EDA Report | Supermarket Sales

Understanding Customer Behavior based on Time Series Data

The success of a business is dependent upon the ability to understand customer behavior. This analysis will try and understand the customer behavior based on Time Series Data to help a supermarket chain in Myanmar developing more profitable customer relationships.

I will use a dataset sourced from Aung Pyae on Kaggle, which contains the historical sales recorded in 3 different branches for 3 months.[1].

Method

To conduct the analysis, I will look into the hourly distribution for both weekdays and weekends.

As a first step I cleaned the dataset and more particularly the date and time columns to extract the opening hours (from 10am to 8pm) and separate the workdays (Monday to Friday) from the weekend (Saturday and Sunday).

I end up with 560 rows of data for wordays and 297 for the weekends. Because of the unbalanced population, I only worked with relative numbers.

To define the customer behavior I will go over the following criterias: the attendance, the product line, the quantity of products bought, the total price of the sale and the ratings.

Results

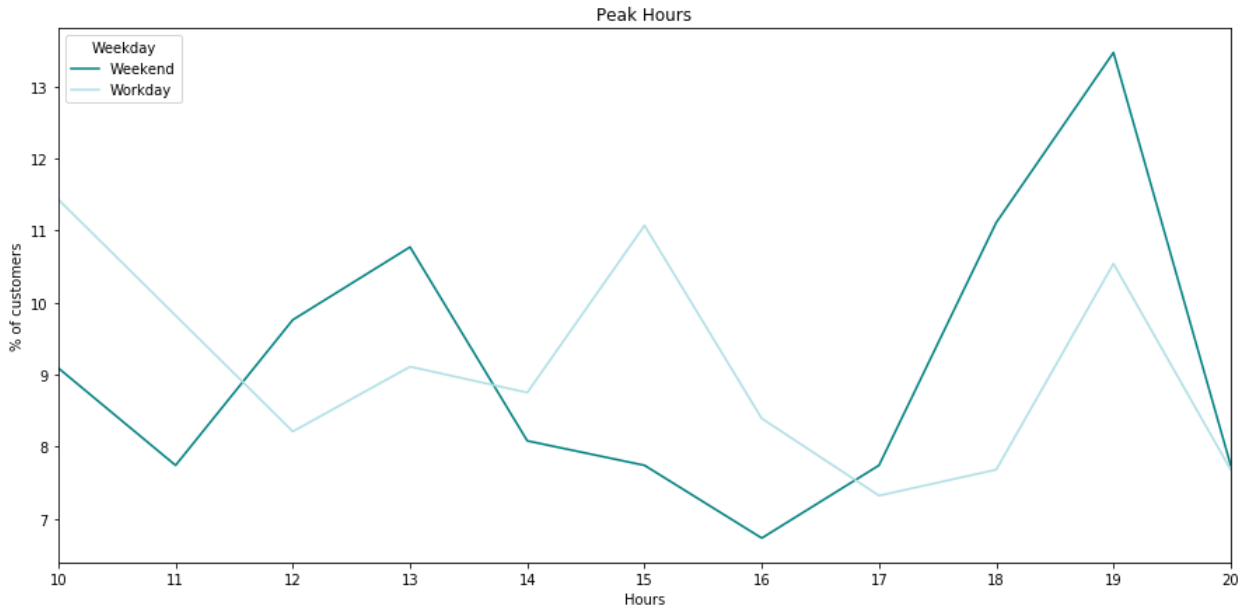
We will start the analysis by trying and defining the potential peak hours over the workdays and the weekends.

The Attendance

65.34% of the supermarket chain customers are shopping during working days.

The below plot shows the average number of customers at the different hours of the day for both weekend and workdays.

[1] KAGGLE, *Supermarket sales* [online], 2019, viewed on 18/02/2020
URL: https://www.kaggle.com/aungpyaeap/supermarket-sales

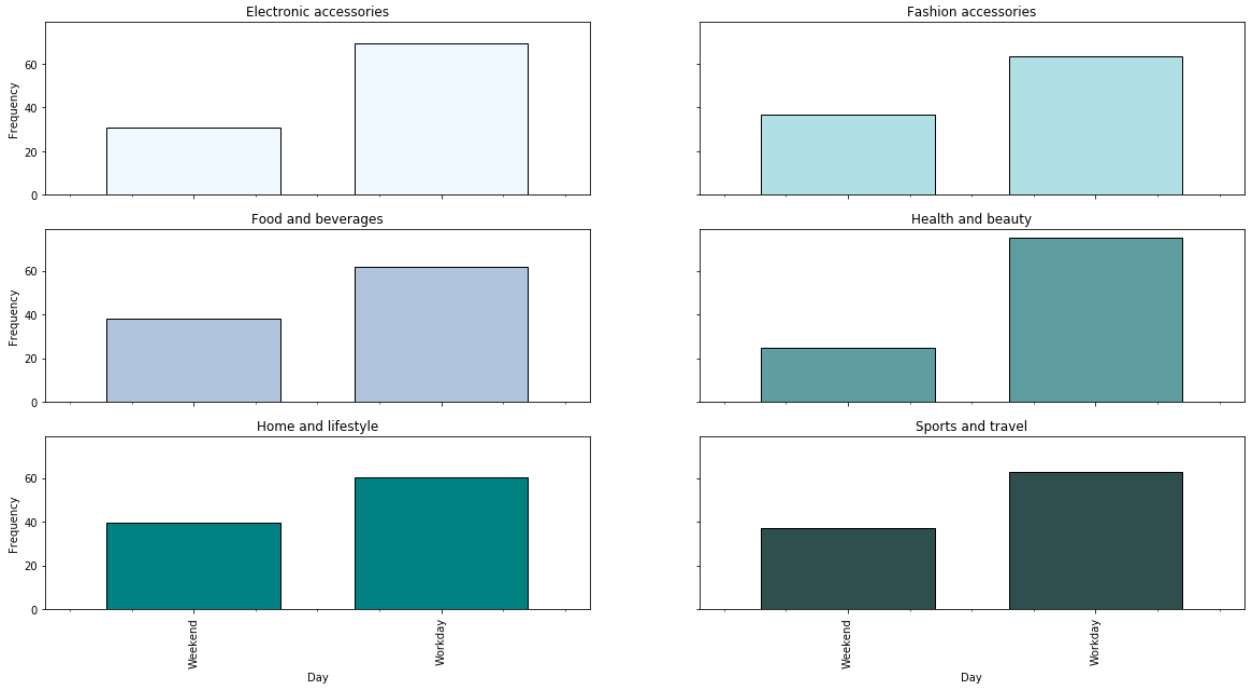


We can see that the pattern differs:

- Workdays: there are 2 peaks, one at 1pm (10.77%) and one at 7pm (13.47 % of the customers). The slowest hour is at 4pm.
- Weekends: if we can see peaks at 10am, 3pm and 7pm, the data is more spread out trough the day. The slowest hours are 4 and 8pm.

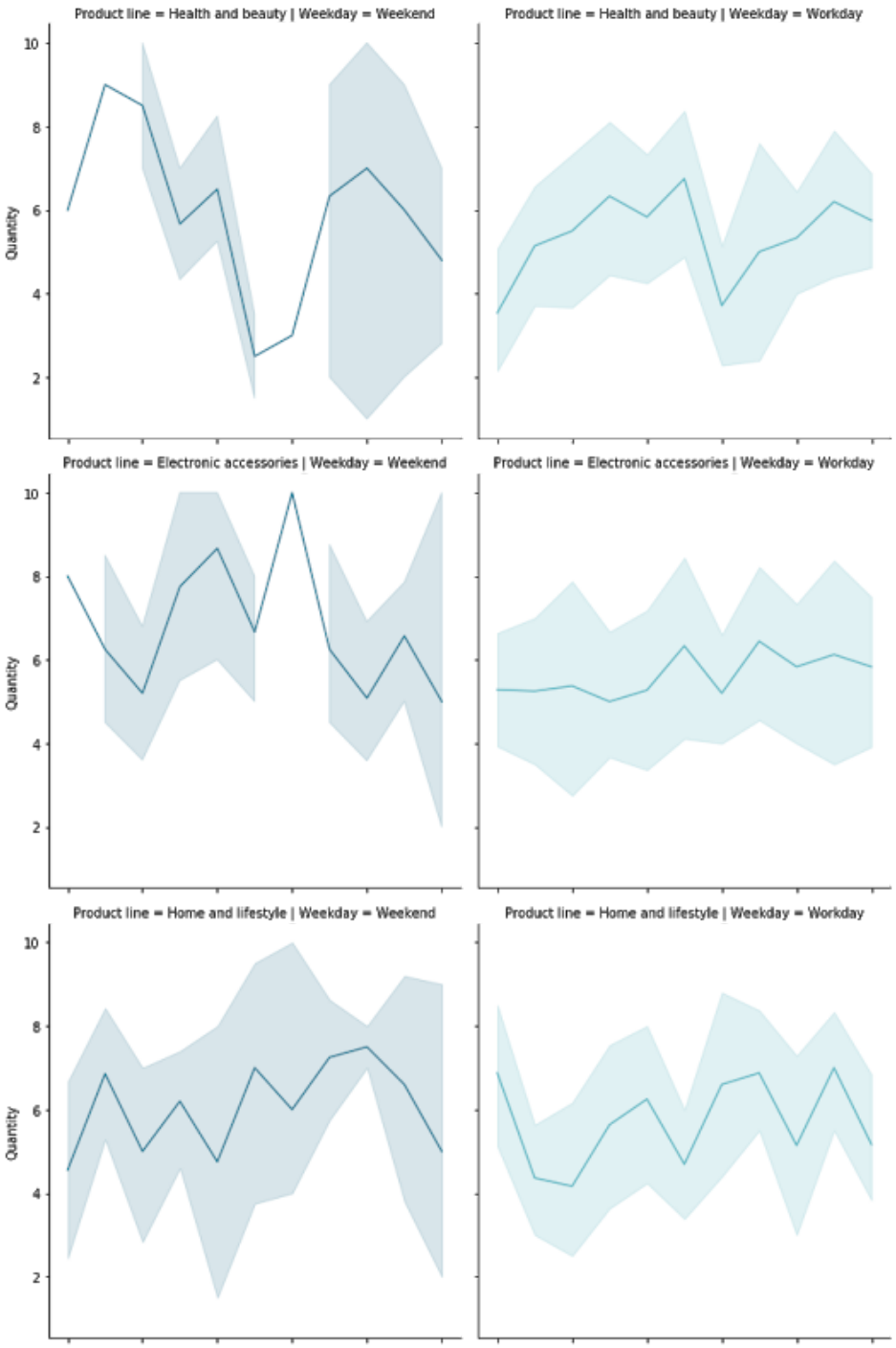
The Product Line

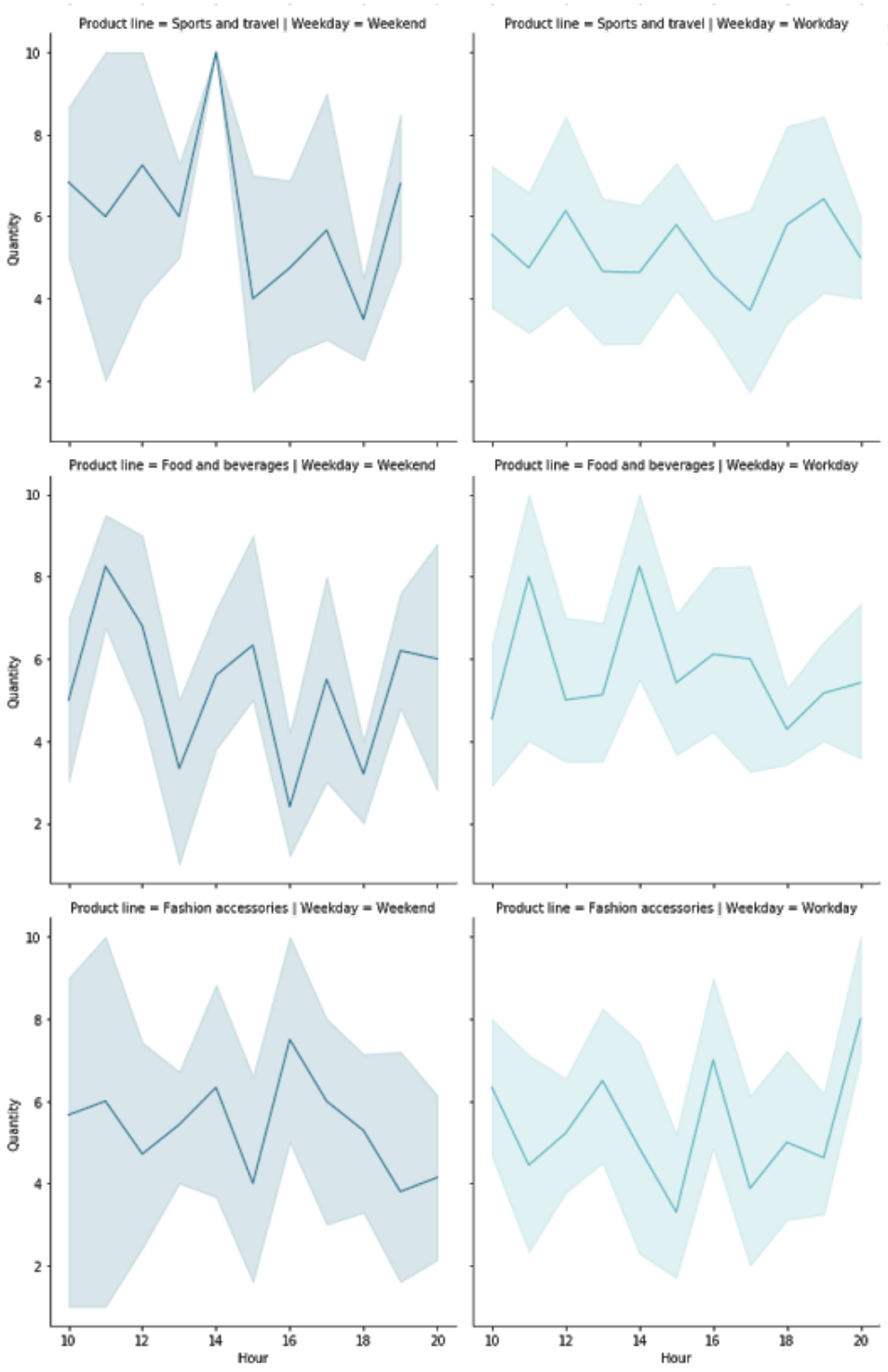
The supermarkets sell six types of products classified as follows: Health and beauty, Electronic accessories, Home and lifestyle, Sports and travel, Food and beverages and Fashion accessories.



For all the product types, we can see that the high majority (more than 60%) of the sales are done over working days. The highest difference being for the health and beauty products where 75.19% of the products are sold during the workdays.

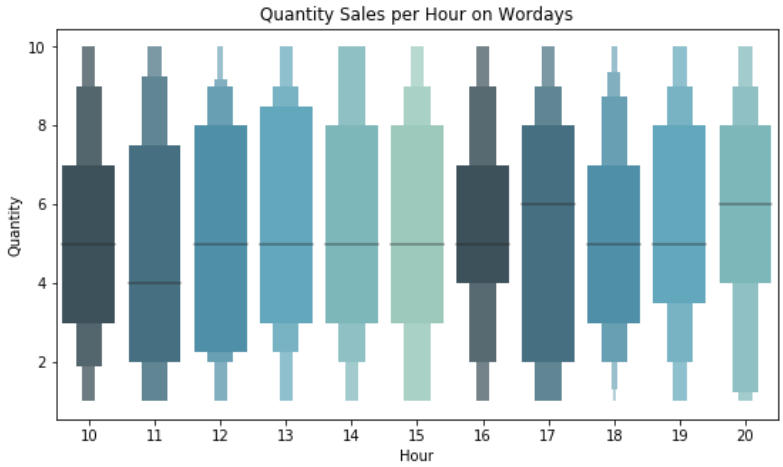
If we look more closely into the hourly distribution, we see that the products are sold more uniformly over the workdays when there are several high peaks and lows other the weekends.

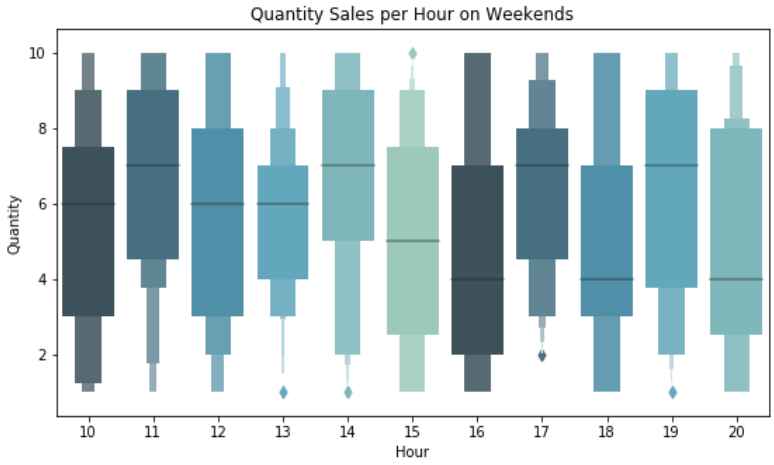



Product line = Sports and travel | Weekday = Weekend
Product line = Sports and travel | Weekday = Workday
Product line = Food and beverages | Weekday = Weekend
Product line = Food and beverages | Weekday = Workday
Product line = Fashion accessories | Weekday = Weekend
Product line = Fashion accessories | Weekday = Workday
Quantity
Hour
10
8
6
4
2
10
12
14
16
18
20

The Quantity

The quantity sold is between 1 and 10 items per purchase.



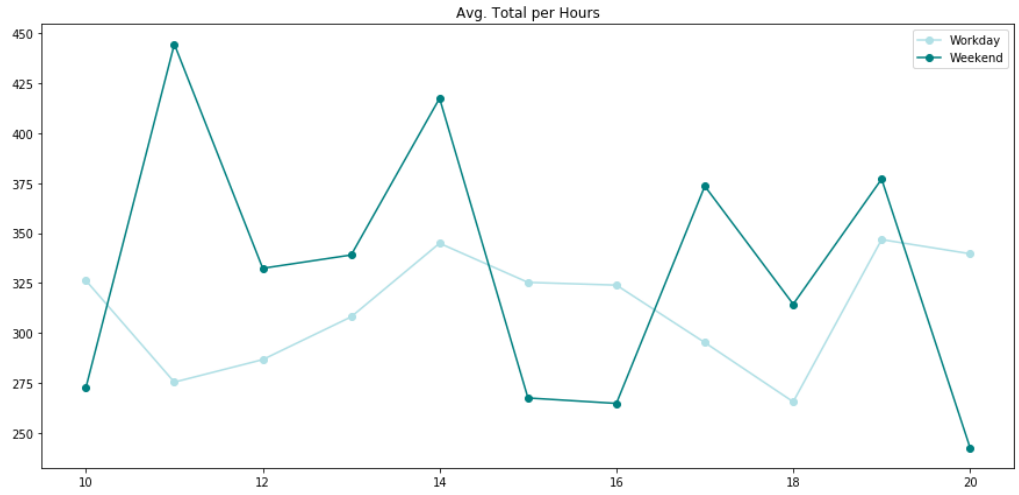


- Workdays: the highest quantity sold is at 5 and 8pm with an average of 6 products and the lowest at 11am with an average of 5.
- Weekends: we can see several peaks at 11am, 2pm, 5pm and 7pm with an average of 7 products.

The Total Purchase Price

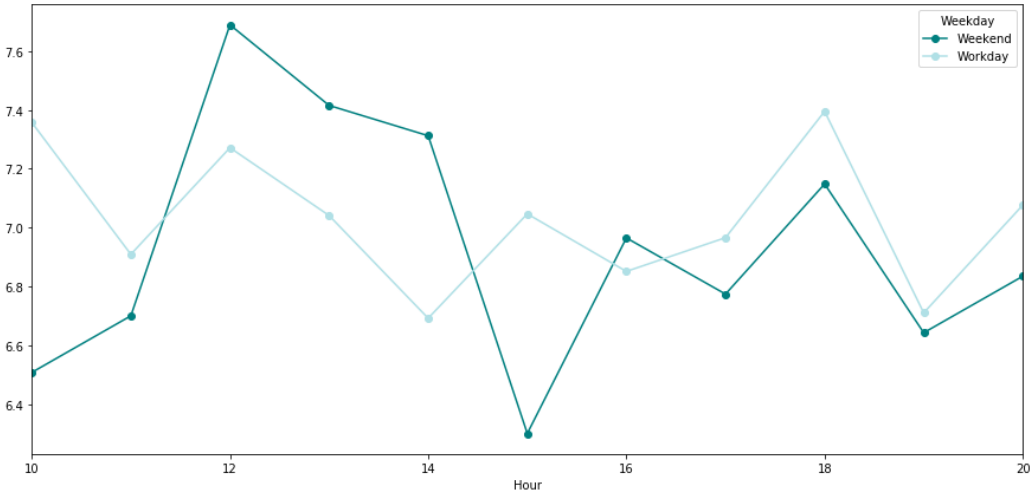
On average, the total of a purchase during the weekend is $334 against $314 during workdays.

If we look more closely into the hourly distribution, we can see that the totals of the purchases over the workdays is more uniformly spread over the day when, during the weekend we can see 4 peaks at 11am, 2pm, 5pm and 7pm which is consistent with the information about the quantity sold per hours (see above).



The Ratings

To conclude our analysis we will study the ratings distribution.
The ratings go from 4 to 10 with an average of 6.95 on the Weekend and 7.03 on the Workdays which is fairly close.



- Weekends: the average rating is different depending on the hour of the day with a peak at 12pm with a score of 7.69 and a low at 3pm with a score of 6.30.
- Workdays: the average is more consistent with a minimum of 6.69 and a maximum of 7.4.

Future Step

To be able to provide counsel to the supermarket management, I would like to analyze data on number of employees working at the different hours and days, the number of registers open and the average time in the queue.