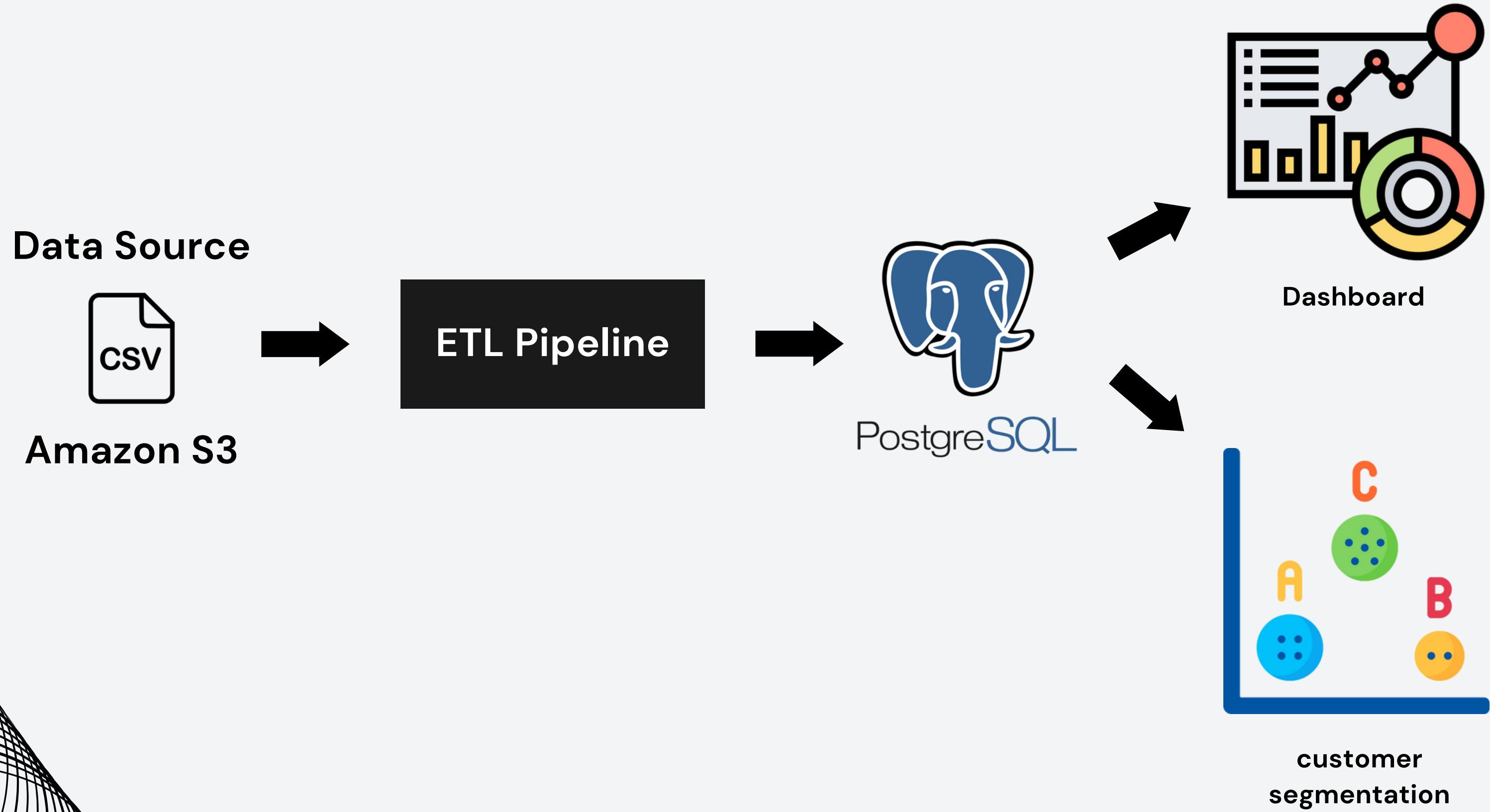


# **LEMONFARM DATA INSIGHTS & CUSTOMER SEGMENTATION PIPELINE**

**KORNKRIT RATTANASAMNIANG**

# PROJECT OVERVIEW



# LEMON FARM DATA

## LF\_CAFE\_DATA

bill_date	bill_time	bill_no	member_id	sku_code	sku_name	qty	sales
2022-09-07	08:35:22	a6bd1288	14bc04ee	7009018	ลาเต้เย็น 240 ml#ชวด	1	79.8579

## CUSTOMER\_DATA

member_id	sex	age	location
14bc04ee	F	59	North

## PRODUCT CAT

code	category
7023	Salad

# ETL PIPELINE

- EXTRACT CSV DATA FROM AMAZON S3
- TRANSFORM DATA
  - > JOIN CUSTOMER\_DATA AND PRODUCT\_CAT ON LF\_CAFE\_DATA
  - > DEVELOP DAY\_NAME AND TIME\_OF\_DAY FEATURE
- LOAD DATA TO POSTGRESQL DATABASE

->LINK TO PYTHON CODE <-

# TRANSFORMED DATA

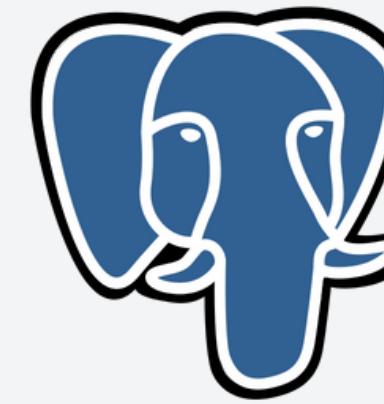
## TRANSFORMED DATA

bill_date	bill_time	bill_no	member_id	sku_code	sku_name	qty	sales	sex	age	location	category	day_name	time_of_day
2022-09-07	08:35:22	a6bd1288	14bc04ee	7009018	ลาเต้ เย็น 240 ml#ขวด	1	79.8579	F	59	North	Salad	Monday	Morning



# DATA DASHBOARD

# DATA DASHBOARD



PostgreSQL



LOOKER STUDIO

->LINK TO DASHBOARD<-

# DATA DASHBOARD

## Transaction data of Lemon Farm Cafe in 14 days (7 Sep - 20 Sep)

Total Revenues  
305.5K

Total Orders  
2,457

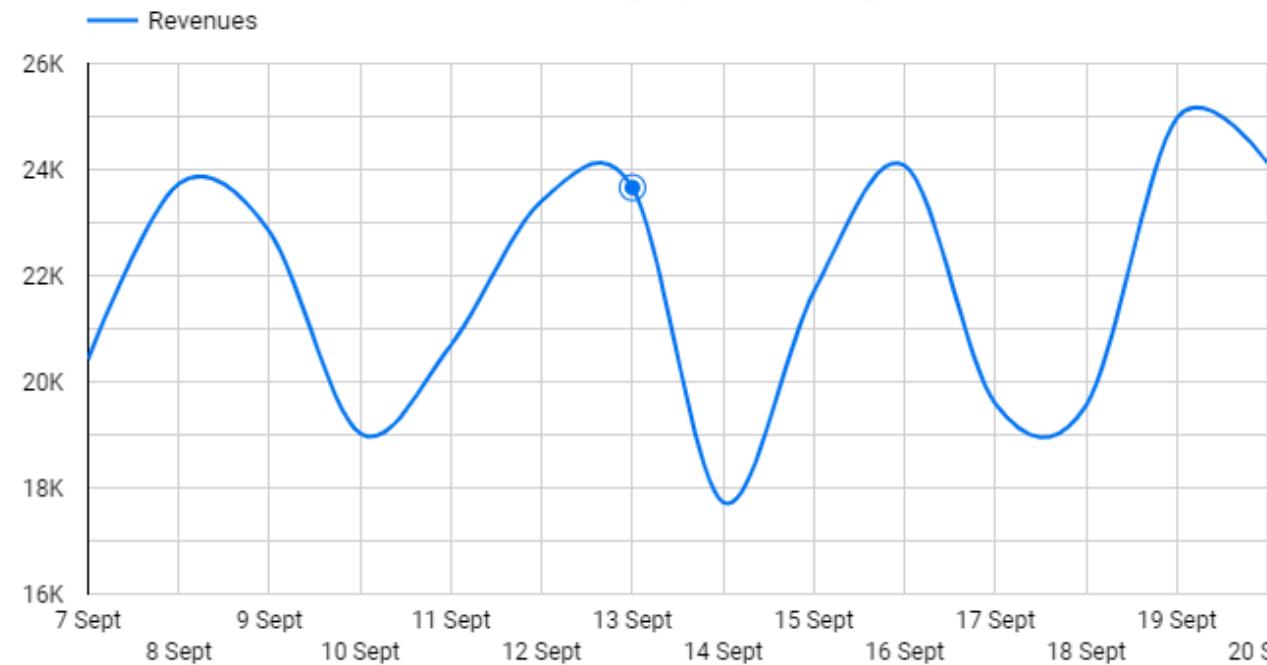
Total Members  
1,043

Avg Daily Revenues  
21.8K

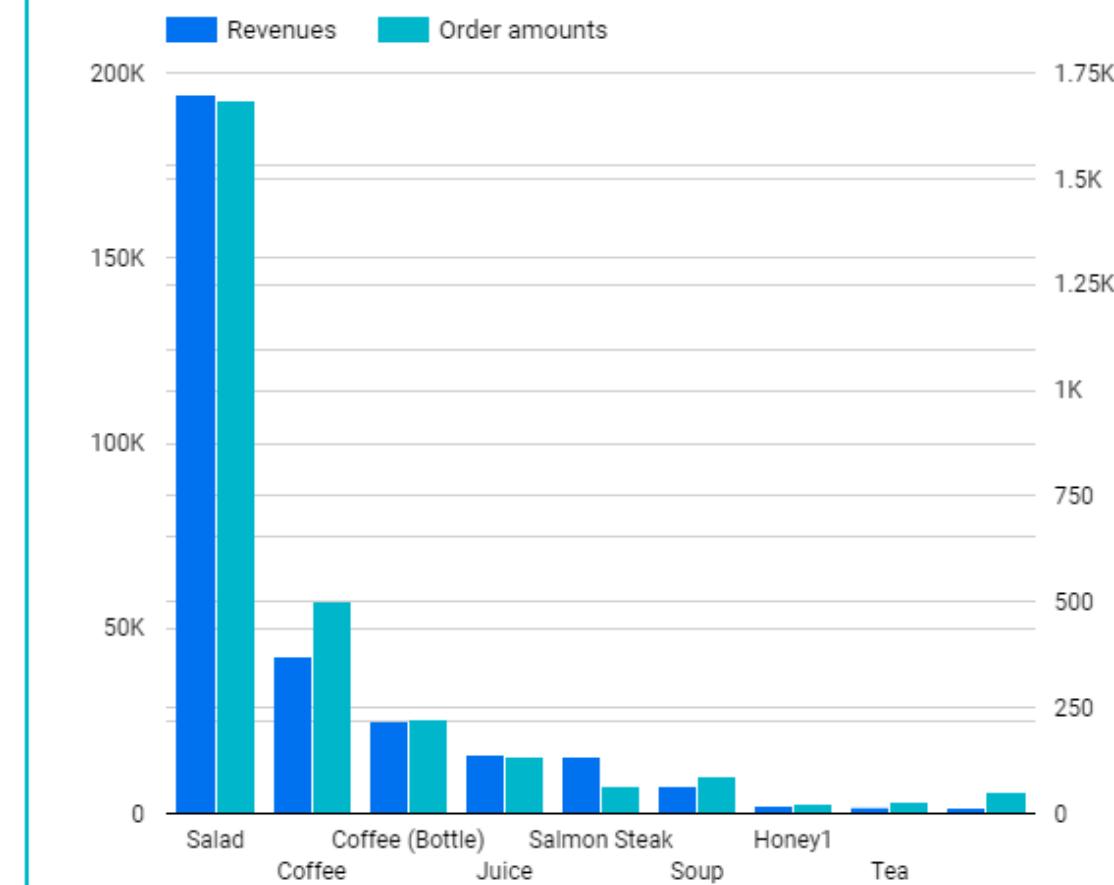
Avg Daily Orders  
176.0

Avg Daily Members  
75.0

Revenues graph in 14 days



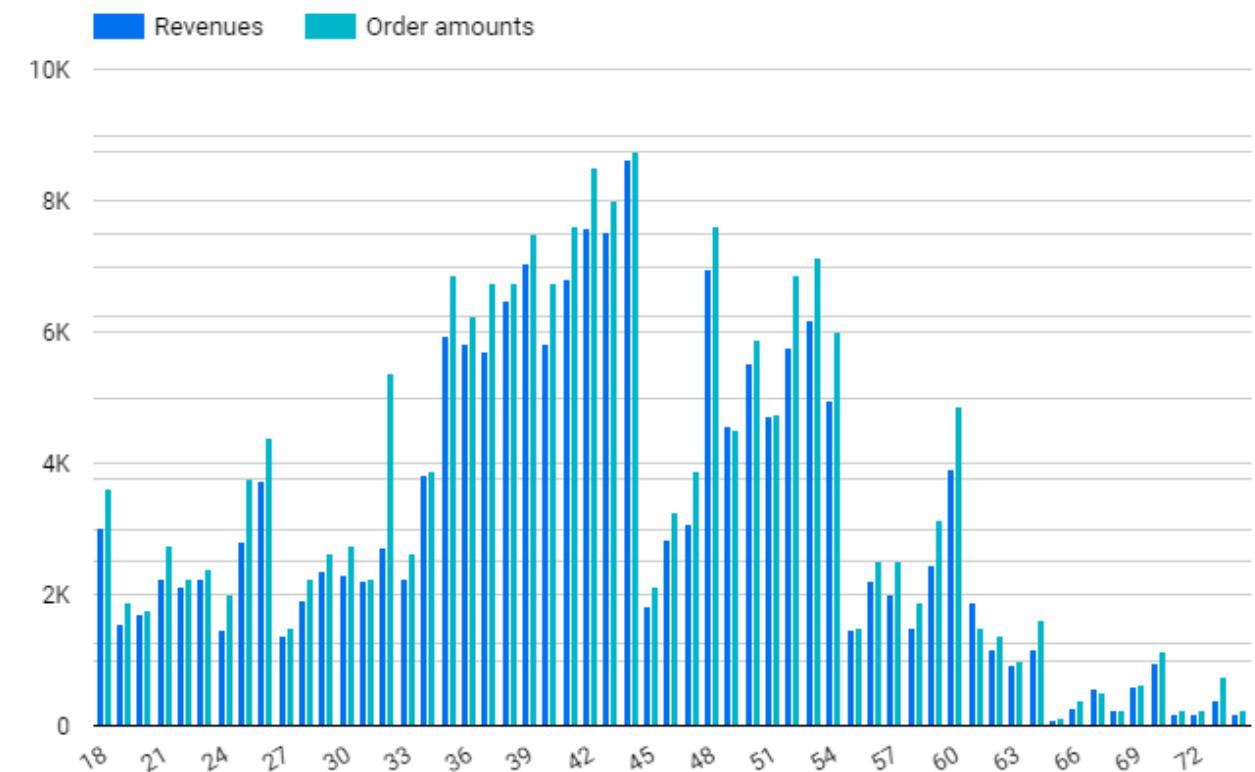
Revenues & Order amounts in each menu categories



# DATA DASHBOARD

## Lemon Farm customer characteristics

Revenues & Order amounts in each age of customer



Revenues & Order amounts in each sex of customer



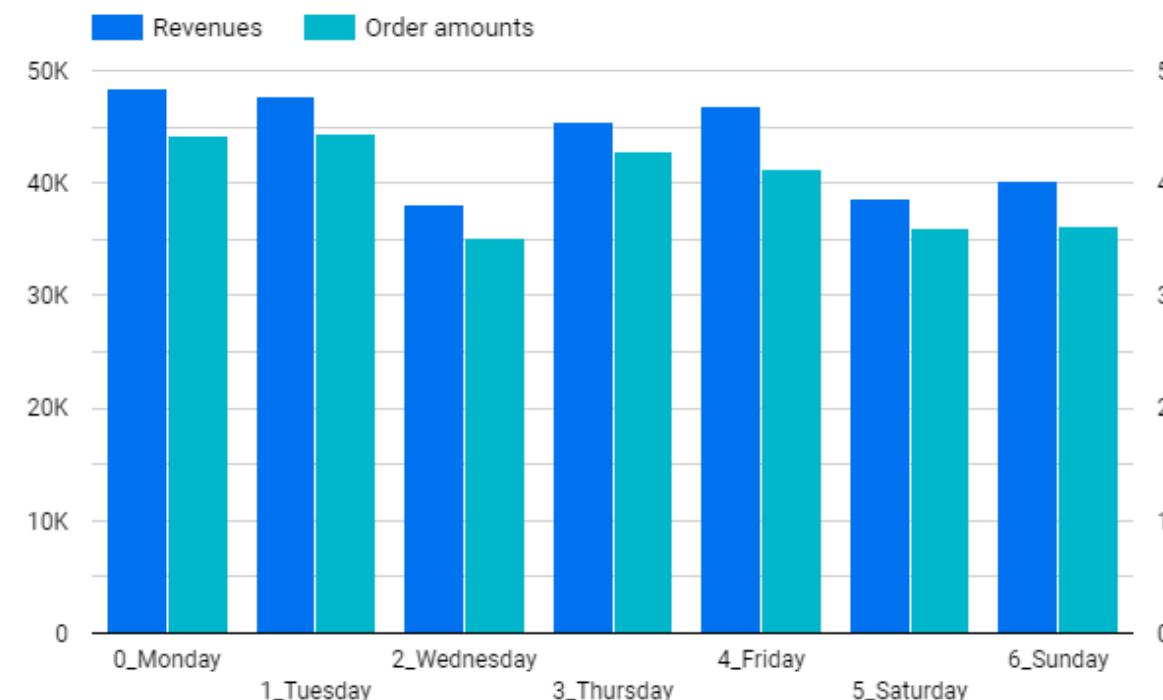
Revenues & Order amounts in each location of customer



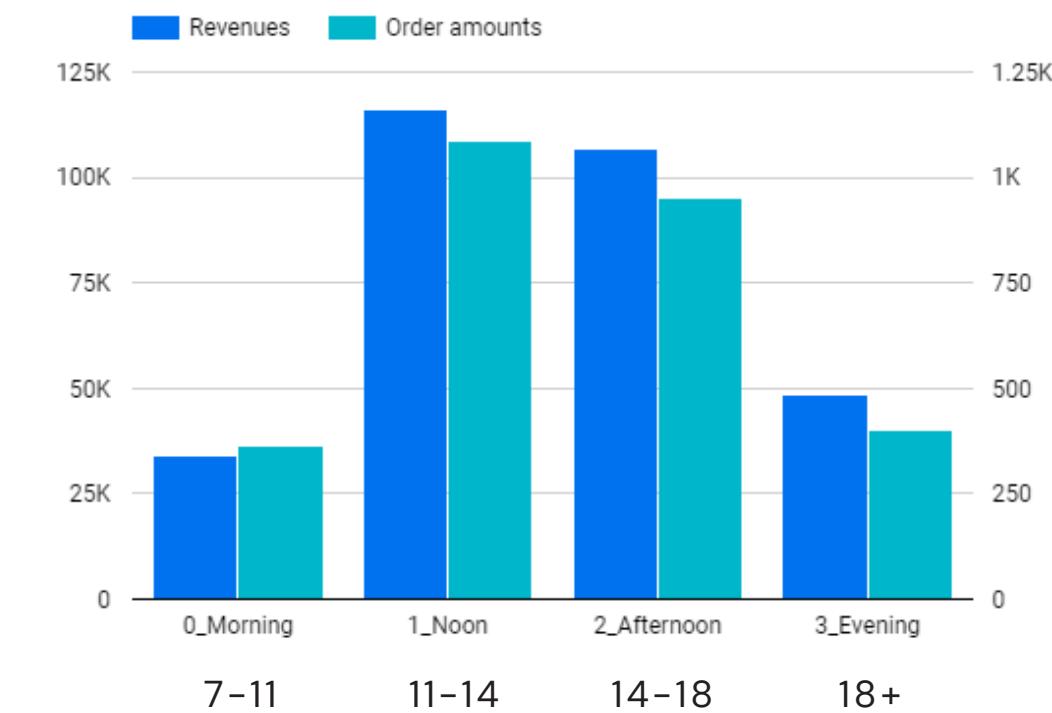
# DATA DASHBOARD

## Transaction Time and Period

Revenues & Order amounts in each day



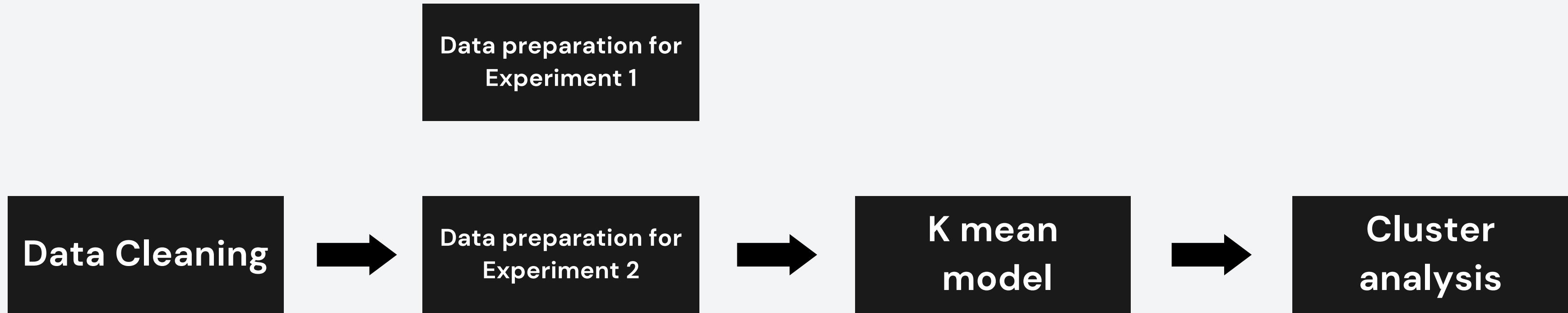
Revenues & Order amounts in each time period





# **CUSTOMER SEGMENTATION**

# CUSTOMER SEGMENTATION



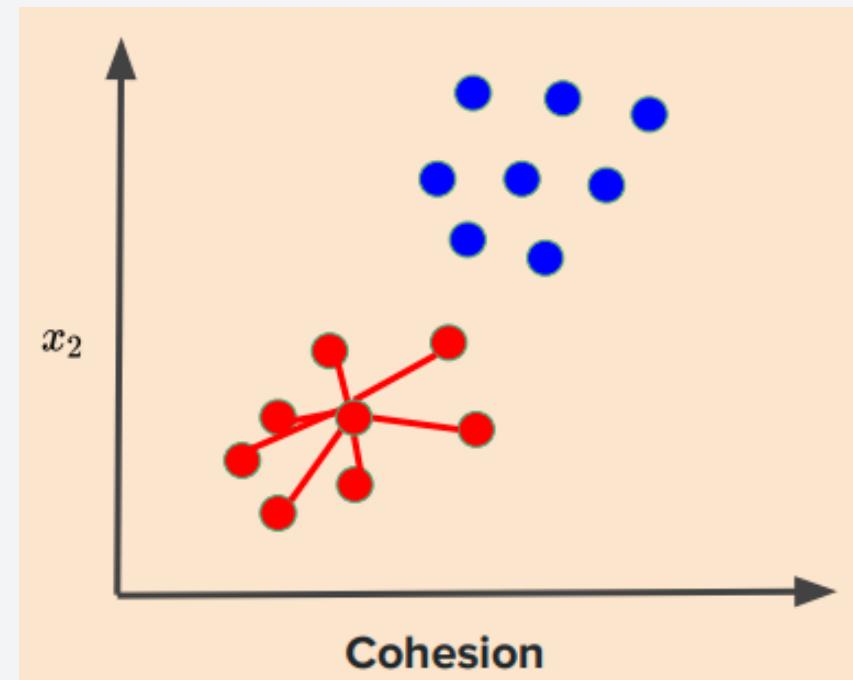
Data preparation for  
Experiment 3

Experiment 1: menu categories clustering  
Experiment 2: menu categories and time\_of\_day clustering  
Experiment 3: customer characteristics clustering

->LINK TO PYTHON CODE <-

# CLUSTER EVALUATION

## WCSS SCORE (WITHIN-CLUSTER-SUM-OF-SQUARES)



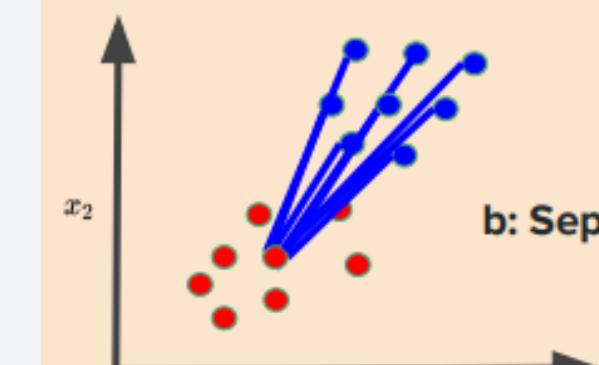
$$\text{WCSS} = \sum_{C_k} \left( \sum_{d_i \text{ in } C_k} \text{distance}(d_i, C_k)^2 \right)$$

Where,

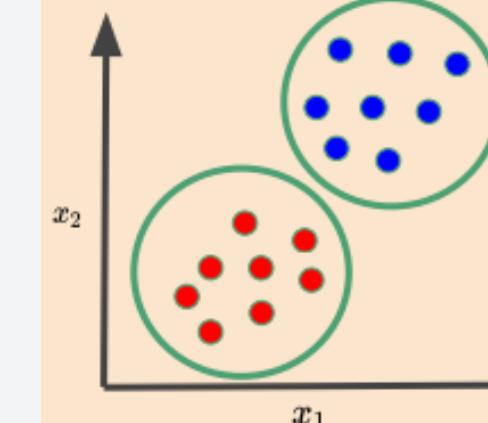
$C$  is the cluster centroids and  $d$  is the data point in each Cluster.

## SILHOUETTE SCORE

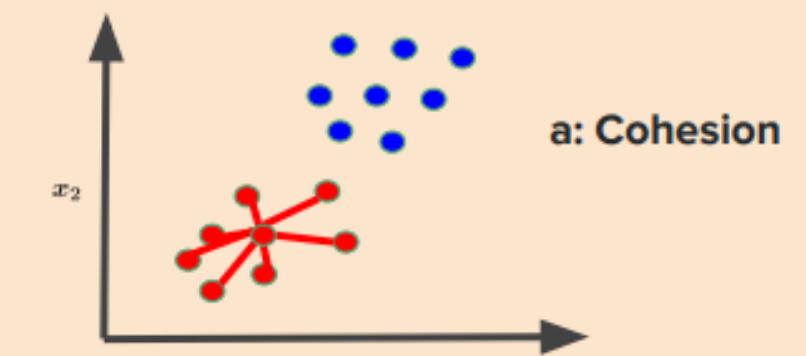
$$s = \frac{b-a}{\max(b,a)}$$



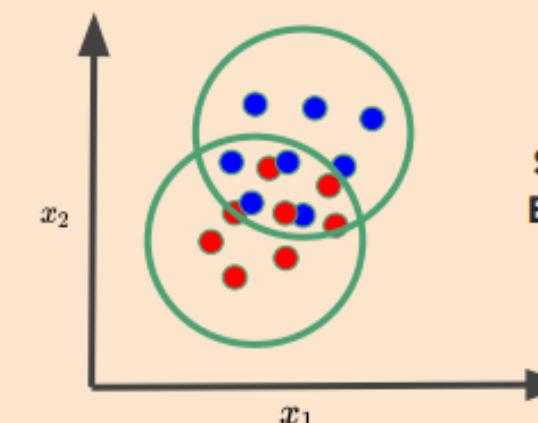
b: Separation



s close to 1  
Good clusters



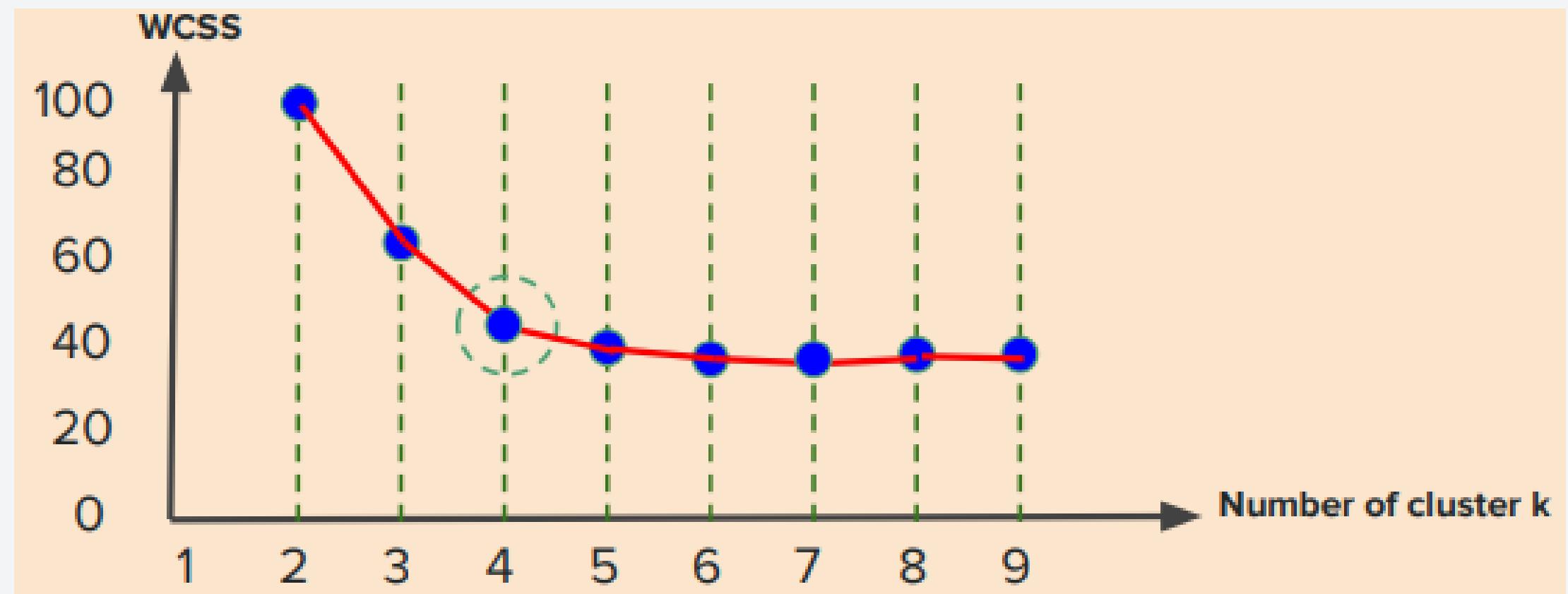
a: Cohesion



s close to 0  
Bad clusters

# CLUSTER EVALUATION

## ELBOW METHOD



# EXPERIMENT 1

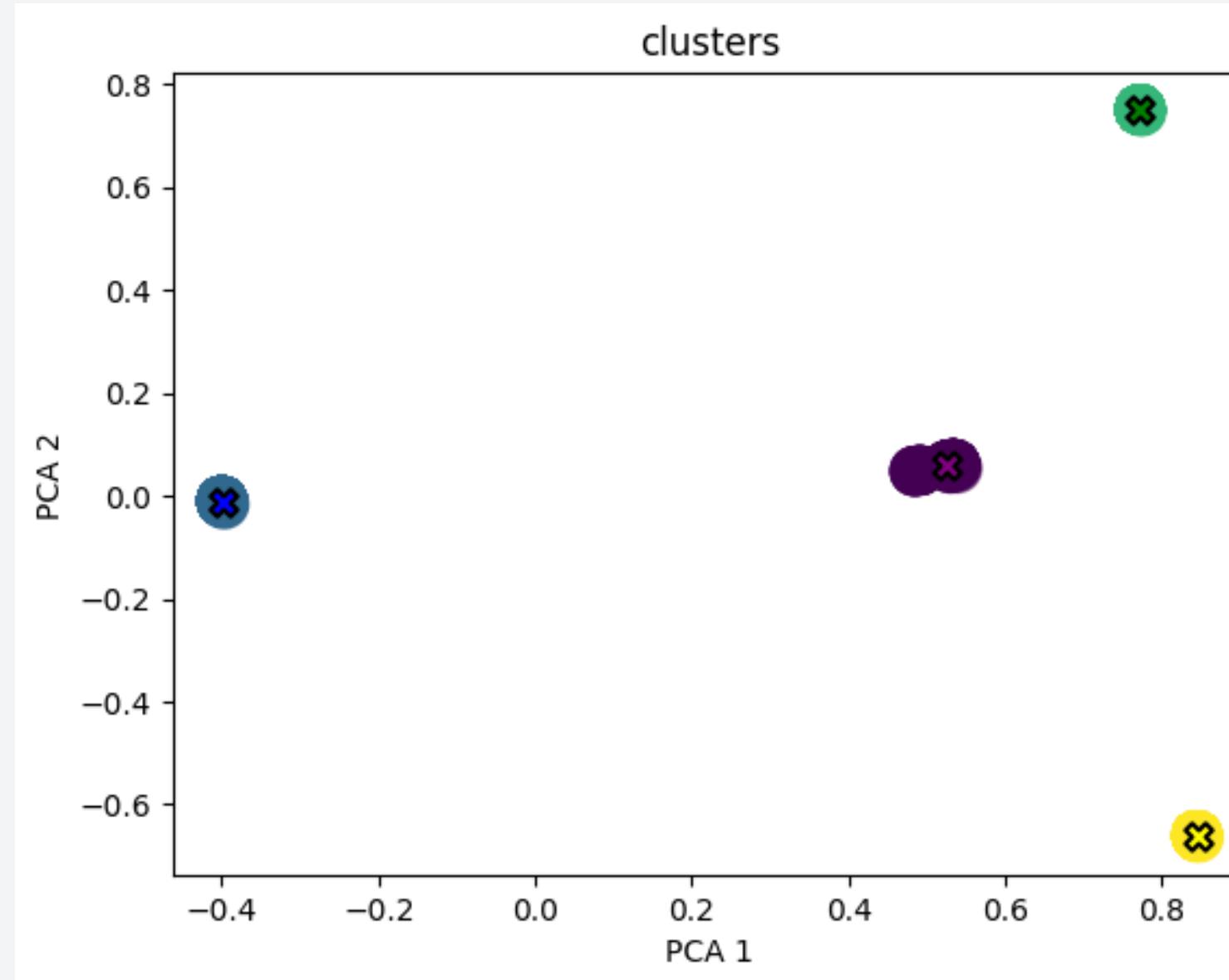
## MENU CATEGORIES CLUSTERING

- > SELECT FEATURE (MEMBER\_ID, SALES, CATEGORY)
- > REMOVE MISSING VALUE
- > GROUP BY MEMBER\_ID (AGGREGATE FUNCTION -> SALES:SUM, CATEGORY:MAX)
- > DROP MEMBER\_ID
- > ONE-HOT ENCODING
- > MIN MAX SCALER

#	Column	Non-Null Count
0	sales	1036 non-null
1	category_Coffee	1036 non-null
2	category_Coffee (Bottle)	1036 non-null
3	category_Honey1	1036 non-null
4	category_Juice	1036 non-null
5	category_Salad	1036 non-null
6	category_Salmon Steak	1036 non-null
7	category_Soup	1036 non-null
8	category_Tea	1036 non-null

# EXPERIMENT 1

## MENU CATEGORIES CLUSTERING



WCSS SCORE = 96.9357 SILHOUETTE SCORE = 0.8710

### BLUE CLUSTER

CLUSTER SIZE: 666  
MEAN SALES: 163.99  
SD SALES: 112.66  
SALAD: 666

### GREEN CLUSTER

CLUSTER SIZE: 118  
MEAN SALES: 132.48  
SD SALES: 98.62  
COFFEE: 118

### PURPLE CLUSTER

CLUSTER SIZE: 118  
MEAN SALES: 250.36  
SD SALES: 245.76  
HONEY: 11  
JUICE: 31  
SALMON STEAK: 35  
SOUP: 34  
TEA: 7

### YELLOW CLUSTER

CLUSTER SIZE: 134  
MEAN SALES: 157.80  
SD SALES: 135.19  
COFFEE (BOTTLE): 134

# EXPERIMENT 2

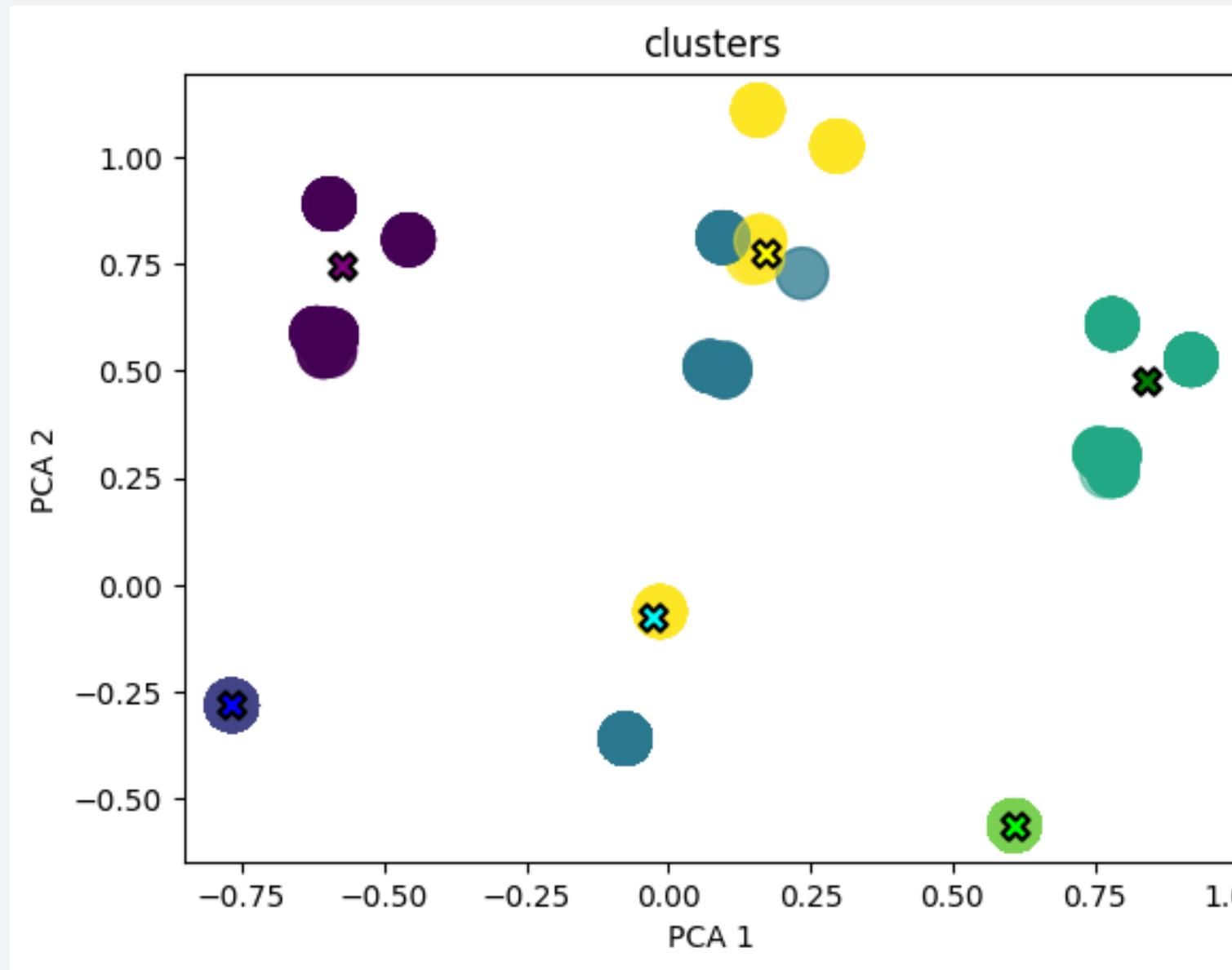
## MENU CATEGORIES AND TIME\_OF\_DAY CLUSTERING

- > SELECT FEATURE (MEMBER\_ID, SALES, CATEGORY, TIME\_OF\_DAY)
- > REMOVE MISSING VALUE
- > GROUP BY MEMBER\_ID (AGGREGATE FUNCTION -> SALES:SUM, CATEGORY:MAX, TIME\_OF\_DAY:MAX)
- > DROP MEMBER\_ID
- > ONE-HOT ENCODING
- > MIN MAX SCALER

#	Column	Non-Null Count
0	sales	1036 non-null
1	category_Coffee	1036 non-null
2	category_Coffee (Bottle)	1036 non-null
3	category_Honey1	1036 non-null
4	category_Juice	1036 non-null
5	category_Salad	1036 non-null
6	category_Salmon Steak	1036 non-null
7	category_Soup	1036 non-null
8	category_Tea	1036 non-null
9	time_of_day_0_Morning	1036 non-null
10	time_of_day_1_Noon	1036 non-null
11	time_of_day_2_Afternoon	1036 non-null
12	time_of_day_3_Evening	1036 non-null

# EXPERIMENT 2

## MENU CATEGORIES AND TIME\_OF\_DAY CLUSTERING



K = 6

WCSS SCORE = 340.2785 SILHOUETTE SCORE = 0.6687

### PURPLE CLUSTER

CLUSTER SIZE: 114  
AFTER NOON: 114  
MEAN SALES: 172.83  
SD SALES: 127.26  
COFFEE: 19  
COFFEE (BOTTLE): 48  
HONEY: 3  
JUICE: 10  
SALMON STEAK: 18  
SOUP: 12  
TEA: 4

### YELLOW CLUSTER

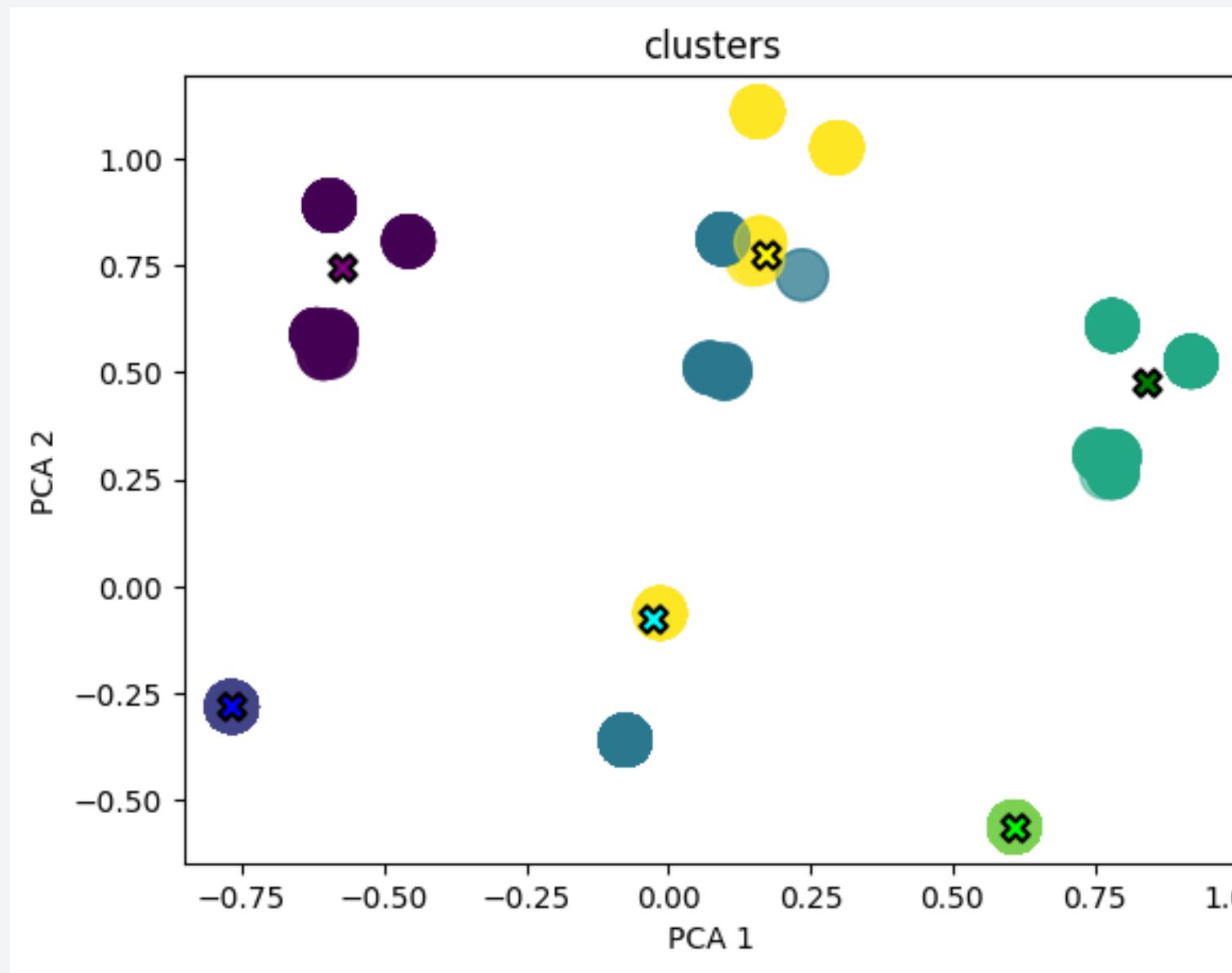
CLUSTER SIZE: 105  
MORNING: 105  
MEAN SALES: 141.02  
SD SALES: 98.59  
COFFEE: 41  
COFFEE (BOTTLE): 30  
HONEY: 2  
JUICE: 11  
SALAD: 24  
SOUP: 5  
TEA: 2

### BLUE CLUSTER

CLUSTER SIZE: 263  
AFTER NOON: 263  
MEAN SALES: 161.92  
SD SALES: 97.45  
SALAD: 263

# EXPERIMENT 2

## MENU CATEGORIES AND TIME\_OF\_DAY CLUSTERING



WCSS SCORE = 340.2785 SILHOUETTE SCORE = 0.6687

### CYAN CLUSTER

CLUSTER SIZE: 184  
EVENING: 184  
MEAN SALES: 208.64  
SD SALES: 197.89  
COFFEE: 2  
COFFEE (BOTTLE): 25  
JUICE: 9  
SALAD: 133  
SALMON STEAK: 11  
SOUP: 4

### GREEN CLUSTER

CLUSTER SIZE: 124  
NOON: 124  
MEAN SALES: 167.01  
SD SALES: 181.70  
COFFEE: 56  
COFFEE (BOTTLE): 31  
HONEY: 6  
JUICE: 11  
SALMON STEAK: 6  
SOUP: 13  
TEA: 1

### LIME CLUSTER

CLUSTER SIZE: 246  
NOON: 246  
MEAN SALES: 159.88  
SD SALES: 111.86  
SALAD: 246

# EXPERIMENT 3

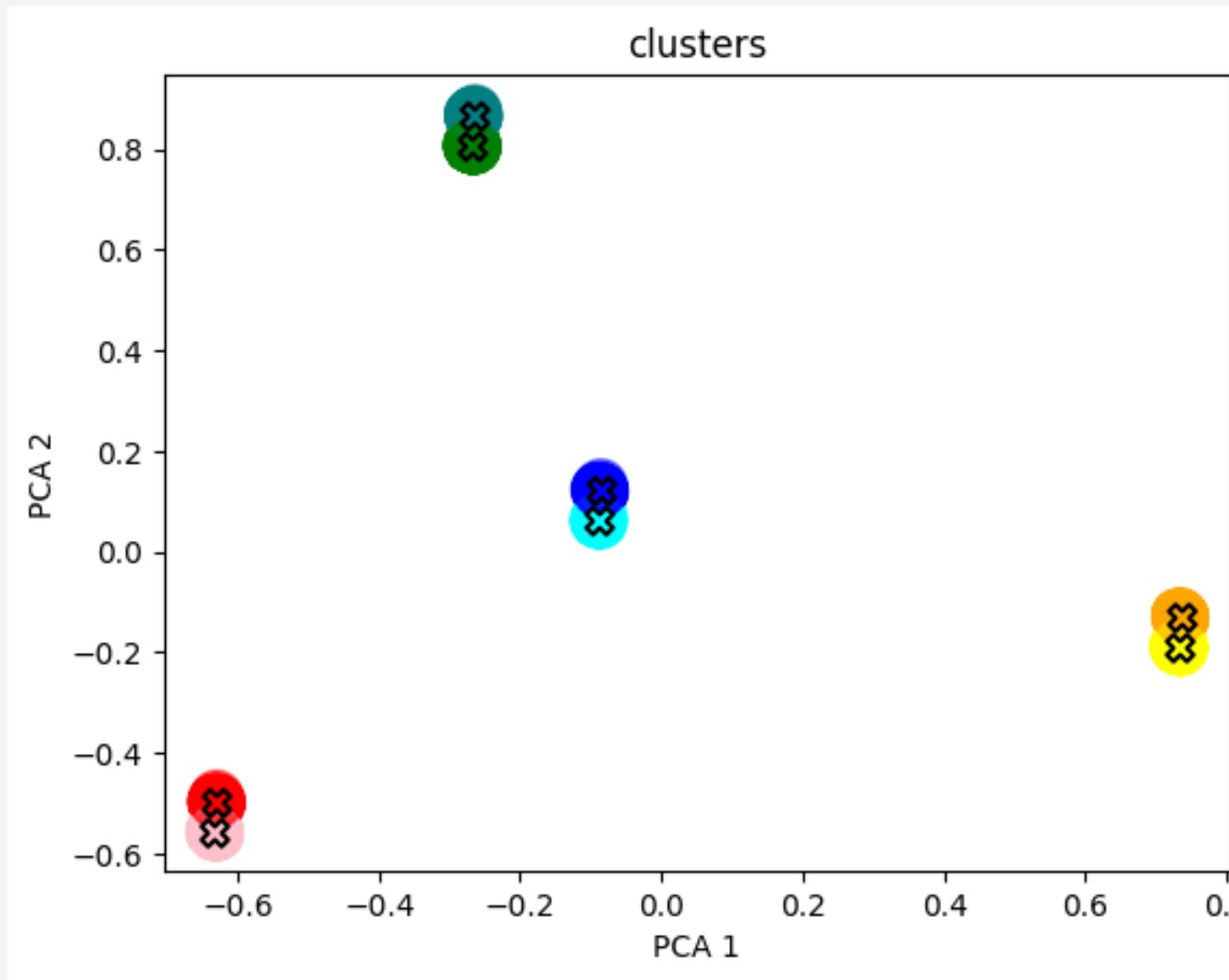
## CUSTOMER CHARACTERISTICS CLUSTERING

- > SELECT FEATURE (MEMBER\_ID, SALES, AGE, SEX, LOCATION)
- > REMOVE MISSING VALUE
- > GROUP BY MEMBER\_ID (AGGREGATE FUNCTION -> SALES:SUM, AGE:MAX, SEX:MAX, LOCATON:MAX)
- > DROP MEMBER\_ID
- > ONE-HOT ENCODING
- > MIN MAX SCALER

#	Column	Non-Null Count
0	sales	910 non-null
1	sex	910 non-null
2	age	910 non-null
3	location_East	910 non-null
4	location_North	910 non-null
5	location_South	910 non-null
6	location_West	910 non-null

# EXPERIMENT 3

## CUSTOMER CHARACTERISTICS CLUSTERING



K = 8

WCSS SCORE = 51.1088    SILHOUETTE SCORE = 0.7340

**RED CLUSTER:** SOUTH MALE → SIZE: 108, SALES: 159.58

**PINK CLUSTER:** SOUTH FEMALE → SIZE: 162, SALES: 153.88

**TEAL CLUSTER:** WEST MALE → SIZE: 93, SALES: 172.88

**GREEN CLUSTER:** WEST FEMALE → SIZE: 136, SALES: 174.70

**BLUE CLUSTER:** EAST MALE → SIZE: 41, SALES: 180.57

**CYAN CLUSTER:** EAST FEMALE → SIZE: 44, SALES: 184.99

**ORANGE CLUSTER:** NORTH MALE → SIZE: 132, SALES: 149.07

**YELLOW CLUSTER:** NORTH FEMALE → SIZE: 194 SALES: 176.40

MEAN AGE OF ALL CLUSTER: 41-43

EAST HAVE FEWEST CUSTOMER BUT HIGHT SALES

NORTH HAVE MORE CUSTOMER THAN OTHER LOCATION

NORTH MALE HAVE FEWEST SALES

FEMALE CUSTOMER MORE THAN MALE CUSTOMER IN EVERY LOCATION

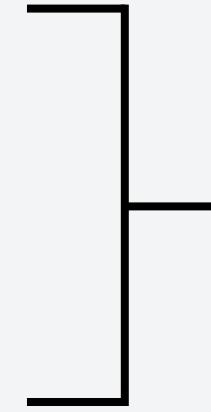
# SUMMARY

SHOULD HAVE CAMPAIGN OR PROMOTION FOR ...

NON SALAD MENU

EVENING CUSTOMER

EAST CUSTOEMER



ORDER AMOUNT IS SMALL BUT SALES ARE HIGH.

**THANK YOU**

