Markov Chain

Super Markets

Markov Chain Monte Carlo (MCMC)
Simulations on “DOODL Supermarket data”

Introduction

How does the behavior of different customer types affect traffic and congestion levels in supermarkets?

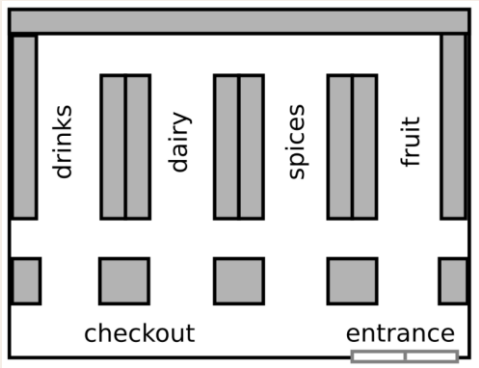
Examining the influence of customer age on shopping patterns and store occupancy.

The DATA

timestamp	customer_no	location
2019-09-02 07:03:00	1	dairy
2019-09-02 07:03:00	2	dairy
2019-09-02 07:04:00	3	dairy
2019-09-02 07:04:00	4	dairy
2019-09-02 07:04:00	5	spices
...	...	...
2019-09-02 21:49:00	1442	checkout
2019-09-02 21:49:00	1444	checkout
2019-09-02 21:49:00	1445	dairy
2019-09-02 21:50:00	1446	dairy
2019-09-02 21:50:00	1447	fruit

4884 rows × 2 columns

5 different data sets
from weekdays



[1]

1 fruit

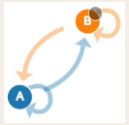
2 fruit

3 drinks

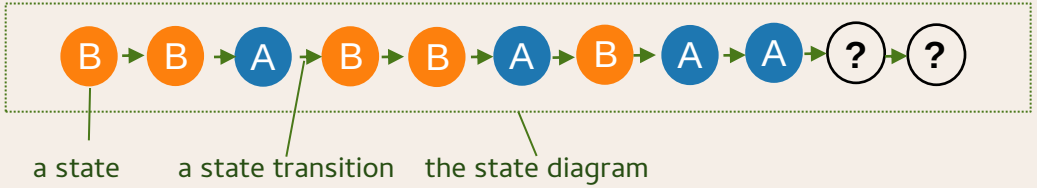
4 spices

5 checkout

Markov Chain Theory



[2]

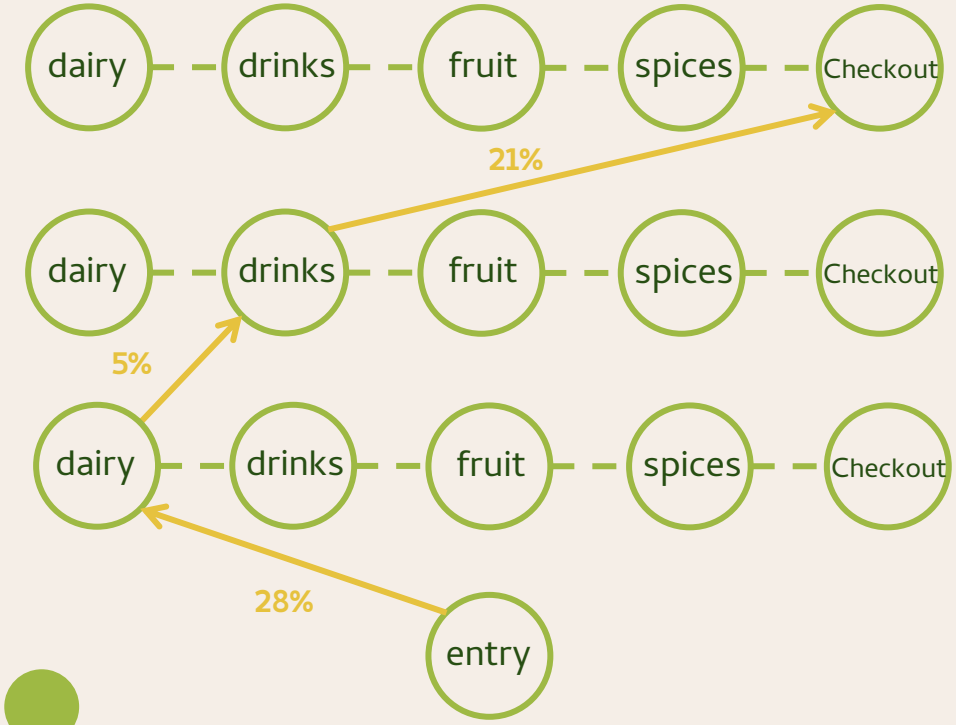


Properties of MC Simulations:

- Markov (Memoryless) Property
- Stationarity Property
- Ergodicity property
- Transition matrix property

	A	B
A	0.61	0.39
B	0.27	0.73

Markov Chain Theory



next_location	checkout	dairy	drinks	fruit	spices
location					
dairy	0.102697	0.738706	0.058134	0.049478	0.050986
drinks	0.215334	0.010880	0.599199	0.087755	0.086832
fruit	0.200564	0.095428	0.054564	0.599029	0.050415
spices	0.149613	0.191839	0.161948	0.090305	0.406294
entry	0.000000	0.287576	0.153526	0.377435	0.181464

Assumptions of MC models:

- there is a finite state space
- a state only depends on the previous state
- no hidden states (all states are known and observable)
- discrete time (time is measured in discrete steps)
- time-homogenous (transition probabilities do not change over time)

Data Preprocessing

timestamp	customer_no	location	day	customer_id
2019-09-02 07:05:00	1	checkout	monday	monday_1
2019-09-02 07:05:00	5	checkout	monday	monday_5
2019-09-02 07:06:00	2	checkout	monday	monday_2
2019-09-02 07:06:00	3	checkout	monday	monday_3
2019-09-02 07:07:00	8	checkout	monday	monday_8
...	...	...	...	...
2019-09-06 21:47:00	1502	checkout	friday	friday_1502
2019-09-06 21:48:00	1501	checkout	friday	friday_1501
2019-09-06 21:48:00	1504	checkout	friday	friday_1504
2019-09-06 21:50:00	1507	checkout	friday	friday_1507
2019-09-06 21:50:00	1508	checkout	friday	friday_1508

7417 rows × 4 columns

1.Monday to Friday

customer_id	timestamp	location	next_location
friday_1	2019-09-06 07:00:00	dairy	dairy
friday_1	2019-09-06 07:01:00	dairy	dairy
friday_1	2019-09-06 07:02:00	dairy	dairy
friday_1	2019-09-06 07:03:00	dairy	spices
friday_1	2019-09-06 07:04:00	spices	checkout
...	...	...	...
wednesday_998	2019-09-04 16:57:00	fruit	fruit
wednesday_998	2019-09-04 16:58:00	fruit	checkout
wednesday_998	2019-09-04 16:59:00	checkout	fruit
wednesday_999	2019-09-04 16:53:00	fruit	checkout
wednesday_999	2019-09-04 16:54:00	checkout	NaN

54192 rows × 3 columns

2.After cleaning and delay function

location	next_location	checkout	dairy	drinks	fruit	spices
dairy		0.102697	0.738706	0.058134	0.049478	0.050986
drinks		0.215334	0.010880	0.599199	0.087755	0.086832
fruit		0.200564	0.095428	0.054564	0.599029	0.050415
spices		0.149613	0.191839	0.161948	0.090305	0.406294

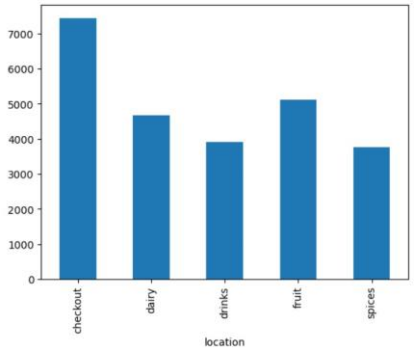
3.Transition Matrix (TM)

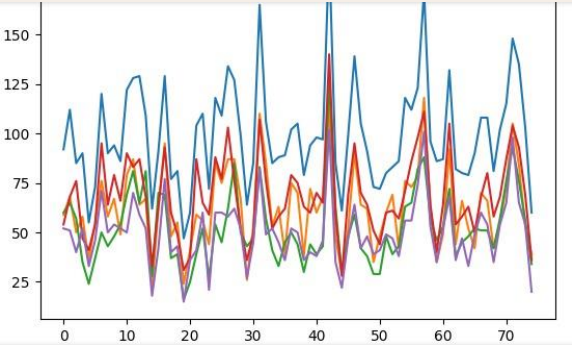
location	next_location	checkout	dairy	drinks	fruit	spices
dairy		0.102697	0.738706	0.058134	0.049478	0.050986
drinks		0.215334	0.010880	0.599199	0.087755	0.086832
fruit		0.200564	0.095428	0.054564	0.599029	0.050415
spices		0.149613	0.191839	0.161948	0.090305	0.406294
entry		0.000000	0.287576	0.153526	0.377435	0.181464

1. Combine all datasets
2. Data cleaning
3. Calculate TM
4. Organize TM
5. Customer Profile Specific TMs

Modeling Customer Traffic in the Supermarket

1. Exploratory Analysis
2. TM Probabilities
3. MCMC Simulation
4. CSV Output





next_location	checkout	dairy	drinks	fruit	spices
location					
dairy	0.102697	0.738706	0.058134	0.049478	0.050986
drinks	0.215334	0.010880	0.599199	0.087755	0.086832
fruit	0.200564	0.095428	0.054564	0.599029	0.050415
spices	0.149613	0.191839	0.161948	0.090305	0.406294
entry	0.000000	0.287576	0.153526	0.377435	0.181464

Junior = I* 0.9
Senior = I / 0.9

Future Directions

01

Collect more data

Get better TM and make customer profiling on solid base.

02

Introduce more types and perks

In model only “junior” and “senior”. Add more specific individual customer types: vegan, trash eater, thirsty throat etc.

03

Compare results & optimize

Compare simulation results with previous data and control in future.

04

Add more functions

Model other qualities, such as distance covered by a customer, bottle necks, stolen goods, etc.

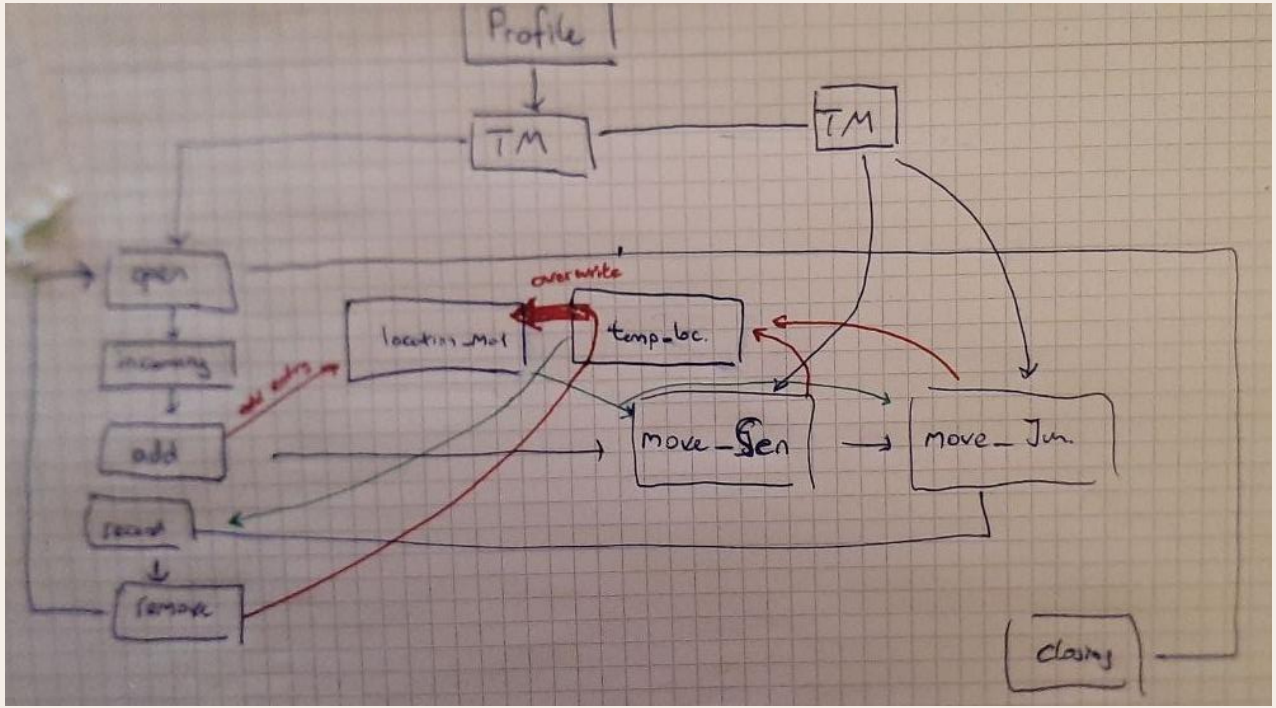
THANKS!

Do you have any questions?

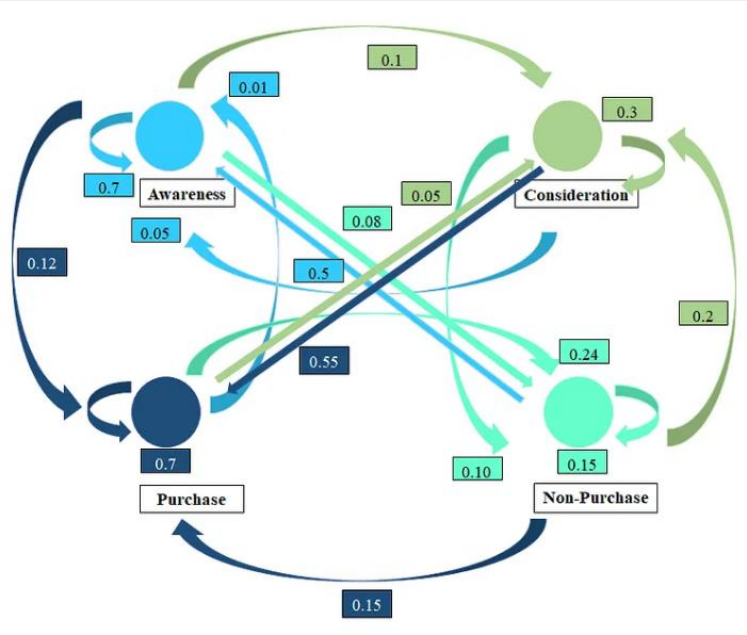
References:

[1] Spiced Academy (https://www.spiced-academy.com/)
[2] Victor Powell (https://setosa.io/ev/markov-chains/)

Code Flow Chart



Marketing Analytics through MC



https://towardsdatascience.com/marketing-analytics-through-markov-chain-a9c7357da2e8