Introduction:

We have a data about sales for Food Stores in Egypt, data contains 8 Columns and 8035 Rows

I started with data analysis process as follow:

1) Identifying data and measures (business Metrix) to use
2) Clean data (incorrect data format – incorrect data type – create custom columns – create columns from selection – remove duplicates – recorrect cashier names based on their cashier number – merge columns – fix datetime and date data)
3) Analyze data and create useful measures that indicates findings
4) Create charts based on the measures we used and design a suitable dashboard

---

Identifying the data:

Store: sales stores in Egypt

Source: kind of product sold in these stores

ReferenceNumber: Indicates the specific order number and should be unique

Value: product price

CreatedAtStore: date and time of creating and put the product in our store

DueDate: last date and time for selling the product

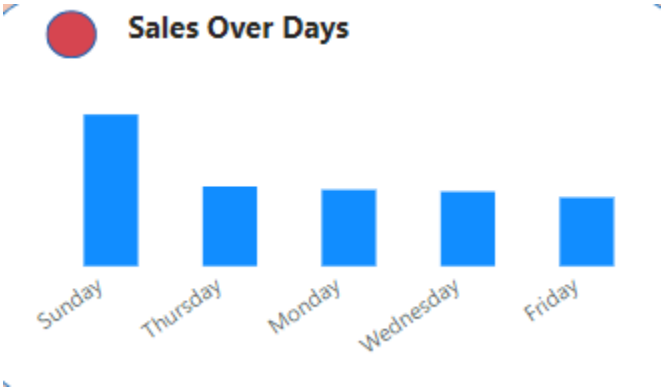
CashedDate: date and time of selling the product

Cashier Number: a unique Id for each cashier that used as a primary Key referee to a specific cashier

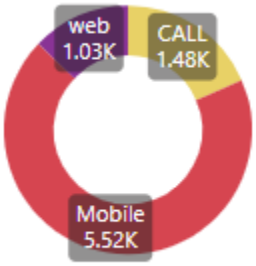
Cashier Name: cashier Name

Used Measures with details:

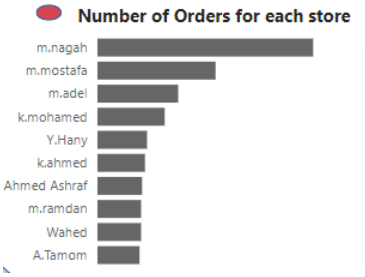
1) Sales by days: I created a column chart that indicates the sales over days of week with a tooltip indicates the average time predicted for selling any product in each day



2) Count number of products sold: A donut chart indicates the number of each product(source) sold from our stores

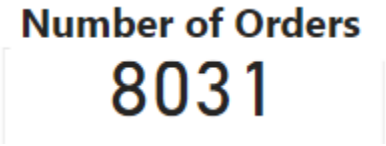


3) Number of orders sold by cashiers: a column chart indicates number of orders sold by each cashier

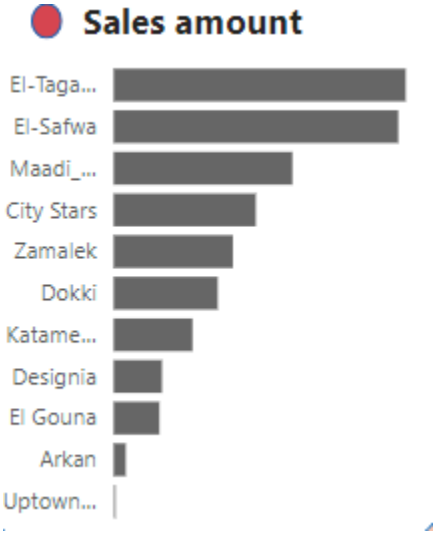


4) Sales amount: a card indicates the total sales amount

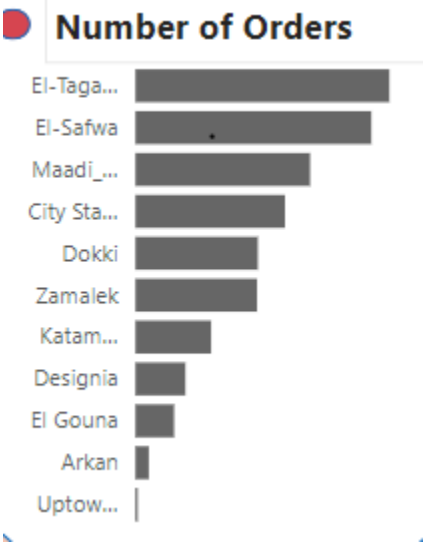
5) Number of orders: a card indicates total number of orders



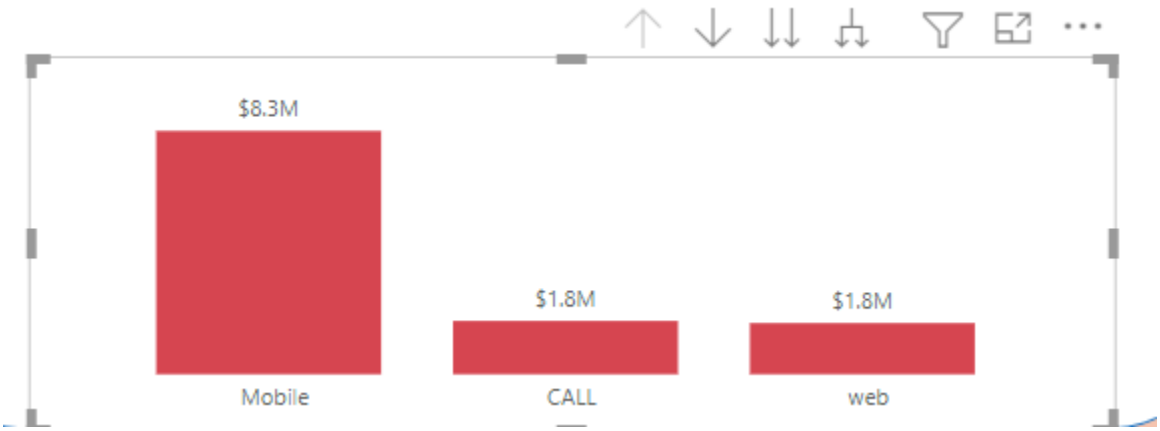
6) Sales amount for each store: a column chart indicates the total sales amount for each store



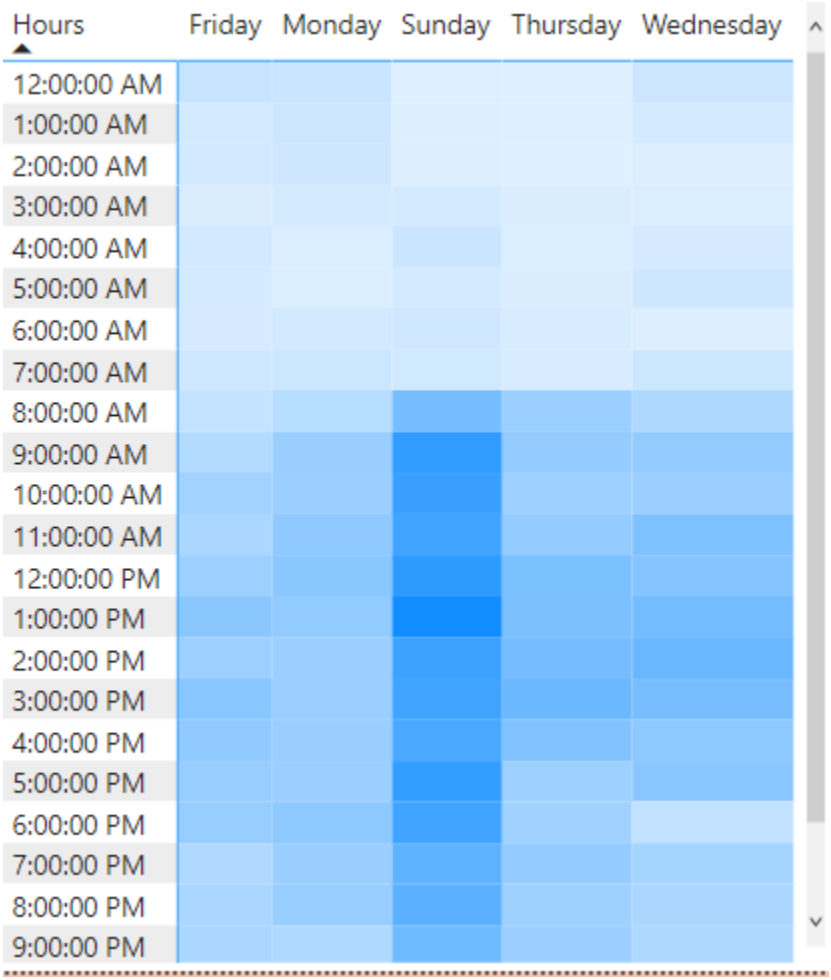
7) Sales amount for each store: a column chart indicates the total sales amount for each store



8) Sales for each source (Product): a bar chart with a drill-down through each product for a year, quarter, month, and day



9) Selling hours for each day: a heatmap indicates from most to lowest time to sale in our stores



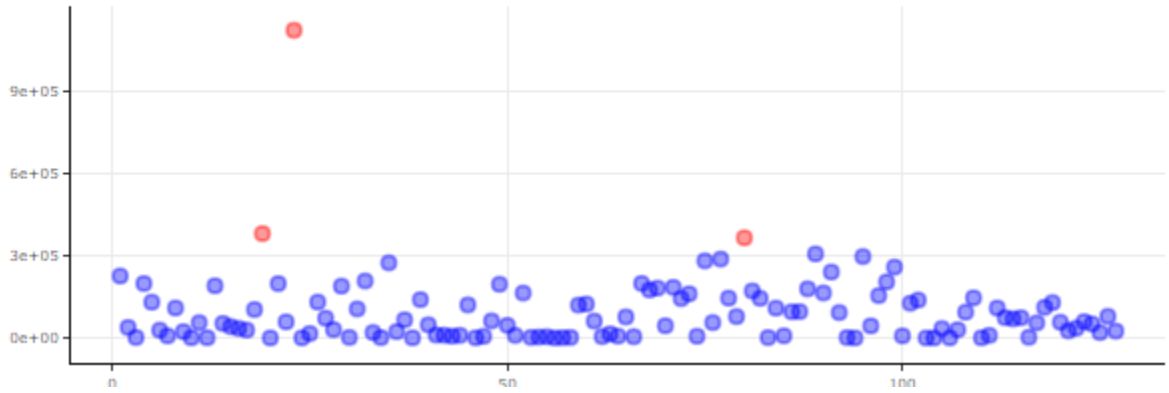
10) hours and number of orders: a table indicates number of products sold for each hour

Hours	Number of Order
1:00:00 PM	669
3:00:00 PM	620
12:00:00 PM	606
2:00:00 PM	604
4:00:00 PM	553
11:00:00 AM	538
5:00:00 PM	538
9:00:00 AM	510
10:00:00 AM	501
6:00:00 PM	471
7:00:00 PM	439

11) average time in store for each cashier: a table indicates the average time that it takes from each cashier to buy the product (cashed – created)

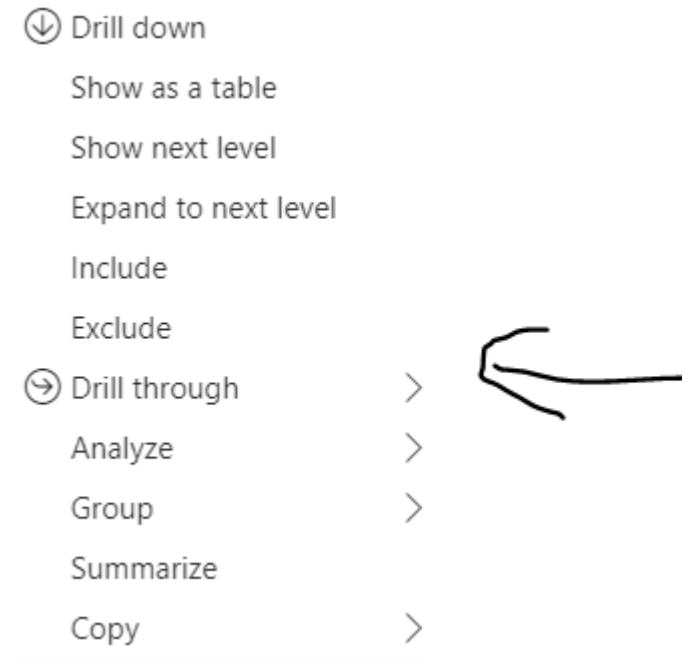
Cashier Name	Average Time In Store
m.medhat	0.10
Fathy Reda	0.15
m.hussien	0.15
Ahmed Hussien	0.23
W.Farouk	0.28
Saleh Gamal	0.32
Omar Hanafy	0.33
b.gamal	0.33

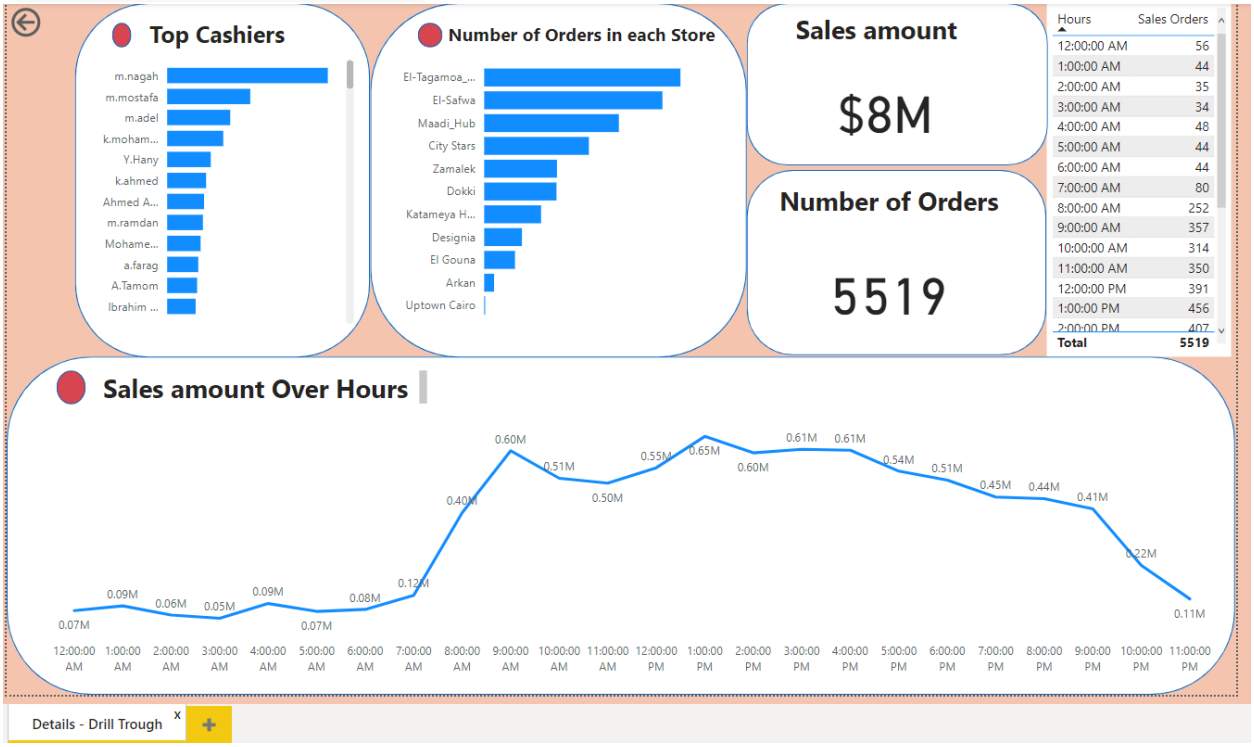
12) sales outliers detection for each cashier: an outlier detection measure used to detect outliers based on values with Zscore method and Standard deviation = 2



13) duplicates slicer: a slicer that indicates whether the orders are duplicated or unique

14) drill-through option: I added a drill-through to shows the related numbers and insights of any specific category





Insights and Findings

1) Sunday is the most sales day with more that 4M total sales and after that comes Thursday and Monday with 2.1M and 2M Respectivly

2) The most sold product is Mobile with 5.5k after that comes Call and Web with 1.5k and 1k respectivly

3) M.nageh is the most active cashier with more than 800 orders

4) El-Taghmoa-Hub is the most selling store with more that 2.5M sales and 1.5K orders

5) Total sales for all stores is quite 12M with more than 8K orders sold

6) Based on the heatmap: Sunday is the most sales day by range from 120 to 240 orders sold

7) 1 PM is the most selling hour

8) M.medhat is the lowest cashier with time in store(time between putting and selling the order from the store) with .10 hour(based on his number of orders sold and it's not a good sign for active cashier because maybe he sold a less orders with a less time[maybe baised so filtering data will tell us his all numbers (orders and sales value)])

9) We have 77 duplicated orders with 80K sales value, most duplicated orders from k.karеam with 5 duplicated orders and from El-Taghmoa-Hub store with 12 duplicated orders

Suggestions and Recommendations

1) Have a meeting with El-Taghmoa-Hub store's manager to discuss why there is a duplicated orders for the same referance number

2) Have a meeting with k.karеam to discuss about duplicated orders

3) Increase mobile products in our stores specifically in El-Taghmoa-Hub store as there is a huge number of customers and sales in this store

4) Increase our store's inventory on Sunday as it's the most shoping day for our customers specifically in El-Taghmoa-Hub and El-Safwaa stores

5) Alert our cashiers to be more attention specifically on 1 PM and (from 11 AM to 5 PM) as it's the most times we have customers on, so they expected to do more efforts

6) Discuss with HR team the performance of m.nageh as he is the most active and productive cashier to give him a reward regardless his great efforts (if it's applicable)

Discuss Data Cleaning and Preparing

First: checking data for duplicates

I found duplicates on the cashiers table (more than 7000 duplicated names and cashiers' numbers)

I found duplicates on the sales table (less than 5 duplicated rows)

Second: Checking for incorrect cashiers' numbers and incorrect names

I found that there are incorrect cashiers' numbers and names for more than one cahier (Ahmed Hasan 791 – m.abdeltawab 2971 – Y.Hany 2992 – and 3 other cahiers) I fixed them based on using the average numbers of names repeated for each cashier number

Third: removing rows with errors

I removed 2 rows with errors on cells values that detected on Power query editor

Fourth: fixing date format:

Split datetime columns to date columns and time columns also creating a column with the start hour for cashed date column to be used on our measures in the dashboard

Fifth: creating columns based on selection and custom columns

Sixth: creating duplicate and unique column

Using Dax to create a specific column that detect whether the order is duplicated or not

Seventh: creating required measures:
I created some measure to be used in our measures

Eighth: building the data model

After cleaning the data, we are able to connect the fact table (Sales) and the dim table (Cashiers) in a [Many to One] Relationship that will smooth our dashboard work, filters, and connections among the different charts in the dashboard

---

in the end:

I let the data cleaning process to the end because it's the base for any accurate and good interactive dashboard