Name: Lithiya Letchumy A/P Ravi @ Thiagu
Email: lithiyaletchumy@gmail.com
Subject: Take home assessment by Sunway Group - Data Science/Analytics/Engineering

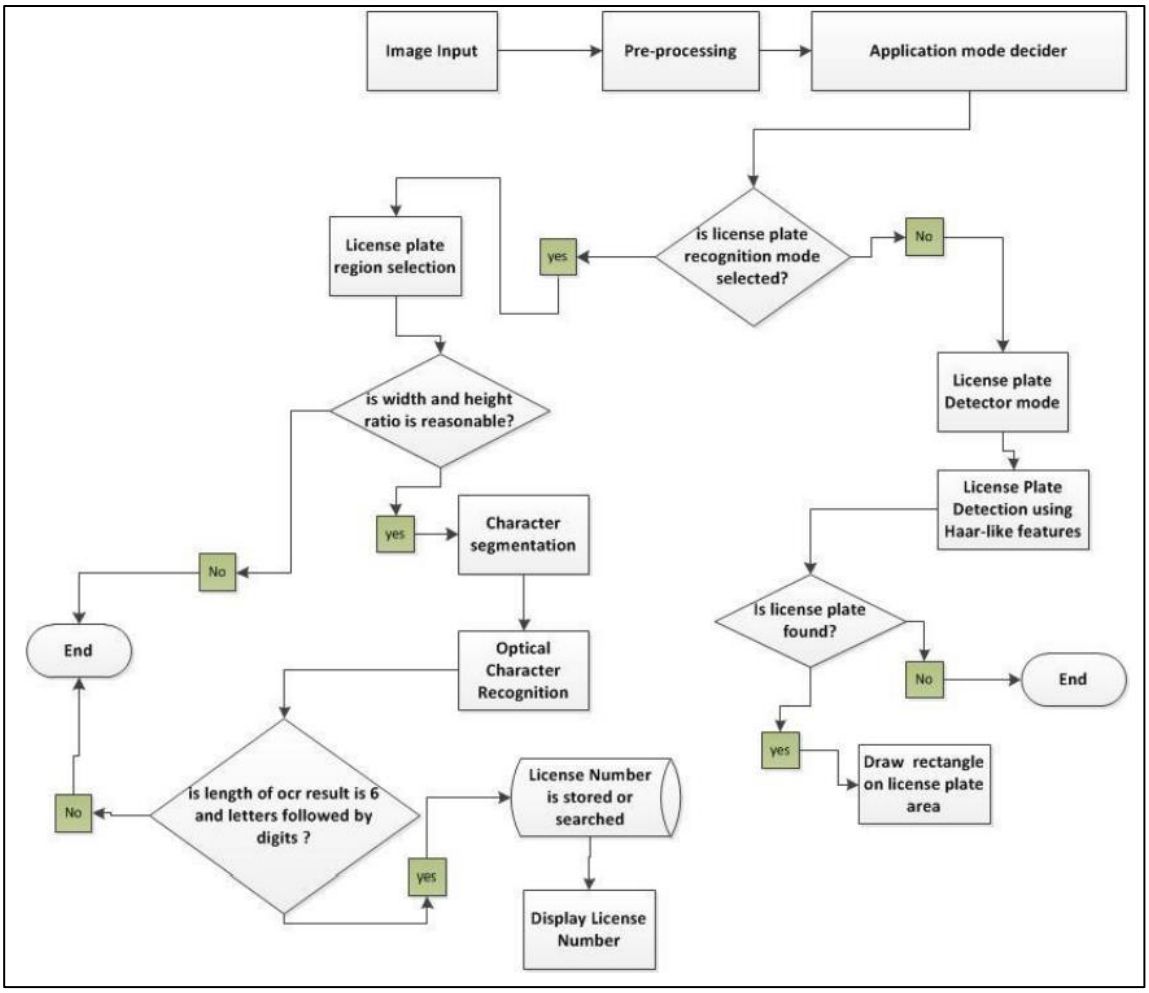
Challenge 1: Data Cleaning, Transformations and ETL pipeline architecture

1. **Find the correlation between both datasets?**
   Both the datasets contain similar customer information such as NAME, CITY, PHONE-NUMBER, EMAIL, and ID.

2. **Explain 5 reasons, why did you join the datasets in that way?**
   - The transaction dataset contains information that's not available in loyalty dataset such as TRANSACTION_ID and AMOUNT.
   - The loyalty dataset contains information that's not available in transaction dataset such as LISENCE_PLATE.
   - The transaction and loyalty datasets were merged to get rid of null values.

3. **Provide a business use case from the datasets provided, that would likely bring benefits to a business.**
   - **License Plate Recognition (LPR)**: Every car has its License Plate, so why not take advantage of this to ensure a safer and more accessible parking experience? First, safety: the license plate serves as a virtual ticket, and grants free-flow access, preventing stolen cars or ticket swapping. Since ticket and license plate must match at entry and exit, it is almost impossible to pursue fraudulent behaviours. With LPR, you can also easily manage blocklists and be in control of all the events and movements at the entrance and exit. LPR is highly reliable: in case of lost tickets, it is easy to determine the entry time from the license plate number. In addition, LPR also allows multiple integrations and customizations, primarily for contract parking, where the recurrent user can smoothly access and egress the car park without any titles other than the vehicle license plate. With LPR, it is possible to enter with a pre-paid ticket or use cashless solutions. Another benefit offered by LPR at entries/exits is the possibility to book a parking space in advance to ensure the driver a hassle-free experience. Lastly, LPR can be retrofitted and added to existing solutions. Now, it is not hard to imagine how LPR can improve parkers' experiences and open a wide range of advantages for car park operators.

4. **Design a cloud/local machine-based pipeline architecture encompassing the business use case provided by you above.**



Challenge 2: Customer engagement

1. **How would you present to a customer regarding the insights derived from their dataset, in your own words?**

   License Plate Recognition System (LPR) offers a seamless parking experience that's cashless, ticketless, and cardless. With this system, drivers will no longer need to stop to collect a ticket or tap their parking pass when entering a car park. As a result, this can help ease congestion at entry and exit points, especially at shopping malls during peak hours. Using cameras powered by AI, it can recognize a vehicle by its number plate. The LPR system has captured and analyzed over a million car plates in Malaysia over the past few years. It claims to have a high accuracy rate above 99%, and it will even work with plates that use non-standard fonts. Each time a car enters the car park, the parking app will issue an e-ticket. When it's time to exit, drivers can pay via the app, which has its eWallet. Alternatively, users can enable an auto-deduct mode where the app automatically deducts the parking fee from your eWallet upon exit. For better security, there's also the option to "lock" in the app so the vehicle can't leave the car park. On top of that, only one vehicle can be tied to one account to prevent cloning. However, you can register up to three vehicles under a single charge. For users that don't have the app installed, drivers can still pay at autopay machines by entering their vehicle number. Unlike Touch' n Go, the app assures no extra surcharge for using the LPR system. Apart from visitors, the LPR system can also be used for season car park users. Users can renew their monthly passes via the app, and this makes it more efficient for car park management companies.

Challenge 3: Scrapping, Data sourcing and Enriching data

1. **What is data enrichment?**

   Data enrichment is the process of enhancing existing information by supplementing missing or incomplete data. Data enrichment is achieved by appending one or more data sets with other attributes and values from different data collections. This standardizes all the merged data to fit our data's quality, which results in refined, improved, and enhanced data grouping. For example, when customer information data from sales calls are combined with the data from the internet, you enrich data. Enriched data gives us more knowledge regarding our business and customers, thus offering us more ammunition to improve our brand's presence.

2. Come up with 1 new use-cases with the core dataset as above by enriching with public data.

To look beyond a license plate, License Plate Recognition (LPR) results can be enhanced with third-party data and contextual information. In computer vision systems, the distinction can be made between internal and external enrichments. Internal enrichments originate from the system by visual interpretation of an image or video. Direction, speed of travel, or origin of the license plate are clear examples of internal processes. External enrichments find their origin in third-party databases. Examples of third-party results that can be used as enriched data are:

a) **Vehicle registration authorities**: All countries have a vehicle registration authority, and, in many countries, governments make vehicle data publicly available. Recognized license plates are linked to corresponding make, model, year of build, and color by a database lookup. This data is immensely informative and important regarding privacy and anonymity.

b) **Vehicle classification**: The distinction between trucks, motors, cars, bicycles, and pedestrians on top of LPR enables refined access or crowd management control. A truck can be granted access to the perimeter of a logistic centre, while an arriving car will not be allowed. Meanwhile, the staff receives a message to announce the arrival of the truck and the details of the load. A matrix board leads the truck driver to a specified docking station. While reading license plates, the ability to detect people, vehicles and other objects provides a complete overview of traffic flow in a tunnel, on a bridge, city landscape, or business park.