Logistics DA Assignment

1) SQL

At Glovo, there is the possibility to bundle two orders from different customers if they both ordered from the same store around the same time. In the case of a bundle, a single courier picks up both orders approximately at the same moment from a store, then proceeds to the first customer address to deliver the first order, and after that delivers the second order to the second customer. In particular, during peak hours, such as dinner time, bundling can increase the capacity of the fleet, leading to lower delivery times on average.

Suppose that you have a table in an SQL database called **orders**, containing data on all customer orders. A snapshot of that table is printed below.

order_id	city_code	store_id	creation_time	pickup_time	enters_delivery	pd_dist	final_status
4596184593	AMS	3372	2021-11-01 23:23:04	2021-11-01 23:33:52	2021-11-01 23:43:17	1503	DeliveredStatus
4569203459	GLV	8844	2021-11-02 11:13:23	NULL	NULL	2004	CanceledStatus
4596020394	GLV	99103	2021-11-01 20:56:01	2021-11-01 21:03:22	2021-11-01 21:11:20	1842	DeliveredStatus
4592303948	PLY	12287	2021-11-01 16:49:18	2021-11-01 16:55:05	2021-11-01 16:55:35	5	DeliveredStatus
4592303949	PLY	12287	2021-11-01 16:50:30	2021-11-01 16:59:45	2021-11-01 17:12:48	1562	DeliveredStatus

In addition, you have a table called **bundled_orders**, which contains information only about orders that were part of a bundle. Find an example of this table below.

order_id	bundle_id	is_bundled	is_unbundled
4395449294	87632847	TRUE	FALSE
4596020394	87632847	TRUE	FALSE
4339452836	87632239	TRUE	TRUE
4592303948	87632239	TRUE	TRUE
4395529454	87633554	TRUE	FALSE

Your goal is to build SQL queries that provide you with the most accurate output to the following questions:

    A. Given the two tables above, you are interested in comparing the percentage of orders that were bundled together in Glovalia (city code GLV) against Playa City (PLY) on November 1st, 2021. Unbundled orders should not be counted as bundled.

    B. Average courier speed from pickup to delivery of each city in the last 30 days. In the case of bundled orders please consider only the trajectory to the first delivery point, always being the one with shortest pickup to delivery distance. The columns **pd_dist**, **pickup_time** and **enters_delivery** represent, respectively: the distance from pickup to delivery; the timestamp when the courier picks up the order and starts heading to the delivery point; the timestamp when the courier enters the delivery area.

Please make your query readable and scalable.

2) Business Case

The purpose of this case study is to test your ability to leverage data to generate useful insights.

Attached you will find [data](#) corresponding to all orders which happened during one week (Delivered and Cancelled) in the city of Glovalia with all important time references.

Note that:

- **Store Address ID** - is the unique id identifying which store the order was for
- **Courier ID** - is the id of the last courier the order was assigned to
- **Vertical**
  - **WALL - Partner** means the order was for a store which is our partner (they receive the orders through the partner web-app before the courier arrives there.
  - **WALL - NonPartner** means the order was for a store which we have as "Fake" (we have no agreements with this store, so the courier needs to go inside and order as a regular customer).
  - **QUIERO** means the order was done through the central "Anything" button on the app where customers can ask us to deliver anything that fits into a Glovo backpack. (operationally similar to the WALL-NonPartner)
- **Transport** is the vehicle type that was used by the courier who was assigned the last to the order

- **Activation** is when the order enters the pool of active orders that need to be assigned to a courier.
- **Last Courier Assigned** is the timestamp of the notification sent to the last courier that was assigned to the order
- **Courier Started Order** is when the courier, upon receiving the order notification, accepts it and starts working on the order by moving towards the pickup location.
- **Courier enters pickup** is the timestamp corresponding to the courier entering the pickup location area (usually 100m radius from the store GPS coordinates)
- **Pickup time**, courier picks the order and heads towards the delivery location
- **Courier enters delivery point**, timestamp when the courier enters the delivery area
- **Termination time**, corresponds to the delivery time (validated by the courier after the customer receives the order) for delivered orders, or the cancellation time for canceled orders.
- **Number of assignments** shows how many times we needed to assign different couriers to an order (e.g. if it is equal to 2, it means that the first courier assigned refused to do the order and we needed to find a second courier, in other words the order was reassigned)
- **Total real distance** stands for the total real distance in KM the courier did to deliver the order

Common Questions

A. Based on the data you were given, provide visibility over the following metrics:
   a. Customer Delivery Time. How long the customer waited for the order.
   b. Delivery Distance. Total distance traveled by the couriers.
   c. Reassignment Probability (whenever we need to assign an order to another courier after the first assignment it counts as a reassignment).
B. Regarding reassignments can you see any trends / correlation between this metric and any of the dimensions in the dataset?
C. Once the courier arrives at the store, the time until the order is picked-up (waiting_time_at_pickup), can present high variation. With the data given, present an analysis showcasing this. Can you identify any clear trends in terms of worst offenders? What could be the possible causes when the courier is waiting for a very long time before picking up the order?

We encourage you to use Python or R and any module/package you see fit to solve the exercise. We expect you to share with us the code used to import, clean and

transform the data as well as the calculations, plots and thought processes that allowed you to reach your conclusions.