



Data mining for small business

Nai-Tan Chang, Yang Fu
MKT 568
Apr 27, 2019

Imagining owning a bakery shop



Generate 1st-hand transaction data daily.

They could be excellent source for some data mining.



A sneak peek of the bakery transaction dataset

	Date	Time	Transaction	Item
1	2016-10-30	09:58:11	1	Bread
2	2016-10-30	10:05:34	2	Scandinavian
3	2016-10-30	10:05:34	2	Scandinavian
4	2016-10-30	10:07:57	3	Hot chocolate
5	2016-10-30	10:07:57	3	Jam
6	2016-10-30	10:07:57	3	Cookies

Data entries: 21,293

- Date: 2016-10-30 to 2017-04-09
- Time: time of transaction
- Transaction: 9684 unique IDs
- Item: 95 unique products

Source of dataset: Transactions from a bakery

<https://www.kaggle.com/sulmansarwar/transactions-from-a-bakery>



Business questions:

Product Performance

- What are our popular products?
- How is the growth of our core products?

Sales Pattern

- Is there a pattern when certain products were heavily purchased around certain time?

Market Basket

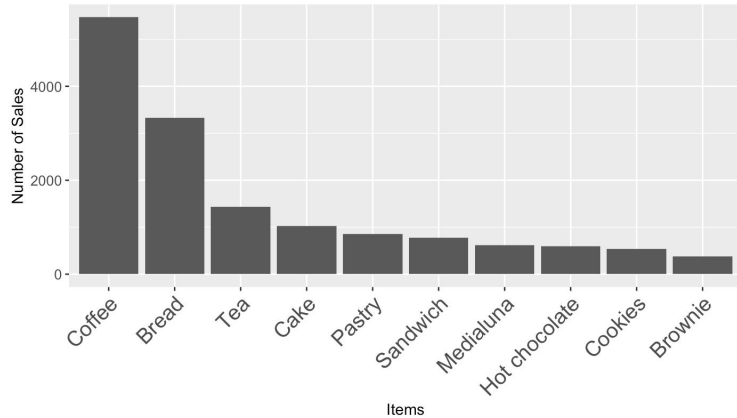
- What kind of products are more likely to be sold together?
- What's is our growth opportunity?



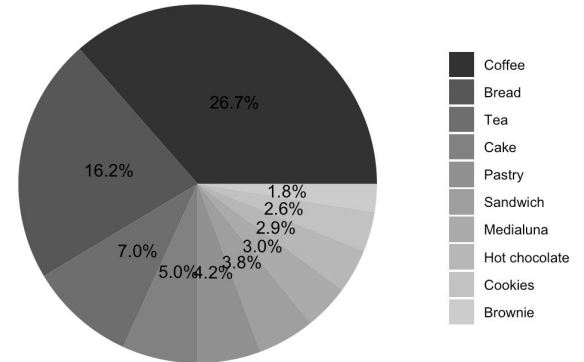
What are our popular products?
How is the growth of our core products?

Product Performance

Top 10 Sales by Number



Top 10 Sales by %



Picture sources

<https://www.telegraph.co.uk/food-and-drink/features/how-coffee-affects-your-body-from-the-first-sip/>

<https://leitesculinaria.com/93789/recipes-5-minute-artisan-bread.html>

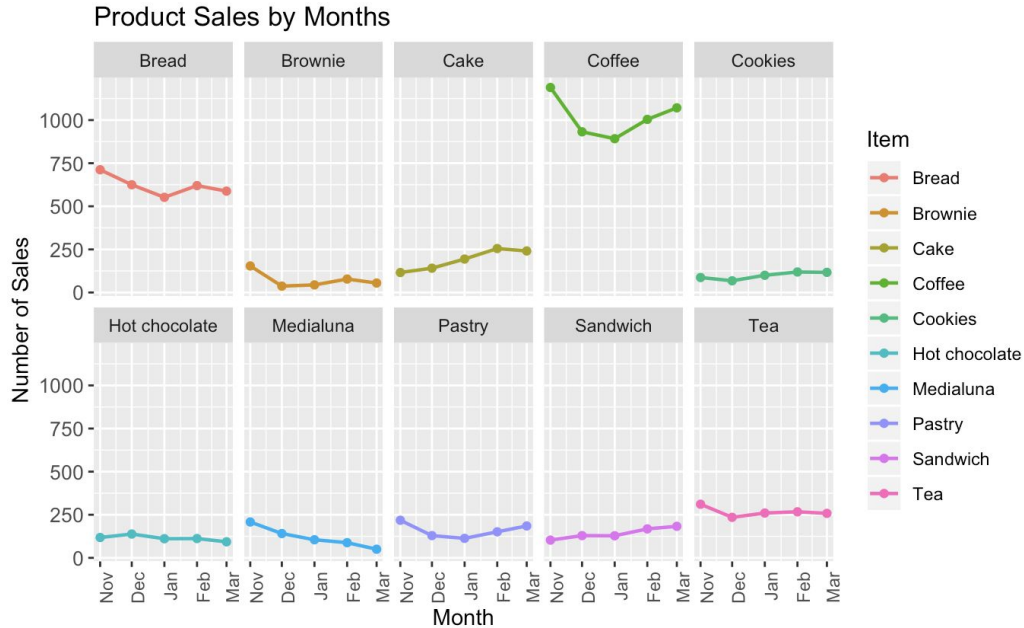
<https://uniquelylocal.co.uk/experiences/afternoon-tea/>

<https://www.quericavida.com/recetas/medialunas-rellenas-con-chocolate/7fc0912a-c33f-4bc9-93aa-2e3a819ec5bf>



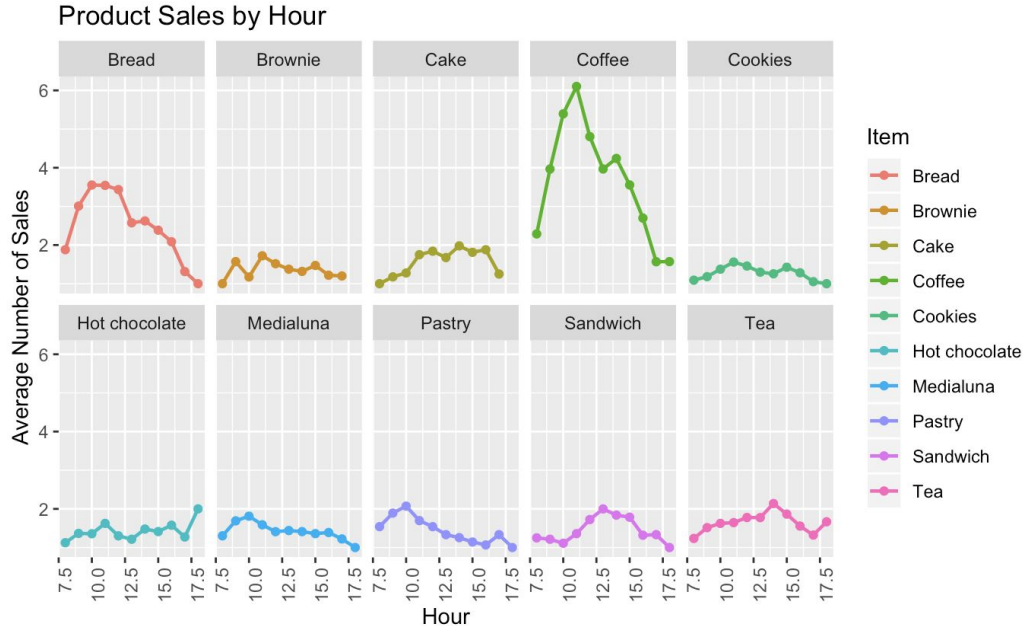
Is there a pattern when certain products were heavily purchased around certain time?

Sales Pattern by Month



- Most of the sales stay flat.
- Slight drop between Dec and Jan (holiday season).
- Some alarming performances. (Medialuna)

Sales Pattern by Hour



Customers tend to buy some products at a specific time during a day.

Morning - coffee, bread, pastry

Noon - sandwich

Afternoon - coffee, tea, bread

Late afternoon - Hot chocolate



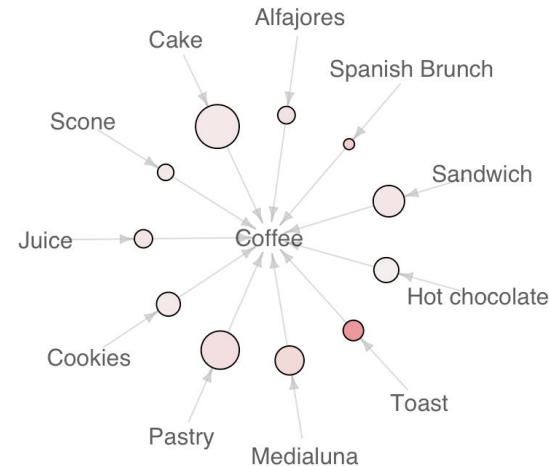
**What kind of products are more likely
to be sold together?**

Market Basket analysis - with coffee

	lhs	rhs	support	confidence	lift	count
[1]	{Spanish Brunch}	=> {Coffee}	0.01088105	0.5988372	1.252451	103
[2]	{Toast}	=> {Coffee}	0.02366364	0.7044025	1.473238	224
[3]	{Scone}	=> {Coffee}	0.01806465	0.5229358	1.093705	171
[4]	{Alfajores}	=> {Coffee}	0.01964927	0.5406977	1.130854	186
[5]	{Juice}	=> {Coffee}	0.02060004	0.5357143	1.120431	195
[6]	{Cookies}	=> {Coffee}	0.02820621	0.5194553	1.086426	267
[7]	{Medialuna}	=> {Coffee}	0.03507289	0.5684932	1.188987	332
[8]	{Hot chocolate}	=> {Coffee}	0.02957955	0.5072464	1.060891	280
[9]	{Sandwich}	=> {Coffee}	0.03824213	0.5323529	1.113401	362
[10]	{Pastry}	=> {Coffee}	0.04743292	0.5515971	1.153650	449
[11]	{Cake}	=> {Coffee}	0.05472216	0.5269583	1.102118	518

- Toast and coffee are **most likely** to be bought together.
- Cake/pastry and coffee were **most frequently** bought together.

Graph for 11 rules

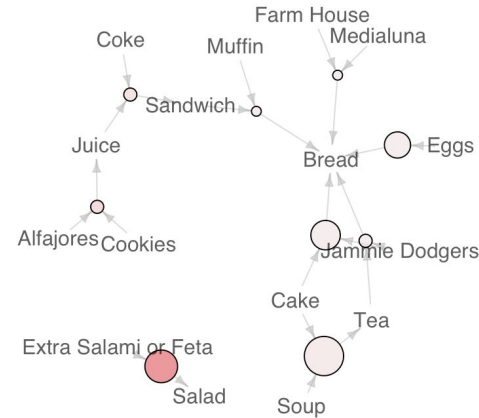


Market Basket analysis - without coffee

	lhs	rhs	support	confidence	lift	count
[1]	{Eggs}	=> {Bread}	0.001692252	0.5000000	1.336942	14
[2]	{Extra Salami or Feta}	=> {Salad}	0.001934002	0.4210526	35.185540	16
[3]	{Cake, Jammie Dodgers}	=> {Bread}	0.001813127	0.5172414	1.383044	15
[4]	{Jammie Dodgers, Tea}	=> {Bread}	0.001208751	0.4166667	1.114119	10
[5]	{Coke, Juice}	=> {Sandwich}	0.001208751	0.4761905	5.793417	10
[6]	{Farm House, Medialuna}	=> {Bread}	0.001087876	0.4090909	1.093862	9
[7]	{Cake, Soup}	=> {Tea}	0.002175752	0.4285714	2.628296	18
[8]	{Muffin, Sandwich}	=> {Bread}	0.001087876	0.4500000	1.203248	9
[9]	{Alfajores, Cookies}	=> {Juice}	0.001208751	0.4347826	9.881749	10

- Salad + Extra Salami or Feta
- Coke + Juice + Sandwiches
- Soup + Cake + Tea

Graph for 9 rules



size: support (0.001 - 0.002)
color: lift (1.094 - 35.186)



Business strategies & growth opportunities

- Best-selling products: Keep quality. Increase variety.
- Prepare and stock products more efficiently based on seasonality.
- Plan promotions before and after the holiday season to counter the Dec and Jan sales drop.
- Set up specific in-store campaign to boost sales at certain times during a day.
- Design combo menu for products that are likely and frequently sold together.



Conclusions

Even small business can utilize their own data to evaluate product performance, discover sales patterns, and gain deeper insights into their business.

Further analysis could be done if more information are available:

- Price: revenue of product.
- Cost: profitability of product..
- Customer: sales target.



Thanks!

Questions and comments?