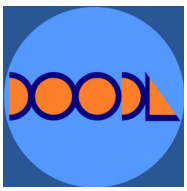


DOODL

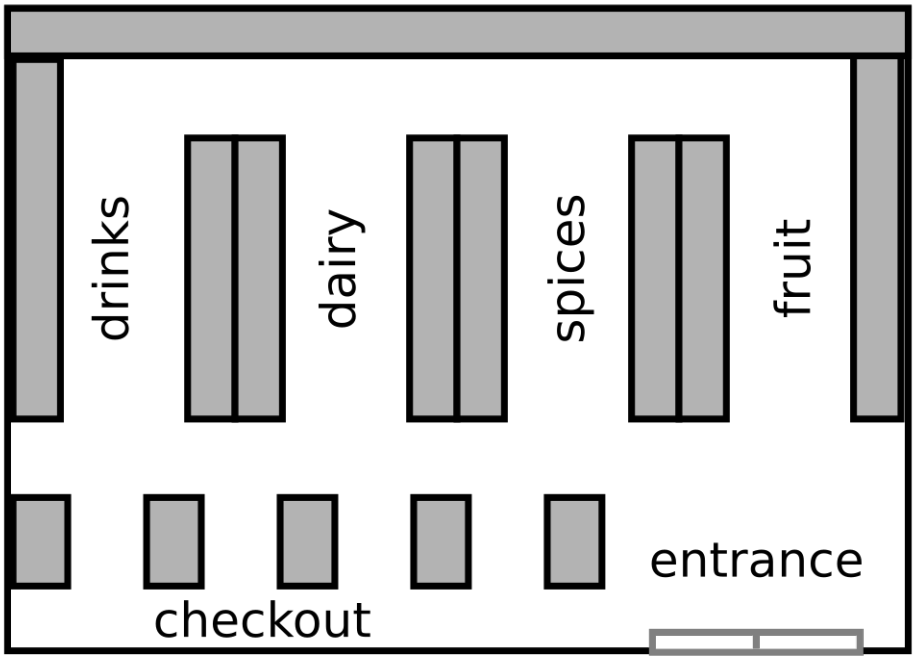
Quality brought close to you

Customer Flow – Project Description

At the DOODL Supermarket chain we would like to understand our customers better in order to optimize the layout, staffing and service of our supermarkets. We would like you and your team to model the way customers move through a representative shop.

1. Market

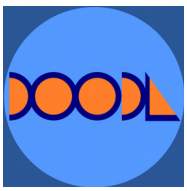
We are considering the following supermarket:



We have divided this DOODL supermarket into four areas: *fruit*, *spices*, *dairy* and *drinks*. Customers move between these areas freely. Sooner or later, they will enter the *checkout* area. Once they do, they are considered to have left the shop.

2. Customer Data

We have anonymously tracked the movement of all DOODL customers during a five-day week. In the files *monday.csv* to *friday.csv* you find tabular data indicating where customers spent their time. The data is complete.



DOODL

Quality brought close to you

3. Project Planning

This project may be complicated to complete on your own. Even in a team, you may want to focus on a few of the tasks. Please do the following:

- Form a team of 2-3 people
- Read the instructions carefully
- Collect tasks, questions and goals on a small task board using post-it's.
- Prioritize the tasks

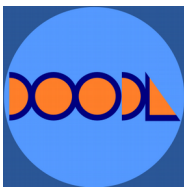
4. Data Analysis

Our sales department is interested in a summary of the collected data. Please generate a report including numbers and diagrams. Note that your audience are not data scientists, so take care to prepare insights that are as clear as possible. We are interested in the following:

- Calculate the total number of customers in each section
- Calculate the total number of customers in each section over time
- Display the number of customers at checkout over time
- Calculate the time each customer spent in the market
- Calculate the total number of customers present in the supermarket over time.
- Our business managers think that the first section customers visit follows a different pattern than the following ones. Plot the distribution of customers of their first visited section versus following sections (treat all sections visited after the first as "following").
- Estimate the total revenue for a customer value using the following table:

section	**revenue per minute**
fruit	4€
spices	3€
dairy	5€
drinks	6€

- Which is the most profitable section according to your data?



DOODL

Quality brought close to you

5. Simulate a single customer

Write a program that uses a transition probability matrix to simulate the journey of a single customer through the market. Use a Markov model to represent the state of a customer. Use one-minute time intervals for the transitions. Once a customer reaches the checkout, consider them *“churned”* – do not simulate them any longer.

- Print all state changes
- Set the transition probabilities manually
- Later add the probabilities extracted from the data
- Extend the model to use separate probabilities for the first location
- Implement the customer as a class

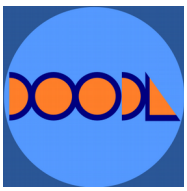
6. Transition Probabilities

We would like to analyze how customers switch between sections of the supermarket. Calculate and visualize the probability of transitions from section A to B by counting all observed transitions.

E.g. if a customer was in the *fruit* section, later in the *spices* section, and went back to *fruit*, we observe two transitions: *fruit* → *spices* and *spices* → *fruit* .

The checkout is a special terminal state, from which customers cannot leave.

- Draw a state diagram
- Display the transition probability matrix
- Visualize the probabilities using an aproppriate library (consider NetworkX or PyGraphViz)



DOODL

Quality brought close to you

7. Simulate a population

Extend the program further to allow for large-scale simulations. Instead of a single customer, simulate multiple customers.

- Run the simulation multiple times for a typical day
- Plot the number of customers in each area over time
- Calculate an idealized population from the transition probabilities
- Compare the average time customers spend inside the supermarket between simulated and real customers. What similarities or differences do you observe?

8. Customer Segmentation

Can you find any relevant customer subgroups in the data? If yes, it would be great to determine separate transition probabilities for these groups.

- Extend the program to represent the groups with their respective probabilities
- How would you describe the found subgroups in plain language?
- Discuss whether inheritance helps to implement the customer segments

9. Checkout Queues

Through surveys we have found that the biggest factor for customer satisfaction is the length of queues at the checkout. On average, a checkout worker has a 50% chance to complete the checkout of a single customer within one minute. In the next minute, there is again a 50% chance and so on.

- What is the name of the underlying distribution?
- Extend the simulation to model the checkout process using the above probability. Manually set the number of checkout workers. Output the number of waitning customers over time.
- How would you estimate the number of necessary checkout workers to make sure the queue never gets longer than 3 customers?

DOODL

Quality brought close to you

10. Visualization

To educate our staff about customer behavior it would be great if you could create an interactive visualization showing how many customers are present in the areas of a shop. Ideally, this would be an animated or interactive diagram plotting the customers into the shop layout, like this: