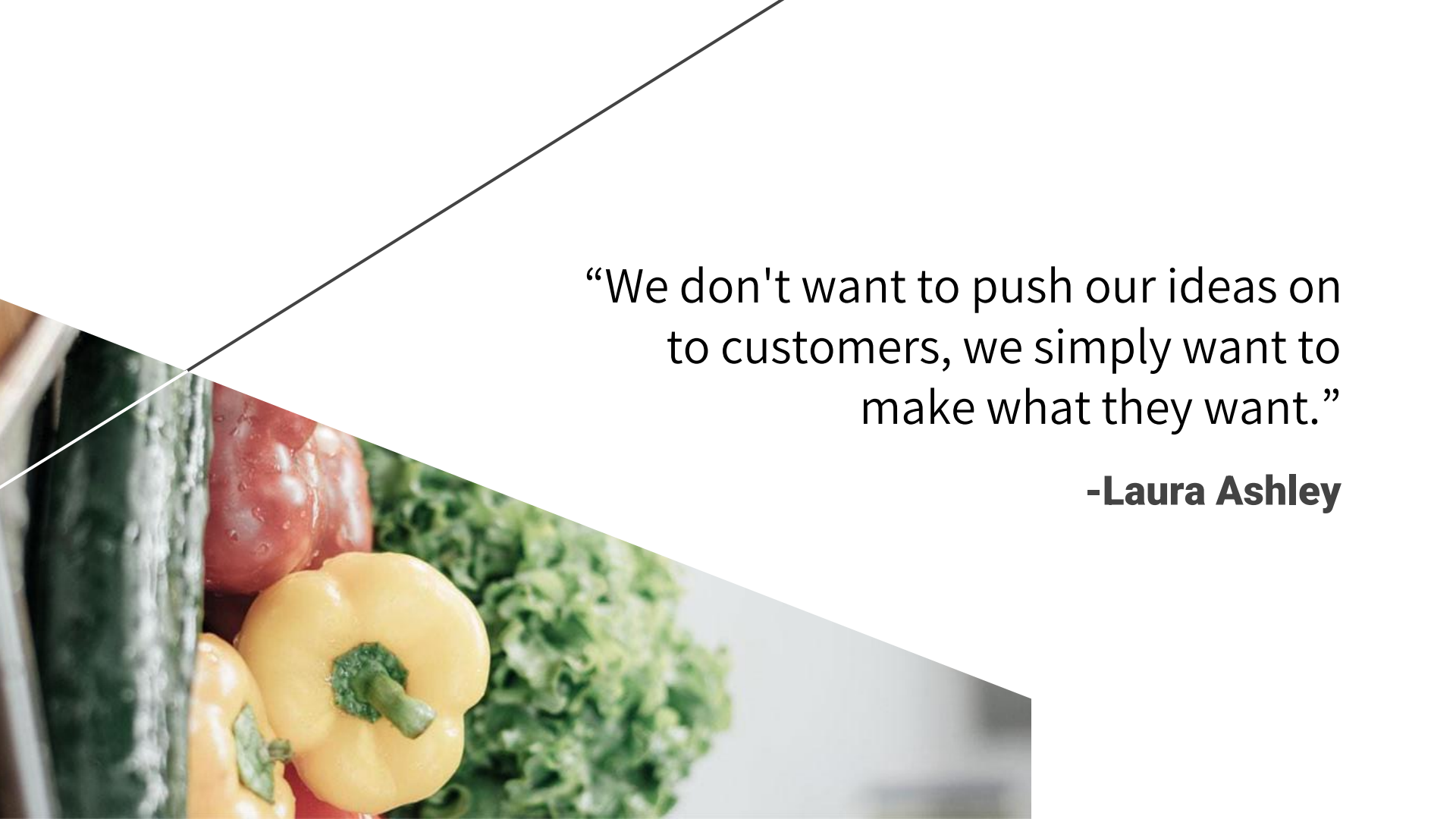**05**

Reliability Test (Chronbach Alpha)

06

Key Outcomes & Implementation

07

Usefulness to the Stakeholders

"We don't want to push our ideas on to customers, we simply want to make what they want."

-Laura Ashley

Introduction

Understanding customer demographics can help companies to not only identify who their customers are, but also how their demographics relate to their purchasing behavior.

- The dataset consists of 29 columns and 2240 rows with the data of 3 years including their age, gender, education level, occupation, and other demographic information of "MSG Fresh Market" company's customer base.

- Contains data compilation of frequency of store or website total conversions since company's establishment in 3rd quarter of 2012 until 2nd quarter of 2014.

- Despite an increase in total new customers for the first 2 quarters since initiation, there has been a compounded decrease of quarterly negative growth rate by -1.69% from Q1/2013 to Q2/2014 (6 quarters) and **20% revenue decrease** in the following quarters.

** Assumption: "MSG Fresh Market" provided this approximation.

Our purpose:

We hope to shed some light on the complex relationship between demographics and consumer behavior, and to provide valuable insights for "MSG Fresh Market" looking to optimize their customer marketing strategies.

We aim to make informed decisions for marketing strategies in order to create products and services that truly meet customer needs and ultimately improve sales.

Objectives

Customer Demographic Analysis

Develop a deep understanding of customer demographics and their impact on purchasing behavior through K-means clustering.

Demographic & Behavior Patterns and Trends

Identify descriptive analytics on demographic & behavior patterns and trends in customer behavior through correlation matrix.
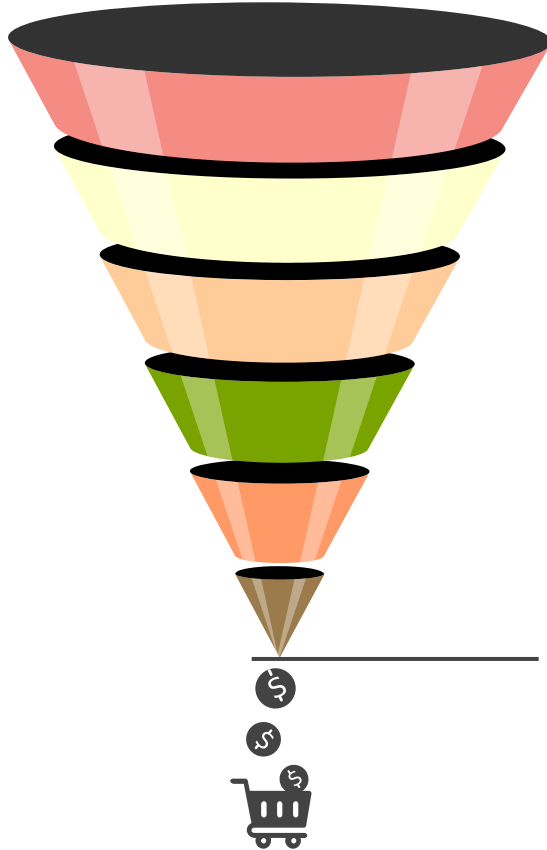
Cross-Selling Opportunities

Identify potential opportunities based on customer demographics and purchasing behavior obtained from correlation matrix.

Reliability

Identify the accuracy and consistency of the data used through Cronbach Alpha Analysis.

Project Strategy

✓ Collected raw data on customer demographics such as age, gender, income, education, and location through Kaggle.

✓ Cleaned and transformed the data using Power Query, then used Power BI to create interactive dashboards to show key trends and patterns.

✓ Use descriptive statistics to define and analyze customer demographics and behavior.

✓ Excel, Python, Power BI visualizations tools.

✓ Communicate insights to stakeholders.

✓ Marketing strategies.

Overview

ETL

Exploratory Analysis

Tools

Interpret Results

Recommend

Analysis Questions

?

1. What distinct customer clusters do our customers fall into?

2. Are there any significant trends in customer behavior that could inform our marketing strategies?

3. Is the data consistent and accurate?

Data analysis and steps

| Correlation Matrix								
Attribute (groups)	AMOUNT SPENT ON						CHILDREN	
Attribute (groups) ▲	Fish	Fruits	Gold Prods	Meats	Sweet Prods	Wines	Kidhome	Teenhome
AMOUNT SPENT ON								
Fish	1.00	0.59	0.42	0.57	0.58	0.40	-0.39	-0.20
Fruits	0.59	1.00	0.39	0.54	0.57	0.39	-0.37	-0.18
Gold Prods	0.42	0.39	1.00	0.35	0.37	0.39	-0.35	-0.02
Meats	0.57	0.54	0.35	1.00	0.52	0.56	-0.44	-0.26
Sweet Prods	0.58	0.57	0.37	0.52	1.00	0.39	-0.37	-0.16
Wines	0.40	0.39	0.39	0.56	0.39	1.00	-0.50	0.00
CHILDREN								
Kidhome	-0.39	-0.37	-0.35	-0.44	-0.37	-0.50	1.00	-0.04
Teenhome	-0.20	-0.18	-0.02	-0.26	-0.16	0.00	-0.04	1.00
INCOME								
Income	0.52	0.51	0.38	0.68	0.52	0.69	-0.51	0.03
PURCHASE TYPE								
Catalog	0.53	0.49	0.44	0.72	0.49	0.64	-0.50	-0.11
Deals	-0.14	-0.13	0.05	-0.12	-0.12	0.01	0.22	0.39
Store	0.46	0.46	0.38	0.48	0.45	0.64	-0.50	0.05
Web	0.29	0.30	0.42	0.29	0.35	0.54	-0.36	0.16

STEP 1: MULTIVARIATE CORRELATION

✓ We can group the numerical variables into 4 groups:
- Amount spent on
- Children
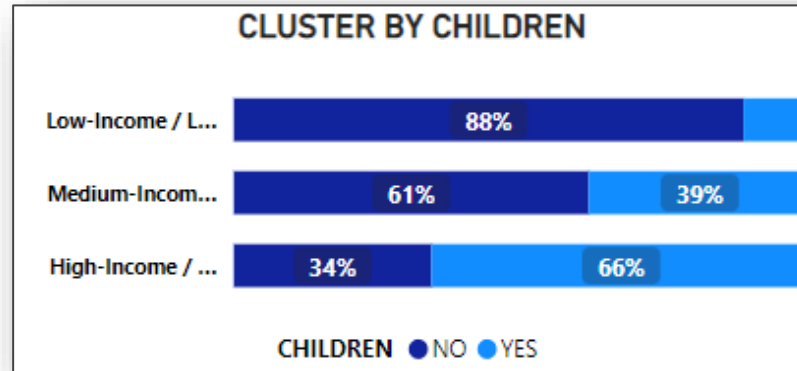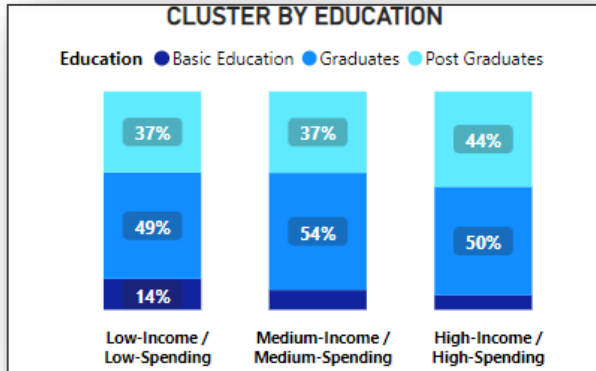- Income
- Purchase type.

Data analysis and steps
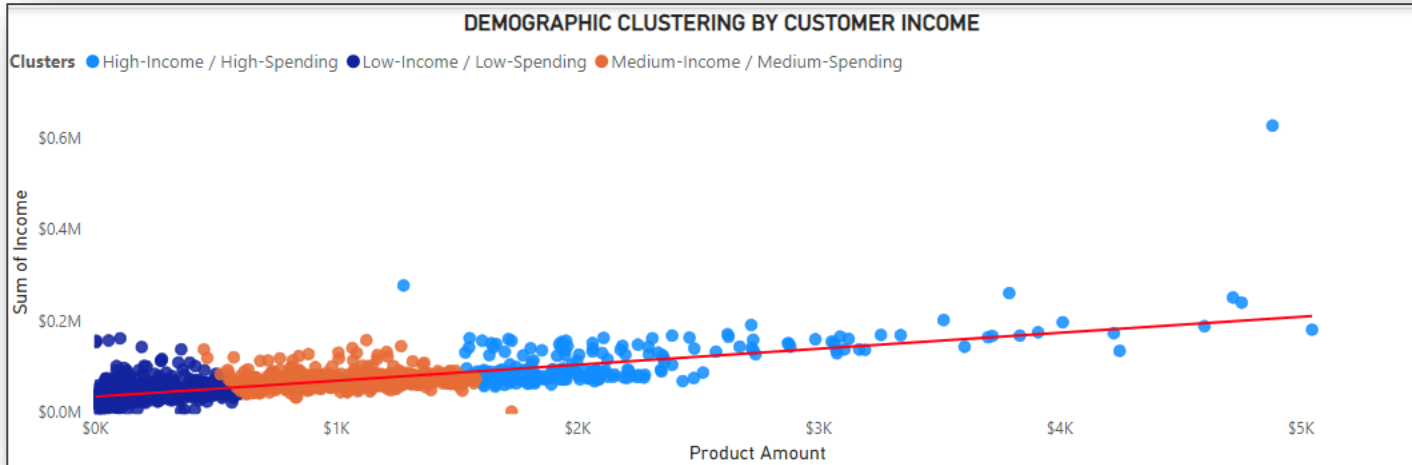
STEP 2: K-MEANS CLUSTERING

✓ As a second step, we used the K-Means Clustering tool of Power BI with the variables that had the strongest correlation: Income and Amount spent on. As a result we obtained the following clusters:

✓ Low-Income / Low-Spending
Medium-Income / Medium-Spending
High-Income / High-Spending

Data analysis and steps

CLUSTER BY EDUCATION

Education ● Basic Education ● Graduates ● Post Graduates

	Low-Income / Low-Spending	Medium-Income / Medium-Spending	High-Income / High-Spending
Post Graduates	37%	37%	44%
Graduates	49%	54%	50%
Basic Education	14%		

CLUSTER BY CHILDREN

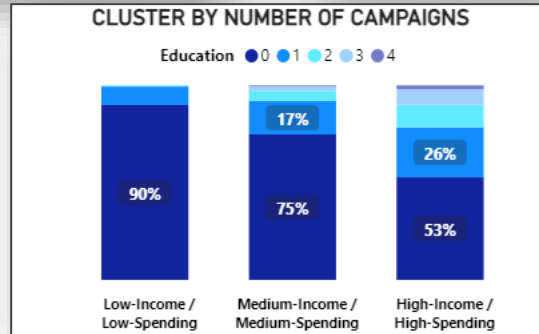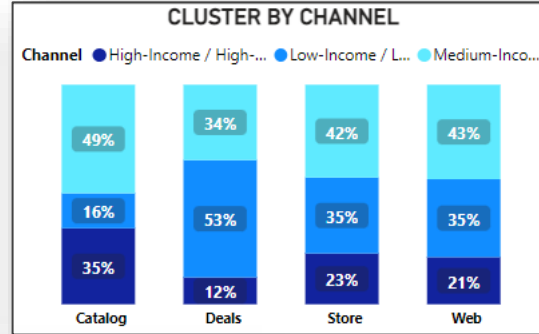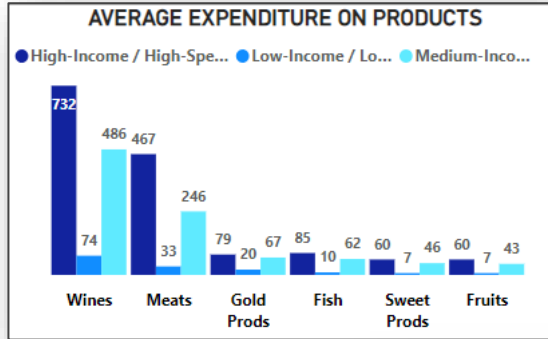|	NO	YES
Low-Income / L...	88%	
Medium-Incom...	61%	39%
High-Income / ...	34%	66%

CHILDREN ● NO ● YES

STEP 3: DEMOGRAPHIC ANALYSIS

✓ These results were obtained when performing the analysis with the quantitative and qualitative demographic variables.

Data analysis and steps

AVERAGE EXPENDITURE ON PRODUCTS

CLUSTER BY CHANNEL

CLUSTER BY NUMBER OF CAMPAIGNS

STEP4: BEHAVIOR ANALYSIS

✓ These results were obtained when performing the analysis with the quantitative and qualitative demographic variables.

Key Outcomes & Implementation

✓ Clusters were obtained for the Income and Amount Spent variables, which have the strongest correlation. For cross-selling strategies, there is a relation of spending on wine and meat products in the 3 clusters.
Implementation suggestion: Use a grid layout to create natural barriers that serve to simultaneously group wine and meat products in order to cross sell.

✓ Regarding another demographic feature, we found that the higher the level of education, the higher the income of the customers.
Implementation suggestion: Get a fellow graduate to give a product recommendation through an online newsletter or Instagram.

✓ For the price and promotion strategies, we found that the Low-Income / Low-Spending Cluster tends to buy more through discounts and the High-Income / High-Spending Cluster makes more purchases through Catalogs.
Implementation suggestion: Provide physical coupons that can be used in specific locations more than once assures the generation of leads.

https://www.shopify.com/retail/the-ultimate-guide-to-retail-store-layouts

Usefulness to the Stakeholders

- ✓ Generating better leads based on identified clusters
- ✓ Improving customer service by adapting products to cluster preferences and usage trends
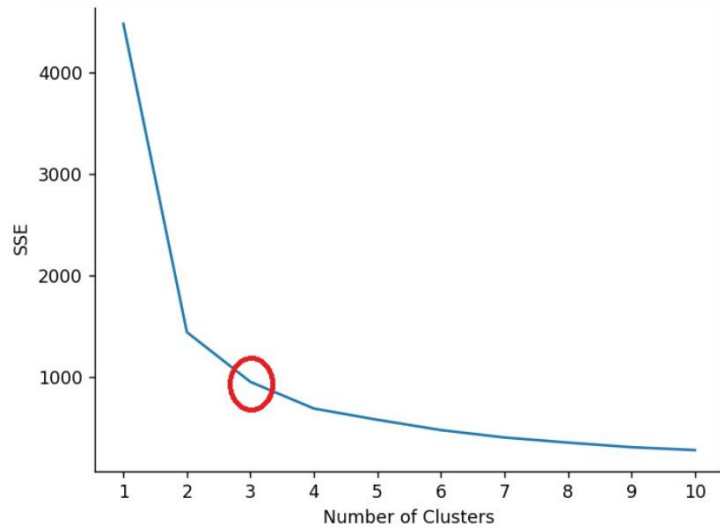- ✓ Ensure the strategy suggested could be relied upon

Feedback from 3/24/23

✓ Add correlation for cross selling in objectives.
    See slide 6.

✓ Please identify the correlation coefficient and re-do the thresholds and the tabling.
    Pearson coefficient used (being acceptable since 0.5 threshold. See below link.

✓ Please also confirm your conclusions by reapplying the correlation analysis after the clustering.
    Done, see slide 13.

✓ Please keep at least 30 iterations for clustering.
    Done. See slide 17 for more details.

https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/#Pearson

venv library root
DogBite.py
KMeans 30 Iterations BI Project.py
main.py
superstore_train.py
Week1.py
Week2.py
Week5.py
Week9.py
Week10.py
Week11.py
External Libraries
< Python 3.10 (pythonProject) > C:\Users\karen\
Binary Skeletons
DLLs
Extended Definitions
Lib
Python310 library root
site-packages
KMeans 30 Iterations BI Project

C:\Users\karen\PycharmProjects\
[[0.28805389 1.67941681]
 [-0.26300308 -0.96127545]
 [0.91765307 0.28267329]
 [-1.18336261 -0.91809432]
 [0.29529603 -0.30525442]]

Figure 1

0
1
2

Income

160000
140000
120000
100000
80000
60000
40000
20000
0

0 500 1000 1500 2000 2500
Amount

x=1077 y=1.375e+05