VoltRan

Kerim Acar, Oktay Ekici, Ömer Elmasrı, Gökçe Ezeroğlu, Tarık Özçelik

26.12.2018

Key Takeaways

- We performed analyses on an FMCG company sales data.
- Main analyses were made on sales figures of categories, cities and segments.
- What we saw was the change of sales figures by loyalty is considerable.
- Another important point declared that Beauty products are the main sale components.
- Also sales by cities can be experienced via our Shiny app on our github page.

Introduction to FMCG Company Sales Analysis World

The following analyses are conducted based on the **sales figures of an FMCG company**.
Data shows us details such as invoice date, sales amount,purchased item,item categories etc.
It also includes info about customer characteristics like gender, age, address, segmentation, loyalty.
Processed data is **31.2 MB**. There are approximately **494K rows** with **15 attributes**.

Data Fields: INVOICEID, ACCOUNTNUM, GENDER, AGE, LOA (loyalty indicator), CITY, CUSTOMER SEGMENT, BUSINESS, CATEGORY, SEGMENT, SALES, UNIT PRICE, QUANTITY, FIRST ORDER, INVOICE DATE

Analysis-Necessary Packages

Here you may find the necessary libraries. The references are indicated on the Citations.

```
library(dplyr)                          #[1]
library(tidyverse)                      #[2]
library(ggplot2)                        #[3]
library(scales)                         #[4]
library(readxl)                         #[5]
library(lubridate)                      #[6]
library(treemapify)                     #[7]
library(magrittr)                       #[8]
```

Downloading the Data
Excel file is downloaded from github to a local data frame (raw_data) and prepared for analysis.
Here the raw data structure can be seen in the manner of the number of variables, variable types and the number of observations.
There are **15 variables** and **494.305 observations** to be conducted.

```
## Observations: 494,305
## Variables: 15
## $ INVOICEID        <dbl> 53927301, 53927301, 53927302, 53927302, 53927...
## $ ACCOUNTNUM       <dbl> 23349830, 23349830, 10915900, 10915900, 10915...
## $ GENDER           <chr> "M", "M", "F", "F", "F", "F", "F", "F", "F", ...
## $ AGE              <dbl> 46, 46, 33, 33, 33, 33, 33, 33, 33, 33, 33, 3...
## $ LOA              <dbl> 25, 25, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
## $ CITY             <chr> "Uşak", "Uşak", "Kırklareli", "Kırklareli", "...
## $ Customer_Segment <chr> "E", "E", "B", "B", "B", "B", "B", "B", "B", ...
## $ BUSINESS         <chr> "BEAUTY", "BEAUTY", "BEAUTY", "BEAUTY", "BEAU...
## $ CATEGORY_        <chr> "BODY", "TOILETRIES", "BODY", "TOILETRIES", "...
## $ SEGMENT_         <chr> "MY BODY", "STYLING", "HAND TREATMENT", "WASH...
## $ SALES            <dbl> 223.400, 15.200, 0.758, 0.888, 0.758, 0.942, ...
## $ UNIT_PRICE       <dbl> 22.340, 3.040, 0.758, 0.888, 0.758, 0.942, 0....
## $ QUANTITY         <dbl> 10, 5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ `FIRST ORDER`    <dttm> 2016-12-23, 2016-12-23, 2018-09-06, 2018-09-...
## $ INVOICE_DATE     <dttm> 2018-09-29, 2018-09-29, 2018-09-29, 2018-09-...
```

```
## [1] 494305      15
```

Exploratory Data Analysis

1. Total Sales vs Customer Segments (1st Graph)

Beauty is more driver than the Fashion and Home. D segment is the highest for contribution to the total sales.

The first graph in following shows us the total sales for 3 business segments (Beauty, Fashion and Home).
Accordingly, the main business category of this company is **'beauty'**, since the majority of sales come from this group. **'Fashion'** has also small share in total sales, while **'home'** share in total is negligible.

From A to J, there are **10 customer segments** in total. The other output out of the first graph is the contribution amount of every single customer segment to total sales.

It is showed that, customers in **D segment** brings **the highest contribution**; while F and J segment follow it,respectively.

Total sales based on Sales Representative Segment
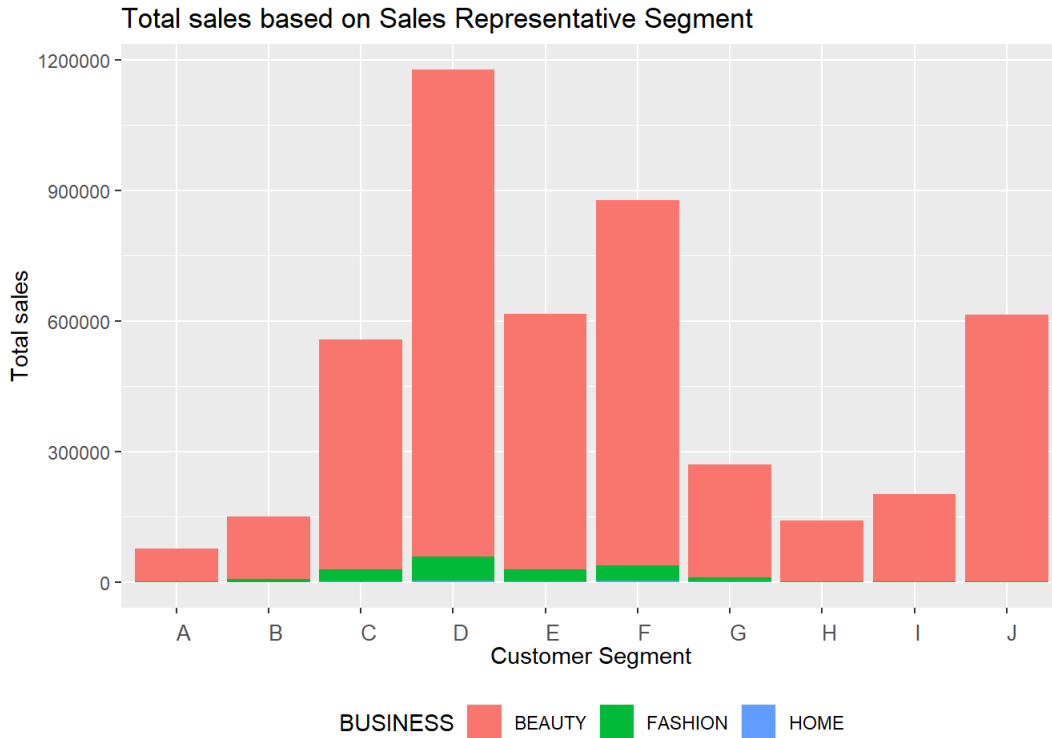

2. Average Sales vs Customer Segments (2nd Graph)

J segment is Atom-Ant. The highest average sales with lower sales contribution to the total sales.

The second graph shows the average sales for each segment. Once we compare it with the previous one, it results in the fact that, the middle segments have the higher contribution to total sales.

However, **J segment**, which is **the premium one**, has the highest average sales among all. The reason is, the number of customers in this premium (J segment) is quite low. But their total sales amount is quite high. Therefore, the average sales of this **J segment** is **the highest one**.
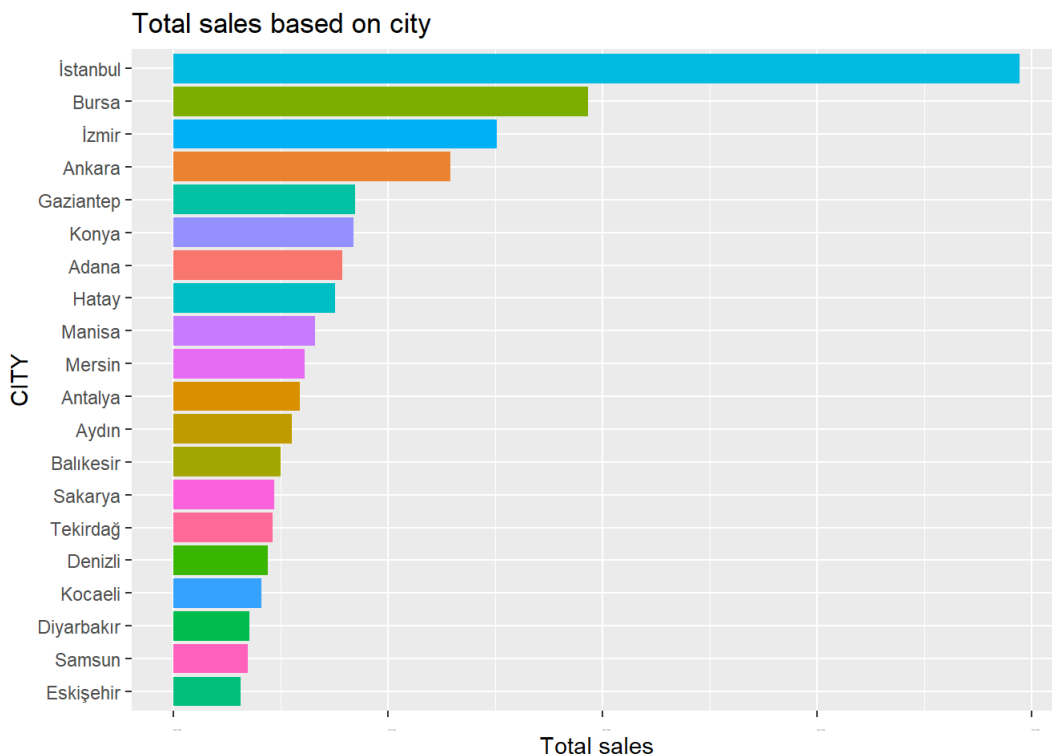
Average sales based on Sales Representative Segment
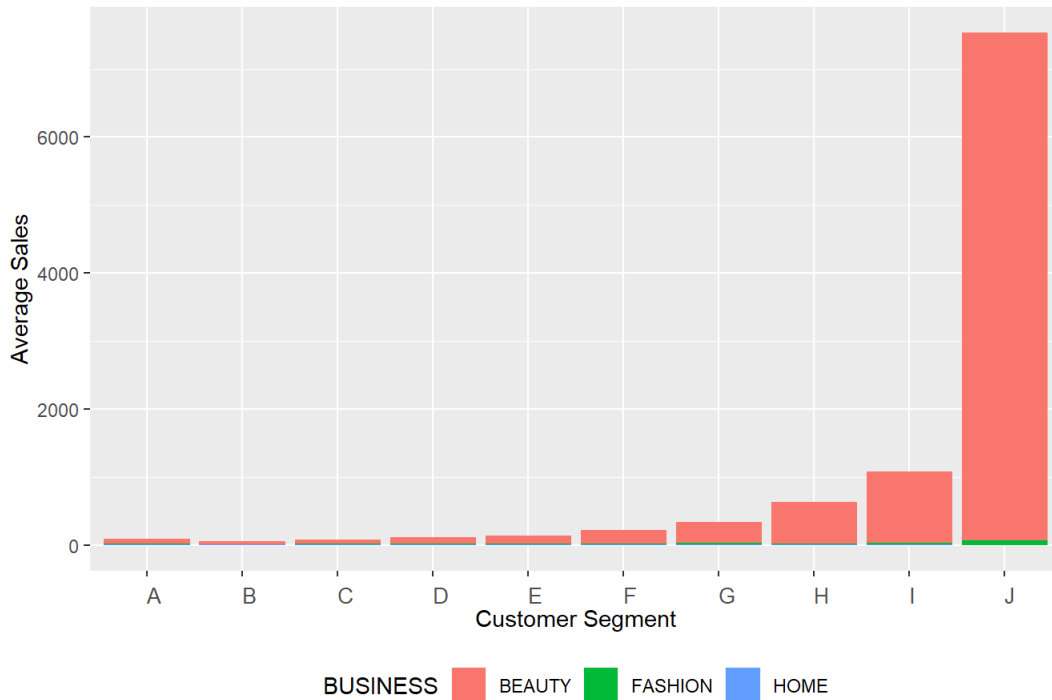
3. Total Sales based on the City (3rd Graph)

Everybody knows, Istanbul has huge market in Turkey. However, do not underestimate Bursa, please. Bursa is runner-up city.

In the third graph, top 20 cities,based on their total sales amount are showed. **Istabul, Bursa** and **Izmir** are in top 3.

Total sales based on city

Elections? No, just the sales figures provided by AA.

In the fourth graph, the sales density can be red from the map. The reddish the city is, the denser the sales are. Or can we say **women spending more money on beauty products?**
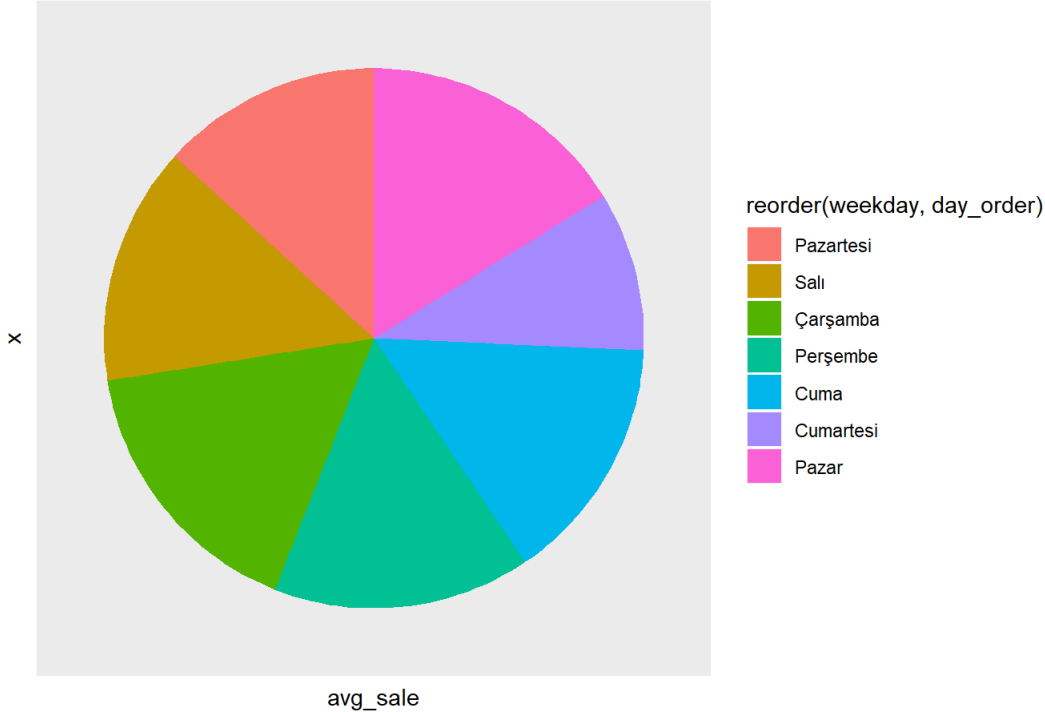
Relative Frequency Distribution of All Sales
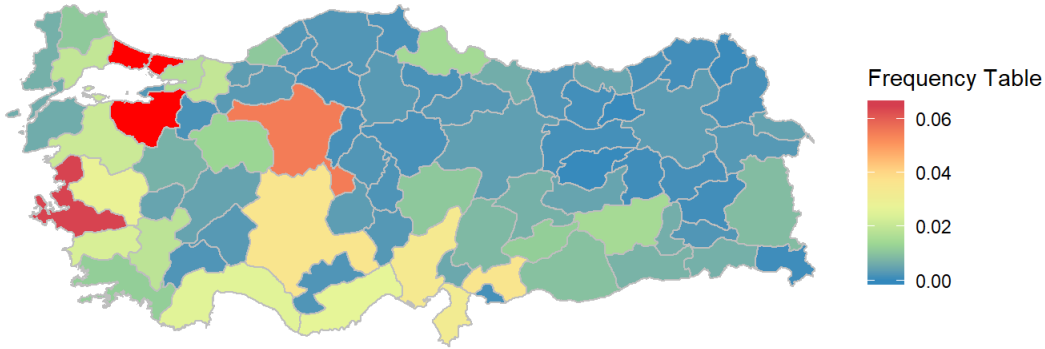
4. Which Days are Popular for Sales? (5th Graph)

The laziest day may be Saturday. Everybody loves Saturday for relaxing.

In the 5th graph, weekdays are desired to compared with respect to their sales contribution, distributon is given in a pie chart. Saturday has slightly lower share in a week.


5. Which Product Category is the Most Famous? (6th Graph)

The Oscar goes to the **Fragrance** . Fragrance is very important. If you think not, Please watch the movie "Perfume: The Story of a Murderer"

There are sub- categories in each of the business units.To illustrate, **'beauty'** business category includes 5 different sub-categories; such as: body, toiletries, face, fragrance, color

6th graph in below shows us the share of all these sub-categories in total sales. Bigger area indicates bigger share.
Accordingly, **fragnance** is the first, and it is followed by **toiletres** and **color** respectively.
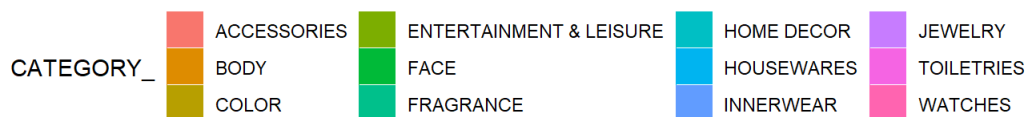
CITIATIONS:

[1]Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.7.
https://CRAN.R-project.org/package=dplyr

[2]Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1.
https://CRAN.R-project.org/package=tidyverse

[3]H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
https://ggplot2.tidyverse.org/

[4]Hadley Wickham (2018). scales: Scale Functions for Visualization. R package version 1.0.0.
https://CRAN.R-project.org/package=scale

[5] Hadley Wickham and Jennifer Bryan (2018). readxl: Read Excel Files. R package version 1.1.0.
https://CRAN.R-project.org/package=readxl

[6]Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25.
http://www.jstatsoft.org/v40/i03/

[7]David Wilkins (2018). treemapify: Draw Treemaps in 'ggplot2'. R package version 2.5.2.
https://CRAN.R-project.org/package=treemapify

[8]Stefan Milton Bache and Hadley Wickham (2014). magrittr: A Forward-Pipe Operator for R. R package version 1.5.
https://CRAN.R-project.org/package=magrittr