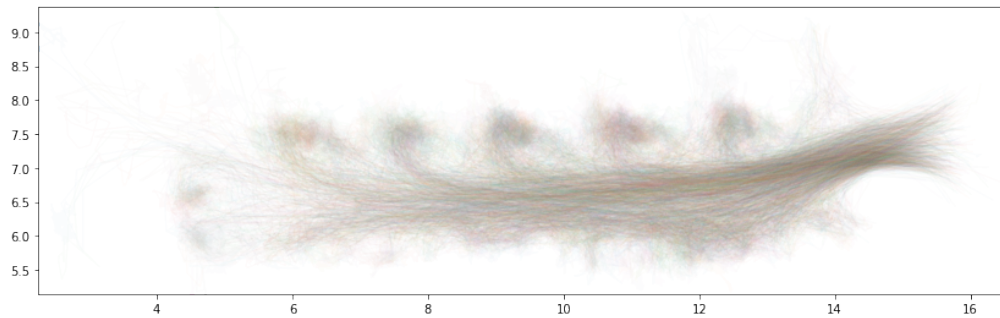


Riešenie úlohy

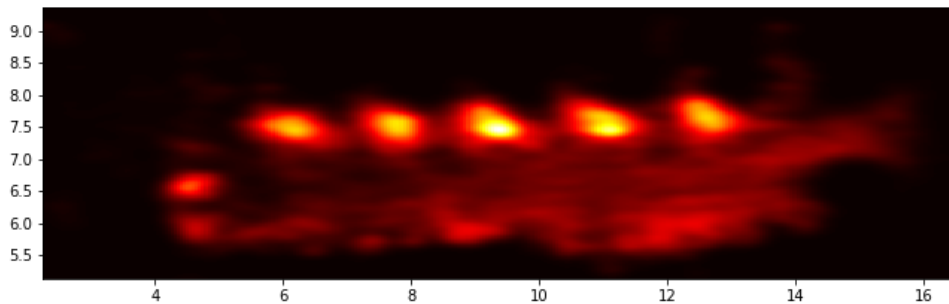
Martin Gažo

1 Vizualizácie

Dve najdôležitejšie vizualizácie sú v obrázkoch 1 a 2, kde je znázornený pohyb návštevníkov po obchode. Tieto obrázky som používal neskôr pri vytváraní metód.



Obr. 1: Znázornenie všetkých trajektórií



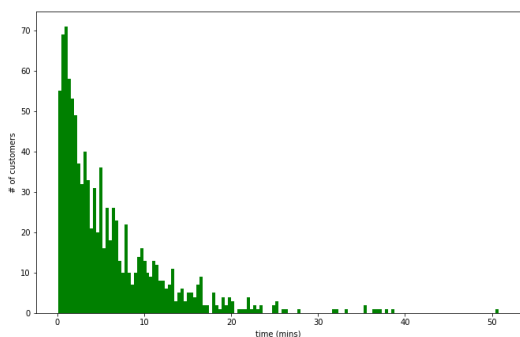
Obr. 2: Heatmapa, ktorá znázorňuje koľko času strávia customers v daných častiach. Každý návštevník má rovnakú váhu.

V Jupyter Notebooku sa nachádza aj funkcia, ktorá vytvorí animáciu pohybu customera po obchode.

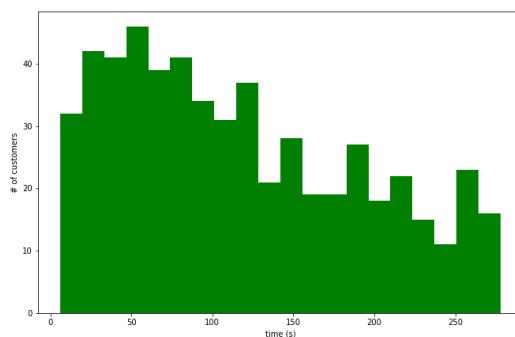
Ďalej som sa ešte pozrel na čas strávený zákazníkmi v obchode v podobne histogramu na obrázkoch 3. Z neho je okamžite vidieť, že viacero zákazníkov odišlo z prevádzky už po relatívne krátkom čase.

2 Metódy

V tomto probléme nemáme k dispozícii správne výsledky na tréning, nevieme presnú dispozíciu obchodu a nevieme ani to, ako to presne v obchode funguje. Preto musíme začať podmienkou podľa ktorej budeme vedieť klasifikovať zákazníkov ručne. Ak sa pozrieme na heatmapu na obrázku 2, je vidieť, že miesta kde sú zákazníci pravdepodobne obsluhovaní sa nachádzajú v najjasnejších miestach. Preto za obslužených návštevníkov budeme považovať tých, ktorí strávia dostatočný čas v tejto oblasti.



(a) Celý dataset



(b) Priblíženie len na zákazníkov, čo nestrávili v obode dlho

Obr. 3: Histogram času stráveného v prevádzke.

Model 1 (jednoduchý)

Tento model je založený assumption, že customers budú čakať v rade na inom mieste, ako tom, kde budú obslúžení a keď vstúpia do územia so stoličkami budú určite obslúžení. Pozerám sa teda len na to, či zákazník niekedy vstúpi do územia, kde sú stoličky, a tých čo vstúpia beriem ako za obslúžených.

Jedným zo základných problémov tohto modelu je fakt, že customers môžu arbitrárne definovaným územím obsluhovania len prejsť, resp. nestráviť tam veľa času a aj tak budú zaznamenaní. Preto výsledok z tohto modelu bude len dolný odhad percenta neobslúžených zákazníkov.

Model 2

V tomto modeli určujem obslúženie zákazníkov tak, že sa pozerám na to, kde ľudia stoja dlhší čas, a potom, podľa toho či to je pri stoličkách alebo v rade, určím, či boli obslúžení.

Na určenie toho, či sa customer pohybuje používam metódu `Model2.is_moving`, ktorá spočíta posunutie za niekoľkokondový časový interval v každom bode trajektórie a podľa jeho veľkosti určí, či sa customer pohybuje.¹

Model 3

V tomto modeli zisťujem, koľko času customers strávili v zóne s stoličkami a podľa dĺžky tohto času ich klasifikujem. Časová hranica bola určená najlepším F1 score s 2. modelom ² a vyšla na 13 sekúnd, čo vyzerá rozumne (t.j. ľudia, čo boli pri stoličkách kratšie ako 13 sekúnd pravdepodobne neboli reálne obslúžení)

Model 4

V tomto modeli skúmam heatmaps vytvorené len od jednotlivých customerov. Algoritmus nájde maximálnu heatmapu a zistí, kde sa nachádzajú tie najsilnejšie. Ak sú nejaké v zóne so stolmi so stoličkami, znamená to, že zákazník bol obslúžený. Narozdiel od metódy 3 tu nesčítavam státie pri dvoch rôznych stoloch, ale beriem tieto časy oddelene.

Ďalšie nápady

Tu stručne popisujem ďalšie nápady, ktoré som nestihol implementovať:

¹Detailnejšie, spočítam rolling averages zložiek rýchlosti, z ktorých potom zistím veľkosť rýchlosti.

²To samozrejme znamená, že navzájom sú tieto modely korelované, ale keďže nemáme informácie o tom ako dlho trvá vybavenie zákazníka je to aspoň nejaká metóda na určenie.

- Sledovanie všetkých ľudí naraz a spracovanie vzájomnej dynamiky – napríklad by sme sledovali aktuálnu veľkosť rady a obsadenosť stolov a na základe toho určovali, či ľudia, čo sa otočili naozaj odišli kôli rade.
- V dátach bolo vidieť, že niektorí návštevníci sa pohybovali po predajni z jednej zóny do druhej. Keďže som si nebol úplne istý, čo takéto správanie môže znamenať som ich nijako špeciálne neriešil, ale je jasné, že by si zaslúžili hlbšie preskúmanie.
- *Hypotetický*: Viacero metód by sa dalo spojiť do jednej (napríklad váhovaným priemerovaním kde by sa váhy nastavovali podľa toho, o aký typ trajektórie sa jedná). Na implementáciu by však bolo treba podrobne sledovať kde sa modeli rozporujú a ideálne to porovnať so správnym riešením (ktoré nemáme ale teoreticky by sa dalo v praxi získať manuálne pri časti dát).

3 Výsledky

Výsledky na skúmenú otázku (percento ľudí, ktorí vyjdú neobslúžení) podľa modelov sú zhrnuté v tabuľke 1.

Model	Výsledok
1.	15.5%
2.	22.4%
3.	24.7%
4.	23.5%

Tabuľka 1: Výsledky podľa modelu

Ako sme už povedali, 1. model je len dolný odhad, čiže výsledná hodnota sa bude pohybovať niekde okolo 23%.