2022/3/26 | By: Ahmed Ramadan

Report (Otlop)

[E3ml business task]

Data Analyst Task (Otlob)

Delivering orders (“Otlob” application).

Table of Contents:¶

Context

We work on generating insights from sales data, finding patterns of orders, Discovery any clear trends or causality. This report provides insights about the top-performing and underperforming Store and delivery, find a solution to the problem of high variation time waiting until the order is picked-up.

Task Requirements:

A- Finding any clear trends / causality?

- What other causes that are not in the given data table?
- can you think of affecting the number of times an order is reassigned?
- Based on your finding propose at least one action to tackle this issue.

B- Once the courier arrives to the store,

- the time until the order is picked-up can present very high variation.
- With the data given, present an analysis showing this, giving as much as detail as possible.
- Can you identify any clear trends in terms of worst offenders?
- What could be causing this high variability?
- What could be the possible causes?
- when is the courier waiting for a very long time before picking up the order?
- Propose at least one action to

C- SQL Test p

Code

Link of code:

https://colab.research.google.com/drive/1s0jslFy8e4K085XaISe9M4eKr54aZX5c?usp=sharing

Dependences

The main dependencies for the project are Tableau, Excel, and Python data analysis libraries.

Guiding questions

What tools are you choosing and why?

- I'm using Tableau, Excel tools Because of the small dataset. And Python for this effective cleaning process of data.

Have you ensured your data's integrity?

Yes, the data is consistent throughout the columns.

What steps I have taken to ensure that your data is clean?

- First the missing values where removed, then the columns formatted to their correct format.

How can you verify that your data is clean and ready to analyze?

- Yes, it's all documented in this python notebook.
- Document the cleaning process.

https://colab.research.google.com/drive/1s0jslFy8e4K085XaISe9M4eKr54aZX5c?usp=sharing

Key tasks

- Check the data for errors, Missing values, Impossible values, Outliers.
- Transform the data so you can work with it effectively.
- Document the cleaning process.

Deliverable

- Documentation of any cleaning or manipulation of data.

Analyze

Putting in a new variable:

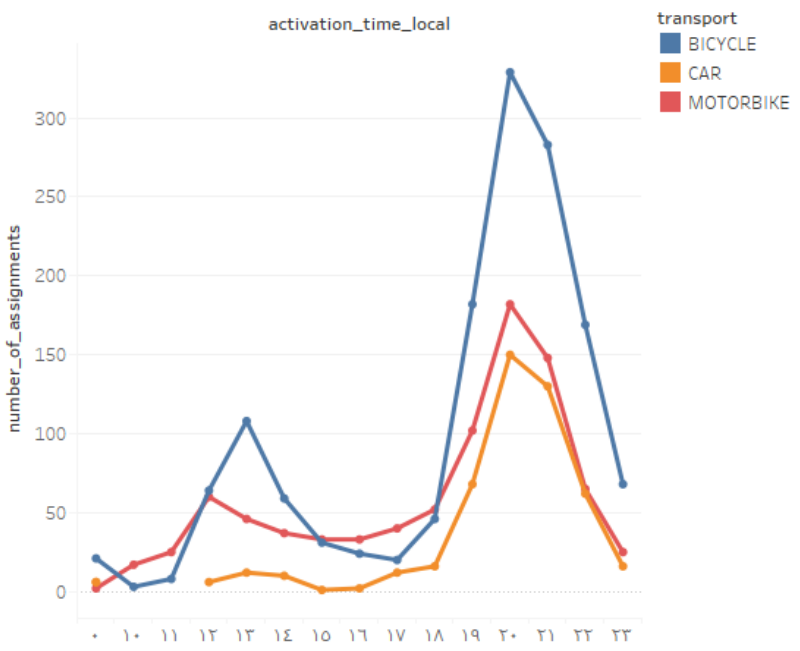
- Time waiting to Pick up order
- Time to Assign
- Distance Time from Store to Customer

Guiding questions

A- Finding any clear trends / causality?

- From 7:00 Pm to 11:00 Pm we have high orders demand peak “Launch time”.
- From 12:00 Pm to 2:00 Pm we have order demand peak “Dinner time”.

From 7:00 to 11:00 Pm > high order demand peak



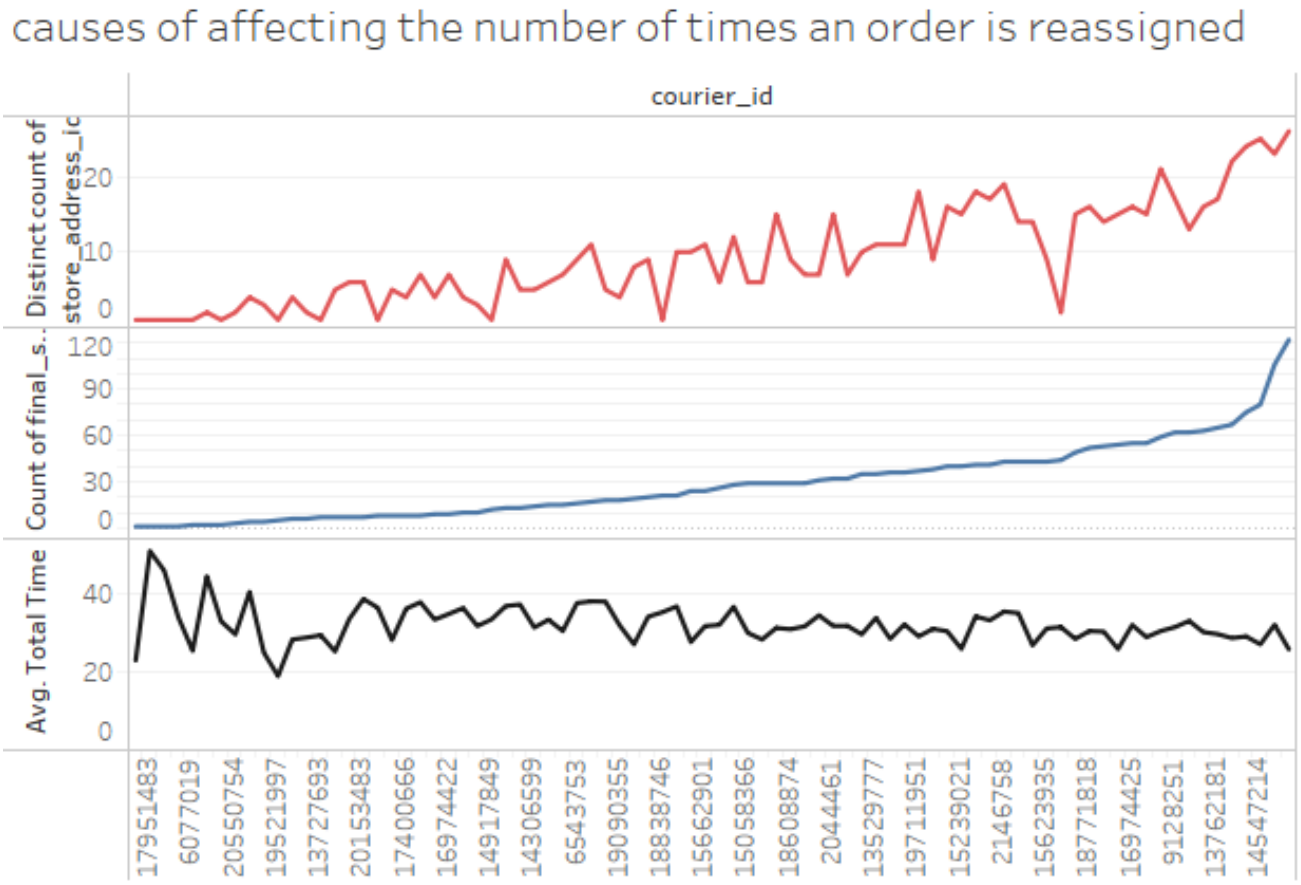
The trend of sum of number_of_assignments for activation_time_local Hour. Colour shows details about transport.

What other causes that are not in the given data table?

- New research shows that the daily diet increases hunger and cravings for sweet, starchy, and salty foods in the evening.

What causes of affecting the number of times an order is reassigned?

- Total time of order is most factor effect on number of assignments
- **curve black:** There is Negative relationship between Avg of Total Order time and Number of assignments When Order time increase the avg of Number of assignments decrease.
- **Red curve**: There is a positive correlation between the count of purchase from the same Store and customer id.
- **Blue curve**: Explain the number of purchases delivered to each individual customer



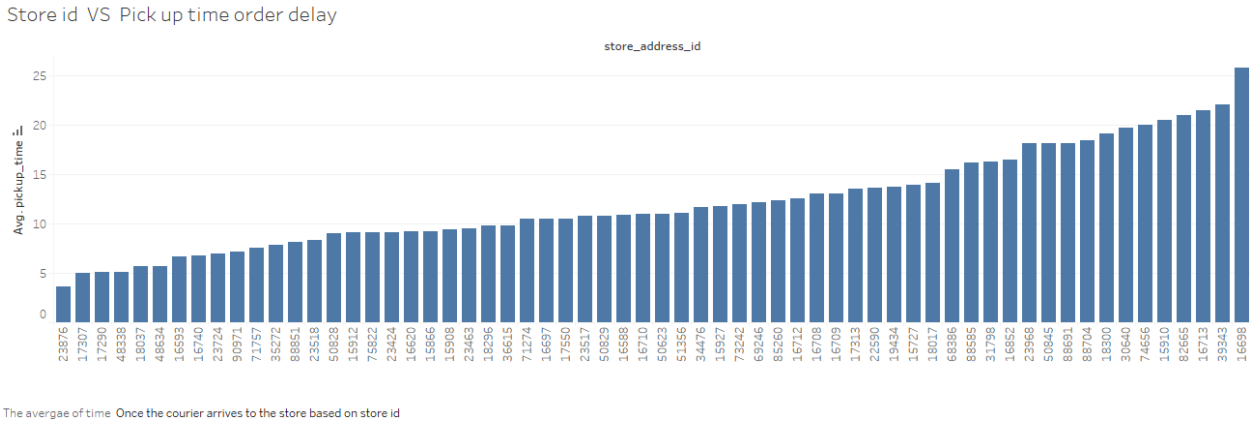
The trends of distinct count of store_address_id,count of final_statusandaverage of Total Time for courier_id.The view is filtered on courier_id, which has multiple members selected.

Based on finding Recommendation action to tackle this issue.

1. During peak hours, such as dinner time, bundling can increase the capacity of the -eat, leading to lower delivery times on average, so increasing the number of couriers in the evening from 7:00 Pm to 11:00 Pm will solve problem
2. Gather Feedback from couriers and store Id to evaluate them.

B- Once the courier arrives to the store

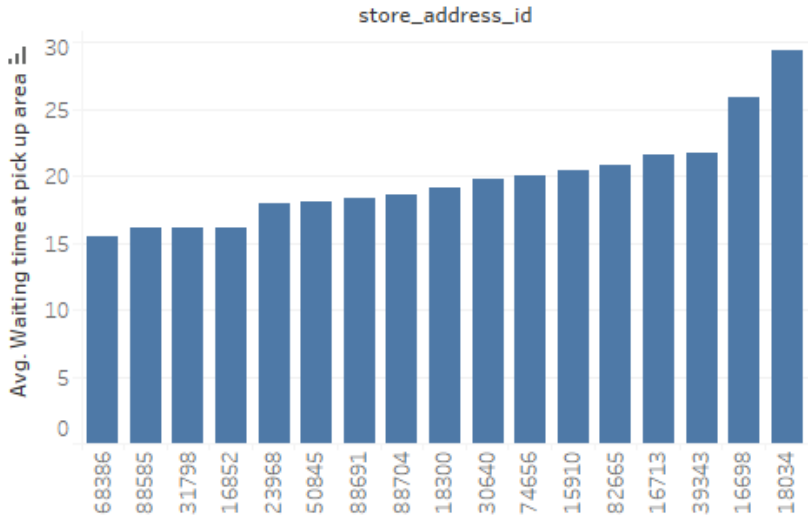
- The time until the order is picked-up can present very high variation. Because of some stores have a delay in prepare orders exceeded 20 to 30 min where the average is 15 min as we can see here in fig:



The avergae of time Once the courier arrives to the store based on store id

Top 17 worst Store?

15 min is the average of Waiting time at pick up area . We have 17 stores delay orders exceeded the 15 min



15 min is the average of Waiting time at pick up area Once the courier entre the area to the store based on store id. We have 17 stores delay orders exceeded the 15 min

What could be causing this high variability?

- Average of Time to assign is 2 min for each courier.
- The view is filtered exceeded the average of Time to assign, which ranges from 2 to 9 min

when is the courier waiting for a very long time before picking up the order?

Courier waits to prepare the order in the pickup area. The delay mostly from stores.

Propose actions to tackle this issue.

- It is extremely important to prioritize your orders so that you can identify certain customers or stops that must be served urgently. Once you define the order's priority, you can accordingly include it during the route optimization process.
- Set delivery priorities: identify certain customers or stops that must be served urgently
- Set realistic courier deadlines set realistic delivery service timelines. Otherwise, you will end up with unsatisfied customers that will not avail of your service again
- Develop a strategy for High volume shipments time: Most courier drivers usually work from 9 a.m. to 6 p.m. You should always bear in mind that delayed delivery often occurs during peak hours
- Awareness Stakeholders of the negative effect of time delay with number of orders
- Evaluate and game fiction of delivery on time for store and courier

What surprises were discovered in the data?

- pick-up area has some numbers of minus numbers. I think this because of courier man already in the area within 100m radius from the Store before the activation time for orders

Key tasks

- Organize and format data.
- Perform calculations.
- Identify trends and relationships.

3- SQL Task

```
-- Using Subquery in SELECT statement

SELECT 100 * COUNT(is_bundled) / COUNT(order_id) AS 'Percent'

FROM
   bundled_orders AS b

WHERE is_bundeld = "true" IN (
     SELECT order_id
     FROM customer_orders AS C
     WHERE city_code = "GIV" AND "PLY"
)
INNER JOIN
   customer_orders ON b.order_id = c.order_id



-- Using CASE statements

SELECT
   CASE
      WHEN is_bundeld = "true" THEN "true"
      WHEN city_code =  "GIV" THEN "GLV"
      WHEN city_code = "PLY" THEN "PLY"
      ELSE "false"
      END AS "type_of_orders"
    100 * COUNT(is_bundled) / COUNT(order_id) AS 'Percent'
FROM
   Customer_orders AS C
INNER JOIN
   orders AS O ON O.order_id = C.order_id
GROUP BY
     1
```