Logistics Performance Analytics

Background

Due to the recent COVID-19 pandemic across the globe, many individuals are increasingly turning to online platforms like Shopee to purchase their daily necessities. This surge in online orders has placed a strain onto Shopee and our logistics providers but customer expectations on the timely delivery of their goods remain high. On-time delivery is arguably one of the most important factors of success in the eCommerce industry and now more than ever, we need to ensure the orders reach our buyers on time in order to build our users' confidence in us.

In order to handle the millions of parcels that need to be delivered everyday, we have engaged multiple logistics providers across the region. Only the best logistics providers that are able to meet Shopee's delivery standards are partnered with us.

The performance of these providers is monitored regularly and each provider is held accountable based on the Service Level Agreements (SLA). Late deliveries are flagged out and penalties are imposed on the providers to ensure they perform their utmost. The consistent monitoring and process of holding our logistics providers accountable allows us to maintain our promise of timely deliveries to our buyers.

Task

Identify all the orders that are considered late depending on the Service Level Agreements (SLA) with our Logistics Provider.

For the purpose of this question, assume that all deliveries are considered successful by the second attempt.

Basic Concepts

- Each orderid represents a distinct transaction on Shopee.
- SLA can vary across each route (A route is defined as Seller's Location to Buyer's Location) - Refer to SLA_matrix.xlsx
- Pick Up Time is defined as the time when the 3PL picks up the parcel and begins to process for delivery. It marks the start of the SLA calculation.
- Delivery Attempt is defined as an attempt made by the 3PL to deliver the parcel to the customer. It may or may not be delivered successfully. In the case when it is unsuccessful, a 2nd attempt will be made. A parcel that has no 2nd attempt is deemed to have been successfully delivered on the 1st attempt.
- All time formats are stored in epoch time based on Local Time (GMT+8).
- Only consider the date when determining if the order is late; ignore the time.
- Working Days are defined as Mon - Sat, Excluding Public Holidays.
- SLA calculation begins from the next day after pickup (Day 0 = Day of Pickup; Day 1 = Next Day after Pickup)

- 2nd Attempt must be no later than 3 working days after the 1st Attempt, regardless of origin to destination route (Day 0 = Day of 1st Attempt; Day 1 = Next Day after 1st Attempt).
- Only consider the date when determining if the order is late; ignore the time.
- Assume the following Public Holidays:
- 2020-03-08 (Sunday);
- 2020-03-25 (Wednesday);
- 2020-03-30 (Monday);
- 2020-03-31 (Tuesday)

Data Description

delivery_orders_march.csv: contains all the relevant information for orders delivered in the month of March.
Columns:

- orderid: Each orderid represents a distinct transaction on Shopee
- pick: Pick Up Time, which is defined as the time (represented by epoch time) when the 3PL picks up the parcel and begins to process for delivery
- 1st_deliver_attempt: Time (represented by epoch time) when 3PL first attempts a delivery.
- 2nd_deliver_attempt: Time (represented by epoch time) when 3PL attempts a delivery again after the 1st attempt has failed. Orders which were successfully delivered the 1st time will not have a 2nd attempt.
- Buyeraddress: buyer's address (Destination)
- Selleraddress: seller's address (Origin)

SLA_matrix.xlsx: contains the Service Level Agreements for the different routes.

Examples

Example 1

Orderid	1955598428
Seller's Address	"Block 2, Lots 2,3,10 & 11, Honest St cor. Determined Street, Calamba Premiere International Park (CPIP), Batino, Calamba, Laguna, Philippines Calamba City Batino Laguna Luzon"
Buyer's Address	"unit 2 seaviews castles, Tambo, Paranaque City, Metro Manila, Metro Manila"
Pick Up Time	1583137548 (Converted to 2020-03-02 4:25:48 PM Local Time)
1st Attempt Time	1583733540 (Converted to 2020-03-09 1:59:00 PM Local Time)

2nd Attempt Time	NaN

Based on the SLA matrix:

- Luzon -> Metro Manila: 1st Attempt must be within 5 days from Pick Up Time.

This means the Logistics Provider has up to 2020-03-07 to attempt a delivery. This order is deemed as late.

Calculating the Days

Day 0 = 2020-03-02 (Pick Up Time)
Day 1 = 2020-03-03
Day 5 = 2020-03-07
Day 6 = 2020-03-09 (we exclude 2020-03-08 from the calculation because it is a Sunday)

Example 2

Orderid	1955598428
Seller's Address	"Block 2, Lots 2,3,10 & 11, Honest St cor. Determined Street, Calamba Premiere International Park (CPIP), Batino, Calamba, Laguna, Philippines Calamba City Batino Laguna Luzon"
Buyer's Address	"unit 2 seaviews castles, Tambo, Paranaque City, Metro Manila, Metro Manila"
Pick Up Time	1583137548 (Converted to 2020-03-02 4:25:48 PM Local Time)
1st Attempt Time	1583412300 (Converted to 2020-03-05 8:45:00 PM Local Time)
2nd Attempt Time	1583850180 (Converted to 2020-03-10 10:23:00 PM Local Time)

Based on the SLA matrix:

- Luzon -> Metro Manila: 1st Attempt must be within 5 days from Pick Up Time
- 2nd Attempt must be within 3 days from the 1st Attempt

This order has a 1st Attempt within 5 days. However, the 2nd Attempt is 4 days from the 1st Attempt (we exclude Sundays from the calculation). This order is deemed late.

Approach

Here are the main difficulties in this problem statement:

1. Processing of large datasets
2. Handling of epoch time

3. Parsing the address fields to extract the relevant states when calculating SLAs
4. Weekend & Public Holiday exemptions when calculating SLAs
5. 1st attempt deliveries & 2nd attempt deliveries

Perhaps the largest challenge would be the exemptions for Weekend & Public Holiday and applying it to both 1st & 2nd attempts

General approach to solving this:

1. Convert all the epoch time format into something that is easy to read and compute. For python, this can be done easily with " pd.to_datetime". However, do be careful as this will convert the time to GMT timezone. We can change it to GMT+8 by simply adding 8 hours in seconds to the epoch time before applying " pd.to_datetime".

2. Extract all the names of the states from both buyer and seller addresses.
   A simple string match or regex should yield the correct states.

3. Get # of days between pickup and 1st delivery attempt
   We ignore the time and consider only the date. Since we have previously converted our epoch time using " pd.to_datetime ", we can continue using Pandas' datetime functions. We can get the difference between the 2 datetimes and apply "dt.days" to get the number of days. Do the same for 1st delivery attempts and 2nd delivery attempts.

4. Check if any dates should be exempted from the calculation
   One way is to check how many weekends/public holidays exist between the start and end. We'll have to get the dates of all the Sundays and Public Holidays. If any of those dates are between the start and end, then we minus it from the # of days previously calculated in (3). Note, this step might take quite some time since we'll have to apply the same calculation to both pickup -> 1st attempt & 1st attempt -> 2nd attempt so do test it on a sample set first

5. Compare the # of days with the respective SLAs based on the SLA matrix
   If the # of days is more than the SLA, it is deemed late. Both the 1st delivery attempt and 2nd delivery attempt have to be on time, if either is late, then the whole order is considered late.

Tip: You are advised to run your tests on a sample of the dataset first.

Submission Format

Check each delivery order and determine whether it is late.
Two columns required:

- orderid.
- is_late: assign value 1 if the order is late, otherwise 0.

orderid	is_late

1955512445	0
1955598428	1

Your submission should have 3,176,313 rows (excluding headers), each with 2 columns.