Cognitive Systems

Monika Wysoczanska, 180817

Boris Mattijssen

Data Exploration

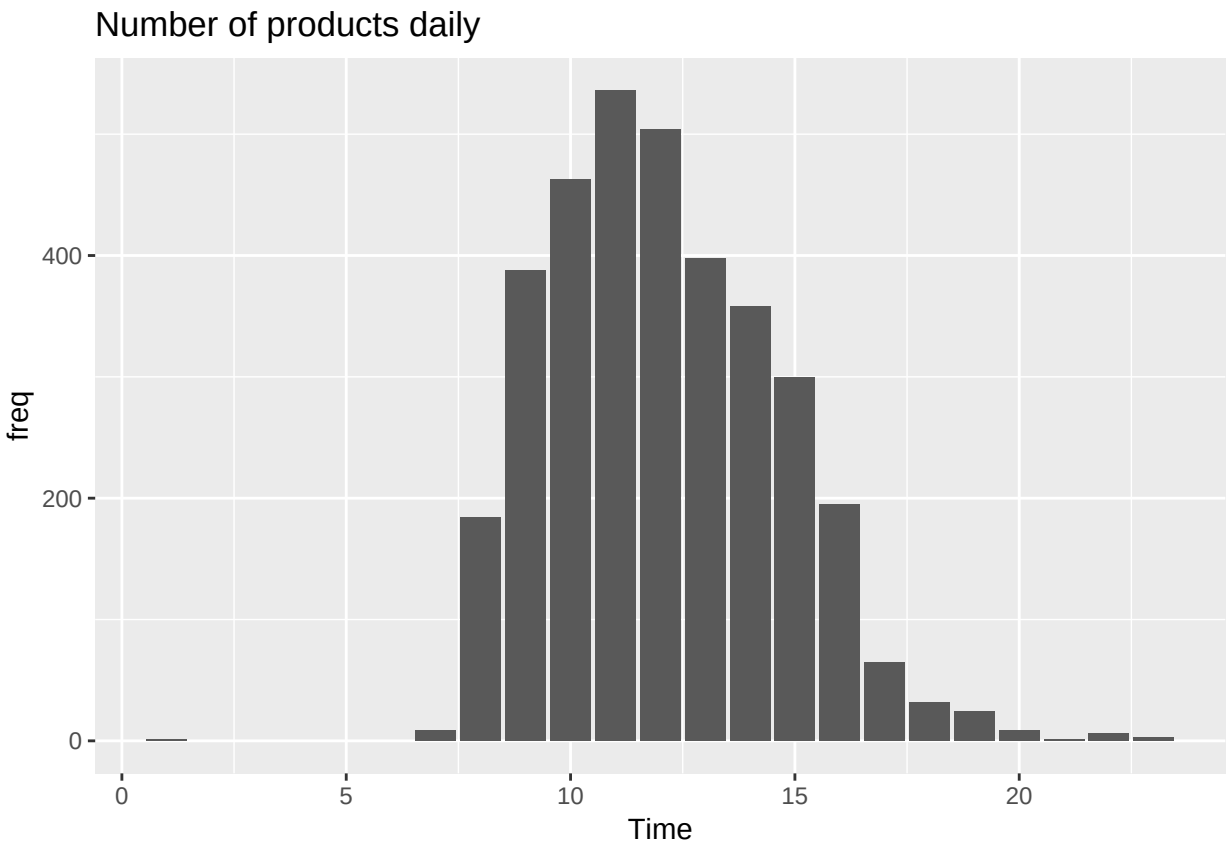
The Basket Bread dataset consists of 9684 transactions of 94 different products offered in the Bakery. Those purchases are discribed by 21293 unique observations, where each one of them represents a product of particular transaction. Note: For 786 observations there is missing information about the kind of product it represents. For each transaction we are also given the exact time of its occurance.

The average number of products per transaction is 2.23.

Now we want to investigate the distribution of purchases along the time visually.

Distribution along time of the number of products

```
## Joining by: Transaction
```

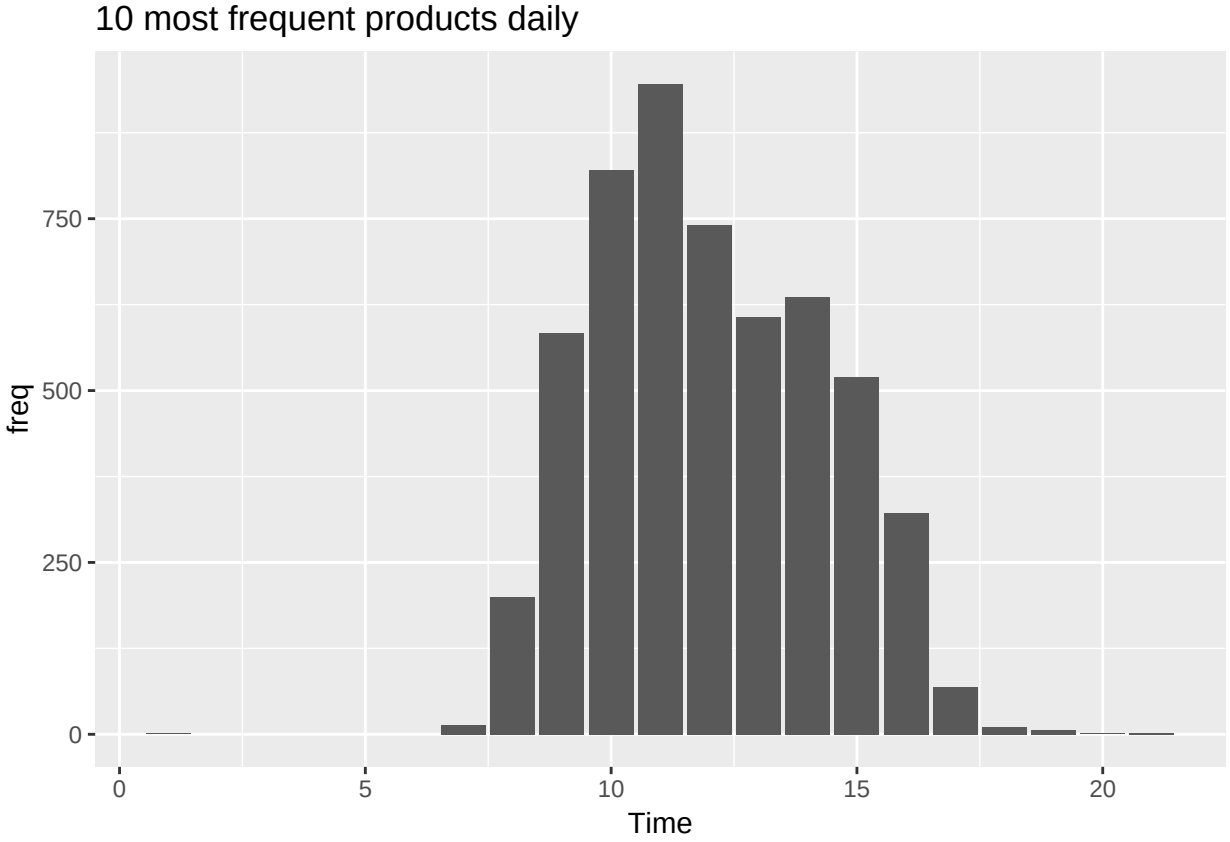


We can observe that the increase of sales starts between 8:00 and 9:00 and reaches the peak between 11:00 and 12:00. After the rush hour the number of sold products constatly decreases and after 5 pm drops significantly.

Distribution of most 10 frequent items

In order to analyze the distribution of the 10 most frequent products we removed NONE product transactions. The 10 most frequent products are given below:

x
Brownie
Cookies
Hot chocolate
Medialuna
Sandwich
Pastry
Cake
Tea
Bread
Coffee

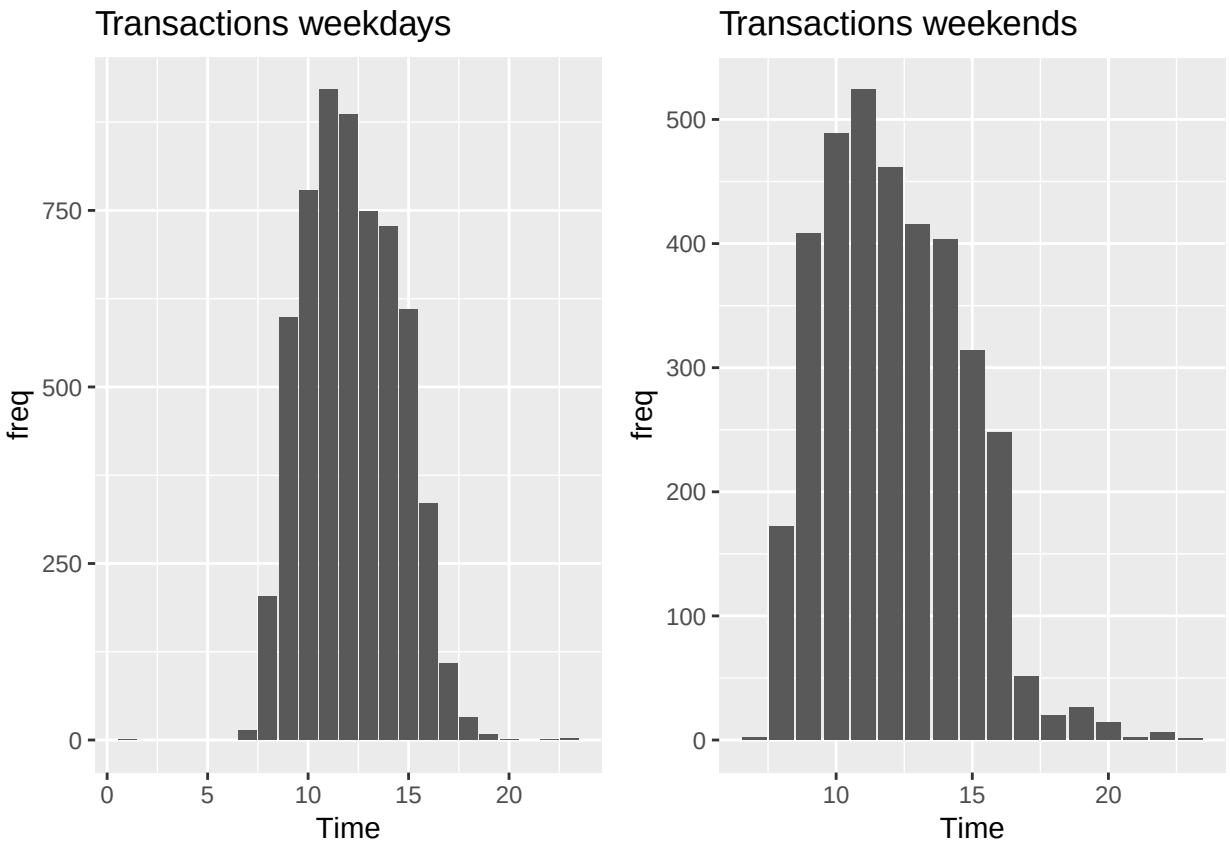


We observe the peak at the same time as general sales increase. Additionally, there is a slight increase of sales between 14:00 and 15:00.

Distribution of transaction/times

The analysis of the distribution of number of transactions over time we divided into parts, as we believe there might be differences between workdays and weekends.

Working days



The distribution of transactions over time does not differ much between working days and weekends. The 'rush hour' in both cases is between 11:00 and 12:00. However, we can see that the second popular time slot for purchases during working days is the one in the afternoon, while for the weekends it's the one in the morning.

Finding apriori rules

In order to find out which products are likely to be sold together the apriori algorithm was used. We aim at finding the group of items that are bought frequently, which means we focus on higher support. We believe that lower confidence leaves the area for improvement, as we can encourage customers by different marketing techniques.

Firstly, we investigate the working days, but before that we remove all the NONE product rows as they do not provide any relevant information in case of apriori rules.

We discovered 12 association rules with the support above 1% and confidence above 50%. The 6 of them with the highest support are presented in the table below.

```
##     lhs                rhs      support    confidence lift     count
## [1] {Cake}          => {Coffee} 0.06019221 0.5622047  1.139978 357
## [2] {Pastry}        => {Coffee} 0.05243635 0.5727440  1.161349 311
## [3] {Sandwich}      => {Coffee} 0.03793627 0.5382775  1.091461 225
## [4] {Medialuna}     => {Coffee} 0.03506997 0.5745856  1.165083 208
## [5] {Hot chocolate} => {Coffee} 0.03068622 0.5229885  1.060460 182
## [6] {Cookies}       => {Coffee} 0.03034901 0.5187320  1.051829 180
```

	lhs		rhs	support	confidence	lift	count
[1]	{Cake}	=>	{Coffee}	0.0601922	0.5622047	1.139978	357
[2]	{Pastry}	=>	{Coffee}	0.0524364	0.5727440	1.161349	311
[3]	{Sandwich}	=>	{Coffee}	0.0379363	0.5382775	1.091461	225
[4]	{Medialuna}	=>	{Coffee}	0.0350700	0.5745856	1.165083	208
[5]	{Hot chocolate}	=>	{Coffee}	0.0306862	0.5229885	1.060460	182
[6]	{Cookies}	=>	{Coffee}	0.0303490	0.5187320	1.051829	180

We conduct the analysis using the same apriori parameters for the Weekends transactions. The 6 most frequent set of products for weekends are given in the table below.

```
##     lhs            rhs      support    confidence lift     count
## [1] {Pastry}    => {Coffee} 0.03933220 0.5110294  1.126624 139
## [2] {Sandwich}  => {Coffee} 0.03876627 0.5229008  1.152796 137
## [3] {Medialuna} => {Coffee} 0.03537068 0.5605381  1.235771 125
## [4] {Brownie}   => {Coffee} 0.02801358 0.5238095  1.154799  99
## [5] {Cookies}   => {Coffee} 0.02461800 0.5178571  1.141676  87
## [6] {Toast}     => {Coffee} 0.02263724 0.6896552  1.520425  80
```

	lhs		rhs	support	confidence	lift	count
[1]	{Pastry}	=>	{Coffee}	0.0393322	0.5110294	1.126624	139
[2]	{Sandwich}	=>	{Coffee}	0.0387663	0.5229008	1.152796	137
[3]	{Medialuna}	=>	{Coffee}	0.0353707	0.5605381	1.235772	125
[4]	{Brownie}	=>	{Coffee}	0.0280136	0.5238095	1.154799	99
[5]	{Cookies}	=>	{Coffee}	0.0246180	0.5178571	1.141676	87
[6]	{Toast}	=>	{Coffee}	0.0226372	0.6896552	1.520425	80

Conclusions

For both weekends and working days all of the frequent itemsets include coffee, so we see that coffee is a common addition to all transactions. As the support for all presented rules is relatively high, we can see that these products customers tend to buy together, although the confidence of the rules is just above 50%, which leaves some area for improvement.